Volume 327, number 2, 213–218 FEBS 12765 July 1993


# Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1

Takuma Hayashi[a,b], Yoshio Ueno[b] and Takashi Okamoto[a]

[a]*Virology Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104, Japan* and [b]*Faculty of Pharmaceutical Sciences, Science University of Tokyo, 12 Ichigaya-Funakawaramachi, Shinjuku-ku, Tokyo 162, Japan*

Received 7 June 1993

A region of human immunodeficiency virus type 1 (HIV-1) genomic RNA known to interact with the viral nucleocapsid protein is one of the crucial components for retroviral genomic RNA packaging. We have investigated the secondary structure for this RNA packaging signal of HIV-1 using RNA mapping techniques and extensive computer analyses. Our results suggest that the RNA sequence containing the packaging signal conforms to a stable stem-loop structure and that a portion of this structure is conserved among twenty independent HIV-1 isolates. The heterogeneity of the RNA packaging sequence was not correlated with the amino acid sequence variability of the viral nucleocapsid protein. These findings have critical implication in understanding the viral maturation process.

Stem-loop structure; Packaging signal; Human immunodeficiency virus type 1; Viral genomic RNA; Gag precursor protein

## 1. INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) shares many common features with other retroviruses [1–3]. The assembly of HIV-1 viral components into virions occurs on the surface of infected cells in parallel with the budding process, during or after which the virions mature [4]. When viral components of RNA virus assemble into virions, Gag protein containing nucleic acid binding motifs (zinc finger motifs) incorporates the viral genomic RNA by interacting with the specific sequence (called 'packaging signal') [5–7]. Since this protein can distinguish the viral genomic RNA from other viral and cellular mRNAs, this protein–RNA interaction is considered to be sequence and/or structure specific [8,9].

It has previously been shown that the portions of viral genomic RNA of human immunodeficiency virus type 1 (HIV-1) from its nucleotide positions +754 to +773 [10,11], +748 to +770 [12], and +754 to +784 [13] are required for incorporation into the virus particles. In the recent paper by Hayashi et al. [9], it was reported that the 46 nucleotide-long RNA stem-loop structure (between nucleotide positions +751 and +796) was necessary and sufficient for packaging into the HIV-1 virions, and that this stem-loop structure specifically was recognized by the p55 Gag precursor protein [9]. Here we demonstrate that this 46 base region forms a stable stem-loop structure using HIV-1 mRNAs prepared from the H9 cells infected with HIV-1 IIIB by RNA mapping analysis. Computer-assisted analysis has confirmed this structure.

*Corespondence address:* T. Okamoto, Virology Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104, Japan. Fax: (81) (3) 3543-2181.

## 2. MATERIALS AND METHODS

### 2.1. *Secondary structure determination of the HIV-1 packaging sequence by RNA mapping*

Polyadenylated RNA were prepared from H9 cells infected with HIV-1 IIIB (H9/IIIB) 48 h post-infection by the method previously described [9]. Preliminary experiments revealed that 84 h after infection turned out to be the most appropriate for preparation of the HIV-1 genomic RNA from H9 cells infected with HIV-1, as reported previously. [14]. (According to Kim et al. [14], during an asynchronous infection, the amounts of HIV-1 RNA, particularly the 9.2-kb genome-size viral RNA, gradually increased 48 h after infection and the ratio of the 2.0-, 4.3-, and 9.2-kb viral mRNAs at this point was approximately 1.0:0.8:2.5.) RNA was dissolved in each reaction buffer at a concentration of 10 mg/ml in diethylpyrocarbonate (DEP) buffer (5 mM sodium cacodylate (pH 7.5), 10 mM $MgCl_2$) or S1 buffer (10 mM Tris-HCl (pH 7.5), 200 mM NaCl, 2 mM $MgCl_2$). RNA was denatured at 72°C for 2 min, followed by gradual cooling to 20°C [15]. The renatured RNA (10 $\mu$g per reaction) was treated with varying concentrations of DEP (0.5, 1.0, 2.0% (v/v)) or digested with increasing amounts of RNAase S1 (0.1, 0.5, 1.0 U). The digested RNA was processed by phenol extraction and precipitated with ethanol. Primer extension was applied to detect the sites of the biochemical modifications by DEP or RNAase S1. The following primers were used in primer extension: primer 1 (5′CTCCCCCGCTTAATACT GAC3′; complementary to the nucleotide sequence from +803 to +822 of HIV-1 IIIB) and primer 2 (5′CCCCCTGGCCTTAACCGAAT3′; complementary to the nucleotide sequence from +843 to +862). Each oligonucleotide DNA primer was annealed to the RNA and extended by AMV reverse transcriptase (Boehringer Mannheim, Germany) with [$\alpha$-$^{32}$P]dCTP as previously described [16]. The extended and labelled single-strand DNA (ssDNA) was recovered by ethanol precipitation and was dissolved in sequencing gel-loading buffer. Samples were then analyzed in a 8% sequencing gel.

