Arch Virol (1988) 102: 131–139

Archives of Virology

Crossover regions in foot-and-mouth disease virus (FMDV) recombinants correspond to regions of high local secondary structure

Brief Report

Valerie Wilson*, **P. Taylor****, and **U. Desselberger*****

Institute of Virology, University of Glasgow, Glasgow, Scotland

Accepted July 5, 1988

Summary. The RNA genome of foot-and-mouth disease virus (FMDV) was analysed for the degree of inverted complementarity and thus potential secondary structure using the procedure of Pustell and Kafatos [Nucleic Acids Res (1982) 10: 4765–4782]. Regions of crossover in 42 FMDV recombinants [King et al. (1985) Virus Res 3: 373–384; Saunders et al. (1985) J Virol 56: 921–929] and regions lacking crossovers were assigned an average secondary structure score against which the number of observed recombinants was plotted. In general it was found that the mean value of potential secondary structure is significantly higher in crossover zones than in recombination-free zones. Recombination increased much more steeply with increasing secondary structure in the part of the genome coding for non-structural proteins than in the 5′ third of the genome coding for structural proteins.

*

Homologous recombination in viruses with non-segmented RNA genomes was recognized in early studies using genetic markers [9, 18, 25], and a considerable amount of genetic evidence for recombination of picornaviruses had accumulated [4, 5, 13, 21] before its existence was confirmed biochemically by protein and oligonucleotide separation techniques [14, 28, 29, 31] and more recently by genome sequence data [15, 27, 32]. As well as in picornaviruses, recombination has recently been found to occur in coronaviruses [10, 19] and in the genome of a multipartite plant virus [2].

* Present address: Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, England.

** Present address: Regional Virus Laboratory, East Birmingham Hospital, Birmingham B9 5ST, England.

*** Member of the MRC Virology Unit.

In theory there are two ways in which RNA can recombine, either by a break-rejoin mechanism as occurs during splicing of cellular and viral messenger RNAs [30] and in self-splicing reactions [3, 33, 34] or by a copy-choice mechanism where the RNA polymerase terminates prematurely and "jumps" to another site where it reinitiates replication. Kirkegaard and Baltimore [15] have recently provided strong evidence for a copy-choice mechanism in poliovirus recombination. Romanova et al. [27] have suggested that recombination occurs at higher frequency in regions of higher secondary structure. They proposed that such regions help to bring together homologous regions of two recombining genomes via formation of intermolecular duplexes enabling the migrating polymerase to "jump" onto the nucleotide chain of the second molecule. Transiently formed secondary structures have also been implicated in DI RNA formation where a nucleotide sequence complementary to sequences adjacent to the 5′ and 3′ ends of the deleted portion could hold together sequences long enough for the RNA polymerase to "jump" ("supporting sequence loop model") [17]. The mechanism of the formation of poliovirus recombinants might also be related to rearrangement events in viroids [11] but can be distinguished from the splicing events mentioned above [30] and also from priming of viral mRNA synthesis by heterologous and heterogeneous mRNA molecules [16, 24].

No detailed models have yet been proposed for recombination of foot-and-mouth disease virus (FMDV). As this virus has a genome structure similar to poliovirus and also shares other characteristics, a copy-choice mechanism of recombination should be considered. The nucleotide sequence of 94 percent of the genomic RNA of FMDV has been determined by Forss et al. [7]; this sequence spans the RNA between the 5′ proximal poly (C) tract and the 3′ terminal poly (A) tail. FMDV recombinants have been extensively characterized since Pringle's [25] original observation, and McCahon's group has recently mapped the crossover sites of 17 classes of FMDV recombinants using biochemical techniques [14, 21, 22, 29]. In their analysis of oligonucleotides of recombinants Saunders et al. [29] used the numbering system of Forss et al. [7], and we have maintained it for determining the boundaries of zones in which recombination events had been observed. Although there were crossover sites found virtually in all parts of the genome they did not seem to be evenly distributed and while this may be an artifact due to the selection systems used to isolate the recombinants, it was tempting to hypothesize that—as in the collection of poliovirus recombinants of Romanova et al. [27] and more recently of Tolskaya et al. [32]—there might be "hotspots" and "coldspots" of recombination which could possibly correspond to regions of high and low secondary structure, respectively.

