The EMBO Journal vol.3 no.12 pp.2985-2993, 1984

Structure and polymorphism of the HLA Class II SB light chain genes

Dietmar J. Kappes, David Arnot[1], Kiyotaka Okada and Jack L. Strominger

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, USA

[1]Present address: Division of Parasitology, New York University Medical Centre, 341 East 25th Street, New York, NY 10016, USA

Communicated by A. Rich

The HLA Class II region contains at least three groups of loci, DR, DC and SB, which play an important role in the immune response. The antigens encoded at these loci are heterodimers composed of an α and a β chain. The sequence of a complete Class II β cDNA clone whose sequence agrees closely with the limited N-terminal protein sequence available for the SBβ chain is reported. In addition the structure and coding sequence of genomic SBβ clones of two different SB haplotypes has been obtained and allows definition of some polymorphic regions. The SBβ gene appears to undergo alternate splicing at its 3′ end, resulting in expression of two different intracytoplasmic regions. Partial sequencing of a second non-allelic SBβ-like gene, SXβ, indicates that it is a pseudogene.

Key words: major histocompatibility complex/Class II SB light chain genes/nucleotide sequence

Introduction

The major histocompatibility complex (MHC) Class II region, located on chromosome 6 in man (HLA-D region) and on chromosome 17 in the mouse (Ia region), is involved in a number of immunological phenomena, including MLR, graft rejection and susceptibility to certain diseases (Kaufman *et al.*, 1984). So far three distinct groups of loci have been mapped to the HLA-D region, namely DR, DC and SB. The products of all of these loci are polymorphic glycoproteins consisting of two subunits, a heavy or α chain and a light or β chain. They are found on the cell surfaces of B cells, activated T cells and macrophages. DR α and β genes are homologous by sequence comparison with the murine I-E genes, while DC α and β genes are homologous to the murine I-A genes. SB appears to lack a murine homologue. Also, while DR and DC are in strong linkage disequilibrium with each other, SB is not.

In contrast to DR and DC, which were originally distinguished serologically, the existence of SB was only recognized when it was found that cells matched at all known HLA loci, including DR and DC, still showed reciprocal stimulation in a secondary mixed lymphocyte reaction (Shaw *et al.*, 1981). Class II antigens isolated with an antibody which blocks this reaction have been subjected to limited N-terminal sequencing and shown to be distinct from DR and DC α and β chains (Hurley *et al.*, 1982).

In order to isolate the genes corresponding to this antigen and so elucidate their molecular structure and organization, both cDNA and genomic libraries from human B cell lines have been constructed. Screening with DC α and β gene probes has allowed identification of clones corresponding to both a complete SBα (Auffray *et al.*, 1984) and a partial SBβ (Roux-Dosseto *et al.*, 1983) gene. This paper presents sequences of a complete SBβ cDNA clone and of the corresponding coding regions of two SBβ genomic clones of different SB haplotypes. Comparisons are drawn both between SBβ alleles and between β genes of the three human Class II loci. In addition, a partial sequence of a second SBβ-like gene, SXβ, is presented which indicates that this gene is not expressed.

Results and Discussion

Isolation and identification of an SBβ cDNA clone

A cDNA library enriched for Class II β sequences by mRNA size selection was prepared from the DRw6,DC1,SB2 homozygous cell line, LB. The library was screened with a DCβ probe, pII-β-1 (formerly pDR-β1, Larhammar *et al.*, 1983a). Thirty hybridizing clones were identified. Restriction site mapping as well as limited nucleotide sequencing and comparison with previously published sequences were used to group clones into several categories including DCβ-like (two clones) and DRβ-like (by far the most abundant class). Of the five clones that did not appear to fall within either of these two groups, one 750-bp clone, pDD2, was selected for further study. Sequence analysis established that the insert in pDD2 (called DD2) was a partial 5′ cDNA clone of a Class II β gene comprising the complete 5′-untranslated region, signal sequence and β1 and β2 domains (Figure 1). A comparison of the translated amino acid sequence of DD2 with partial N-terminal protein sequences of SBβ and DRβ (Hurley *et al.*, 1982), as well as the deduced N-terminal sequence of pDD1, whose insert was a DRβ cDNA clone from the LB cDNA library (Figure 2), showed that DD2 encodes an SBβ-like chain. DD2 and protein-SBβ agree at nine out of 10 of the positions available for comparison (as opposed to two out of 10 for DD2 and protein-DRβ). The single change (residue 9) is presumed to be due to allelic variation, as the sequences being compared stem from different SB types (protein sequence SB3, cDNA sequence SB2).

Further proof of the identity of DD2 is provided by comparing it with a clone previously isolated in this laboratory, pHAβ (Roux-Dosseto *et al.*, 1983). Southern blot analysis of SB-typed cell lines, as well as of deletion mutants and members of recombinant families has shown pHAβ to be a clone of a gene closely linked or identical to the SBβ gene. The nucleotide sequences of HAβ and DD2 are identical within the region of 400 bp over which the clones overlap (HAβ starts at residue 51; see Figure 3B). It is therefore extremely likely that they are derived from the same gene, and probably from the same allele, SB2 (HAβ was derived from the cell line JY which has been typed as heterozygous SB2/4; DD2 is derived, as mentioned, from a homozygous SB2 cell line).

