Volume 287, number 1,2, 1–4 FEBS 09979 August 1991

ADONIS 001457939100689N

Review Letter

How do viral reverse transcriptases recognize their RNA genome?

Jean-Michel Mesnard and Geneviève Lebeurier

Institut de Biologie Moléculaire des Plantes du C.N.R.S., Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cédex, France

Received 3 May 1991

Reverse transcription is not solely a retroviral mechanism. Hepadnaviruses and caulimoviruses have RNA intermediates that are reverse transcribed into DNA. Moreover non-viral retroelements, retrotransposons, use reverse transcription in their transposition. All these retroelements encode reverse transcriptase but each group developed their own expression modes capable of assuring a specific and efficient replication of their genomes.

Retrovirus; Pararetrovirus; Reverse transcription; Encapsidation; Viroplasm

1. INTRODUCTION

The common mode of information transfer in biological systems is DNA to RNA to protein. Twenty years ago it was established that retroviruses used information transfer from RNA to DNA catalyzed by an enzyme called reverse transcriptase [1,2]. Several years later it was found that reverse transcriptase was distributed in bacteria, yeast, insects, vertebrates and plants. An archiving role for reverse transcriptase has been suggested in the putative ancient transition from an RNA world to a DNA world.

Other viruses, hepadna- and caulimoviruses, also use reverse transcriptase during the replication of their genomes. Based on this fact, these two virus families have been regrouped under the name of pararetroviruses. Retroviruses and pararetroviruses replicate from a genome-length RNA transcript which is the template for reverse transcription. This type of replication necessitates a strict recognition between the reverse transcriptase and the viral RNA. In fact, in the absence of a specific enzyme–template interaction, the reverse transcriptase could interact with the cellular RNAs causing, on the one hand, a deregulation in the translation of the mRNA of the host cell and, on the other hand, a drop in the rate of replication of the viral genome. How does the reverse transcriptase recognize the viral RNA? Although similar in their modes of replication and even in their genomic organizations, the retro-, hepadna- and caulimoviruses have developed a reverse transcriptase-viral RNA recognition strategy which is particular to each group.

Abbreviations: CaMV, cauliflower mosaic virus; CoYMV, commelina yellow mottle virus; HBV, hepatitis B virus; HIV-1, human immunodeficiency virus 1; HTLV-II, human T-cell leukemia virus type II; MMTV, mouse mammary tumor virus; MoMLV, Moloney murine leukemia virus; MPMV, Mazon-Pfizer monkey virus; ORF, open reading frame; RSV, Rous sarcoma virus

Correspondence address: G. Lebeurier, Institut de Biologie Moléculaire des Plantes du C.N.R.S., Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cédex, France

2. IN RETROVIRUSES, REVERSE TRANSCRIPTASE IS GUIDED BY THE VIRAL CAPSID PROTEIN

The retrovirus replication mechanism involves the synthesis of a terminally redundant genomic RNA that is packaged into a viral nucleocapsid and reverse transcribed into a DNA genome. However, the onset of reverse transcriptase is delayed until the mature virus re-enters a host cell. This explains why retroviruses contain an RNA genome.

The retroviral reverse transcriptase is encoded by the ORF pol, located downstream from the ORF gag, which codes for the virus capsid proteins. The expression of pol necessitates the synthesis of a gag-pol polyprotein, later matured in capsid proteins, in protease, in reverse transcriptase and in integrase. The maturation is catalyzed by an aspartic protease coded either by ORF gag for RSV [3], or by ORF pol for MoMLV [4] and HIV-1 [5,6], or by a particular ORF, pro for HTLV-II [7], MPMV [8] and MMTV [9]. Retroviruses use two strategies to generate a fusion protein of gag-pol: either by read-through of a termination codon as in the case of the MoMLV [10], or by

ribosomal frameshifting in the −1 direction as for the RSV [11] and HIV-1 [12]. The gag-pol polyprotein is produced about 5 percent as efficiently as the translation product of gag alone [13]. The synthesis of such a fusion protein presents two advantages. Firstly, structural proteins (gag) are synthesized in a greater quantity compared to enzymatic proteins (pol): for 5 molecules of reverse transcriptase, 100 molecules of capsid proteins are produced. Secondly, reverse transcriptase can be directed to viral capsids since gag codes for a zinc finger-like protein which binds the genomic RNA in the virion [14]. For the foamy retroviruses, gag protein does not contain the cysteine motif of the nucleic acid-binding proteins; instead gag encodes a strongly basic protein which likely binds the genomic RNA [15]. Thanks to the gag-pol fusion protein, reverse transcriptase is specifically attached to the viral RNA matrix and not to the RNAs of the host cell. Thus, in retroviruses, reverse transcriptase is strongly dependent on capsid proteins in translation as well as in replication.

