Two distinct polypeptides may be translated from a single spliced mRNA of the X genes of human T-cell leukemia and bovine leukemia viruses

Noriyuki Sagata, Teruo Yasunaga+ and Yoji Ikawa

Laboratory of Molecular Oncology and +*Computation Center, The Institute of Physical and Chemical Research, Wako, Saitama 351-01, Japan*

Received 29 August 1985

Human T-cell leukemia and bovine leukemia viruses have a potential transforming gene, termed X. In addition to the major open reading frame known to encode a functional protein, the X gene harbors another short open reading frame which overlaps this major one. Both of these open reading frames are found on a single spliced X mRNA in a potentially functional form. Circumstantial evidence strongly suggests that they are both translated from the single X mRNA molecule, showing striking similarity to the translation mechanism of an adenovirus Elb gene mRNA. We note that the short open reading frame has the capability to encode a putative nuclear protein with structural features similar to those of an AIDS virus *trans*-acting protein.

HTLV *BLV* *X gene* *Spliced mRNA* *Dual translation* *Nuclear protein*

1. INTRODUCTION

Human T-cell leukemia virus type I (HTLV-I) and bovine leukemia virus (BLV) cause adult T-cell leukemia [1] and enzootic bovine leukosis [2], respectively. These retroviruses and their close relatives simian T-cell leukemia virus (STLV) and human T-cell leukemia virus type II (HTLV-II) all have a potential transforming gene, termed X, between the *env* gene and 3′-long-terminal repeat (LTR) [3–6]. The major long open reading frame in the HTLV X genes encodes a protein of 38–40 kDa [7–9]. We have recently shown that an analogous open reading frame of the BLV X gene also encodes a 38 kDa protein, which is localized in the nucleus of an infected cell [10]. These proteins probably function for viral replication and/or cellular transformation [11–13].

In addition to the major long open reading frame, the X genes of these retroviruses commonly contain a short open reading frame which overlaps this major one [3–6]. Here we show the protein-coding capability of the short open reading frame and suggest that both the long and short open reading frames are translated from a single spliced X mRNA.

2. MATERIALS AND METHODS

2.1. *Northern blotting analysis of BLV RNA*

Poly(A)$^+$ RNA was obtained from BLV-producing cells (FLK-BLV [14]) [10] and subjected to Northern blotting analysis [15] using various portions of the cloned BLV DNA as probes [4].

2.2. *Calculation of nucleotide difference at synonymous sites*

For each pair of aligned nucleotide sequences of homologous genes, the nucleotide difference at synonymous sites was calculated as described [16].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

3. RESULTS AND DISCUSSION

3.1. *Protein-coding capability of the short open reading frame X-II*

Fig.1A schematically shows the structure of the X genes of BLV [4], HTLV-1 [3], STLV [5] and HTLV-II [6]: in addition to the functional major open reading frame (designated X-I frame), all the X genes usually contain another short open reading frame (designated X-II frame) which overlaps the 5′-half of the X-I frame. In all X genes, neither X-I nor X-II frame has an initiator ATG codon. However, a splice acceptor sequence occurs at the 5′-end of each X-I frame, suggesting that at least this frame is expressed as a spliced mRNA.

To assess whether the overlapping X-II frame also encodes a functional protein, we compared the nucleotide sequences of the *gag*, *pol*, *env* and X genes for 3 pairs of viruses: BLV-J (Japanese isolate [4,17]) vs BLV-B (Belgian isolate [18,19]), HTLV-I vs STLV and HTLV-I vs HTLV-II. From these comparisons, we calculated the nucleotide differences at synonymous sites of the respective genes [16]; for the X gene, the calculation was done separately for overlapping (i.e. X-II region) and non-overlapping regions of the X-I frame. Fig.1B shows that the nucleotide differences at synonymous sites of all the genes or regions, except for the X-II region, give nearly identical values in a given pair of viruses; this is consistent with the notion that evolutionary rate of the nucleotide

