MOLECULAR AND CELLULAR BIOLOGY, Oct. 1985, p. 2677–2683
0270-7306/85/102677-07$02.00/0

Vol. 5, No. 10

Sequence Analysis of the Human Major Histocompatibility Gene SXα

JEREMY M. BOSS, ROSEMARIE MENGLER, KIYOTAKA OKADA,† CHARLES AUFFRAY,‡ AND JACK L. STROMINGER*

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Received 20 May 1985/Accepted 25 July 1985

The DP subregion of the human major histocompatibility complex contains two closely linked gene pairs, DPα, DPβ and SXα, SXβ. The exon-intron organization and the complete DNA sequence of the SXα gene are reported here. There are several mutations within the SXα gene which strongly suggest that it is a pseudogene. These include two frameshift mutations, one in the α1 domain and the other in the cytoplasmic domain. A 5′ splice site mutation at the end of the α1 exon also exists. DNA sequence homology between DPα and SXα suggests that these genes arose through a gene duplication event.

The class II genes of the major histocompatibility complex (MHC) encode cell surface glycoproteins that function in the regulation of the immune response by association with "processed" foreign antigen to provide a recognition unit for helper T cells as well as for class II-directed cytotoxic T cells. Class II MHC proteins are heterodimers (an α chain of 33 kilodaltons and a β chain of 29 kilodaltons [21]) and in humans are found on the surfaces of B cells, macrophages, and activated T cells. In humans the class II genes are located within the HLA-D region of chromosome 6. This region is homologous to the Ia region of chromosome 17 in mice. To date six human α-chain genes and seven β-chain genes have been cloned as cDNAs or as genomic clones (for reviews see reference 8 and A. J. Korman, J. M. Boss, T. Spies, R. Sorrentino, K. Okada, and J. L. Strominger, Immunol. Rev., in press). The cloning and subsequent sequence analyses of some of these genes have helped organize the genes into three families. The DR region contains one α- and three β-chain genes (22a); the DQ (formally DC) region contains the DQα and DQβ genes as well as a homologous pair of DQ-like genes called DXα and DXβ (16a). The DR and DQ families have analogous counterparts in mice, namely, IE and IA, respectively. The third family, DP (formally SB), has been shown to contain four genes which are closely linked, i.e., within 120 kilobases (kb) of DNA (17, 20, 24). This cluster contains the genes encoding the DPα and DPβ proteins as well as two homologous genes which we have termed SXα and SXβ.(These genes are also referred to as DPα2 and DPβ2 by some groups. The nomenclature DPα1 [for SBα], DPα2, DPβ1 [for SBβ], and DPβ2 introduces confusion both with the numerical naming of alleles and with the naming of the domains of the chains as α1 or β1 and α2 or β2. We therefore prefer to retain SXα and SXβ until these genes have been named officially.) In addition to these families, an α-chain cDNA clone termed DO/DZα has been isolated by several groups (22; H. Inoko, A. Ando, H. Kimura, S. Ogata, and K. Tsuji, *in Histocompatibility Testing* 1984, Springer-Verlag, in press). DO/DZα has equal homology to DRα, DQα, and DPα and appears to be in a family by itself. It is not presently known whether there is a β chain associated with or linked to this gene.

Although many class II genes are encoded in the HLA-D region, it is not certain that all these genes are expressed at the level of either mRNA or protein on the cell surface. One of the DRβ genes has been shown to be a pseudogene owing to two in-frame stop codons and several splice site mutations (10). Likewise, the SXβ gene may also be a pseudogene for the same reason as the DRβ gene metioned above (7). In this paper we report the structure of the SXα gene from two different DP haplotypes, DP3 and DP4, and present evidence which suggests that SXα is a pseudogene in both of these haplotypes. These observations are further supported by the previously reported sequence of an SXα gene (termed DPα2) from an untyped individual (20).

MATERIALS AND METHODS

Isolation and subcloning. Clones were obtained by screening a cosmid library constructed from the human lymphoblastoid cell line Priess (DR4,4; DQ3,3; DP3,4) with DPα and DPβ cDNA clones as probes as previously described by Okada et al. (17). Two of the many positive clones, cG8A (from the DP3 haplotype) and cS2B (from the DP4 haplotype), were found to have two α-chain genes, DPα and SXα, surrounding a β-chain gene, DPβ. Subclones pSX17 (from cG8A) and pS2B2 (from cS2B) containing a 5-kb *Eco*RI fragment which encodes three exons of the SXα gene were inserted into pUC13 to facilitate detailed restriction fragment mapping and sequence analysis. An additional *Eco*RI subclone, pSX20, containing the 1.8-kb fragment adjacent to and upstream of pSX17 was also cloned. An ordered set of deletions were constructed from pSX17 by the DNase I method as previously described (5) to allow rapid sequencing of part of the gene.

Sequencing. DNA fragments were prepared for sequencing by 3′ end labeling with either the Klenow fragment of DNA polymerase (Bethesda Research Laboratories) with α^{32}P-labeled deoxynucleoside triphosphates or by terminal transferase [α-^{32}P]ddATP, using (Amersham Corp.). Labeled fragments were digested with a second restriction enzyme and isolated from low-meling-point agarose by extraction with phenol and ether. Sequence reactions were carried out by the method of Maxam and Gilbert (14).