3. TEMPLATE-REVERSE TRANSCRIPTASE RECOGNITION IN RETROTRANSPOSONS, A MECHANISM SIMILAR TO THAT OF RETROVIRUSES

Ty elements are a family of transposable elements which are dispersed throughout the yeast genome. They transpose by a retroviral-like mechanism within a virus-like particle synthesized from the product of the ORF TYA, analog of the retroviral ORF gag [16,17]. The ORF TYB, located downstream from the ORF TYA, encodes a polymerase which includes sequence homologies to retroviral reverse transcriptase, integrase and protease. TYB is synthesized in the form of a TYA:TYB fusion protein equivalent to the gag-pol protein of retroviruses. For Ty1 the expression of TYB is very efficient resulting in 20% read-through into TYB from TYA [18], thanks to a ribosomal frameshifting in the +1 direction [18,19]. Due to these analogies with the retroviruses, the Ty elements have been regrouped in the retrotransposon family.

Other retrotransposons have been characterized by the genome of various organisms such as insects, plants and mammals. All these elements transpose by reverse transcription and express their reverse transcriptase fused to the capsid protein. The *Drosophila* transposable element copia [20,21], the *Arabidopsis thaliana* retroelement Ta1 [22] and the *Nicotiana tabacum* retroelement Tnt1 [23] possess in fact only a single ORF coding for a polyprotein having sequence homologies to retroviral zinc finger-like protein, reverse transcriptase, integrase and protease. Thus, as the retroviruses, retrotransposons synthesize their reverse transcriptase combined with a capsid protein in order to specifically direct the enzyme to the retroelement RNA.

4. IN HEPADNAVIRUSES, REVERSE TRANSCRIPTASE IS EXPRESSED INDEPENDENTLY OF THE CAPSID PROTEIN

Although they contain a DNA genome, hepadnaviruses resemble retroviruses in that they replicate via reverse transcription of an RNA pregenome. In the hepatitis B virus, or HBV, a typical member of the hepadnavirus group, reverse transcriptase is coded by the ORF P located downstream from the ORF C coding for the viral capsid. This genomic organization, common to all hepadnaviruses [24–26] is similar to that of the retroviruses and the retrotransposons. However, the analogies stop here as the hepadnaviruses have neither zinc finger-like protein, nor integrase nor aspartic protease [27,28]. Moreover, it has been demonstrated that the synthesis and the encapsidation of hepadnavirus reverse transcriptase do not require formation of capsid-polymerase fusion proteins [29,30]. Following these studies, it was suggested that translation of the hepadnavirus reverse transcriptase was regulated by a cap-independent mechanism, the translation being initiated at the ATG of ORF P [31]. A mechanism different from that of the retroviruses takes place in the reverse transcriptase-RNA viral recognition of hepadnaviruses.

The hepadnaviral genome contains a genome-linked protein attached to the 5′-end of its minus strand, suggesting that reverse transcription of the hepadnaviral RNA is primed by a protein and not by a tRNA molecule as is the case for the retroviruses. However, this primer protein corresponds to the N-terminal part of the polymerase encoded by the ORF P [32,33] and is therefore fused to the reverse transcriptase. Thanks to such a protein, the RNA of hepadnaviruses is recognized by its reverse transcriptase. This mechanism is possible in hepadnaviruses as the capsid protein does not control the replication stage as strictly as in retroviruses although the reaction involving reverse transcription of hepadnaviral RNA is segregated from the cytosol within a subviral particle. In fact, the polymerase is not only required for reverse transcription but also for RNA packaging, the encapsidation function of the enzyme being separated from its reverse transcriptase activity [34,35].

Unlike the retroviruses and the retrotransposons, the reverse transcriptase of the hepadnaviruses is expressed independently of the capsid protein due to the synthesis of a polymerase capable of interacting specifically with the viral RNA, of controlling the RNA packaging and of carrying out the reverse transcription stage.

