

A pattern of partially homologous recombination in mouse L cells

(DNA-mediated gene transfer/thymidine kinase/pBR322/plasmid rescue)

RICHARD A. ANDERSON*, SHINGO KATO, AND R. DANIEL CAMERINI-OTERO

Molecular Genetics Section, Genetics and Biochemistry Branch, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, Building 10, Room 9D-15, National Institutes of Health, Bethesda, MD 20205

Communicated by Elizabeth F. Neufeld, September 19, 1983

ABSTRACT Herpes simplex virus thymidine kinase gene and pBR322 DNA (in large excess to the thymidine kinase gene) were introduced into mouse L cells by calcium phosphate DNA-mediated gene transfer. DNA fragments encompassing six junctions between the exogenous DNAs have been cloned and their nucleotide sequences determined. Analysis of these sequences has shown that stretches of partial homology involving from 20–50 base pairs are present near the points at which joining occurs between the donor molecules. The structure of the junction sequences suggests that the recombination event involves the alignment of the two donor DNA molecules at partially homologous regions followed by staggered cutting and joining. One donor molecule is always cut in the region of partial homology, while the second is cut at some distance that is a small multiple of 13.5 ± 0.5 base pairs away (at 0, 14, 27, 39, 41, and 54 base pairs). In the three junctions where the second cut is far from the region of homology, a 17- to 19-base-pair segment of DNA separates the donor sequences. In all cases the origin of this “filler” DNA appears to be oligonucleotides derived from pBR322.

Unlinked DNA fragments introduced into the same mammalian cells and integrated into cellular DNA are usually found linked to each other (1, 2). The mechanisms involved in the linking of the fragments to each other (or, for that matter, those pertaining to the insertion of the foreign DNA into the genome) are unknown. Previous studies relating to the joining of exogenous DNA sequences to each other have looked both at the linkage between homologous sequences and non-homologous fragments. That homologous recombination is the predominant mode of DNA joining after transfection is shown by the high frequency of recombination between co-transferred, overlapping gene fragments (3, 4) and by the formation of long head-to-tail tandem arrays of simian virus 40 (SV40) (5, 6) and plasmid sequences (2, 7). In these cases it has not been easy to determine the exact position of the recombinant joint. Although this has not been a problem in the study of the joining of nonhomologous viral DNA sequences, no pattern or mechanism, other than ligation, has yet been implicated in the linking of these DNAs (8–16). In particular, none of these studies has reported any local features in the DNA (including statistically significant stretches of local sequence homology) surrounding the joining events.

To elucidate the mechanism of the joining of DNA molecules we have used the herpes simplex virus (HSV) thymidine kinase gene fragment (*tk*) as our selectable gene and pBR322 plasmid DNA in large excess to the *tk* as the only other DNA added to our cells. These DNAs were chosen because the complete nucleotide sequence of each is known (ref. 17; see legend to Fig. 2). The use of the plasmid DNA has also enabled us to use the technique of plasmid rescue (2, 18) to recover the sequences that have been recombined in mammalian cells. We suggested in a previous communica-

tion (2) that, in the one junction examined, a small area of partial homology might be involved in the process that linked pBR322 to *tk*. Here we report the rescue of three new independent plasmids from two additional cell lines and the nucleotide sequence across all (six) the junctions between the exogenous DNAs obtained by us.

MATERIALS AND METHODS

DNA-Mediated Gene Transfer and Plasmid Rescue. The recipient Ltk^-aprt^- mouse cell line, the gene transfer protocol, the preparation and sources of the transforming DNAs, the cloning of the cell lines, and the preparation of high molecular weight mouse L-cell DNA have been described (2). Our method for carrying out plasmid rescues has also been published (2).

DNA Sequence and Analysis. Sequences were determined by using the method of Sanger *et al.* (19) (modified by using reverse transcriptase and deoxyinosine when necessary; unpublished observations) on appropriate fragments cloned into M13 bacteriophage vectors (20). Except for J5, all junction fragments were subjected to sequence analysis in both directions. “E” values were calculated by multiplying the two lengths of the sequences being compared and the probability values (*P*) for the random occurrence of the sequences. These *P* values were obtained from the SEQ program using the parameters AfterDis = 2 and LoopOut = 3 (21). In this case we searched 50 nucleotides on either side of each donor end so $E = P(100)^2$. $E = 0.03$ implies that 3% of comparisons of random sequences 100 base pairs (bp) long taken two at a time would be expected to yield a similar match. ΔG values were calculated by using the approximation of Tinoco *et al.* (22) and are expressed in kcal/mol.

