Complete Nucleotide Sequence of a Milk-Transmitted Mouse Mammary Tumor Virus: Two Frameshift Suppression Events Are Required for Translation of *gag* and *pol*

R. MOORE,¹ M. DIXON,¹ R. SMITH,¹ G. PETERS,² AND C. DICKSON^{1*}

Imperial Cancer Research Fund Laboratories, London WC2A 3PX,¹ and Imperial Cancer Research Fund Laboratories, St. Bartholomew's Hospital, London EC1A 7BE,² United Kingdom

Received 1 August 1986/Accepted 30 October 1986

We sequenced two recombinant DNA clones constituting a single provirus of the milk-transmitted mouse mammary tumor virus characteristic of BR6 mice. The complete provirus is 9,901 base pairs long, flanked by 6 base-pair duplications of cellular DNA at the site of integration. Five extensive blocks of open reading frame corresponding to the *gag* gene, the presumed protease, the *pol* and *env* genes, and the open reading frame *orf* within the long terminal repeat of the provirus were readily discernible. Translation of *gag*, protease, and *pol* involved three different translational reading frames to produce the three overlapping polypeptide precursors Pr77, Pr110, and Pr160 found in virus-infected cells. Synthesis of the reverse transcriptase and endonuclease therefore required two separate frameshifts to suppress the termination codons at the ends of the Pr77 and Pr110 domains. Direct evidence is presented for translational readthrough of both stop codons in an in vitro protein synthesis system.

Retroviruses are widely dispersed among vertebrate species but are unified by their genome organization and mode of replication. An essential step in their replication is the reverse transcription of the viral RNA into a double-stranded DNA intermediate that becomes integrated as a provirus within the chromosomal DNA of the host cell. As a result, analyses of retroviral genomes have recently focused on the isolation and characterization of molecular clones of proviral DNA to the extent that complete nucleotide sequences are now available for several avian and mammalian retroviruses (18, 40, 42, 44, 46-49). Such techniques and the opportunity to perform site-directed mutagenesis have also led to the identification of previously unrecognized or poorly understood viral functions over and above the well-characterized components encoded by the *gag*, *pol*, and *env* genes. These include a protease, required for processing of the viral polypeptide precursors; an endonuclease, required for proviral integration; and specific sequences needed for the initiation of viral DNA synthesis, the integration of viral DNA, and the packaging of genomic RNA (5, 12, 16, 24, 27, 31, 32, 43, 50, 55). Whereas direct evidence for these functions is restricted to a few specific examples, the general uniformity of retrovirus architecture and conservation of particular localized sequence motifs suggest that most retroviruses conform, with only minor embellishments, to a basic prototype.

Mouse mammary tumor virus (MMTV) is unique among the *Retroviridae* because it represents a distinct morphological subclass (B type) and is one of the causative factors in a specific epithelial neoplasm (9, 29). However, despite being the first mammalian retrovirus isolated, MMTV has remained among the most refractory to molecular analysis. Two reasons can be cited for this: its strong tropism for the mammary alveoli, which may underlie the difficulties encountered in tissue culture manipulation of the virus, and its apparent resistance to molecular cloning as a complete DNA provirus (2, 11, 13, 25, 26, 53). Although large segments of

the genome have been previously characterized (6, 11, 13, 14, 20, 25, 26, 37), here we report the successful isolation and DNA sequencing of recombinant phage clones constituting what we believe to be an entire milk-transmitted MMTV provirus. This provides the first complete comparison between the genome organization of MMTV and that of other retroviruses and clarifies the disposition of the open reading frames within the *gag* and *pol* domains. Previous experiments from our group and others (for a review, see reference 9) indicating three overlapping polypeptide precursors were confirmed by the DNA sequence and extended by the direct demonstration of two independent translational frameshifts in vitro.

MATERIALS AND METHODS

Cloning of MMTV proviral DNA. High-molecular-weight DNA from BR6 mouse mammary tumors was digested to completion with *EcoRI*, for which there is a single cleavage site in the MMTV provirus, and ligated into the purified arms of the λ gtWES $\cdot$ λ B vector (33). Packaged phage was plated on *Escherichia coli* LE392 and screened by hybridization to a probe for the MMTV long terminal repeat (LTR) by standard procedures. Positive phage was then plaque purified and hybridized to 5'- and 3'- specific MMTV probes to distinguish the respective virus-cell DNA junctions (33). From one particular library prepared from the tumor designated W26 (30), we recovered recombinants corresponding to both halves of a single integrated provirus. These were matched initially by preparing restriction fragment probes specific for the cellular DNA adjacent to the viral sequences in each clone. The procedures used for the labeling of probes and blot hybridization have been adequately detailed elsewhere (33). Although the 3' junction fragment was readily transferred into a plasmid vector to facilitate further analysis, difficulties were encountered in subcloning the corresponding 5' junction (1a).