2.2. *Computer-assisted estimation of RNA secondary structure*

The programs employed were FOLD and SQUIGGLES (University of Wisconsin Genetics Computer Group) running on a VAX computer RNA secondary structure predictor (using the latest version 1.5 released in 1993) based on the free-energy minimization algorithm. The calculation was carried out with the folding windows of 20 bases at 10-base-pair intervals from the nucleotide position +455 (RNA CAP site of HIV-1 mRNA) to +1183.

## 3. RESULTS

### 3.1. *Structural mapping of the RNA packaging signal of HIV-1*

RNA secondary structure analysis using the latest version of FOLD program on VAX computer revealed that the 5′ region of HIV-1 RNA (nucleotide position +455 (CAP site) to +1183 of HIV-1 IIIB) containing the putative packaging signal (nucleotide position of +750 to +795) could form a stable stem-loop structure (Gibbs free energy value of −172.0 kcal) (Fig. 1). The region of the packaging signal also showed a stable stem-loop structure as shown in the inset of Fig. 1. It was previously shown that this region was functionally crucial for packaging of HIV-1 RNA into the virions [9].

To confirm the validity of this structure, we carried out RNA structure mapping analyses using authentic viral mRNA prepared from the H9 cells freshly infected with HIV-1 IIIB (H9/IIIB). Total RNA was purified from H9/IIIB cells 48 h after infection and was digested with single-strand specific reagents, DEP or RNAase S1 (the inset in Fig. 1 indicates the putative cleavage sites by these reagents according to the secondary structure predicted by computer). RNA structure mapping was performed using DEP or RNAase S1. Two different primers, primer 1 and 2, were used for each chemical modification study. The results of the RNA structure mapping analysis are presented in Fig. 2a. The bands corresponding the A or U residues in the bulge or loop regions were clearly demonstrated as indicated in Fig. 2b summarizing the actual cleavage sites by these reagents. The cleaved A's and U's corresponded to the unpaired A and U residues in the predicted RNA secondary structure model (Fig. 1). These results have indicated that the packaging region of HIV-1 genomic RNA synthesized in the virus-infected H9 cells actually conforms a stable stem-loop structure as predicted.

### 3.2. *Comparison of the RNA packaging signals from twenty independent HIV-1 isolates*

We then investigated the heterogeneieties of the packaging sequences and of the secondary structures among

