The Nucleotide Sequence of a Type 3 Poliovirus Isolated During a Recent Outbreak of Poliomyelitis in Finland

By Pamela J. Hughes, David M. A. Evans, Philip D. Minor, Geoffrey C. Schild, Jeffrey W. Almond and Glyn Stanway

Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, National Institute for Biological Standards and Control, Holly Hill, London NW3 6RB and Department of Microbiology, University of Reading, London Road, Reading RG1 5AQ, U.K.

(Accepted 12 June 1986)

SUMMARY

We have cloned and sequenced the complete genome of a strain of poliovirus type 3 (23127) isolated during an outbreak of poliomyelitis in Finland. The genome is 7435 nucleotides long excluding the 3' poly(A) stretch and is 95.5% homologous at the amino acid level to the previously sequenced type 3 strain, P3/Leon/37. The most striking feature of the presented sequence is the extent of amino acid substitutions relative to P3/Leon/37 and other type 3 strains in areas of known antigenic importance. The major antigenic determinant for virus neutralization (site 1), located at residues 89 to 100 of VP1, has three amino acid substitutions and there are six substitutions in site 3, a composite site made up of sequences from VP1 and VP3. The variation in these regions probably accounts for the observed unusual antigenicity and may explain why the virus was able to spread in a well-vaccinated community. Sequence comparisons imply that the virus is not derived from the currently used live attenuated vaccine.

INTRODUCTION

Paralytic poliomyelitis has been largely eliminated from developed countries through the use of effective vaccines against the causative agents, the three poliovirus serotypes. However, in late 1984 there was an outbreak of the disease in Finland, a country where inactivated poliovirus vaccines have been used successfully for about 25 years and where the vaccine acceptance rate is in excess of 97% (Weekly Epidemiological Record, 1985; Lapinlemu & Stenvik, 1981). Nine paralytic cases occurred, each associated with infection by a poliovirus designated type 3. However, further serological characterization revealed that the viruses isolated during the outbreak were antigenically unusual in their response to both polyclonal and monoclonal antisera raised against the other type 3 strains. This may have contributed to the ability to spread in a well-vaccinated community (Magrath et al., 1986). The antigenic properties were partly explained by mutations detected in a region of the poliovirus capsid known to play a major role in virus neutralization (Evans et al., 1983). This finding is surprising because wild strains of poliovirus are usually highly conserved in this region (Ferguson et al., 1985).

Poliovirus is a member of the picornaviridae family and has a positive-sense, single-stranded RNA genome of approximately 7500 nucleotides, enclosed within a protein capsid made up of 60 copies of each of four proteins, VP1 to VP4. The 5'-most 740 nucleotides of the genome are presumed to be non-translated and an AUG located after this region initiates the synthesis of a long polyprotein which is cleaved by virus proteases to generate all the known virus proteins. As part of our research on poliovirus antigenicity and the molecular relationship between picornaviruses, we have determined the complete nucleotide sequence of a strain of the virus responsible for the Finnish outbreak. The sequence presented here shows that the regions of the poliovirus capsid known to be antigenically important exhibit marked differences from the
previously sequenced poliovirus type 3 strains and this is consistent with the unusual antigenicity. In addition, the sequence gives an indication of the degree of molecular heterogeneity within a poliovirus serotype.

METHODS

Virus strain. The virus under study (strain 23127) was isolated during the outbreak from an asymptomatic individual who was a close relative of a victim of paralytic poliomyelitis.

Molecular cloning and nucleotide sequencing. Purified viral RNA (5 μg) was reverse-transcribed and cloned into Escherichia coli JA221 by the cDNA : RNA hybrid method (Cann et al., 1983; Stanway et al., 1984a) using pBR322 as the vector. Of the recombinants obtained, 800 were analysed by hybridization using portions of the previously cloned poliovirus type 3 genome as probes (Cann et al., 1983). A set of five overlapping subgenomic clones were selected which together spanned the genome (data not shown). The cDNA inserts were sequenced by the dideoxynucleotide method after generation of random fragments and cloning into M13 mp8 as previously described (Stanway et al., 1984b). This methodology generated the majority of the sequence data while residual, small stretches were sequenced after cloning AluI or Sau3AI fragments into M13 mp8 digested with SmaI or BamH1 respectively. Approximately 80% of the sequence was obtained in both orientations and all regions were sequenced at least twice. Sequence data were assembled and analysed using published computer programs (Staden, 1980).

