ELSEVIER

FEBS Letters 338 (1994) 118–121

FEBS 13569

FEBS LETTERS

Hypothesis

The origin of human immunodeficiency virus type-1 *rev* gene

An evolutionary hypothesis

Satoshi Kubota[a,b], Stephen Oroszlan[a], Masakazu Hatanaka[b,*]

[a]*Laboratory of Molecular Virology and Carcinogenesis, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, USA*

[b]*Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606, Japan*

Received 18 November 1993

Abstract

The Rev protein of human immunodeficiency virus type-1 is an RNA-binding posttranscriptional transregulator encoded by an accessory gene that is distinct from retroviral oncogenes and whose origin is unclear. We hypothesize that the *rev* gene was generated by duplication of a viral RNA segment having a secondary-structure that evolved into the Rev-responsive element (RRE). This hypothesis is based on the following findings. First, accumulated data on functional mapping of Rev, Tat, and the transmembrane protein of Env suggested that the major coding exon of *rev* should have been inserted into the transmembrane region of *env* during the course of its evolution. Experiments with equine infectious anemia virus, another complex retrovirus, also indicate that a large portion of *rev* is located within the dispensable transmembrane region of *env*. Second, base usage analysis suggests the same origin for *rev* and RRE. Our hypothesis may provide a new insight into the evolutionary aspect of RNA-binding transactivators.

Key words: HIV-1; Rev protein; RNA binding protein; lentivirida

Human immunodeficiency virus type-1 (HIV-1) is shown as the primary causative agent of acquired immunodeficiency syndrome (AIDS) [1]. This virus is one of the complex retroviruses that possess several accessory genes in addition to the common structural genes. The *Rev* protein is an RNA-binding transregulator that has been shown to induce cytoplasmic accumulation of unspliced viral RNA via a structural RNA target termed the Rev-responsive element (RRE) [2,3]. The *rev* gene consists of three exon: of the three, exon 2 encoding the N-terminal 26 amino acid residues and exon 3 encoding the C-terminal 90 residues [4]. Exon 3 of *rev* is overlapped with the coding region of the envelope transmembrane glycoprotein (gp41). The minor exon 3 of *tat* resides within this area as well [5]. Indeed, a certain region of the *env* gene is translated into three different proteins using all coding frames. However, frames that are indispensable functionally do not overlap with each other, as shown in Fig. 1. Although *rev* seems to be completely overlaid by the gp41 coding region, its functional domain is located within a region of gp41 that is not essential for viral replication [5,6]. In contrast, the C-terminal amino acid residues 91–116 of Rev are dispensable for its function and they overlap an indispensable region of gp41 (amino acid residues 793–847 in Env; see Fig. 1) [4]. Additionally, the minor exon 3 of *tat* overlapped by *rev* is wholly dispensable for its TAR-dependent transactivating ability [7,8]. Thus, this region, to be termed the 'Rev-dominant unit', may have evolved primarily for Rev function. It should be noted that the gp41 cytoplasmic domain including the Rev-dominant unit is extraordinarily long compared to that of simpler retroviruses [9], implying that such a long cytoplasmic tail may not be essential in the retroviral life cycle. This theory is also supported by a recent report showing different evolutionary pathways for *env* and *rev* [10]. To examine whether the Rev-dominant unit is present in other complex retroviruses, we investigated the genomic organization of EIAV (Fig. 2).

EIAV possesses a gene encoding an extraordinarily long envelope transmembrane glycoprotein (gp45) in addition to the *rev* gene [9,11]. The major exon of *rev* is located in the gene for the transmembrane glycoprotein. For viral maturation, gp45 is processed into the N-terminal mature transmembrane protein gp32 or gp35 and C-terminal unglycosylated protein p20 [12], cleaving coincidentally near the start site of the major *rev* exon (Fig. 2A). Although frequent mutations terminate the translation of gp45 upstream of the cleavage site, without major

*Corresponding author. Fax: (81) (75) 761-5626.

