MOLECULAR AND CELLULAR BIOLOGY, Aug. 1986, p. 3005–3009
0270-7306/86/083005-05$02.00/0

Vol. 6, No. 8

Nonhomologous Recombination in the Parvovirus Chromosome: Role for a CT$^{A}_{T}$TT$^{C}_{T}$ Motif

AILEEN HOGAN† AND EMANUEL A. FAUST*

Cancer Research Laboratory and Department of Microbiology and Immunology, University of Western Ontario, London, Ontario N6A 5B7, Canada

Received 10 February 1986/Accepted 28 April 1986

The mechanism of nonhomologous recombination in murine cells infected with the parvovirus minute virus of mice (MVM) has been investigated by analysis of DNA sequences at recombination junctions in naturally occurring deletion variants of the virus. We report here that nonhomologous recombination in the MVM chromosome is characterized by short homologies, by insertion at recombination junctions of foreign DNA sequences that are enriched for preferred eucaryotic topoisomerase I cleavage sites, and by an association with a common DNA sequence motif of the type 5′-CT$^{A}_{T}$TT$^{C}_{T}$-3′. Additional analyses of broken MVM chromosomes provided evidence for specific enzymatic cleavage within 5′-CTTATC-3′ and 5′-CTATTC-3′ sequences. The results indicate that the 5′-CT$^{A}_{T}$TT$^{C}_{T}$-3′ motif is an important genetic element for nonhomologous recombination in the parvovirus chromosome.

In mammalian cells, nonhomologous recombination is important for evolutionary variation in gene families (13), chromosome translocations (16), gene amplification (27), movement of retroviruses (29), and integration of pseudogenes (6). Short homologies that are often associated with nonhomologous recombination events are believed either to stabilize intermediates that arise by slippage (13, 28) or to facilitate end joining of broken DNA molecules (22). Although the molecular mechanism remains obscure, eucaryotic topoisomerase I has been implicated in some aspects of nonhomologous recombination on the basis of an association of the preferred topoisomerase I cleavage sites CTT and GTT with excisional recombination crossover points (7, 8).

High-frequency deletion in the 5.0-kilobase linear single-stranded parvovirus chromosome proceeds via a nonhomologous recombination pathway that is mediated by 4- to 10-base-pair homologies (18). The deletions are nonrandom and range in size from approximately 2.0 to 4.5 kilobases (15, 26). Despite this extensive deletion, the *cis*-dominant genetic elements necessary for viral DNA replication and encapsidation are selectively conserved in or near palindromic regions located at the termini of the parvovirus chromosome (9, 15; E. A. Faust and A. Hogan, *in* P. Tijssen, ed., *Handbook of Parvoviruses*, in press). Deletion variants are therefore readily obtained in encapsidated form and are separable from standard infectious virus by equilibrium density gradient centrifugation in CsCl (9, 15). The DNA can be recovered from purified virus particles and cloned in plasmid (pBR322) or phage (M13) vectors, and the DNA sequences at the deletion junctions can be determined (18). Our analysis of three recombinants has shown that deletions averaging 3 kilobases in length occur between pairs of perfectly homologous 4- to 10-base-pair direct repeats such that one copy of the repeated sequence remains at the recombination junction (18). In the present study we analyzed eight independently cloned recombination junctions obtained from naturally occurring deletion variants of the murine parvovirus minute virus of mice (MVM). Our analysis reveals that the recombination borders are associated with a DNA sequence motif, 5′-CT$^{A}_{T}$TT$^{C}_{T}$-3′, that has close homology to the cleavage site for an M_r 60,000 topoisomerase I-like protein found covalently bound to the 5′ termini of intracellular replicating forms of MVM DNA (3, 10). The motif also includes a preferred cleavage site, CTT, for eucaryotic topoisomerase I (7, 8).

