#### SHORT COMMUNICATION

Shiga Hasuike • Kiyonori Miura • Osamu Miyoshi Toshinobu Miyamoto • Norio Niikawa • Yoshihiro Jinno Mutuo Ishikawa

# Isolation and localization of an IDDMK1,2-22-related human endogenous retroviral gene, and identification of a CA repeat marker at its locus

Received: March 29, 1999 / Accepted: May 21, 1999

**Abstract** We intended to confirm genetically the involvement of the IDDMK1,2-22 gene in the pathogenesis of insulin-dependent diabetes mellitus (IDDM). For this purpose, we isolated a human endogenous retrovirus gene, possibly corresponding to IDDMK1,2-22. The isolated gene showed 99.8% and 99.7% homologies in nucleotide sequences to a part of the *env* region and of the 3'-LTR region, respectively, compared to those of IDDMK1,2-22 deposited in GenBank. The gene also showed a close relation to *HERV-K18*, of which the 3'-LTR sequence gave 99.5% homology. It seemed likely that these genes represented the same single gene. The newly isolated gene was present in the first intron of the *CD48* gene and was located on chromosome 1q21.2–q22. A CA repeat marker was found approximately 20kb upstream from the 5'-end of the 5'-LTR of the gene.

**Key words** Insulin-dependent diabetes mellitus · Gene isolation · Chromosomal localization · CA repeat marker · IDDMK1, 2-22 · HERV-K18

# Introduction

Insulin-dependent diabetes mellitus (IDDM) is an autoimmune disease characterized by T-lymphocyte-dependent destruction of the insulin-producing  $\beta$  cells in the pancreas.

e-mail: sirius@med.u-ryukyu.ac.jp

K. Miura

Department of Obstetrics and Gynecology, Nagasaki University School of Medicine, Nagasaki, Japan

S. Hasuike · T. Miyamoto · M. Ishikawa Department of Obstetrics and Gynecology, Asahikawa Medical College, Asahikawa, Japan

As indicated by the 30%–50% concordance in monozygotic twins, its inheritance is affected and complicated by environmental factors as well as by genes. The incidence of IDDM shows seasonal and geographic variations. Biased male-to-female ratios in patients are also observed (Karvonen et al. 1997; Cucca et al. 1998). In addition, differential parent-to-offspring transmission of genes predisposed to diabetes has been observed in several studies, suggesting a possible role for genomic imprinting (Vadheim et al. 1986; Julier et al. 1991). Genome-wide scans of affected families have detected more than 10 susceptibility loci with IDDM1/MHC being the major genetic component (Todd 1995). Conrad et al. (1994) suggested the involvement of a superantigen in IDDM, and recently isolated a novel human endogenous retrovirus named as IDDMK1,2-22 as the causative superantigen from the supernatants of cultured islet cells of two IDDM patients (Conrad et al. 1997).

Human endogenous retroviruses (HERVs) are multiple DNA sequences closely related to infectious retroviruses and compose an integral part of the human genome. The human genome contains several families of HERV elements, ranging in copy number from 1 to 100 to  $10^4$  (solitary long terminal repeat, LTR). Full length HERVs show the same basic genome organization as exogenous infectious retroviruses, possessing regions with sequence similarity to the LTRs and major open reading frames (ORFs) of retroviruses, namely gag, pol, and env. Active HERV transcription has been observed in various normal tissues, including peripheral blood mononuclear cells and the placenta, as well as in various tumors. Although most HERV loci are translationally defective because of random mutations, it is now clear that complete ORFs exist in several HERV families (reviewed by Patience et al. 1997).