RESULTS

Plasmid Rescue. Our experimental protocol involves the transfection of mouse Ltk^-aprt^- cells with the HSV *tk* fragment [3.5-kilobase-pair (kbp) *Bam*HI Q fragment of HSV-1 CL101 DNA containing the *tk*] and the use of pBR322 DNA (covalently closed supercoiled circles) as carrier DNA. Genomic DNA is isolated from cloned cell lines that have become stably transformed to the tk^+ phenotype. In most cases these DNAs contain 1 or 2 copies of *tk* and between 5 and 10 copies of pBR322 per haploid genome (2). The technique of plasmid rescue (2, 18) has enabled us to recover some of these recombined exogenous sequences in the form of plasmids that are capable of propagation in bacteria. In a typical plasmid rescue experiment 100 μ g of high molecular weight DNA yields about 10–20 plasmids, many of which are duplicates.

In all cases we were able to rescue only a subset of the sequences integrated into the cellular genome. A total of sev-

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.

Abbreviations: SV40, simian virus 40; *tk*, thymidine kinase gene fragment; HSV, herpes simplex virus; bp, base pair(s), kbp, kilobase pair(s).

*Present address: Department of Medicine, University of Iowa Hospitals and Clinics, Iowa City, IA 52242.

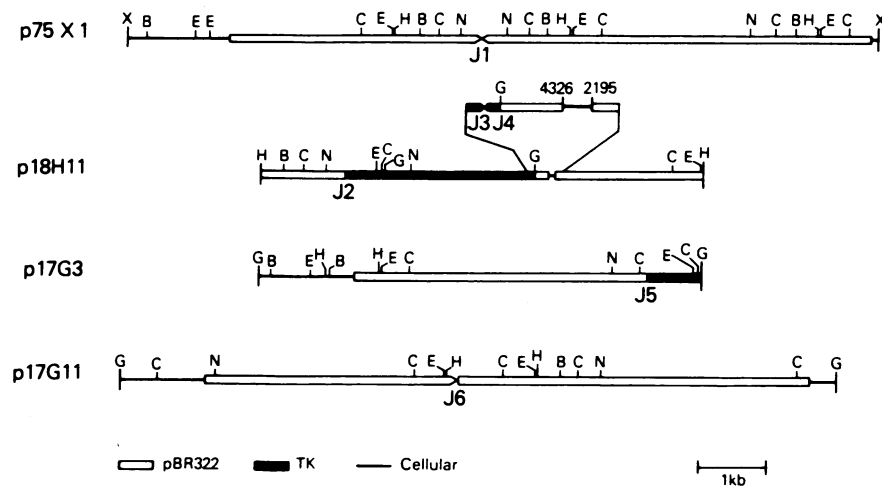


FIG. 1. Restriction endonuclease maps of the rescued plasmids. The linear plasmid maps start and end with the recognition sequences for the restriction endonuclease used in the rescue. These DNA fragments were rescued from the genomic DNA of transformant cell lines 1-2-75, 1-2-18, and 1-2-17. The identity of the cellular sequences in p75X1, p17G3, and p17G11 was confirmed by demonstrating their ability to hybridize with mouse lymphocyte DNA on Southern blots (unpublished results). In the case of p18H11 the sequence of the 70 bp of DNA shown in the inset does not resemble pBR322 or *tk* and is presumably cellular in origin. The junctions between fragments of pBR322 or *tk* are numbered J1–J6. The junctions involving cellular DNA will be discussed elsewhere. Abbreviations for the restriction endonucleases are: B, *Bam*HI; C, *Hinc*II; E, *Eco*RI; G, *Bgl* II; H, *Hind*III; N, *Nru* I; and X, *Xba* I. kb, Kilobase.