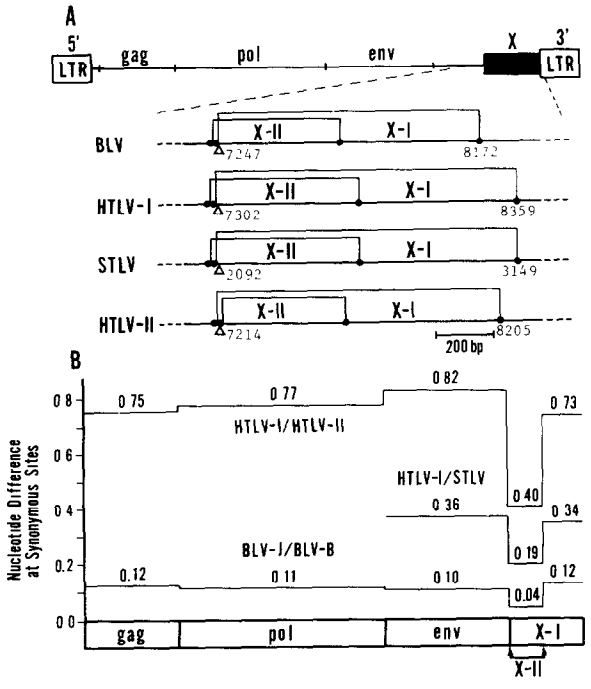

Fig.1. Structures of the X genes of the HTLV/BLV family (A) and the protein-coding capability of their short open reading frames X-II (B). Structures are from [4], BLV; [3], HTLV-I; [5], STLV; [6] HTLV-II. In A, (●) termination codon; (△) splice acceptor site. (B) Calculation according to Miyata and Yasunaga [16]. See text for detailed explanation.

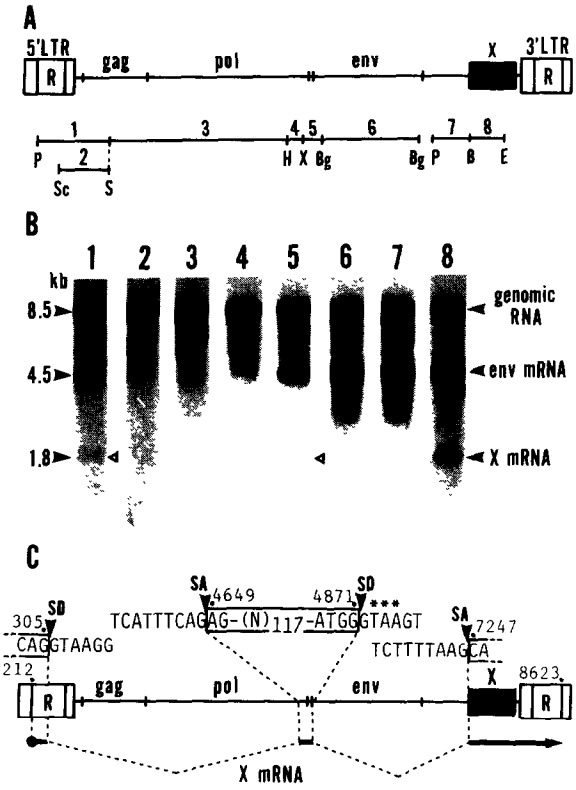

Fig.2. Identification and structure of the BLV X mRNA. (A) Location of probes used for Northern blotting analyses. P, *Pvu*II; Sc, *Sac*I; S, *Sal*I; H, *Hin*dIII; X, *Xba*I; Bg, *Bgl*II; B, *Bam*HI; E, *Eco*RI. (B) Northern blotting analyses of the FLK-BLV poly(A)$^+$ RNA (lane numbers coincide with probe numbers in A). (△) Weak hybridization with the X mRNA. Genomic RNA and *env* RNA are also indicated. (C) Spliced structure of the X mRNA. SD, splice donor; SA, splice acceptor. TAA with stars is a termination codon of the *pol* gene [4]. See text for details.

substitution of synonymous sites is constant among different genes [20]. In contrast, in any pair of viruses the nucleotide difference at synonymous sites of the X-II region is extremely reduced as compared with those of the other genes or regions. Since extreme reduction in synonymous substitutions is typical of overlapping genes of certain prokaryotic and eukaryotic viruses [20,21] and since the X-II region makes an open reading frame commonly in all members of the HTLV/BLV family (fig.1A), it is most likely that the X-II region, or the X-II frame, is an overlapping gene encoding an unknown protein.

