Molecular Cloning and Complete Nucleotide Sequence of an Adult T Cell Leukaemia Virus/Human T Cell Leukaemia Virus Type I (ATLV/HTLV-I) Isolate of Caribbean Origin: Relationship to Other Members of the ATLV/HTLV-I Subgroup

By KARIM T. A. MALIK,* JOS EVEN† AND ABRAHAM KARPAS

Department of Haematological Medicine, Cambridge University Clinical School, Hills Road, Cambridge CB2 2QL and 1Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QL, U.K.

(Accepted 5 April 1988)

SUMMARY

We report the first complete nucleotide sequence of an adult T cell leukaemia virus/human T cell leukaemia virus type I (ATLV/HTLV-I) isolate from a British patient of Caribbean origin. Sequence comparisons of our proviral clone (HS-35) with other molecular clones are shown. We note the strong sequence conservation between isolates of Caribbean and Japanese origin (2-3% divergence), but demonstrate the higher homologies existing between isolates originating from similar geographical areas (approximately 1% divergence). Implications for the origin, evolution and dissemination of the ATLV/HTLV-I subgroup are discussed. Analysis of defective proviral clones isolated from the same genomic library is also reported, and suggests a pattern of proviral sequence deletions during the biogenesis of defective proviruses.

INTRODUCTION

Adult T-cell leukaemia virus/human T cell leukaemia virus type I (ATLV/HTLV-I) is an exogenous human retrovirus implicated in the aetiology of adult T cell leukaemia (ATL) (Hinuma et al., 1981; Yoshida et al., 1982), a distinct form of human leukaemia first recognized in Japan (Uchiyama et al., 1977). ATL patients generally have antibodies to ATLV/HTLV-I structural proteins and epidemiological studies have revealed a high incidence of ATL in south-west Japan (Tajima et al., 1981; Hinuma et al., 1982). Subsequently, a virus similar to ATLV/HTLV-I was found in black people from the West Indies (Blattner et al., 1982; Catovsky et al., 1982) and Africa (Fleming et al., 1983; Hunsmann et al., 1983), where a much lower degree of infection was observed.

The molecular characterization of several ATLV/HTLV-I isolates has been documented. The first provirus cloned was from the MT-1 cell line derived from the leukaemic cells of a Japanese ATL patient (Miyoshi et al., 1980). The proviral clone was designated ATM-1, and its long terminal repeat (LTR) region was sequenced (Seiki et al., 1982). The first complete nucleotide sequence of an integrated provirus, ATK, from a Japanese patient's leukaemic cells was later published (Seiki et al., 1983), and still represents the only complete sequence available. Partial sequences have been reported for an American isolate, CR-1 (Josephs et al., 1984), a Zairian isolate, MC-1 (Ratner et al., 1985), and another Japanese isolate, HY-4 (Hiramatsu et al., 1987). The sequences are strongly conserved between isolates, and ATLV/HTLV-I has been shown to be related to HTLV-II (Shaw et al., 1984), simian T cell leukaemia virus (STLV) (Komuro et al., 1984), and more distantly to bovine leukaemia virus (BLV) (Sagata et al., 1984).

The origin of ATLV/HTLV-I remains unclear. It has been reported that the Japanese and American isolates are the same strain of retrovirus (Watanabe et al., 1984), are more closely related to the African subtype of STLV than the Asian subtype (Watanabe et al., 1986), and that the human virus has been disseminated from an African origin (Ito, 1985).
We have previously reported the establishment of two ATLV/HTLV-I-transformed umbilical cord blood T cell lines, Karpas 1010 and Karpas 1025, each producing a separate ATLV/HTLV-I isolate (Karpas et al., 1987). Both were isolated from British patients of Caribbean origin. In this paper, we report the first complete nucleotide sequence of a proviral clone (HS-35) of non-Japanese origin, with a view to understanding the relationship, origin and evolution of ATLV/HTLV-I isolates from diverse geographical regions. The isolation and possible significance of integrated defective proviruses is also discussed.

METHODS

Cells. The ATLV/HTLV-I-producing cell lines Karpas 1025 and Karpas 1010 were established by co-cultivation of umbilical cord blood cells with leukaemic cells from ATL patients of Caribbean origin (Karpas et al., 1987). Maintenance of cells was in RPMI 1640 medium containing 10% foetal calf serum.