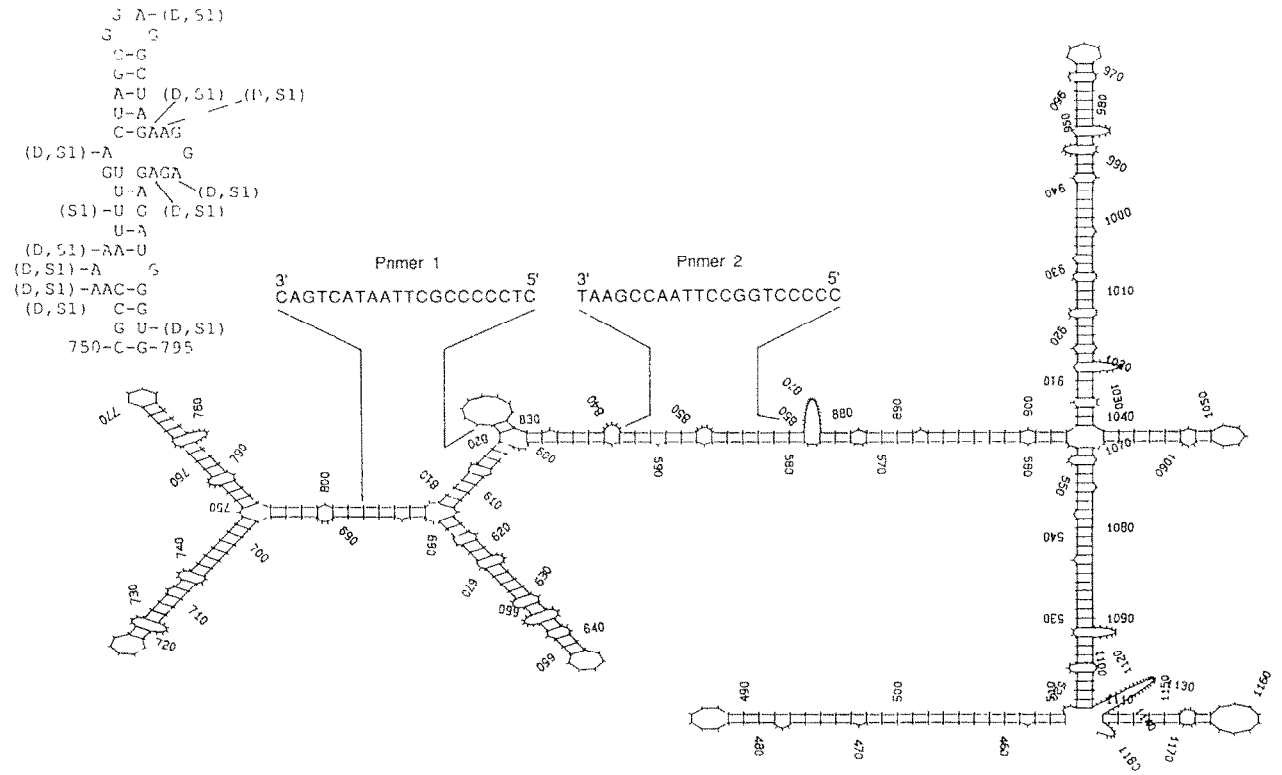


Fig. 1. Putative secondary structure of the packaging signal of HIV-1 RNA. Calculation was carried out with the sequence from HIV-1 IIIB strain sequence (from bases +455 (CAP site) to +1183) using FOLD program running on a VAX computer. The latest version of RNA secondary structure predictor (version 1.5; 1993) was applied for this analysis. The SQUIGGLES program was used for drawing the secondary structure The regions of the complementary primers, primer 1 and 2, used for RNA mapping analyses indicated. Inset indicates the putative secondary structure of the packaging signal and the expected cleavage sites by diethylpyrocarbonate (DEP) (A-specific) and RNase S1 (A and U-specific) D, possible site of cleavage by DEP. S1, possible site of enzymatic digestion by RNase S1.

