

Large Number of Polymorphic Nucleotides and a Termination Codon in the *env* Gene of the Endogenous Human Retrovirus ERV3

H.B. Rasmussen and J. Clausen

*Department of Life Science and Chemistry,
Roskilde University, PO Box 260, DK-4000
Roskilde, Denmark*

ABSTRACT: The terminal portion of the *pol* gene and the entire *env* gene of the human endogenous retrovirus ERV3 was screened for polymorphic nucleotides. For this purpose fragments amplified from the desired regions of ERV3 were subjected to single strand conformational analysis (SSCP analysis). Using this approach, we detected 13 polymorphic nucleotides, namely four in the *pol* gene and nine in the *env* gene. Three of the nucleotide substitutions were synonymous (not affecting the amino acid code). One of the non-synonymous nucleotide substitutions changed an arginine codon to a termination codon. The alleles at the different polymorphic sites could be arranged into five ERV3 haplotypes, two of which were new.

To evaluate the possible significance of the termination codon, which precludes expression of a putative immunoregulatory factor, we examined samples of DNA from patients with multiple sclerosis, a demyelinating disease of presumed autoimmune etiology. We did not find an association between the ERV3 allele with the termination codon and this disease.

Perhaps the presence of a stop codon combined with the high number of non-synonymous nucleotide substitutions in the reading frame of the *env* gene reflects absence of selective constraints during evolution. Obviously, our findings contradict the assumption that the reading frame of the ERV3 *env* gene has been conserved throughout evolution.

KEYWORDS: human endogenous retrovirus, polymorphisms

INTRODUCTION

Vertebrates carry a high number of endogenous proviruses in their genomes, including humans [13]. To a large extent their significance in health and disease remains obscure. Attention has been focused upon endogenous retroviruses in studies of malignant diseases [5,22]. Findings that endogenous retroviruses possess immunoregulatory properties have opened up a new line of research to examine etiologies of autoimmune diseases [19]. Furthermore, endogenous retroviruses are capable of interacting with exogenous counterparts through a number of different mechanisms thereby affecting the outcome of retroviral infections [18].

The human endogenous retrovirus ERV3 is a full-length retroviral segment with in-frame stop codons in the *gag* and *pol* regions [16]. However, the *env* gene reading frame of this endogenous retrovirus, encoding an envelope surface unit (SU) and a transmembrane protein (TM), is generally assumed to have persisted throughout 30 million years of primate evolution [1]. This would probably not have been possible in the absence of selective forces [1].

Transcription of the ERV3 *env* gene has been detected in a variety of normal tissues including placentas [3]. An interesting aspect is that ERV3 is transcribed into a downstream sequence, encoding a *Krüppel*-like zinc-finger protein [10]. Absence of ERV3 transcripts in choriocarcinoma has led to speculations that ERV3 or the downstream zinc-finger sequence encodes a tumour suppressor [3,10]. Although this possibility has now been dismissed [1,15], there is reason to believe that ERV3 possesses certain

functions [1]. Perhaps its TM protein has immunoregulatory properties and participates in down-regulation of the maternal immune response at the placental barrier [1,25]. Recently, antibodies to a synthetic peptide from the *env* region of ERV3 were detected in Sjogren's syndrome and in pregnant women, particularly in mothers giving birth to babies with congenital heart block [12]. These findings indicate that the products of the *env* gene of ERV3 are potential autoantigens.

Table 1
Amplification primers

Sequence	Position*
TGAAAGACTGTTTCGGCAAGC	17–36
CCAAGGCACAATTCCTGCAA	301–320
CTTGTCTATTCCCACTGTTG	275–294
GACTTCAGATTTCTAACCAC	580–599
AAGCTTGAGTCATCATCAGG	551–570
GTTTTAGTCATGATGTTCCC	799–818
ACTGCACCCACACTACGTGT	777–796
AAGGATACGACATCCTTGCC	1024–1043
CCAAACCAAGGTATTTCCCT	1002–1021
CCTAGCGGTGTCTTTAAACC	1330–1349
CAGATCAGCCCATATGGACA	1307–1326
GCCATTGGTCTCCCATGTTC	1581–1600
AGCCTGCACGTTGCTTCATG	1543–1562
CTTTTGCCCTCCATAAAGTC	1809–1829
TATTACAACGAGACACTAGG	1786–1805
TCATCATAGATGGGGTATCC	2059–2078
CCTAAAACAGGGAGAAGCCT	2037–2056
CTGGGCAAGCAGATTCAAGG	2282–2301
ATTACCAATGAACTGCAGG	2260–2279
TCCTCCAAGGGATAAGAACC	2528–2547
CCAGATTCCCTTTTCAGAGG	2506–2525
GAGAAAGCATCACTATCTTC	2743–2762

* Numbering of nucleotides is in accordance with previous sequence data [2] and GenEMBL (accession number M12140). The *pol* coding region ends at nucleotide 494. Probable start of the *env* reading frame is at nucleotide 808.