Although the involvement of HERVs in autoimmune diseases has been suggested, a clear association with a disease remains to be shown. In addition, their biological significance is not well understood. While DNA methylation is thought to play an important role to control allele-specific expression of imprinted genes, Yoder et al. (1997) argued

S. Hasuike · K. Miura · O. Miyoshi · N. Niikawa · Y. Jinno<sup>1</sup> (⊠) Department of Human Genetics, Nagasaki University School of Medicine, Nagasaki, Japan Tel. +81-98-895-3331 (ext. 2783); Fax +81-98-895-5782

Present affiliation and address

<sup>&</sup>lt;sup>1</sup>Department of Molecular Biology, Ryukyu University School of Medicine, Nishihara, Okinawa 903-0215, Japan

that suppression of parasitic sequence elements was the primary function of cytosine methylation. As mentioned, the inheritance in IDDM seems to suggest a role for genomic imprinting. Thus, we initiated isolation of a genomic sequence corresponding to IDDMK1,2-22 to examine the causative link of IDDMK1,2-22 to IDDM and to delineate the imprinting mechanism by investigating the controlling mechanism of HERVs.

# **Materials and methods**

# PCR amplification of HERV-K10-related sequences

Based on comparison of HERV-K10 (accession no. M14123) and IDDMK1,2-22 (accession no. AF012336) sequences, four sets of primers were prepared from four regions that showed suitable restriction site differences (RSD1 through RSD4) to discriminate the sequences with the sites of IDDMK1,2-22 from other HERV-K10-related sequences. PCR was carried out in a 50-µl reaction volume containing the standard reaction buffer (1.5mM MgCl<sub>2</sub>/ 50mMKCl/10mM Tris-Cl, pH 8.3/200µM each dNTP/1µM each primer/1.5U Taq polymerase) and 0.2µg genomic DNA as template for 25 cycles at 94°/52°/72°C for 30/30/30s. The gag and LTR regions were also amplified to examine the presence or absence of their sequences in the isolated cosmid clones. Primer sequences are as follows: RSD1A/K10Pa, 5'-GTAAGCGGGATGTCACTCAG; RSD1B, 5'-ACCGCACTATTGGCCACACA; RSD2A, 5'-TTACTGTGGCCTCACACCA; RSD2B, 5'-TCACC-AGCAGAATACGGTG; RSD3A, 5'-CACCGTATTCT-GCTGGTGAG; RSD3B, 5'-CTGAACAGAAGAGTGC-AATG; RSD4A, 5'-GCCCATTTAAATTTGGTGCCA; RSD4B, 5'-GTCATCATGGCCCGTTCTC; GAGA, 5'-CCGGTAATGGCTCAGTCAAC; GAGB, 5'-GGCA-GATAGCTCTAACTTGC; LTRA, 5'-ATGTGCTGTGT-CAACTCAGG; LTRB, 5'-ATGTGTCAGGGTCACAA-GAC.

#### Cosmid library screening

A human placental genomic cosmid library was obtained from Clontech (Palo Alto, CA, USA). A total of  $6.5 \times 10^5$ cosmid clones was screened with a 20-*mer*-oligonucleotide probe (OPDV). Hybridization was carried out overnight at 55°C in a solution containing 6× SSC/5× Denhardt's solution/0.01 M sodium phosphate/1 mM EDTA/0.5% SDS. The membranes were washed three times with 6× SSC/0.1% SDS at room temperature for 20min and finally once at 55°C for 15min (OPDV, 5'-ACAAGATATCGTAAGCC-ATT).

#### Southern blot analysis

Peripheral blood or placental DNA,  $30\mu g$  each, was digested with  $5-10 U/\mu g$  DNA of restriction enzymes at  $37^{\circ}C$ 

overnight, fractionated on a 0.8% agarose gel, and transferred to a nylon membrane. The membrane was hybridized with an oligonucleotide probe OPDV or SSDM. Hybridization and washing were carried out similarly as in the cosmid library screening (SSDM, 5'-TTCCATTCTAACGGTT-CCT).

### Sequencing analysis

The insert DNA of the cosmid clone no. 14 was variously subcloned into pUC19 vector based on the restriction map of the cosmid. DNA sequences were determined using an ALF DNA sequencer (Pharmacia Biotech, Piscataway, NJ, USA) with AutoRead Sequencing kits and AutoCycle sequencing kits. Sequencings were read on multiple reactions and runs in both directions in the *env* and 3'-LTR regions.