* Corresponding author.
† Present address: Department of Biophysics and Biochemistry, Facultys Science, University of Tokyo, Tokoy 113, Japan.
‡ Present address: Institut d'Embryologie, 94130 Nogent sur Marne, France.

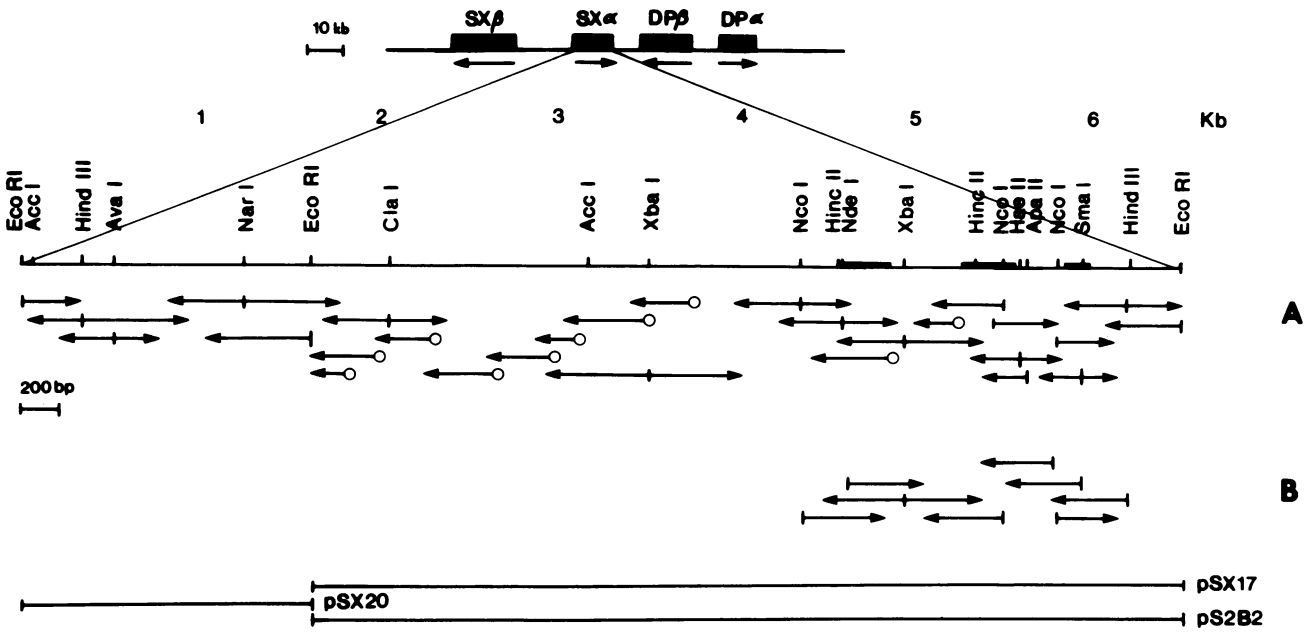

FIG. 1. Restriction map and DNA sequence strategy. The arrangement and orientation (arrows) of the genes in the DP subregion are shown in the top line, along with cosmid clones G8A and S2B. (A) Maxam and Gilbert (14) sequence strategy for clones pSX17 and pSX20. Arrows with open circles indicate sequences of clones obtained from deletions. (B) Sequence strategy for clone pS2B2.

RESULTS

Cloning, subcloning, and sequencing. The cosmid clones cG8A and cT10B were isolated from a cosmid library constructed from the human lymphoblastoid cell line Priess and were previously characterized by Okada et al. (17). Each of these cosmids as shown in Fig. 1 contains three class II genes, DPα, DPβ, and SXα. Southern blot experiments with DPα cDNA as a probe showed that the SXα gene was highly homologous, but not identical, to the DPα gene, since it hybridized less strongly. Polymorphic restriction fragment mapping and comparison of a partial DPβ amino acid sequence with the DNA sequence of the DPβ gene separated the DP3 haplotype on cG8A from the DP4 haplotype on cS2B.

The *Eco*RI fragments containing the SXα gene were subcloned into pUC13 for sequence analysis. Figure 1 shows a restriction enzyme map of clones pSX17 (SX3) and pS2B2 (SX4) as well as pSX20, the subclone adjacent to pSX17. Sequencing strategies for these clones are shown in Fig. 1.

A comparison of the nucleotide sequence and amino acid sequence of SX3α with the DPα cDNA sequence (1) showed that the SX3α gene contains a total of three nucleotide deletions, two 1-base-pair (bp) deletions (3′ to nucleotides 4764 and 5905) and one 3-bp deletion (3′ to nucleotide 5891). Maximum homology with the DPα amino acid and cDNA sequence was obtained by placing gaps, represented by asterisks, in the sequence to correct the reading frame at the deletions (Fig. 2). Individual deletions and their possible effects are discussed below.