DNA sequence analysis. The 6.7-kilobase (kb) 5' junction fragment, containing around 1 kb of flanking cellular DNA, and the corresponding 6.4-kb 3' fragment were excised from

* Corresponding author.

their respective vectors with *EcoRI* and recovered by preparative agarose gel electrophoresis. The fragments were then self-ligated, randomly sheared by sonication, and blunt-end ligated into the *SmaI* site of the M13 vectors mp8 and mp9 (30). Recombinant M13 phage was grown in small-scale liquid cultures to prepare single-stranded DNA templates for dideoxynucleotide chain termination sequencing procedures (30). DNA sequences were compiled by the DBUTIL computer program (51) and, apart from a small section of the 3' LTR, were obtained for both DNA strands.

In vitro synthesis of MMTV-specific RNAs and proteins. A segment of MMTV proviral DNA extending from the *PstI* site at nucleotide 2580 to the *EcoRI* site at nucleotide 5803 was inserted into the polylinker of the pSP65 plasmid vector (Promega Biotec, Madison, Wis.). The orientation was such that MMTV-specific RNA of positive sense could be transcribed in vitro from the SP6 phage promoter with the SP6-specific RNA polymerase (28). Conditions for RNA synthesis were essentially as described by Melton et al. (28). RNAs of different length were obtained by cleavage of the plasmid at the specific *KpnI*, *BglII*, and *EcoRI* sites in the MMTV proviral DNA (see Fig. 3). The size and integrity of each RNA preparation was verified by Northern blot analysis with an MMTV specific probe (not shown). After synthesis, the products were treated with RNase-free DNase, phenol extracted, and ethanol precipitated. The protein coding capacity of each RNA was then assessed by translation in vitro in a nuclease-treated reticulocyte lysate (8, 10). The analysis of the resultant [³⁵S]methionine-labeled products by acrylamide gel electrophoresis, immune precipitation, and tryptic peptide mapping followed protocols described in detail previously (7, 8, 10).

RESULTS

Cloning and DNA sequencing of the MMTV provirus. A recombinant phage library of *EcoRI*-digested DNA from a BR6 mouse mammary tumor was screened with a probe specific for the MMTV LTR (33). Since *EcoRI* cuts the MMTV provirus at a single site, each positive plaque identified was expected to represent a junction between viral and cellular DNA sequences. During plaque purification, these recombinants were hybridized with probes specific for either the 5' or 3' portions of the provirus (defined relative to the *EcoRI* site). In contrast to our previous experience and that of other laboratories in cloning MMTV DNA from GR and C3H mouse mammary tumors, we recovered 5' and 3' junction fragments with approximately equal frequency from BR6 tumors (11, 25, 26, 53). Moreover, many of these recombinants represented proviral DNA of the milk-transmitted MMTV characteristic of BR6 mice rather than endogenous sequences (35). This was initially shown by preparing unique sequence probes specific for the cellular DNA adjacent to each provirus as described elsewhere (33, 35). Such probes also permitted the matching of corresponding 5' and 3' junctions, since the probes identified signature restriction fragments derived from the unoccupied site in normal DNA.

As a result of these analyses, we were in a position for the first time to determine the complete sequence and genome organization of a single, potentially infectious provirus of milk-borne MMTV. The derived sequence for two matched *EcoRI* halves and the cellular DNA immediately flanking the LTRs are depicted in Fig. 1. Six base pairs of cellular DNA were duplicated at the site of integration, confirming the linkage between these two junction fragments. It should be

stressed, however, that although the sequence was obtained for both DNA strands, it did not cross the *EcoRI* site at nucleotide 5803, and we therefore cannot formally exclude additional sequence at this position (e.g., two closely apposed *EcoRI* sites). Other features of the sequence are described below and in Discussion.