### Chromosomal localization

Fluorescent in situ hybridization (FISH) was performed as described previously (Ohta et al. 1993);  $0.4\mu$ g of the cosmid DNA labeled with biotin-16-dUTP (Boehringer Mannheim, Tokyo, Japan) was hybridized to human Rbanded metaphase chromosomes in 20µl of the hybridization mixture containing 50% formamide, 2× SSC, 1% dextran sulfate, bovine serum albumin (BSA), and 2µg human Cot-1 DNA (GIBCO-BRL, Rockville, MD, USA). After hybridization, the slides were washed in 50% formamide/2× SSC for 13 min at 37°C.

Radiation hybrid mapping was performed using the Stanford G3 panel (Research Genetics, Huntsville, AL, USA). PCR was carried out in  $10\mu$ l standard reaction buffer containing  $1.5\mu$ l of hybrid DNA for 35 cycles at 94°/ 50°/30°C for 30/30/30s. Primers were as follows: BBGF2, 5'-ACCTCTGTATATGTGAGGTG; BBGR2, 5'-GGATG-AAGAAAGAGGGTGAG.

CA repeat isolation and polymorphism analysis

A CA repeat sequence was found in a cosmid clone flanking the original clone no. 14 by Southern blot hybridization probed with the Poly(dA-dC)·Poly(dG-dT) (Pharmacia) and confirmed by sequencing. Polymorphism was detected in 60 chromosomes of unrelated Japanese individuals by 35 cycles of PCR at 94°/55°/72°C for 30/30/30s. Primer sequences were: CDCA5, 5'-CTGCTTGTTTATCCCATT-CGC; CDCA3, 5'-CCCTCACTTTTCTCAGTGGA.

### Results

Isolation and characterization of an endogenous retrovirus gene corresponding to IDDMK1,2-22

A retroviral sequence closely related to that of the human endogenous retrovirus K10 (*HERV-K10*) gene was isolated

by RT-PCR from IDDM patients (Conrad et al. 1997). We first examined the possibility that the corresponding gene was present in the human genome. From sequence comparison between IDDMK1,2-22 and HERV-K10, we chose four regions that showed restriction site differences (RSDs 1-4) and prepared primer sets to amplify the related sequences (Fig. 1). Expected restriction products were obtained from all PCR products in the four RSD regions. Among them, *Eco*RV digestion of the PCR product in the RSD2 gave the least amount of the expected band relative to other restriction products (data not shown). We further inspected this possibility by sequencing cloned DNA of the expected band of the EcoRV digest. Among 30 clones sequenced, 9 clones showed 100% identity to IDDMK1,2-22 through the full length of the inserts with 200 bp. Thus, we concluded that the IDDMK1,2-22 gene was possibly present in the human genome.

Based on comparison of sequences between IDDMK1,2-22, *HERV-K10*, and the cloned DNA, an oligonucleotide probe (OPDV), which differed from the *HERV-K10* sequence in 2 of 20 nucleotides, was prepared to screen a

cosmid library. This probe detected a band with the expected sizes and a fainter extra band in genomic Southern blot hybridization (data not shown). Among  $6.5 \times 10^5$ clones screened, four clones were isolated. When they were examined at the restriction sites in the RSD regions by restriction of the PCR products, three clones showed the same restriction patterns as expected in the IDDMK1,2-22, sequence (Table 1). One clone, which had three more restriction sites otherwise absent in IDDMK1,2-22, carried a truncated endogenous retroviral sequence devoid of the 5'-LTR and gag regions. Cosmid clone no. 14 gave positive bands with expected sizes in various restriction digestions by Southern blot hybridization (data not shown). Thus, clone no. 14 seemed to carry the gene corresponding to IDDMK1,2-22. Restriction mapping of the cosmid clones revealed that clones no. 213 and 223 were overlapping clones of clone no. 14 (Fig. 2).

