Molecular Cloning of the Closed Circular Provirus of Human T Cell Leukaemia Virus Type I: A New Open Reading Frame in the gag–pol Region

By KEICHI HIRAMATSU, JUNJI NISHIDA, AYAKO NAITO AND HIROSHI YOSHIKURA

1 Department of Microbiology, Faculty of Dentistry, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, 2 The Third Department of Internal Medicine and 3 Department of Bacteriology, Faculty of Medicine, Tokyo University, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

(Accepted 24 September 1986)

SUMMARY

A DNA clone of human T cell leukaemia virus type I (HTLV-I) was isolated from extrachromosomal closed circular copies in chronically infected promyelocytic leukaemia HL60 cells. The new HTLV-I isolate had an intact reading frame in the gag–pol region which could encode protein of 234 amino acids. This open reading frame has not been observed in previous HTLV-I isolates, although similar open reading frames have been reported in the corresponding locations in the related bovine leukaemia virus and HTLV type II. We consider that this open reading frame codes for the virus-encoded protease, on the basis of the homology of the predicted amino acid sequence with those of previously identified retrovirus proteases.

The cloning and sequence determination of two independent molecular clones of human T cell leukaemia virus type I (HTLV-I), λATK-I and λMC-I, has been reported. These clones have provided us with the general gene structure of the first human cancer-causing retrovirus. It seems premature, however, to conclude that they represent intact genomes of HTLV-I. For example, the pX-I open reading frame is present in λATK-I but absent from λMC-I (Ratner et al., 1985b). It is not unreasonable to suspect that these molecular clones could be defective pseudotypes in view of the fact that they were cloned from the integrated proviruses of adult T cell leukaemia (ATL) cells. It has been reported that the integrated proviruses of ATL patients' lymphocytes frequently have deletions (as high as 60% in those from the Nagasaki area; Yoshida et al., 1985). We have also shown that, during in vitro transmission, as many as 40% of the integrated proviruses are defective (Hiramatsu & Yoshihara, 1986). The reported inability of the clone λATK-I to produce infectious virus by DNA transfection might be due to some defect in the provirus genome (Inoue et al., 1986). In view of this, we have attempted to clone the HTLV-I genome from extrachromosomal copies, because the majority of them may represent intact virus copies, being active replication intermediates, and also because they are uniform in size in contrast to the great diversity of genome sizes among their integrated counterparts in chronically infected HL60 cells (Hiramatsu et al., 1986).

We have reported the in vitro transmission of an infectious HTLV-I isolate from the peripheral blood lymphocytes of an ATL patient to a human diploid fibroblast line, IMR90 (Yoshikura et al., 1984). Ninety-five percent of the patient's peripheral blood lymphocytes were leukaemic cells of monoclonal origin which contained a single copy of HTLV-I in the cellular genome. HTLV-I was further transmitted to HL60 cells by co-culture with the virus-infected IMR90. Some of the HL60 clones thus obtained were shown to produce infectious HTLV-I and also to contain extrachromosomal copies of the provirus even after 6 months' culture (Hiramatsu et al., 1986). The extrachromosomal proviruses were extracted (Hirt, 1967) from one of the HL60 clones, clone 39, and circular proviruses were further purified by CsCl-ethidium bromide isopycnic centrifugation, linearized and cloned into the Xbal site of the λWESAB phage vector.
Fig. 1. Restriction maps of \( \lambda HY-4 \) and two other isolates of HTLV-I and the strategy for sequencing the \( gag-pol \) region. For the purpose of comparison, the insert of \( \lambda HY-4 \) was rearranged into the genome order of the other isolates. Only the restriction enzyme sites of \( \lambda ATK-1 \) are fully described. Common enzyme sites are indicated by interrupted lines. The arrows under the map indicate the direction and extent of the DNA sequences which were determined by the dideoxynucleotide method (Schreier & Cortese, 1979); a modified method using dITP was employed to confirm the entire sequence (Mills & Kramer, 1979).

Fig. 2 shows the sequence of the \( gag-pol \) region of clone \( \lambda HY-4 \) in comparison with the two other clones (see Fig. 1 for the sequence strategy). It was noted that there was an open reading frame which overlapped 37 bp of the \( gag \) coding frame and 256 bp of the \( pol \) frame, utilizing a total of 702 bp in a reading frame (frame 3) different from those of the \( gag \) (frame 1) and \( pol \) (frame 2) genes. The previously isolated molecular clones did not have this long open reading frame. A premature termination codon was present at position 388 in \( \lambda ATK-1 \), and a single base deletion had occurred at position 244 in \( \lambda MC-1 \). The latter deletion altered the reading frame and caused premature termination at position 266. Apart from this major difference, the nucleotide sequences of the three independent HTLV-I clones were very similar (99% homology between \( \lambda HY-4 \) and \( \lambda ATK-1 \); 98% homology between \( \lambda HY-4 \) and \( \lambda MC-1 \)).

Recently, the corresponding region of simian T cell leukaemia virus (STLV) has been shown to have an open reading frame considered to encode a protease (Inoue et al., 1986). According to the sequence data, the STLV open reading frame started at the same position as that of HTLV-I (37 bp upstream of the \( gag \) termination), and ended 199 bp downstream of the \( pol \) start site. The codon usage and location of the STLV open reading frame were similar to those of HTLV-I. The nucleotide sequence homology of the STLV reading frame with those of \( \lambda ATK-1 \) and \( \lambda HY-4 \) was 90% and 91%, respectively. This extensive conservation of nucleotide sequences among different isolates of HTLV-I and STLV signifies the biological importance of the protein encoded by this open reading frame.

