

Molecular map of the human HLA-SB (HLA-DP) region and sequence of an SB α (DP α) pseudogene

Bo Serenius, Kenth Gustafsson, Eva Widmark, Eva Emmoth, Göran Andersson, Dan Larhammar, Lars Rask¹ and Per A. Peterson

Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, and ¹Swedish University for Agricultural Sciences, Box 562, S-751 22 Uppsala, Sweden

Communicated by P.A. Peterson

The human major histocompatibility complex contains the genes for at least three different types of class II antigens, DR, DC and SB (DR, DQ and DP). They are all composed of an α and a β chain. We have cloned a chromosomal region of 70 kb containing the SB (DP) gene family in overlapping cosmid clones. This segment contains two α genes and two β genes, located in the order SB α 1, SB β 1, SB α 2 and SB β 2. The orientation of the α genes is reversed compared with that of the β genes. This organisation suggests that the SB region has arisen by duplication of a chromosomal segment encompassing one α and one β gene. Partial nucleotide sequences of the SB α 1 and SB β 1 exons demonstrate that the genes correspond to SB α and β cDNA clones. Consequently these genes are expressed. In contrast nucleotide sequence determination of the SB α 2 gene shows that it is a pseudogene.

Key words: gene family/histocompatibility antigens/HLA-SB/HLA-DP

Introduction

The major histocompatibility complex (MHC) of all vertebrate species investigated to date encodes a class of polymorphic cell-surface molecules that are expressed mainly on cells belonging to the immune system. These molecules, the class II antigens, are involved in the regulation of the immune response (Benacerraf, 1981). They are composed of an α chain of ~34 000 daltons and a β chain of ~28 000 daltons (Cullen *et al.*, 1974; Cresswell and Geier, 1975; Klareskog *et al.*, 1977). The class II antigen chains are encoded by a subregion of the MHC, the HLA-D region in man and the H-2 I-region in the mouse. Immunochemical studies of class II antigens and analyses of cDNA and genomic clones show that two types of class II molecules are expressed in the mouse, called I-A and I-E. This notion is compatible with the available molecular map of the I-region (Steinmetz and Hood, 1983). The I-A and I-E molecules are homologous but can nevertheless be easily distinguished from each other.

The sequence of the human counterparts of the I-A and I-E molecules, the DC and DR antigens, have been deduced from cDNA clones (Larhammar *et al.*, 1982a, 1982b; Auffray *et al.*, 1982; Long *et al.*, 1983; Gustafsson *et al.*, 1984b). In addition to these molecules, the existence of a third type of class II antigen, called SB, was inferred from primed lymphocyte typing (Shaw *et al.*, 1980). cDNA clones encoding SB α and β chains have now been isolated and sequenced. The SB molecules were found to be as related to DC and DR antigens as are the latter to each other (Roux-Dosseto *et al.*, 1983;

Gustafsson, 1984a; Gorski *et al.*, 1984; Auffray *et al.*, 1984).

DNA-hybridization experiments with SB α and β cDNA probes have failed to identify an SB counterpart in the mouse (Steinmetz *et al.*, 1982). The SB gene family may therefore be a distinguishing feature between the human and the murine class II regions, and the SB molecules may accordingly have a function partly different from that of the other class II antigens. To analyse the structural relationships between SB, DR and DC genes, we decided to try to establish a detailed molecular map of the chromosomal region encoding these molecules. Here we describe the characteristics of four genes of the SB locus encoded by a 70-kb chromosomal region.

Recently the nomenclature for human class II antigens was changed. While DR is still termed DR, DC and SB are now called DQ and DP, respectively. In this communication the older nomenclature is used.

Results and Discussion

Isolation and characterization of cosmid clones containing SB gene

A cDNA clone, pII- β -7, weakly hybridizing to DR and DC β cDNA probes was isolated from a cDNA library by colony hybridization. Sequence analysis showed that pII- β -7 was distinct from DR and DC β cDNA clones (Gustafsson *et al.*, 1984a). The clone was truncated in its 5' end which precluded comparison with the amino-terminal sequence of an SB β chain (Hurley *et al.*, 1982). A restriction fragment from pII- β -7 was found to hybridize to a cosmid, cosII-412, that had been isolated from a cosmid library screened with a DR β cDNA probe (Gustafsson *et al.*, 1984b). The β gene of cosII-412 was subcloned and characterized. The nucleotide sequence of the exon encoding the first domain overlapped the 5' end sequence of pII- β -7. The amino-terminal amino acid sequence predicted from cosII-412 agreed almost completely with the available sequence of an SB β chain (Hurley *et al.*, 1982). Thus, pII- β -7 and cosII-412 were identified as SB β cDNA and genomic clones, respectively.

Several cosmids containing β genes were isolated after further screening of the cosmid library with the DR β cDNA probe. Four of these cosmids hybridize strongly to a fragment from pII- β -7. They were denoted cosII-2301, -2602, -2901 and -3103. These clones and cosII-412 were initially digested with two restriction enzymes, *Bam*HI and *Hpa*I. The resulting fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose filters which were separately hybridized to probes isolated from SB β , DR α and SB α cDNA clones and to a cosmid vector probe. The results are summarized in Table I.