Open reading frames in the MMTV provirus. From the MMTV DNA sequence and known features of the LTRs (11, 13, 14, 20, 21, 26), we deduced that the viral genome RNA must extend for 8,585 nucleotides, beginning at nucleotide 1196 (Fig. 1). Since retroviral RNA is in the positive sense the sequence as presented was directly translatable into the viral proteins. A computed translation of the MMTV provirus in the three reading frames is depicted in Fig. 2, indicating the positions of all potential termination codons. It was immediately apparent that the viral genome RNA could encompass five substantial protein-encoding domains. The three largest were presumed to encode the viral *gag*, *pol*, and *env* functions, as expected of a prototype retrovirus. An additional segment of open reading frame characteristic of MMTV and designated *orf* began immediately proximal to the boundary of the 3' LTR as previously reported (10, 11, 14, 20, 26, 37). However, the presumed *gag* and *pol* domains, nominally ascribed to reading frames 1 and 2, respectively, were not contiguous, and it was clear that continuity in the generation of a *gag-pol* precursor would require the inclusion of amino acids encoded in reading frame 3. The situation therefore paralleled that described recently for Mason-Pfizer monkey virus, bovine leukemia virus, and human T-cell leukemia virus types I (HTLV-I) and II (HTLV-II) in which the viral protease bridges the *gag* and *pol* domains in a different frame (38, 40, 44, 46, 49, 60; see below). The data also concur with previous immunobiochemical and in vitro translation experiments in which three overlapping precursors were identified with antisera to the viral structural proteins (for a review, see reference 9). Designated Pr77, Pr110, and Pr160, these precursors probably share the same amino terminus, initiating at the methionine codon at nucleotide 1508 and terminating at nucleotides 3281, 4087, and 6771, respectively (Fig. 2). Synthesis of the Pr160 *pol* precursor would therefore necessitate two separate -1 translational frameshifts, switching from frame 1 to 3 at the Pr77-Pr110 boundary and from frame 3 to 2 at the end of Pr110.

Demonstration of translational frameshifting in vitro. We and others have previously shown that rabbit reticulocyte lysates programmed with purified MMTV genome RNA can support synthesis of Pr77, Pr110, and Pr160 in roughly the same relative proportions as detected in lysates from infected cells (8, 45). However, these studies could never rigorously exclude the presence of minor RNA species in virion preparations which might direct synthesis of the longer readthrough products. To circumvent these problems, we followed the example of Jacks and Varmus (19) and constructed recombinant plasmids in which appropriate segments of MMTV RNA could be transcribed in vitro with the specific RNA polymerase of phage SP6 (28). Thus, a segment of MMTV proviral DNA extending from the *PstI* site at nucleotide 2580 to the *EcoRI* site at 5803 was inserted into the polylinker of the vector pSP65 (Fig. 3). Linearization of the resultant plasmid at either the *KpnI*, *BglII*, or *EcoRI* sites within the MMTV sequences would be expected to yield RNA transcripts of 0.9, 1.8, and 3.2 kb, which could discriminate between the proposed translational frameshifts at the ends of Pr77, Pr110, and Pr160. Although this fragment lacks the normal initiation codon at the start of the *gag* gene,

FIG. 1. Complete nucleotide sequence of a milk-transmitted MMTV provirus. Two *EcoRI* fragments representing the 5' and 3' ends of a single MMTV provirus were recovered as recombinant clones and sequenced by shotgun cloning into M13 vectors and dideoxynucleotide chain termination procedures (30, 33). The sequence across the *EcoRI* site was not determined, but with the exception of a small section of the 3' LTR, all sequences were obtained for both strands. Shown are 9,901 nucleotides of proviral sequence, numbered from base 1 of the 5' LTR and flanked by segments of the cellular DNA adjacent to the provirus (shown in lower case). The boundaries of the LTRs are indicated by rectangular brackets, and the six bases of cellular DNA duplicated during proviral integration and underlined (~~~~). The beginnings of long open reading frames are also shown (►, accompanied by a number referring to the translation frame used). Other sequence features indicated are the 5' end of the viral genome RNA (cap); the binding site for tRNA₃^{lys} (pbs); the splice donor (sd) and acceptor (sa) sites for *env* and *orf* mRNAs; the known amino-terminal sequences of p27, p14-p30, gp52, and gp36 (┐), and the potential inverted repeat sequences distal to translational frameshift sites (overlined) (11, 13, 14, 17, 20, 21, 25, 26, 34, 37, 54, 57).

FIG. 1—Continued.

FIG. 2. Disposition of open reading frames in the MMTV proviral sequence. (a) Computed translation of the MMTV DNA sequence of Fig. 1 in the three possible reading frames. The positions of all potential termination codons are indicated relative to the linear map from nucleotides 1 to 9901. (b) Open reading frames shown in panel a directly correlated with the precursor polyproteins identified in MMTV-infected cells and characterized by immune precipitation and tryptic peptide mapping (9). The different reading frames are depicted by the symbols: , frame 1; , frame 3; , frame 2. Thus Pr77 is translated in frame 1, Pr110 spans frames 1 and 3, and Pr160 uses frames 1, 3, and 2. The open reading frame *orf* in the LTR is not included in this figure, since the predicted protein has not been detected in MMTV-infected cells (8, 10, 45).

11 was acquired in the readthrough from the 51- to 92-kDa products. Moreover, the other major peptides were entirely consistent with the initiation of translation within the coding region of virion protein p27.