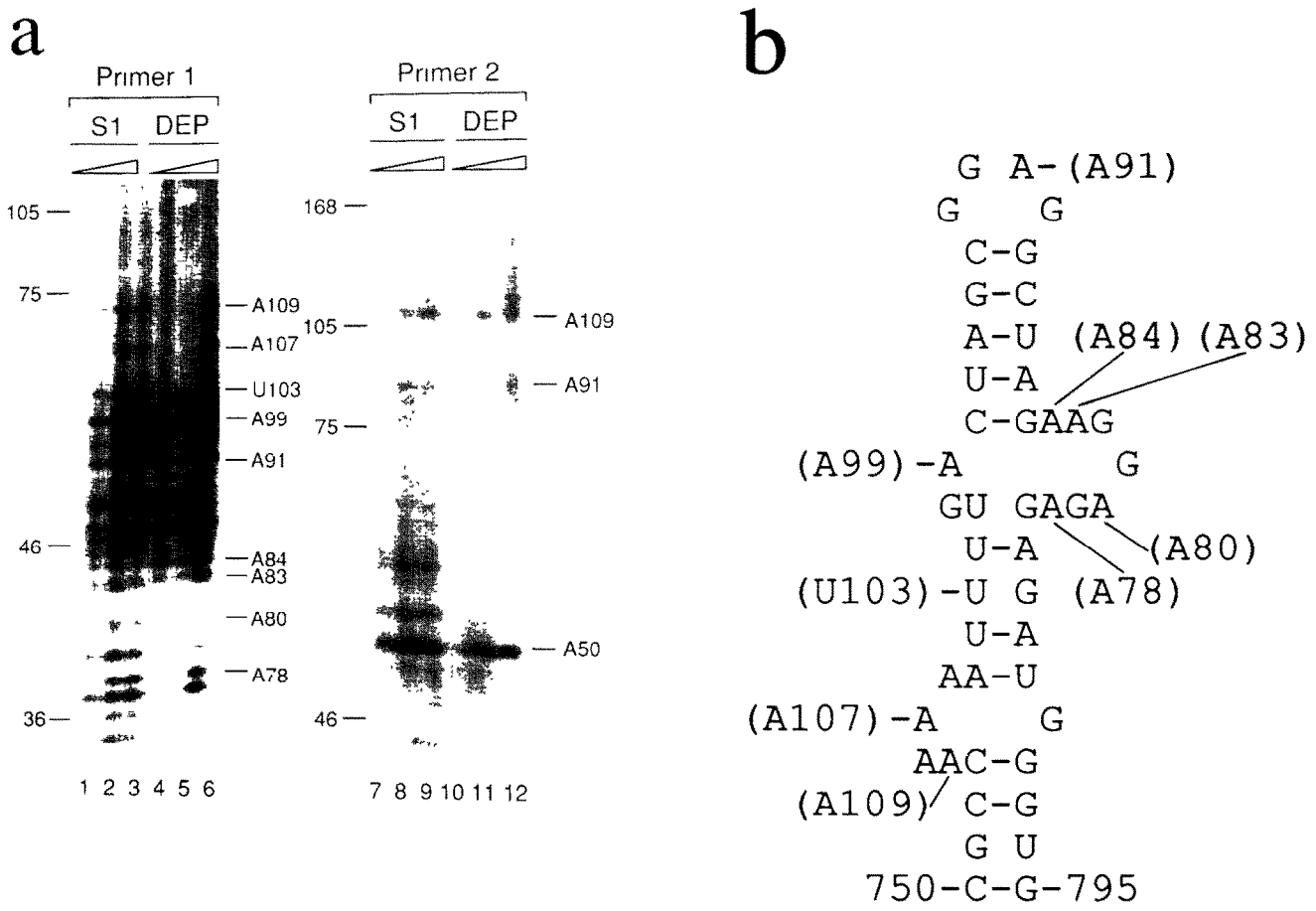


Fig. 2. (a) The results of RNA biochemical modification analysis of HIV-1 genomic RNA. The RNA was prepared from the H9 cells infected with HIV-1 IIIB. The RNA was denatured at 72°C for 2 min and re-annealed by gradual cooling to 20°C. The RNA modifications by two single-strand specific reagents, DEP and RNAase S1, were made. Each treatment (indicated above the lanes) was performed with increasing amounts of these reagents. Cleavage-modification sites were detected by primer extension using two primers (Fig. 1). Number on the left indicates nucleotide length. Positions of the cleaved nucleotides were indicated on the right (numbers following A or U indicate the locations from the end of primer 2 (nucleotide position +862 of HIV-1 IIIB)) (b) Putative secondary structure of the RNA packaging signal indicated from the RNA mapping analysis. Locations of the cleaved nucleotides within the packaging signal are indicated as in Fig. 2a.

different HIV-1 virus isolates to further characterize the critical region for the viral RNA packaging. Fig. 3a illustrates the sequence alignment of the packaging sequences from twenty HIV-1 strains which were independently isolated. Among nine of these strains the sequence was identical. The other eleven strains have sequence variability in either 5′ or 3′ region or in both periphery of the packaging signal (Fig. 3a). The central 22 nucleotides were conserved among all the HIV-1 strains examined here.