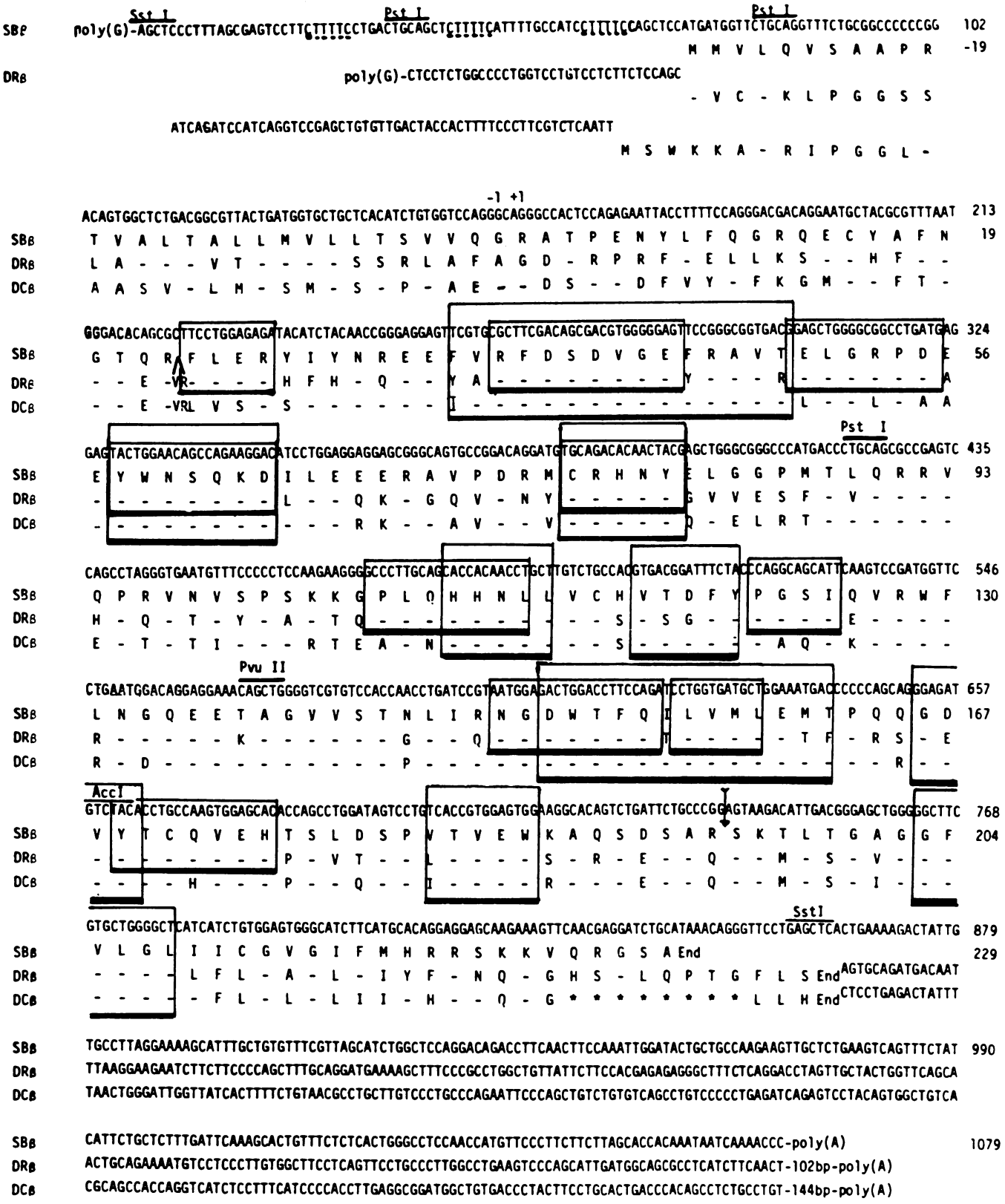

Fig. 1. Nucleotide and predicted amino acid sequence of the SB2β cDNA clone, pSB2β. The amino acid sequence is compared with the sequences of DRβ (Long *et al.*, 1983) and DCβ (Boss and Strominger, 1984) (in the case of the 3′ and 5′ untranslated regions comparison is, of course, between the nucleotide sequences). Numbering refers to the SBβ sequence. Amino acid position 1 represents the first residue of the mature protein. The Λ after amino acid 23 represents the position of the two amino acid deletion in SBβ (VR in both DRβ and DCβ). Boxes indicate identity between nucleotide (not amino acid) sequences of >12 bp. Boxes with long vertical sides indicate identity between SBβ and DCβ; boxes with short vertical sides indicate identity between SBβ and DRβ. The arrow after base pair 738 indicates the end of the DD2 insert; the remainder of the sequence is derived from HAβ (Roux-Dosseto *et al.*, 1983).

A full-length cDNA clone, pSB2β, was generated from DD2 and HAβ by splicing the two clones at a shared AccI site (Figure 3B). The insert in pSB2β is 1079 bp long and contains all protein domains as well as the 5′ and 3′-untranslated regions of SB2β including the poly(A) signal.

Comparison of SBβ cDNA sequence with DRβ and DCβ genes

The complete nucleotide sequence of the insert in pSB2β is shown in Figure 1 (the vertical arrow after base pair 738 marks the end of DD2). For purposes of comparison, its translated amino acid sequence is aligned with DRβ (Long *et al.*, 1983) and DCβ (Boss and Strominger, 1984) chains. To maximize homology between the three chains, it is necessary to introduce a two amino acid gap in pSB2β after positon 23, also generated in a computer homology matrix comparison.

	7	9	16	17	18	24	26	28	30	32
DRβ cDNA (pDD1)	*	-	-	F	F	-	F	-	-	-
DRβ protein	F	-	-	F	F	-	(F)	-	Y	Y
SB2β cDNA (pDD2)	Y	F	Y	-	F	F	-	Y	Y	-
SB3β protein	Y	Y	Y	-	F	F	-	Y	Y	-
SB3β genomic (G17A)	Y	Y	Y	-	F	F	-	Y	Y	-
SB4β genomic (T10B)	Y	F	Y	-	F	F	-	Y	Y	-

Fig. 2. Comparison of partial N-terminal sequences of protein-SBβ and protein-DRβ (Hurley *et al.*, 1982) with the predicted amino acid sequences of cDNA clones DD1 and DD2 and the genomic clones of SBβ. Since only the positions of tyrosine and phenylalanine residues were determined for the protein sequences, no other amino acids are included in this comparison. A dash represents an amino acid other than tyrosine or phenylalanine; a star indicates a position at which no information is available.

Homology comparisons between the nucleotide sequences of pSB2β and representative DRβ and DCβ genes (Long *et al.*, 1983; Larhammar *et al.*, 1983b) were carried out for individual exons and/or protein domains. Table I summarizes the results of this analysis in terms of percentage nucleotide homology between each pair of genes (SBβ/DRβ, SBβ/DCβ and DRβ/DCβ). The 5′-untranslated (5′UT) regions of SBβ, DRβ and DCβ are very weakly if at all related (29–43% homologous) within the 30-bp region available for comparison (the DRβ 5′UT is incomplete and, for the purposes of this comparison, only the corresponding sequences represented in all three clones have been used). Also note that the coding strand of the pSB2β 5′UT region is exceptionally pyrimidine rich (75% C or T) and contains three repeats of the hexamer CTTTTC, one copy of which also occurs in the DCβ 5′UT region. The signal sequences of the SBβ, DRβ and DCβ chains are 56–61% and the β1 and β2 domains 73–80% homologous. Throughout these regions DCβ consistently shows greater homology to SBβ than to DRβ by 4–6%, SBβ/DRβ homology lies between these two values). For exon 4 (CP, TM and part of CY), the relationships are partly reversed. DCβ and DRβ are most closely related (78%), followed by SBβ and DCβ (71%) and then SBβ and DRβ (68%). It seems, then, that in the 4th exon, as well as in