Fig. 3 shows the predicted amino acid sequence of the above open reading frame in comparison with those of protease coding frames of HTLV-I related retroviruses, HTLV type II (Shimotohno et al., 1985) and bovine leukaemia virus (BLV) (Rice et al., 1985). There is an extensive homology of the amino acid sequences (59% homology with HTLV-II, 41% homology with BLV). It was also noted that there were three stretches of amino acid sequences well conserved among the proteases of more distantly related retroviruses, i.e. murine leukaemia
Fig. 2. Nucleotide sequence of the \textit{gag-pol} region of λHY-4 and comparison with two other isolates of HTLV-I. The sequence from 40 bp upstream of the \textit{gag} termination to 259 bp downstream of the \textit{pol} start site is shown. There is a long open reading frame (fr.) from positions 1 to 702. For λMC-1 and λATK-1, only the nucleotides different from those of λHY-4 are designated. The stop codons are indicated by *. The stop codon at position 389 is a nonsense substitution in λATK-1. A deletion in the λMC-1 sequence is indicated by d at position 244.
Fig. 3. Predicted amino acid sequence of the new reading frame of 2HY-4 compared with those of the corresponding regions of other retroviruses. Computer analysis for the homology of the sequence was done using the 'gap' program (Needleman \& Wunsch, 1970). Common amino acids are boxed. The region corresponding to the purified mature BLV protease is underlined. The closed circles under amino acid residues indicate conservation in RSV, MLV and HTLV-III/LAV also, although these sequences are not shown: large circles indicate residues conserved in these three viruses; small circles indicate residues identical in two of the three and substituted in the other by an amino acid of the same nature.

virus (MLV) (Shinnick \textit{et al.}, 1981), Rous sarcoma virus (RSV) (Schwartz \textit{et al.}, 1983), and HTLV-III/LAV (Ratner \textit{et al.}, 1985\textit{a}; Sanchez-Pescador \textit{et al.}, 1984). (In the case of HTLV-III/LAV, the first 318 bp of the \textit{pol} region were used for comparison, although the occurrence of a protease encoded by this virus remains inconclusive.)

The virus-encoded protease of BLV was recently isolated, and its amino acid sequence has been determined (Yoshinaka \textit{et al.}, 1986). The sequence corresponded well to the nucleotide sequence located in the middle part of the BLV open reading frame discussed above (shown underlined in Fig. 3). Based on these observations it is highly probable that the open reading frame of 2HY-4 encodes a virus-specific protease.

A proteolytic activity required for processing p55\textsubscript{gag} to p24\textsubscript{gag} was present in the cells from which 2HY-4 was isolated. Fig. 4 is a fluorograph of immunoprecipitated samples from two HL60 clones infected with HTLV-I from the same patient’s leukaemic cells. Both HL60 clones were labelled with [\textsuperscript{3}H]leucine and [\textsuperscript{35}S]methionine for various periods, and culture fluids and cell extracts were subjected to immunoprecipitation using the patient’s serum which has been shown to be reactive with p24\textsubscript{gag} (Yoshikura \textit{et al.}, 1984). After 6 h of labelling, p24\textsubscript{gag} and p55\textsubscript{gag} proteins were precipitated from the lysate of clone 39 (lanes 7 and 8). From the culture fluid, only p24\textsubscript{gag} (presumably in the released HTLV-I particles) was precipitated (lanes 2 and 3). The same results were obtained with clone 9 (only 14 h data are shown; lanes 4 and 9). These results indicate that this HTLV-I isolate retained a functional virus-specific protease activity, because, as has been pointed out, proteolytic cleavage of p24\textsubscript{gag} from the p55\textsubscript{gag} precursor cannot be achieved by host cell proteases (Oroszlan \& Copeland, 1985).

We thank M. Yoshida for a DNA probe, and M. Kato for technical assistance. We also thank N. Iizuka and K. Yamamoto for the computer analysis. This work was supported by Grants-in-Aids from the Ministry of Education and the Ministry of Health and Welfare.
Fig. 4. Immunoprecipitation of gag gene products from clones 39 and 9. Immunoprecipitation was done with the following samples: lanes 1 to 3, culture fluids of clone 39 after 1, 6 and 14 h radiolabelling with 50 μCi each of [3S]methionine and [3H]leucine, respectively; lane 4, clone 9 culture fluid after 14 h labelling; lane 5, HL60 culture fluid after 14 h labelling; lanes 6 to 8, clone 39 cell lysates after 1, 6 and 14 h labelling, respectively; lane 9, clone 9 cell lysate after 14 h labelling; lane 10, HL60 cell lysate after 14 h labelling. The cell number used for radiolabelling was 1 × 10^6 for each clone. The fluorograph was exposed for 12 h. Molecular weights are shown × 10^{-3}. The arrows indicate gag gene products corresponding to p55^a- (upper arrow) and p24^a- (lower arrows).

REFERENCES


Short communication

an AIDS-associated retrovirus (ARV-2). Science 227, 484-492.


sequence of the provirus genome integrated in leukemia cell DNA. Proceedings of the National Academy of
Sciences, U.S.A. 80, 3618-3622.

SHIMOTOHNO, K., TAKAHASHI, Y., SHIMIZU, N., GOJOBORI, T., GOLDE, D. W., CHEN, I. S. Y., MIWA, M. & SUGIMURA, T.


New York: Springer-Verlag.

from Japanese adult T-cell leukemia patients in human diploid fibroblast strain IMR90 and the biological

purification, chemical analysis, and in vitro processing of gag precursor polyproteins. Journal of Virology 57,
826-832.

(Received 22 July 1986)