The exon-intron arrangement of the SXα gene. The exon-intron arrangement of SXα was determined first by hybridization to various restriction fragments from DPα cDNA and later by direct amino acid and nucleotide sequence comparison with the DPα cDNA sequence (1). Exon 1 (nucleotides 4550 to 4794), exon 2 (nucleotides 5189 to 5670), and exon 3 (nucleotides 5786 to 5949) corresponded to the α1 domain (amino acids 5 through 94), the α2 domain (amino acids 95 through 178), and the connecting peptide-transmembrane (TM) domain-cytoplasmic domain (amino acids 179 through 232) of the DPα gene, respectively. The 3′ end of the TM exon was placed at the first 5′ splice site after the termination codon, as is the case with DPα, DQα, and DRα (1, 19). The three exons were spaced at similar distances, as are the exons in the DQα and DRα genes. For example, the distances between the α1 and α2 domains of SXα, DQα, and DRα were 395, 491, and 354 nucleotides, respectively.

Promoter and signal sequence region. Sequencing the 5′ end of the DNA on clone pSX17 revealed little homology to the DP signal sequence and to the highly conserved class II upstream promoter sequence (UPS) (13, 18; Okada et al., in press). Additional sequencing of the upstream *Eco*RI fragment in pSX20 revealed a region with a high degree of homology to the UPS (Fig. 3) and suggested that this is the promoter for the SXα gene. The distance between this UPS and the α1 exon (3 to 4 kb) was similar to that of the DRα (19), DQα, and DXα genes (Okada et al., in press). As with other α-chain genes, typical TATA and CCAAT sequences were not found downstream of the UPS. The first ATG following the UPS was 73 bp 3′ to the end of the conserved region and is probably too close, since for DRα (19) and DQβ (2) the distance is about 138 bp. The next ATG was located about 197 bp downstream of the UPS and may represent the beginning of a signal sequence which is 18 amino acids long before it would have to splice out to avoid a termination codon. This "signal sequence" (Fig. 2) or the other translation reading frames in this region do not share homology with the DPα signal sequence (H. Erlich, personal communication).

A region homologous to the 3′-untranslated sequence of DPα was not found in the DNA fragment (SX17) sequenced, either by direct sequence comparison or through hybridization on Southern blots with a 3′ UT DPα probe. In addition, a similar hybridization analysis on the 3′ adjacent *Eco*RI