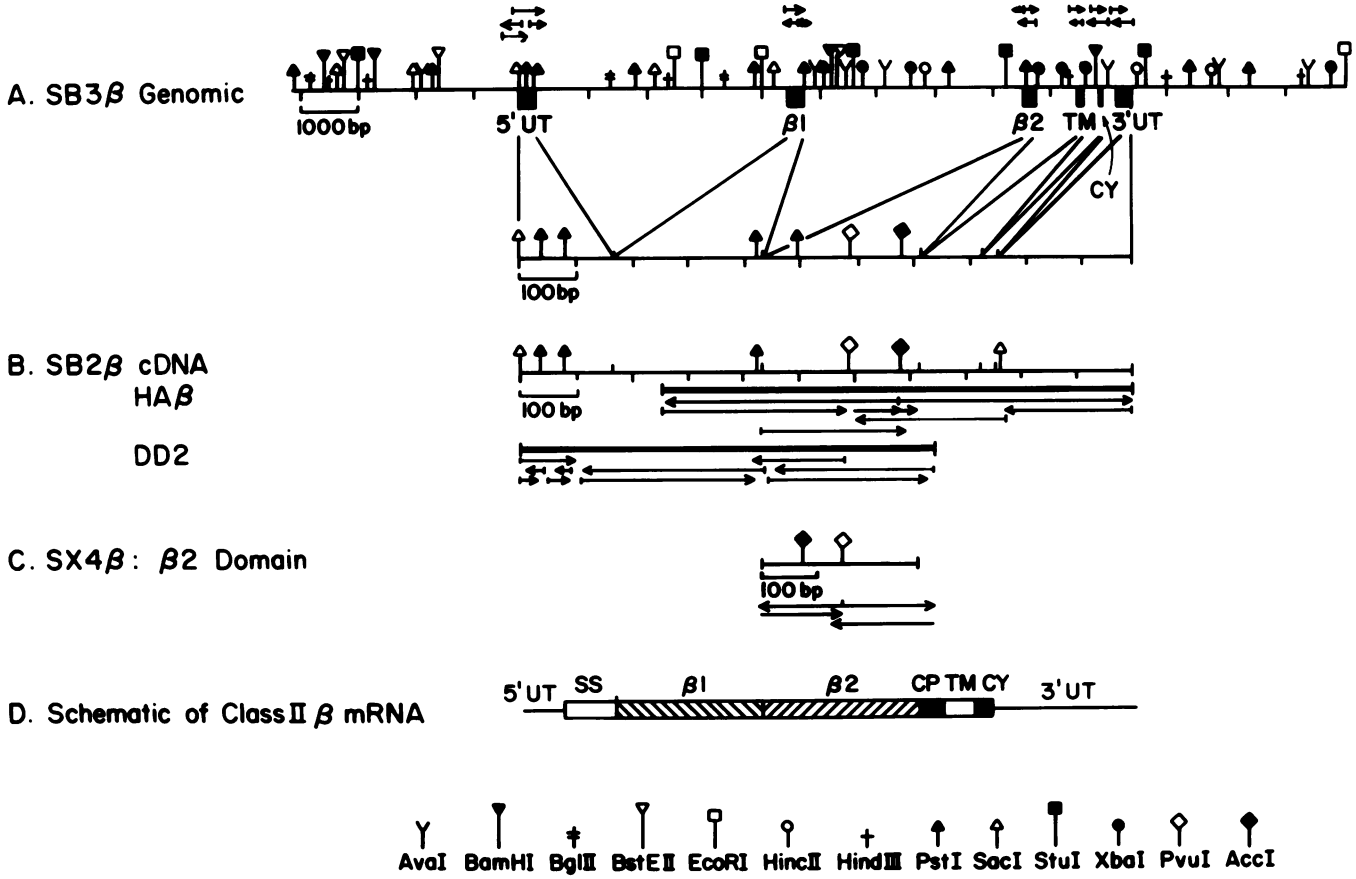

Fig. 3. Restriction maps for, (**A**) the SB3β genomic clone and its predicted cDNA, (**B**) pSB2β and the partial clones pHAβ and pDD2, (**C**) the β2 domain subclone of SX4β. Line (**D**) is a schematized representation of a Class II β mRNA drawn to scale with the SB2β and SB3β cDNA and SX4β maps. The arrows denote the sequencing strategy.

Table I. % Nucleotide homology of exons of SB, DR and CBβ

	Exon 1		Exon 2	Exon 3	Exon 4	Exon 5	Exon 6
	5'UT	SS	β1	β2	CP,TM,CY	CY	CY,3'UT
SB/DR	43	56	76	77	68	46	29
SB/DC	37	61	77	80	71	(54)	30
DR/DC	29	56	73	74	78	(58)	29

the 5th exon, SBβ has diverged significantly from the other two chains. [Exon 5 is not expressed in DCβ cDNA, but a recognizably homologous sequence is found in the genomic DNA. This sequence has been used to obtain the values shown in Table I.] The 6th exons of the three genes which contain mostly non-coding information (3'UT) are virtually unrelated. The first 14 bp, however, which encode the end of the cytoplasmic domain in some chains (see Figure 6) are 71–79% conserved.

Closer examination shows that the nucleotide differences between the three isotypic chains within particular exons are not evenly distributed. Specifically, in β1 there are distinct blocks of variability in the regions encoding amino acids 1–14, 68–76 and 82–87. In β2 differences are concentrated to a large extent at the beginning of the domain in the region encoding amino acids 94–108, which corresponds to an important polymorphic region in the HLA Class I α3 domain as well (Lopez de Castro *et al.*, 1982). Conversely, there are also distinct blocks of identity between the three genes in the β1 and β2 domains. Boxed regions in Figure 1 represent blocks of identity which comprise 12 nucleotides or more (boxes with long vertical sides indicate identity between SBβ and DCβ, boxes with small vertical sides indicate identity between SBβ and DRβ). The SBβ gene appears to consist partly of a patchwork of DRβ and DCβ sequences. Some of the conserved regions are shared between all three genes (in β1, for instance, nucleotide regions 265–290, 328–348 and 385–400), while others are shared only between two genes. In the latter case, the gene which does not share the conserved sequence can be quite divergent (in β1, for instance, nucleotide regions 226–237 and 303–322).

This pattern is presumably due in part to random point mutations which have altered a sequence in one gene while leaving the corresponding regions in the others coincidentally unaffected. Positive selection acting on the amino acid sequence probably also plays a role in preserving sequences shared by all three genes, some of which may have an essential structural function (for instance, a general role in heterodimer stabilization). However, positive selection does not explain the lack of silent mutations in these regions. In order to explain fully the pattern of conserved nucleotide sequences it may be necessary to postulate some other mechanism, for instance gene conversion-like events which act to homogenize short stretches of corresponding sequences between genes. In this connection, it is interesting to speculate that the two codon deletion found in the β1 domain of SBβ might stem from a conversion event involving a DRβ gene as donor. The deletion is bounded on its 3' side by a 12-bp sequence which is identical to the corresponding sequence from DD1, a partial DRβ clone from the LB cDNA library (Kappes, unpublished data), as well as the DRβ cDNA of Long *et al.* (1983). Also, in both DRβ clones the sequence corresponding to the codons deleted from SBβ (GTG CGG) is an almost exact copy of the preceding two codons (GAG CGG). Possibly this duplication led to an error in recombination during the hypothetical gene conversion-like event resulting in the deletion of one of the two copies.

Finally, it is intriguing that SBβ appears to be unique among Class II light chain genes in that it possesses a second glycosylation site located at the beginning of the β2 domain at amino acid 98 encoded by the sequence N-V-S, as well as the glycosylation site N-G-T in β1 at amino acid 19 which is conserved in all murine and human Class II β chains.