RESULTS AND DISCUSSION

The complete nucleotide sequence and predicted amino acid sequence of strain 23127 are shown in Fig. 1. The genome comprises a 5' non-coding region of 746 nucleotides, a single open reading frame of 6618 nucleotides (2206 codons), a short 3' non-coding region of 71 nucleotides and a poly(A) tract. Sequence comparisons indicate that strain 23127 is highly homologous to the representatives of each of the three poliovirus serotypes sequenced to date. These are the type 1 strain P1/Mahoney (Kitamura et al., 1981) and type 3 strain P3/Leon/37 (Stanway et al., 1984b), which are wild-type strains, and the Sabin type 2 vaccine strain P2/P712, Ch, 2ab (Toyoda et al., 1984). The amino acid homologies between the individual proteins of strain 23127 and these viruses are shown in Table 1. Overall, in confirmation of the serotyping, 23127 most closely resembles P3/Leon/37. This is particularly evident in the capsid proteins where the level of homology ranges from 91.7% in VP1 to 100% in VP4. The corresponding figures for the VP1 proteins of the type 1 and type 2 strains are only 71.3% and 73.7% respectively. Taking the genome as a whole there is 95.5% amino acid homology and 80.7% nucleotide homology to P3/Leon/37. Given the immunological relationship between strain 23127 and P3/Leon/37, the degree of nucleotide sequence divergence is quite high. This is consistent with the generally accepted view of the high mutability of an RNA genome (for review, see Domingo et al., 1985).

Although strain 23127 is more closely related to P3/Leon/37 than it is to the type 1 and type 2 strains, the pattern of homology is not that which might be expected from previous comparisons between poliovirus strains. In particular, the 5' untranslated region and the non-structural proteins show an unexpectedly high degree of difference to P3/Leon/37. A comparison between the 5' untranslated regions of strain 23127 and the three poliovirus serotypes is presented diagrammatically in Fig. 2 and the degree of homology is shown in Table 2. This region is often highly conserved among related picornaviruses (Stanway et al., 1983, 1984c; Toyoda et al., 1984) and this is also true for strain 23127 and the other polioviruses. The conservation is of the order of 85% in each case. The 100 nucleotides prior to the AUG which initiates the long open reading frame are highly variable between poliovirus serotypes and if this region is ignored the sequence conservation is around 90%. In contrast to the results seen for the capsid proteins, the 5' non-coding region of strain 23127 is not obviously more similar to the type 3 strain in terms of overall homology. Indeed, it is slightly more similar to P1/Mahoney (Table 2). This is due largely to two particular stretches of nucleotides (positions 122 to 187 and 473 to 508) where strain 23127 is strikingly similar to P1/Mahoney. In the former case there is perfect conservation of 66 nucleotides between these viruses while P3/Leon/37 differs at 13 positions. There is only one difference to P2/P712, Ch, 2ab in the same area. In the stretch of nucleotides from positions 473 to 508 there are only three differences between strain 23127 and P1/Mahoney while there are 14 differences between the type 3 strains. P2/P712, Ch, 2ab differs at 11 positions in this area. Strain 23127 is most similar to P3/Leon/37 outside these regions (Table 2).
Table 1. Percentage amino acid homology between the proteins of strain 23127 and previously sequenced polioviruses