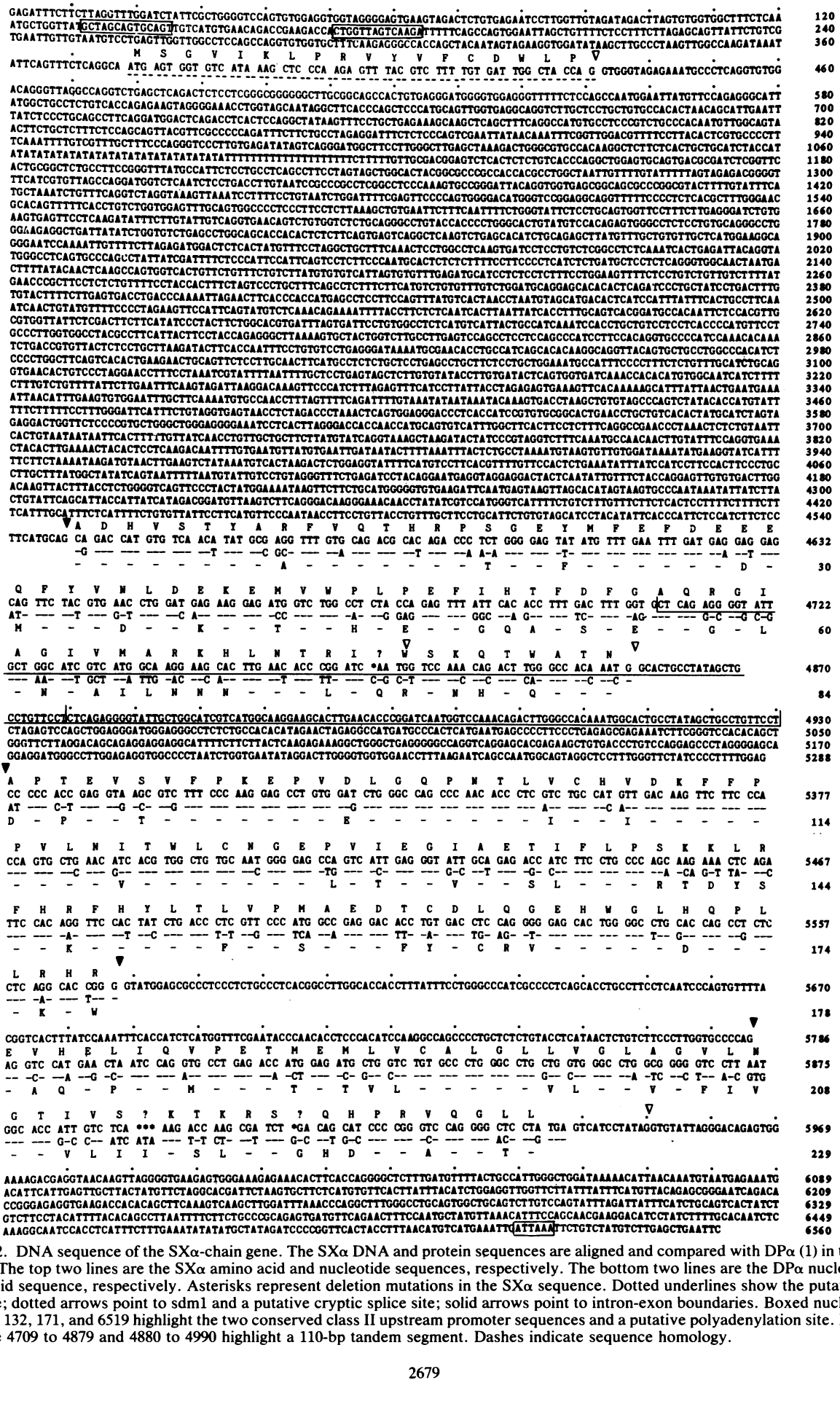

FIG. 2. DNA sequence of the SXα-chain gene. The SXα DNA and protein sequences are aligned and compared with DPα (1) in the coding regions. The top two lines are the SXα amino acid and nucleotide sequences, respectively. The bottom two lines are the DPα nucleotide and amino acid sequence, respectively. Asterisks represent deletion mutations in the SXα sequence. Dotted underlines show the putative signal sequence; dotted arrows point to sdm1 and a putative cryptic splice site; solid arrows point to intron-exon boundaries. Boxed nucleotides at positions 132, 171, and 6519 highlight the two conserved class II upstream promoter sequences and a putative polyadenylation site. Bracketed sequence 4709 to 4879 and 4880 to 4990 highlight a 110-bp tandem segment. Dashes indicate sequence homology.

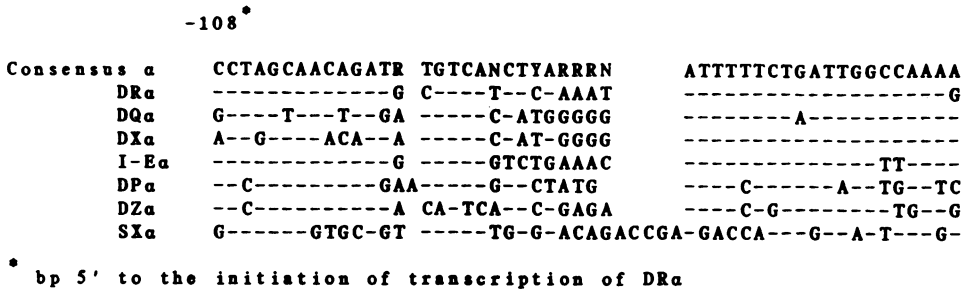

FIG. 3. Upstream promoter sequence. The upstream promoter sequences for class II α-chain genes are compared. Homology with a consensus sequence is indicated by dashes, and gaps are placed to maximize the homology. DRα (19), DQα, DXα (Okada et al., in press); I-Eα (11), DPα, and DZα (9). R, purine; Y, pyrimidine.

fragment to pSX17, which contains the 3′ end of the DPβ chain gene in the opposite orientation, also gave a negative result even under relaxed hybridization conditions. This result was not surprising because 3′ untranslated regions can typically be used as probes to distinguish closely related genes. In the DRα chain gene the distance from the transmembrane domain to the 3′UT is approximately 1,100 bp. A scan of the DNA sequence 3′ to the TM exon revealed a single polyadenylation site ATTAAA, at nucleotides 6528 through 6533, which is 589 bp 3′ to the end of the transmembrane region, suggesting the possibility that this is the end of a putative 3′UT. Although AATAAA polyadenylation sites are more frequent, ATTAAA has been found in several MHC genes, such as DCβ (2) and DRβ (12).

Mutations. For the purposes of presenting and discussing the mutations in the SXα gene and their effects on the structure of a class II α chain, it will be assumed that SXα can be transcribed, processed, and translated, although there is no evidence to support even transcription. In fact, it appears that SXα cannot be properly processed if transcribed. Deletions are referred to by a Δ followed by the number of the previous 5′ bp. For example, Δ4764 is in the α1 domain and is 3′ to nucleotide 4764 (Fig. 2, represented by an asterisk).

Mutation Δ4764 located in the α1 exon altered the reading frame of only the last 10 amino acids of that domain such that there was no amino acid sequence homology to DPα (one of nine residues; SX* in Fig. 4). These altered residues also showed little homology to other class II α chains.

The next mutation is the 5′ splice site 3′ to the α1 domain (smα1). The canonical dinucleotide at 5′ splice sites, GT (16), was changed to GC. A few examples of GC 5′ splice sites have been reported. Both the duck and chicken α^D-globin genes normally have a GC 5′ splice site in intron 2 (3, 4) and are thought to be expressed properly. Other 5′ splice site mutations have been found in some of the human β thalassemias. In these cases, splicing from the mutated 5′ splice sites is reduced or abolished and cryptic 5′ splice sites are activated (for a review, see reference 23). None of the β thalassemia splice site mutations are the same as SXα (i.e., GT to GC). However, in vitro mutagenesis of the rabbit β-globin IVS2 5′ splice site to a GC was shown to activate cryptic splice sites (28). It remains to be determined whether the GC dinucleotide in SXα can function as an efficient 5′ splice site or whether normal splicing is abolished, with or without activation of cryptic splice sites. Since all 5′ splice sites are GT in the closely related gene DPα, the splicing apparatus may not recognize the GC in SXα.