Structure of SBβ genomic clones

Extensive screening of cosmid libraries prepared from the human lymphoblastoid cell line Priess (SB3/4) has resulted in the isolation of genomic clones covering the entire SB region of both haplotypes (Okada *et al.*, in preparation). The SBβ gene is located between two α genes, SBα and SXα. The coding regions of the two allelic SBβ clones, from the cosmids G17A and T10B, have been identified on the basis of cross-hybridization with the SB2β cDNA clone and nucleotide sequencing. The restriction map of the G17A SBβ gene and its exon/intron organization are shown in Figure 3A. The gene is ~11 kb long and consists of six exons. The distance between the 5'UT/SS exon and the β1 exon (4.4 kb) is unusually large for a Class II β chain gene or for that matter any HLA gene. However, since the putative 5'UT/SS region from the genomic clone is entirely homologous to the corresponding cDNA sequence (there is no other sequence located closer to the β1 domain exhibiting detectable cross-hybridization) and since consensus sequences characteristic of the promoters of Class II genes are located at the expected distances upstream of this region (see below), there can be little doubt that it actually represents the first exon of SBβ.

Figure 4 shows the nucleotide sequence of the coding regions of the G17A and T10B SBβ genes as well as some flanking and intron sequences. The translated amino acid sequence of the G17A gene is also given. The splice sites which are all in agreement with the GT-AG rule, were determined by comparison with the SB2β cDNA sequence. The first exon comprises the 5'UT region, the signal sequence and the first four amino acids of the mature protein. The transcriptional initiation site has not been explicitly determined. However, it must lie within the 8-bp region bounded by the genes' presumptive TATA box (see below) and the 5' end of the SB2β cDNA clone. It is possible that the 5' end of the cDNA actually represents the transcription initiation site (for the DCβ gene, transcription has been determined to initiate 13–19 bp downstream from the TATA box; Boss and Strominger, 1984). The next two exons, β1 and β2, comprise amino acids 5–94 and 95–186, respectively. The fourth exon comprises the connecting peptide, transmembrane region and six amino acids of the cytoplasmic region. The remaining six amino acids of the cytoplasmic region are encoded in the

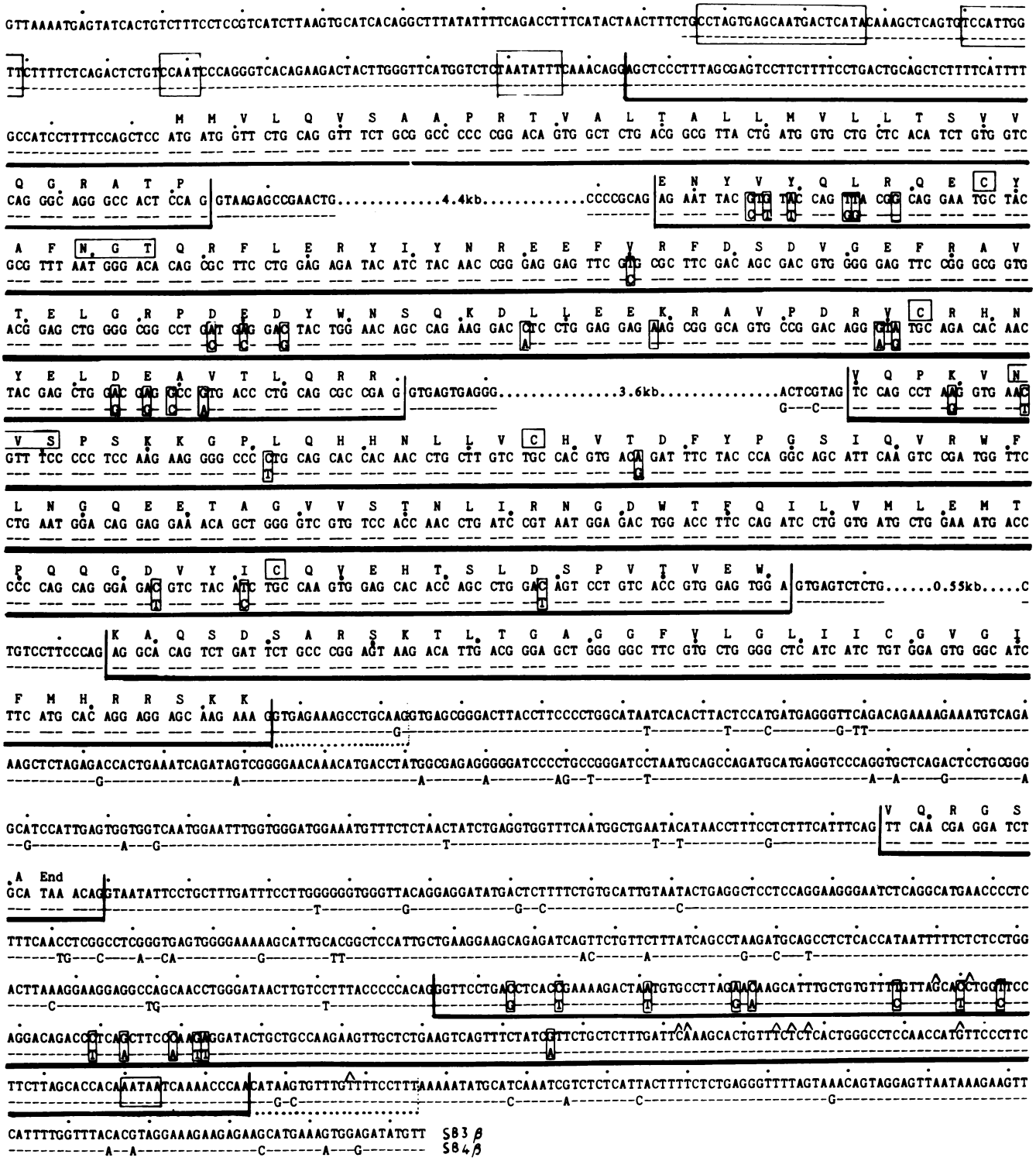

Fig. 4. Nucleotide and predicted amino acid sequence of the SB3β (top line of nucleotide sequence) and SB4β (lower line) genomic clones. Undetermined intron sequences are indicated by spaces. Boxed sequences in the 5'-flanking region of the gene are putative regulatory sequences as well as the TATA and CAT boxes. Also boxed, in the β1 and β2 domains are the glycosylation sites and conserved cysteine residues, and in the 3' untranslated region the poly(A) addition signal. Nucleotide sequences present in pSB2β are solidly underlined. Polymorphic nucleotide positions between SB3β, SB4β and SB2β are indicated by small boxes. Sequences present in pII-β-7 (Gustafsson *et al.*, 1984a), but not in pSB2β, are denoted by dotted underlining. The translated amino acid sequence is that of SB3β.

small, 24-bp, 5th exon, which also contains the translational stop codon and 4 bp of the 3'UT region.

Analysis of the putative promoter region upstream of the 1st exon of the SBβ gene shows the expected TATA and CAT boxes at positions -9 and -53 relative to the first nucleotide of the SB2β cDNA. In addition, the sequences TCCATTG-GTT and CCTAGTGAGCAATG, which are closely homologous to conserved sequences of possibly regulatory significance found in the promoter regions of other Class II genes (Mathis *et al.*, 1983; Saito *et al.*, 1983), are located at

positions −76 and −105, respectively. It should be pointed out that the SBβ sequence shows more differences from the consensus sequence than any of the other Class II genes presently available for comparison (including DCβ, I-Aβ and I-Eβ, as well as DCα, DXα, DRα and I-Eα). The promoter regions of the G17A and T10B SBβ genes are, incidentally completely conserved.