DISCUSSION

The majority of previous studies on the MMTV provirus and sequence have concentrated on the milk-borne viruses characteristic of the GR and C3H strains of mice or on endogenous elements. Although there have been numerous reports on the sequence of the LTRs, the *env* gene, and parts of the *pol* gene, none of these has provided a comprehensive picture of the MMTV genome (6, 11, 13, 14, 20, 25, 26, 37). Our success in obtaining recombinant clones encompassing the 5' half of the genome may in some measure reflect our choice of the BR6 mouse strain. The exact origin of the milk-transmitted MMTV in these mice remains unclear, since they were derived initially from a cross between an RIII male and a nonviremic C57BL female (15), but we assume that this virus represents the one that is characteristic of RIII mice. However, the so-called poison sequences which have hampered many studies on MMTV are not completely abrogated in this strain, since we encountered difficulties whenever attempts were made to transfer cloned 5' junction fragments from λ vectors into plasmids (1a). It was therefore impracticable to reconstruct a complete provirus by joining the two *EcoRI* junction fragments, which thus precluded any tests on the biological activity of this provirus. Nevertheless, all the features of the derived DNA sequence would fulfil the expectations for an infectious virus and would be entirely consistent with the known biochemical data.

The 5' LTR and leader sequences. The LTR of the provirus shown in Fig. 1 extends for 1,328 base pairs and is bounded by the expected 6-base-pair inverted repeats (11, 14, 20, 25). The various elements required for initiation of transcription and glucocorticoid regulation have been described in detail by others (for a review, see reference 39) and are maintained in our sequence with only scattered base changes. Flanking the LTR to the 5' side is the hexanucleotide CGTCAG duplicated upon integration of the provirus into cellular DNA (11, 20, 25), and at the 3' boundary is the 18-nucleotide binding site for the DNA synthesis primer tRNA^{Lys} (25, 34). The sequences distal to the primer binding site are loosely described as leader sequences, since they precede the start of the *gag* gene, but they encompass the splice donor site for *env* mRNA at nucleotide 1484 (Fig. 1) and presumably all or part of the packaging signal for MMTV genome RNA (13, 26, 27, 54, 55).

***gag* gene.** Identification of the ATG codon at nucleotide 1508 as the initiator for the *gag* gene products is based on a number of considerations: it is the first ATG encountered after the LTR (Fig. 1), it conforms to the GNNATG consensus (23), it is followed by 590 codons in frame (Fig. 1 and 2), and it predicts the amino terminal Met-Gly sequence consistent with the observed myristylation of the MMTV *gag* precursor (13, 41). However, we as yet have no information that would allow us to align the amino-terminal virion protein, p10, with the DNA sequence. Partial amino acid sequencing has allowed us to match the N terminus of p27, the major core protein, with the proline codon at nucleotide 2315, and upstream of this point are two domains of around 20 amino acids, one of which is composed of 50% acidic residues, the other being 50% basic. This latter region is reminiscent of the highly basic virion protein, p8, described

FIG. 3. Translational frameshifting in vitro. (a) Partial restriction map of MMTV proviral DNA. The positions of significant (but not all) *Pst*I (P), *Kpn*I (K), *Bgl*II (Bg), and *Eco*RI (E) sites are indicated. The segment from *Pst*I to *Eco*RI depicted by the bold line was subcloned into the polylinker of the pSP65 vector such that transcription of the resultant plasmid DNA by the SP6 phage RNA polymerase initiated at the specific promoter (◆) and copied the MMTV sequences into positive-sense RNA (19, 28). Linearization of the plasmid at the *Kpn*I, *Bgl*II, or *Eco*RI sites yielded specific RNA transcripts (~~~~) of 0.9, 1.8, and 3.2 kb, respectively. When these were used to direct protein synthesis in a rabbit reticulocyte lysate (8, 10), the expected products had the sizes (in kilodaltons) and composition shown in the diagram. The symbols for each translational frame correspond to those used in Fig. 2. Termination occurred either at the normal stop codons or at the 3' end of the specific RNA transcripts. (b) In vitro translation products generated from RNAs terminating at the *Kpn*I (K), *Bgl*II (Bg), or *Eco*RI (E) sites in the MMTV sequence. These products were analyzed by electrophoresis in 10% polyacrylamide gels (7, 8), labeled by the addition of [³⁵S]methionine to the reticulocyte lysates, and visualized by autoradiography. The positions and molecular sizes (in kilodaltons) of the relevant products are indicated. The No RNA lane shows the background of protein synthesis observed in the

FIG. 4. Tryptic peptide maps of in vitro translation products. Two-dimensional fingerprints were prepared of tryptic peptides from the ³⁵S-labeled in vitro products depicted in Fig. 3b (7, 8). Shown are the 25-kDa (a), the 51-kDa (b), and the 92-kDa (c) products and a schematic diagram of the previously reported fingerprint of in vitro synthesized Pr160 (d) (8). Numbered peptides correspond to those discussed in earlier publications and are represented according to the frame symbols used in Fig. 2. The symbols Δ and ▲ indicate additional peptides found in the 51-kDa and 92-kDa products, respectively.

