Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1539–1543, March 1984
Medical Sciences

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$_1$b

(attenuation/mutation)

GLYN STANWAY*, PAMELA J. HUGHES*, ROGER C. MOUNTFORD*, PETER REEVE†‡, PHILIP D. MINOR†, GEOFFREY C. SCHILD†, AND JEFFREY W. ALMOND*

*Department of Microbiology, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom; and †National Institute of Biological Standards and Control, Holly Hill, Hampstead, London NW3 6RB, United Kingdom

Communicated by F. Sanger, November 8, 1983

ABSTRACT **As part of a study into the molecular basis of attenuation and reversion to neurovirulence in the Sabin poliovirus vaccines, we have determined the complete nucleotide sequence of a cloned DNA copy of the genome of P3/Leon/37, the neurovirulent progenitor of the type 3 Sabin vaccine strain, P3/Leon 12a$_1$b. Comparison of the sequence with that which we previously obtained for the vaccine strain [Stanway, G., Cann, A. J., Hauptmann, R., Hughes, P., Clarke, L. D., Mountford, R. C., Minor, P. D., Schild, G. C. & Almond, J. W. (1983) *Nucleic Acids Res.* 11, 5629–5643] indicates that attenuation has been brought about by a maximum of 10 point mutations, at least 5 of which are likely to be of minor significance. Predicted amino acid sequences of all the known virus-encoded proteins show that amino acid substitutions have occurred at only three positions. Two of these are in structural proteins (i.e., Ser → Phe in VP3 and Lys → Arg in VP1), and the third, Thr → Ala, is in the nonstructural protein P2-3b. The distribution and nature of nucleotide and amino acid sequence differences suggest that a single base substitution may be responsible for the attenuated phenotype of the vaccine strain.**

Over the past 20 years, the use of live, attenuated oral poliovirus vaccines has played a major role in reducing the incidence of paralytic poliomyelitis in many areas of the world (1). The live, attenuated vaccine strains in current use were developed by Albert Sabin in the 1950s by protracted passage of wild-type virus in monkey tissue *in vivo* and *in vitro* (2). Although the Sabin vaccines have proved highly effective and generally safe, occasional cases of paralytic disease do occur among vaccinees and their susceptible contacts (3). Viruses isolated from such cases can often be designated "vaccine-like" on the basis of serology and two-dimensional electrophoretic maps of ribonuclease T1 oligonucleotides (1, 4–7). The majority of these viruses are of serotype 3. Despite the widespread use of live, attenuated vaccines, little is known about the molecular basis of attenuation and why reversion to neurovirulence apparently varies among vaccines of the three distinct serotypes (3). Prospects for improving the safety of live, attenuated vaccines may be enhanced by a clearer understanding of these processes.

Polioviruses are typical members of the family Picornaviridae, and are composed of a small (27-nm diam.), icosahedral particle containing four virus proteins, VP1–VP4, surrounding a single-stranded positive-sense RNA genome of approximately 7500 nucleotides (8). The biological properties of different strains of virus must be determined ultimately by the nucleotide sequences of their genomes. One approach to the elucidation of the molecular basis of attenuation is, therefore, to compare the nucleotide sequences of neurovirulent and attenuated polioviruses (9, 10).

In the present study, we have determined the nucleotide sequence of the poliovirus type 3 strain, P3/Leon/37, the neurovirulent progenitor of the currently used Sabin type 3 live, attenuated vaccine (2). The complete nucleotide sequence of the vaccine strain itself has been published (11). A comparison of these two genome sequences reveals all of the mutations that have occurred during the attenuation process, which must include those responsible for the attenuated phenotype.

MATERIALS AND METHODS

Virus. P3/Leon/37 was isolated originally from the brain and spinal cord of a fatal case of paralytic poliomyelitis and is the neurovirulent progenitor from which the attenuated vaccine strain, P3/Leon 12a$_1$b, was derived by 53 passages *in vitro* interspersed with 21 passages *in vivo* (2). The isolate of P3/Leon/37 used in this study was obtained from A. Sabin by way of O. M. Kew of the Center for Disease Control (Atlanta, GA) as the strain from which the vaccine was developed. This isolate is significantly different in its molecular and biological properties from that deposited under the same name in the American Type Culture Collection (unpublished data). A plaque-purified derivative no. 960 was characterized by its two-dimensional electrophoretic map of ribonuclease T_1 oligonucleotides and found to be indistinguishable from the parental stock and from plaque-purified isolate no. 411 of the vaccine strain. P3/Leon/37 960 was shown to be neurovirulent in monkeys (10).