Comparison of the N-terminal translated amino acid sequence of the G17A genomic SBβ clone with the limited SBβ protein sequence data (Figure 2) demonstrates complete homology. In view of this correlation and the fact that this protein sequence was derived from material isolated with an SB3 specific antibody from the same cell line (Priess) used to generate the cosmid clones described here, this gene is presumed to be SB3β. The T10B genomic SBβ clone, which differs at one position from the SB3β protein sequence, is presumed to be SB4β.

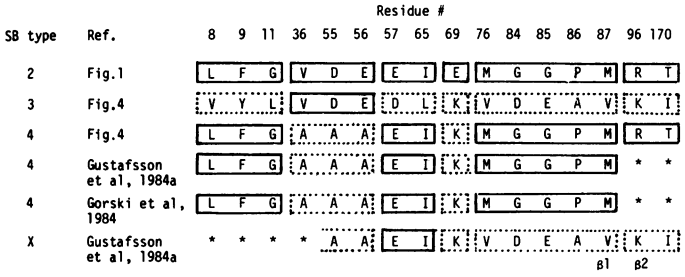

Fig. 5. Comparison of the variable residues in alleles of SBβ. Only those residues are included in this comparison which vary between two of these sequences (i.e., all positions not referred to are identical among all six sequences). Stars denote unassigned positions. Blocks of residues which are identical between different sequences are indicated by solid or dotted boxes.

Allelic variation in the SBβ gene

A comparison of the nucleotide sequences of the SB2β cDNA with the corresponding portions of the genomic clones shows that allelic polymorphism is restricted to 39 of 1079 positions (indicated by small boxes in Figure 4). In other words, 96.4% of the sequence is conserved between these three alleles. The differences, furthermore, are not evenly distributed. The first (5′UT/SS), fourth (TM) and fifth (CY) exons, which together comprise 30% of the mature mRNA contain no polymorphic positions whatsoever. The third (β2) and sixth (3′UT) exons, which comprise 282 and 231 bases, contain seven and 14 polymorphic sites respectively. Interestingly, SB2β and SB4β are identical at each of these positions. [SB3β, in turn, is identical in β2 to a recently published partial SBβ cDNA of unknown haplotype, pII-β-7 (Gustafsson *et al.*, 1984a). In the 3′UT region SB3β and pII-β-7 are identical at the 14 sites at which SB3β differs from SB2β and SB4β; however, pII-β-7 differs from SB3β at eight additional positions (four substitutions and four deletions, indicated by Λ in Figure 4).] β1 contains 18 polymorphic positions; all but two of the substitutions at these positions cause amino acid changes. This is in contrast to the β2 domain in which the substitutions at five out of seven polymorphic sites are silent. This suggests that the β2 domain, as well as the completely invariant 5′UT/SS, TM and CY domains are subject to markedly stronger conservative selective pressure than β1. For purposes of comparison, the introns flanking the CY exon and the 3′-flanking region have diverged by 8% between the

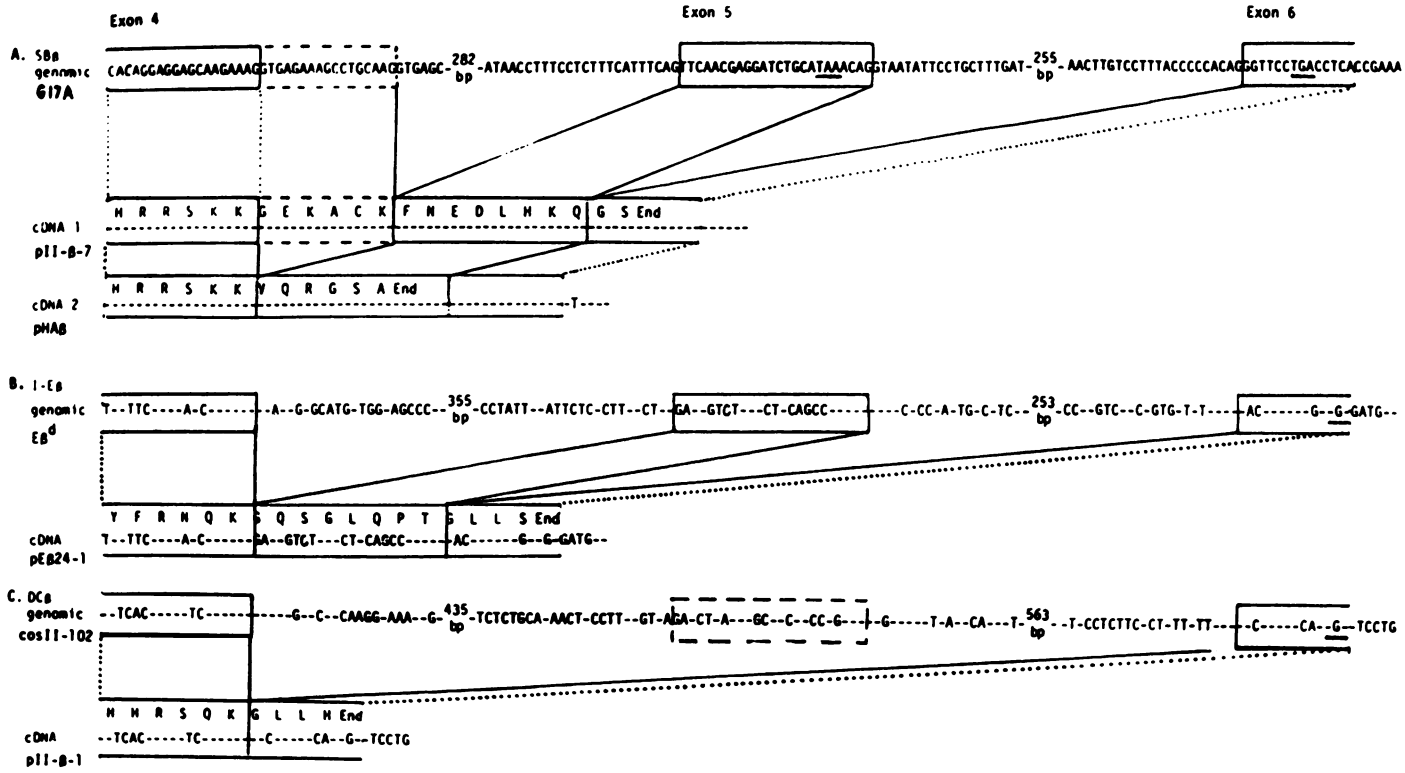

Fig. 6. Comparison of the cytoplasmic region of genomic and cDNA clones of SBβ, I-Eβ and DCβ. The beginning of the cytoplasmic domain is defined in principle by the transition from hydrophobic to hydrophilic amino acid residues at the carboxy-terminal end of the transmembrane domain. For purposes of this diagram, the boundary is placed after residue 217 of SBβ (219 of I-Eβ and DCβ). Dashes in the I-Eβ and DCβ sequences indicate homology with the corresponding regions of the SBβ genomic sequence (references; I-Eβ, Saito *et al.*, 1983; DCβ, Larhammar *et al.*, 1982; Larhammar *et al.*, 1983b; SBβ, Gustafsson *et al.*, 1984a).