The DNA sequences surrounding recombination borders in eight deletion variants of MVM are depicted in Fig. 1. With the exception of DL-382, recombination junctions were flanked by homologous repeats 4 to 10 bases long. Junctional homology was also present in DL-382, but in this case the homologous repeats were displaced from the recombination junction by several nucleotides. DL-382 was also unusual in that it contained a 23-nucleotide insertion of foreign DNA at the recombination junction. The origin of this DNA is not clear, since this sequence is not found in the standard MVM genome. Interestingly, a 46-nucleotide DNA segment encompassing the 23-nucleotide insertion in DL-382 (including nucleotides 974 through 989 and 4209 through 4215) contained eight copies of 5′-CTT-3′ or its inverted complement 5′-AAG-3′. In addition, the sequences 5′-AAG-3′ and 5′-AAC-3′ were present just upstream of the right recombination border (nucleotides 4202 through 4207). The sequences 5′-CTT-3′ and 5′-GTT-3′ (or their inverted complements) were also prevalent at recombination junctions in several of the other deletion variants examined, namely DL-338, DL-412, DL-750, DL-965, and DL-966 (Fig. 1). The significance of the apparent enrichment for these specific sequences at recombination junctions is discussed below.

Further examination of the DNA sequences surrounding the recombination borders depicted in Fig. 1 revealed an additional common repeating element that conforms (with a few exceptions) to a DNA sequence motif of the type 5′-CT$^{A}_{T}$TT$^{C}_{T}$-3′. This motif was derived by comparing 16 different but related sequences, 11 of which were located within 9 nucleotides of a junctional homology. One variant,

* Corresponding author.
† Present address: Imperial Cancer Research Fund Developmental Biology Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3P5, England.