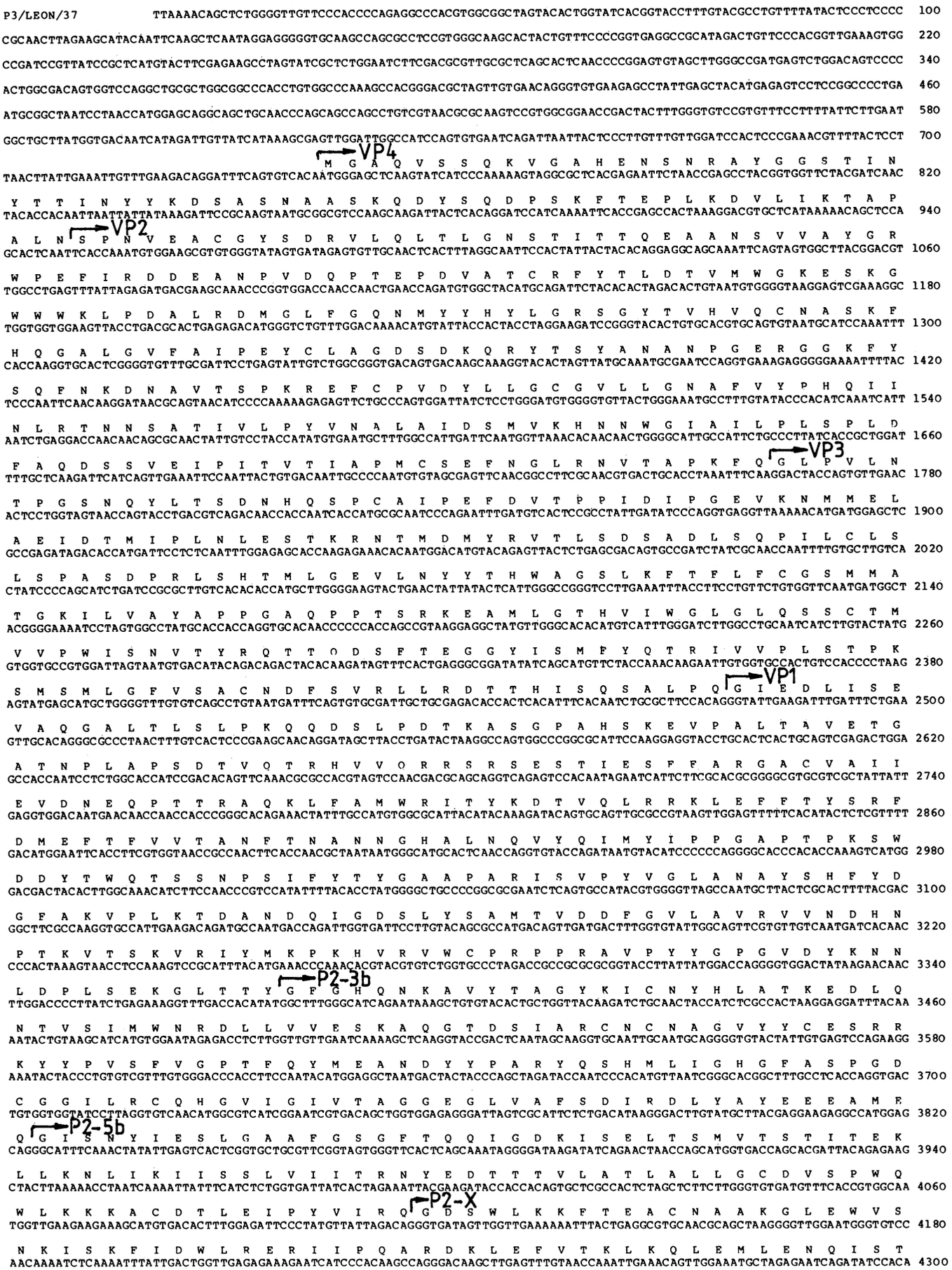
cDNA·RNA hybrids prepared from plaque-purified isolate 960 were cloned in *Escherichia coli*, and a plasmid containing a complete copy of the genome (pOLIO-Leon) was constructed (ref. 12; unpublished data).