3.2. *Structure of X mRNA*

We previously predicted the existence of a spliced X mRNA of BLV [17]. To verify this, we performed Northern blotting analyses of the poly(A)$^+$ RNA isolated from BLV-producing FLK cells, using various portions of the BLV proviral DNA as probes (fig.2A). Fig.2B shows that, of the 3 BLV-specific mRNAs detected, a subgenomic 1.8 kb mRNA preferentially hybridizes with an X gene-specific probe (lane 8), indicating that it is an X mRNA. This subgenomic X mRNA also hybridizes, although very weakly, with both 5′-LTR and *pol-env* junction probes (lanes 1 and 5), which implies that small portions of these are also exons of the X mRNA. Thus, the BLV X mRNA has a dually spliced structure, Δ5′-LTR-*pol/env* junction-X. Based on our published BLV DNA sequence [4], 2 potential splice donor sites are located in the R region of the 5′-LTR and immediately upstream of the 3′-end of the *pol* gene, while 2 potential splice acceptor sites are located about 170 bp upstream of the *env* gene and at the 5′-end of the X-I frame (fig.2C); most of these splice donor/acceptor sites are consistent with our earlier predictions [4,17]. Very recently, an almost identical structure of X mRNA was accurately determined for HTLV-I [22] and HTLV-II [23]. HTLVs and BLV therefore have a common and novel splicing mechanism of RNA for expression of their X genes.

3.3. *Existence of both X-I and X-II frames on a single spliced X mRNA*

Fig.3 shows the deduced structure of the spliced

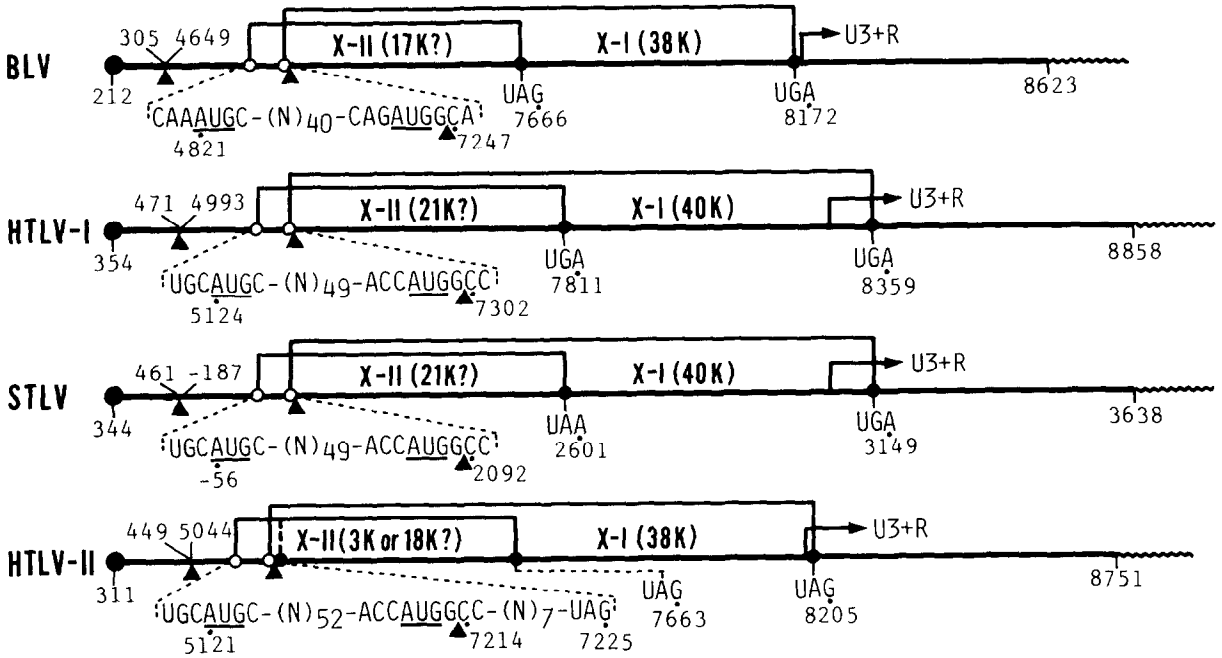

Fig.3. Fine structures of the X mRNAs of the HTLV/BLV family. In each X mRNA, the first underlined AUG codon is an X-II initiation codon, while the second one is an X-I initiation codon. (●) Cap, (▲) splice site, (○) initiation codon, (●) termination codon, (——) poly(A) tail. See text for details.

