Nonhomologous Recombination in the Parvovirus Chromosome: Role for a CT\textsuperscript{A}TTT\textsuperscript{C} 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\textsuperscript{A}TTT\textsuperscript{C}-3'. Additional analyses of broken MVM chromosome provided evidence for specific enzymatic cleavage within 5'-CTTATC-3' and 5'-CTTATTT-3' sequences. The results indicate that the 5'-CT\textsuperscript{A}TTT\textsuperscript{C}-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 and homology. One variant, DL-382, in 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'-AAC-3'. In addition, the sequences 5'-AAC-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'-GT-3' (or their inverted complements) were also prevalent at recombination junctions in several of the other deletion variants examined, namely DL-338, DL-421, 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\textsuperscript{A}TTT\textsuperscript{C}-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,
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 CTTTTT motif (underlined) are shown. Deleted regions are indicated by \( \texttt{\textbackslash V} \). 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'-CTTTT-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 CTTTTT 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 \( d1395 \) and \( d120 \), respectively, were recovered from gels, and their DNA sequences were determined. The DNA sequence of \( d120 \) revealed that it consisted of 120 nucleotides from the 3' end of the standard MVM genome. The 5' end of \( d120 \) 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 \( d120 \) 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 \( d1395 \) was determined by labeling the 3' end of the DNA with the Klenow fragment of \( Escherichia coli \) DNA polymerase I and \( [\alpha-32P]\)dGTP (4). The 5' terminus of \( d1395 \) was labeled with \( [\gamma-32P]ATP \) 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 \( d1395 \) 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 \( d1395 \), unlike that of \( d120 \), is not unique, but the spacing of the bands in the sequencing gel indicated that \( d1395 \) 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 \( d1395 \) is thus almost identical to that found at the 5' end of \( d120 \), the only difference being an inversion of the central AT nucleotide pair. These results strongly suggest that \( d1395 \) and \( d120 \) 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 CTATT 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 c-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 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 replication-form intermediates (1, 3). A likely candidate for the site-specific endonuclease is the M, 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. 2. Detection of dl95 and dl120. 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 radio- labeled with [γ-32P]ATP and T4 polynucleotide kinase. The DNA was analyzed by electrophoresis in a 6% polyacrylamide 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 MboI 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).
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 Mₖ 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, which regulates DNA supercoiling and catalyzes nonhomologous recombination in E. coli (19), the Mₖ 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 Mₖ 60,000 topoisomerase I-like protein (1, 3) and to the 5’ CTPTT-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 CAACCTTN↓NT sequence (11). Whether this intriguing similarity will turn out to reflect an additional role for the Mₖ 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