5. THE CAULIMOVIRUSES, A SPECIAL CASE AMONG THE RETROELEMENTS

The most studied virus among the caulimoviruses is the CaMV [36]. The reverse transcriptase of CaMV is

coded by the ORF V located downstream from ORF IV coding for the major capsid protein. As with the retroviruses, the capsid protein possesses a zinc finger-like domain capable of interacting with the viral RNA [37]. Moreover, ORF V also codes for aspartic protease located in the N-terminal position of the reverse transcriptase [38], as in the MoMLV and the HIV-1. It thus seems possible that CaMV expresses its reverse transcriptase in the form of a precursor of the gag-pol type which is later matured by proteasic activity. However, an ORF IV-V fusion protein could not be detected in infected plants [39,40]. Furthermore, translation experiments in vitro [41] or in a yeast system [42] did not make it possible to characterize such a fusion protein. Ultimately, thanks to different mutations located in the CaMV ORF IV/V overlapping region, Schultze et al. [43] showed that the reverse transcriptase is translated independently from the capsid ORF in much the same way as in the hepadnaviruses.

The case of the caulimoviruses remains different nevertheless from the hepadnaviruses as the CaMV reverse transcription is not primed by a protein but by a $tRNA^{met}$. The CaMV thus developed a mechanism different from the hepadnaviruses to direct the reverse transcriptase toward the viral RNA. In fact, it has a particular ORF, ORF VI, coding for the major protein of the viroplasms. Characteristic of the caulimoviruses, viroplasms are cytoplasmic inclusions where the replication and the encapsidation of viral genome are localized. Thus the reverse transcription of CaMV RNA is segregated from the cellular cytosol within the viroplasms. This compartmentalization would facilitate a specific recognition between the reverse transcriptase and its RNA template.

The above hypothesis has been confirmed by the characterization of another plant pararetrovirus which does not belong to the caulimovirus group, the CoYMV [44]. The genome of this pararetrovirus is a double-stranded DNA possessing three ORFs, of which one codes for a polymerase, having sequence homologies to CaMV capsid protein and protease/reverse transcriptase polypeptide. The CoYMV would seem similar to retroviruses because the capsid protein is fused to the reverse transcriptase, the polypeptide being matured by an aspartic protease. But contrary to CaMV, the CoYMV does not have an ORF coding for a viroplasmic or equivalent protein. In the absence of such a structure responsible for a cellular compartmentalization, the CoYMV would therefore use a replication-encapsidation mechanism similar to that of the retroviruses.

6. CONCLUSION

Recently, Xiong and Eickbush [45] studied the origin and the evolution of retroelements based upon their reverse transcriptase sequences. For the authors, the retroviruses may represent retrotransposable elements which have acquired an envelope gene making it possible for them to leave the cell. On the other hand, the origins of the hepadnaviruses and the caulimoviruses were more difficult to explain. The authors suggested that the pol gene was acquired by pre-existing viruses in the pararetroviruses. This fragment would also have included the protease domain in the case of the caulimoviruses.

In such a hypothesis, it is understandable that pararetroviruses express their reverse transcriptase independently of the capsid. In the absence of a fusion protein, the hepadnaviruses and the caulimoviruses would have developed their own expression modes capable of assuring a specific and efficient replication of their genomes: a primer protein for the hepadnaviruses and a cellular compartmentalization for the caulimoviruses due to their viroplasmic structures.

Acknowledgement: The advice, help and criticism of Professor L. Hirth is gratefully acknowledged.

REFERENCES

[1] Baltimore, D. (1970) Nature 226, 1209-1211.
[2] Temin, H.M. and Mizutami, S. (1970) Nature 226, 1211-1213.
[3] Schwartz, D.E., Tizard, R. and Gilbert, W. (1983) Cell 32, 853-869.
[4] Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981) Nature 293, 543-548.
[5] Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal, F. (1985) Nature 313, 277-284.
[6] Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. and Alizon, M. (1985) Cell 40, 9-17.
[7] Shimothono, K., Takahashi, Y., Shimizu, N., Gojobori, T., Goldo, D.W., Chen, I.S.Y., Miwa, M. and Sugimura, T. (1985) Proc. Natl. Acad. Sci. USA 82, 3101-3105.
[8] Sonigo, P., Barker, C., Hunter, E. and Wain-Hobson, S. (1986) Cell 45, 375-385.
[9] Moore, R., Dixon, M., Smith, R., Peters, G. and Dickson, C. (1987) J. Virol. 61, 480-490.
[10] Jones, D.S., Nemoto, F., Kuchino, Y., Masuda, M., Yoshikura, H. and Nishimura, S. (1989) Nucleic Acids Res. 17, 5933-5945.
[11] Jacks, T., Madhani, H.D., Masiarz, F.R. and Varmus, H.E. (1988) Cell 55, 447-458.
[12] Wilson, W., Braddock, M., Adams, S.E., Rathjen, P.D., Kingsman, S.M. and Kingsman, A.J. (1988) Cell 55, 1159-1169.
[13] Jacks, T. and Varmus, H.E. (1985) Nature 320, 1237-1242.
[14] Gorelick, R.J., Henderson, L.E., Hanser, J.P. and Rein, A. (1988) Proc. Natl. Acad. Sci. USA 85, 8420-8424.
[15] Maurer, J., Bannert, H., Darai, G. and Flügel, M.R. (1988) J. Virol. 62, 1590-1597.
[16] Mellor, J., Fulton, S.M., Dobson, M.Y., Wilson, W., Kingsman, S.M. and Kingsman, A.J. (1985) Nature 313, 243-246.
[17] Clare, J.J. and Farabaugh, P.J. (1985) Proc. Natl. Acad. Sci. USA 82, 2829-2833.
[18] Clare, J.J., Belcourt, M. and Farabaugh, P.J. (1988) Proc. Natl. Acad. Sci. USA 85, 6816-6820.