BLV X mRNA, along with those of the other viruses [22,23]. The BLV X-I frame, which originally had no initiator ATG codon in the proviral genome (fig.1A), now displays a potentially functional form on the spliced mRNA: an AUG triplet, initially located just upstream of the second splice donor site or the 3′-end of the *pol* gene but in a different reading frame from those of both *pol* and *env* genes (fig.2C), is jointed in frame to the 5′-end of the original X-I frame (fig.3). The resulting X-I frame, supplied with only one initiator methionine codon, can thus encode a 309-amino-acid residue protein, which has recently been identified as p38 (X_{BL}) [10]. As shown in fig.3, X mRNAs of HTLV-I, STLV and HTLV-II have similarly organized X-I frames, each being supplied with an initiator AUG codon, which is also located just upstream of the second splice site; it is notable that these AUG codons, unlike that of BLV, are themselves *env* start codons [22,23].

A closer examination of the nucleotide sequence of the BLV X mRNA reveals another AUG triplet which is located 44 nucleotides upstream of the X-I AUG start codon (fig.3). Interestingly, this is an *env* start codon [4,18], is a 5′-proximal AUG triplet, and is in frame with the X-II frame. Thus, the BLV X-II frame also has a potentially functional form on the spliced mRNA, with the first 51 *env*-coding nucleotides at its 5′-end (fig.3); it should be noted that this 51-nucleotide sequence encodes an NH_2-terminal half of an *env* signal peptide [4]. Inspection of the nucleotide sequences of HTLV-I and STLV X mRNAs also reveals an AUG triplet which is located 53 nucleotides upstream of the respective X-I AUG start codons (fig.3). To our surprise these AUG triplets are also 5′-proximal AUG codons and are in frame with the respective X-II frames, although unlike the putative BLV X-II start codon, they are within, but in a different reading frame, the *pol* gene [3,5]. Thus, in all these viruses, both X-I and X-II frames are on a single spliced X mRNA in an apparently functional form.

The HTLV-II X mRNA also contains an AUG triplet, 56 nucleotides upstream of the X-I AUG start codon; again, this is a 5′-proximal AUG codon and is in frame with the original X-II frame (fig.3). In this case, however, the resulting X-II frame is disrupted by an in-frame UAG terminator codon located at the 5′-end of the original X-II frame. The reason why only the HTLV-II X-II frame is prematurely terminated in spite of its strong protein-coding capability (fig.1) is not known at present.

3.4. *Possible dual translation of the X mRNA*

A plasmid construct containing a cDNA clone of the HTLV-I X mRNA has recently been shown to direct synthesis of the major X-I product $p40^{x}$ in transfected cells [22], clearly showing that the X-I frame is translatable from the X mRNA in vivo. In contrast, no direct experimental data on the translatability of the X-II frame from the same X mRNA are available. Recently, however, sera from BLV-infected cattle have been shown to react with an in vitro translated X-II-related protein, suggesting the existence of an X-II product in vivo [24]. Furthermore, using a peptide antiserum against the BLV X-II frame, a polypeptide with a size similar to that of the predicted BLV X-II product (17 kDa, fig.3) has been preliminarily observed in BLV-producing cell lines (N. Rice, personal communication). We believe that these putative X-II products are those which are translated from the BLV X mRNA since no other BLV mRNA which could encode them has so far been observed. Thus, these observations and the common structure of the X mRNA in all members of the HTLV/BLV family (figs 1–3) strongly suggest that in all these viruses both the X-I and X-II frames are translatable from the single X mRNA in vivo.

If the unusual dual translation of the X mRNA truly occurs, then how can ribosomes recognize both the X-I and X-II AUG start codons on the same single X mRNA? For the selection by eukaryotic ribosomes of an initiator AUG codon, both proximity of the AUG codon to the 5′-end of the mRNA and the sequences that flank the AUG codon are important (modified scanning model [25]). In this respect, the X-II AUG start codon of every X mRNA is a 5′-proximal AUG and is in a suboptimal sequence context, whereas the X-I AUG codon is a second AUG and is in an optimal or nearly optimal sequence context (cf. sequences that flank the X-I and X-II AUG start codons, shown in fig.3, with those in the optimal sequence context, $^{A}_{G}$XXAUGG, proposed by Kozak [25]). Thus, according to the modified scanning model, some 40 S ribosome subunits would stop and in-

