Evolution of a rare vaccine-derived multirecombinant poliovirus

Ioannis Karakasiliotis,† Eleni Paximadi2 and Panayotis Markoulatos2

1Department of Virology, Hellenic Pasteur Institute, 127 Vas. Sofias Avenue, 11521 Athens, Greece
2Department of Biochemistry and Biotechnology, University of Thessaly, 26 Ploutonos and Aiolou Street, 41221 Larissa, Greece

Recombination is one of the mechanisms by which viral genomes evolve. A vaccine-derived multirecombinant poliovirus strain was isolated from a 5-month-old child with vaccine-associated paralytic poliomyelitis after oral poliovirus vaccine administration. The isolate had an S2/S1/S2/\text{S1} primary genomic structure as revealed by restriction fragment length polymorphism and sequencing analysis. Recombination of the middle S1/S2 region is extremely rare and one of the few characterized types of recombination with Sabin type 1 as a 5′ partner. An attempt was made to perform evolutionary analysis of the contributing sequences using the identified mutations in comparison with the original Sabin sequences. A hypothesis is proposed for the order in which the identified recombination events occurred.

INTRODUCTION

Polioviruses are considered to be among the most rapidly evolving viruses, and use two main evolutionary strategies, mutation (Domingo & Holland, 1997; Drake & Holland, 1999; Gavrilit et al., 2000; Liu et al., 2000) and genomic rearrangement (reviewed by Agol, 1997). Homologous recombination events are the most frequently observed genomic rearrangements and occur mainly in the non-structural protein-coding regions of the viral genome (Cuervo et al., 2001; Georgopoulou & Markoulatos, 2001; Karakasiliotis et al., 2004). However, in some very rare cases, recombination events occur in the extremities of the capsid-coding region (P1), supporting a theory that requires the maintenance of capsid integrity for each poliovirus serotype (Georgescu et al., 1995; Blomqvist et al., 2003). The extensive application of oral poliovirus vaccine (OPV) worldwide has resulted in an increase in the number of studies on recombination in polioviruses. During OPV administration, three live, but attenuated, Sabin strains coinfect the cells of the intestine, resulting in a wide variety of recombinant viruses (Georgescu et al., 1994; Cuervo et al., 2001).

Recombinants of all three Sabin types have been investigated from healthy vaccinees, and patients with vaccine-associated paralytic poliomyelitis (VAPP) or other vaccine-associated diseases. The presence of Sabin type 1 recombinants is extremely rare as are any S1/Sx (S1/S2 or S1/S3) recombinations in general. However, sequences of Sabin type 1 origin are regularly found in Sx/S1 recombinant strains (Driesel et al., 1995; Gavrilit et al., 2000; Cuervo et al., 2001). The reasons for this profound imbalance are still unknown. The explanation for the dominance of some recombination types, and the localization of their recombination sites in selected genomic regions, also remains to be revealed (Guillot et al., 2000; Cuervo et al., 2001). The preservation of viable RNA secondary structures has been proposed as a possible reason for this selection (Karakasiliotis et al., 2004).

The vast majority of recombinants have a single site of recombination, but multiple sites of recombination have been identified in one isolate, reflecting the occurrence of multiple recombination events (Georgescu et al., 1994, 1995; Guillot et al., 2000; Cuervo et al., 2001; Georgopoulou & Markoulatos, 2001; Kew et al., 2002; Blomqvist et al., 2003; Liu et al., 2003; Karakasiliotis et al., 2004). Multirecombinants show the high flexibility of the poliovirus genome. Moreover, they also provide information about interactions between functional regions of the genomic RNA (Georgescu et al., 1994).

In this study we investigated an isolate from a case of VAPP in a 5-month-old child, using the RT-PCR-RFLP (restriction fragment length polymorphism) sequencing method described previously by our group (Karakasiliotis et al., 2004). We studied the primary structure of the recombination sites, and characterized the genomic organization of the isolate as an S2/S1/S2/S1 tri-recombinant.
METHODS

Viruses and cells. The clinical isolate was derived from a case of VAPP after the first OPV dose, in 1982 in Greece. The symptoms started approximately 20 days after OPV administration. The patient was a 5-month-old child, and the virus was isolated from stools (clinical strain I$_{13}$). The Sabin reference strains, maintained in our laboratory, and the clinical isolate were grown in HEp-2 cells in round-bottomed plastic tubes containing 2 ml MEM-D medium for about 6 days. When a complete cytopathic effect (CPE) was observed, the tubes were frozen at $-20^\circ$C. Serotyping was achieved through a sero-neutralization test with rabbit polyclonal antibodies (National Institute for Public Health and the Environment, RIVM, The Netherlands). To avoid having a mixture of viruses, serial dilutions of the sample were grown in HEp-2 cells, and the last dilution with complete CPE was used for redetermining the serotype, as described above.