SB3β and SB4β genomic clones (the lengths of the introns are, however, completely conserved).

When the amino acid sequences of SB2β, SB3β, SB4β and pII-β-7 are compared 16 amino acid positions can be identified which vary between alleles. Figure 5 shows the residues encoded at these positions for each of these sequences. From this limited data pool it appears that at each polymorphic position there is a choice between only two alternate residues. Furthermore, certain adjacent polymorphic residues seem always to be associated with particular residues at adjacent polymorphic positions (e.g., 36, 55 and 56; 76, 84, 85, 86 and 87). It would be consistent with this data to postulate that these four or five alleles have all arisen from two hypothetical primordial alleles as a result of multiple crossing-over and/or gene conversion-like events (at least, three cross-overs are necessary, specifically, between residues 11 and 36, residues 56 and 57 and residues 65 and 76, or two gene conversion-like events). The two partial SB4β sequences included in Figure 5 (Gustafsson *et al.*, 1984a; Gorski *et al.*, 1984), which were derived from two different SB4 homozygous cell lines, are incidentally, completely identical in the region of overlap with the SB4β genomic clone presented here not only at the amino acid level but also at the nucleotide level. This suggests that there is very little if any intra-allelic variation, at least for the SB4β allele.

As in the case of SBβ, the amino acid polymorphism of DCβ and DRβ is confined largely to the β1 domain (Boss and Strominger, 1984; Gustafsson *et al.*, 1984b). However, the level of polymorphism is significantly higher in DCβ and DRβ. Comparisons of three allelic DCβ sequences and four allelic or pseudo-allelic DRβ sequences (Gustafsson *et al.*, 1984b) show that the β1 domains of these genes contain 25 and 23 polymorphic amino acid positions, respectively, twice the number observed for SBβ. Furthermore, at many of these positions there are three alternate residues. The location of the so-called polymorphic regions within β1 also varies somewhat between SBβ, DCβ and DRβ. Three of the four DCβ polymorphic regions (Boss and Strominger, 1984) are also evident in SBβ at amino acids 55–57, 65–76 and 84–87 (57–59, 67–78 and 86–89 in the DCβ numbering sequence). The fourth polymorphic region in DCβ (residues 26–30) appears conserved in SBβ (this region is adjacent to the deletion in SBβ at residues 24 and 25). SBβ, however, has a fourth polymorphic region at residues 8–11, absent in DCβ but also found in I-Aβ (see Boss and Strominger, 1984) and DRβ (Gustafsson *et al.*, 1984b). DRβ, in turn, lacks the second (57–59) DCβ polymorphic region. It appears then that evolution has driven the polymorphism of SRβ, DCβ and DRβ somewhat differently.

Differential splicing at the 3' end of the SBβ gene

Comparison of SB3β genomic and SB2β cDNA nucleotide sequences with another SBβ cDNA sequence recently published, pII-β-7 (Gustafsson *et al.*, 1984a), indicates the occurrence of alternative splicing at the 3' end of the transmembrane exon (Figure 6A). In the case of pII-β-7, splicing from the fourth (TM) to the fifth (cytoplasmic) exon occurs at an alternate splice site 17 bp downstream of the donor splice site used in pHAβ, resulting in a frame shift which moves the translational termination signal from the fifth (cytoplasmic) into the sixth (3'UT) exon. This increases the size of the cytoplasmic region from 12 to 22 amino acid residues. Both the pHAβ and pII-β-7 splice donor sites are present in the SB3β genomic clone. In addition, the splice donor site used in pHAβ is also present in the gene from which pII-β-7 was transcribed as it is encoded within the cDNA (pII-β-7 apparently represents a distinct alelle of SBβ, but its sequence agrees at most polymorphic positions within β2 and the 3'UT region with SB3β). It appears then that two different products can be generated from the same SBβ gene, at least for the SB3β allele and the allele corresponding to pII-β-7. The SB2β genomic sequence is not known, so that it is unclear whether the alternative splice donor site used in pII-β-7 is present in this allele. In the SB4β genomic sequence the alternative splice site is changed from CAAG/GTGA to CAGG/GTGA. Although, in principle, this remains an acceptable donor site, it is possible that splice site selection might be affected.

Comparison of the available sequences for DRβ, I-Eβ, I-Aβ and DCβ (Long *et al.*, 1983; Saito *et al.*, 1983; Larhammar *et al.*, 1982, 1983a, 1983b; Robinson *et al.*, 1983; Boss and Strominger, 1984) indicates that the splice donor site used at the end of the fourth exon (TM) is always in a position corresponding to that used in pHAβ (Figure 6B and C). None of the genomic sequences (available only for I-Eβ, I-Aβ and DCβ, not DRβ) contain a usable splice donor site at a position homologous to that used in pII-β-7. As for the fifth and sixth exons, DRβ, I-Eβ and I-Aβ all use the same splice junctions as those shared by pII-β-7 and pHAβ (Figure 6B). However, since the translational stop codon is located in the sixth exon of these genes spanning nucleotide positions 12–14 following the splice acceptor site (as opposed to nucleotides 18–20 following the splice acceptor site of the fifth exon for pHAβ), an 18 amino acid cytoplasmic tail results. DCβ (Figure 6C) represents an unusual case in that the small cytoplasmic exon (exon 5 in SBβ) is not expressed due to a defective splice acceptor sequence (Larhammar *et al.*, 1983b; Boss and Strominger, 1984). The splice acceptor at the beginning of the 3'UT exon (exon 6 in the other genes) is in the same position as that used in all other β genes. Since the fifth exon consists of 24 bp (a multiple of three), its omission does not result in a frame shift and translation terminates at a stop codon within the 3'UT exon at a position corresponding to that used in DRβ, I-Eβ and I-Aβ. As eight codons from the cytoplasmic exon are missing, this results in a 10 amino acid cytoplasmic region for DCβ. Except then for the omission of the fifth exon in DCβ (clearly an alteration of relatively recent evolutionary origin, as I-Aβ, the murine homologue of DCβ, still expresses this exon) all four of these genes, DRβ, I-Eβ, I-Aβ and DCβ, use the same splice junctions and terminate translation on homologous stop codons. In SBβ the corresponding codon is changed from TGA to TCA. Presumably, this mutation arose after the appearance of a stop codon further upstream in exon 5 rendered downstream stops obsolete; it is less likely that the stop codon was originally located in the fifth exon, as the subsequent evolution of correct splice signals for expression of a sixth exon (containing no translated information at this point) seems intuitively improbable.