To investigate this hypothesis, the first step was to use a dot matrix computer program taken from "RINTHOM" [26] to plot the degree of inverted complementarity, and thus of possible secondary structure, along the genome. Successive stretches of 21 nucleotides were checked at a minimum score level of 62 percent matching, and the matrix obtained in this way is shown in Fig. 1.

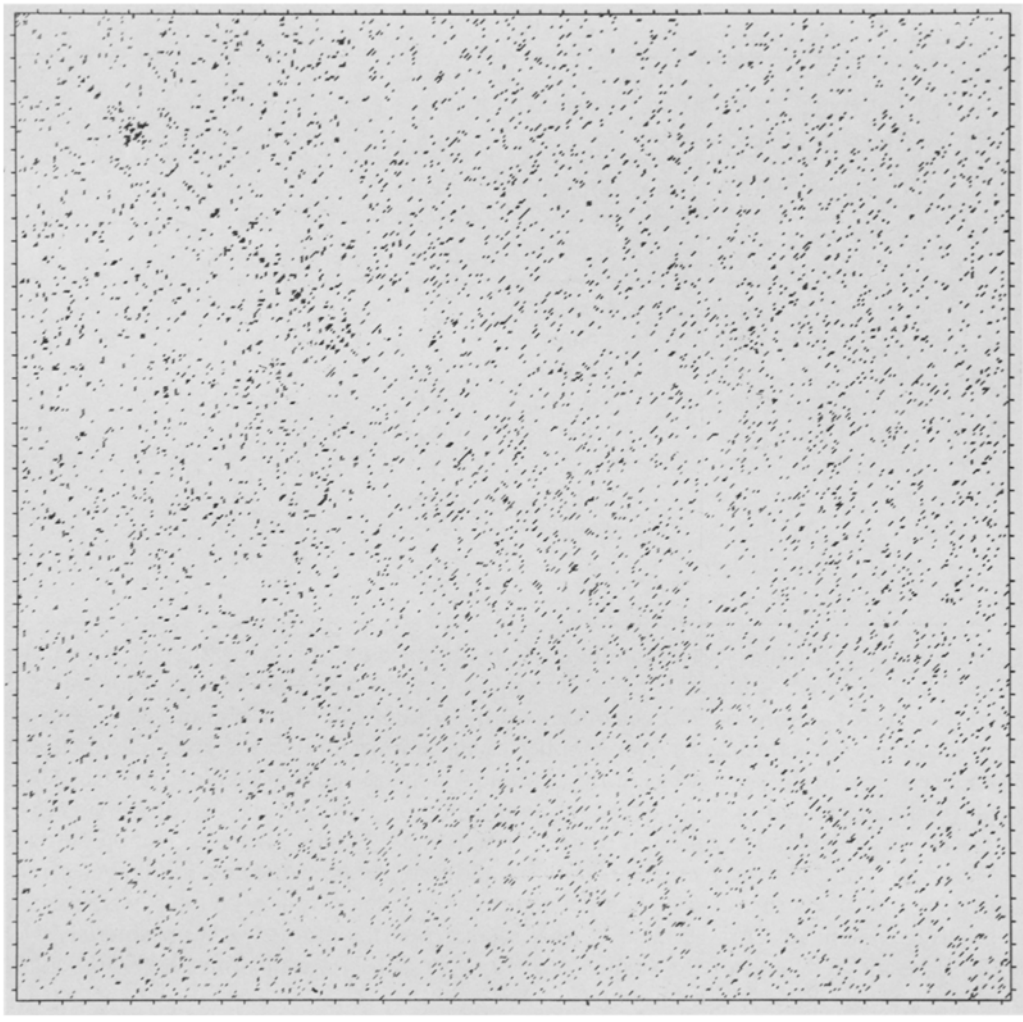

Fig. 1. Reversed homology matrix plot of the entire FMDV genome sequence [7] on the horizontal axis compared with its complement on the vertical axis. The program, modified from RINTHOM as described by Pustell and Kafatos [26], was run on a PDP11 computer with the following settings: Window size: 21 nucleotides; Scale: 0.95; Minimum score: 62; compression: 18. A dot thus indicates a match of above 62 percent. X axis and Y axis: Nucleotide sequence of FMDV from base no. 1 to base no. 7,804; respectively

Several clusters of reverse homology are visible along the main diagonal. This implies that the FMDV genome has the potential to form short-range stem/loops and hairpins in numerous but defined regions. (Increase of the minimum score level to 66% greatly reduced the background but still left clusters of homology along the diagonal; results not shown).