<table>
<thead>
<tr>
<th>Protein</th>
<th>Strain 23127</th>
<th>P3/Leon/37</th>
<th>P1/Mahoney</th>
<th>P2/P712, Ch, 2ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP4</td>
<td>92.7</td>
<td>100</td>
<td>92.7</td>
<td></td>
</tr>
<tr>
<td>VP2</td>
<td>83.0</td>
<td>95.2</td>
<td>80.0</td>
<td>95.2</td>
</tr>
<tr>
<td>VP3</td>
<td>82.8</td>
<td>95.0</td>
<td>84.9</td>
<td>95.0</td>
</tr>
<tr>
<td>VP1</td>
<td>71.3</td>
<td>91.7</td>
<td>73.7</td>
<td>91.7</td>
</tr>
<tr>
<td>P2-A</td>
<td>94.0</td>
<td>93.3</td>
<td>93.3</td>
<td>93.3</td>
</tr>
<tr>
<td>P2-B</td>
<td>84.5</td>
<td>91.8</td>
<td>88.7</td>
<td>91.8</td>
</tr>
<tr>
<td>P2-C</td>
<td>96.0</td>
<td>97.3</td>
<td>96.4</td>
<td>97.3</td>
</tr>
<tr>
<td>P3-A</td>
<td>97.7</td>
<td>98.9</td>
<td>97.7</td>
<td>98.9</td>
</tr>
<tr>
<td>VPg</td>
<td>95.7</td>
<td>95.7</td>
<td>100</td>
<td>95.7</td>
</tr>
<tr>
<td>Protease</td>
<td>96.7</td>
<td>95.6</td>
<td>96.2</td>
<td></td>
</tr>
<tr>
<td>Polymerase</td>
<td>98.3</td>
<td>97.4</td>
<td>97.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Percentage nucleotide homology between strain 23127 and other polioviruses in the non-coding regions

<table>
<thead>
<tr>
<th>Region</th>
<th>Strain 23127</th>
<th>P3/Leon/37</th>
<th>P1/Mahoney</th>
<th>P2/P712, Ch, 2ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall 5' non-coding</td>
<td>85.9</td>
<td>84.0</td>
<td>85.9</td>
<td>84.0</td>
</tr>
<tr>
<td>Nucleotides 1–640</td>
<td>90.8</td>
<td>89.8</td>
<td>90.8</td>
<td>89.8</td>
</tr>
<tr>
<td>Excluding nucleotides</td>
<td>89.6</td>
<td>91.5</td>
<td>89.6</td>
<td>91.5</td>
</tr>
<tr>
<td>122–187 and 473–508</td>
<td>98.6</td>
<td>97.2</td>
<td>98.6</td>
<td>97.2</td>
</tr>
</tbody>
</table>

In view of the demonstration that polioviruses can undergo recombination in man (Kew & Nottay, 1985; Minor et al., 1986a), a possible explanation for this pattern of homology is that strain 23127 is a recombinant in which the 5' untranslated region has been derived from a type 1 strain. This is, however, unlikely given that the type 1-specific regions are flanked by regions where strain 23127 is more homologous to P3/Leon/37. Thus, several recombination events would need to have taken place to generate the final virus. An alternative explanation is that the Finnish strain is closer to the poliovirus type 3 progenitor strain than is P3/Leon/37 and that the latter strain has accumulated several additional mutations in these areas since its divergence. The obvious implication of this is that strain 23127 is not on a direct genetic lineage from P3/Leon/37. Since the type 3 vaccine strain used throughout much of the western world was produced from P3/Leon/37 and differs at only 10 nucleotide positions (Stanway et al., 1984b), these results argue that the virus is not vaccine-derived. It is presumably a wild strain which has circulated independently.

In contrast to the 5' non-coding region, the 3' non-coding region of strain 23127 sheds little light on the origin of the strain and its relationship to the other polioviruses. This region is highly homologous between the other sequenced polioviruses and this is also observed with strain 23127. There is just one nucleotide difference from the type 1 and type 3 strains and two from the type 2 strain. The open reading frame is terminated by two stop codons in tandem in each of the polioviruses sequenced to date. Interestingly, the single nucleotide substitution in the 3' non-coding region of strain 23127 destroys the second of these stop codons, indicating that this is not required for the viability of the virus.

In addition to the 5' non-coding region, the other seemingly anomalous aspect of the presented sequence is the degree of homology in the non-structural proteins. These are highly conserved between the poliovirus type 1 to 3 strains sequenced to date, as might be expected of proteins which perform the same function remote from the immunological pressures upon the capsid proteins. The capsid proteins of strain 23127 are most similar to P3/Leon/37 and presumably in evolutionary terms these two viruses are closely related. It might therefore be expected that their non-structural proteins would be particularly similar. However, this is not the case since strain 23127 shows approximately the same number of amino acid differences from each of the three
Sequence of a poliovirus Finland strain
Sequence of a poliovirus Finland strain