Clone no. 14 was further examined for sequencing analysis and chromosome mapping. Sequences in a part of the *env* region (the SAG region) and the 3'-LTR were carefully determined to compare with those of IDDMK1,2-22 depos-

RSD 2 RSD 3

**Fig. 1.** Restriction site differences (*RSD*) between the *HERV-K10* and IDDNK1,2-22 sequences in the *env* region. Restriction enzymes and their locations are shown on and *below the line*. Primers are indicated by *arrows* 



Fig. 2. Restriction map of cosmid clones. Restriction sites are indicated by *vertical lines* on the *horizontal lines* that represent the genomic sequences. Restriction sites depicted below the line are a partial map made on a plasmid subclone, which were demonstrated to compare with the map of *HERV-K18*. The endogenous retrovirus gene is shown by *boxes*. *Hatched boxes* are the LTRs; the first exon of *CD48* is indicated by a *filled box*. *E*, *Eco*RI; *H*, *Hind*III; *K*, *Kpn*I; *N*, *Not*I; *P*, *Pst*I; *S*, *Sal*I

Table 1. Characterization of cosmid clones by PCR amplification and restriction of the PCR produc
---

		2	1		1					
	RSD1		RSD2		RSD3		RSD4			
	MboI	HinfI	EcoRV	Sa11	NspI	HphI	ScaI	ClaI	GAG	LTR
IDDMK1,2-22	+	_	+	_	+	_	+	_	(+)	+
Cosmid clone 14	+	_	+	_	+	_	+	_	+	+
Cosmid clone 223	+	_	+	_	+	_	+	_	+	+
Cosmid clone 219	+	+	+	_	+	+	+	+	_	+
Cosmid clone 213	+	-	+	-	+	-	+	-	+	+

RSD 1

+ or -, the presence or absence of the restriction sites (or the PCR products for GAG and LTR)

RSD 4

ited in GenBank (accession nos. AF012337, AF012336, and AF012335). Of a total of 2547 nucleotides (nt) in the regions (accession no. AF134984), inconsistencies were found in 7nt (99.7% homology; 99.76% in the SAG region and 99.65% in the 3'-LTR region). The amino acid sequence deduced from the nucleotide sequence showed 100% identity to that presented as the SAG sequence (Conrad et al. 1997). One of the discrepancies was inspected by Southern blot analysis of genomic DNA restricted with BspHI, the recognition sequence of which was present in the cosmid sequence and absent in the IDDMK1,2-22 sequence. Another oligonucleotide probe, SSDM, detected only the expected single fragment in various restriction digests except the RsaI digest, and BspHI completely cleaved all these fragments (data not shown). This result indicates that either the difference in the BspHI recognition sequence represents a polymorphism or the IDDMK1,2-22 sequence is not present in the human genome. The RsaI fragment detected in the genomic DNA was larger than that in the cosmid and that expected from both sequence data; this is likely to result from a polymorphism. Another closely related sequence was HERV-K18, which showed 99.5% homology (864/868nt) in the 3'-LTR sequence (accession no. M12852). In addition, the restriction maps in the outside region downstream to the 3'-LTR seemed to be consistent with each other as well as to that in the HERV gene (Ono 1986). Thus, it is highly probable that the HERV sequence carried in cosmid clone no. 14 represents the HERV-K18 gene, although the primary sequence data are available only in the LTR sequences of HERV-K18. It is also conceivable that the three sequences represent the same gene.

In a preliminary sequencing analysis, it was incidentally revealed that the *HERV* gene isolated in the present work is located in the first intron of the *CD48* gene in the opposite direction.

#### Chromosome mapping

Localization of the newly isolated *HERV* gene was determined by FISH and the Stanford radiation hybrid cell panel. As cosmid clone no. 14 could not give a definite signal on a single locus, a walking clone extending downstream to the original clone was obtained. This walking clone yielded a signal on chromosome 1q21.3-q22 (Fig. 3). Among 24 metaphase cell spreads with 39 signals, 17 showed a signal on each of the two chromatids at the locus. Radiation hybrid mapping placed the gene just at the marker SHGC-30228 locus (LOD = 1 000; 0.00 cR<sub>10000</sub>).

