Diabetologia (2000) 43: 1429–1434

Diabetologia

Genomic structure of mouse *IA-2:* comparison with its human homologue

K. Saeki, J. Xie, A. L. Notkins

Experimental Medicine Section, Oral Infection and Immunity Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA

Abstract

Aims/hypothesis. IA-2 is a transmembrane protein with a tyrosine phosphatase (PTP)-like structure and a major autoantigen in Type I (insulin-dependent) diabetes mellitus. Because the nucleotide sequence of human and mouse *IA-2* cDNA are closely related, it seemed likely that the genomic organization of the two molecules would be similar. To test this possibility the current experiments were initiated to characterize and compare the genomic structure of mouse and human *IA-2*.
Methods. *IA-2* cDNA was used to screen a 129SVJ mouse genomic library. We selected and mapped 7 overlapping clones. The subcloned inserts were used to determine intron-exon junctions by direct sequencing. Polymerase chain reaction and restriction mapping were used to estimate the size of the introns.
Results. The mouse *IA-2* gene and the 5' upstream regulatory region were isolated and the intron-exon junctions determined. Mouse IA-2 encompasses approximately 20 kb and encodes 23 exons. Both the 3' and 5' ends were mapped by rapid amplification of cDNA ends (RACE) and a 2 kb 5'-upstream region was shown to have functional promoter activity.
Conclusion/interpretation. Comparison of the genomic structure of mouse and human *IA-2* shows that they have the same number of exons and nearly identical intron-exon junctions. The region around the major transcription start site of mouse *IA-2* is similar to human *IA-2* and other transmembrane protein tyrosine phosphatases. It is concluded that human and mouse *IA-2* are highly conserved and derived from a common ancestral gene. [Diabetologia (2000) 43: 1429–1434]

Keywords *IA-2*, insulin-dependent diabetes mellitus, autoantibodies, genomic structure, protein tyrosine phosphatase, intron-exon junctions.

The protein tyrosine phosphatase-like molecule IA-2 is a new member of the transmembrane protein tyrosine phosphatase (PTP) family [1,2]. It is distantly related to other transmembrane PTPs with 35–40 % sequence similarity in the intracellular domain. It is also closely related to another new member of the transmembrane PTP family, IA-2β, with 74 % identity in the intracellular domain [3,4]. Human IA-2 and mouse IA-2 are, however, very closely related to each other showing 98 % identity in the intracellular domain and 92 % identity in the extracellular domain [5]. Although IA-2 is a member of the PTP family based on sequence analysis, it lacks enzymatic activity with conventional PTP substrates [2].

Over the last few years IA-2 has generated considerable interest because it was found to be a major autoantigen in Type I (insulin-dependent) diabetes mellitus [6–12]. Approximately 70 % of pateints with newly diagnosed Type I diabetes have autoantibodies to IA-2. These autoantibodies appear years before the development of clinical disease and, thus, serve

Received: 25 April 2000 and in revised form: 27 July 2000

Corresponding author: A. L. Notkins, Experimental Medicine Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Building 30, Room 121, 30 Convent Drive MSC 4322, Bethesda, MD 20892–4322, USA
Abbreviations: PTP, protein tyrosine phosphatase; RACE, rapid amplification of cDNA ends; IA-2, protein tyrosine phosphatase-like molecule.