Using restriction fragment analysis a genetic association between ERV3 and rheumatoid arthritis was previously reported [21]. A more recent study did not support this observation [24]. Based upon polymorphisms in the upstream and downstream long terminal repeats (LTRs) of ERV3, we have previously defined three allelic forms or haplotypes of this endogenous retrovirus [17]. Examination of their distributions in healthy individuals and patients with multiple sclerosis (MS), a disease of presumed autoimmune etiology, did not reveal a significant difference [17]. In the present study we have extended our search for polymorphisms to the terminal part of the *pol* gene and the entire *env* gene of ERV3, including the segment encoding the TM protein. Our aims were to provide a more detailed characterization of ERV3 and to further examine a possible association between this endogenous retrovirus and autoimmune diseases, exemplified by MS.

MATERIALS AND METHODS

Samples of genomic DNA were derived from 79 healthy individuals (blood donors) and 80 patients with MS. These patients all fulfilled the criteria of clinically definite MS [23]. In a previous study, ERV3 genotypes of most of these individuals (patients as well as blood donors) had been determined on the basis of polymorphisms in the 5' and 3' LTRs [17]. All individuals were European Caucasians.

The region under examination included the terminal portion of the *pol* gene and the entire *env* gene of ERV3 corresponding to a stretch of about 2700 nucleotides [2]. Fragments amplified from this region by means of polymerase chain reaction (PCR) were subjected to single strand conformational analysis (SSCP analysis) [14]. Amplifications were carried out with a panel of eleven different primer pairs producing overlapping segments of about 300 base pairs (Table 1).

PCR fragments suspected of containing polymorphic nucleotides were ligated into plasmids for subsequent sequencing.

Table 2
DNA polymorphisms in the *pol* and *env* gene of ERV3

Position*	Nucleotide substitution**
71 (<i>pol</i>)	T (leu) → C (leu)
458 (<i>pol</i>)	C (leu) → G (leu)
459 (<i>pol</i>)	C (arg) → T (trp)
478 (<i>pol</i>)	G (arg) → A (his)
956 (<i>env</i> , SU)	T (ile) → C (thr)
1262 (<i>env</i> , SU)	G (cys) → A (tyr)
1354 (<i>env</i> , SU)	C (arg) → T (stop)
1394 (<i>env</i> , SU)	A (tyr) → G (cys)
2129 (<i>env</i> , TM)	G (ser) → A (asn)
2393 (<i>env</i> , TM)	G (ser) → A (asn)
2436 (<i>env</i> , TM)	T (val) → C (val)
2534 (<i>env</i> , TM)	T (leu) → C (ser)
2666 (<i>env</i> , TM)	G (arg) → A (gln)

*The nucleotides are numbered in accordance with the published sequence [2]. SU and TM are abbreviations for envelope surface unit and transmembrane protein, respectively.

** In brackets amino acid encoded by the triplet containing the polymorphic nucleotide.

Alternatively, PCR fragments were sequenced without prior cloning. Sequencings were carried out with the Thermo Sequenase™ radiolabelled terminator cycle sequencing kit (Amersham Life Science Ltd, Little Chalfont, UK).

Haplotypes were defined on the basis of the linkage disequilibrium relationships between polymorphic sites. Genotypes of MS patients and healthy individuals were determined at polymorphic sites defining new haplotypes. Distribution of alleles in MS patients and healthy control subjects was compared by chi-square test.

RESULTS

A total of 13 polymorphic sites were found in the examined region of ERV3 (Table 2). Twelve of these nucleotide substitutions were transitions and one was a transversion. Four nucleotide substitutions, i.e. two synonymous and two substitutions with amino acid replacement, were located in the *pol* region. Of the remaining nine nucleotide substitutions, four were present in region of the *env* gene encoding the SU while

five were detected further downstream in the region encoding the TM protein. One of the nine nucleotide substitutions in the *env* gene was synonymous.

The C → T variation at position 1354 in the *env* gene of ERV3 is associated with a change of an arginine codon to a stop codon. The non-synonymous nucleotide substitution at position 2393 was located in the highly conserved and hydrophilic region of the transmembrane protein, encoded by the nucleotides between 2332 and 2409. The two nucleotide substitutions at 2534 and 2666 in the *env* gene were located downstream of the termination codon formed by nucleotides 2500–2502 [2].

