Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat

Woo Jung Lee, Hyun Jin Kwun and Kyung Lib Jang

Department of Microbiology, College of Natural Sciences, Pusan National University, Pusan 609-735, South Korea

INTRODUCTION

All humans carry human endogenous retroviruses (HERVs) as an integral part of their genome. Up to 5% of the genome consists of HERV sequences, corresponding to about 1500–2000 proviruses and at least 20 000–40 000 copies of solitary long terminal repeats (LTRs), considered to be relics of ancient proviruses resulting from internal recombination (Wilkinson et al., 1994). They have been amplified during evolution by repeated reintegration of reverse-transcribed mRNA into the DNA of germ line cells (Jurka, 1998; Leib-Mösch & Seifarth, 1996; Löwer et al., 1996; Patience et al., 1997). By virtue of their ability to integrate randomly into the host genome, HERVs are considered to be an important class of insertion mutagens (Di Cristofano et al., 1995).

HERV gene expression is mostly under control of their LTRs. HERV solitary LTRs can often be found in close vicinity to functional genes (Sverdlov, 1998). In addition, the U3 region of HERV LTRs contains all the sequences required for initiation of transcription, such as promoters, enhancers and transcription factor (TF) binding sites (Domansky et al., 2000; Knössl et al., 1999; Schön et al., 2001); thus it not only regulates transcription of viral genes but also influences the expression of neighbouring cellular genes. Indeed, some HERVs are transcribed, and HERV proteins as well as replication-defective virus particles have been detected in several tissues (Boller et al., 1993; Seifarth et al., 1995; Lyden et al., 1994). Furthermore, HERVs might cause a range of disease processes including neoplasia, auto-immunity, encephalitis and foetal malformations (O’Reilly & Singh, 1996; York et al., 1992), although their biological significance awaits clarification.

Members of the HERV-W family, originally isolated from monocytes and leptomeningeal cells from patients with multiple sclerosis, are preferentially expressed in normal human placenta (Blond et al., 1999) where they might play a role in morphogenesis by mediating trophoblast fusion (Mi et al., 2000). Computer sequence analysis suggested the presence of several putative TF binding sites in the U3 promoter region and its promoter activity varies significantly in different human cell types (Schön et al., 2001). However, the precise transcriptional regulatory mechanism is still obscure because little is known about the action of each putative TF binding site on transcriptional initiation from the HERV-W LTR. Therefore, in the present study, we tried to identify a regulatory sequence(s) that is important for HERV-W LTR-directed transcription in order to understand transcriptional regulation of the LTR. Furthermore, we analysed nucleotide sequences in the regulatory sites of several HERV-W LTR isolates and compared their promoter activities. In addition, we investigated which regulatory site is responsible for the cell type-specific promoter activity of HERV-W LTR.

METHODS

Cloning and sequence analysis of HERV-W LTR. Six different HERV-W LTR isolates were obtained by PCR amplification from
genomic DNA samples prepared from blood cells of Korean schizophrenia patients. A primer pair, HW-LTR forward (5'-AGG ACA AGC TTG ATT TCC TAG G-3'; nt 1-22) and HW-LTR reverse (5'-TGG AGG TAC TCT CAT GGT T-3'; nt 847-828), was derived from the consensus sequences of HERV-WLTR (Blond et al., 1999). An initial denaturation at 94 °C for 5 min was followed by 30 cycles with denaturation at 94 °C for 1 min, annealing at 52 °C for 1 min, elongation at 72 °C for 1 min, and a final extension at 72 °C for 5 min. PCR products cloned into pBluescript II Blue vector (Novagen) were sequenced using an automated DNA sequencing model (373A). The accession numbers of the isolates obtained from the GenBank database by using the BLAST network server (Altschul et al., 1997) were AY208132 for HW1, AY208133 for HW2, AY208134 for HW3, AY208135 for HW4, AY208136 for HW5 and AF531174 for HW6, respectively.