previously (7, 9). The amino-terminal sequences of virion protein p14 and the minor protein p30 both align perfectly with the DNA sequence beginning at nucleotide 2996 (N. Totty, M. Waterfield, R. Moore, M. Dixon, R. Smith, G. Peters, and C. Dickson, unpublished results). P14 is the major nucleic-acid-binding protein of MMTV, a conclusion that is supported by the presence of two of the motifs (N, N + 3, N + 13) of cysteine residues identified in other systems (1, 4). However, any function of p30 remains a mystery, since to our knowledge, no equivalent protein has been described for other retroviruses. As p30 represents a *trans*-frame protein, sharing an amino terminus with p14 in frame 1 and a carboxy terminus with Pr110 in frame 3, it will clearly be interesting to determine the complete amino acid sequence of p30 to define exactly where in the nucleotide sequence the frameshift occurs.

Protease. One consequence of expressing viral structural proteins in the form of precursors is the requirement for a specific protease to process these polypeptides into the mature forms. It is therefore hardly surprising that retroviruses encode their own enzymes to perform this role, but it has become apparent that they have adopted various strategies for implementing this function. Avian leukosis viruses, for example, include the protease within the *gag* gene; in murine and feline leukemia viruses, the enzyme is formally

nuclease-treated lysate. We noted that several of the major in vitro products migrated as triple rather than single entities in acrylamide gels, but we have no explanation for this phenomenon, nor have we been able to distinguish differences between the various species.

FIG. 5. Amino acid sequence homologies among retrovirus *pol* genes. Sequence comparisons are presented between the *pol* genes of MMTV (this paper), Mason-Pfizer monkey virus (49), Rous sarcoma virus (42), Moloney murine leukemia virus (47), HTLV-I (44), and Visna virus (48). Residues homologous to the MMTV sequence are shaded. The distinction between polymerase (a) and endonuclease (b) domains is based on the features of these proteins in Rous sarcoma virus and Moloney murine leukemia virus.

part of the *pol* gene, occurring downstream of the termination codon separating *gag* and *pol*; and in HTLV-I, HTLV-II, bovine leukemia virus, and Mason Pfizer monkey virus, the protease is encoded in a separate reading frame between the *gag* and *pol* genes (5, 18, 19, 38, 40, 42, 44, 46, 47, 49, 58-60). The situation in MMTV most closely parallels that of the last virus group in that the sequence LLDTGADK, which is homologous to the consensus active site of retroviral and some cellular acid proteases (52), occurs in frame 3 (centered on nucleotide 3821) in the bridge between the *gag*

and *pol* domains (Fig. 1 and 2). Although we are not yet in a position to define the boundaries or exact size of the MMTV protease, it is clear from these data that its expression requires at least one translational frameshift.

Polymerase and endonuclease. Characterization of the MMTV reverse transcriptase has been and remains a contentious topic, with conflicting reports of a 100-kDa monomer or 85- and 55-kDa dimers as the active moiety (for a review, see reference 9). Now that the full DNA sequence is known, it should be possible to resolve this issue by obtain-

FIG. 6. Sequences involved in translational frameshifting. The DNA sequences spanning the boundaries between Pr77 and Pr110 (a) and between Pr110 and Pr160 (b) are displayed in a simplified form. The amino acids encoded in the relevant frames are indicated by the single letter code, with stop codons also shown (■). The sequence motifs discussed in the text are underlined.

ing and aligning terminal amino acid sequences. A similar strategy will be required to characterize the endonuclease domain, since the presence of poison sequences has precluded the types of mutagenesis studies applied to other retroviruses (12, 16, 32, 43). As noted by others, the coding domains of polymerase and endonuclease are nevertheless readily discernible in the MMTV sequence simply by homologies to other known sequences (3, 6, 38, 40, 42, 44, 46–49). The degree of conservation is most striking within the polymerase itself (Fig. 5), particularly between MMTV and Mason-Pfizer virus, where the homology in the region shown is around 65% (Fig. 5) (49).

env gene. We previously reported the *env* gene sequence of the GR strain of MMTV and discussed its implications in terms of the viral glycoproteins and their precursor Pr73 (37). The nucleotide sequence of the corresponding region of the BR6 provirus differs at 60 positions, reflected in only 10 amino acid differences. Five of these changes are localized within the leader preceding the gp52 domain (17, 26, 37), but their significance is unclear, since we do not yet know which ATG codon serves as the initiator for Pr73 (26, 37). None of the essential features of the two mature glycoproteins, such as the number and position of glycosylation sites or the nature of the transmembrane anchor, is affected.