Disregarding the nucleotide variation at the polymorphic sites, we found other differences from the published ERV3 sequence [2]. First, our sequence contained an additional G in between nucleotide 795 and 796. Second, we found the nucleotides AG at position 1338–1339 instead of GA. Third, our sequence analyses revealed a G and not an A at position 1364. Fourth, the nucleotide at position 1440 was a C according to our findings as opposed to a G in the published sequence. These findings were done with all fragments we sequenced.

The linkage disequilibrium relationships between alleles at the polymorphic sites in the *pol* and *env* genes allowed us to define five haplotypes or allelic forms of ERV3 (Figure 1). Taking the previously determined markers in the LTRs of this endogenous retrovirus into account, we found that two of the five haplotypes were new. They were defined on the basis of genetic variation at nucleotide position 1354 and 1394, respectively.

The C → T substitution at position 1354 creates a *Nla*III recognition sequence, facilitating determination of the two alleles at this polymorphic site. Based upon numbers of the three *Nla*III genotypes in MS patients and healthy control subjects, we calculated the corresponding allele frequencies (Table 3). The frequency of the *Nla*III-positive allele (i.e. the allele with the termination codon) amounted to 0.12 in MS patients and 0.09 in healthy individuals, implying that the *Nla*III negative allele appeared at

frequencies of 0.88 and 0.91 in these two groups. A statistical examination did not provide evidence of a significant difference in the distributions of *Nla*III alleles between MS patients and healthy individuals ($p = 0.47$). Subsequently, the frequencies of the alleles defined by the A \rightarrow G substitution at 1394 was examined. The allele with the G at that position appeared at a frequency of 0.05 in healthy individuals and 0.03 in MS patients.

DISCUSSION

A relatively high number of polymorphisms were detected in the region of ERV3 under examination. The actual number could even be higher since some polymorphisms might escape detection by SSCP [9]. However, the efficiency of this technique to detect polymorphisms is above 90% under most conditions [4], depending upon electrophoretic parameters [6] and size of PCR fragments [9]. Our findings of an excess of transitions are not unexpected since transitional

changes are known to occur at a higher frequency than transversal changes [11].

Genetic drift is a result of the random process which occurs in each generation when gametes unite. The effect of random genetic drift on alleles frequencies are important in small populations in which mutations more often become fixed than in large populations [8]. Probably, the formation of the five ERV3 haplotypes reported here reflects mutations in separate subpopulations in which some ERV3 alleles have been fixed while others have gone extinct.

An excess of synonymous substitutions over that of non-synonymous substitutions is often believed to reflect selective constraints acting to preserve a functionally important coding sequence [11]. In the present study only one of the seven base pair substitutions in the open reading frame of ERV3 encoding envelope proteins was synonymous, arguing against the assumption that this endogenous retrovirus has important although yet unknown functions [1,25]. The high number of non-synonymous nucleotide

Table 3
Distribution of *Nla*III genotypes of ERV3 in MS patients and healthy individuals*

	<i>Nla</i> III ⁻ / <i>Nla</i> III ⁻	<i>Nla</i> III ⁺ / <i>Nla</i> III ⁻	<i>Nla</i> III ⁺ / <i>Nla</i> III ⁺
MS patients	62	17	1
Healthy individuals	66	12	1

* *Nla*III was used to distinguish between the two alleles at nucleotide 1354 in the *env* region of ERV3. Individuals carrying the *Nla*III positive allele has a stop codon.

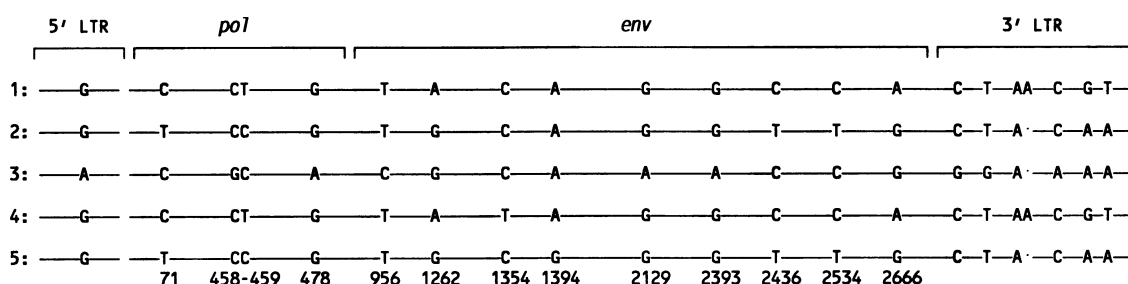


Fig. 1. Five different haplotypes of ERV3. The polymorphic nucleotides are indicated. Numbering of nucleotides is in accordance with the published sequence [2]. This sequence corresponds to haplotype 2. Most of the *pol* gene and the entire *gag* gene have not yet been screened for polymorphic nucleotides and are therefore not depicted. The polymorphisms in the LTRs were detected in previous study [17].

substitutions in ERV3 contrast with observations of another human endogenous retrovirus, namely HRES-1, which does not seem to carry any non-synonymous nucleotide substitutions in a coding region of about 600 nucleotides [20].

