Genetic divergence of poliovirus strains isolated in the Karachi region of Pakistan

Anita Huovilainen, Mick N. Mulders, Mubina Agboatwalla, Tuija Pöyry, Mirja Stenvik and Tapani Hovi

Enterovirus Laboratory, National Public Health Institute, Mannerheimintie 166, 00300 Helsinki, Finland, Laboratory of Virology, National Institute of Public Health and Environmental Protection, WHO Collaborating Centre for Reference and Research on Poliomyelitis (RIVM), Bilthoven, The Netherlands and Department of Paediatrics II, Civil Hospital, Karachi, Pakistan

Seventy-seven wild poliovirus strains isolated from poliomyelitis cases in the Civil Hospital of Karachi in Pakistan in 1989–1993 were selected for partial sequence analysis covering the VP1/2A junction region of the viral genome to study the genetic relationships and epidemiological links between strains. Viral RNA was partially amplified by RT–PCR and sequenced by a solid phase method. Computer analysis revealed genetic divergence of the strains within each serotype. Most of the nucleotide differences between strains were silent: only a few specific amino acid substitutions were seen in the sequenced region. Three genotypes of poliovirus type 1 and two of poliovirus type 3 were co-circulating, while type 2 strains were represented by a single genotype. Representatives of all the genotypes present have been found among previously or concurrently characterized strains isolated elsewhere, but direct epidemiological links were found only in the case of serotype 1. Many of the epidemics caused by poliovirus type 1 in other countries were genetically linked to Pakistan. This study clearly shows the endemic circulation and wide variation of all three poliovirus serotypes in southern Pakistan and indicates the need for more effective vaccination programmes to prevent the further spread of these wild viruses.

Introduction

Genetic characterization of the concurrently circulating poliovirus strains is important for proper monitoring of the progress of the polio eradication programme of the World Health Organization (WHO). Published studies have mostly been concerned with the genetic divergence of poliovirus type 1 (Rico-Hesse et al., 1987; Kew et al., 1990; Zheng et al., 1993; van Nickerk et al., 1994; Mulders et al., 1995; Lipskaya et al., 1995), while the evolutionary branching of poliovirus type 3 (Kew et al., 1990; Lipskaya et al., 1995; Pöyry et al., 1990; Oostvogel et al., 1994) and especially that of type 2 (Kew et al., 1990) are less well characterized. All these studies have shown great variability of polioviruses and made it possible to find epidemiological links between geographically distant strains.

In spite of the remarkable progress of the WHO programme for global eradication of poliomyelitis, endemic poliovirus circulation still persists in many Asian countries including Pakistan. Patients with acute poliomyelitis from all over the province of Sindh in Pakistan have been referred to the Civil Hospital Karachi, and since 1989 they have been followed clinically (Agboatwalla, 1992) and examined virologically (Isomura et al., 1993; Roivainen et al., 1993). The problems with the efficiency of the oral poliovirus vaccine (OPV) programmes, often faced in developing countries (Patriarca et al., 1991), can also be found in Pakistan. The reported OPV coverage is about 80%, but the percentage of vaccine failure cases was estimated to be 20% in the Karachi region (Isomura et al., 1993). More than 80% of unvaccinated children were seropositive against all poliovirus serotypes at the age of five (Isomura et al., 1993), indicating a widespread circulation of wild polioviruses. In addition, there was an outbreak caused by serotype 1 in autumn 1990 and another caused by serotype 2 in 1991 (Isomura et al., 1993).

In this study we have analysed genetic relationships between wild poliovirus strains within all three serotypes isolated from children with acute poliomyelitis in the Civil Hospital Karachi, in 1989–1993, by using RT–PCR amplification and solid phase sequencing. Sequence data for the VP1/2A junction region were obtained and compared with those of previously isolated strains.
Methods

RNA extraction, RT-PCR and sequencing of PCR products

Enrolment of patients, collection of faecal specimens and isolation of virus strains in cell cultures as well as identification of the isolates by neutralization with type specific hyperimmune rabbit antisera were as described previously (Roivainen et al., 1993). Serotype 1 was most often isolated from the specimens. Because of the greater number of type 1 strains, only about one-third of them were sequenced instead of two-thirds for serotypes 2 and 3.