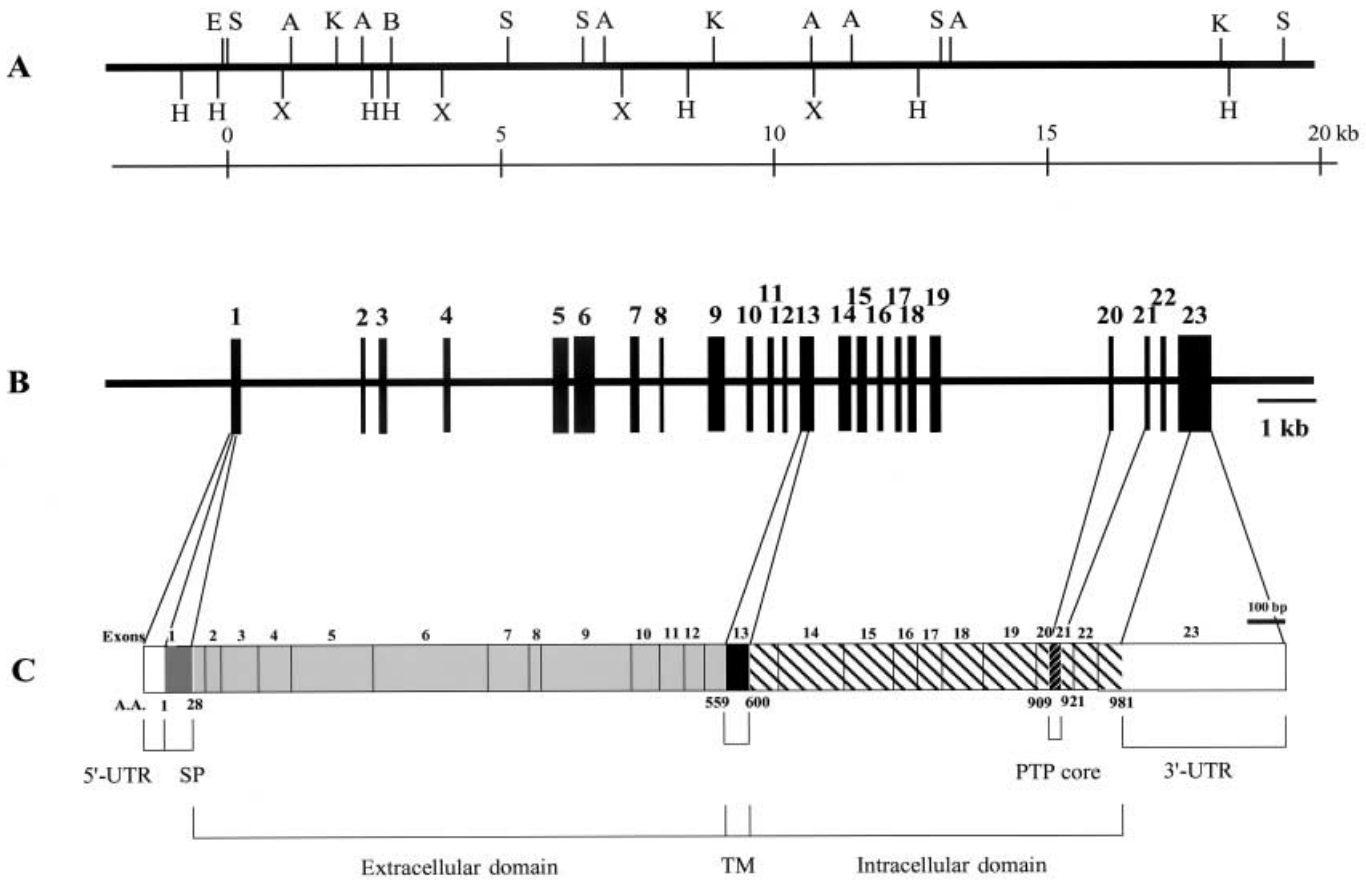

Fig. 1 A–C. Schematic representation of the genomic structure of mouse *IA-2* (*mIA-2*). **A** Restriction map of *mIA-2*. Seven genomic clones were used to characterize the gene: *Apa* I (A); *Bam* HI (B); *Eco* RI (E); *Hin*d III (H); *Kpn* I (K); *Sac* I (S); *Xho* I (X). **B** Genomic structure of m*IA-2*. Approximate position and relative size of exons (vertical bars) and introns (horizontal lines). Number of each exon is indicated. **C** mIA-2 protein. Protein regions encoded by exons are indicated. The open bars in exons 1 and 23 represent 5' and 3' untranslated regions (UTR), respectively. Transmembrane (TM)

as predictive markers for identifying people at high risk of becoming diabetic [10]. The function of IA-2 is still not known. For a better understanding of this molecule, with the long-term goal of knocking out the gene, knowledge of its genomic structure is essential. We report the complete genomic structure of mouse *IA-2* and compare it with the recently reported genomic structure of human *IA-2* [13] (GenBank accession numbers L18983 and AF042285).

Materials and methods

Screening of genomic libraries. A 129SVJ mouse genomic library, constructed in lambda FIX II vector (Stratagene, La Jolla, Calif., USA), was screened by enzyme-derived chemillkuminescence non-radioactive plaque hybridization and detection method (Amersham, Arlington Heights, Ill., USA). Positive clones were confirmed by restriction enzyme mapping and Southern blot hybridization with a mouse *IA-2* cDNA probe.