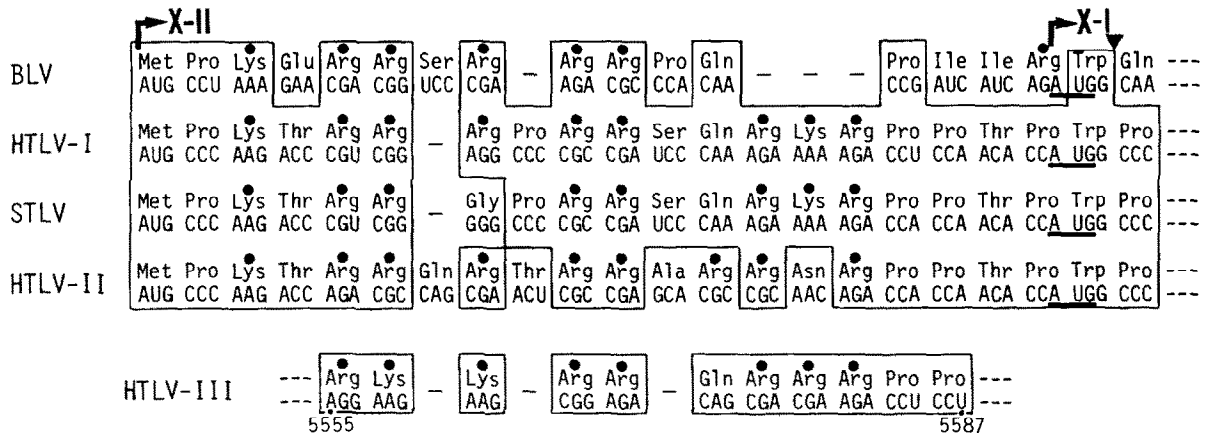

Fig.4. Comparison of the NH_2-terminal sequences of the putative X-II products of HTLV/BLV family. Sequence alignment is performed with several gaps. Highly basic amino acid residues are given closed circles and homologous residues are boxed. Overlapping X-I frames start at underlined AUGs (▼, splice site). A highly basic amino acid sequence of the HTLV-III *trans*-acting protein [31,32] is also aligned.

itiate at the 5′-proximal X-II AUG codon, but some would bypass that site because of its suboptimal sequence context and initiate at the downstream X-I AUG codon which is in an optimal sequence context, thereby resulting in the dual translation of the single X mRNA. There are indeed a few other examples of eukaryotic viral and cellular mRNAs that are translated from more than one AUG codon [26–28]; among these, the most notable example is a spliced mRNA of an adenovirus E1b gene, which also encodes 2 distinct proteins in different reading frames [28].

3.5. *Nature of the putative X-II product: a nuclear regulatory protein?*

The X-II frames are capable of encoding proteins of 17–21 kDa (for BLV, HTLV-I, and STLV) or only 3 kDa (for HTLV-II, assuming premature termination) (fig.3). These putative X-II products apparently do not show amino acid sequence homology between BLV and the other viruses (not shown). However, they commonly share a very high proline residue content (20–23% of total residues). More interestingly, their NH_2-terminal 20-residue sequences, which are derived from second exons (fig.3), are usually highly basic with many arginine residues and show even sequence homology between BLV and the other viruses (fig.4), despite the fact that they are read from different reading frames of the respective proviral genomes (see section 3.3). These findings suggest that the highly basic NH_2-terminal sequences are functional domains of the respective X-II products. In certain cellular and viral proteins, such a highly basic amino acid sequence has DNA-binding activity [29] and is a nuclear transport signal [30], raising the possibility that the putative X-II product is a nuclear protein.

Very recently, the acquired immune deficiency syndrome (AIDS)-associated retrovirus (HTLV-III) has been shown to encode a nuclear *trans*-acting protein which also carries a number of proline residues and a striking cluster of arginine-lysine residues [31,32]. This highly basic amino acid cluster is also read from the *pol-env* junction of the proviral genome [31,32] and appears to have significant sequence homology with the basic NH_2-terminal sequences of the putative X-II products (fig.4), suggesting analogous function between the HTLV-III *trans*-acting protein and the HTLV/BLV X-II products. We postulate that the X-II product is a nuclear regulatory protein which is involved in viral replication and/or cellular transformation, possibly interacting with the known X-I product.

ACKNOWLEDGEMENTS

We thank Mrs J. Tsuzuku-Kawamura for manuscript preparation. This work was supported partly by Research Grant of the Princess Takamatsu Cancer Research Fund and partly by a grant from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Gallo, R.C., Essex, M.E. and Gross, L. (1984) Human T-cell Leukemia/Lymphoma Virus, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[2] Burny, A., Bruck, C., Chantrenne, H., Cleuter, Y., Dekegel, D., Ghysdael, J., Kettmann, R., Leclercq, M., Leunen, J., Mammerickx, M. and Portetelle, D. (1980) in: Viral Oncology (Klein, G. ed.) pp.231–289, Raven, New York.