For sequence analysis, one culture tube of GMK cells (a continuous cell line of green monkey kidney origin) was infected with the test virus at the first passage level and grown until CPE was complete. The cells were then frozen and thawed, centrifuged at low speed, and treated with proteinase K (Boehringer Mannheim) at a final concentration of 100 μg/ml for 1 h at 55 °C. After phenol extraction and ethanol precipitation the RNA was dried, resuspended in 50 μl of RNase free water and 40 μl of RNase inhibitor (Promega) was added.

Six microlitres of the suspension was used for cDNA synthesis with 100 pmol of primer 612 (Table 1). The 40 μl reaction mixture contained 50 mM-Tris-HCl pH 8.3, 70 mM-KCl, 10 mM-MgCl₂, 0.5 mM-dNTPs, 40 U of RAAs inhibitor and 40 U of AMV reverse transcriptase (Life Sciences). The mixture was incubated for 1 h at 55 °C and for an additional 10 min at 65 °C to inactivate the enzyme. Two microalitres of cDNA-reaction mixture was used for PCR. The variation of the virus strains made it necessary to use more than one primer pair for each serotype. Type 1 strains were amplified with primers 4069-B3480 or 4079-B3480 (Table 1). Usually, 100 pmol of primer was used per reaction, although in some cases the biotinylated primer (B479 or B3480) was used in a 3-10-fold excess (100 pmol) as compared to the other primer of the pair (10-30 pmol). The 100 μl reaction mixture contained a 200 μM concentration of each dNTP, 0.01% gelatin, 0.1% Triton X-100, 50 mM-Tris-HCl pH 8.8 and 4 U of Taq polymerase. Temperature cycling was performed using MJ Research PTC-100 or Pharmacia LKB thermal cyclers. For the first five cycles low stringency PCR conditions were used to amplify cDNA from as many virus strains as possible. The temperatures and times for denaturation, annealing and elongation were 94 °C for 1 min, 30 °C for 1 min and 72 °C for 1 min, respectively, for the first five cycles, and 94 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min for the next 35 cycles. After completion of the cycling programme, an additional 5 min incubation at 72 °C was used to ensure the completion of final extensions.

Biotinylated amplicons were captured on avidin-coated polystyrene beads (Avidin–Fluoricon Assay particles, IDEXX) as described earlier (Salminen, 1992), and sequenced by using the Sequenase 2.0 kit (United States Biochemical). Internal primers were used for sequencing the amplification product: 3861 or 4069 for type 1 strains, 4045 or 3683 for type 2 strains and 3429 or 3484 for type 3 strains.


Computer analysis of sequences. The sequence analysis was based on 150 nucleotides in the VP1/2A junction region and was performed on a VAX computer. The nucleotide identities were calculated with the GAP program of the Genetics Computer Group Package (Devereux et al., 1984). Dendrograms were created with the program Clustal W (Higgins et al., 1992). Some published VP1/2A sequences were used in sequence comparisons: P1/16006/IND82 (Rico-Hesse et al., 1987), P1/7064/IND86 (Kew et al., 1990), P1/1372/CHN90 (Zheng et al., 1993), P2/7079/IND86, P2/15340/IND82, P2/11996/SWE77, P2/WHO6683/YUG78, P2/7834/PER83, P3/105/IND86, P3/7095/IND86 (all Kew et al., 1990), and P3/23127/FIN84 (Hughes et al., 1986).

Results

Genetic divergence in the VP1/2A region

Thirty-five poliovirus type 1 strains isolated in Pakistan during 1989–1993 were partially sequenced. The strains