Besides the differences due to genetic variation, our sequence differed from the published ERV3 sequence with regard to other nucleotides. A possible explanation to these disagreements is that the published ERV3 sequence represents a haplotype different from those in the present study. Sequencing errors is another and perhaps more likely possibility.

The appearance of the two new haplotypes reflected genetic variation at nucleotide 1354 and 1394. The frequency of the allele at nucleotide 1354, encoding a truncated version of the *env* protein, was not significantly increased in the MS group in comparison with healthy individuals. It is noteworthy that two individuals, namely one patient and one healthy individual, were homozygous for the stop codon, suggesting that the envelope proteins of ERV3 do not perform absolutely vital functions. The haplotype defined on the basis of genetic variation of nucleotide 1394 was very rare in MS patients as well as healthy individuals. Taken together none of the alleles defined by the nucleotide substitutions at 1354 or 1394 appeared to be associated with MS.

Several of the nucleotide substitutions could have a significant impact on protein conformation, e.g. the nucleotide substitution at 1262 in the SU region of the *env* gene, changing an aliphatic amino acid (cysteine) to an aromatic amino acid (tyrosine). More importantly, the appearance of a stop codon at 1354 precludes expression of the TM protein. If this protein acts by downregulating immune responses at the placental barrier or in other tissues and organs, its absence due to an upstream stop codon could have serious consequences. The possible effect of the amino acid substitution in the conserved and hydrophilic region of the TM protein caused by the nucleotide substitution at 2393 is difficult to assess. In many retroviruses this region harbours an immunosuppressive motif [7]. However, the significance of the TM protein has not been adequately investigated in endogenous

retroviruses such as ERV3.

The relatively significant degree of individual variation in the *env* gene of ERV3 with many non-synonymous nucleotide substitutions and the presence of a stop codon in some individuals could be of importance to future studies of this endogenous retrovirus. For example, some ERV3 peptides might not be reliable for detection of autoantibodies in all individuals. Our findings are also of relevance to production of antibodies against ERV3, in particular monoclonal antibodies and antibodies produced with short synthetic peptides as antigens.

Still the significance and possible functions of ERV3 are not known. We did not detect an association between ERV3 and MS. Whether there is an association between this endogenous retrovirus and other diseases such as congenital heart block is a possibility which can not be dismissed. The haplotyping system defined in this study could be valuable for detection of such associations.

Note added in proof

We have been informed that another study of DNA polymorphisms in ERV3 recently was published by Nathalie de Parseval and Thierry Heidmann: Physiological knockout of the envelope gene of the single-copy ERV-3 human endogenous retrovirus in a fraction of the Caucasian population. *J. Virol.* **72**, (1998) 3442–3445.

Acknowledgments

The John and Birthe Meyer Foundation is thanked for support. This study was part of the retrovirus research programme of the Danish Multiple Sclerosis Society.

References

- [1] Boyd, M.T., Bax, C.M.R., Bax, B.E., Bloxam, D.L. and Weiss, R.A. The human endogenous retrovirus ERV-3 is upregulated in differentiating