Restriction enzyme mapping. Phage DNA was prepared from culture lysates with a Wizard Lamda DNA preparation kit (Promega, Madison, Wis., USA). We selected and mapped seven overlapping clones by the restriction enzymes *Apa* I, *BamH* I, *EcoR* I, *Hind* III, *Sac* I, *Xho* I and *Kpn* I with a FLASH non-radioactive gel mapping kit (Stratagene, La Jolla, Calif., USA). The position of the restriction enzyme sites was determined by mapping from both ends and confirmed by analysis of the band pattern from complete digestion.

Intron-exon boundaries. Cloned genomic DNA lambda phage was digested by different restriction enzymes, separated on agarose gels and Southern blotted with cDNA probes. Exon-containing fragments were subcloned into a pBluescript II vector (Stratagene). Nested deletions of cloned inserts were prepared by exonuclease digestion, the religated plasmids were amplified and used as templates for sequencing. Each exon-intron junction was determined by sequence analysis and compared with mouse *IA-2* cDNA (GenBank accession number Ul1872). The size of introns was estimated by the sequence data or calculated by the distance from restriction enzyme sites.

Mapping 3' and 5' ends of mRNAs. The 3' cDNA end and 5' transcription start sites were determined with RNA prepared from mouse β-TC-1 cells. Total RNA was extracted from these cells [14] and then subjected to reverse transcription with (dT) 17-adaptor primer (3' mapping) or mIA-2 anti-sense primer

Table 1. Exon-intron junctions of mouse IA-2

Exon No.	Exon Size (bp)	Intron No.	Intron Size (bp)	Sequence at exon-intron junctions												Amino acid position
				5'-Splice donor						3'-Splice acceptor						
				S	A	H					G	C	L	F		**42**
1	~184	1	~2100	AGT	GCC	CAC	G	**gt**caggagaaacttgatggg	.. ..	gagaatcttctctttcac**ag**	GC	TGT	CTG	TT		
				C	I	Q					D	G	L	F		**59**
2	51	2	183	TGT	ATT	CAG	G	**gt**gagtagggccctggcata	.. ..	caagacctttgtgcttttt**ag**	AT	GGC	TTG	TT		
				M	S	Q					G	L	S	W		**97**
3	114	3	~1000	ATG	TCC	CAA	G	**gt**gaggtctggttgcattgg	.. ..	cctacccttcctcctttc**ag**	GC	TTG	TCC	TG		
				P	R	D	R					S	G	L		**129**
4	97	4	~1900	CA	AGG	GAC	AG	**gt**aggcagcctttcccaact	.. ..	ctctctcctttcctccct**ag**	G	TCT	GGT	TTG		
				F	H	Q					F	G	S			**215/216**
5	259	5	82	C	TTC	CAC	CAG	**gt**gagacctccaacctggcc	.. ..	acaactctgtctccccac**ag**	TTT	GGC	TCC	C		