Next, in order to compare the degree of potential secondary structure along the genome with the frequency of crossover sites and events, the following approach was taken.

As the recombination zones analysed [14, 21, 29] are of different size, new programs were written to sum directly all secondary structure scores above a

Table 1.

Zone	Nucleotide interval	Number of nucleotides	Secondary structure score		Number of recombinants
			total per interval	per nucleotide in interval	
1	2	3	4	5	6
1	1– 700	700	2,006	2.87	0
2	701–1,600	900	5,080	5.64	2 (16, 17[a])
3	1,601–2,000	400	896	2.24	0
4	2,001–2,500	500	3,019	6.04	4 (9, 10, 11)
5	2,501–3,700	1,200	1,909	1.59	0
6	3,701–5,100	1,400	6,059	4.33	16 (1, 3, 5, 8, 12, 13[a], 15[a])
7	5,101–5,300	200	61	0.30	0
8	5,301–6,300	1,000	2,708	2.71	9 (2, 4)
9	6,301–7,500	1,200	6,499	5.42	11 (6, 7, 13[b], 14, 15[b], 17[b])
10	7,501–7,804	304	611	2.01	0

Total secondary structure scores per interval were obtained by adding together all percentage matches above 62 percent occurring in the zones of length indicated by nucleotide interval along the diagonal, and of 50 nucleotides width to each side. In Column 5 the ratios of total secondary structure score over number of nucleotides are given. The frequencies of crossover events in different areas of the genome which are summarized in Table 1 of King et al. [14] and defined in more detail by Saunders et al. [29] are indicated in Column 6. Zones of crossovers in the different classes were defined by oligonucleotide mapping data of Saunders et al. [29] and by protein analyses and were between 500 and 1,400 nucleotides long (Column 3). As the oligonucleotides were numbered by Saunders et al. [29] using the numbering system of Forss et al. [7], this system was maintained. (Approximately 550 nucleotides at the 5′ end of the sequence of Forss et al. [7] are missing; the exact length of that part of genome is not known)

[a] Left-hand (5′ end) crossover

[b] Right-hand (3′ end) crossover

threshold value of 62 percent in successive groups of 100 nucleotides length and 50 nucleotides width to each side of the main diagonal. These scores were then added together for each recombination zone, and the sums divided by the total number of nucleotides to yield an average secondary structure score per nucleotide for each different zone. The results are listed in Table 1. From this in general it can be deduced that lack of crossover sites correlates with a low secondary structure score per nucleotide in the zone (0.30 to 2.87) whereas the average secondary structure score per nucleotide is between 2.71 and 6.04 in recombination zones.

In order to investigate whether or not there is a quantifiable relationship between the secondary structure score per nucleotide and the number of crossovers per zone, the values were plotted against each other as shown in Fig. 2. When *all* paired values were subjected to the analysis of linear regression, no significant correlation was found ($r = 0.490$; $P > 0.1$; results not shown).

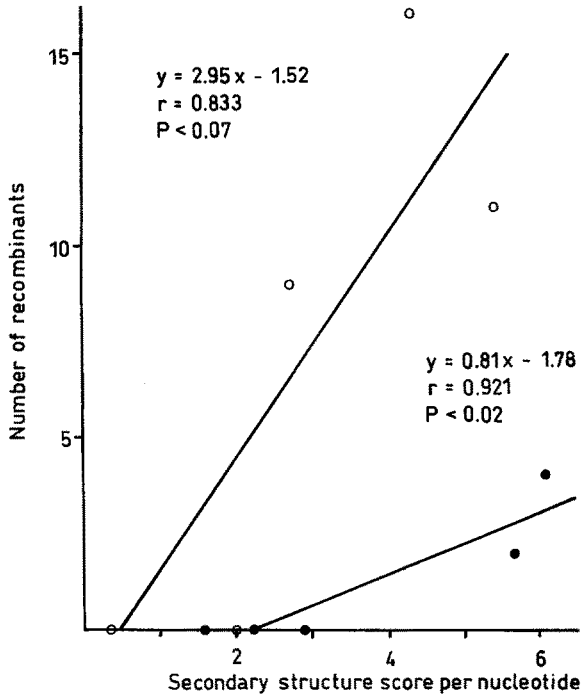

Fig. 2 Plot of number of recombinants against the secondary structure score per nucleotide of zones of recombination as indicated in Table 1. ● Paired values of zones 1–5, ○ paired values of zones 6–10. The straight lines fitting best the paired values of zones 1–5 and 6–10 were calculated by linear regression analysis. The slopes, correlation coefficients (r), and P values of significance are indicated