SSDI 0014-5793(93)E1486-6

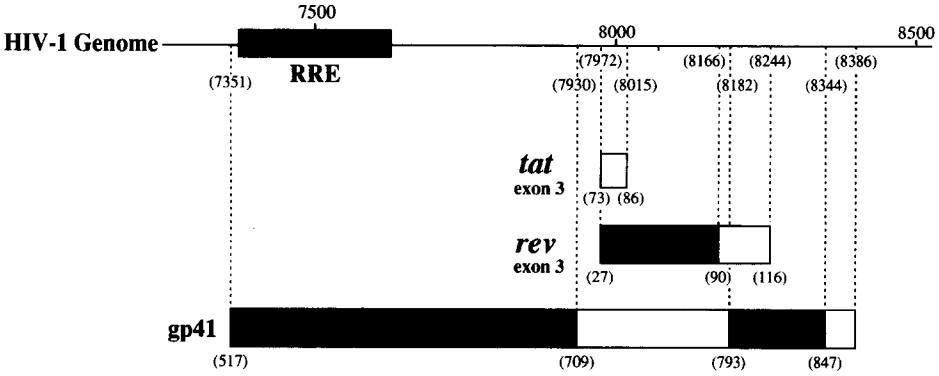

Fig. 1. Aligned mapping of functional domain structures of three genes encoded in different frames of the HIV-1 *env* region. The solid line represents a part of HIV-1 genomic RNA. Numbers above the line indicate nucleotide positions for the LAV-1 isolate (GenBank entry: HIVBRUCG). Boxes represent protein-coding regions: shaded areas indicate regions that are functionally essential and open areas indicate dispensable regions. Numbers within parentheses denote residues counted from the initiator methionine of each protein. The location of the RRE in the genome is also displayed.

alteration of infectivity, the premature termination of *rev* has never been observed in proviruses integrated in cells producing infectious EIAV [12] (Fig. 2B). Since the C-terminal domain of Env (p20) is not essential for virus replication, it is likely that the 3′ region of the *env* gene (immediately downstream of gp32) was evolved into the present form by the insertion of the ancestor of *rev* as observed for HIV-1.