Extraction of viral RNA. After further passaging of the viruses in HEp-2 cells, with an m.o.i. ranging from 1 to 5 p.f.u. cell$^{-1}$, a CPE greater than 75% was usually observed 96 h after inoculation. Frozen and twice-thawed cell culture (120 ml) was incubated for 10 min at room temperature with 480 ml lysis buffer (4 M guanidine thiocyanate, 0.5% N-lauryl sarcosine, 1 mM dithiothreitol, 25 mM sodium citrate and 40 ml glycogen per tube) (Casas et al., 1995). Cold 2-propanol (600 ml) was added, and the tubes were allowed to stand for 15 min on ice. After centrifugation for 10 min at 14,000 g at 4°C, the 2-propanol was removed and the pellet was washed by the addition of 1 ml 70% ethanol. The pellet was dried and dissolved in 50 ml sterile, double-distilled water (ddH$_2$O).

RT-PCR-RFLP and sequencing for the determination of the recombinant nature of the isolate. The extracted RNAs were used for an RT-PCR-RFLP procedure, as described previously (Karakasiliotis et al., 2004), with the UG$_{52}$/UC$_{23}$, UG$_1$/UC$_1$, UG$_{23}$/UC$_{15}$ and UG$_{16}$/UC$_{12}$ primer pairs, which roughly characterizes a strain as a recombinant or not (Table 1). The four PCR products were digested with restriction enzymes at 37°C for 1 h: the UG$_{52}$/UC$_{23}$ product with HaeII, NcoI, Ddel and Aval (Promega), the UG$_1$/UC$_1$ product with Hpall (New England BioLabs), HaelII and Ddel, the UG$_{23}$/UC$_{15}$ product with Hinfl (New England BioLabs) and Rsal (Invitrogen), and the UG$_{16}$/UC$_{12}$ product with Hinfl, Rsal and Ddel. These digestion products were analysed in a 4% agarose gel, with HaelII-digested phage $\Phi$X174 DNA (Invitrogen) as a molecular mass marker. The RFLP analysis of the UG$_{52}$/UC$_{23}$ PCR product was used, as described by Georgopoulou et al. (2000), to identify the Sabin type composition of the isolate, and its genotype. The PCR products found to be putative recombinants were purified with a NucleoSpin Extract kit (Macherey-Nagel), and were sequenced by Macrogen.

PCR and sequencing of the 3’ half of the viral genome. Following the results of the preliminary study, an extended region of the genome was amplified with overlapping primer pairs, in order to include all the crossover sites. The primer pairs used were as follows: S$_{97}$/S$_{598}$, S$_{48}$/S$_{911}$, S$_{70}$/S$_{299}$, S$_{71}$/S$_{450}$ and S$_{198}$/S$_{885}$ (Table 1). The PCRs were performed in a Perkin Elmer Gene Amp PCR 9600 thermal cycler for 40 cycles. Each cycle consisted of 20 s denaturing at 94°C, 10 s annealing at 45°C for S$_{97}$/S$_{598}$ and S$_{48}$/S$_{911}$ primer pairs, and at 55°C for S$_{70}$/S$_{299}$, S$_{71}$/S$_{450}$ and S$_{198}$/S$_{885}$ primer pairs, and 20 s extension at 74°C, followed by a final extension at 78°C for 15 min. The reaction mixture consisted of 3 ml cDNA, 2 ml each primer pair (50 pmol ml$^{-1}$), 5 ml 10 × $Taq$ reaction buffer, 2 mM MgCl$_2$, 1 mM dNTPs, 1.5 U Bio $Taq$ DNA Polymerase and ddH$_2$O up to 50 ml final reaction volume. The PCR products were purified and sequenced as described above.