Utilization of the putative GC 5′ splice site would result in stop codons in the α2 domain, since the reading frame in the rest of the gene would be changed owing to frameshift Δ4764. If sdm1 resulted in inactivation of this splice site and the activation of cryptic splice sites, then a structurally homologous class II protein may be possible. For example,

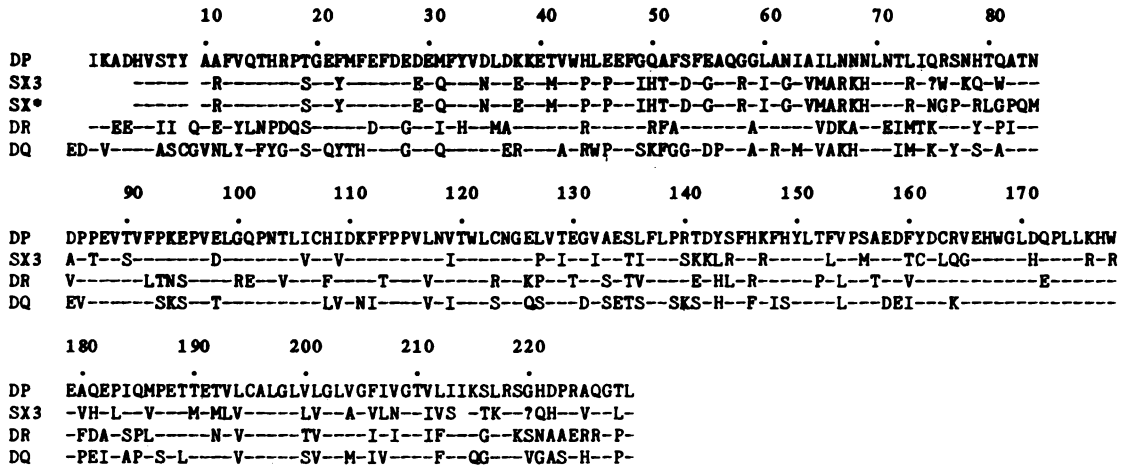

FIG. 4. Comparison of class II α chains with DPα. Dashes indicate homology with the DPα sequence. The SX3 amino acid sequence is corrected for the frameshift mutation, while SX* is not. DPα (1); DRα (19); DQα (1).

TABLE 1. Comparison of haplotypes[a]

Location	Nucleotide(s)	Nucleotide(s)		Amino acids	
		SX3	SX4	SX3	SX4
5′ to α1	4479	A	G		
5′ to α1	4517	A	G		
α1	4572, 4573	AG	GA	R	E
α1-α2 intron	5036	C	—[b]		
α2	5402	T	C	C	R
α2	5470	C	A	F	L
α2	5567	C	T	R	W
α2-TM intron	5663	G	—[b]		
α2-TM intron	5671	C	T		
α2-TM intron	5758	C	T		
TM exon	5843	G	C	L	L
3′ to TM exon	6037	—	GCATTGA[c]		
3′ to TM exon	6124	G	T		
3′ to TM exon	6198	G	A		

[a] Differences at both the amino acid and nucleotide levels are shown. Nucleotide positions are for SX3 (Fig. 2).
[b] Deletion in SX4.
[c] Insertion in SX4.

if splicing occurred after Δ4764 at bp 4769 (open arrow in Fig. 2), the domain would be shortened by only eight residues and would not incorporate any of the frameshifted residues into the protein. The effects of the frameshift would then be nullified. Alternatively, other putative cryptic splice sites could be used, such as the one located 19 nucleotides downstream of smα1, the use of which would incorporate seven additional amino acids into the peptide.

One additional point involves the occurrence of a 110-bp tandem repeat (nucleotides 4609 through 4719 and 4720 through 4830; underlined in Fig. 2) spanning the α1 exon-intron region. Each repeat contained half of the α1 domain, the frameshift mutations, Δ4764, and smα1. The second repeat was in frame with the coding sequence, so that if cryptic splicing did not occur within the first copy of the repeat, as discussed above, it probably would not occur in the second, resulting in an even larger transcript and protein. There are many GT dinucleotides 3′ to the second repeat, and it is difficult to assess which, if any, of these could be used as a splice site.

The rest of the deletions occurred in exon 4, which encodes three structural domains, the connecting peptide (nucleotides 5887 through 5824), the transmembrane domain (nucleotides 5825 through 5893), and the cytoplasmic domain (nucleotides 5894 through 5936). The connecting peptide was intact. The transmembrane domain contained three deletions in both haplotypes. Deletion Δ5890, which occurred at the end of the transmembrane domain, was a 3-bp deletion (Fig. 2) and did not affect the reading frame; therefore, both coding sequences were intact within this domain. The last deletion, Δ5905, occurred in the cytoplasmic domain of both haplotypes and altered the reading frame. This mutation frameshifted the normal stop codon so that translation would continue and the next termination codon would be read 123 bp downstream, with the result that 45 amino acid residues would be added to the cytoplasmic domain. Considering the above mutations in this gene, the abnormally long cytoplasmic tail is least likely to have an effect. It has been shown by exon deletion and shuffling experiments that the cytoplasmic tail is not necessary for class I MHC molecules to function properly in the systems assayed (29).

Another interesting difference between the DPα and SXα chains involves the cysteine residues which would form a disulfide loop in the α2 domain. However, one of the Cys residues (DPα position 159) was changed to a Leu in SXα, and a reciprocal mutation changing a Tyr to a Cys at position 157 occurred. The size of the putative disulfide loop would therefore be decreased by two amino acids. The SX4 allele also shared these changes but only had two Cys residues in this domain owing to a Cys-to-Arg change at position 123 (see below). It is not clear whether changing the size of the disulfide loop or removing the extra Cys residue would make a difference in the function of this α chain.