Construction of HERV-W LTR reporter plasmid. The full-length HERV-W LTR fragment of HW6 was subcloned into the KpnI and HindIII sites of the luciferase reporter vector pGL2-basic (Promega) to create HW\_6\_447. For construction of HW\_1\_772 and HW\_2\_447, the sequence either downstream or upstream of the SacI site at nt 772, respectively, were deleted from HW\_1\_847. For construction of HW\_1\_575, the sequence downstream of the Spel site at nt 575 was deleted from HW\_1\_847. For HW\_1\_347, the U3R fragment from nt 1 to 347 was amplified using as primer set HW-LTR forward and HW347R. For HW\_3\_47, the 74 bp fragment at the 3' terminus of the LTR was cloned into the KpnI and HindIII sites of the luciferase vector. Five additional U3R-containing luciferase constructs, HW1 to HW5, were prepared by the same procedure. For construction of HW\_3\_47, the sequence amplified using as primer set HW347F (5'-CTG CAC TCT TCT TGT CCG T-3') and HW-LTR reverse was cloned into the Smal and SacI sites of the luciferase vector. For HW\_1\_447 and HW\_3\_47, the LTR fragments of HW\_1\_847 and HW\_1\_347, respectively, were cloned in reverse orientation into the HindIII and KpnI sites of pG2L2 vector. For HW\_1\_259 and HW\_1\_347, the sequence either downstream or upstream of the Nael site at nt 259 was deleted from HW\_1\_347. For construction of HW\_1\_143 and HW\_1\_347, the sequence either downstream or upstream of the SacI site at nt 143 was deleted. For HW\_1\_347, the fragment from 161 to 347 was amplified using as primer set HW161F (5'-AAA ATG CTA GCT AGC CAA AA-3') and HW347R, and was cloned into the Nhel and HindIII sites of the luciferase vector. For HW\_1\_347, the fragment from 191 to 347 was amplified using as primer set mCAAT-F (5'-AAAT AGC TAG CCA TCT ATT GCC TGA-3') and HW347R, respectively, were cloned in reverse orientation into the HindIII and KpnI sites of the luciferase vector. For HW\_1\_259 and HW\_1\_347, the sequence either downstream or upstream of the SacI site at nt 259 was deleted from HW\_1\_347. For construction of HW\_1\_143 and HW\_1\_347, the sequence either downstream or upstream of the SacI site at nt 143 was deleted. For HW\_1\_347, the fragment from 161 to 347 was amplified using as primer set HW161F (5'-AAA ATG CTA GCT AGC CAA AA-3') and HW347R, and was cloned into the Nhel and HindIII sites of the luciferase vector. For HW\_1\_347, the fragment from 191 to 347 was amplified using as primer set mCAAT-F (5'-AAAT AGC TAG CCA TCT ATT GCC TGA-3') and HW347R, and was cloned into the Nhel and HindIII sites of the luciferase vector. HW\_1\_143, which has an internal deletion from nt 144 to 160, was constructed by ligating the two fragments from HW\_1\_143 and HW\_1\_347. For HW\_AGC/EBP, the fragment from 1 to 178 was amplified using as primer set HW161F (5'-AAA ATG CTA GCT AGC CAA AA-3') and HW347R.

**RESULTS**

Analysis of regulatory sequences responsible for the transcription from HERV-W LTR

Initially, we measured the promoter activity of HERV-W LTR in HeLa cells, which express HERV-W at a relatively high level (Schön et al., 2001). For this purpose the reporter construct HW\_1\_847, which contains the full-sized HERV-W LTR of HW6 (GenBank accession no. AF531174) in forward orientation relative to the reporter gene, was transfected into HeLa cells. The promoter activity of the LTR was measured by luciferase assay and compared with that of the extract from the promoterless vector (pGL2) transfected cells. Results from this experiment clearly indicate that the HERV-W LTR can enhance the basal expression of the luciferase gene approximately 13-fold in HeLa cells (Fig. 1A).