Nucleotide Sequence Determinations. pOLIO-Leon (10 μg) was digested with *Pst* I, and the two poliovirus-specific fragments were isolated. These were circularized by ligation and then sheared by sonication. The random fragments thus generated were end-repaired by treatment with the large fragment of *E. coli* DNA polymerase I and fractionated on a 1.5% agarose gel. DNA of size 300–1000 base pairs was electroeluted, subcloned into *Sma* I-digested, phosphatase-treated M13 mp8 and sequenced by the dideoxynucleotide method (13, 14).

Nucleotide sequences thus generated were compared with the sequence of P3/Leon 12a$_1$b by using computer programs developed by Staden (15).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡Present address: Smith Kline & French Laboratories, 1500 Spring Garden Street, P.O. Box 7929, Philadelphia, PA 19101.



(Fig. 1 continues on the next page.)

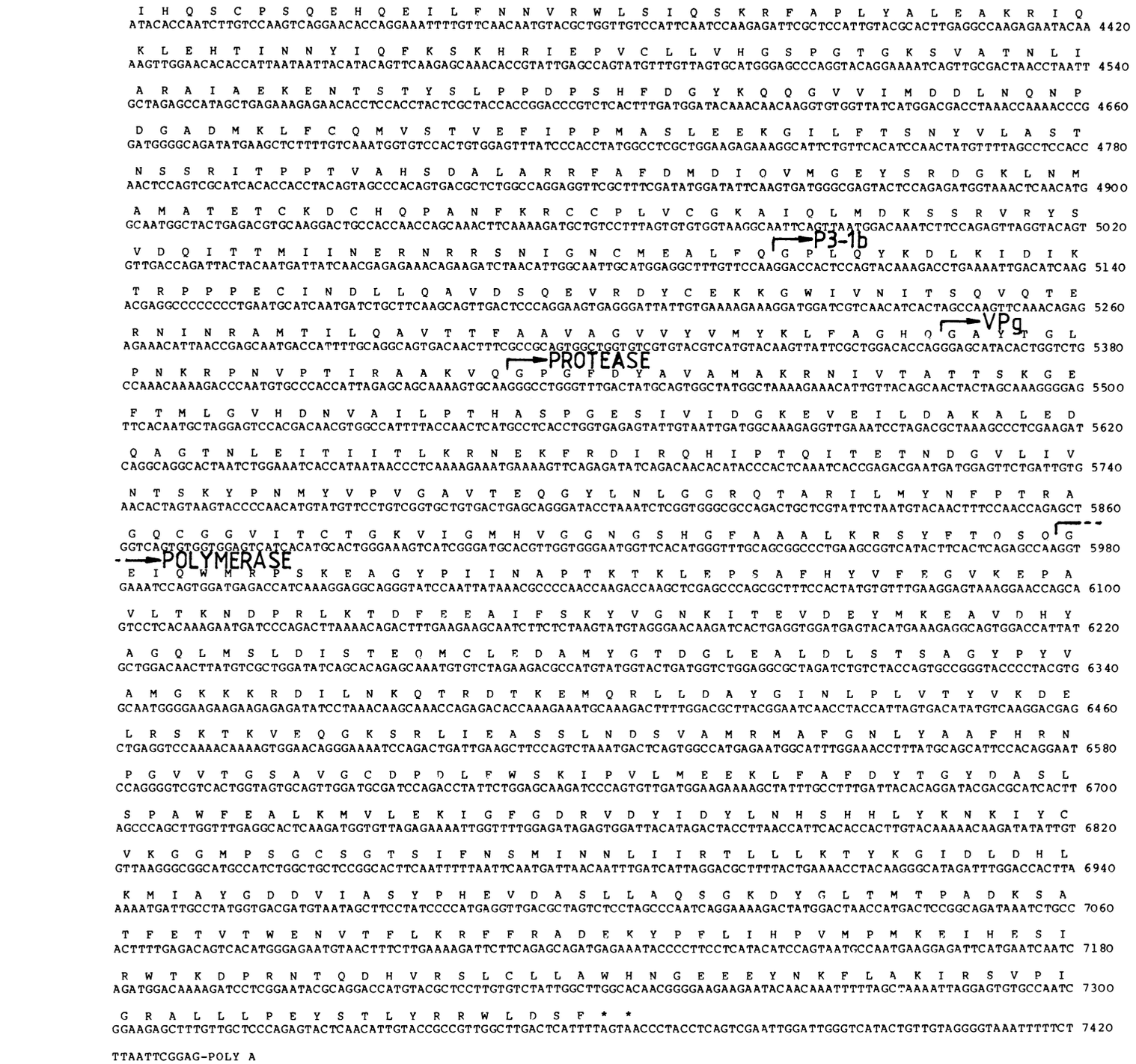


FIG. 1. Complete nucleotide sequence of poliovirus type 3 P3/Leon/37. A cloned DNA copy of the genome was subjected to random shearing by sonication and was subcloned into bacteriophage M13. Nucleotide sequence was determined by the dideoxynucleotide method (13, 14).

RESULTS

To safeguard against the possibility that an individual, cloned DNA copy of the genome could be derived from a subpopulation unrepresentative of the biological properties of the original virus stock, P3/Leon/37 was plaque-purified, and its highly neurovirulent phenotype was confirmed in monkeys prior to gene cloning (10). The vaccine strain, P3/Leon $12a_1b$ plaque isolate 411, was similarly tested and shown to be attenuated. Therefore, it is highly likely that nucleotide sequences derived from cloned cDNAs of these strains reflect their neurovirulence properties. In addition, the close similarity in serological and biochemical properties of the strains, including the identical migration of 55 characteristic T_1-oligonucleotides (10), suggest that neutral mutations in the genomes are minimal and, therefore, that sequence differences have a high probability of correlating with biological differences.