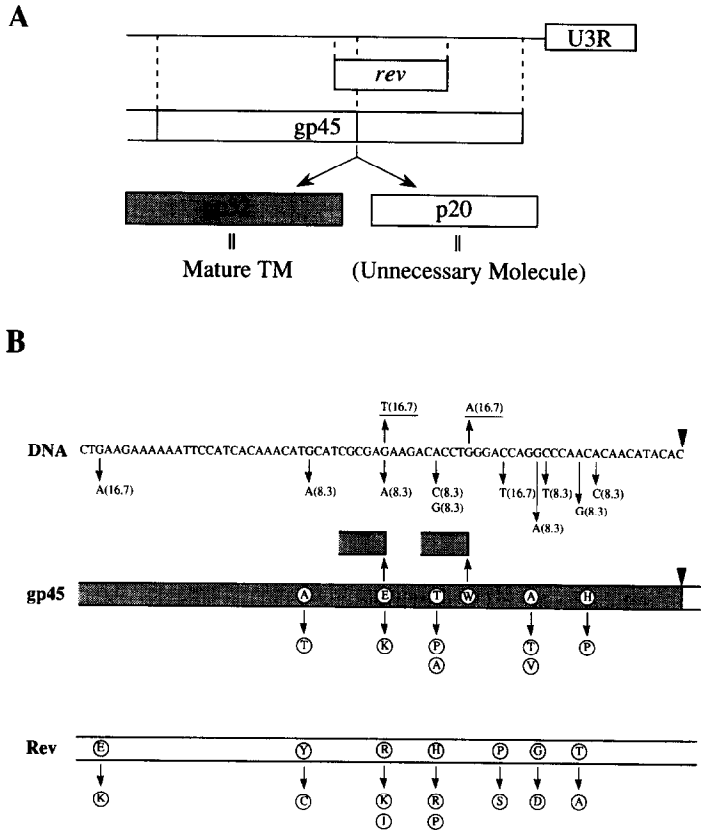

Fig. 2. 'Cutoff' of the C-terminal region of EIAV gp45 overlapping Rev (A) 'Cutoff' at the protein level. Boxes represent proteins translated from corresponding regions of the EIAV genome. Arrows denote the cleavage of gp45. (B) 'Cutoff' at the DNA level. A serial prototypic DNA sequence upstream of the gp32–p20 cleavage site is serially shown. Solid arrows above and under the DNA sequence, indicate variations of DNA sequences obtained by polymerase chain reaction (PCR) with genomic DNA from cells producing infectious EIAV. Numbers within parentheses indicate the percentages of the incidence of each mutation. The mutations that cause premature termination of gp45 are underlined. Premature translational terminations of gp45 are displayed by small boxes and arrows. Amino acid changes in gp45 and Rev are shown in single letter code circled and by arrows. Arrowheads indicate the gp32–p20 cleavage site and its corresponding locus on the DNA sequence.

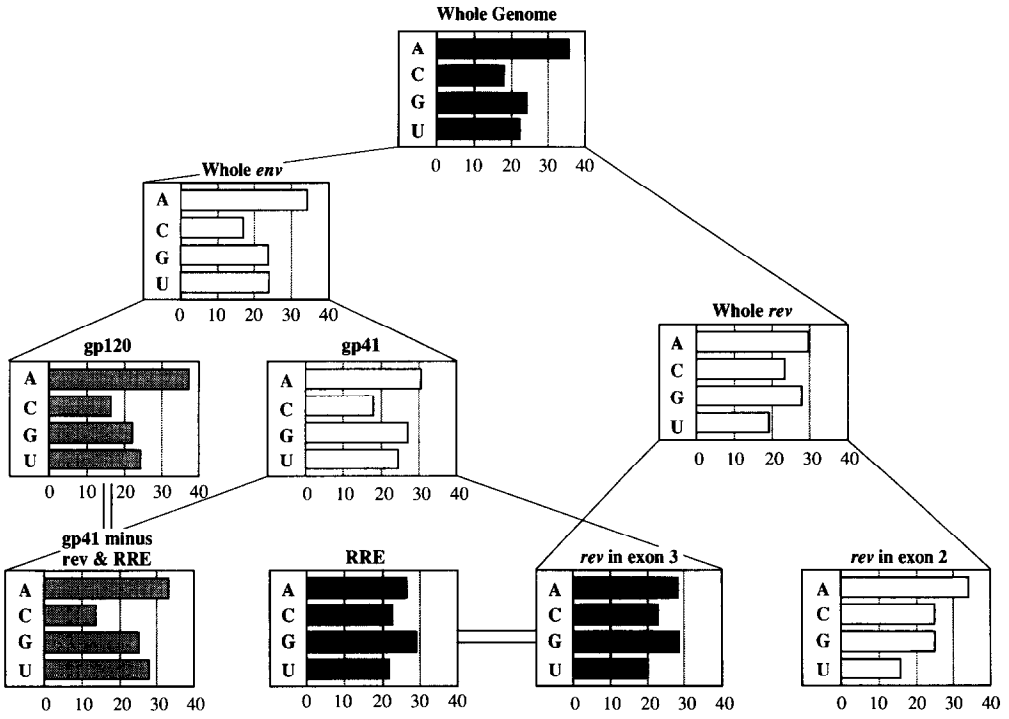

Fig. 3. Base-usage analysis of an HIV-1 isolate, LAV-1. The nucleotide sequence of LAV-1 from GenBank was analyzed by DNASIS (HITACHI SK). All histograms except those for the whole genome and RRE were derived from information about the coding regions. Light shadings are for *env*-patterns and dark shadings are for RRE and *rev*-patterns. Solid black was used only for the pattern of the whole genome. Numbers at the bottom of each histogram indicate the percentages of bases.