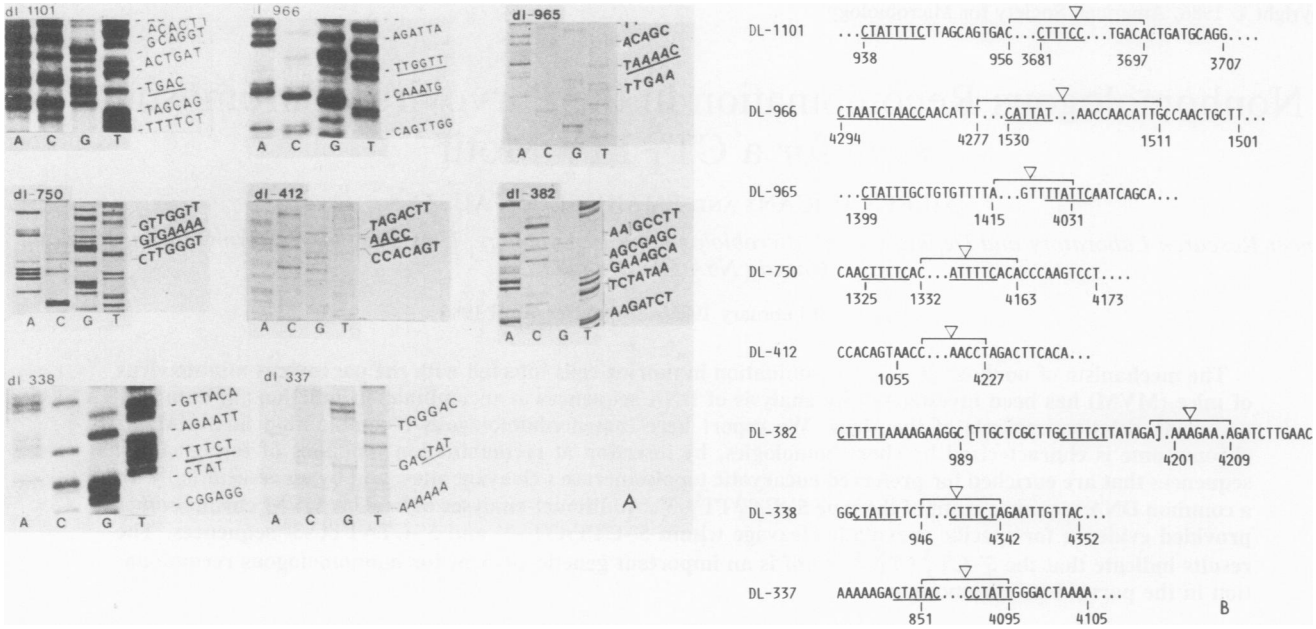

FIG. 1. DNA sequences at recombination junctions in deletion variants of MVM. (A) Recombinant M13 phages (mp9, mp8, and mp18) with MVM inserts were isolated (18), and the dideoxynucleotide DNA sequencing method of Sanger et al. (23) was used to determine the nucleotide sequence at recombination junctions. (B) The DNA sequences surrounding recombination borders are those from the plus strand of the MVM genome, except for DL-966, in which the nucleotide sequence depicted is from the minus strand. Nucleotides are numbered by the convention of Astell et al. (4). Nucleotides found juxtaposed at recombination junctions are connected by solid lines. Directly repeated sequences present immediately adjacent to recombination borders and sequences that conform to a $CT^{A}_{T}TT^{C}_{T}$ motif (underlined) are shown. Deleted regions are indicated by ▽. The extra 23 nucleotides at the insertion junction in DL-382 are bracketed.

DL-412, was not associated with a nearby motif sequence. Nucleotides represented in the motif were found at each position in the following proportions: position 1, 13/16; position 2, 15/16; position 3, 16/16; position 4, 14/16; position 5, 10/16; and position 6, 15/16. The motif sequence therefore reflects sequences found at or near recombination borders at least 81% of the time for five of the six positions in the motif. Position 5 exhibits a greater degree of variability and is accurately represented only 63% of the time. In addition, in 15 of 16 cases, the sequences 5′-CTA-3′ or 5′-TTC-3′ are found at least once.

The probability that DNA sequences conforming to a 5′-$CT^{A}_{T}TT^{C}_{T}$-3′ motif occurs near recombination junctions by chance can be calculated on the basis of the random occurrence of a hexanucleotide sequence in which two positions may be occupied by either one of two nucleotides. Taking into account the base composition of MVM DNA (4), this value is 0.0023. The observed frequency for a perfect match was 0.058 (five perfect matches occurred in 86 nucleotides searched 5′ to junctional homologies). The occurrence of these motif sequences 5′ to junctional homologies in MVM deletion variants is therefore unlikely to be due simply to chance.

The occurrence of sequences that conform to a $CT^{A}_{T}TT^{C}_{T}$ motif at or very near recombination borders does not by itself allow any firm conclusion to be drawn about any specific role in recombination. However, additional analyses of broken MVM DNA molecules (described below) and the inclusion within this motif of the preferred eucaryotic topoisomerase I cleavage site CTT point to these sequences as preferred sites for DNA cleavage.

When MVM variant DNA was analyzed at high resolution in polyacrylamide gels, discrete DNA species of approximately 400 and 120 nucleotides were routinely observed (Fig. 2). These DNA species from mature virions, referred to here as *dl*395 and *dl*120, respectively, were recovered from gels, and their DNA sequences were determined. The DNA sequence of *dl*120 revealed that it consisted of 120 nucleotides from the 3′ end of the standard MVM genome. The 5′ end of *dl*120 consisted of the sequence 3′-CTAT(T)-5′, corresponding to nucleotides 116 through 120 of MVM DNA (Fig. 3). A C residue was present at nucleotide 121, so that the progenitor to *dl*120 appeared to have been cleaved between the T and C residues at positions 120 and 121, respectively (Fig. 4).

The sequence at the 3′ end of *dl*395 was determined by labeling the 3′ end of the DNA with the Klenow fragment of *Escherichia coli* DNA polymerase I and [α-^{32}P]dGTP (4). The 5′ terminus of *dl*395 was labeled with [γ-^{32}P]ATP in the polynucleotide kinase reaction, and the DNA sequence was determined by the chemical degradation method of Maxam and Gilbert (20). This analysis showed that *dl*395 begins at the 3′ end of the standard viral genome and terminates within a 3′-CTTATC-5′ sequence that is present at nucleotides 392 through 397 (Fig. 4). The 5′ terminus of *dl*395, unlike that of *dl*120, is not unique, but the spacing of the bands in the sequencing gel indicated that *dl*395 terminates at nucleotides 396 and 397, just downstream of a run of T's that could be read unambiguously.