A complete DNA copy of the genome of P3/Leon/37, plaque-isolate 960, was subjected to random shearing by sonication and subcloned into bacteriophage M13 (14). Overlapping sequences were determined by the dideoxynucleotide chain-termination method (13) and together comprise the complete sequence presented in Fig. 1. This 7431-nucleotide sequence excluding the poly(A) tract is 1 base shorter

than that derived from the vaccine strain P3/Leon $12a_1b$ and differs from it at only 10 positions as summarized in Table 1. The level of sequence conservation between the two strains is surprising in view of the number of passages between them [53 *in vitro* and 21 *in vivo* (2)] and current ideas on the high mutability of RNA genomes [one error per 10^3–10^4 incorporated nucleotides (16)]. The result contrasts with that obtained in a study comparing the poliovirus type 1 strains, Sabin vaccine P1/LSc,2ab, and its neurovirulent precursor P1/Mahoney (9). Between these strains 57 point mutation differences were observed, 21 of which gave rise to amino acid changes scattered throughout the genome. It should be noted that in the case of type 1, there is no reported plaque-purification and neurovirulence testing of strains immediately prior to molecular cloning and/or nucleotide sequence determination (9, 17, 18).

Although the predicted amino acid sequences of poliovirus types 1 and 3 show close homology (*ca.* 90%) (11), none of the mutations observed in the type 3 vaccine strain have an identical counterpart in the type 1 vaccine strain, indicating that the mutational events responsible for attenuation are different in these two serotypes.

Of the 10 changes listed in Table 1, 5 are unlikely *per se* to be of significance to the attenuated phenotype. Thus, the changes at 871, 4064, 6127, and 7165 are all silent and fall within the large open-reading frame believed to encode all the known virus-specified polypeptides (17, 18). Although it is possible that these changes could influence biological properties by destabilizing RNA secondary structure, we have no evidence for the involvement of these regions in secondary structure of any significance. The change at position 7432 (A → G) corresponds in P3/Leon/37 to the first A of the poly(A) tract. Variation at this position has been observed in poliovirus type 1 between two isolates of the same strain, both of which are believed to be neurovirulent (17, 19).