Haplotypes and homology to other class II genes. The mutations listed above have serious effects which, as discussed below, raise questions whether SX3α would be an expressed gene. To see whether these mutations existed in another haplotype, a second haplotype was subcloned and the SX4α chain gene was partially sequenced. The region sequenced as compared with SX3α covered nucleotides 4400 to 6220 in Fig. 2, including the α1, α2, and TM exons and their introns. There were 21 differences beween these alleles, none of which involved any of the deletions or the splice site mutation, i.e., the same deletions were present in SX4α as in SX3α. Of the 21 changes, 4 affected amino acid coding sequences, and the rest were found within the introns. All differences and their effects are listed in Table 1.

A partial sequence of the SXα gene has recently been published (20). This sequence most closely matches the SX4α sequence presented here and is also suggested to be a pseudogene by the authors. There are, however, several differences between the two sequences that are worth mentioning. The two frameshift mutations, Δ4764 and Δ5905, are placed 6 and 3 bp downstream, respectively, of those found here. For Δ4764, their placement results in three amino acid differences with the DPα (1) sequence instead of two. The placement of Δ5905 has no effect on the homology comparison. A nucleotide substitution at position 5402 (T to C) changes the Cys residue shared with DPα to an Arg residue, thus removing the extra Cys residue in this domain. The substitution of a C by a T in the TM exon at nucleotide 5923 causes a stop codon to form, but only when aligned with the DPα reading frame. This mutation has no effect on the present reading frame of SXα, owing to the frameshifts in both the α1 and TM exons.

If the deletions above are ignored by inserting gaps in the sequence, a comparison between SX3α and DPα can be carried out at the DNA or at the protein level (Fig. 2 and 4; Table 2). The α1, α2, and transmembrane exon domains between these genes have 72, 82, and 70% nucleotide homology and 63, 70, and 52% amino acid homology, respectively. The β-chain genes in the DP-region DPβ and SXβ genes are much more closely related, having 93% nucleotide and 85% amino acid homology in their coding regions (7). This high degree of homology, as well as the

TABLE 2. Amino acid homology[a]

Genes	% Homology between α chairs for:			
	α1	α2	TM	Total
DP/SX	63	70	52	64
DR/SX	38	52	46	47
DQ/SX	34	52	42	45
DQ/DX	77	99	92	89
DP/DR	54	68	52	60
DP/DQ	41	65	56	55

[a] References are DPα, DXα, and DQ4α (1) and DRα (19).

proximity of these genes to one another (50 kb) suggests that one of these two pairs of genes arose by a gene duplication event. This observation has been previously noted (7, 9, 17, 20, 24). For comparison, the amino acid homology between DQα and DXα is 89% and that between DQβ (1) and DXβ in the polymorphic β1 domain (the only sequence available in DXβ; Okada et al., in press), is 73%. The DQ/DX genes, although located farther apart than the DP/SX genes, have probably arisen by an analogous gene pair duplication event. Homology between the different α chains can be seen in Fig. 4 and is shown in Table 2. DPα and SXα are much more closely related to each other than either is to the DRα and DQα genes. There are, however, regions which have remained conserved between all these genes. Most of these regions occur in the α2 domain and may determine basic structural features in the skeleton of a class II α chain.

DISCUSSION

The SXα gene probably represents a pseudogene, that is, a gene whose product is no longer made or functional in the cell. This is supported by the following line of evidence from the DNA sequence data reported here. First, there are two 1-bp deletions which alter the reading frame. The first of these deletions occurs after 80 amino acids of the α1 domain and radically changes the last 10 amino acids. Second, the 5′ splice site at the end of what would be the α1 domain is aberrant; this would probably result in improper splicing of the mRNA precursor by activation of cryptic sites or greatly reduce proper splicing at this site. A further complication in this region involves the occurrence of a 110-bp tandem duplication which includes the first two mutations (i.e., the frameshift Δ4764 and smα1). It is difficult to predict what role the repeat may play if this exon were expressed. If cryptic splicing occurs before the end of the exon, as discussed above, the repeat would play no role in expression. On the other hand, if splicing at a site within the repeat or 3′ to it occurs, a partial duplication of the α1 domain of this protein would result. Additional reading frame problems would also arise if splicing occurred at the normal site or any other site which did not place the α2 domain in frame. Third, the frameshift in the cytoplasmic domain significantly alters the length of this domain. Because it is not clear what happens before the cytoplasmic domain, it is difficult to predict which reading frame would be used if the mRNA were translated, but if the proper reading frame was encoded, an additional 43 amino acids would be encoded as a result of Δ5906. The effect of additional amino acids in the cytoplasmic tail of class II molecules is unknown, but deletion experiments involving cytoplasmic domains of other MHC antigens suggest that it is not needed for many functions (29). Overall, the combined mutations would result in a frameshifted, improperly spliced mRNA which, if translated, would result in a protein resembling a class II α chain for only the first 80 amino acids, unless, as discussed, a proximal alternative splice site was used, shortening the α1 domain by eight amino acids. Interestingly, SXβ is also a pseudogene, although it could also be corrected by translating a shortened β2 domain with a putative cryptic splice site (7).