Location of the recombination sites. The sequences of the isolate were aligned with the reference strains cited in GenBank, with the multiple online alignment program ClustalW (http://www.ebi.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer sequence and position</th>
<th>Polarity</th>
<th>Region</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UG$_{52}$*</td>
<td>(162)-CAAGCACCCTTGTGTTCCCGG-(182)†</td>
<td>Sense</td>
<td>5’ UTR</td>
<td>434</td>
</tr>
<tr>
<td>UG$_{23}$*</td>
<td>(595)-TTGTCACATAAGCAGCCA-(577)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UG$_1$*</td>
<td>(2402)-TTTGGTTCAGCGTGAATGA-(2421)†</td>
<td>Sense</td>
<td>VP1</td>
<td>480</td>
</tr>
<tr>
<td>UG$_c$*</td>
<td>(2882)-GAATTCATGGCATCAAATCTAGA-(2862)†</td>
<td>Sense</td>
<td>2C</td>
<td>797</td>
</tr>
<tr>
<td>UG$_{23}$*</td>
<td>(4169)-AAGGGATTGGAGTGGGTGTC-(4188)†</td>
<td>Sense</td>
<td>2C–3A</td>
<td>502</td>
</tr>
<tr>
<td>UG$_{15}$*</td>
<td>(4965)-CATCTCTGAATTTGTGTCG-(4946)†</td>
<td>Sense</td>
<td>2C–3C</td>
<td>864</td>
</tr>
<tr>
<td>S$_{97}$</td>
<td>(4777)-TTAGCCTCACAACCTCGCAG-(4796)†</td>
<td>Sense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{598}$</td>
<td>(5278)-TGATGTTCTCCTGTTGTAAGA-(5156)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{48}$</td>
<td>(4968)-TCTTTTATGGTGCGCAAGG-(4987)†</td>
<td>Sense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{911}$</td>
<td>(5831)-GTTTTCGTCCACCGAGATT-(5812)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{70}$</td>
<td>(5770)-TATGTATGTTCCTGTGCTGCTG-(5791)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{299}$</td>
<td>(5999)-ACTGATTTCACCTGACTCTG-(5978)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UG$_{16}$*</td>
<td>(5921)-GTTGCTGGGAAACGTTACA-(5940)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UC$_{12}$*</td>
<td>(6516)-TCAAATATGCTGTTTTCCTCG-(6494)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_1$</td>
<td>(6427)-ATATGGAATCACCCTCCCAC-(6447)†</td>
<td>Sense</td>
<td>3D</td>
<td>444</td>
</tr>
<tr>
<td>S$_{450}$</td>
<td>(6870)-ATTGAATGTCCTGAACAAAC-(6851)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{198}$</td>
<td>(6728)-GCTAAAGAAGTGGTCCTGAGAAA-(6751)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$_{885}$</td>
<td>(7415)-CTACACACGTATGGACCCATCCAA-(7392)†</td>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Guillot et al. (2000).
†Numbers refer to positions in the Sabin type 1 strain sequence (GenBank no. AY184219).
‡Numbers refer to positions in the Sabin type 2 strain sequence (GenBank no. AY184220).
Evolutionary investigation of the isolate. The evolution of the isolate was studied through alignment of its sequence with a sequence constructed by the combination of the reference strains, Sabin type 1 (Rezapkin et al., 1994) and Sabin type 2 (Rezapkin et al., 1999), based on the located recombinations. The alignment of the nucleotide and amino acid sequences of both clinical and constructed sequences was performed with ClustalW. For a rough estimation of the strain’s age, three parameters were calculated according to the method of Li et al. (1985), and used for the estimation of the evolutionary rates of polioviruses (Gavrilin et al., 2000). The three parameters were the percentage of mutated synonymous sites among all synonymous sites (kS), the percentage of mutated non-synonymous sites among all non-synonymous sites (kA), and the percentage of mutated sites among all sites (kA). The mean of the three age values estimated using the above parameters is given, followed by the SD.

RESULTS

Preliminary study of the isolate

Through the sero-neutralization test with the polyclonal antisera, the serotype of the isolate was identified as a type 2 poliovirus. Through the genotyping method (Georgopoulou et al., 2000), the 5′ UTR (UG23/UC33) product of the strain was found to have a Sabin type 2 origin (data not shown). The isolate did not present any modified restriction site or a sum for the RFLP bands greater than expected (434 nt), either of which could indicate the presence of a mixture of polioviruses. The PCR product from the UG1/UC1 primers, which amplifies a region in VP1, was digested with HpaII, HaeIII and DdeI. The RFLP profiles of HpaII (Fig. 1a), HaeIII and DdeI (Fig. 1b) restriction enzymes resembled the profiles of the reference strain Sabin type 2.