However, from considering Table 1 it is obvious that there are many more recombinants mapping in the 3′ two thirds of the genome (N = 36) than there are recombinants mapping to the 5′ third (N = 6) in spite of a high secondary structure score per nucleotide in the two recombination zones nearest the 5′ end. It occurred to us that there may be more constraints for recombination in the part of the genome coding for the structural proteins (nucleotide positions 1,365–3,524) [7]. Therefore we performed linear regression analysis separately for paired values of zones 1–5 (nucleotides 1–3,700) and 6–10 (nucleotides 3,701–7,804). The correlation coefficients for the two groups of paired values were 0.921 and 0.833, respectively, being significant for $P < 0.02$ and $P < 0.07$. Thus the linear regression line for the 3′ end recombination zone is of borderline significance, only. However, the mean value of potential secondary structure per nucleotide is significantly higher in crossover zones than in recombination-free zones of both groups [crossover zones of 5′ end group: 5.84 ± 0.20 (N = 2), recombination-free zones of 5′ end group: 2.23 ± 0.52 (N = 3), different (t-test) for $P < 0.001$; crossover zones of 3′ end group: 4.15 ± 1.11 (N = 3), recombination-free zones of 3′ end group; 1.15 ± 0.86 (N = 2), different (t-test) for $P < 0.05$]. The slopes of the two straight lines (Fig. 2) were 0.811 and 2.940, respectively, indicating that recombination in the part of the genome coding

for non-structural proteins (nucleotides 3,525–7,709) [7] increased much more steeply with potential secondary structure of a zone. This foregoing analysis was hampered by the fact that the number of isolated recombinants was limited (N = 42) and that the limits of crossover sites were very large in some cases. It remains to be seen whether or not the results of this analysis hold when more precise and more extensive data become available.

Data by Currey et al. [6] have provided evidence that there is good agreement between secondary structure of psoralen-crosslinked plus strand RNA of poliovirus as visualized by electron microscopy and computer-predicted maps of secondary structure. The copy choice mechanism as developed by Romanova et al. [27] requires that the RNA polymerase will pause, probably at regions of high secondary structure, and that this will free the enzyme to which the nascent RNA is attached. The in vitro products of poliovirus polymerase include a ladder of free "strong-stop" molecules whose 3′ ends may correspond to regions of high secondary structure [8]. Similarly, in coronavirus-infected cells free nascent RNAs have been detected the 3′ ends of which correspond to the regions predicted to be of higher secondary structure [1, 19]. Thus, in order to construct a model for RNA recombination using the copy-choice mechanism, the proximity of the parental genomes is not a necessary prerequisite. Kirkegaard and Baltimore [15] conclude from their data that poliovirus RNA recombination results from a template switch of the viral polymerase during negative strand synthesis, i.e. early in infection, and that it is not particularly site specific within a small range of 190 nucleotides: "The apparent preference of the polioviral replicase to switch templates during negative strand rather than positive strand synthesis results from different structures or availabilities of the templates for negative and positive strand synthesis during replication" [15]. For FMDV it has also been found that recombination events between strains occur early in infection [23].

The secondary structure analysis of poliovirion RNA by Currey et al. [6] revealed a higher degree of secondary structure towards the 5′ end. If the polymerase merely paused at any region containing a higher degree of secondary structure, the frequency of recombination would be expected to be higher towards the 5′ end. In the analysis of FMDV recombination frequency it was certainly found that a relatively high degree of potential secondary structure of the RNA near the 5′ end is not correlated with the observation of a particularly high number of recombinants in this area, but this observation may merely be due to negative selection.

Simple pausing of the polymerase at secondary stem/loops and subsequent dissociation of nascent RNA into the cytoplasm followed by reinitiation of synthesis on another template may not be the only factor determining recombination along the FMDV genome. The model by Romanova et al. [27] predicts not only pausing of the polymerase at stem/loop structures but also opening of hairpins to allow the recombining parents to pair in opposite orientations and bring together the recombination sites. We have indicated that secondary

structure may be important in recombination but this does not necessarily argue for the version of the copy-choice model of Romanova et al. [27]. It could be assumed that secondary structure of the template is only important to stop the polymerase.