Fig. 1. Complete nucleotide sequence and predicted amino acid sequence of strain 23127, a type 3 poliovirus isolated during the recent outbreak of poliomyelitis in Finland. Amino acid differences from the previously sequenced type 3 strain, P3/Leon/37, are circled.
the 49 amino acid differences in the non-structural region, at 17 positions the Finnish strain has the same sequence as both the type 1 and the type 2 strains. This seems to imply that at these positions mutations have occurred in the P3/Leon/37 lineage since its divergence from the lineage which eventually produced strain 23127. This again suggests that strain 23127 is not descended from P3/Leon/37 or from the vaccine strain P3/Leon 12a1b.

In the capsid proteins, the relationship between strain 23127 and the previously sequenced strains of poliovirus is clearer than in other areas of the genome. Here, there are 50 amino acid differences from P3/Leon/37, which is only about one-third of the differences from P1/Mahoney and P2/P712, Ch, 2a,b and the amino acid homologies are much greater within each of the individual proteins (Table 1). In addition, there is a very clear relationship to P3/Leon/37 in regions where there is hypervariability between the three previously sequenced strains such as the antigenic sites and the N terminus of VP1. This accounts for the serotyping of 23127 and the other Finnish strains as type 3 polioviruses. Variations between the type 3 strains seem to be scattered throughout the capsid proteins, with the exception of VP4 where there are no amino acid differences. This may be due to the fact that VP4 is apparently not exposed on the surface of the virus particle and is therefore not subject to immunological pressure (Hogle et al., 1985). When the differences were plotted on the recently derived three-dimensional structure of P1/Mahoney (Hogle et al., 1985), the great majority were found to be in loops and other elaborations outside the β-barrel core structure of the three major capsid proteins, VP1 to VP3 (data not shown). This is consistent with the β-barrel structure being of central importance in determining the morphology of the capsid. Another feature of the sequence comparison with P3/Leon/37 is the concentration of amino acid differences in the N-terminal region of VP1. This part of the protein lies within the interior of the virus particle and therefore cannot be involved in the antigenicity of the virus. However, it is highly variable between the known polioviruses and is possibly flexible in terms of conformation since its structure is not observable by X-ray crystallography (Hogle et al., 1985). This possibly explains why a number of amino acid differences have been able to accumulate between the two type 3 strains.
Table 3. Amino acid sequences of P3/Leon/37 and strain 23127 in regions of antigenic importance

<table>
<thead>
<tr>
<th>Location</th>
<th>Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1: 89-100 P3/Leon/37 (EVDNEQP_TTRAQ) 23127 (EVDNEQPATNYQ)</td>
<td></td>
</tr>
<tr>
<td>Site 2a: 217-223 P3/Leon/37 TDA(NDOI) 23127 SDA(NDOY)</td>
<td></td>
</tr>
<tr>
<td>Site 2b: 164-172 P3/Leon/37 (_~AVTSPKRE) 23127 (.T_AVTSPK RE)</td>
<td></td>
</tr>
<tr>
<td>Site 3a: 286-290 P3/Leon/37 (KNNLD) 23127 (KDGLA)</td>
<td></td>
</tr>
</tbody>
</table>

*The previously identified antigenic sites (Minor et al., 1986b) are enclosed in parentheses.