The DNA sequence at the 5′ end of *dl*395 is thus almost identical to that found at the 5′ end of *dl*120, the only difference being an inversion of the central AT nucleotide pair. These results strongly suggest that *dl*395 and *dl*120 arose via specific enzymatic cleavage and that the endonuclease involved can recognize either 3′-CTTATC-5′ or 3′-

CTATTC-5′ sequences. An identical degree of flexibility is seen when comparing the sequences at the 3′ and 5′ ends of MVM DNA that are cleaved by a site-specific endonuclease during viral DNA replication (1, 3, 4, 10) (Fig. 4).

The features described here for recombination junctions in MVM deletion variants, namely, the 23-nucleotide insertion in the DL-382 variant that is enriched for preferred eucaryotic topoisomerase I cleavage sites, the CT$\overset{A}{T}$TT$\overset{C}{T}$ motif, and the homologous repeats that flank the deletions, place the DNA rearrangements in the parvovirus chromosome in the same general category of nonhomologous recombination described previously for excisional recombination in the simian virus 40 (SV40) chromosome (7, 8), breakage and reunion of DNA that has been transfected into mammalian cells (22), and spontaneous deletions in the β-globin gene family (13). As in excisional recombination of SV40 DNA (7, 8), deletions in the MVM chromosome may involve cleavage by eucaryotic topoisomerase I, since the preferred topoisomerase I cleavage sites CTT and GTT are prevalent at MVM recombination junctions.

The insertion junction that we observed in DL-382 is reminiscent of similar results obtained after transfection of broken SV40 DNA molecules into monkey cells, in which five extra nucleotides of unknown origin appeared at a recombination junction (22). Junctional insertions have also been demonstrated at the breakpoints of *myc* gene translocations (16). In the present study, one insertion junction was found among eight independently cloned MVM deletion variants, indicating that insertion junctions are relatively common. The mechanism for the creation of insertion junctions remains a mystery, although a role for terminal transferase has been suggested (12). The 23-nucleotide insertion in DL-382 contains several 5′-CTT-3′ repeats as well as a 5′-TTCTT-3′ sequence that is a perfect inverted complement

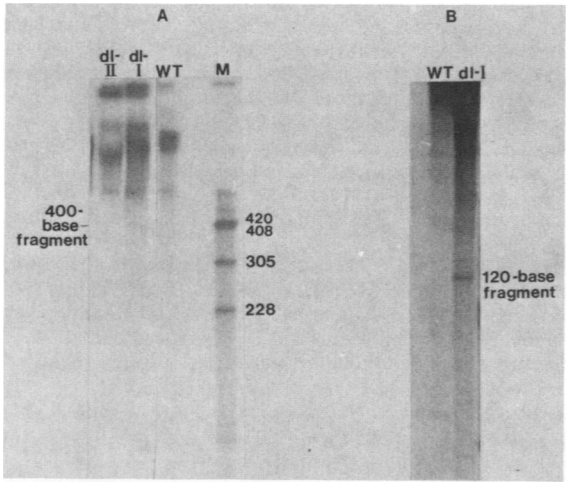

FIG. 2. Detection of *dl*395 and *dl*120. Variant DNA from purified virions was fractionated by hydroxylapatite chromatography into *dl*-I and *dl*-II as described previously (15). The 5′ termini of *dl*-I, *dl*-II, and standard MVM single-stranded genomic DNA were radiolabeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. The DNA was analyzed by electrophoresis in a 6% polyacylamide gel that contained 8 M urea. An autoradiogram was then prepared by exposing the wet gel to Kodak XAR-5 film at −70°C (A). Standard MVM replicative-form DNA was cleaved with *Mbo*I and radiolabeled as above for use as size markers (lane M). *dl*-I DNA and standard MVM DNA were analyzed by electrophoresis in an 8% polyacrylamide gel, and an autoradiogram was prepared as described above (B).