Considering the origin of the *rev* sequence, we can speculate that its ancestor might have been derived from a cellular transcript by recombination during reverse transcription, comparable to retroviral oncogenes. However, the structure of the *rev* gene is genomic and requires splicing for its expression. Furthermore, the *rev* homolog has not been found in host cells. Another source of the *rev* gene may be the viral genome itself. A fragment of viral genome might have been inserted into the *env* gene at a certain stage of evolution. We propose that the ancestor of *rev* was derived from a secondary structure-bearing viral RNA fragment that evolved further to become the RRE.

It should be noted that the base-usage patterns of RRE and the major *rev* coding region in an HIV-1 strain (LAV-1) are strikingly similar, yet quite different from those of the other portion of *env* or the whole genome (Fig. 3). These data may reflect the footprint of their common origin, albeit with less sequence homology between *rev* and RRE. Similarly the resemblance of base-usage patterns of gp120 and gp41 nucleotide sequences, excluding the *rev* and RRE regions, suggests their close relationship as *env*.

Complex retroviruses might have evolved by the formation of structured regions on genomic RNAs from prototypic primitive retroviruses. If so, then proteins that bind to such secondary structure regions of RNA might have evolved to favor the survival of these retroviruses. Since host cells have many RNA-binding proteins, some viruses could utilize such cellular proteins as binding molecules. Conversely, viruses could evolve to become complex retroviruses by encoding their own RNA-binding proteins. We assume that the structured region is duplicated during replication which could result in the evolution of a gene encoding a binding protein for the structured region and, ultimately, the RNA-binding posttranscriptional transregulator. In the case of both HIV-1 and EIAV, such a gene, the *rev* gene, is located in *env* as a spacer sequence.

Acknowledgements: We thank Takashi Gojobori and Rika Furuta for helpful discussions, Masumi Fujiki and Cheri Rhoderick for preparation of the manuscript and Ann Arthur for her editorial assistance.

References

[1] Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rouzenbaum, W. and Montagnier, L. (1983) Science 220, 868–871.
[2] Malim, M.H., Hauber, J., Le, S.-Y., Maizel, J.V. and Cullen, B.R. (1989) Nature 338, 254–257.
[3] Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. and Singh, M. (1991) Proc. Natl. Acad. Sci. USA 88, 7366–7370.
[4] Malim, M.H., Böhnlein, S., Hauber, J. and Cullen, B.R. (1989) Cell 58, 205–214.
[5] Gabuzda, D.H., Lever, A., Terwilliger, E. and Sodroski, J. (1992) J. Virol. 66, 3306–3315.

[6] Lee, S.-J., Hu, W., Fisher, A.G., Looney, D.J., Kao, V.F., Mitsuya, H., Ratner, L. and Wong-Staal, F. (1989) AIDS Res. Hum. Retroviruses 5, 441–449.
[7] Kubota, S., Endo, S., Maki, M. and Hatanaka, M. (1989) Virus Genes 2, 113–118.
[8] Endo, S., Kubota, S., Siomi, H., Adachi, A., Oroszlan, S., Maki, M. and Hatanaka, M. (1989) Virus Genes 3, 99–110.
[9] Gallaher, W.R., Ball, J.M., Garry, R.F., Griffin, M.C. and Montelaro, R.C. (1989) AIDS Res. Hum. Retroviruses 5, 431–440.
[10] Martins, L.P., Chenciner, N., Åsjö, B., Meyerhans, A. and Wain-Hobson, S. (1991) J. Virol. 65, 4502–4507.
[11] Stephens, R.M., Derse, D. and Rice, N.R. (1990) J. Virol. 64, 3716–3725.
[12] Rice, N.R., Henderson, L.E., Sowder, R.C., Copeland, T.D., Oroszlan, S. and Edwards, J.F. (1990) J. Virol. 64, 3770–3778.