The SXα gene sequenced by Servenius et al.(20) is closely related to the SX4 allele presented here. Some of the differences involve our placement of the deletions, as discussed above, while others represent actual nucleotide differences. Conservation of DNA sequence information between these alleles is remarkable. A comparison of the SX3α and SX4α genes shows that there are only 21 nucleotide differences over 1,820 bp, of which 6 are within the coding region (Table 1). This gives a level of homology of just under 99%. The homology among any of the DRβ or DQβ alleles is much lower. This observation may be interpreted in many ways. For example, the expressed SXα gene may have been similar to the DRα gene in that polymorphism between haplotypes is extremely limited or even suppressed. The polymorphism seen today in SXα would then represent the recent decay of an inactive gene. Alternatively, the conservation of DNA sequences within pseudogenes may reflect the need for expression and selection to generate polymorphisms.

To investigate the expression of this gene at the level of transcription, Northern blots were carried out on B cell and fibroblast RNA by using a genomic probe spanning the end of the α2 and TM exons. The results of this experiment as well as the screening of a large B-cell cDNA library were negative. These results imply that the SXα gene is not transcribed in B cells and fibroblasts. On the other hand, the gene could be developmentally regulated or even unstable owing to splicing difficulties. However, the accumulated data presented here suggest that in these two haplotypes the SXα gene is a pseudogene.

To date, there are three known pseudogenes in the class II region of the MHC, i.e., one of the DRβ genes (10), the SXβ gene (7), and the SXα gene presented here and in reference 20. In the mouse Ia region, the Aβ3 gene is also a pseudogene (G. Widera and R. A. Flavell, submitted for publication). The DXα and DXβ genes could also be pseudogenes, because their transcripts and protein products have not as yet been detected, although they are structurally intact, including intact promoter regions and, at least for DXα, a demonstration of normal splicing (A. Korman, personal communication).

Why do pseudogenes occur in the human MHC? This genetic region has undergone a moderate expansion relative to that of the mouse and may be as much as three times as large. It includes a minimum of six α- and seven β-chain genes (compared with two α- and five β-chain genes in mice), of which eight are known to be expressed (four in mice). What could be the function of the remaining genes, and why have they persisted? One possibility is that some of these genes are expressed only in a tissue-specific or developmentally specific manner. Another possibility is that they serve as reservoirs of gene sequences for the generation of polymorphism within this family of genes. The class I and class II MHC genes are the most polymorphic known to occur in humans. One mechanism for generation of this polymorphism is the copy repair mechanism analogous to gene conversion, which can occur in multigene families and which has been shown to operate on both class I and class II MHC antigens (15, 26, 27). Thus the pseudogenes in the class II region may simply serve as reservoirs of genetic information. Lastly, the genes in this region may have duplicated in response to some ancient environmental pressure, and then, in the absence of continued pressure, the nonessential genes may have degenerated. Pseudogenes have also been shown to occur in other gene families, including the class I MHC genes (25), the immunoblogulin genes (6), and the human β-globin genes (11).

ACKNOWLEDGMENTS

We thank James Sonner for technical assistance and A. Korman, D. Dialynas, G. Blanck, and A. Krainer for helpful discussions.

This research was supported by Public Health Service research grant AM-30241 to J.L.S. from the National Institutes of Health.

K.O. is a recipient of a long-term overseas research fund from the Ministry of Education of Japan. J.M.B. is supported by Damon Runyon-Walter Winchell and National Research Service Award (5 F32 AI06860) fellowship grants.

LITERATURE CITED

1. **Auffray, C., J. W. Lillie, D. Arnot, D. Grossberger, D. Kappes, and J. L. Strominger.** 1984. Isotypic and allotypic variation of the class II human histocompatibility antigen α chain genes. Nature (London) **308**:327–333.
2. **Boss, J. M., and J. L. Strominger.** 1984. Cloning and sequence analysis of a DC-3β gene. Proc. Natl. Acad. Sci. USA **81**:5199–5203.
3. **Dodgen, J. B., and J. D. Engel.** 1983. The nucleotide sequence of the adult chicken α globin gene. J. Biol. Chem. **258**:4623–4628.
4. **Erbil, C., and J. Niessing.** 1983. The primary structure of the duck α^D-globin gene: an unusual 5′ splice junction sequence. EMBO J. **2**:1339–1343.
5. **Frisohaut, A. M., H. Garott, and H. Lehrach.** 1980. A subcloning strategy for DNA sequence analysis. Nucleic Acids Res. **8**:5541–5549.
6. **Hisajima, H., Y. Nishida, S. Nakai, N. Takahashi, S. Ueda, and T. Honjo.** 1983. Structure of the human immunoglobulin $C_\varepsilon 2$ gene, a truncated pseudogene: implications for its evolutionary origin. Proc. Natl. Acad. Sci. USA **80**:2995–2999.
7. **Kappes, D. J., D. Arnot, K. Okada, and J. L. Strominger.** 1984. Structure and polymorphism of the HLA class II SB light chain genes. EMBO J. **3**:2985–2993.
8. **Kaufman, J. F., C. Auffray, A. J. Korman, D. A. Shackelford, and J. L. Strominger.** 1984. The class II molecules of the human and murine major histocompatibility complex. Cell **36**:1–13.
9. **Kelly, A., and J. Trowsdale.** 1985. Complete nucleotide sequence of a functional HLA-DPβ gene and the region between the DPβ1 and DPα1 genes: comparison of the 5′ ends of HLA class II genes. Nucleic Acids Res. **13**:1607–1621.
10. **Larhammar, D., B. Servenius, L. Rask, and P. A. Peterson.** 1985. Characterization of an HLA-DRβ pseudogene. Proc. Natl. Acad. Sci. USA **82**:1475–1479.
11. **Little, P. F. R.** 1982. Globin pseudogenes. Cell **28**:683–684.
12. **Long, E. O., C. T. Wake, J. Gorski, and B. Mach.** 1983. Complete sequence of an HLA-DRβ chain deduced from a cDNA clone and identification of multiple nonallelic DRβ chains. EMBO J. **2**:389–394.
13. **Mathis, D. J., C. O. Benoist, V. E. Williams II, M. R. Kanter, and H. O. McDevitt.** 1983. The murine Eα immune response gene. Cell **32**:745–754.
14. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
15. **McIntyre, K., and J. Seidman.** 1984. Nucleotide sequence of mutant I-$A\beta^{bm12}$ gene is evidence for genetic exchange between mouse immune response genes. Nature (London) **308**:551.
16. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.

