The nucleotide sequence of poliovirus type 3 Leon 12 a,b: comparison with poliovirus type 1

Glyn Stanway, Alan J Cann, Rudolf Hauptmann*, Pamela Hughes, Lindsay D Clarke, Roger C Mountford, Philip D Minor, Geoffrey C Schild and Jeffrey W Almond

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

Received 5 July 1983; Accepted 19 July 1983

ABSTRACT
The complete nucleotide sequence of the genome of the Sabin vaccine strain of poliovirus type 3 (P3/Leon 12 a,b) has been determined from cDNA cloned in E.coli. The genome comprises a 5' non-coding region of 742 nucleotides, a large open reading frame of 6618 nucleotides (89% of the sequence) and a 3' non-coding region of 72 nucleotides. There is 77.4% base-sequence homology and 89.6% predicted amino-acid homology between types 1 and 3. Conservation of all glutamine-glycine and tyrosine-glycine cleavage sites suggests a mechanism of polyprotein processing similar to that established for poliovirus type 1.

INTRODUCTION
Polioviruses, the causative agents of paralytic poliomyelitis, are members of the enterovirus group of the family picornaviridae [1]. The three distinct serotypes, 1, 2 and 3, cause identical disease and are closely similar in structure and composition. In each case the 27nm diameter icosahedral capsid consists of 60 copies each of four virus polypeptides, VP1-VP4, surrounding a single-stranded, positive-sense RNA genome of approximately 7,500 nucleotides [2]. The RNA contains a small virus-coded protein, VPg, covalently linked to its 5'-terminus [3,4] and a poly-A tract at its 3'-terminus [5]. The genome is most probably expressed as a single translational unit encoding a large precursor polyprotein which undergoes specific cleavage to give rise to recognized structural and non-structural virus polypeptides [6,7].

Over the past twenty years epidemic poliomyelitis has been almost eliminated from developed countries, mainly through the use of vaccines. The most widely adopted of these are based on
the live-attenuated strains developed by Dr. Albert Sabin in the 1950's [8]. Although these vaccines have been very successful, a low incidence of disease persists in vaccinated communities and there is evidence that instability of the type 3 vaccine may be a contributing factor [9]. Gene cloning and nucleic-acid sequencing studies offer a better understanding of the molecular basis of attenuation of live-vaccines and may suggest ways in which they can be improved.

We have previously described the molecular cloning of poliovirus type 3 using a method based on RNA-cDNA hybrids [10]. This paper describes the complete nucleotide sequence of the type 3 Sabin vaccine, P3/Leon 12 a,b, derived from these clones. The nucleotide sequence is compared with those previously published for type 1 [7,11,12].

MATERIALS AND METHODS

Virus
A plaque-purified derivative of the P3/Leon 12 a,b Sabin vaccine strain, isolate #411, was used in these experiments [13,14]. Details of the molecular cloning of this virus have been published previously [10].

Nucleotide sequence analysis
Overlapping cDNAs contained in plasmid vector pAT 153 were sub-cloned into bacteriophage M13 mp7, mp8 and mp9, prior to sequence determination [15]. The cDNAs were purified from vector DNA by digestion with Pst I and electrophoresis on agarose gels. After electrophoretic elution from the gel, the cDNAs were digested with Alu I, Hae III or Sau 3A to provide overlapping fragments for sub-cloning [10].

Nucleotide sequence was obtained using the dideoxynucleotide chain-termination method [16] and the data was collated and assembled with the aid of a Digital PDP 11/44 computer using programs developed by Staden [17].

Other methods
Restriction endonuclease digestions were carried out according to manufacturers instructions. Ligations, restriction mapping and preparation of DNA was as described previously [10].
RESULTS AND DISCUSSION

Overlapping cDNA clones, together spanning the entire genome of poliovirus P3/Leon 12 a₁b, were characterized by restriction endonuclease mapping, using the method of Smith and Birnsteil [18]. The results obtained (not shown) determined our strategy for sub-cloning into bacteriophage M13 prior to nucleotide sequencing. This detailed restriction map, together with the close homology to poliovirus type 1 in nucleotide and/or predicted amino-acid sequence over most of the genome, allowed the relative location of individual gel readings to be discerned. Where corresponding sequence data were obtained from more than one cDNA clone, perfect matching was observed, providing further evidence of the high fidelity of the RNA-cDNA hybrid cloning method [10].

The total sequence of 7432 nucleotides of the P3/Leon 12 a₁b genome derived from cloned cDNA is presented in figure 1. We are confident that the sequence presented is accurate. All regions were covered by at least two independent sequencing reaction and comparison with published poliovirus type 1 sequence [11] provided a useful check on reading-frame errors. Furthermore, the sequence has been verified by comparison with cDNAs from closely related type 3 strains of almost identical sequence [14].