The significance of sequence changes at positions 220 and 472 is difficult to assess. This 5′ region of the genome, which is presumed to be noncoding (17), shows extensive sequence conservation between types 1 and 3, suggesting that it plays a critical role in the replicative cycle of the virus (11). However, its precise function and the constraints this imposes on sequence changes within the region remain obscure. It is unlikely that position 220 is involved in a translational function because there are no AUG codons in the preceding sequence (11). Position 472, on the other hand, could function in two short open-reading frames. In one of these, positions 461–595, the C → T mutation would be silent. In the second, positions 322–519, the mutation would give rise to a Ser → Pro amino acid substitution. However, this second frame is not conserved between poliovirus types 1 and 3 and, therefore, is of doubtful significance (11). Indeed, at the present time, there is no evidence for the use of any translational reading frames other than that encoding NCVPOO (17, 19).

Table 1. Differences in nucleotide sequence between the genomes of poliovirus type 3 P3/Leon/37 and the Sabin vaccine derivative P3/Leon $12a_1b$ (2, 11)

	Change from parent → vaccine		
Position	Nucleotide	Amino acid	Altered protein
220	G → T	—	—
472	C → T	?	—
871	G → A	—	—
2034	C → T	ser → phe	VP3
3333	A → G	lys → arg	VP1
3464	A → G	thr → ala	P2-3b
4064	T → C	—	—
6127	T → C	—	—
7165	G → A	—	—
7432	A → G	—	—

The mutations most likely to determine the attenuated phenotype are those at positions 2034, 3333, and 3464, which give rise to amino acid substitutions in VP3, VP1, and nonstructural protein P2-3b, respectively. The most chemically conservative of these changes is the Lys → Arg change in VP1. We have argued previously that this change is unlikely to be the major determinant of attenuation, mainly on the evidence that a neurovirulent revertant of the Sabin type 3 vaccine does not show a back-mutation at this point (10). Although our arguments are still valid, the observation of only two additional coding changes in the entire genome suggests that involvement of the Lys → Arg change in attenuation should not be totally ruled out. Indeed, it has been argued previously that surface changes may play an important role in attenuation; thus, VP1 is a likely site for attenuating mutations (9, 20). A similar case may be made for the structurally most drastic of the three coding changes (i.e., Ser → Phe in VP3). It is conceivable that this change could modify the surface properties of the virion in a way that would reduce its neurotropism.

The final coding change, Thr → Ala (position 3464), causes an alteration in the nonstructural protein P2-3b. This protein is the precursor of two other nonstructural proteins observed in poliovirus-infected cells, P2-5b and P2-X (17). However, the region of P2-3b incorporating the change is removed upon formation of these derivatives. It is noteworthy that this change is juxtaposed between two mutations that are 12 amino acids apart in the Sabin type 1 vaccine strain (i.e., Asp → Glu at 3460 and Ser → Asn at 3492 of that strain) (9). Therefore, these mutations may affect the same structural domain of P2-3b and, thereby, possibly represent a common basis of attenuation in the two serotypes. However, an explanation of the molecular mechanism would necessitate an understanding of the function of P2-3b.

DISCUSSION

The nucleotide sequence of P3/Leon/37 presented here indicates that only 10 mutations have occurred during the derivation of the Sabin type 3 vaccine strain, P3/Leon $12a_1b$. Unambiguous identification of the mutations responsible for the attenuated phenotype is not yet possible, but it is likely that one or more of the three coding changes are involved.

It might be predicted, considering the extensive passage regime used to derive the Sabin poliovirus vaccines, that attenuation would be the result of an accumulation of mutations acting concertedly (2). The number of mutations observed in the type 1 vaccine and the observed genetic stability of this strain favor such a conclusion (3, 9). For serotype 3, however, the sparsity of the mutations and their disparate locations raise the possibility that attenuation has been effected by a single point mutation. The observed frequency of reversion of the type 3 vaccine (at least 10 times greater than that of type 1) (3, 21, 22) is consistent with this conclusion. Profound biological effects resulting from single point mutations have been observed in other viruses. In the case of rabies virus, a single amino acid substitution in the surface glycoprotein can dramatically reduce neurovirulence (23). In influenza viruses, a single change may affect virus interaction with the cell surface (24).

We have reported the partial nucleotide sequence of revertants of the type 3 vaccine in an attempt to shed further light on the attenuating change(s) (10). It is likely that a parallel genetic approach based on the construction of recombinants from cDNA *in vitro* will also be required (25). A precise definition of attenuating mutations in Sabin vaccine types 1, 2, and 3 and an explanation of their mechanism of action may raise the possibility of designing derivatives that are incapable of reverting to neurovirulence.