16a. **Okada., K, J. M. Boss, H. Prentice, T. Spies, R. Mengler, C. Auffray, J. Lillie, D. Grossberger, and J. L. Strominger.** 1985. Gene organization of DC and DX subregions of the human major histocompatibility complex. Proc. Natl. Acad. Sci. USA **82**:3410–3414.

17. **Okada, K., H. Prentice, J. M. Boss, D. Levy, D. Kappes, T. Spies, R. Raghupathy, R. Mengler, C. Auffray, and J. L. Strominger.** 1985. SB subregion of the human major histocompatibility complex: gene organization, allelic polymorphism and expression in transformed cells. EMBO J. **4**:739–748.
18. **Saito, H., R. A. Maki, L. K. Clayton, and S. Tonegawa.** 1983. Complete primary structures of the Eβ chain and its gene of the mouse major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**:5520–5524.
19. **Schamboeck, A., A. J. Korman, A. Kamb, and J. L. Strominger.** 1983. Organization of the transcriptional unit of a human class II histocompatibility antigen: HLA-DR heavy chain. Nucleic Acids Res. **11**:8663–8674.
20. **Servenius, B., K. Gustafsson, E. Widmark, E. Emmoth, G. Andersson, D. Larhammar, L. Rask, and P. A. Peterson.** 1984. Molecular map of the human HLA-SB (HLA-DP) region and sequence of an SBα (DPα) pseudogene. EMBO J. **3**:3209–3214.
21. **Shackelford, D. A., J. F. Kaufman, A. J. Korman, and J. L. Strominger.** 1982. HLA-DR antigens: structure, separation of subpopulations, gene cloning and function. Immunol. Rev. **66**:133–187.
22. **Spielman, S. R., J. Lee, W. F. Bodmer, J. G. Bodmer, and J. Trowsdale.** 1984. Six HLA-D region α chain genes on human chromosome 6: polymorphisms and associations of DCα-related sequences with DR types. Proc. Natl. Acad. Sci. USA **81**:3461–3465.

22a. **Spies, T., R. Sorrentino, J. M. Boss, K. Okada, and J. L. Strominger.** 1985. Structural organization of the DR subregion of the human major histocompatability complex. Proc. Natl. Acad. Sci. USA **82**:5165–5169.

23. **Treisman, R., S. H. Orkin, and T. Maniatis.** 1983. Structural and functional defects in β-thalessemia, p. 99–121. *In* G. Etamatoyannopoulos and A. W. Nienhius (ed.), Globin gene expression and hematopoietic differentiation. Alan R. Liss, Inc., New York.
24. **Trowsdale, J., A. Kelly, J. Lee, S. Carson, P. Austin, and P. Travers.** 1984. Linkage map of two HLA-SBβ and two HLA-SBα-related genes: an intron in one of the SBβ genes contains a processed pseudogene. Cell **38**:241–249.
25. **Weiss, E. H., L. Golden, K. Fahrner, A. L. Mellor, J. S. Devlin, H. Bullman, H. Tiddens, H. Bud, and R. A. Flavell.** 1984. Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. Nature (London) **310**:650–655.
26. **Weiss, E. H., A. Mellor, L. Golden, K. Fahrner, E. Simpson, J. Hurst, and R. A. Flavell.** 1983. The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. Nature (London) **301**:671–674.
27. **Widera, G., and R. A. Flavell.** 1984. The nucleotide sequence of the murine I-$E\beta^b$ immune response gene: evidence for gene conversion events in class II genes of the major histocompatibility complex. EMBO J. **3**:1221–1225.
28. **Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann.** 1983. Unusual splice sites revealed by mutgenic inactivation of an authentic splice site of the rabbit β-globin gene. Nature (London) **301**:38–43.
29. **Zuniga, M. C., B. Malissen, M. McMillan, P. R. Brayton, S. S. Clark, J. Forman, and L. Hood.** 1983. Expression and function of transplantation antigens with altered or deleted cytoplasmic domains. Cell **34**:535–544.