The calculated molecular weight of the genome is 2.66 \times 10^6 and the proportions of the four nucleotides are A=28.9\%, G=23.5\%, C=23.4\%, U=24.1\%. Computer-assisted analysis of the sequence in both polarities reveals just one open reading-frame large enough to encode the known virus proteins [19]. This spans 2206 consecutive codons starting at position 743 and ending with two stop codons in tandem at position 7361.

Homology with poliovirus type 1

The complete nucleotide sequences of two strains of poliovirus type 1 have been determined in three independent laboratories [7,11,12]. These data, together with analyses of infected-cell and virus-structural polypeptides including \textit{N} and C-terminal sequencing, has allowed a detailed genetic map of poliovirus type 1 to be established [6,7,20,21,22]. This information has aided greatly the interpretation of the poliovirus type 3 sequence presented here, since, based on homology, the two
sequences can be aligned almost perfectly. The overall homology in nucleotide sequence between the two types is 77.4%, higher than some estimates [23] but in agreement with the recent results obtained using a method based on cDNA hybridization [24]. The size of the 5' and 3' non-coding regions are identical to type 1 at 742 and 72 bases respectively, but the coding region of type 3 contains three fewer triplets than that of type 1.

Non-coding regions
The sequence of the 5'-terminus of the cDNA presented in fig. 1 is identical to the previously published sequence of 20 nucleotides determined directly from genomic RNA of poliovirus type 3 [25]. This sequence, plus 15 succeeding nucleotides up to a Hae III recognition site, has been determined from five independent cDNA clones and is in each case immediately preceded by a homopolymer tract of about 20 dG residues. This confirms that the 5'-sequence of the cDNA depicted in fig. 1 actually corresponds to the 5'-terminal sequence of poliovirus type 3 genomic RNA, and that the RNA-cDNA hybrid method permits direct cloning of the 5'-terminus [10].

The 5' non-coding region shows close homology to poliovirus type 1. There is only one difference in the first 54 nucleotides (A at position 22 in type 1). This difference forms part of the loop in an otherwise totally conserved stable hairpin structure previously recognized for type 1 and postulated to play a role in ribosomal recognition or in polymerase binding [26]. From position 54 until the start of the large open reading-frame at position 743, the overall homology is approximately 80%. There is a notable stretch of 65 identical bases from 509-573, the significance of which remains obscure. Although it appears that several insertions and deletions have occurred in this 5' region, the size, 742 nucleotides, is identical in the two types. Before the start of the 2206 codon reading-frame at

Figure 1. Complete Nucleotide sequence of the genome of poliovirus type 3 Leon 12 a_1 b as determined from cloned cDNA, plus the predicted amino-acid sequence from the major open reading frame. Cleavage sites analagous in location to those used in poliovirus type 1 are indicated.
Table 1. Codon usage in the long open-reading frame of poliovirus type 3. STOP = termination codons.

<table>
<thead>
<tr>
<th>2nd</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>U</td>
<td>F</td>
<td>40</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>38</td>
<td>66</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>STOP</td>
<td>0</td>
<td>STOP</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>7</td>
<td>0</td>
<td>W</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>L</td>
<td>17</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>21</td>
<td>31</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>59</td>
<td>Q</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>16</td>
<td>37</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>I</td>
<td>65</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>43</td>
<td>61</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>47</td>
<td>K</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>10</td>
<td>55</td>
<td>18</td>
</tr>
<tr>
<td>G</td>
<td>V</td>
<td>28</td>
<td>A</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>44</td>
<td>60</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>59</td>
<td>E</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>16</td>
<td>64</td>
<td>38</td>
</tr>
</tbody>
</table>

position 743, there are seven other AUG codons which could be used to initiate translation. Most of these establish reading-frames which are not conserved free of nonsense-codons between poliovirus types 1 and 3 and are therefore of doubtful significance. There is one conserved potential reading-frame located at 239-322 which could code for a peptide of 28 amino-acids. However, there is at present no evidence that this region is translated in poliovirus-infected cells.

The non-coding region at the 3'-terminus of the genome is identical in sequence to that determined for poliovirus type 1 Mahoney as has been discussed previously [10]. Although the conservation of this region implies that its primary structure is functionally important, it is noteworthy that other enteroviruses show substantial divergence in this region, [27, our unpublished work].