The 3' LTR and *orf*. The most significant feature of the 3' LTR of MMTV is the existence of a substantial open reading frame designated *orf*, which begins at an A residue in the polypurine tract immediately preceding the boundary of the LTR and continues for 962 residues in the U3 portion of the genome (8, 10, 11, 14, 20, 26, 45). We previously argued that the maintenance of *orf* in several endogenous and milk-borne strains of MMTV indicates it has a functional role (9). The sequence presented in Fig. 1 extends this argument, since the integrity of the reading frame is maintained despite 94 single-base differences between this sequence and the GR virus LTR. Moreover, the putative splice acceptor site mapped approximately 50 nucleotides upstream of the LTR is also represented in our sequence (54, 57) (Fig. 1).

Suppression by translational frameshifting. Although it is still unclear how or why *orf* may be expressed, synthesis of the *env* gene from a spliced, subgenomic mRNA presumably reflects the need for high levels of the viral glycoproteins, roughly equivalent to the levels of the *gag* structural pro-

teins. In contrast, the protease, polymerase, and endonuclease are nonstructural components, and although they may occur in virions, they are required in only catalytic amounts. It has long been recognized that infected cells express the *gag-pol* precursor at only around 5% of the level of the *gag* precursor, and it has recently become apparent that retroviruses have adopted an unusual strategy, namely suppression of termination codons, for controlling the levels of these functions (19, 36, 58–60). In MMTV, there are three precursors, now identified as Pr77^{gag}, Pr110^{gag pro}, and Pr160^{gag pro pol}, so that two termination codons must be bypassed during synthesis of the *pol* gene. This required two frameshifts (Fig. 2), and both events could occur in rabbit reticulocyte lysates primed with defined segments of MMTV RNA (Fig. 3 and 4). From the ratios of readthrough products to terminated products, we estimated that each frameshift occurred with an efficiency of around 15 to 25% in vitro, which is reasonably consistent with the levels of Pr77, Pr110, and Pr160 observed for infected cells (7, 9). However, these studies cannot formally exclude the possibility of self-splicing or splicing catalyzed by components of the reticulocyte lysate.

The sequences at which frameshifting occurs are displayed in more detail in Fig. 6. At present, we have insufficient information to determine where within the overlaps between frames 1 and 3 in Fig. 6a and between frames 3 and 2 in Fig. 6b the actual switches occur. Resolution of this issue awaits protein sequence data for the virion p30 and protease, the two likely candidates for products which span these frameshifts. However, we note that both these frameshift sequences comply with precedents set by other retroviruses, in that one occurs at or near the sequence ATTTA, as in avian sarcoma virus (19, 42), and the other probably involves the AAAAAAC sequence identified as the potential frameshift site in bovine leukemia virus, HTLV-I, and HTLV-II and paralleled in some procaryote systems (22, 38, 46, 56, 60). Moreover, both frameshift sites are closely followed by potential stem-loop structures in the virion RNA. Although we have not assessed the thermodynamic stability of such secondary structural features relative to others in the sequence, it is interesting to speculate that they may contribute in *cis* to frameshifting (38, 46, 60). Whether such processes have wider significance in the general control of gene expression remains an intriguing possibility.

ACKNOWLEDGMENTS

We thank S. Brookes for additional technical help; L. Crawford, J. Wyke, and J. Witkowski for comments on the manuscript; and A. Kessler for its preparation.

LITERATURE CITED

- Arthur, L. O., C. W. Long, G. H. Smith, and D. L. Fine. 1978. Immunological characterization of the low molecular weight DNA binding protein of mouse mammary tumor virus. *Int. J. Cancer* 22:433–440.
- Brookes, S., M. Placzek, R. Moore, M. Dixon, C. Dickson, and G. Peters. 1986. Insertion elements and transitions in cloned mouse mammary tumour virus DNA: further delineation of the poison sequences. *Nucleic Acids Res.* 14:8231–8245.
- Buetti, E., and H. Diggelmann. 1981. Cloned mouse mammary tumor virus DNA is biologically active in transfected cells and its expression is stimulated by glucocorticoid hormones. *Cell* 23:335–345.
- Chiu, I. M., R. Callahan, S. R. Tronick, J. Schlom, and S. A. Aaronson. 1984. Major *pol* gene progenitors in the evolution of oncornaviruses. *Science* 223:364–370.
- Copeland, T. D., S. Oroszlan, V. S. Kalyanaraman, M. G.