Molecular cloning. High molecular weight DNA was extracted from Karpas 1025 and 1010 cells by the method of Blin & Stafford (1976), and was then partially digested with MboI, and size fractionated on a 0.45% low melting point agarose gel. Gel fractions varying from 16 to 23 kb were excised, melted and purified by chromatography through NACS-52 PREPAC cartridges (Gibco-Bethesda Research Laboratories). Genomic insert DNA was ligated to BamHI-digested bacteriophage λ2001 DNA (Karn et al., 1984), packaged in vitro and plated on Escherichia coli host strain Q359 (Karu et al., 1983). ATLV/HTLV-I probes (kindly supplied by Dr M. Hatanaka) were nick-translated (sp. act. approx. 108 c.p.m./μg) according to Rigby et al. (1977), and used to screen 2 x 106 recombinant phage. Hybridizations were at 65 °C in 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 6 × standard saline citrate (SSC; 0.9 M-NaCl, 0.01 M-trisodium citrate), 0.01 M-EDTA and 0.5% SDS. Filters were washed in 0.1 x SSC and 0.5% SDS at 65 °C, and exposed on preflashed X-ray film at -70 °C. Positive clones were analysed by restriction mapping using ATLV/HTLV-I subgenomic probes, and proviral restriction fragments were subcloned into appropriately digested pUC plasmids.

Nucleotide sequencing strategy. Viral fragments derived from a full-length clone (HS-35) were sequenced by the dideoxy chain terminator procedure (Sanger et al., 1977) after shotgun cloning in bacteriophage M13mp18 (Yanisch-Perron et al., 1985). Sequence data were compiled and analysed by using the computer programs of Staden (1982, 1984) and any regions of ambiguity were resolved using dITP (Mills & Kramer, 1979).

RESULTS

Screening of 106 recombinant phage comprising the Karpas 1025 genomic library resulted in the isolation of 30 ATLV/HTLV-I-positive clones. Only one of these clones, HS-23, was found to contain pol gene sequences. Mapping of the remaining defective proviral clones indicated that the predominant deletion extended from within the gag gene to the env-pX-I region. The LTRs, part of the gag gene, and most of the sequences downstream of the env gene tended to be conserved; all clones had LTR sequences, and only three clones lacked the pX sequences from nucleotides 6736 to 7286 (Fig. 2) homologous to the pX region probe employed (data not shown). This amplification of defective proviruses is in accordance with the observations of Hiramatsu & Yoshikura (1986), who showed approximately 40% defective proviral insertions following experimental transmission of ATLV/HTLV-I, although our library yielded a far higher percentage of defective proviruses displaying an apparent pattern of deletion.

As clone HS-23 also proved to be defective, a second library was constructed and screened with a pol gene-specific probe, in order to isolate a full-length provirus. Two more clones were isolated from screening of a further 106 recombinant phage, and one clone, designated HS-35, was judged to contain a full-length provirus, on the basis of restriction enzyme analysis using ATLV/HTLV-I subgenomic probes. Fig. 1 shows the alignment of the HS-35 LTR sequence with other LTR sequences. 1010/3 represents sequences derived from a defective proviral clone isolated from the Karpas 1010 cell line. The promoter and tat-responsive sequences (Seiki et al., 1982; Fujisawa et al., 1986; Rosen et al., 1987) are strongly conserved between isolates, as is the proline tRNA primer-binding site 3' of the 5' LTRs. The degree of base substitution between LTR sequences is summarized in Table 1.