# CA repeat isolation

A CA repeat marker was found in another walking clone extending upstream to the original clone no. 14 approximately 20kb distant from the 5'-end of the 5'-LTR. Four alleles were detected in 60 chromosomes of unrelated Japanese individuals. The observed heterozygosity was 0.68. The size and frequency of the four alleles are shown in



**Fig. 3.** Chromosomal localization of the newly isolated endogenous retrovirus gene by FISH. *Arrows* indicate the hybridization siganals on human chromosome 1q21.3-q22

**Table 2.** Allele frequencies of the CA repeat marker locus

Allele	Size (bp)	Frequency		
1	119	0.09		
2	117	0.16		
3	111	0.40		
4	109	0.36		

Table 2. Codominant inheritance was observed in four nuclear families.

# Discussion

We obtained a human endogenous retrovirus gene in careful efforts to isolate the genomic sequence corresponding to IDDMK1,2-22 that was isolated from the supernatants of pancreatic islets cell cultures of two IDDM patients by multistep procedures of RT-PCR (Conrad et al. 1997). A thorough screening of  $6.5 \times 10^5$  cosmid clones essentially resulted in the isolation of only one clone with the complete structure of the *HERV* gene. Sequence analyses of the gene revealed a high homology but not 100% identity to the IDDMK1,2-22 sequence. Among 2,547 nucleotides compared, 7 nucleotides (0.3%) did not match. On the other hand, the gene also showed similar high homology to *HERV-K18*. Although only the 3'-LTR sequence was available to be compared, the homology reached 99.5%. In addition, a partial restriction map on the unique sequence outside the repetitious gene was very consistent as well as that on the gene body. Thus, it is likely that the newly isolated gene and *HERV-K18* are the same gene. The mismatches of 0.5% between the two genes may have resulted from polymorphisms and sequencing errors. Therefore, it is also possible that all three sequences (our HERV, HERV-K18, and IDDMK1,2-22) are the representatives of the same single gene.

While the involvement of IDDMK1,2-22 in the pathogenesis of type 1 diabetes is controversial (Lan et al. 1998; Lower et al. 1998; Murphy et al. 1998; Muir et al. 1999), it is interesting that the gene was mapped to chromosome 1q21.3-q22 regardless of whether it is the genomic sequence corresponding to IDDMK1,2-22. Recent study in IDDM families from the United Kingdom detected several regions with more than 1.5 of maximum lod score (MLS) that have not been reported previously (Mein et al. 1998). The chromosome region 1q12-q24 is such an example. We identified a CA repeat marker at the newly isolated gene locus that would contribute to determining the involvement of this endogenous retrovirus gene in the pathogenesis of IDDM.

It is also intriguing that the gene was placed in the first intron of the CD48 gene. CD48 is a glycosylphosphatidylinositol-linked cell-surface molecule. It has two immunoglobulin-like domains and is one of the ligands for CD2. Although its physiological roles are not known in detail, mice homozygous for the CD48 mutation are severely impaired in CD4<sup>+</sup> T-cell activation (Gonzalez-Cabrero et al. 1999). CD4 has been linked and associated with type 1 diabetes in Danish families (Kristiansen et al. 1998). The mouse Cd2 gene is localized to a region on chromosome 3 where one of the type 1 diabetes loci (Idd10) has been mapped (Vyse and Todd 1996). Its human syntenic region, 1p13, is encompassed in the 1p21-p12 region detected at MLS = 1.8 as susceptibility loci to IDDM (Mein et al. 1998). Although Conrad et al. hypothesized that an endogenous retroviral superantigen (SAG) triggers polyclonal activation of the V $\beta$ -restricted T-cell subset leading to initiation of organ-specific tissue destruction by autoreactive T cells within the subset of SAG-activated T lymphocytes, it is also conceivable that CD molecules themselves are causatively involved in the pathogenesis of IDDM. In the latter case, *HERV* could affect expression of the CD genes through its regulatory sequences in the LTR or trans-acting products of the gene.

Acknowledgments We appreciate Dr. Jacobs for her help during manuscript preparation. The present study was supported by Research on Human Genome and Gene Therapy from the Ministry of Health and Welfare of Japan.