				A	Q	P					E	L	S	L		**233**
6	352	6	522	GCA	CAA	CC A	G	**gt**agcagtggcctcagcctc	.. ..	ccgtgtctggctcccggc**ag**	AA	TTG	AGT	CT		
				R	N	L					E	G	A	V		**377**
7	132	7	412	AGA	AAT	CTT	G	**gt**gagtgtccagtcagggag	.. ..	tgctttttgtgtttctac**ag**	AA	GGG	GCT	GT		
					V	K	K				T	I	Q			**389/390**
8	38	8	779	T	GTC	AAG	AAA	**gt**gagttcctagcacctagc	.. ..	tgcatcctgtcttataac**ag**	ACA	ATA	CAA	C		
				T	D	Q	K					P	L	S		**481**
9	275	9	415	CT	GAC	CAG	AA	**gt**gagtgctgcatcctgggg	.. ..	tgcctgctcccctcctgc**ag**	A	CCC	CTG	AGC		
				I	N	I	S					V	V	G		**510**
10	87	10	276	TC	AAC	ATC	AG	**gt**ggggttgccattgtggtg	.. ..	acatatttcctatttggc**ag**	T	GTG	GTG	GGA		
				Q	Q	A					G	L	V	K		**537**
11	80	11	129	CAG	CAA	GCT	G	**gt**aaatagcaccctgtcttc	.. ..	ggcttgcaccttcctgc**ag**	G G	CTG	GTG	AA		
					V	G	Q				R	E	E			**558/559**
12	65	12	354	G	GTG	GGA	CAG	**gt**aagcactaactgtagaat	.. ..	ttctggttttgttcccac**ag**	AGG	GAG	GAA	G		
					E	Y	Q				D	L	C			**631/632**
13	219	13	468	T	GAG	TAC	CAG	**gt**gtgaaactatagaagctc	.. ..	acccaatccttcctctgc**ag**	GAC	CTG	TGT	C		
				M	I	L					A	Y	M			**698/699**
14	201	14	100	C	ATG	ATT	CTG	**gt**caggaaagggcccaggcc	.. ..	ggctgtggtgccccctac**ag**	GCA	TAC	ATG	G		
				F	L	P					Y	D	H	A		**748**
15	148	15	168	TTC	CTA	CCC	T	**gt**aagcgagtcgtgggccaa	.. ..	catttgtgagttctgtgc**ag**	AT	GAC	CAT	GC		
					S	P	I				I	E	H			**772**
16	74	16	225	C	AGC	CCC	ATC	**gt**gagcagctccgattccca	.. ..	catatcctctctgcccttag	ATC	GAG	CAT	G		
					F	W	Q				M	V	W			**798/799**
17	78	17	92	C	TT C	TGG	CAG	**gt**gcacgcttggagaggggt	.. ..	acctggatctctgtccttag	ATG	GTG	TGG	G		
					V	Y	E				V	N	L			**838/839**
18	120	18	338	C	GTC	TAT	GAG	**gt**cagagtacccaacagtaa	.. ..	ccgtgcccgatatctgac**ag**	GTG	AAC	CTG	G		
				D	F	R	R					K	V	N		**894**
19	167	19	~3100	AC	TTC	CGC	AG	**gt**gagcaagctgccagctgc	.. ..	acactcagctctcctttc**ag**	G	AAA	GTG	AAC		
				V	H	C	S					D	G	A		**912**
20	54	20	559	TG	CAC	TGC	AG	**gt**gggtaaggacctgccaca	.. ..	gtcaccatctcctactgc**ag**	T	GAC	GGT	GCA		
				M	A	K					G	V	K	E		**934**
21	65	21	191	ATG	GCC	AAA	G	**gt**aaggtcgtgtccccacag	.. ..	accccacccctattcccc**ag**	GA	GTG	AAG	GA		
					R	S	K				D	Q	F			**958/959**
22	74	22	216	C	CGT	TCT	AAG	**gt**gacagtggctccctgccg	.. ..	ctgccactggccttgcac**ag**	GAC	CAG	TTT	G		
23	578			.. ..GTGAATAAAGTTAGTGTGTTGTCTGTGCAGCTGC aacacgg												
				polyA signal						**polyA tail**						