en plasmids were obtained from the genomic DNAs of three cloned tk^+ transformant lines. Of these, four plasmids contained six recombinant junctions (J1–J6) between different sequences of pBR322 or *tk* (Fig. 1). The plasmid p18H11 (Fig. 1) was rescued by using the restriction endonuclease *Hind*III and included junctions J2, J3, and J4. We have previously presented the structure of this plasmid at the level of restriction endonuclease mapping (2). The plasmid p75X1 (Fig. 1) was rescued by using the restriction enzyme *Xba* I from the DNA of a tk^+ cloned cell line obtained as the result of the cotransfer of the *tk* fragment and plasmid pde19 (a derivative of pBR322 deleted from nucleotide 1745 to 2505; unpublished results). p75X1 includes a head-to-head linkage between pBR322 sequences at J1 (Fig. 1), a tandem head-to-tail linkage of pBR322 sequences, and stretches of cellular DNA containing the two integration sites. Plasmids p17G3 and p17G11 (Fig. 1) were both rescued from another genomic DNA by using the restriction endonuclease *Bgl* II. Plasmid p17G3 is composed of a nearly full-length pBR322 sequence linked to the *tk* sequence (J5). Plasmid p17G11 contains a linkage of two different pBR322 segments in the same orientation (J6).

To confirm that the plasmids represent the structure of exogenous DNA as found incorporated in the eukaryotic genome, a series of double restriction endonuclease digests of plasmid and cellular DNA, always including the endonuclease used in the plasmid rescue, were compared by using Southern blotting techniques. In all cases all of the bands that hybridize to the pBR322 probe in each of the plasmid digests are present in the genomic digests (data not shown). This type of data for junctions J2, J3, and J4 in plasmid p18H11 was presented in much greater detail in an earlier paper (2).

Several additional lines of reasoning indicate that the linkages were the result of events in the mouse cell. First, J2, J4, and J5 involve the joining of pBR322 and *tk*. As the latter is present in only one or two copies in our tk^+ transformant mouse cells, it is highly unlikely that the covalent links were made in the *Escherichia coli*. Second, J3 and J4 surround a new *Bgl* II site (A-G-A-T-C-T at the *tk*-“filler” boundary in J4; see Fig. 2) that was created by the recombination event and that has previously been mapped to this location in the genome of the original mouse cell line (2). Third, filler DNA (see below), present in J1, J2, and J4, has not been described

in the joining of heterologous DNAs in bacteria (25–27).

Nucleotide Sequences of the Recombinant Junctions. Based on the restriction endonuclease maps shown in Fig. 1 we subcloned fragments encompassing the six DNA junctions, inserted them into M13 bacteriophage (20), and determined their nucleotide sequence (19). Because the DNA sequences of pBR322 (17) and the full *tk* fragment are known (see legend to Fig. 2), the location of the six junctions could be assigned without any ambiguity. In three of the six junctions, the donor molecules are not directly linked to each other but are joined via 17–19 bp of spacer or filler DNA; these sequences are underlined in the figure. Above the sequences of the recombinant junctions we show the sequences of the parental molecules. We have aligned them at regions of partial sequence homology shared by the donor molecules near the points of joining. Regions of partial homology are present near all of the six junctions that we have investigated. These homologies were found by a computer search using the SEQ program (21). In effect, this computer search allowed us to first align the nucleotide sequences of the intact donor molecules with their respective ends in the recombinant and then slide the donor sequences relative to each other until a significant partial homology was found. The alignment shown in Fig. 2 is the best found between the fragments for two stretches of 100 bp centered on the nucleotide denoting the boundaries of each donor. Except for junction J4, the matches have an expectation value, *E* (see *Materials and Methods*), of <0.05 . For J4 the value of *E* is higher; also there is another 19 out of 30 bp match between pBR322 at nucleotides 4068–4097 and *tk* at nucleotides 2929–2956 (but reverse orientation) that has a similar value of *E* (and ΔG) to the J4 alignment shown in Fig. 2. However, this alternate alignment would have neither of the recombination donor ends in the region of homology and would separate the cuts necessary for recombination by about 80 bp. In the context of the structure of the other alignments we believe that the arrangement portrayed in Fig. 2 for J4 is probably that which is significant.

As a measure of the physical significance of the homologies, we used the parameters of Tinoco *et al.* (22) to calculate the free energies of the postulated heteroduplex structures. The results tabulated in Fig. 2 show a wide range of values, but all are conducive to duplex formation.