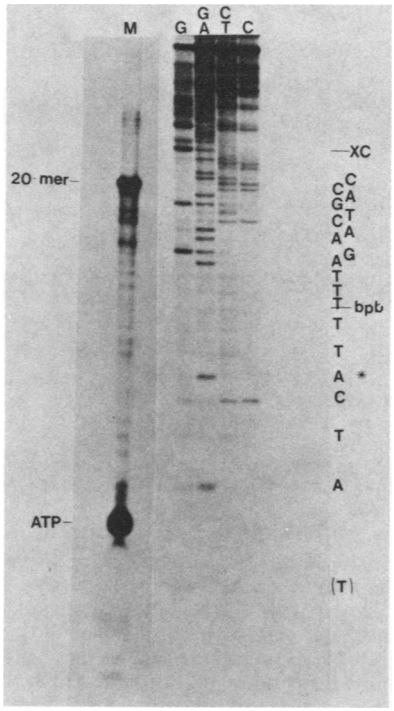

FIG. 3. DNA sequence of *dl*120. *dl*-I MVM variant DNA was radiolabeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase. After electrophoresis, a 120-base fragment (Fig. 2B) was excised from an 8% polyacrylamide gel, extracted with phenol-chloroform (1:1; vol/vol), and precipitated with spermine. The sequence of the DNA was obtained by the chemical degradation method of Maxam and Gilbert (20). The cleavage products were electrophoresed in a 20% polyacrylamide gel (0.3 mm thick) containing 8 M urea, and an autoradiogram was prepared from the gel. The positions of the tracking dyes xylene cyanol (XC) and bromophenol blue (bpb) are indicated. A preparation of 5′-end-labeled 20-mer oligonucleotide was run as a size marker (lane M). The A residue at nucleotide 115 of the standard MVM genome is indicated (*).

of a 5′-AAGAA-3′ sequence located 5′ to the insertion (Fig. 1). This configuration of sequences suggests that at least some of the sequences in the insertion may be related to the nearby 5′-AAGAAG-3′ sequence and could have been produced by a template-directed mechanism.

During MVM DNA replication, resolution of the 5′-terminal hairpin region occurs by a process, termed hairpin transfer, in which a single-strand break exposes a 3′-OH terminus oriented so that it can be extended toward the nearby molecular terminus by DNA polymerase (3, 4, 21). On the basis of an analysis of the DNA sequence at the 5′ terminus of the extended conformation, Astell et al. deduced that a site-specific endonuclease cleaves the DNA 18 nucleotides inboard of the 5′-terminal palindrome at a 5′-CTATT ↓ C-3′ sequence and remains covalently bound to the 5′ end of the DNA at the cleavage site (1, 3, 10) (Fig. 4). Cleavage at a similar sequence, namely, 5′-CTTATC-3′, located immediately adjacent to the 3′-terminal palindrome, is believed to be important for processing dimeric replicative-form intermediates (1, 3). A likely candidate for the site-specific endonuclease is the M_r 60,000 topoisomerase I-like protein found covalently bound via a phosphotyrosyl residue to the 5′ ends of MVM parvovirus replicative-form DNA (1, 3, 10). The genetic origin of this protein is not