- placental trophoblast cells. *Virology* **196**, (1993) 905–909.
- [2] Cohen, M., Powers, M., O'Connell, C. and Kato, N. The nucleotide sequence of the *env* gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. *Virology* **147**, (1985) 449–458.
- [3] Cohen, M., Kato, N. and Larsson, E. ERV3 human endogenous provirus mRNAs are expressed in normal and malignant tissues and cells, but not in choriocarcinoma tumor cells. *J. Cell. Biochem.* **36**, (1988) 121–128.
- [4] Fan, E., Levin, D.B., Glickman, B.W. and Logan, D.M. Limitations in the use of SSCP analysis. *Mutat. Res.* **288**, (1993) 85–92.
- [5] Foulds, S., Wakefield, C.H., Giles, M., Gillespie, J., Dye, J.F. and Guillou, P.J. Expression of a suppressive p15E-related epitope in colorectal and gastric cancer. *Br. J. Cancer* **68**, (1993) 610–616.
- [6] Glavac, D. and Dean, M. Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Hum. Mutat.* **2**, (1993) 404–414.
- [7] Haraguchi, S., Good, R.A., Cianciolo, G.J., Engelman, R.W. and Day, N.K. Immunosuppressive retroviral peptides: immunopathological implications for immunosuppressive influences of retroviral infections. *J. Leukocyte Biol.* **61**, (1977) 654–666.
- [8] Hartl, D.L. and Clark, A.G. *Principles of population genetics*. Sunderland, MA, USA: Sinauer Associates, Inc. (1989).
- [9] Hayashi, K. and Yandell, D.W. How sensitive is PCR-SSCP? *Hum. Mutat.* **2**, (1993) 338–346.
- [10] Kato, N., Shimotohno, K., VanLeeuwen, D. and Cohen, M. Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger protein related to Krüppel. *Mol. Cell. Biol.* **10**, (1990) 4401–4405.
- [11] Li, W.-S. and Graur, D. *Fundamentals of molecular evolution*. Sunderland, MA, USA: Sinauer Associates, Inc. (1991).
- [12] Li, J.-M., Fan, W.S., Horsfall, A.C., Anderson, A.C., Rigby, S., Larsson, E. and Venables, P.J.W. The expression of human endogenous retrovirus-3 in fetal cardiac tissue and antibodies in congenital heart block. *Clin. Exp. Immunol.* **104**, (1996) 388–393.
- [13] Löwer, R., Löwer, J. and Kurth, R. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **93**, (1996) 5177–5184.
- [14] Mashiyama, S., Sekiya, T. and Hayashi, K. Screening of multiple DNA samples for detection of sequence changes. *Technique* **2**, (1990) 304–306.
- [15] Matsuda, T., Sasaki, M., Kato, H., Yamada, H., Cohen, M., Barrett, J.C., Oshimura, M. and Wake, N. Human chromosome 7 carries a putative tumor suppressor gene(s) involved in choriocarcinoma. *Oncogene* **15**, (1997) 2773–2781.
- [16] O'Connell, C., O'Brien, S., Nash, W.G. and Cohen, M. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. *Virology* **138**, (1984) 225–235.
- [17] Rasmussen, H.B., Heltberg, A., Lisby, G. and Clausen, J. Three allelic forms of the human endogenous retrovirus, ERV3, and their frequencies in multiple sclerosis patients and healthy individuals. *Autoimmunity* **23**, (1996) 111–117.
- [18] Rasmussen, H.B. Interactions between exogenous and endogenous retroviruses. *J. Biomed. Sci.* **4**, (1997) 1–8.
- [19] Rasmussen, H.B. and Clausen, J. Possible involvement of endogenous retroviruses in the development of autoimmune disorder, especially multiple sclerosis. *Acta Neurol. Scand.* **169** (suppl.), (1997) 32–37.
- [20] Rasmussen, H.B. and Clausen, J. A novel haplotype of the endogenous retrovirus, HRES-1, in patients with multiple sclerosis and healthy individuals. *Autoimmunity* (in press).
- [21] Rubin, L.A., Siminovitch, K.A., Shi, M.H. and Cohen, M. A novel retroviral gene association with rheumatoid arthritis. *Arthritis Rheum.* **34** (suppl.), (1991) S60.
- [22] Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dölken, G., Lemm, I., Buck, M., Best, B., Neumann-Haefelin, D. and Mueller-Lantzsch, N. Human endogenous retrovirus K10: expression of gag protein and detection of antibodies in patients with seminomas. *J. Virol.* **69**, (1995) 414–421.
- [23] Schumacher, G.A., Beebe, G., Kibler, R.F., Kurland, L.T., Kurtzke, J.F., McDowell, F., Nagler, B., Sibley, W.A., Tourtelotte, W.W. and Willmon, T.L. Problems of experimental trials of therapy in multiple sclerosis: report by the panel on the evaluation of experimental trials of

- therapy in multiple sclerosis. *Ann. N.Y. Acad. Sci.* **122**, (1965) 552–568.
- [24] Takeuchi, K., Katsumata, K., Ikeda, H., Minami, M., Wakisaka, A. and Yoshiki, T. Expression of endogenous retroviruses, ERV3 and lambda 4-1, in synovial tissues from patients with rheumatoid arthritis. *Clin. Exp. Immunol.* **99**, (1995) 338–344.
- [25] Venables, P.J.W., Brookes, S.M., Griffiths, D., Weiss, R.A. and Boyd, M.T. Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. *Virology* **211**, (1995) 589–592.