In all of the recombinant junctions that we have examined,

with the exception of one base pair in J2, the nucleotide sequences of the donor molecules have been conserved and joined either to filler DNA or directly to each other without apparent scrambling. When the *tk* sequence near J2 in the recombinant molecule was compared to the sequence obtained by us for the *tk* fragment used for these experiments, one sees that, upstream from the junction with filler, a guanine residue has replaced the thymine residue originally present (both are overlined in Fig. 2). This replacement could be the result of either a mutation in *tk* (in our transformant mouse cells or during the bacterial propagation of p18H11) or a mismatch repair event in the heteroduplex implied in Fig. 2.

DISCUSSION

A Pattern of Recombination. In this study we have described the sequences of six recombinant junctions formed between two fully sequenced DNAs (pBR322 and *tk*) transferred into mouse L cells. An evident feature of half of the junctions is the presence of a 17- to 19-bp length of filler DNA separating the parental ends. Less evident, but clearly present, is a patchy homology occurring between the donor sequences near the joining points in all of the sequences. We propose that these sequence homologies are not only statistically and energetically significant but also biologically important because they are related to the points of joining by the pattern summarized in Fig. 3. At the left of Fig. 3 the donor molecules are represented as being aligned at the patchy homologies (as the sequences are aligned in Fig. 2). At the right of Fig. 3 a recombinant is diagrammed. Examining the data of Fig. 2 in relation to the pattern in Fig. 3, one observes that in all cases at least one donor sequence ends within the homology (*x* in Fig. 3); the other donor molecule is apparently cut at *y*, at a variable distance (*N*) away from the first cut. In all cases *N* is a small multiple (including the trivial zero multiple) of 13.5 ± 0.5 bp—that is, 0, 14, 27, 39, 41, and 54 bp (see Fig. 2). Once cuts are made on each donor, one in the homology and another at a multiple of 13 to 14 bp away, the ends are brought close to each other. Thus, in all cases, except when *N* is zero (J3), DNA is deleted. The deleted DNA can include part of the potential heteroduplex as well as flanking DNA. The 17- to 19-bp segment of filler DNA separates the donor sequences in those three junctions where the second donor DNA is cut far from the region of homology (J1, J2, and J4).

Although the reason for the pattern of cutting in multiples of a repeat of 13 to 14 bp is unknown, one mechanism we would like to suggest is the following. The first step would be the formation of an extended protein–nucleic acid complex that partially unwinds both of the donor DNAs from 10.6 bp [seen in B-form DNA (28)] to 13.5 ± 0.5 bp in each turn of the partially unwound helix. The DNA could then be particularly susceptible to nucleases at multiples of the helix repeat. This form of one-sided accessibility is common for DNAs bound to a matrix (28, 29). The prokaryotic recombination protein *recA* does form a complex with double-stranded DNA, in which the DNA is unwound to the extent that one helix turn encompasses 18.6 bp (30). A protein that seems to be required in recombination and has some similarities to *recA* has been isolated from the fungus *Ustilago maydis* (31).

Origin of the Filler DNA. DNA that is not obviously contributed by the donor sequences has been documented in many eukaryotic recombinant junctions. Up to 9 new bp are present in a number of immunoglobulin splice junctions (32, 33). Inserted sequences, usually 4–8 bp (12, 13, 15) but 40 bp in one case (10), have been found at viral recombinant joints. The origin of these new sequences is obscure.

In our junctions we have found DNA sequences identical to the filler DNA in regions of pBR322 DNA generally far removed from the region of joining. The most dramatic ex-

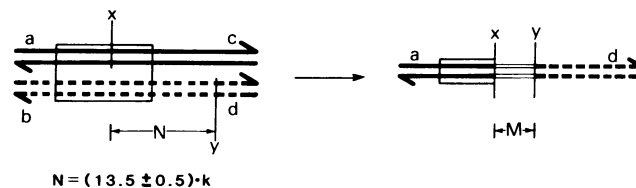


FIG. 3. A pattern of partially homologous recombination. Solid and dashed lines represent the strands of the two donor molecules. These contain segments a, c and b, d, respectively, in the parental configuration on the left. The structure on the right represents the recombinant product in which a and d have been joined. The box denotes a region of shared sequence homology. *x* and *y* mark points at which the donor DNA molecules are apparently cleaved and then joined. *N* represents the distance between the cuts at *x* and *y* in the aligned donor DNAs. The value of *N* is in all cases a small multiple ($k = 0, 1, 2, 3, \text{ or } 4$) of 13.5 ± 0.5 bp. *M* is the length of filler DNA. *M* is either 0 (no filler) or 17–19 in all cases examined.