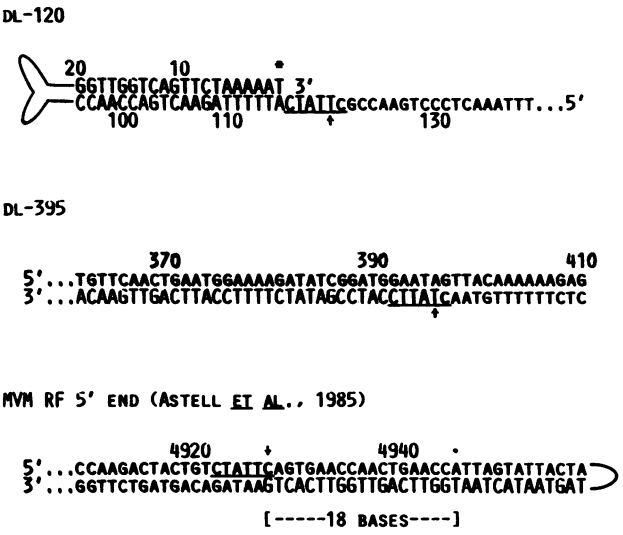

FIG. 4. Comparison of *dl*120, *dl*395, and the 5'-terminal hairpin region of standard MVM replicative-form (RF) DNA. For *dl*120, the 3' hairpin of standard MVM DNA is shown in its fold-back configuration (2). The 3'-terminal T residue (position 1) of the MVM genome is indicated (*). Larger type denotes the sequence of *dl*120, and the 5'-terminal nucleotide of *dl*120 is indicated by the arrow (↑). For *dl*395, the region of MVM replicative-form DNA from nucleotides 362 to 410 is shown (4). The arrow (↑) indicates the 5'-terminal nucleotide of *dl*395 (see text), and the 5' portion of *dl*395 is shown in larger type. For the MVM 5' end, the 5'-terminal hairpin of MVM replicative-form DNA is shown in its fold-back configuration (3, 4). The 5'-terminal nucleotide found in encapsidated MVM DNA (·) is indicated. The arrow (↓) indicates the site cleaved by a "nickase" during viral DNA replication (1, 3).

known, but it appears to be unrelated to any of the known viral proteins (3, 10). Since it remains bound to the 5' side of a DNA break and is attached to its DNA substrate via a phosphotyrosyl residue, its role in MVM DNA replication has been likened to that of the gene-A protein of φX174 (3, 14, 24). Our results suggest that the M_r 60,000 topoisomerase I-like protein is also important in nonhomologous recombination, since its preferred cleavage site is associated with the recombination borders in deletion variants of the MVM chromosome and with the ends of broken viral chromosomes. Thus, like DNA gyrase, whch regulates DNA supercoiling and catalyzes nonhomologous recombination in *E. coli* (19), the M_r 60,000 topoisomerase I-like protein appears to have a dual role in the parvovirus replicative cycle.

Finally, we note that sequences surrounding the cleavage sites for several site-specific recombinases show a striking resemblance to the cleavage site for the M_r 60,000 topoisomerase I-like protein (1, 3) and to the 5' CT$^{A}_{T}$TT$^{T}_{C}$-3' motif described here. For example, the FLP recombinase encoded by the 2 μm plasmid of *Saccharomyces cerevisiae* (5) cleaves at 5'-CTATTC↓-3' and 5'-CTATAC↓TTTC-3' sequences (17, 25), and the λ Int protein cleaves between bases 7 and 8 of a CAACTTN↓NT sequence (11). Whether this intriguing similarity will turn out to reflect an additional role for the M_r 60,000 topoisomerase I-like protein in specialized site-specific recombination reactions awaits its purification and biochemical characterization.

We thank David Denhardt and George Chaconas for critically reading the manuscript and Dale Marsh and Linda Bonis for preparing the figures and typing the manuscript.

This work was supported by grants to E.A.F. from the Bickell Foundation and the Medical Research Council of Canada. Additional support was provided by the National Cancer Institute of Canada. A.H. was the recipient of an Ontario Graduate Scholarship.