Coding region
The open reading-frame from nucleotide 743 to 7360 comprises 2206 consecutive codons and is the only region of the genome which can encode the known virus polypeptides [19]. This single reading-frame probably represents the entire coding function of
Figure 2. Diagrammatic representation of the amino-acid sequence differences between poliovirus type 3 Leon 12 a b and poliovirus type 1 (Mahoney strain). A single vertical line represents a single amino-acid substitution, insertion or deletion. Lines of double thickness represent two adjacent changes, etc. The map of the genome is based on that determined for poliovirus type 1 [7].

The AUG at position 743 is the eighth such codon from the 5'-terminus, so presumably a mechanism exists which allows ribosomes to bypass the preceding seven. It is noteworthy that poliovirus mRNA lacks a 5'-terminal cap [28,29] and therefore its translation is likely to be initiated by a mechanism different from that used by cellular mRNA [30].

Codon usage is illustrated in table 1 and shows only minor differences from that of poliovirus type 1 [11]. The overall homology in predicted amino-acid sequence is 89.6% although there is substantial variation around this figure between individual virus proteins. Figure 2 illustrates the distribution of amino-acid sequence differences between types 1 and 3. Most are the result of point mutations but there are also several insertions and/or deletions of whole triplets. In general non-structural proteins are more highly conserved than structural proteins. This may be due to a greater selection pressure for structural variants of the virus particle e.g. immunological, or may simply reflect a lower sequence flexibility in non-structural proteins whose role is mainly enzymatic.

The largest region showing no homology to type 1 spans...
nucleotides 2480-2542 encoding the N-terminus of VP1 and possibly the C-terminus of VP3. This sequence contains two glutamine-glycine potential proteolytic-cleavage sites at positions 2477 and 2509, either or both of which may be used to generate VP1 and the VP0-VP3 precursor [7]. Amino-acid sequence analysis of the C-terminus of VP3 and the N-terminus of VP1 is in progress to resolve this question.

Another region of substantial sequence divergence includes the major antigenic site against which neutralizing antibodies are directed (encoded by nucleotides 2754-2776) [31]. We have previously shown that this sequence of eight amino-acids in the VP1 protein plays a critical role in the neutralization of poliovirus type 3 [32]. Because of the similarity in sequence over most of VP1 between the serotypes, it is likely that the eight analogous amino-acids from poliovirus type 1 constitute its equivalent antigenic site. Five differences out of the eight amino-acids comprising this region is a level of homology consistent with the serological distinctness of the two types. The relative similarity in sequence arrangement however, (also for type 2, A. Nomoto personal communication), suggests that the structure of this site may be less flexible than, for example, the equivalent region in the VP1 protein of foot-and-mouth disease virus, where more drastic variations probably account for the greater number of serotypes of this virus [33].

As discussed previously [10], the putative protease (P3-7c) [34] encoded by nucleotides 5428-5944 is highly conserved between types 1 and 3 (98% amino-acid homology), suggesting an identical function. Conservation of all glutamine-glycine pairs including those shown to be involved in processing of the type 1 polyprotein [7] (fig. 3) lends further support to this conclusion. Similarly all tyrosine-glycine pairs and the asparagine-serine cleavage site between VP4-VP2 are conserved, suggesting that host involvement in processing is identical in the two types [34].

Concluding remarks

The complete nucleotide sequence presented here is the first to be derived from a poliovirus type 3 strain. Knowledge of this sequence will be helpful in many respects. Sequence
Figure 3. Location of potential proteolytic-cleavage sites in the polyprotein encoded by poliovirus type 3 Leon 12 a1b. Arrows indicate those sites used in poliovirus type 1 [7]. Proteins indicated are those whose N-termini are generated by the cleavages shown.

▲ = tyrosine-glycine
▲ = glutamine-glycine
〇 = asparagine-serine (only that known to be used, VP2-VP4, is indicated).
? = potential alternative site for the generation of VP1, (see text).

Information is an essential prerequisite for elucidation of the molecular basis of the biological features of the virus including neurovirulence, antigenicity and tissue-tropism. The possibility of recovering live poliovirus from cloned cDNA has opened up new possibilities for genetic studies on these viruses [35]. The current live-attenuated type 3 poliovirus vaccine has certain disadvantages, particularly with respect to reversion of the virus to virulence on passage in tissue-culture [36] and in the occurrence of vaccine-associated disease [9, 37, 38]. Manipulation of the genome using techniques such as site-directed mutagenesis, recombination in vitro and the insertion or deletion of genetic material, may lead to the production of safer vaccines. Work in this area is currently in progress.

Acknowledgements
We would like to thank Dr A. Nomoto for communicating unpublished results. This work was supported by the Medical Research Council of Great Britain (Project grant No. G80/0815/2CA).
REFERENCES