Table 1. Target sequences of primers used in cDNA synthesis, PCR and sequencing

<table>
<thead>
<tr>
<th>Primer</th>
<th>Location</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>612</td>
<td>7407–7428*</td>
<td>5' GAAAAATTTACCCCTACAACG 3'</td>
</tr>
<tr>
<td>3430</td>
<td>2612–2640*</td>
<td>5' ACTGCTGTGTACCTGCTACTAA(T/C)C 3'</td>
</tr>
<tr>
<td>4069</td>
<td>3212–3228*</td>
<td>5' GTCAATGATCACAACC 3'</td>
</tr>
<tr>
<td>3861</td>
<td>3238–3256*</td>
<td>5' CACTTCCACAGGATAGATG 3'</td>
</tr>
<tr>
<td>B479</td>
<td>2955–2975**</td>
<td>5' GTATG/G/TTTCCTCCA/G(C)/GG(A/C/G/T)G(G)(G/C/G/T)CC 3'</td>
</tr>
<tr>
<td>3429</td>
<td>3211–3230**</td>
<td>5' GTATG/GATGACCTAAATCC 3'</td>
</tr>
<tr>
<td>4045</td>
<td>3246–3264**</td>
<td>5' CAAGATTTAGGATTTACATG 3'</td>
</tr>
<tr>
<td>3683</td>
<td>3486–3503**</td>
<td>5' AGTAGATCCCTGGTCACC 3'</td>
</tr>
<tr>
<td>2785</td>
<td>3703–3722**</td>
<td>5' GGA/G(A/C/G/T)G/G(T/T)G/GAAATA(C/T)T 3'</td>
</tr>
<tr>
<td>3484</td>
<td>3161–3179***</td>
<td>5' AGTGCAAATGACTGATGA 3'</td>
</tr>
<tr>
<td>B3480</td>
<td>3478–3495***</td>
<td>5' ATACAGGTCTGTCTTCCAC 3'</td>
</tr>
</tbody>
</table>

† Depending on the serotype the numbering is according to Sabin 1 (*) (Nomoto et al., 1982), Sabin 2 (**) (Toyoda et al., 1984) or Sabin 3 (***) (Stanway et al., 1984) strains.
could be divided into three genotypes: A, B and C (Fig. 1), using the designated criterion that a strain represents a new genotype if it has less than 85% nucleotide identity with previous ones (Rico-Hesse et al., 1987). Two of the genotypes (B and C) contained two discernible clusters (a and b).

The recently published data concerning the molecular epidemiology of wild poliovirus type 1 genotypes circulating in Europe and in the Middle East (Mulders et al., 1995) give an overview of related strains for all Pakistani poliovirus type 1 genotypes. In this study we took a closer look at genotype C. Several direct epidemiological links, by definition having a minimum nucleotide identity value of 98% (Rico-Hesse et al., 1987), were found when comparing the strains belonging to cluster b of genotype C with strains isolated recently elsewhere (Fig. 2): strain P1-8768 isolated in 1988 in Oman shared 98.7% nucleotide identity with strain P1-046 representing the Pakistani poliovirus type 1 epidemic strains. There is 98% nucleotide identity between P1-2742 isolated in 1992 in Jordan and Pakistani strain P1-193 isolated in 1992. The closest related strain to Pakistani poliovirus type 1 strains was found to be strain P1-9261 isolated in 1992 in the United Arab Emirates. It shared 99.3% nucleotide identity with Pakistani strain P1-181 isolated in 1991. The other cluster (a) of genotype C was distantly (with about 90% nucleotide identity) related to strains isolated in 1990 and 1991 in northwestern China (Zheng et al., 1993) (Fig. 2). Cluster Ca circulated in Pakistan at least until 1992.

All 29 sequenced poliovirus type 2 strains isolated in Pakistan in 1990–1993 formed a single genotype having

Fig. 2. Dendrogram showing the genetic links between the Pakistani poliovirus type 1 genotype C, and the type 1 strains isolated elsewhere. The following country abbreviations are used in Figs 2, 4 and 6: ARM, Armenia; BUL, Bulgaria; CHN, China; CYP, Cyprus; EGY, Egypt; FIN, Finland; FRA, France; GEO, Georgia; IND, India; ISR, Israel; JOR, Jordan; MAA, Malaysia; NET, The Netherlands; OMA, Oman; PAK, Pakistan; PER, Peru; ROM, Romania; RUS, Russia; SRL, Sri Lanka; SWE, Sweden; SYR, Syria; TAJ, Tadjikistan; TRK, Turkmenistan; TUN, Tunisia; TUR, Turkey; UAE, United Arab Emirates; UKR, Ukraine; USA, United States of America; YUG, Yugoslavia. Pakistani strains are marked by an asterisk (*).
a minimum nucleotide identity of about 88% (Fig. 3). Strains circulating in 1990 (P2-009, P2-37 and P2-56) belonged to two clusters. These clusters differed from each other by a maximum of 12.5%, while the strains isolated in 1991 showed relatively little variation (0–5.3%). Strains belonging to the latter cluster were isolated until 1993. The VP1/2A region sequences of the Pakistani poliovirus type 2 strains were compared with
other type 2 strains isolated over several decades in different parts of the world (Fig. 4). The closest relatives were found in India: for example, strain P2-8671 isolated in 1991 in India had 94.7% nucleotide identity with Pakistani strain P2-037 isolated in 1990.