The region amplified with the UG23/UC15 primers was digested with HinfI and Rsal enzymes. Both restriction enzymes gave a profile identical to the one from a respective digestion of the Sabin type 2 reference strain. In Fig. 1(c), the digestion with HinfI is illustrated.

The region amplified with the UG16/UC12 primers, amplifying a region at the 3C–3D junction, was digested with HinfI, Ddel and Rsal restriction enzymes. The RFLP profile with HinfI was identical to the Sabin type 1 profile. The digestion with Ddel resulted in a profile that didn’t resemble any known RFLP profile (Fig. 1d), while the digestion with Rsal resulted in a pure Sabin type 2 profile (Fig. 1e). The RFLP results for the region amplified by UG16/UC12 primers showed a putative recombination site between the Sabin type 1 and Sabin type 2 sequences.

Sequencing of the last genomic region revealed the presence of two recombination sites. The first was from an S1/S2 recombination event with the crossover site located between nt 5995 and 6004, while the second site was from an S2/S1 recombination event located between nt 6388 and 6389. Both sites were located in the 3D-coding region, which means that the 3D-coding region had an S1/S2/S1 primary structure (Fig. 2).

Sequence of the 3′ half of the virus

Given the result of the RFLP analysis for the UG23/UC15 PCR product (Sabin type 2 profile), another crossover site was expected to be located between the 2C- and 3D-coding regions. Thus, we conducted extended sequencing of the 3′ half of the virus. We located a third crossover site between nt 4985 and 5000 into the 3′ end of the 2C-coding region. The accession number for the complete 3′ half sequence in GenBank is AY830710 (nt 4189–7368; Sabin type 2 numbering), and the accession number for the 5′ partial sequence of the VP1 region is AY830709. By carefully scrutinizing the chromatograms we were unable to locate any ambiguous base calls within the sequenced region,

![Fig. 1. Preliminary RFLP results for the isolate I34. (a) HpaII digestion of the VP1 region, (b) Ddel digestion of the VP1 region, (c) HinfI digestion of the 2C region, (d) Ddel digestion of the 3C–3D region and (e) Rsal digestion of the 3C–3D region. In every assay Sabin type 1 and Sabin type 2 reference strains were used for the original, serotype-specific RFLP profiles.](http://vir.sgmjournals.org)
supporting our conclusion that the sequence is derived from a clonal virus isolate.

Mutations and evolution of the virus

Comparison of the I34 sequence with the sequences of the reference Sabin strains revealed a number of mutations and amino acid substitutions (Table 2). The $K_s$, $K_a$, and $K_t$ parameters were calculated for the entire sequenced genomic region. The calculated values were $K_s$ 1-02 %, $K_a$ 0-11 % and $K_t$ 0-31 %. These values were used to estimate the age of the strain by relating them to existing reference sequenced region of the I 34 genome.

The calculated values for the three parameters were $K_s$ 0-63 %, $K_a$ 0-06 % and $K_t$ 0-18 %, corresponding to 35, 23 and 25 days with a mean of 28 days (4 weeks ± 6 days).

DISCUSSION

We came across a case of a 5-month-old child that contracted flaccid paralytic poliomyelitis about 20 days after administration of the OPV, during an investigation of VAPP cases for the potential implication of recombinant strains. Following preliminary RT-PCR-RFLP sequencing (Karakasiliotis et al., 2004), the dominant poliovirus strain isolated from the stools of the child was shown to have two recombinant sites in the 3D region. The first site was an S1/S2 recombination site located between nt 5995 and 6004 (Sabin type 1 numbering), and the second one was an S2/S1 recombination site located between nt 6388 and 6389 (Sabin type 2 numbering). Considering the RFLP analysis results for the preceding regions, which were shown to correspond to a Sabin type 2 sequence, a third recombination site was expected to be found between the 2C and 3D regions. This site was located, through sequencing the whole 3' half of the genome, between nt 4985 and 5000 (Sabin type 2 numbering), and it was, as expected, from an S2/S1 recombination event. With this final characterization, the genome of the isolate was found to have an S2/S1/S2/S1 primary structure. An S2/S1 recombination event occurring in nt 6388 and 6389 has been previously identified (Cuervo et al., 2001). The S1/S2 recombination site, located in the 5' end of the 3D-coding region, is an extremely rare recombination type. Its crossover site may not have been incidentally located in that region. The very few recombinations that implicate Sabin type 1 as first partner are mainly S1/S3 type, and show that such a recombination (S1/Sx) occurs mainly in the same region as in the I34 isolate (Georgescu et al., 1994; Driesel et al., 1995; Cuervo et al., 2001). That provides additional evidence for the non-random distribution of recombination events, even amongst rarely occurring recombination types.