A higher degree of conservation is apparent within the two ‘Japanese type’ clones (ATK and ATM), and the two ‘Caribbean type’ clones (HS-35 and 1010/3). Approximately twice the base substitution rate is seen when comparing either of our Caribbean isolates with either of the
Fig. 1. Sequence alignments of LTRs derived from different isolates (see text). Lower case letter and vertical arrow denote an insertion, (d) denotes deletions. Conserved imperfect 21 bp repeats are marked by horizontal arrows. The primer-binding site (PBS) is underlined.
Fig. 2. Complete remaining sequence of clone HS-35 compared with available sequences; (>) and (<) mark the start and end points, respectively, of the other sequences. The viral genes are labelled with PR-ORF representing the protease ORF, and pX representing the third pX cDNA exon. Deletions are shown by (d) or by (---). (***) marks the stop protease ORF codon eliminated in Caribbean isolates.
Table 1. Summary of sequence variations between ATLV/HTLV-I isolates

<table>
<thead>
<tr>
<th></th>
<th>LTR</th>
<th>Protease</th>
<th>env</th>
<th>env-pX</th>
<th>pX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS-35</td>
<td>1010</td>
<td>ATK</td>
<td>ATM</td>
<td>CR-1</td>
</tr>
<tr>
<td>HS-35</td>
<td>* 10</td>
<td>22 20 24</td>
<td>* 4</td>
<td>11 14</td>
<td>* 9† 13‡ 21† 13‡</td>
</tr>
<tr>
<td>1010</td>
<td>1.3 * 20</td>
<td>19 24 2 9</td>
<td>12</td>
<td>1† 4‡</td>
<td>* 20 14‡</td>
</tr>
<tr>
<td>ATK</td>
<td>2.8 2.6</td>
<td>* 7 16 4 2</td>
<td>* 7</td>
<td>7† 4‡</td>
<td>7† 3‡ * 3.0 3.2 * 52</td>
</tr>
<tr>
<td>ATM</td>
<td>2.6 2.5</td>
<td>0.9 * 15</td>
<td>–§ – – –</td>
<td>– –</td>
<td>– – –</td>
</tr>
<tr>
<td>CR-1</td>
<td>3.2 3.2</td>
<td>2.1 2.0</td>
<td>– – – –</td>
<td>– –</td>
<td>– – –</td>
</tr>
<tr>
<td>HY-4</td>
<td>– – –</td>
<td>– – –</td>
<td>6 4 4</td>
<td>– – –</td>
<td>– – –</td>
</tr>
<tr>
<td>MC-1</td>
<td>– – –</td>
<td>– – –</td>
<td>– – –</td>
<td>7.2 7.2</td>
<td>8.3 9 9</td>
</tr>
</tbody>
</table>

* Above asterisk, values indicate number of base changes between respective isolates. Values below asterisk indicate either percentage variation (for non-coding regions LTR and env-pX) or number of amino acid differences (for protease, env and pX regions) between the various isolates.
† Changes in surface glycoprotein (gp46).
‡ Changes in transmembrane protein (p21).
§ – Indicates no available data.
Japanese isolates. The American isolate, CR-1, is shown to be distinct from the Japanese and Caribbean isolates at this level of analysis.

The whole remaining HS-35 sequence is shown in Fig. 2, ending at the termination codon of p40x within the 3' LTR (Seiki et al., 1985). Comparisons with other published sequences are shown.

The gag gene sequences show 2-2% variation between isolates HS-35 and ATK. Assignment of the three gag proteins is according to Oroszlan & Copeland (1985). Only four amino acid (aa) changes are seen. The nucleic acid-binding protein p15 is the most strongly conserved with no changes, whereas the major core protein, p24, and p19 show one aa and three aa changes respectively.

Analysis of the gag-pol region of HS-35 reveals an extension to the open reading frames (ORFs) previously reported for HY-4 (Hiramatsu et al., 1987) and PT (STLV) (Inoue et al., 1986), arising from a point mutation at nucleotide position 2050 (see Fig. 2). This results in a codon alteration of TGA to CGA (termination to arginine), allowing an ORF commencing at nucleotide position 1849 to be read through into the postulated protease-encoding domain. Clone 1010/3 is deleted in the gag-pol region, but another molecular clone representative of the I010 ATLV/HTLV-I strain, referred to as 1010/2, allowed the analysis of the protease ORF of the 1010 provirus. Sequencing of the gag-pol region of 1010/2 revealed an equivalent mutation, suggesting that it may be characteristic of Caribbean strains of ATLV/HTLV-I. An analysis of protease gene base substitutions and aa changes between isolates is shown in Table 1. The conservation within isolates of similar geographical origin is again apparent, the nucleotide variation being between 0-6% (HS-35 and 1010/2) and 1% (HY-4 and ATK), whereas HS-35 differs from ATK and 1010/2 by 1-6% and 2% respectively. 1010/2 is marginally more similar to the Japanese clones (1.3 and 1-7% variation from ATK and HY-4 respectively). These results are reflected to a lesser degree in the aa substitutions. Four-hundred and twenty-five bp of the protease gene from the Zairian isolate, MC-1, are also shown in Fig. 2. Surprisingly, this isolate is missing one nucleotide from the same position as the error in the original sequence of clone ATK (Seiki et al., 1983; Inoue et al., 1986), and is presumably defective. Comparison of the available 425 bp indicates a variation of 1-6% from ATK, 2-4% from HY-4 and 1010/2, and 2-8% from HS-35.