ample of this is shown in Fig. 4 for J4, where a 17-bp sequence occurring 2,000 bp away on the pBR322 molecule from the junction sequence is identical to 15 bp of filler DNA and 2 flanking bp of the pBR322 donor. The random occurrence of a 17-bp perfect match between this filler DNA and the whole of pBR322 is a very low probability event ($E = 0.000015$; see *Materials and Methods*). That the two bases on the right of this 17-bp sequence match the two bases of the donor DNA highlights the obvious ambiguity in assigning an exact length to any of the fillers. These data taken together with that for the other filler DNAs in Fig. 4 suggest that all of the inserted sequences could actually be 19 bp in length and that they may arise from combinations of oligonucleotides derived from excess carrier DNA.

Although it is possible that the filler is the residue of more than one recombination event having occurred at these junctions, this appears unlikely because (i) the fillers are all virtually the same length and (ii) we have not found significant homology between the 100 bp of pBR322 sequence centered on the filler oligonucleotides (Fig. 4) and the 100 bp centered on the ends of the corresponding flanking donor sequences. Thus, our data suggest that fillers arise from the insertion of fragments of excess carrier DNA into a gap formed during a single joining event.

Relationship to Other Recombination Systems. A few investigators (11, 12, 15) have reported the existence of patchy sequence homologies involving stretches of 4–7 bp near the

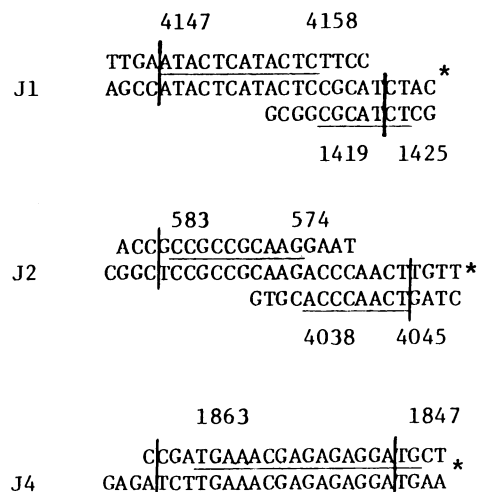


FIG. 4. The origin of filler DNA. Sequence across junction; sequences above and below represent segments of pBR322 sequence (17), near map positions indicated. Vertical lines delineate filler. Sequence matches are noted by horizontal lines.

joining points of exogenous, nonhomologous molecules in eukaryotic cells. In these reports the importance of these short sequences in recombination has been discounted because of the high probability of their random occurrence and because they have not been found in a reproducible relationship to the joints.

The obvious differences between our results and those of other investigators may have a number of origins. Our system differs from all of the systems studied in that we are not using virus-infected cells. Our experimental protocol differs in that we have no functional constraints on size or reading frame preservation that may select the recombination events that are seen by others. Also, we are looking at intermolecular recombination as opposed to the intramolecular deletion events examined in many other reports. Furthermore, our criteria for homologies between donors are less stringent than those in most reports, where only absolute matches are taken into account. In addition, the homologies near our junctions were seen in some cases only by sliding the alignment over a 50-bp range in either direction from the actual ends of the parental sequences.

We have re-examined some of the published junction sequences in light of our results and have found partially homologous patches on the order of those we have reported in some, but not all. However, the matches that we have found do not have the same relationship to the joining points that we see in the data reported here.

The occurrence of homologies on the order of those we have formed are not statistically rare events. Using the value of E of 0.4 for J4 (Fig. 2) we would expect to find a match like this between two random sequences 150 bp in length. For lower probability values, such as that for J6 ($E = 0.001$), we would expect such matches for two sequences 3,000 bp in length. To account for the relative genetic stability of mammalian cells we must suppose that the cellular genome is not a frequent substrate for the form of recombination we have studied. Whether this is because the mechanism implied here is highly regulated or because it is only used on foreign DNAs entering a cell is not clear. To determine whether a similar pattern of recombination is discernible in the junctions between exogenous DNA and cellular DNA we need to sequence the segments of unperturbed cellular DNA involved in the recombination events.