LITERATURE CITED

1. **Astell, C. R., M. B. Chow, and D. C. Ward.** 1985. Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. J. Virol. **54**:171–177.
2. **Astell, C. R., M. Smith, M. B. Chow, and D. C. Ward.** 1979. Structure of the 3' hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at origins of DNA replication. Cell **17**:691–703.
3. **Astell, C. R., M. Thomson, M. B. Chow, and D. C. Ward.** 1983. Structure and replication of minute virus of mice DNA. Cold Spring Harbor Symp. Quant. Biol. **47**:751–762.
4. **Astell, C. R., M. Thomson, M. Merchlinsky, and D. C. Ward.** 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. Nucleic Acids Res. **11**:999–1018.
5. **Babineau, D., D. Vetter, B. J. Andrews, R. M. Gronostajski, G. A. Proteau, L. G. Beatty, and P. D. Sadowski.** 1985. The FLP protein of the 2-micron plasmid of yeast. J. Biol. Chem. **260**:12313–12319.
6. **Bernstein, L. B., S. M. Mount, and A. M. Weaver.** 1983. Pseudogenes for human small nuclear RNA U3 appear to arise by integration of self-primed reverse transcripts of the RNA into new chromosomal sites. Cell **32**:461–472.
7. **Bullock, P., J. J. Champoux, and M. Botchan.** 1985. Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. Science **230**:954–958.
8. **Bullock, P., W. Forrester, and M. Botchan.** 1984. DNA sequence studies of simian virus 40 chromosomal excision and integration in rat cells. J. Mol. Biol. **174**:55–84.
9. **Carter, B. J.** 1984. Variant and defective interfering parvoviruses, p. 209–258. *In* K. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
10. **Chow, M., J. W. Bodnar, M. Polvino-Bodnar, and D. C. Ward.** 1986. Identification and characterization of a protein covalently bound to DNA of minute virus of mice. J. Virol. **57**:1094–1104.
11. **Craig, N. L., and H. A. Nash.** 1983. The mechanism of phage λ site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell **35**:795–803.
12. **Desiderio, S. V., G. D. Yancopoulos, M. Paskind, E. Thomas, M. A. Boss, N. Landau, F. Alt, and D. Baltimore.** 1984. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature (London) **311**:752–755.
13. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissmann, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Barelle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
14. **Eisenberg, S., and A. Kornberg.** 1979. Purification and characterization of φX174 gene A protein. A multifunctional enzyme of duplex DNA replication. J. Biol. Chem. **254**:5328–5332.
15. **Faust, E. A., and D. C. Ward.** 1979. Incomplete genomes of the parvovirus minute virus of mice: selective conservation of genome termini, including the origin for DNA replication. J. Virol. **32**:276–292.
16. **Gerondakis, S., S. Cory, and J. M. Adams.** 1984. Translocation of the *myc* cellular oncogene to the immunoglobulin heavy chain locus in murine plasmacytomas is an imprecise reciprocal exchange. Cell **36**:973–982.
17. **Gronostajski, R. M., and P. D. Sadowski.** 1985. Determination of DNA sequences essential for FLP-mediated recombination by a novel method. J. Biol. Chem. **260**:12320–12327.
18. **Hogan, A., and E. A. Faust.** 1984. Short direct repeats mediate spontaneous high-frequency deletions in DNA of minute virus of mice. Mol. Cell. Biol. **4**:2239–2242.

19. **Ikeda, H., and M. Shiozaki.** 1984. Nonhomologous recombination mediated by Escherichia coli DNA gyrase: possible involvement of DNA replication. Cold Spring Harbor Symp. Quant. Biol. **49**:401–409.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **Revie, D., B. Y. Tseng, R. H. Grafstrom, and M. Goulian.** 1979. Covalent association of protein with replicative form DNA of parvovirus H-1. Proc. Natl. Acad. Sci. USA **76**:5539–5543.
22. **Roth, D. B., T. N. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5**:2599–2607.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
24. **Sanhueza, S., and S. Eisenberg.** 1984. The cleavage of single-stranded DNA by the φX174 A* protein: the A*–single-stranded DNA covalent linkage. Proc. Natl. Acad. Sci. USA **81**:4285–4289.
25. **Senecoff, J. F., R. C. Bruckner, and M. M. Cox.** 1985. The FLP recombinase of the yeast 2-μm plasmid: characterization of its recombination site. Proc. Natl. Acad. Sci. USA **82**:7270–7274.
26. **Senepathy, P., and B. J. Carter.** 1984. Molecular cloning of adeno-associated virus variant genomes and generation of infectious virus by recombination in mammalian cells. J. Biol. Chem. **259**:4661–4666.
27. **Stark, G. R., and G. M. Wahl.** 1984. Gene amplification. Annu. Rev. Biochem. **53**:447–491.
28. **Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye.** 1966. Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31**:77–84.
29. **Varmus, H. E.** 1983. Retroviruses, p. 411–503. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.