Nucleotide sequences at the intron (lower case) and exon (upper case) junctions. Exons are numbered from the 5' end as illustrated in Figure 1. The sizes of introns and exons are indicated in base pairs. The amino acids interrupted by an intron are indicated and the positions are shown

(gttccatctcctgggagatcacatgct) (5' mapping). The procedure for rapid amplification of cDNA ends (RACE) was done as described previously [15] and repeated twice using appropriate nested primers. The final products were gel purified and subcloned into a pCRscript vector (Stratagene). Plasmid DNA from clones picked at random was subjected to DNA sequencing and aligned with the genomic sequence.

DNA sequencing. All subcloned genomic fragments and cloned RACE products were sequenced by an automated DNA sequencer managed by the core facilities in the National Institute of Dental and Craniofacial Research.

Results

To determine its genomic structure, mouse *IA-2* cDNA probes were used to screen a 129SVJ mouse genomic library. Seven overlapping clones were se-

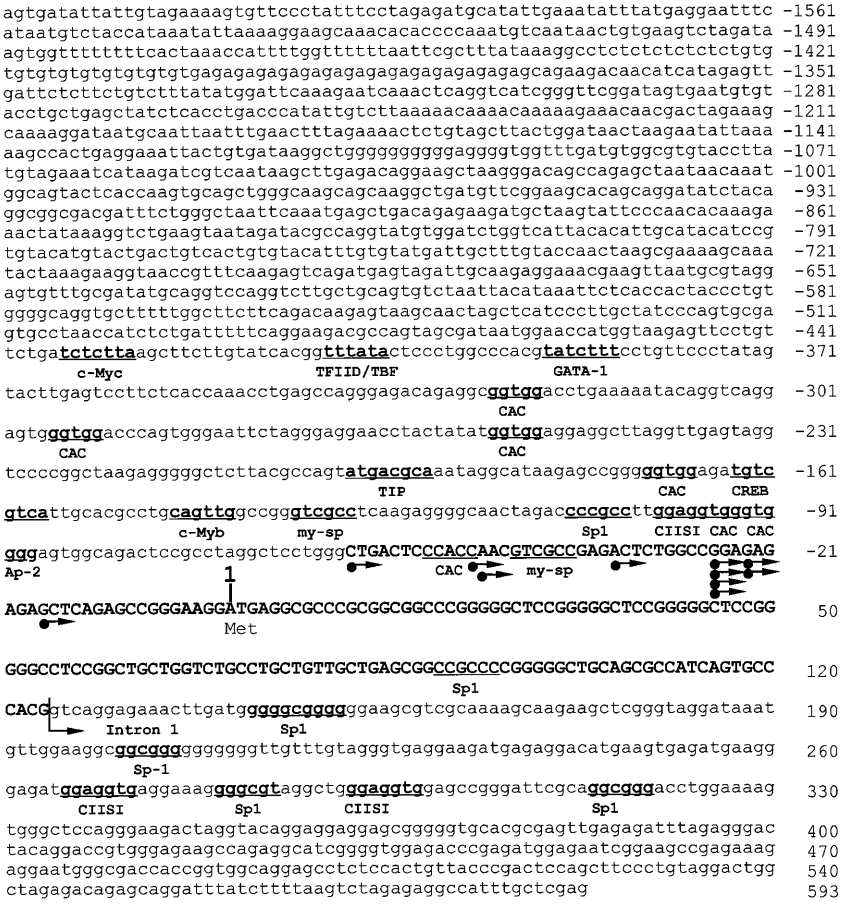

Fig. 2. Nucleotide sequence of putative promoter region of the mouse *IA-2* gene (GenBank accession number AF288816). First exon and flanking region are indicated by upper and lower case letters, respectively. Potential transcription start sites, determined by 5'-RACE, are indicated by a dot followed by an arrow (11 clones examined). The potential binding sites for transcription regulatory factors, such as Sp1, in the proximal 1 kb of the promoter are underlined. The adenosine of the translated initiating codon is assigned position + 1. Nucleotides are indicated by numerals on the right end of each lane

lected and mapped by the restriction enzymes. The genomic fragments corresponding to the whole *IA-2* cDNA were assembled and the restriction map is shown in Figure 1.

The intron-exon junctions of the *IA-2* gene were determined by directly sequencing nested deletion clones of subcloned genomic fragments. Table 1 shows the size of each exon and intron, the sequence at each exon-intron junction and the position of the amino acid where the intron is located. The mouse *IA-2* gene consists of 23 exons. The relative position of each exon is indicated (Fig. 1). The size of the exons range from the shortest (exon 8) with 38 nucleotides to the longest (exon 23) with 578 nucleotides in the 3' UTR region. A long exon in the 3' UTR also is found in other members of the PTP family [16–18]. The entire 5'-untranslated region and the first 42 amino acids are encoded by exon 1. The extracellular region is encoded by exons 1 to 13, the transmembrane region by exon 13 and the intracellular region by exons 13 to 23. All of the splice acceptor and donor sequences agree with the GT/AG rule [19] and analysis of the intron sequences showed a microsatellite sequence with 20 TG repeats in intron 3 (data not shown).

The 3' cDNA end and 5' transcription start sites were determined by rapid amplification of cDNA ends (RACE). Based on sequence analysis of 11 clones, the most likely transcription start site is located at position –26 (Fig. 2). Comparison of both the 3' and 5' RACE data with the known cDNA sequence of IA-2 indicates that there are no introns in either the 5' or 3' UTR regions. The start of the poly A tail of mIA-2 cDNA was found 25 nucleotides downstream of the last polyA signal (AATAAA) as indicated at the bottom of Table 1.