[19] Wilson, W., Malim, M.H., Mellor, J., Kingsman, A.J. and Kingsman, S.M. (1986) Nucleic Acids Res. 14, 7001–7016.
[20] Mount, S.M. and Rubin, G.M. (1985) Mol. Cell. Biol. 5, 1630–1638.
[21] Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. and Saigo, K. (1985) Nature 315, 773–776.
[22] Voytas, D.F. and Ausubel, F.M. (1988) Nature 336, 242–244.
[23] Grandbastien, M.-A., Spielmann, A. and Caboche, M. (1989) Nature 337, 376–380.
[24] Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. and Charnay, P. (1979) Nature 281, 646–650.
[25] Galibert, F., Chen, T.N. and Mandart, E. (1982) J. Virol. 41, 51–65.
[26] Mandart, E., Kay, A. and Galibert, F. (1984) J. Virol. 49, 782–792.
[27] Jean-Jean, O., Salhi, S., Carlier, D., Elie, C., De Recondo, A.-M. and Rossignol, J.-M. (1989) J. Virol. 63, 5497–5500.
[28] Nassal, M., Galle, P.R. and Schaller, H. (1989) J. Virol. 63, 2598–2604.
[29] Schlicht, H.-J., Radziwill, G. and Schaller, H. (1989) Cell 56, 85–92.
[30] Chang, L.-J., Pryciak, P., Ganem, D. and Varmus, H.E. (1989) Nature 337, 364–368.
[31] Ou, J.-H., Bao, H., Shih, C. and Tahara, S.M. (1990) J. Virol. 64, 4578–4581.
[32] Bartenschlager, R., Junker-Niepmann, M. and Schaller, H. (1990) J. Virol. 64, 5324–5332.
[33] Bosch, V., Bartenschlager, R., Radziwill, G. and Schaller, H. (1988) Virology 166, 475–485.
[34] Hirsch, R.C., Lavine, J.E., Chang, L.-J., Varmus, H.E. and Ganem, D. (1990) Nature 344, 552–555.
[35] Bartenschlager, R. and Schaller, H. (1988) EMBO J. 7, 4185–4192.
[36] Mesnard, J.M., Lebeurier, G. and Hirth, L. (1990) in: Chromosomes: Eukaryotic, Prokaryotic, and Viral, Vol. III (Adolph, K.W., ed.), pp. 143–162, CRC Press, Boca Raton, FL.
[37] Covey, S.N. (1986) Nucleic Acids Res. 14, 623–633.
[38] Torruella, M., Gordon, K. and Hohn, T. (1989) EMBO J. 8, 2819–2825.
[39] Pietrzak, M. and Hohn, T. (1987) Virus Genes 1, 83–96.
[40] Kirchherr, D., Albrecht, H., Mesnard, J.M. and Lebeurier, G. (1988) Plant Mol. Biol. 11, 271–276.
[41] Gordon, K., Pfeiffer, P., Fütterer, J. and Hohn, T. (1988) EMBO J. 7, 309–317.
[42] Wurch, T., Guidasci, T., Geldreich, A., Lebeurier, G. and Mesnard, J.M. (1991) Virology 180, 837–841.
[43] Schultze, M., Hohn, T. and Jiricny, J. (1990) EMBO J. 9, 1177–1185.
[44] Medberry, S.L., Lockhart, B.E.L. and Olszewski, N.E. (1990) Nucleic Acids Res. 18, 5505–5513.
[45] Xiong, Y. and Eickbush, T.H. (1990) EMBO J. 9, 3353–3362.