Multirecombinant vaccine-derived poliovirus strains have been characterized, either with full vaccine origin sequence...
(Georgescu et al., 1994), or with both vaccine and wild origin sequence (Kew et al., 2002; Liu et al., 2003). The latter ones have been implicated in VAPP cases in Hispaniola and China, where they emerged. Strains with two or more sites of recombination arise after a series of recombination events, either in the same host or in a different host. In the cases involving wild poliovirus sequences (Kew et al., 2002; Liu et al., 2003), the serial recombination events were found to occur at different times, in different hosts, and even in different, but close, geographical regions. In the case of clinical strain I34, the sequence was found to be purely in different, but close, geographical regions. In the case of the same recombination events, either in the same host or in a different host. In the cases involving wild poliovirus sequences (Kew et al., 2002; Liu et al., 2003), the serial recombination events were found to occur at different times, in different hosts, and even in different, but close, geographical regions. In the case of clinical strain I34, the sequence was found to be purely vaccine-derived. An attempt to determine the age of the virus was made using the evolutionary curves constructed by Gavrilen et al. (2000). These curves show the evolution of polioviruses through time applying the three evolutionary parameters used by Li et al. (1985), which are the $K_s$, $K_a$, and $K_r$ values. The calculation of these parameters showed a mean age of 7 (± 1) weeks for the isolate. This value was at least double that expected, as the virus had the chance to replicate in the child’s intestine for only about 3 weeks. This impairment led us to create a map of the mutations identified in the genome (Fig. 3). The mutation map showed an aggregation of the mutations in the region between the first and the third crossover sites. Following the same method, we calculated the $K_s$, $K_a$, and $K_r$ parameters for the region inside the crossover sites, and the regions outside them. The mean ages derived were approximately 10 and 4 (± 1) weeks, respectively.

The results for age estimation of the two sequence parts guide us towards the reconstruction of a potential history of this isolate. The age ($\sim 4 \pm 1$ weeks) calculated for the sequences outside the region flanked by the two recombination sites in 3D matched with the 3 weeks of virus replication in the child. The age calculated for the region between the recombination sites showed a different evolution start point from the rest of the genome ($\sim 10$ weeks) (Fig. 3). As can be expected for short periods of time for a multifactorial phenomenon, both stochastic and non-stochastic in nature, such as mutagenesis, the statistical error can be relatively high. Considering the limited knowledge and the complexity that characterize the mutation rates in polioviruses, we cannot exclude the possibility that all the recombination events occurred in the same child after vaccination. Nevertheless, considering the small dispersion of the ages estimated using the three parameters, a possible explanation can be proposed. Thus, a recombinant (S1/S2) virus that emerged before the vaccination of the child possibly could have been recombined with the new Sabin strains introduced with the vaccination, something that may have resulted in the different estimated ages for the two parts of the genome. There is no evidence about the order in which these final recombination events could have happened, but it is obvious from the sequence of the isolate that a Sabin type 2 and a Sabin type 1 strain were the donors of the 5' and 3' moieties, respectively, of the final S2/S1/S2/S1 multirecombinant virus (Fig. 3).

Even if we cannot propose that the isolated virus caused the VAPP (Georgescu et al., 1994, 1997), our results indicate that serious consideration should be taken before planning the future of worldwide anti-poliovirus vaccination, especially regarding the presence of vaccine-derived polioviruses that have been produced and spread into the environment during the past four decades of mass OPV administration. The presence of vaccine-derived polioviruses in immunodeficient patients (Bellmunt et al., 1999; Khetsuriani et al., 2003; MacLennan et al., 2004), civil sewage (Blomqvist et al., 2004; Vinje et al., 2004) or currently vaccinated populations clearly supports the existence of this problem. Finally, rare poliovirus isolates with multiple recombinations, like the one discussed here, may constitute a real threat to the vision of a polio-free world.

ACKNOWLEDGEMENTS

The present work was supported by research grants from the Direction des Affaires Internationales de l’Institut Pasteur Paris, ‘grant AC02 Vaccin Polio Oral’ and ‘grant Resistance aux Anti-Infectieux’.

REFERENCES