The size of the HS-35 pol gene ORF is identical to that of ATK, but differs in its sequence by 2-1%, giving rise to 11 aa changes. The homologies of the ATLV/HTLV-I polymerase to other retrovirus polymerases are, however, maintained (Haseltine et al., 1985).

The pattern of low sequence divergence between the Caribbean isolates is continued across the env gene (see Table 1): the HS-35 and 1010/3 genes differ by 1.5%, whereas comparison of both these isolates with the ATK sequence shows a variation of 2.4%. We note that the sequence coding for the transmembrane env protein is not preferentially conserved between the Caribbean isolates (2.4% divergence), compared with strong conservation of the surface glycoprotein coding sequence (1% divergence). The base substitutions displayed by ATK span both env protein domains (Oroszlan & Copeland, 1985), causing seven aa changes in the env surface glycoprotein; the 1010/3 env gene contains an amber mutation, and shows one conservative aa change from the HS-35 surface glycoprotein. These variations in the surface glycoprotein primary structure may be important to the design of immunogens for immunological diagnosis and therapy of ATL; env gene aa sequences are shown in Fig. 3. The transmembrane protein sequences vary by three to four aa amongst all isolates, but homologies with other retrovirus transmembrane proteins remain (Ciancolo et al., 1984; Patarca & Haseltine, 1984).

The pX genes of ATLV/HTLV-I are expressed by means of a two-step splicing mechanism (Seiki et al., 1985), with one polycistronic mRNA species being translated into both the p40x and p27x proteins (Nagashima et al., 1986). Both are nuclear proteins (Slamon et al., 1985; Kiyokawa et al., 1985) essential for the efficient replication of ATLV/HTLV-I (Inoue et al., 1987). The observed variation across the third exon of the pX mRNA between the Caribbean isolates and ATK is very low (see Table 1), exhibiting a sequence divergence of only 1%. HS-35 and 1010/3 differ by a similar degree. However, clone MC-1 varies by 2-3 to 2-8% from the other three sequences, this being reflected in a greater degree of aa substitution.

ATLV/HTLV-I nucleotide sequence
In view of the key role of \( pX \) gene products in ATLV/HTLV-I replication (Inoue et al., 1987) and oncogenesis (Maruyama et al., 1987; Nerenberg et al., 1987), it was of interest that many of the defective proviruses from the 1025 genomic library appeared similar in genome organization, especially in the conservation of the \( pX \) sequences. We examined the sequences demarcating deletions in seven of the 26 defective clones displaying an LTR–gag–env/\( pX \)–LTR genome structure, and found that the deletion commenced at, or near, the pentanucleotide ATTAA at nucleotide position 989 in five of the seven defective proviruses (see Fig. 4). These five clones could be further divided into two subtypes, based on the 3' end of the deletion. HS-1, HS-2 and HS-17 all used a repeat of the 5' motif (ATTAA) at nucleotide position 6701 at the 3' end of the deletion, whereas HS-8 and HS-16 recommenced at position 6869. The host flanking sequences of the five clones do not appear to be shared, suggesting that the proviruses represent independent integrations (data not shown). Additionally, HS-8 and HS-16 differ by a point mutation 5' to the deletion at position 975. The deletion in HS-10 commenced near the sequence GTTAA at position 937, and ended at nucleotide position 6564. It should be emphasized that the precise locations of the deletions cannot be ascertained due to the direct repeats at their ends. Clone HS-13 was also deleted between the \( gag \) and \( pX \) sequences, but did not share the sequence specificities at the deletion termini (Fig. 4).
DISCUSSION

Although we have no data on the biological activity of HS-35, we believe that it constitutes a complete provirus on the basis of its length and genomic integrity. Fig. 5 shows a comparison of the ORFs of HS-35 and ATK (Seiki et al., 1983). The sequence of ATK has been corrected for the nucleotide that was missing in the protease gene of the original sequence (Inoue et al., 1986).