- Sarngardharan, and R. C. Gallo. 1983. Complete amino acid sequence of human T-cell leukemia virus structural protein p15. *FEBS Lett.* **162**:390-395.
5. Crawford, S., and S. P. Goff. 1985. A deletion mutation in the 5' part of the *pol* gene of Moloney murine leukemia virus blocks proteolytic processing of the *gag* and *pol* polypeptides. *J. Virol.* **53**:899-907.
6. Deen, K. C., and R. W. Sweet. 1986. Murine mammary tumor virus *pol*-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus *pol* gene. *J. Virol.* **57**:422-432.
7. Dickson, C., and M. Atterwill. 1979. Composition, arrangement and cleavage of the mouse mammary tumor virus polyprotein precursor Pr77^{gag} and p110^{gag}. *Cell* **17**:1003-1012.
8. Dickinson, C., and G. Peters. 1981. Protein-coding potential of mouse mammary tumor virus genome RNA as examined by in vitro translation. *J. Virol.* **37**:36-47.
9. Dickson, C., and G. Peters. 1983. Proteins encoded by mouse mammary tumour virus. *Curr. Top. Microbiol. Immunol.* **106**:1-34.
10. Dickson, C., R. Smith, and G. Peters. 1981. In vitro synthesis of polypeptides encoded by the long terminal repeat region of mouse mammary tumour virus DNA. *Nature (London)* **291**:511-513.
11. Donehower, L. A., A. L. Huang, and G. L. Hager. 1981. Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *J. Virol.* **37**:226-238.
12. Donehower, L. A., and H. E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* **81**:6461-6465.
13. Fasel, N., E. Buetti, J. Firzlaff, K. Pearson, and H. Diggelmann. 1983. Nucleotide sequence of the 5' noncoding region and part of the *gag* gene of mouse mammary tumor virus; identification of the 5' splicing site for subgenomic mRNAs. *Nucleic Acids Res.* **11**:6943-6955.
14. Fasel, N., K. Pearson, E. Buetti, and H. Diggelmann. 1982. The region of mouse mammary tumor virus DNA containing the long terminal repeat includes a long coding sequence and signals for hormonally regulated transcription. *EMBO J.* **1**:3-7.
15. Foulds, L. 1949. Mammary tumours in hybrid mice: the presence and transmission of the mammary tumour agent. *Br. J. Cancer* **3**:230-239.
16. Grandgenett, D., T. Quinn, P. J. Hippenmeyer, and S. Oroszlan. 1985. Structural characterization of the avian retrovirus reverse transcriptase and endonuclease domains. *J. Biol. Chem.* **260**:8243-8249.
17. Henderson, L. E., R. Sowder, G. Smythers, and S. Oroszlan. 1983. Terminal amino acid sequences and proteolytic cleavage sites of mouse mammary tumor virus *env* gene products. *J. Virol.* **48**:314-319.
18. Herr, W. 1984. Nucleotide sequence of AKV murine leukemia virus. *J. Virol.* **49**:471-478.
19. Jacks, T., and H. E. Varmus. 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science* **230**:1237-1242.
20. Kennedy, N., G. Knedlitschek, B. Groner, N. E. Hynes, P. Herrlich, R. Michalides, and A. J. J. van Ooyen. 1982. Long terminal repeats of endogenous mouse mammary tumour virus contain a long open reading frame which extends into adjacent sequences. *Nature (London)* **295**:622-624.
21. Klemenz, R., M. Reinhardt, and H. Diggelmann. 1981. Sequence determination of the 3' end of mouse mammary tumor virus RNA. *Mol. Biol. Rep.* **7**:123-126.
22. Kohno, T., and J. R. Roth. 1978. A salmonella frameshift suppression that acts at runs of A residues in the messenger RNA. *J. Mol. Biol.* **126**:37-52.
23. Kozak, M. 1984. Compilation and analysis of sequences upstream from the transcriptional start site in eucaryotic mRNAs. *Nucleic Acids Res.* **12**:857-872.
24. Lobel, L. I., and S. P. Goff. 1985. Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. *J. Virol.* **53**:447-455.
25. Majors, J. E., and H. E. Varmus. 1981. Nucleotide sequences at host-proviral junctions for mouse mammary tumour virus. *Nature (London)* **289**:253-258.
26. Majors, J. E., and H. E. Varmus. 1983. Nucleotide sequencing of an apparent proviral copy of *env* mRNA defines determinants of expression of the mouse mammary tumor virus *env* gene. *J. Virol.* **47**:495-504.
27. Mann, R. S., R. C. Mulligan, and D. Baltimore. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **32**:871-879.
28. Melton, D. A., P. A. Krieg, M. R. Robagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7035-7056.
29. Moore, D. H., C. A. Long, A. A. Vaidya, J. B. Sheffield, A. S. Dion, and E. Y. Lasfargues. 1979. Mammary tumor viruses. *Adv. Cancer Res.* **29**:347-418.
30. Moore, R., G. Casey, S. Brookes, M. Dixon, G. Peters, and C. Dickson. 1986. Sequence, topography and protein coding potential of mouse *int-2*: a putative oncogene activated by mouse mammary tumour virus. *EMBO J.* **5**:919-924.
31. Panganiban, A. T., and H. M. Temin. 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. *Cell* **36**:673-679.
32. Panganiban, A. T., and H. M. Temin. 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* **81**:7885-7889.
33. Peters, G., S. Brookes, R. Smith, and C. Dickson. 1983. Tumorigenesis by mouse mammary tumor virus: evidence for a common integration region for provirus integration in mammary tumors. *Cell* **33**:369-377.
34. Peters, G., and C. Glover. 1980. tRNA's and priming of RNA-directed DNA synthesis in mouse mammary tumor virus. *J. Virol.* **35**:31-40.
35. Peters, G., M. Placzek, S. Brookes, C. Kozak, R. Smith, and C. Dickson. 1986. Characterization, chromosomal assignment, and segregation analysis of endogenous proviral units of mouse mammary tumor virus. *J. Virol.* **59**:535-544.
36. Philipson, L., P. Anderson, U. Olshevsky, R. Weinberg, D. Baltimore, and R. Gesteland. 1978. Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the *gag-pol* polypeptide with yeast suppressor tRNA. *Cell* **13**:189-199.
37. Redmond, S. M. S., and C. Dickson. 1983. Sequence and expression of the mouse mammary tumour virus *env* gene. *EMBO J.* **2**:125-131.
38. Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden. 1985. The *gag* and *pol* genes of bovine leukemia virus: nucleotide sequence and analysis. *Virology* **142**:357-377.
39. Ringold, G. M. 1983. Regulation of mouse mammary tumor virus gene expression by glucocorticoid hormones. *Curr. Top. Microbiol. Immunol.* **106**:79-103.
40. Sagata, N., T. Yasunaga, J. Tsuzuku-Kawamura, K. Ohishi, Y. Ogawa, and Y. Ikawa. 1985. Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. *Proc. Natl. Acad. Sci. USA* **82**:677-681.
41. Schultz, A. M., and S. Oroszlan. 1983. In vivo modification of retroviral *gag* gene-encoded polypeptides by myristic acid. *J. Virol.* **46**:355-361.
42. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* **32**:853-869.
43. Schwartzberg, S., J. Colicelli, and S. P. Goff. 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* **37**:1043-1052.
44. Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida. 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc. Natl. Acad. Sci. USA* **80**:3618-3622.
45. Sen, G. C., J. Racevskis, and N. H. Sarkar. 1981. Synthesis of