By automated DNA sequencing, 1631 bp of the 5'-upstream region from the transcription start site were determined (Fig. 2). To test for promoter activity, this region was fused to the luciferase report gene (pGL3 basic Promega). Transient transfection of PC12 and RIN cells resulted in a 10 to 15-fold increase in luciferase activity. The region around the major transcription start site, similar to human *IA-2* and other

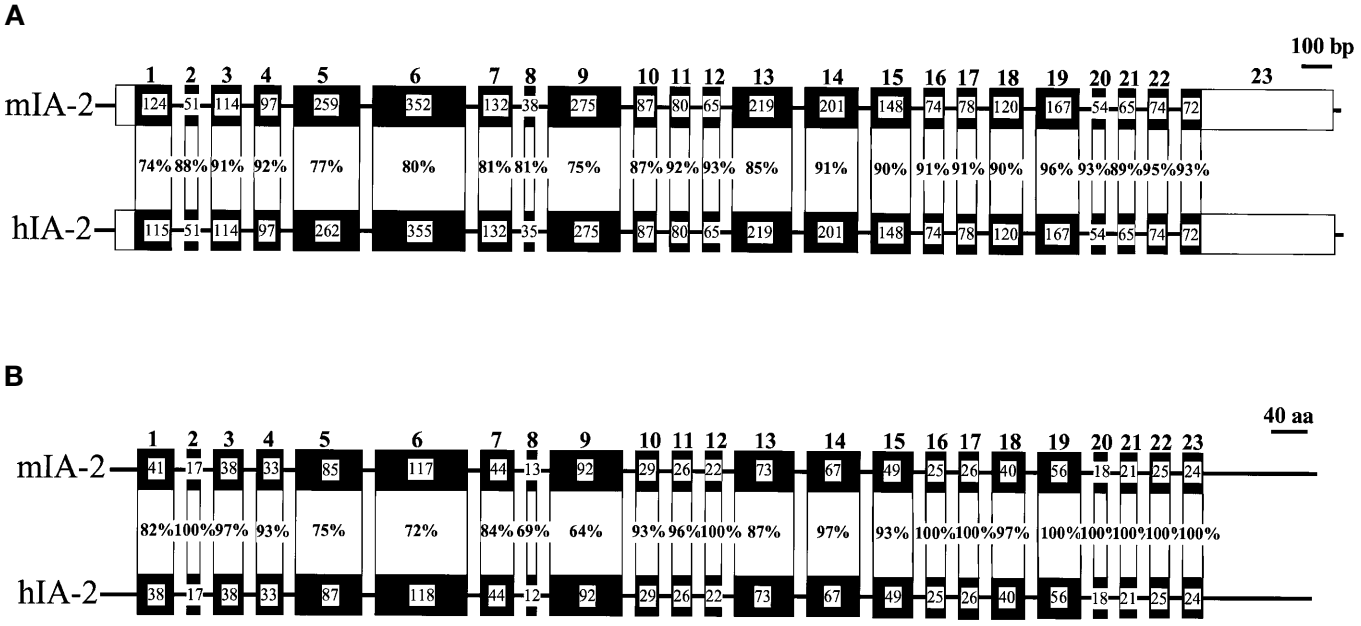

Fig. 3 A, B. Comparisons of the nucleotide (**A**) and amino-acid (**B**) sequence of mouse and human *IA-2*. Number of nucleotides or amino acids in each exon shown in boxes. Degree of identity indicated by percentage

transmembrane PTPs [16–18], was found to be highly "GC" rich and contained multiple putative Sp1 binding sites. Several Sp1 sites also were found in the 5' end of the first intron (Fig. 2). The 5'-upstream region of both the human and mouse *IA-2* gene lack a TATA box and show similar potential transcription binding sites.

Discussion

Comparison of the genomic structure of mouse and human *IA-2* shows that both species have 23 exons and like many mammalian genes [20,21] show very similar intron-exon junctions (Fig. 3). The coding region of 19 of the 23 mouse and human exons possess the same number of nucleotides with the majority showing greater than 90% identity. Similarly, 19 of the 23 mouse and human exons possess the same number of amino acids with 13 of the 19 showing greater than 96% identity. The greatest discrepancy is found in the extracellular region where the amino acids of human and mouse exons 1 and 5–9 are only 64% to 84% identical. The amino acids in the exon encoding the transmembrane region (exon 13) are 87% identical, whereas the amino acids in the exons encoding the intracellular region (exons 13 to 23) of mouse and human IA-2 are 93% to 100% identical. This high degree of sequence and structural identity indicates that the *IA-2* gene is derived from a highly conserved ancestral gene family. Studies in progress (Cai, T. et al.) have found homologues of *IA-2* in *C. elegans*, Drosophila and zebrafish.