The results of this paper emphasize the close relationship between isolates of ATLV/HTLV-I from diverse geographical regions, but also show that subdivisions, demonstrable by nucleic acid sequencing, exist in this group of retroviruses. The homology of the Caribbean isolates is consistently greater than the homology between Caribbean and Japanese isolates, indicating that the viruses from these two regions are evolutionarily distinct. Fig. 6 qualitatively summarizes the nucleotide variations between different strains of ATLV/HTLV-I with respect to the prototype sequence of Seiki et al. (1983), and demonstrates that the Caribbean isolates display a pattern of sequence divergence spanning the proviral genome.

In terms of gene organization, the most interesting difference observed is the 5' extension of the protease ORF of HS-35 and 1010/2. Although the mechanism of ATLV/HTLV-I protease gene expression is not yet established, frameshift suppression of the gag termination codon has been suggested for ATLV/HTLV-I (Rice et al., 1985), allowing the expression of a gag-protease fusion protein as in the case of BLV (Yoshinaka et al., 1986), mouse mammary tumour virus (Moore et al., 1987), and human immunodeficiency virus type I (Jacks et al., 1988). The 5' end of the extended ORF of HS-35 and 1010/2 also contains the sequence AAAAAAC implicated in ribosomal frameshifting, and is followed by an imperfect repeat (CCCCCCCC/GGTCCGGGG) that may form a stem-loop structure to facilitate frameshifting and expression of the extended ORF. However, we have not evaluated the thermodynamic stability of this structure, and we note that BLV (Sagata et al., 1985) and HTLV-II (Shimotohno et al., 1985) do not conserve frameshift consensus sequences at these positions. The possible expression of these sequences must await further experimental evidence.
Detailed comparisons with other known sequences of ATLV/HTLV-I have not enabled us to delineate unequivocally the source and pattern of ATLV/HTLV-I dissemination, although several points of interest emerge. We have compared all the LTR sequences used in Fig. 1 with the LTRs of chimpanzee and African green monkey STLV isolates (Watanabe et al., 1986), and found all the human sequences to be equally divergent from the African subtype STLV LTRs, confirming and extending the findings of Watanabe et al. (1986) to human isolates from the Caribbean region.

To our surprise, the American isolate, CR-1, appears more closely related to the Japanese isolates (ATK and ATM) than to the Caribbean isolates (HS-35 and 1010/3). If this observation is confirmed by the sequencing of further American isolates, it seems probable that the ATLV/HTLV-I associated with ATL in the U.S.A. (Blayney et al., 1983) has evolved separately from the Caribbean type retrovirus, and is not associated with populations migrating between these regions (Catovsky et al., 1982; Vyth-Dreese et al., 1983).

Comparisons with MC-1 sequences show no consistent pattern of homology to other sequences. In the protease ORF, MC-1 seems most related to ATK as opposed to the other Japanese isolate, HY-4, and our Caribbean isolates, HS-35 and 1010/2. The inconsistency of the relationship between MC-1 and the two Japanese isolates is surprising in view of our results demonstrating a characteristic variation between isolates from the Caribbean and Japan, spanning the entire genome. This may be attributable to the length of the sequence available for comparison.
A summary of sequence divergence covering the region between the \textit{env} gene and the third \textit{pX} exon (Seiki \textit{et al.}, 1985) is shown in Table 1. Clone MC-1, contrary to the homologies displayed in the protease ORF, shows greatest divergence from clone ATK. We note that the deletions shown by MC-1 in this region predate this clone on an evolutionary scale with respect to other human isolates, as the Asian subtype of STLV also has deletions in this region (Watanabe \textit{et al.}, 1985). The region of strongest genetic conservation between Caribbean and Japanese isolates is within the \textit{pX} exon beginning at nucleotide position 7304 and ending at position 8358 (1\% divergence), supporting the results of Inoue \textit{et al.} (1987), which suggest that proteins expressed by this region are essential for viral replication. The sequence of MC-1 is marginally more similar to that of clone ATK.