Based on the results of free energy calculations for each sequence, the secondary structures of the 46-base region obtained from these twenty HIV-1 strains were determined as illustrated in Fig. 3b. It was noted that the first bulge structure ('I') and the stem-loop region (from +762 to +778) are conserved while both peripheries are heterogeneous., while both peripheries are heterogeneous. It was also noted that this central region appeared to conform a stable structure.

### 3.3. *Comparison of amino acid sequences of the zinc finger motif regions in p55 Gag precursor protein among the HIV-1 isolates*

Since it is already known that HIV-1 genomic RNA is recognized by p55 Gag precursor protein specifically [5,9,12], we compared the amino acid sequence of the RNA-binding region containing the zinc finger motif in p55 Gag precursor protein of eighteen HIV-1 strains (the sequences of two HIV-1 strains, NEFAAE and WMJ21, were not available). The amino acid sequences of the nucleic acid binding motifs of p55 appeared to be well conserved and the variation was rather salient. It was noted that the amino acid sequence variability did not correlate with either the nucleotide sequence or the

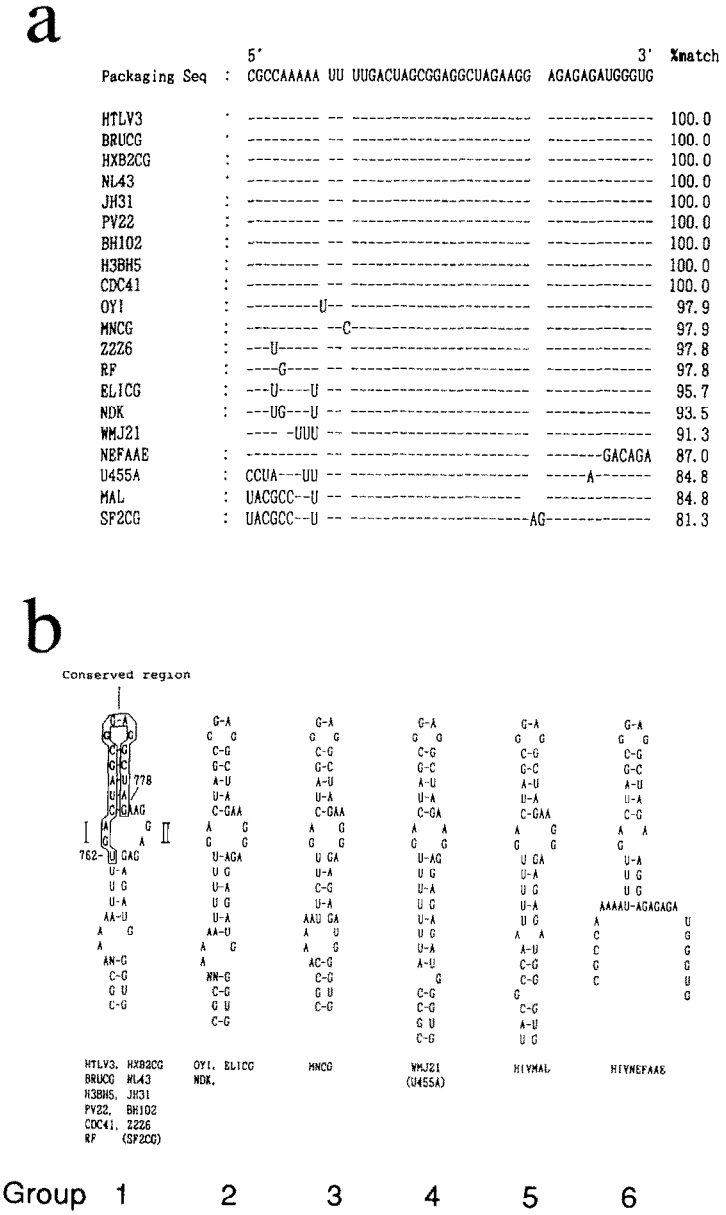


Fig. 3. (a) Alignment of the RNA packaging signal sequences from twenty independent isolates of HIV-1. Matched nucleotides are indicated by horizontal lines. Percent homology was indicated on the right. (Sequence data from this article have been deposited with the DDBJ under Accession No. D13115). (b) Possible secondary structures of the RNA packaging signal from twenty HIV-1 strains. The structural motifs were classified into six groups according to the nucleotide sequence in the second bulge (II). The location of the conserved RNA structure is indicated by a box.

structure heterogeneity of the packaging signal (Fig. 4). Most of the amino acid changes are among similar amino acid species except a few residues. These results have indicated that the amino acid sequence variation dose not account for the heterogeneity of the structure of RNA packaging signal.