Comparison of partial sequence of SXβ with SBβ

A second SBβ-like gene, named SXβ, has also been isolated from the Priess cosmid library (Figure 3C and Okada *et al.*, in preparation). Restriction mapping demonstrates that the SXβ and SBβ genes account for all bands which light up strongly in Southern blots probed with SBβ cDNA under stringent conditions. The 5.6/6.0-kb *Hind*III polymorphism previously

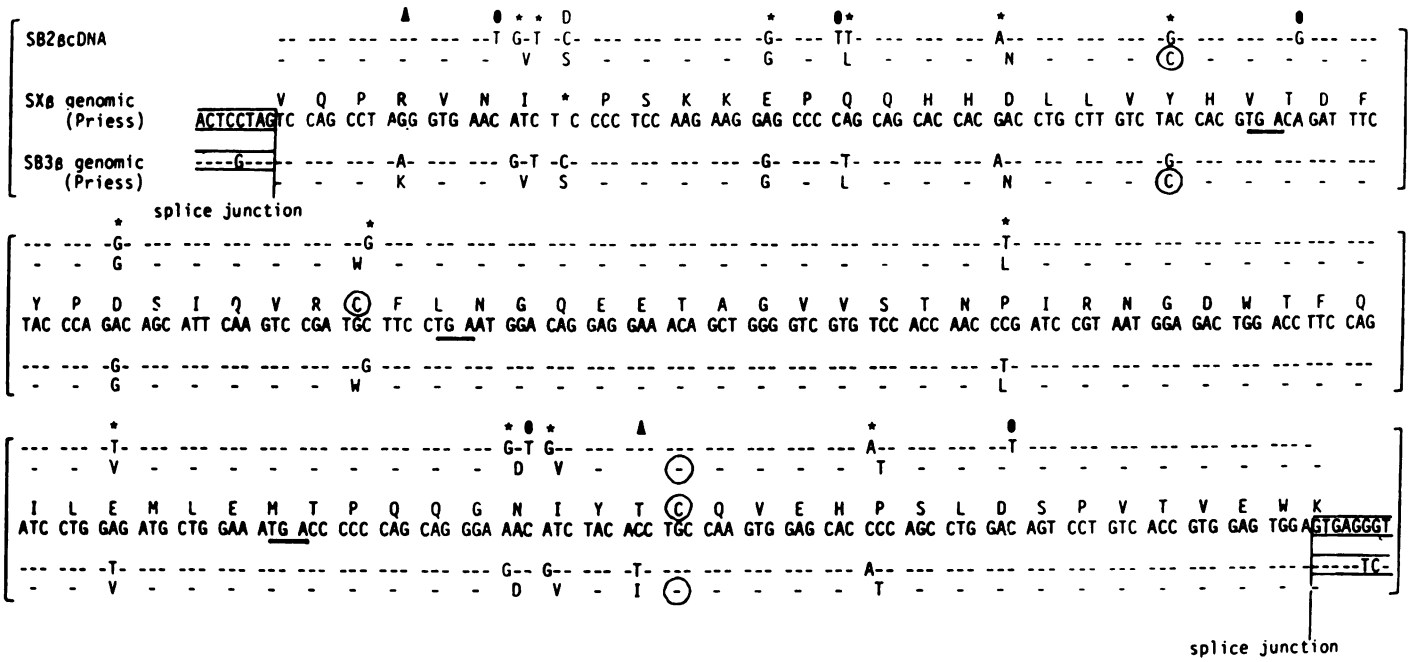

Fig. 7. Comparison of the nucleotide and predicted amino acid sequences of SX4β with SB2β and SB3β. Cysteine residues are indicated by circles. Nucleotide residues unique to SX4β are indicated by *, those to SB2β by ● and those to SB3β by ▲. Potential in-frame stop codons caused by the 1-bp deletion near the 5′ end of the SX4β β2 domain (indicated by D) are underlined.

revealed by Southern blotting with the SB2β cDNA probe (Roux-Dosseto *et al.*, 1983) is due to the SXβ gene, not the SBβ gene itself, which is located on the non-polymorphic 1.9-and 8.3-kb fragments. Since the probe used in these blots, pHAβ lacked the 5′ end, the third *Hind*III genomic fragment containing the 1st exon, which is 5.2 kb long in SB3β, was not picked up. This fragment also appears to be non-polymorphic. It should, however, be pointed out that SBβ does exhibit restriction enzyme polymorphism for a number of other enzymes (Okada *et al.*, in preparation). The SXβ gene analyzed here contains the 6.0-kb *Hind*III fragment and is thus linked to SB4β. It is referred to subsequently as SX4β.

The β2 domain of SX4β has been subcloned into the *Xba*I site of pUC12 and sequenced (Figure 3C). The nucleotide sequence and translated amino acid sequence of the β2 domain of SX4β has been compared with the homologous sequences of the SB3β genomic clone and the SB2β cDNA (Figure 7). Correct splice signals can be found at positions corresponding to those used in the SB3β gene. Within the translated region, 21 bp from the splice acceptor site, however, a 1-bp deletion shifts the remainder of the translated region out of frame. Three translational stop codons occur in the new reading frame (underlined in Figure 7). Considering how close to the beginning of the domain the deletion is located, it is conceivable, however, that an alternative splice acceptor site located directly 3′ to the mutation, of which there are potentially several, could be used, which would allow for the bulk of the coding information to be preserved.

When the β2 sequence of SX4β is aligned with the SB3β and SB2β sequences to maximize homology, by ignoring the 1-bp deletion as in Figure 7, SX4β is found to be 95% and 94% homologous, respectively, as opposed to 98% when comparing the β2 domains of SB3β and SB2β. 262 out of 282 nucleotide positions in β2 are invariant between all three sequences. At each of the remaining 20 positions one of the sequences differs uniquely from the other two, which are identical. SX4β differs uniquely at 13 positions (12 of which cause amino acid changes), SB2β at five positions (all silent) and SB3β at two positions (both cause amino acid changes). Because 12 of the 13 changes are productive, SX4β is significantly less related at the amino acid level to SB2β and SB3β (87% and 85%, respectively) than SB2β and SB3β are to each other (98%). Significantly, seven of the 12 productive changes in SX4β occur at positions which are strictly conserved between SBβ, DRβ and DCβ, as well as I-Eβ and I-Aβ. Furthermore, of these seven, two result in charge differences and two others shift the location of the first cysteine residue (one mutating Cys 23 and the other creating Cys 37) such that a possible β2 disulfide loop is reduced from 55 to 41 amino acids. In view of these drastic changes, it is highly unlikely that SX4β could produce a functional product even if an alternative splice acceptor site were used.

Interestingly, the β2 domain of SX4β shares five of the seven allelic nucleotide differences between SB2β and SB3β with SB3β (all silent) and two with SB2β (both productive). Possibly, this intermediate position is due to the fact that SXβ duplicated from SBβ before the SBβ alleles developed, so that it has preserved the sequence of the original SBβ gene at these positions. It is unlikely that SXβ duplicated from SBβ subsequent to the divergence of the SBβ alleles, as in this case some SBβ alleles would be expected to lack an SXβ gene. Southern blotting analysis shows that every SB type tested has two SBβ-like genes. It is also improbable that SXβ has acquired portions of different SBβ alelles by gene conversion-like events, as the similarities to particular alleles are not clustered but are scattered throughout the β2 domain.