The degree of genetic conservation within geographically and evolutionarily distinct members of the ATLV/HTLV-I subgroup, especially in genes critical for viral replication and host–virus interaction (Inoue \textit{et al.}, 1987; Maruyama \textit{et al.}, 1987) suggests that the variation in infection rates shown by ATLV/HTLV-I in different endemic areas (Ito, 1985) may be attributable to racial, ethnic and/or environmental cofactors (Ito \textit{et al.}, 1983). However, we cannot evaluate phenotypic effects which may be exerted by alterations such as those observed in the \textit{env} gene (see Fig. 3).

Although the mode of generation of defective proviruses is unclear, Olsen & Swanstrom (1985) suggested that aberrant cleavage of circular proviral DNA (which serves as a precursor to
linear, integrated DNA) may be responsible for creating defective proviruses. We note the homology of the sequence ATTAA/GTTAA at the 5′ end of deletions in six of seven defective clones (and also repeated at the 3′ end in three of seven clones) with the nucleotides at the LTR–LTR junctions of Moloney murine leukaemia virus, spleen necrosis virus and avian leukosis virus circular proviral DNAs (Panganiban et al., 1985). This supports the theory that defective proviruses may be generated by cleavage of circular proviruses, probably by the pol gene-associated endonuclease (Grandegenett et al., 1986). Interestingly, the distribution of the ATTAA motif coincides with the deletions observed in the defective proviruses, i.e. all nine ATTAA sites occur between nucleotides 989 and 6701. The 3′ ends of the deletions in HS-8, HS-10, HS-13 and HS-16 may represent secondary, but related specificities (Grandegenett et al., 1986). These observations may explain the apparent conservation of the pX sequences in our defective isolates, and those identified by Hiramatsu & Yoshikura (1986).

The amplification of defective proviruses and their remarkable pattern of deletion suggests that they may have a biological function. Although we have not sequenced the defective genomes in their entirety, our mapping studies indicate that none of the clones discussed have deleted sequences other than those defined. It is also unlikely that the remaining sequences are affected by pronounced base alterations, as we detected only two point mutations (in HS-13 and HS-16) in the survey of seven deletion zones, both of these occurring near the 5′ end of the deletion. Expression from defective proviral templates remains to be investigated.

It has been suggested that deletions within the human pvt-like region may occur as a result of excision of ATLV/HTLV-I integrations, and that these deletions may contribute to the development of ATL (Mengle-Gaw & Rabbitts, 1987). We therefore examined whether the amplified defective proviruses might act as insertional mutagens in the pvt-like region. However, none of the 32 ATLV/HTLV-I-positive clones from the 1025 genomic library hybridized to probes spanning the pvt-like region (kindly supplied by Dr L. Mengle-Gaw). The proviral sequences were also negative for myc gene sequences (data not shown).

Our results show that it may be possible to delineate the precise distribution of individual isolates of ATLV/HTLV-I. Clearly such molecular characterization studies with further isolates of ATLV/HTLV-I and STLV, especially those of African origin, are possible and, indeed, necessary to define the ancestry and evolution of this unique group of retroviruses. Furthermore, the amplification of defective proviruses following in vitro transmission of ATLV/HTLV-I may reflect specific early events following infection in vivo and may contribute to the development of ATL.

We would like to thank Dr G. Peters and Dr G. Winter for helpful discussions and comments, Drs L. Mengle-Gaw and T. H. Rabbitts for pvt-like region and c-myc probes, and Dr M. Hatanaka for ATLV/HTLV-I probes. We would also like to thank A. Keller, A. McIntosh, Y. Nagai and S. Powell for assistance in data compilation, and M. Jefsioutine for assistance in manuscript preparation. K.T.A,M, was the recipient of a Medical Research Council (U.K.) Research Studentship.

REFERENCES