## 4. DISCUSSION

Molecular interaction between the p55 Gag precursor protein and the viral genomic RNA of HIV-1 is the critical step for the production of infectious virions [4,5,9,12]. In this study, we have attempted to determine the secondary structure of the packaging signal on the viral genomic RNA by employing two independent methods, a computer-assisted prediction of the RNA structure and biochemical methods using two reagents that cleave the unpaired bases in the RNA molecule. Both of these methods have delineated the same secondary structure of the viral packaging signal (Figs. 1 and 2b). Furthermore, the sequence and structural heterogeneity of the packaging signal was investigated by comparing twenty HIV-1 isolates. The analysis has revealed

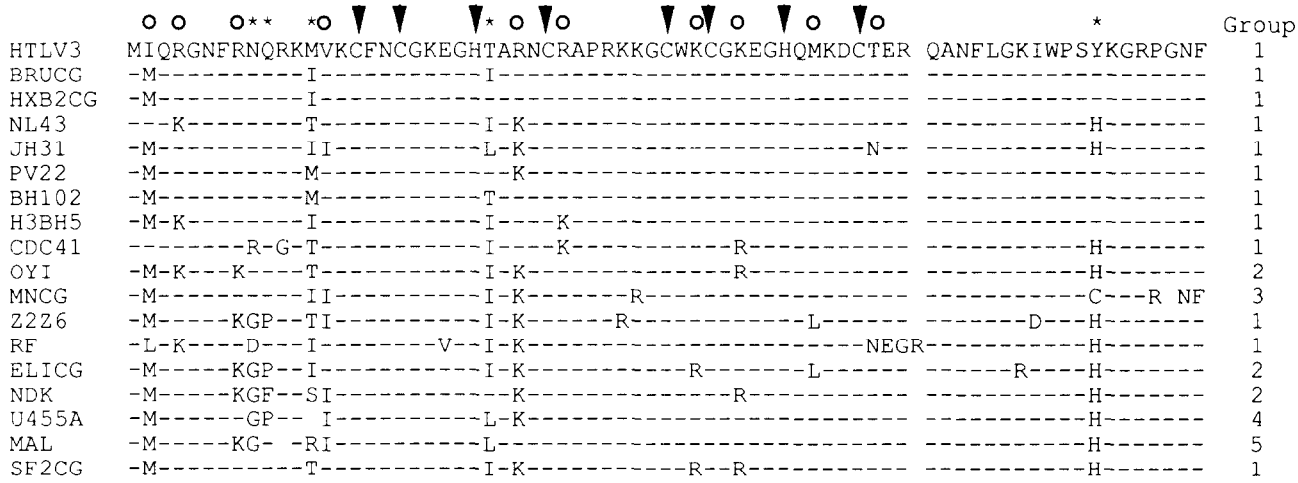


Fig. 4. Amino acid sequence alignment among the regions of two copies of zinc finger motif in p55 Gag precursor protein. Matched amino acid residues are indicated by horizontal lines. The 'group' number on the right indicates the group category classified according to the secondary structure of the packaging signal (Fig. 3). Cysteines within the zinc finger motif are indicated by arrowheads. Amino acid residues indicated by open circles or asterisks indicate substitution between conserved amino acids and non-conserved amino acids, respectively.

a conserved secondary structure containing a bulge and stem-loop region within the RNA packaging signal (Fig. 3). Similarly, the amino acid sequences of the p55 Gag precursor proteins from these virus isolates were compared (Fig. 4). Since variation of amino acid sequence in p55 Gag did not correlate with that of the packaging signal (Figs. 3 and 4), it was suggested that the salient amino acid sequence variation in the nucleic acid binding region of p55 Gag precursor protein and the peripheral portions of the packaging sequence might not affect the specificity of this protein–RNA interaction.