To characterize the regulatory sequences in the HERV-W LTR responsible for the activation of transcription, several reporter constructs, each of which contained a series of progressive 5' - or 3' - deletion mutants of the HERV-W LTR, were generated (Fig. 1). Deletion of the LTR from the 3' end up to nt 348 increased the LTR activity approximately 2-fold, suggesting the presence of a negative regulatory element(s) in the U5 region of the HERV-W LTR. Deletion of the 74 bp fragment located at the 3' end of the U5 region, as in HW\_1\_772, was sufficient to increase the LTR promoter activity. To confirm that the fragment defined above has a negative regulatory element responsible for the repression of transcription from the LTR, either the entire U5 region or the 74 bp fragment at the 3' terminus of the LTR was cloned into pG2L2-luc vector to generate HW\_3\_47 or HW\_772, respectively. Consistently with the deletion analysis described above, both constructs exhibited much lower luciferase activity compared to the empty vector, confirming
Fig. 1. Analysis of transcriptional regulatory sequences in the HERV-W LTR. Schematic diagrams of the luciferase constructs containing various HERV-W LTR fragments are presented on the left. Five putative transcription factor-binding sites and a TATA box are indicated in (B). The construction procedures are described in detail in Methods. Each construct was transfected into HeLa cells for luciferase assay. Luciferase activity was normalized to the β-galactosidase activity measured in the corresponding cell extracts. The average and standard deviations (error bars) from three independent experiments are presented.
that a negative silencer-like regulatory element(s) is present in the U5 region of the HERV-W LTR. In addition, Fig. 1(A) shows that the HERV-W LTR possesses bidirectional promoter activity. When the reporter constructs containing either the full-sized LTR or U3R sequence in reverse orientation towards the reporter gene were tested in a parallel transfection assay no significant differences in promoter activity were observed between luciferase constructs containing the LTR in either the forward or reverse orientation.

**Identification of transcriptional regulatory elements in the U3 region of HERV-W LTR**

According to the above results, most of the transcription stimulatory activity of the LTR resides in the U3R region. To determine whether the R region has an enhancer activity, the sequence from nt 259 to 347 corresponding to the R region was deleted to generate HW1–259. Such a deletion did not change the basal luciferase activity at all, suggesting the absence of enhancer activity in the R region. In addition, the R region alone did not increase the basal luciferase activity of pG5L2, as demonstrated with HW259–347. Therefore, we concluded that the U3 region contains all of the regulatory sequences required for the transcriptional activation of HERV-W LTR and decided to analyse it in detail in order to elucidate the transcriptional activation mechanism of the LTR.

Based on computer sequence analysis, Schön *et al.* (2001) suggested the presence of five putative transcription factor binding sites in addition to a TATA box in the U3 region of the LTR, which are also completely conserved in our isolate (Fig. 1B). To determine which of these regulatory sites is critical for the transcriptional regulation of the LTR, additional luciferase constructs were prepared as shown in Fig. 1(B). Deletion of the sequence between nt 1 and 143 hardly affected the LTR activity, as demonstrated with HW1–143. In addition, the sequence between nt 1 and 143 had little effect on the basal luciferase activity of pG5L2, suggesting that the three putative TF binding sites located in the 5’ half of the U3 region might be dispensable for the LTR activity. Interestingly, further deletion up to nt 160 increased the LTR activity up to 3-fold, suggesting the presence of a negative regulatory element between nt 143 and 160. This was also demonstrated with HWdOct1, which contains an internal deletion of the sequence from nt 144 to 160. Conversely, additional deletion of 30 nt between nt 161 and 190 reduced the LTR activity approximately 5-fold, indicating the presence of an enhancer sequence within this region. This was also demonstrated with HWdC/EBP, which has an internal deletion of the sequence from nt 179 to 190. HW191–347, which includes a TATA box and the R region of the LTR, still exhibited a relatively high luciferase activity, approximately 8-fold higher than that of empty vector. Considering the absence of a regulatory activity in the R region, most of the luciferase activity from HW191–347 might be due to the sequence from 161 to 259, which includes a TATA box.

Recently, Schön *et al.* (2001) suggested a negative regulatory role of the Sp1 site identified at nt 220 to 233 in all the inactive HERV-W LTR isolates, but absent from all the active HERV-W LTR isolates. As none of the HERV-W LTR isolates we obtained contain such a regulatory site, we artificially introduced it into HW1–347 at the corresponding site and tested its effect on the LTR activity to determine whether the Sp1 site actually has a repressive activity for transcription. However, the luciferase activity from HWsp1 was approximately 3-fold that of HW1–347 (Fig. 1B), which is opposite to the prediction of Schön *et al.* (2001).