References

1. Lan MS, Lu J, Goto Y, Notkins AL (1994) Molecular cloning and identification of a recptor-type protein tyrosine phosphatase, IA-2, from human insulinoma. DNA Cell Biol 13: 505–514
2. Notkins AL, Lan MS, Leslie RD (1998) IA-2 and IA-2β: The immune response in IDDM. Diabetes Metab Rev 14: 85–93
3. Lu J, Li Q, Xie H et al. (1996) Identification of a second transmembrane protein tyrosine phosphatase, IA-2β, as an autoantigen in insulin-dependent diabetes mellitus: Precursor of the 37-kDa tryptic fragment. Proc Natl Acad Sci USA 93: 2307–2311
4. Notkins AL, Zhang B, Matsumoto Y, Lan MS (1997) Comparison of IA-2 with IA-2β and with six other members of the protein tyrosine phosphatase family: Recognition of antigenic determinants by IDDM sera. J Autoimmun 10: 245–250
5. Lu J, Notkins AL, Lan MS (1994) Isolation, sequence and expression of a novel mouse brain cDNA, mIA-2, and its relatedness to members of the protein tyrosine phosphatase family. Biochem Biophys Res Commun 204: 930–936
6. Gorus FK, Goubert P, Semakula C et al. (1997) IA-2-autoantibody complement GAD-65-autoantibodies in new-onset IDDM patients and help predict impending diabetes in their siblings. The Belgian Diabetes Registry. Diabetologia 40: 95–99
7. Hawa MI, Rowe RE, Lan MS et al. (1997) Value of antibodies to islet protein tyrosine phosphatase-like molecule in predicting type 1 diabetes. Diabetes 46: 1270–1275
8. Kulmala P, Savola K, Petersen JS et al. (1998) Prediction of insulin-dependent-diabetes mellitus in siblings of children with diabetes: A population-based study. The Childhood Diabetes in Finland Study Group. J Clin Invest 101: 327–336

9. Lan MS, Wasserfall C, Maclaren NK, Notkins AL (1996) IA-2, a transmembrane protein of the protein tyrosine phosphatase family, is a major autoantigen in insulin-dependent diabetes mellitus. Proc Natl Acad Sci USA 93: 6367–6370
10. Leslie RG, Atkinson MA, Notkins AL (1999) Autoantigens IA-2 and GAD in Type I (Insulin-Dependent) Diabetes. Diabetologia 42: 3–14
11. Maclaren N, Lan M, Schatz D et al. (1999) Only multiple autoantibodies to islet cells (ICA), insulin, GAD65, IA-2 and IA-2β predict immune-mediated (type 1) diabetes in relatives. J Autoimmun 12: 279–287
12. Verge DF, Gianani R, Kawasaki E et al. (1996) Prediction of type 1 diabetes in first-degree relatives using a combination of insulin, GAD and ICA512bdc/IA-2 autoantibodies. Diabetes 45: 926–933
13. Xie J, Zhang B, Lan MS, Notkins AL (1998) Genomic structure and promoter sequence of the insulin-dependent diabetes mellitus autoantigen, IA-2 (PTPRN). Genomics 54: 338–343
14. Chomcyznski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156–159
15. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85: 8998–9002
16. Hall LR, Streuli M, Schlossman SF, Saito H (1988) Complete exon-intron organization of the human leukocyte common antigen (CD45) gene. J Immunol 141: 2781–2787
17. Johnson NA, Meyer CM, Pingel JT, Thomas ML (1989) Sequence conservation in potential regulatory regions of the mouse and human leukocyte common antigen gene. J Biol Chem 264: 6220–6229
18. Wong EC, Mullersman JE, Thomas ML (1993) Leukocytes common antigen-related phosphatase (LRP) gene structure: conservation of the genomic organization of transmembrane protein tyrosine phosphatases. Genomics 17: 33–38
19. Mount SM (1982) A catalogue of splice junction sequence. Nucleic Acids Res 10: 459–472
20. Chilov D, Kukk E, Taira S et al. (1997) Genomic organization of human and mouse genes for vascular endothelial growth factor C. J Biol Chem 272: 25176–25183
21. Fingert JH, Ying L, Swiderski RE et al. (1998) Characterization and comparison of the human and mouse GLC1A glaucoma genes. Genome Res 8: 377–384