Harrison et al. recently reported the results of similar experiments [8]. However, the putative RNA secondary structure that they depicted was different from what we demonstrated here. This discordance can be explained as the followings: first, the viral RNA preparations used, analysis were different. They used a segment (from bases 541 (86 bases downstream from the CAP site) to 1086) of the HIV-1 RNA which was in vitro synthesized by riboprobe system, while we used authentic viral RNA freshly isolated from the HIV-1-infected cells. Second, the computer program for free-energy minimization they used was an old version (version 1.22; 1987), as compared with this study (version 1.5; 1993), which resulted in the limitation of the maximum sequence for RNA secondary structure determination. Therefore, their results may represent a partial solution with the limited extent of the RNA sequence. Using the same sequence they utilized for computation we obtained the same results. It may not be relevant to analyze only a portion of the viral RNA since all the viral transcripts are synthesized from the CAP site in H9 cells infected with HIV-1.

The incorporation of viral genomic RNA into the viral particles is one of the crucial steps for production of infectious virions. Protein–RNA interaction between p55 Gag precursor protein and the stable stem-loop RNA structure within the viral packaging signal could be a good target for therapy against AIDS. Moreover, the information obtained from this study should also be useful for the development of a novel retrovirus vector based on HIV-1.

## 4. NOTE ADDED IN PROOF

During the preparation of this manuscript, Baudin et al. [17] have reported the results with similar RNA mapping study using in vitro synthesized HIV-1 RNA. They demonstrated the same secondary structure of the packaging signal, thus confirming our observations using the authentic viral genomic RNA.

*Acknowledgements:* This study was supported by Grants-in-Aid from the Ministry of Health and Welfare and the Ministry of Education, Science and Culture of Japan.

## REFERENCES

[1] Cullen, B.A. (1991) J. Virol. 65, 1053–1056.
[2] Haseltine, W.A. (1988) J. Acquired Immune Deficiency Syndrome 1, 217–240.
[3] Kohl, N.E., Enimi, E.A., Schleif, W.A., Davis, L.F., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M., Signal, I.S. (1988) Proc. Natl. Acad. Sci. USA 85, 4686–4690.
[4] Peterlin, B.M. and Luciw, P.A. (1988) AIDS 2, 29–40.
[5] De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M.C., Roques, B. and Darlix, J.L. (1992) Proc. Natl. Acad. Sci. USA 89, 6472–7476.
[6] Dupraz, P. and Spahr, P.F. (1992) J. Virol. 66, 4662–4670.
[7] Gorelick, R.J., Nigida, S.M., Bess, J.W., Arthur, L.O., Henderson, L.E. and Rein, A. (1990) J. Virol. 64, 3207–3211.

[8] Harrison, G.P. and Lever, A.M. (1992) J. Virol. 66, 4144–4153.
[9] Hayashi, T., Shioda, T., Iwakura, Y. and Shibuta, H. (1992) Virology 188, 590–599
[10] Lever, A., Gottlinger, H., Haseltine, W. and Sodroski, J. (1989) J. Virol 63, 4085 4087.
[11] Poznansky, M., Lever, A., Bergeron, L., Haseltine, W. and Sodroski, J. (1991) J Virol . 65, 532–536.
[12] Aldovini, A. and Young, R.A. (1990) J. Virol. 64, 1920–1926.
[13] Clavel, F. and Orenstein, J M. (1990) J. Virol. 64, 5230–5234
[14] Kim, S., Byrn, R., Groopman, J. and Baltimore, D (1989) J. Virol. 63, 3708–3713.
[15] Berkhout, B., Silverman, R.H. and Jeang, K.T. (1989) Cell 59, 273–282.
[16] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, 2nd edn , Cold Spring Harbor Laboratory Press. pp. 7.79–7.83
[17] Baudin, F , Marquet, R., Isel, C., Darlix, J.-L., Ehresmann, B. and Ehresmann, C. (1993) J Mol. Biol 229, 382–397.