Concluding remarks

The SB locus differs from the other two human Class II loci, DR and DC, in that no murine homologue can be identified. I-Eβ2, a possible homologue of SBβ on the basis of cross-hybridization, has recently been sequenced and is, in fact, no more homologous to SBβ than are I-Eβ and I-Aβ (Long *et al.*, 1984; Braunstein and Germain, cited in Long *et al.*,

1984). The Class II region in man may then have undergone a significant gene expansion. Since this must by definition have occurred subsequent to the divergence of primates and rodents and consequently also to the divergence of the I-E/DR and I-A/DC loci, the SB locus should be considerably more related to either the DR or DC locus, one of which must be its immediate progenitor, than it is to the other. However, as shown above for the SBβ gene, this is not the case. It, therefore, seems possible that the mouse has undergone a deletion of the MHC subregion corresponding to SB. In this connection it is also interesting that a Class I mouse gene, H-2K, has apparently been translocated from the Class I gene cluster into the most centromeric position in the mouse MHC, the position which in man is occupied by the SB genes (Kaufman *et al.*, 1984).

The human SB region contains two β genes, SBβ and SXβ. As shown above, the β2 domains of SBβ and SXβ are highly homologous. Sequencing studies in progress indicate that other SXβ domains, particularly β1 and also the 5'UT/SS and TM domains, which are invariant in SBβ, have diverged markedly. It appears, furthermore, that the second SXβ allele from the Priess cell line, SX3β, is also a pseudogene. It thus seems likely that there is only one functional SBβ-like gene, despite suggestive evidence from primed lymphocyte stimulation studies that a second SB-like locus may be expressed in some individuals (Pawelec *et al.*, 1982). In this connection, it is interesting to note that the SXα gene from the SB3 haplotype of the Priess cell line, which has been completely sequenced (Boss *et al.*, in preparation), is also a pseudogene (in this case due to defects in its splice site). SXα is a potential partner for SXβ in heterodimer formation.

The occurrence of alternate splicing, finally, at the end of the SBβ gene is intriguing. Changing the length of the cytoplasmic tail could potentially have significant effects on interactions of the SB antigen with intracellular components. It will be important to determine whether both forms of the cytoplasmic region are generally expressed or whether one form is expressed only aberrantly or in particular cell lines.

Materials and methods

Libraries

A cDNA library was constructed from the DRw6,DC1,SB2 homozygous human B cell line LB. mRNA was isolated by the LiCl/urea method and fractionated on sucrose gradients. Double-stranded cDNA was made by the standard methods, tailed with poly(C) and annealed with pBR322 tailed with poly(G) at the *Pst*I site. These molecules were then used to transform *Escherichia coli* strain DH1. Details of these procedures can be found in Arnot *et al.* (1984). Cosmid libraries were constructed from the DR4/4,DC4/4,SB3/4 B cell line Priess in the vectors pGNC and pTCF (Grosveld *et al.*, 1982). Details of the screening and isolation procedures as well as the organization of the clones G17A (containing the SB3β gene) and H14D (containing the SX4β gene) can be found in Okada *et al.* (in preparation).

Subcloning and sequencing

Subclones were prepared from cDNA and cosmid clones by insertion of restriction fragments from these clones into the vectors pUC9 and pUC12 (Vieira and Messing, 1982). Sequencing was performed according to the method of Maxam and Gilbert (1980).

Acknowledgements

We are grateful to members of our laboratory, in particular Jeremy Boss, Rosemarie Mengler and Holly Prentiss, for advice, reagents and technical assistance. We thank Alice Furumoto-Dawson for assistance in preparing the manuscript. This work was supported by a research grant from NIH (AM-30241). The DCβ clone, pII-β-1, was kindly provided by Dr. Per Peterson.

References

Arnot,D., Lillie,J., Auffray,C., Kappes,D. and Strominger,J. (1984) *Immunogenetics*, in press.
Auffray,C., Lillie,J., Arnot,D, Grossberger,D., Kappes,D. and Strominger, J. (1984) *Nature,* **308**, 327-333.
Boss,J. and Strominger,J. (1984) *Proc. Natl. Acad. Sci. USA,* **81**, 5199-5203.
Gorski,J., Rollini,P., Long,E. and Mach,B. (1984) *Proc. Natl. Acad. Sci. USA,* **81**, 3934-3938.
Grosveld,F., Lund,T., Murray,E., Mellor,A., Dahl,H. and Flavell,R. (1982) *Nucleic Acids Res.,* **10**, 6715-6732.
Gustafsson,K., Emmoth,E., Widmark,E., Bohme,J., Peterson,P.A. and Rask,L. (1984a) *Nature,* **309**, 76-78.
Gustafsson,K., Wiman,K., Emmoth,E., Larhammar,D., Bohme,J., Hyldig-Nielsen,J., Ronne,H., Peterson,P. and Rask.L. (1984b) *EMBO J.,* **3**, 1655-1661.
Hurley,C., Shaw,S., Nadler,L., Schlossman,S. and Capra,J. (1982) *J. Exp. Med.,* **156**, 1557-1562.
Kaufman,J., Auffray,C., Korman,A., Shackelford,D. and Strominger,J. (1984) *Cell,* **36**, 1-13.
Larhammar,D., Schenning,L., Gustafsson,K., Wiman,K., Claesson,L., Rask,L. and Peterson,P. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 3687-3691.
Larhammar,D., Hammerling,U., Denaro,M., Lund,T., Flavell,R., Rask,L. and Peterson,P. (1983a) *Cell,* **34**, 179-188.
Larhammar,D., Hyldig-Nielsen,J., Servenius,B., Andersson,G., Rask,L. and Peterson,P. (1983b) *Proc. Natl. Acad. Sci. USA,* **80**, 7313-7317.
Long,E., Wake,C., Gorski,J. and Mach,B. (1983) *EMBO J.,* **2**, 389-394.
Long,E.O., Gorski,J. and Mach,B. (1984) *Nature,* **310**, 233-235.
Lopez de Castro,J., Strominger,J., Strong,D. and Orr,H. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 3813-3817.
Mathis,D., Benoist,C., Williams,V., Kanter,M. and McDevitt,H. (1983) *Cell,* **32**, 745-754.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.
Pawelec,G., Shaw,S. and Wernet,P. (1982) *Immunogenetics,* **15**, 187-198.
Robinson,R., Germain,R., McKean,D., Mescher,M. and Seidman,J. (1983) *J. Immunol.,* **131**, 2025-2031.
Roux-Dosseto,M., Auffray,C., Lillie,J., Boss,J., Cohen,D., DeMars,R., Mawas,C., Seidman,J. and Strominger,J. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 6036-6040.
Saito,H., Maki,R., Clayton,L. and Tonegawa,S. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 5520-5524.
Shaw,S., Kavathas,P., Pollack,S., Charmot,D. and Mawas,C. (1981) *Nature,* **293**, 745-747.
Vieira,J. and Messing,J. (1982) *Gene,* **19**, 259-268.

Received on 14 August 1984