# References

- Conrad B, Weidmann E, Trucco G, Rudert WA, Behboo R, Ricordi C, Roddrquez-Rilo H, Finegold D, Trucco M (1994) Evidence for superantigen involvement in insulin-dependent diabetes mellitus aetiology. Nature 371: 351–355
- Conrad B, Weissmahr RN, Boni J, Arcari R, Schupbach J, Mach B (1997) A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. Cell 90: 303–313
- Cucca F, Goy JV, Kawaguchi Y, Esposito L, Merriman ME, Wilson AJ, Cordel HJ, Bain SC, Todd JA (1998) A female-male bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. Nat Genet 19: 301–302
- Gonzalez-Cabrero J, Wise CJ, Latchman Y, Freeman GJ, Sharpe AH, Reiser H (1999) CD48-deficient mice have a pronounced defect in CD4(+) T cell activation. Proc Natl Acad Sci USA 96: 1019–1023
- Julier C, Hyer RN, Merlin F, Soularue P, Briant L, Cathelineau G, Deshamps I, Rotter JI, Froguel P, Boitard C, Bell JI, Lathrop GM (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. Nature 354: 155–159
- Karvonen M, Pitkaniemi J, Kohtamaki K, Tajima N, Tuomilehto J (1997) Sex difference in the incidence of insulin-dependent diabetes mellitus: an analysis of the recent epidemiological data. Diabetes Metab Rev 13: 275–291
- Kristiansen OP, Zamani M, Johannesen J, Mandrup-Poulsen T, Cassiman JJ, Pociot F (1998) Linkage and association between a CD4 gene polymorphism and IDDM in Danish IDDM patients. Diabetes 47: 281–283
- Lan MS, Mason A, Coutant R, Chen Q-Y, Vargas A, Rao J, Gomez R, Chalew S, Garry R, Maclaren NK (1998) HERV-K10s and immunemediated (type 1) diabetes. Cell 95: 14–16
- Lower R, Tonjes RR, Boller K, Denner J, Kaiser B, Phelps RC, Lower J, Kurth R, Badenhoop K, Donner H, Usadel KH, Miethke T, Laoatschek M, Wagner H (1998) Development of insulin-dependent diabetes mellitus does not depend on specific expression of the human endogenous retrovirus HERV-K. Cell 95: 11–14
- Mein CA, Esposito L, Dunn MG, Johnson GCL, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ, Pritchard LE, Cucca F, Barnett AH, Bain SC, Todd JA (1998) A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nat Genet 19: 297–300
- Muir A, Ruan QG, Marron MP, She JX (1999) The IDDMK(1,2)22 retrovirus is not detectable in either mRNA or genomic DNA from patients with type 1 diabetes. Diabetes 48: 219–222
- Murphy VJ, Harrison LC, Rudert WA, Luppi P, Trucco M, Fierabracci A, Biro PA, Bottazzo GF (1998) Retroviral superantigens and type 1 diabetes mellitus. Cell 95: 9–11
- Ohta T, Tohma T, Soejima H, Fukushima Y, Yoshiura K, Jinno Y, Niikawa N (1993) The origin of cytologically unidentifiable chromosome abnormalities: six cases ascertained by targeted chromosomeband painting. Hum Genet 92: 1–5
- Ono M (1986) Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to type A and B retrovirus genes. J Virol 58: 937–944
- Patience C, Wilkinson DA, Weiss RA (1997) Our retroviral heritage. Trends Genet 13: 116–120
- Todd JA (1995) Genetic analysis of type 1 diabetes using whole genome approaches. Proc Natl Acad Sci USA 92: 8560–8565
- Vadheim CM, Rotter JI, Maclaren NK, Riley WJ, Anderson CE (1986) Preferential transmission of diabetic alleles within the HLA gene complex. N Engl J Med 20: 1314–1318
- Vyse TJ, Todd JA (1996) Genetic analysis of autoimmune disease. Cell 85: 311–318
- Yoder JA, Walsh CM, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13: 335–340