**The importance of Oct-1 and C/EBP sites for the transcriptional regulation of HERV-W LTR**

The silencer and enhancer sequences defined in this study contain transcription factor binding sites for Oct-1 and CCAAT/enhancer binding protein (C/EBP), respectively (Fig. 1B; Schön *et al.*, 2001). To prove that the Oct-1 binding site located between nt 161 and 190 is responsible for the silencer activity, we constructed HWmOct1, which contains GAT instead of ATG at the first 3 nt of the Oct-1 binding site (Fig. 2). Such nucleotide substitutions severely damage this binding site (Douville *et al.*, 1995). As expected, the luciferase activity from HWmOct1 was approximately 3-fold higher than that of HW1–347. In addition, destruction of the CAAT-box by nucleotide substitutions as in HWmC/EBP-1 and HWmC/EBP-2 significantly reduced the LTR activity, confirming that the C/EBP binding site acts as an enhancer for transcription (Fig. 2). HWmC/EBP-1, which contains substitutions at both TF binding sites, exhibited a luciferase activity intermediate between that of HWmOct1 and HWmC/EBP-2, probably due to the loss of both silencer and enhancer activities.

HERV-W LTR isolates can be divided into two groups depending on their LTR strength (Schön *et al.*, 2001). It is possible that the difference in LTR strength results from the sequence variation in the two regulatory sequences defined in this study. To prove this hypothesis, we obtained five additional HERV-W LTR isolates from blood cells of Korean schizophrenia patients. Sequence comparison of the isolated HERV-W LTRs showed that they had 88–92 % sequence identity to the HERV-W-CL6 LTR (Komurian-Pradel *et al.*, 1999). Most of the putative TF binding sites described in Fig. 1(B), including the Oct-1 binding site and CAAT-box, are relatively well conserved among the LTR isolates, except for HW1 which has a substituted A instead of G at the third nucleotide position of Oct 1 site (Fig. 2). HW1 had approximately 2-fold higher promoter activity whereas other clones showed similar luciferase activity to HW1–347. Furthermore, the increase of luciferase activity brought about by complete destruction of the Oct-1 site in HW1–347 was not clear in the case of HW1, probably because HW1 already had an impaired Oct-1 site. This result provides further evidence that the Oct-1 site acts as a negative regulatory sequence for transcription. We could not obtain an HERV-W LTR clone with substitutions in the CAAT-box, possibly reflecting its important role for the activation of HERV-W LTR.
**Cell type-specific promoter activity of HERV-W LTR**

To characterize the cell type-specific promoter activity of the HERV-W LTR, the luciferase construct HW_{1-347} was transfected into several human cell lines, including teratocarcinoma (Tera-1), cervical carcinoma (HeLa), colon cancer (HCT116), hepatocarcinoma (HepG2), embryonal kidney (HEK-293) and breast cancer (MCF-7) cells. The promoter activity of the LTR was measured by luciferase assay and was normalized to the β-galactosidase activity measured in the corresponding cell extracts to compensate for the differences in transfection efficiency and expression rate. The luciferase activity of HW_{1-347} was highest in Tera-1 cells and lowest in MCF-7 cells (Fig. 3A). The activity in HeLa was higher than in MCF-7, which is consistent with the previous report by Schönh et al. (2001). When the amount of HERV-W RNA was measured by slot blot analysis, Tera-1 had approximately 2-fold higher HERV-W transcripts compared to HeLa cells (Fig. 3B), which is consistent with the results of the luciferase assay shown in Fig. 3(A).

Finally, we tried to define the element responsible for the cell-type specificity of the LTR. For this purpose, we compared the promoter activity of some mutant constructs described in Fig. 1 after transfecting them into either Tera-1 or HeLa cells. According to the results of the luciferase assay shown in Fig. 3(C), the cell type-specific activity of the LTR was maintained even in HW_{191-347}, which does not contain all of the putative TF binding sites shown in Fig. 1(B), including the Oct-1 binding site and CAAT-box. In addition, the luciferase activity from HW_{191-Oct} and HW_{191-C/EBP}, in which the binding sites for Oct-1 and C/EBP, respectively, have been destroyed, was also cell-type specific. Therefore, the two regulatory sites important for the regulation of basal level promoter activity of HERV-W LTR might not be responsible for determining the cell type-specificity. In contrast, no significant differences in the promoter activity of HW_{259-347}, which contains only the R region, were observed between the two cell lines. Therefore, we concluded that the minimal region responsible for the cell-type specificity of HERV-W LTR is the 3′ terminus of U3 from nt 191 to 259, which includes a TATA box.