- murine mammary tumor viral proteins in vitro. *J. Virol.* 37:963-975.
46. Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura. 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. *Proc. Natl. Acad. Sci. USA* 82:3101-3105.
 47. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* 293:543-548.
 48. Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson. 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* 42:369-382.
 49. Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson. 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* 45:375-385.
 50. Sorge, J., and S. H. Hughes. 1982. Polypurine tract adjacent to the U3 region of the Rous sarcoma virus genome provides a *cis*-acting function. *J. Virol.* 43:482-488.
 51. Staden, R. 1980. A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* 8:3673-3694.
 52. Toh, H., M. Ono, K. Saigo, and T. Miyata. 1985. Retroviral protease-like sequence in the yeast transposon Ty1. *Nature (London)* 315:691.
 53. Ucker, D. S., S. R. Ross, and K. R. Yamamoto. 1981. Mammary tumor virus DNA contains sequences required for its hormone regulated transcription. *Cell* 27:257-266.
 54. Van Ooyen, A. J. J., R. J. A. M. Michalides, and R. Nusse. 1983. Structural analysis of a 1.7-kilobase mouse mammary tumor virus-specific RNA. *J. Virol.* 46:362-370.
 55. Watanabe, S., and H. M. Temin. 1982. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5' long terminal repeat and the start of the *gag* gene. *Proc. Natl. Acad. Sci. USA* 79:5986-5990.
 56. Weiss, R. B. 1984. Molecular model of ribosome frameshifting. *Proc. Natl. Acad. Sci. USA* 81:5797-5801.
 57. Wheeler, D. A., J. S. Butel, D. Medina, R. D. Cardiff, and G. L. Hager. 1983. Transcription of mouse mammary tumor virus: identification of a candidate mRNA for the long terminal repeat gene product. *J. Virol.* 46:42-49.
 58. Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan. 1985. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. USA* 82:1618-1622.
 59. Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan. 1985. Translational readthrough of an amber termination coding during synthesis of feline leukemia virus protease. *J. Virol.* 55:870-873.
 60. Yoshinaka, Y., I. Katoh, T. D. Copeland, G. W. Smythers, and S. Oroszlan. 1986. Bovine leukemia virus protease: purification, chemical analysis, and in vitro processing of *gag* precursor polyproteins. *J. Virol.* 57:826-832.