The model of Romanova et al. [27] predicts that short range rather than more complex secondary structure should be more important for intergenomic pairing for the following reasons:

— A simple hairpin will open out to give a higher number of intergenomic base pairs than a branched hairpin structure with the same number of original base pairs.

— A hairpin the base of which is at a highly branched structure is less likely to hold a recombination site than one which is relatively free of this structure as the recombination sites in the intergenomic duplex are no longer in proximity.

It would seem that the evidence of the present analysis supports the model by Romanova et al. [27] for FMDV recombination, since our results only related to possible 21-nucleotide long hairpins or stem-loops and up to 50 nucleotides separating them. When we increased the width of search for inverted complementarity along the diagonal to 90 nucleotides on each side (thus effectively increasing the complexity of potential secondary structure) the confidence limits of significance in relation of crossover sites to secondary structure dropped (results not shown). However, neither the precise sequences of the crossover regions nor the specific secondary structures at the recombination sites are available yet nor do we know the secondary structure of native FMDV RNA; thus a detailed correlation of particular secondary structure of FMDV RNA with recombination sites is not possible at present.

Recently, Tolskaya et al. [32] have precisely determined the crossover sites of 15 intertypic (type 3/type 1) poliovirus recombinants by the primer extension method. Recombination occurred in the middle of the genomes over a distance of appr. 1200 nucleotides. It was found that all sites of recombination were within potential hairpin-like structures and compatible with the notion of the importance of secondary structure in the copy choice mechanism of poliovirus recombination [27]. The non-randomness of distribution of recombination sites observed by Tolskaya et al. [32] could be attributable in part to the requirement of secondary structure but also to the possibility that recombination, although structurally permitted, may result in the generation of unviable progeny or of progeny with considerably lowered viability [32].

RNA recombination of picornaviruses is not only an academically fascinating subject of great theoretical interest but, besides being observed in vitro, may frequently occur in nature. Kew and Nottay [12] and Minor et al. [20] found recombination of different poliovirus types after vaccination of a child with a trivalent live attenuated vaccine. The significance of such events for the emergence of new, possibly pathogenic strains remains to be investigated.

Acknowledgements

The authors thank Dr. Andrew M. Q. King for extensive discussion, advice, and suggestions, Mr. Edward Tobias for critical and lively debate, Drs. Gareth Westrop and Richard M. Elliott for comments on the manuscript and Fiona Conway for the typing.

References

1. Baric RS, Shieh CK, Stohlman SA, Lai MMC (1987) Analysis of intracellular small RNAs of mouse hepatitis virus: evidence for discontinuous transcription. Virology 156: 342–354
2. Bujarski JJ, Kaesberg P (1986) Genetic recombination between RNA components of a multipartite plant virus. Nature 321: 528–531
3. Cech TR, Bass BL (1986) Biological catalysis by RNA. Annu Rev Biochem 55: 599–629
4. Cooper PD (1968) A genetic map of poliovirus temperature sensitive mutants. Virology 35: 584–596
5. Cooper PD (1977) Genetics of picornavirus. In: Fraenkel-Conrat H, Wagner R (eds) Comprehensive virology, vol 9. Plenum, New York, pp 133–207
6. Currey KM, Peterlin BM, Maizel JV Jr (1986) Secondary structure of poliovirus RNA: correlation of computer-predicted with electron microscopically observed structure. Virology 148: 33–46
7. Forss S, Strebel K, Beck E, Schaller H (1984) Nucleotide sequence and genome organization of foot-and-mouth disease virus. Nucleic Acids Res 12: 6587–6601
8. Hey TD, Richards OC, Ehrenfeld E (1986) Synthesis of plus- and minus-strand RNA from poliovirion RNA template in vitro. J Virol 58: 790–796
9. Hirst GK (1962) Genetic recombination with Newcastle disease virus, poliovirus and influenza. Cold Spring Harbor Symp Quant Biol 27: 303–308
10. Keck JG, Stohlman SA, Soe LH, Makino S, Lai, MMC (1987) Multiple recombination sites on the 5′-end of murine coronavirus RNA. Virology 156: 331–341
11. Keese P, Symons RH (1985) Domains in viroids: evidence of intermolecular RNA rearrangements and their contribution to viroid evolution. Proc Natl Acad Sci USA 82: 4582–4586
12. Kew OM, Nottay BK (1984) Evolution of the oral polio vaccines in humans occurs by both mutation and intermolecular recombination. In: Chanock R, Lerner R (eds) Modern approaches to vaccines. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 357–362
13. King AMQ, McCahon D, Slade WR, Newman JWI (1982) Recombination in RNA. Cell 29: 921–928
14. King AMQ, McCahon D, Saunders K, Newman JWI, Slade WR (1985) Multiple sites of recombination within the RNA genome foot-and-mouth disease virus. Virus Res 3: 373–384
15. Kirkegaard K, Baltimore D (1986) The mechanism of RNA recombination in poliovirus. Cell 47: 433–443
16. Krug RM (1985) The role of RNA priming in viral and trypanosomal mRNA synthesis. Cell 41: 651–652
17. Kuge S, Saito I, Nomoto A (1986) Primary structure of poliovirus defective-interfering particle genomes and possible generation mechanisms of the particles. J Mol Biol 192: 473–487
18. Ledinko N (1963) Genetic recombination with poliovirus type 1. Studies of crosses between a normal horse serum-resistant mutant and several guanidine-resistant mutants of the same strain. Virology 20: 107–119