The pattern of recombination that we see here cannot be easily explained by the generally accepted models that have been proposed to account for the recombination events seen in fungi and in prokaryotes (34–36). Further insight into the generality of the events reported here may be gained from the use of novel substrates in similar types of experiments, but an understanding of the mechanisms will probably require the ability to re-create these events under defined conditions using partially purified cellular extracts.

We are grateful to W. Summers for allowing us to use his results prior to publication. We thank J. Maizel and D. Lipman for their help with the homology search computer programs, D. Bick and J. Kerpelman for their excellent technical help, J. McGhee, K. Mizuuchi, D. Brenner, and G. Grimaldi for valuable advice and discussion, and J. Glass for her help in the preparation of the manuscript.

1. Perucho, M., Hanahan, D. & Wigler, M. (1980) *Cell* **22**, 309–317.
2. Anderson, R. A., Krakauer, T. & Camerini-Otero, R. D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2748–2752.
3. De Saint Vincent, B. R. & Wahl, G. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2002–2006.
4. Small, J. & Scangos, G. (1983) *Science* **219**, 174–176.
5. Botchan, M., Topp, W. & Sambrook, J. (1976) *Cell* **9**, 269–287.
6. Chia, W. & Rigby, P. W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6638–6642.
7. Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. R. (1982) *Mol. Cell. Biol.* **2**, 1372–1387.
8. Zain, B. S. & Roberts, R. J. (1978) *J. Mol. Biol.* **120**, 13–31.
9. Wakamiya, T., McCutchan, T., Rosenberg, M. & Singer, M. (1979) *J. Biol. Chem.* **254**, 3584–3591.
10. Botchan, M., Stringer, J., Mitchinson, T. & Sambrook, J. (1980) *Cell* **20**, 143–152.
11. Gutai, M. W. (1981) *Virology* **109**, 344–352.
12. Gutai, M. W. (1981) *Virology* **109**, 353–365.
13. Ling, L. E., Manos, M. M. & Gluzman, Y. (1982) *Nucleic Acids Res.* **10**, 8099–8112.
14. Westin, G., Visser, L., Zabielski, J., van Mansfield, A. D. M., Pettersson, U. & Rozijn, Th. H. (1982) *Gene* **17**, 263–270.
15. Wilson, J. H., Berget, P. B. & Pipas, J. M. (1982) *Mol. Cell. Biol.* **2**, 1258–1269.
16. Johnson, A. D., Barkan, A. & Mertz, J. E. (1982) *Virology* **123**, 464–469.
17. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
18. Perucho, M., Hanahan, D., Lipsich, L. & Wigler, M. (1980) *Nature (London)* **285**, 207–210.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Messing, J., Crea, R. & Seeburg, P. (1981) *Nucleic Acids Res.* **9**, 310–321.
21. Brutlag, D. L., Clayton, J., Friedland, P. & Kedes, L. H. (1982) *Nucleic Acids Res.* **10**, 279–294.
22. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246**, 40–41.
23. Wagner, M. J., Sharp, J. A. & Summers, W. C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1441–1445.
24. Sharp, J. A., Wagner, M. J. & Summers, W. C. (1983) *J. Virol.* **45**, 10–17.
25. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. (1978) *J. Mol. Biol.* **126**, 847–863.
26. Ikeda, H., Aoki, K. & Naito, A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3724–3728.
27. Marvo, S. L., King, S. R. & Jaskunas, S. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2452–2456.
28. Rhodes, D. & Klug, A. (1980) *Nature (London)* **286**, 573–578.
29. Noll, M. (1974) *Nucleic Acids Res.* **1**, 1573–1578.
30. Stasiak, A. & Capua, E. (1982) *Nature (London)* **299**, 185–186.
31. Kmiec, E. & Holloman, W. K. (1982) *Cell* **29**, 367–374.
32. Alt, F. W., Rosenberg, N., Enea, V., Siden, E. & Baltimore, D. (1982) *Mol. Cell. Biol.* **2**, 386–400.
33. Kurosawa, Y. & Tonegawa, S. (1982) *J. Exp. Med.* **155**, 201–218.
34. Meselson, M. S. & Radding, C. M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 358–361.
35. Radding, C. M. (1978) *Annu. Rev. Biochem.* **47**, 847–880.
36. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983) *Cell* **33**, 25–35.