[3] Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1983) Proc. Natl. Acad. Sci. USA 80, 3618–3622.

[4] Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985) Proc. Natl. Acad. Sci. USA 82, 677–681.

[5] Watanabe, T., Seiki, M., Tsujimoto, H., Miyoshi, I., Hayami, M. and Yoshida, M. (1985) Virology 144, 59–65.

[6] Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M. and Sugimura, T. (1985) Proc. Natl. Acad. Sci. USA 82, 3101–3105.

[7] Kiyokawa, T., Seiki, M., Imagawa, K., Shimizu, F. and Yoshida, M. (1984) Gann 75, 747–751.

[8] Miwa, M., Shimotohno, K., Hoshino, H., Fujino, M. and Sugimura, T. (1984) Gann 75, 752–755.

[9] Lee, T.H., Coligan, J.E., Sodroski, J.G., Haseltine, W.A., Salahuddin, S.Z., Wong-Staal, F., Gallo, R.C. and Essex, M. (1984) Science 226, 57–61.

[10] Sagata, N., Tsuzuku-Kawamura, J., Nagayoshi-Aida, M., Shimizu, F., Imagawa, K. and Ikawa, Y. (1985) Proc. Natl. Acad. Sci. USA, in press.

[11] Sodroski, J.G., Rosen, C.A. and Haseltine, W.A. (1984) Science 225, 381–385.

[12] Rosen, C.A., Sodroski, J.G., Kettmann, R., Burny, A. and Haseltine, W.A. (1985) Science 227, 320–322.

[13] Fujisawa, J., Seiki, M., Kiyokawa, T. and Yoshida, M. (1985) Proc. Natl. Acad. Sci. USA 82, 2277–2281.

[14] Van der Maaten, M.J. and Miller, J.M. (1976) Bibl. Haematol. 43, 360–362.

[15] Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201–5205.

[16] Miyata, T. and Yasunaga, T. (1980) J. Mol. Evol. 16, 23–36.

[17] Sagata, N., Yasunaga, T., Ohishi, K., Tsuzuku-Kawamura, J., Onuma, M. and Ikawa, Y. (1984) EMBO J. 3, 3231–3237.

[18] Rice, N.R., Stephens, R.M., Couez, D., Deschamps, J., Kettmann, R., Burny, A. and Gilden, R.V. (1984) Virology 138, 82–93.

[19] Rice, N.R., Stephens, R.M., Burny, A. and Gilden, R.V. (1985) Virology 142, 357–377.

[20] Miyata, T., Yasunaga, T. and Nishida, T. (1980) Proc. Natl. Acad. Sci. USA 77, 7328–7332.

[21] Yasunaga, T. and Miyata, T. (1982) J. Mol. Evol. 19, 72–79.

[22] Seiki, M., Hikikoshi, A., Taniguchi, T. and Yoshida, M. (1985) Science 228, 1532–1534.

[23] Wachsman, W., Golde, D.W., Temple, P.A., Orr, E.C., Clark, S.C. and Chen, I.S.Y. (1985) Science 228, 1534–1537.

[24] Yoshinaka, Y. and Oroszlan, S. (1985) Biochem. Biophys. Res. Commun., in press.

[25] Kozak, M. (1983) Microbiol. Rev. 47, 1–45.

[26] Kozak, M. (1981) Curr. Top. Microbiol. Immunol. 93, 81–123.

[27] Mardon, G. and Varmus, H.E. (1983) Cell 32, 871–879.

[28] Bos, J.L., Polder, L.J., Bernards, R., Schrier, P.I., Van den Elsen, P.J., Van der Eb, A. and Van Ormondt, H. (1981) Cell 27, 121–131.

[29] Saenger, W. (1983) Principles of Nucleic Acid Structure, pp.385–431, Springer, New York.

[30] Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984) Cell 39, 499–509.

[31] Arya, S.K., Guo, C., Josephs, S.F. and Wong-Staal, F. (1985) Science 229, 69–73.

[32] Sodroski, J., Patara, R., Rosen, C., Wong-Staal, F. and Haseltine, W. (1985) Science 229, 74–77.