19. Makino S, Keck JG, Stohlman SA, Lai MMC (1986) High-frequency RNA recombination of murine coronaviruses. J Virol 57: 729–737
20. Minor PD, John A, Ferguson M, Icenogle JP (1986) Antigenic and molecular evolution of the vaccine strain of type 3 poliovirus during the period of excretion by a primary vaccinee. J Gen Virol 67: 693–706
21. McCahon D (1986) The genetics of foot-and-mouth disease virus. Rev Sci Tech Office Int Epizooties 5: 279–297
22. McCahon D, King AMQ, Roe DS, Slade WR, Newman JWI, Cleary AM (1985) Isolation and biochemical characterization of intertypic recombinants of foot-and-mouth disease virus. Virus Res 3: 87–100
23. McCahon D, Slade WR (1981) A sensitive method for the detection and isolation of recombinants of foot-and-mouth disease virus. J Gen Virol 53: 333–342
24. Patterson JL, Holloway B, Kolakofski D (1984) La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. J Virol 52: 215–222
25. Pringle CR (1965) Evidence of genetic recombination in foot-and-mouth disease virus. Virology 25: 48–54
26. Pustell J, Kafatos FC (1982) A high speed, high capacity homology matrix: zooming through SV 40 and polyoma. Nucleic Acids Res 10: 4765–4782
27. Romanova LI, Blinov VM, Tolskaya EA, Victorova EG, Kolesnikova MS, Guseva EA, Agol VI (1986) The primary structure of crossover regions of intertypic poliovirus recombinations: a model of recombination between RNA genomes. Virology 155: 202–213
28. Romanova LI, Tolskaya EA, Kolesnikova MS, Agol VI (1980) Biochemical evidence for intertypic recombination of polioviruses. FEBS Lett 118: 109–112
29. Saunders K, King AMQ, McCahon D, Newman JWI, Slade WR, Forss S (1985) Recombination and oligonucleotide analysis of guanidine-resistant Foot-and-Mouth-Disease Virus mutants. J Virol 56: 921–929
30. Sharp PA (1985) On the origin of RNA splicing and introns. Cell 42: 397–400
31. Tolskaya EA, Romanova LI, Kolesnikova MS, Agol VI (1983) Intertypic recombination in poliovirus: genetic and biochemical studies. Virology 124: 121–132
32. Tolskaya EA, Romanova LI, Blinov VM, Victorova EG, Sinyakov AN, Kolesnikova MS, Agol VI (1987) Studies on the recombination between RNA genomes of poliovirus: the primary structure and nonrandom distribution of crossover regions in the genomes of intertypic poliovirus recombinants. Virology 161: 54–61
33. Zaug, AJ, Cech TR (1986) The intervening sequence of *Tetrahymena* is an enzyme. Science 231: 470–475
34. Zaug, AJ, Been MD, Cech TR (1986) The *Tetrahymena* ribozyme acts like an RNA restriction endonuclease. Nature 324: 429–433

Authors' address: U. Desselberger, Regional Virus Laboratory, East Birmingham Hospital, Birmingham B9 5ST, England.

Received June 26, 1988