Finally, it is worth noting that the sequence of P3/Leon/37 presented here and that of P3/Leon12a$_1$b presented previously (11) together constitute one of the few examples where total RNA sequences have been obtained from RNA genomes separated by a known number of passages (although the number of replication cycles within one passage and the number of replications of a genome within one infected cell remain unknown). The level of change observed here suggests either that the vast majority of mutations are disadvantageous or that the spontaneous mutation rate in these RNA genomes is significantly lower than previously supposed (16).

We thank Dr. A. T. Bankier for advice on nucleotide sequencing, Roger Key for assistance with computing, and Alan J. Cann for helpful discussion. This work was supported by the Medical Research Council of Great Britain (Project Grant G80/0815/2CA).

1. World Health Organization (1981) *J. Biol. Stand.* **9**, 163–184.
2. Sabin, A. & Boulger, L. (1973) *J. Biol. Stand.* **1**, 115–118.
3. World Health Organization (1983) *Bull. W.H.O.* **60**, 231–242.
4. Cossart, Y. E. (1977) *Br. Med. J.* **1**, 1621–1623.
5. Nottay, B. K., Kew, O. M., Hatch, H. M., Heyward, J. T. & Obijeski, J. F. (1981) *Virology* **108**, 405–423.
6. Kew, O. M., Nottay, B. K., Hatch, M. H., Nakano, J. H. & Obijeski, J. F. (1981) *J. Gen. Virol.* **56**, 337–347.
7. Minor, P. D. (1982) *J. Gen. Virol.* **59**, 307–317.
8. Scaba, D. G. (1979) in *The Molecular Biology of Picorna Viruses*, ed. Perez-Bercoff, R. (Plenum, New York), pp. 1–24.
9. Nomoto, A., Omata, T., Toyoda, H., Kuge, S., Horie, H., Kataoka, Y., Genba, Y., Nakano, Y. & Imura, N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5793–5797.
10. Stanway, G., Cann, A. J., Hauptmann, R., Mountford, R. C., Clarke, L. D., Reeve, P., Minor, P. D., Schild, G. C. & Almond, J. W. (1983) *Eur. J. Biochem.* **135**, 529–533.
11. Stanway, G., Cann, A. J., Hauptmann, R., Hughes, P., Clarke, L. D., Mountford, R. C., Minor, P. D., Schild, G. C. & Almond, J. W. (1983) *Nucleic Acids Res.* **11**, 5629–5643.
12. Cann, A. J., Stanway, G., Hauptmann, R., Minor, P. D., Schild, G. C., Clarke, L. D., Mountford, R. C. & Almond, J. W. (1983) *Nucleic Acids Res.* **11**, 1267–1281.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Deininger, P. (1983) *Anal. Biochem.* **129**, 216–223.
15. Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694.
16. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982) *Science* **215**, 1577–1584.
17. Kitamura, N., Semler, B., Rothberg, P. G., Larsen, G. R., Adler, C. J., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, J. J., van der Werf, S., Anderson, C. W. & Wimmer, E. (1981) *Nature (London)* **291**, 547–553.
18. Dorner, A. J., Dorner, L. F., Larsen, G. R., Wimmer, E. & Anderson, C. W. (1982) *J. Virol.* **42**, 1017–1028.
19. Racaniello, V. R. & Baltimore, D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4887–4891.
20. Wetz, K. & Habermehl, K. O. (1979) *J. Gen. Virol.* **44**, 525–534.
21. Boulger, L. R., Marsden, S. A., McGrath, D. I., Taffs, L. F. & Schild, G. C. (1979) *J. Biol. Stand.* **7**, 97–111.
22. Marsden, S. A., Boulger, L. A., McGrath, D. I., Reeve, P. & Taffs, L. F. (1980) *J. Biol. Stand.* **8**, 303–309.
23. Dietzschold, B., Wunner, W. H., Wiktor, T. J., Lopes, A. D., Lafon, M., Smith, C. L. & Koprowski, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 70–74.
24. Rogers, G. N., Paulson, J. C., Daniels, R. S., Skehel, J. J., Wilson, I. A. & Wiley, D. C. (1983) *Nature (London)* **304**, 76–78.
25. Racaniello, V. R. & Baltimore, D. (1981) *Science* **21**, 916–919.