Extensive analysis of the antigenicity of polioviruses in the past has revealed the presence of several regions important in eliciting neutralizing antibodies (Evans et al., 1983; Minor et al., 1986b). The Finnish strain shows amino acid sequence differences from P3/Leon/37 in many of these regions (Table 3). As already observed (Magrath et al., 1986), the major antigenic site for poliovirus neutralization, termed site 1 and located at positions 89 to 100 in VP1, contains three amino acid substitutions in strain 23127. Our results show that there are also drastic differences in a composite antigenic site (site 3; Minor et al., 1986b), made up of amino acids 286 to 290 of VP1 and several amino acids from VP3 (positions 58 to 60, 70, 71, 77, 79). Strain 23127 shows six amino acid substitutions relative to P3/Leon/37 in this antigenic site. These are at positions 286, 287 and 289 of VP1 and positions 59, 71 and 77 of VP3. In addition, position 80 of VP3 is different. While this residue has not previously been included in the antigenic site, its close proximity suggests that a mutation here may also affect antigenicity. Overall, more than half the residues known to be involved in this antigenic site are different from those of P3/Leon/37. It would therefore be expected that this site would not be recognized by antibodies directed against the corresponding site in P3/Leon/37. In type 3 poliovirus, this antigenic region is far less immunodominant than site 1 (Minor et al., 1986b) but these differences must contribute, to some extent, to the distinctive overall antigenicity of strain 23127. Interestingly, two of the different amino acids (positions 286 and 287 of VP1) are identical to the corresponding amino acids of P1/Mahoney and P2/P712, Ch, 2ab. It is possible that this is a similar situation to that of the homology in the non-coding region described above which suggested that strain 23127 is closer to a type 3 progenitor than is P3/Leon/37. The observation adds further weight to the idea that strain 23127 is not on a direct lineage from P3/Leon/37 and that it is not vaccine-derived. The homologous two amino acids are followed by an amino acid insertion in P1/Mahoney relative to both of the type 3 strains and to P2/P712, Ch, 2ab. The insertion event presumably occurred after the divergence of the poliovirus serotypes. The other antigenic site identified on the poliovirus capsid, site 2, is also composite and is made up of amino acids 226 to 290 of VP1 and, in poliovirus type 3, residues 164 to 172 of VP2 (Minor et al., 1986b). Surprisingly, these regions are highly conserved in strain 23127, the single amino acid difference from P3/Leon/37 being located at position 164 of VP2. Although the VP1 component of this site is identical in the two type 3 strains, two changes are seen flanking the region (at positions 217 and 223) and these may affect the antigenicity. However, substitutions in this antigenic site would seem to play less of a role in the overall antigenic differences than substitutions in site 1 and site 3. Taken together, these amino acid substitutions in the three areas of known antigenic importance are consistent with the unusual antigenicity of the virus and probably contributed to its ability to spread in a well-vaccinated community.

Finally, there is some suggestion that the virus drifted antigenically during the outbreak since some of the viruses isolated show only one or two amino acid differences to P3/Leon/37 in the region of the major antigenic site, rather than the three seen in strain 23127 (Magrath et al., 1986). It would be interesting to analyse antigenic sites 2 and 3 of these isolates, to monitor the
extent of drift there. If such drift has occurred over the relatively short period of the outbreak it
would indicate that the poliovirus capsid is able to accommodate relatively readily the amino
acid differences required. It is thus surprising that to date only three poliovirus serotypes have
been identified and that wild strains are usually conserved antigenically. It remains to be seen
whether the circulating viruses will continue to diverge and whether ultimately a new poliovirus
serotype will emerge. This would have important implications for the future requirements for
poliovirus vaccines.

This work was supported by the Medical Research Council, grant number G83242556CB.

REFERENCES

encoding the protease and polymerase proteins. Nucleic Acids Research 11, 1267–1281.

DOMINGO, E., MARTINEZ-SALAS, E., SOBRINO, F., DE LA TORRE, J. C., PORTADA, A., ORTIN, J., LOPEZ-GALINDEZ, C.,
The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance –


broadly reactive, type specific neutralizing antibody to poliovirus type 3 by synthetic peptides. Virology 143,
505–515.

Science 229, 1358–1365.

KEW, O. M. & NOTTAY, B. K. (1985). Evolution of the oral polio vaccine strains in humans occurs by both mutation
and intermolecular recombination. In Modern Approaches to Vaccines: Molecular and Chemical Basis of Virus
Spring Harbor Laboratory.


Developments in Biological Standardization 47, 241–246.

HOVI, T. (1986). Antigenic and molecular properties of type 3 poliovirus responsible for an outbreak of


Acids Research 8, 3673–3694.


of the genomes of poliovirus type 3 strains by the cDNA : RNA hybrid method. Archives of Virology 81, 67–78.

Comparison of the complete nucleotide sequences of the genomes of the neurovirulent poliovirus P3/Leon/37
and its attenuated Sabin vaccine derivative P3 Leon 12a, b: Proceedings of the National Academy of Sciences,
U.S.A. 81, 1539–1543.


nucleotide sequences of all three poliovirus serotype genomes: implication for genetic relationship, gene


(Received 18 April 1986)