Twelve poliovirus type 3 strains isolated in 1990-1993 in Pakistan belonged to one genotype with a minimum nucleotide identity of 89.3% (Fig. 5). One strain (P3-085, isolated in October 1990) belonged to an additional genotype and shared 90% nucleotide identity with strain P3-0670 isolated in 1991 in Oman (Fig. 6). No direct epidemiological links between the Pakistani type 3 strains and the strains isolated during the last decade elsewhere were found. The closest relatives to the main genotype in Pakistan were the epidemic strains isolated in 1992 in the Netherlands (represented by strain P3-16260) (Oostvogel et al., 1994), strains isolated in 1991-1992 in India and a strain isolated in 1992 in France (P3-19638) (Fig. 6).

**Amino acid replacements in the VP1/2A junction region of Pakistani poliovirus strains**

Most of the nucleotide differences between strains were silent. The variable positions and the amino acid residues differing from those of Sabin vaccine strains, found among the sequenced Pakistani strains, are shown in Fig. 7. Pakistani poliovirus type 1 strains 162, 177, 185 and 190 belonging to genotype C, cluster a, had the same amino acid composition in the VPI region as Sabin 1 vaccine strain. All Pakistani poliovirus type 2 strains had a substitution of tyrosine for phenylalanine-280 of VP1 of the Sabin 2 strain. One Pakistani poliovirus type 3 strain (085) had a substitution of serine for alanine-274 of the Sabin 3 strain. At antigenic site 3A (underlined in Fig. 7) all Pakistani poliovirus type 3 strains differed totally from the Sabin 3 vaccine strain. Most of them had amino acids 286-290 lysine-aspartic acid-glycine-leucine-alanine (KDGLA) instead of arginine-asparagine-asparagine-leucine-aspartic acid (RNNLD). Strain 80 belonging to the minority Pakistani genotype differed from the other Pakistani strains at antigenic site 3A by having a glutamic acid instead of glycine at position 288 of VP1.

In protease 2A amino acid 2 varied, being phenylalanine, tyrosine, leucine, histidine or cysteine in type 1 strains and phenylalanine or tyrosine in type 2 strains. Other variable positions were amino acid 6 in type 1 and type 3 strains and amino acid 19 in type 1 strains (Fig. 7).
Discussion

Partial sequence covering the VP1/2A junction region is regularly used in molecular epidemiological studies on polioviruses, because of the considerable data bank available for comparison. In order to study the range and patterns of the genetic variation of currently circulating wild-type polioviruses, we sequenced 77 wild poliovirus strains representing all three serotypes. Strains were isolated from paralytic patients in Pakistan (mostly in the Karachi region), where polio is endemic. Most of the strains isolated were found to be genetically related to strains isolated previously or concurrently in neighbouring countries.

Three poliovirus type 1 genotypes (A, B and C) with separate clusters (Ba, Bb, Ca and Cb), possibly in the course of segregating to new genotypes, were found to have circulated during the years monitored. Cluster Cb included a number of strains, isolated between June and October in 1990, that had at least 98% nucleotide identity, indicating that these strains were the causative agents of the poliovirus type 1 outbreak described earlier (Isomura et al., 1993). A sign of an emerging epidemic might be that three poliovirus type 1 strains isolated in 1993 showed very high nucleotide identity (100–99.3%, Fig. 1). Strains related to cluster Cb have circulated for over 10 years in Europe, the Middle East and the Indian subcontinent (Mulders et al., 1995). Many of the epidemics caused by strains belonging to this genotype (Lipskaya et al., 1995; Sutter et al., 1991; WHO, 1993) were genetically linked to Pakistan.