**DISCUSSION**

In the present study, we tried to understand the transcriptional regulatory mechanism of the HERV-W LTR. According to the LTR analysis, two TF binding sites for Oct-1 and C/EBP seem to be important for the regulation of transcription from the LTR although the TATA box alone showed a significant stimulatory effect on transcription. In addition, the 74 bp sequence located at the 3′ terminus of the U5 region exhibited a repressive activity. Such an inhibitory effect of the element(s) in the U5 region was also described for other retroviruses such as HERV-K (Domansky et al., 2000; Kwun et al., 2002), human foamy virus (Yang et al., 1997) and human T-cell leukaemia virus type 1 (Okumura et al., 1996). Studies to identify the elements and the inhibitory mechanisms are under way.

The HERV-W LTR possesses bidirectional promoter activity, acting in either the forward or reverse orientation. A similar reverse promoter activity of the U3 region has also been reported in several cases including the HERV-K LTR (Domansky et al., 2000), the HERV-H LTR (Feuchter & Mager, 1990), the murine IAP LTR (Christy & Huang, 1998) and the HIV-LTR (Peeters et al., 2000). Therefore, at least some retroviral LTRs might promote transcription of downstream as well as upstream genes. Possible functions of this reverse promoter activity in the retroviral LTRs, as suggested by Domansky et al. (2000), may include involvement in inhibition of flanking cellular gene transcription, production of double-stranded RNA or the regulation of synthesis of positive-strand-derived proteins by antisense RNA.

Recently, Schönh et al. (2001) divided HERV-W LTRs into two groups by sequence comparison and phylogenetic analysis. They found that all LTR sequences strong in promoter activity belong to group I whereas all weak LTRs...
sequences were allocated to group II. In addition, to explain the difference in promoter activity between two groups, they argued for a negative regulatory role of the Sp1 site identified at nt 220 to 233 in the inactive HERV-W LTR isolates, but absent in the active HERV-W LTR isolates. However, according to our results, the Sp1 site, at least in our context, did not exhibit such a silencer activity but showed an enhancer activity, stimulating transcription from the LTR. Therefore, the Sp1 site might be not responsible for the weak promoter activity observed in the members of group II. Instead, our present study suggests that the Oct-1 and C/EBP binding sites are important for determining the LTR strength. The repressive role of the Oct-1 binding site was demonstrated by the high promoter activity of HW1, which has an impaired Oct-1 site. In addition, all of the inactive HERV-W LTRs identified by Schön et al. (2001) had an altered CAAT-box with a substituted G at the fifth nucleotide, i.e. GCCAGT, whereas all active ones did not contain such a substitution. Therefore, it is possible that the presence or absence of functional binding sites for Oct-1 and C/EBP may determine whether the HERV-W LTR is active or not. We obtained much lower promoter activity when the fifth nucleotide of the CAAT-box was substituted, A into G.

Our present study also demonstrated that the promoter activity of the HERV-W LTR varies significantly depending on the cell type. Neither the Oct-1- nor the C/EBP binding site was required for the cell type-specific activity of the HERV-W LTR, excluding the possibility that it results from differential distribution of Oct-1 or C/EBP depending on the cell type. Instead, the basic promoter, including a TATA box located at the 3' end of the U3, was enough to confer cell type specificity, suggesting that the efficiency of assembly of basic transcription machinery at the TATA box of HERV-W might be different depending on the cell type. However, the possibility that other unidentified regulatory sites present at nt 191 to 259 are responsible for the specificity cannot be excluded.

In conclusion, in this study we analysed the regulatory sequences of the HERV-W LTR in detail in order to understand the cell type-specific activity as well as the transcriptional regulatory mechanism of the LTR. Further studies might be necessary to elucidate the exact mechanism of transcription from the HERV-W LTR. First of all, it is important to clarify whether and how binding of Oct-1 to the site defined in this study represses the promoter activity of the HERV-W LTR. The other question to be solved is how the TATA alone confers cell type specificity. In addition, the repressive activity of the U5 region was partially defined in this study. It might be necessary to analyse more HERV-W isolates to divide HERV-W LTRs into two groups based on the regulatory sequences defined in this study.

ACKNOWLEDGEMENTS
This work was supported by grant no. 2000-2-20500-001-3 from the Basic Research Program of the Korea Science & Engineering Foundation.
REFERENCES