Only a few strains belonging to genotypes A and B could be found among a large number of sequenced strains, suggesting that those genotypes may not be truly endemic for the Karachi region. One of the four strains (21196-9) isolated in Islamabad, in the northern part of the country, belonged to genotype B and the others to cluster b of genotype C indicating that the same genotypes were circulating in Karachi and Islamabad. Recently, representatives of six different genotypes of poliovirus type 1 were found among strains collected during 1982–1991 in China (Zheng et al., 1993). We found the strains belonging to cluster a of genotype C to be distantly related (about 90% nucleotide identity, Fig. 2) to the genotype CHN-XJ90 circulating in the northeastern part of China.

Molecular epidemiology of contemporary polioviruses in the other neighbouring countries, India, Iran and
Tadjikistan is poorly known. So far, only one poliovirus type 1 genotype has been described for Iran, and another for India and Tadjikistan (Mulders et al., 1995; Lipskaya et al., 1995). Strains belonging to both of those genotypes were also isolated in Pakistan, clusters Bb and Cb. No molecular epidemiological data on polioviruses circulating in Afghanistan are available. The polio immunization coverage there was only 33% in 1990 (WHO, 1991) and wild polioviruses of all three serotypes are presumably co-circulating.

Circulation of poliovirus type 2 in a vaccinated population is considered to be a sign of vaccine or vaccination failure, since circulation of type 2 poliovirus is supposed to be best inhibited by vaccines. Kew et al. (1990) reported several poliovirus type 2 genotypes, two of which were endemic to India. All studied strains...
isolated recently in India and Pakistan belonged to the same genotype, which has been circulating for several years. Two separate clusters were found in Pakistan; the cluster dominating in 1991–1992 was responsible for the outbreak, while strains belonging to the other cluster were found only before the epidemic.

Poliovirus type 3 strains formed a minority among the strains isolated from the Pakistani specimens indicating low circulation of type 3 strains. No epidemic strains (with at least 98% nucleotide identity) like type 1 and 2 were found during the years examined. A strain related to Pakistani strains was isolated in 1992 in France. The outbreak in 1992–1993 among an unvaccinated religious group in the Netherlands (Oostvogel et al., 1994) was also caused by a strain belonging to the same genotype as the Pakistani strains.

The designated antigenic site 3A is considered to be important only for poliovirus type 3 (Minor et al., 1986). The observed wide variation of Pakistani type 1 polioviruses at and near this antigenic site suggested, however, that the carboxyl terminus of VP1 may also be antigenic in serotype 1. On the other hand, the epidemic type 1 strains isolated in 1990 had amino acids in the carboxyl terminus of VP1 identical with those of the Sabin 1 vaccine strain.

Protease 2A catalyses the cleavage of the large polyprotein at the structural/nonstructural protein junction (Toyoda et al., 1986). The amino acid composition of 2A, and especially that of its N-terminal part is highly conserved among known polioviruses (Palmenberg et al., 1990). Only one of the variable positions of Pakistani strains (amino acid 19) in the amino-terminal end of protease 2A was located near to the catalytic triad (Cys-109, His-20 and Asp-38) in the hypothesized structure of 2A (Bazan & Fletterick, 1988) and possibly might have some influence on enzyme activity. Consistent with the earlier published data (Rico-Hesse et al., 1987), we found that amino acid 2 especially is variable in poliovirus type 1
Among the studied strains, type 1 presented greatest variation. In spite of a relatively large number of strains sequenced no new genotypes were found. Strains belonging to all the genotypes described have been previously or concurrently described elsewhere, but direct epidemiological links were found only in the case of genotype C of poliovirus type 1. There have been several outbreaks caused by this genotype in the Middle East and in south-western Asia during the last few years.

We thank Dr Olen Kew and Dr Galina Lipskaya for providing some unpublished data and Dr Leena Kinnunen for fruitful discussions through the work. Excellent technical assistance was provided by Ms Päivi Hirtti. A. H. T. P. and T. H. were supported by grants from the Academy of Finland and from the Sigrid Juselius Foundation, Helsinki, Finland. M. N. M. was supported by the Praeventiefonds (grant 28-3225).

References


(Received 11 April 1995; Accepted 18 August 1995)