The cosmid segment hybridizing to the SB β probe was found to be located on *Bam*HI fragments of 4.8 kb in cosII-412, -2901 and -3103. In addition, SB β hybridization was found on a 9.5-kb fragment in cosII-2301, and in a vector-containing fragment of 22 kb in cosII-2602. The DR α probe hybridized under low stringency conditions to a 3.8-kb *Bam*HI fragment in cosII-412. Weak DR α hybridization

Table I. Restriction fragment lengths (kb) of cosmids containing SB α and SB β genes

cosII-	-412	-2301	-2602	-2901	-3103
<i>Bam</i> HI	13 d	15 d	22 a,(b,c),d	>20 d	19 d
	12 (b,c)	9.5 a	7.5	12 (b,c)	12 (b,c)
	8.5 (a)	7.4	7.4	7.5	7.5
	4.8 a	6.8	6.8	4.8 a	4.8 a
	3.8 b,c		2.6	2.6	2.6
	2.6		1.9	1.9	1.9
	1.9				
<i>Hpa</i> I	>20 b,c,d	>20 d	>20 d	>20 d	17 d
	10 (b,c)	12 a,d	10 (b,c)	10 (b,c)	10 (b,c)
	3.7 a		7.5 a	7.2	7.8
	3.3		5.2	3.7 a	3.7 a
	2.9		2.9	2.9	2.9
	1.3		1.3	1.3	1.3

a, b, c, and d denote the restriction fragments hybridizing to: SB β , DR α , and SB α cDNA and vector probes, respectively. () indicates weak signals.

signals of a 12-kb *Bam*HI fragment were later verified by hybridization with a fragment from an SB α cDNA clone as the probe (J. Young, personal communication). (Also, the SB α cDNA probe hybridized more strongly to the 3.8-kb *Bam*HI fragment than to the 12-kb fragment.) The SB α hybridizing *Bam*HI fragment of 12 kb was also found in cosII-2901 and -3103, and it corresponds to a vector-containing band of 22 kb in cosII-2602. No reactivity towards the SB α cDNA probe was found among the fragments of cosII-2301. The cosmid clones were subsequently mapped for restriction sites of seven enzymes by single and double cleavages (Figure 1B). It was evident from the restriction maps that the five cosmids overlap and together encompass 70 kb of the SB region. The two *Bam*HI fragments which hybridize to the SB β cDNA probe are located ~38 kb apart with an SB α sequence located in between the two SB β genes. The second SB α - and DR α -reactive segment was found at one end of the cosmid cluster. Thus, the SB region contains at least two α and two β genes. Provisionally, these genes are called SB α 1, SB β 1, SB α 2 and SB β 2 (Figure 1A).

Characterization of the SB β genes

The 4.8-kb *Bam*HI fragment containing the SB β 1 gene was subcloned to yield p412-4 (Figure 1C). Hybridization experiments showed that this subclone did not contain the first domain exon. Therefore, two contiguous *Hind*III fragments of 7.5 and 1.9 kb, respectively, which overlap the *Bam*HI subclone p412-4 were cloned to give p412-3 and p412-5 (Figure 1C). The three subclones were partially sequenced and the exon-intron organization determined by comparison of the sequence with that of an SB β cDNA clone (Roux-Dosseto *et al.*, 1983) (Figure 2A). In the portions available for comparison between the cDNA and genomic sequences no nucleotide substitutions were found. However, this cDNA clone as well as two other SB β cDNA clones described are truncated in their 5' parts (Gustafsson *et al.*, 1984a; Gorski *et al.*, 1984). The boundaries of the first domain exon (exon 2) were therefore deduced from a comparison with a DC β cDNA clone (Schenning *et al.*, 1984). Several mismatches were found between the translated portions of the SB β gene and the DC β cDNA sequence, but nevertheless they could be unequivocally aligned.

3210

Five exons encoding the two extracellular domains (exons 2 and 3), the membrane-spanning segment (exon 4), the cytoplasmic portion (exon 5) and the 3'-untranslated segment (exon 6) were identified (Figures 1C and 2A). The signal-peptide exon (exon 1) was not present in the 2.2-kb stretch 5' of the first domain exon in p412-3. It is therefore evident that the SB β gene spans >9 kb. The organization of the gene is analogous to that of a DC β gene (Larhammar *et al.*, 1983) with intron-exon junctions conforming to the GT-AG rule (Breathnach and Chambon, 1981).

The SB β -hybridizing *Bam*HI fragment of 9.5 kb containing the SB β 2 gene was subcloned from cosII-2301 (Figure 1C) (p2301-3). However, the cDNA probe used to identify this fragment was truncated in its 5' end. Therefore, a probe corresponding to the intron 5' of the first domain exon of the SB β 1 gene was excised from p412-3 and hybridized to *Bam*HI-digested cosII-2301. The sequence of the SB β 2 gene corresponding to the hybridization probes is present in a 7.4-kb *Bam*HI fragment, which was subcloned to give p2301-1. An overlapping *Hind*III subclone, p2301-2, of 3.4 kb was also isolated. To locate the exons within the SB β 2 gene two probes were isolated from the SB β 1 gene subclones p412-3 and p412-5. One probe corresponds to the first domain exon while the other encompasses the membrane-spanning portion, the cytoplasmic and the 3'-untranslated segments. Hybridization with these exon-specific probes to single and double digests of the subclones p2301-1 and p2301-3 defined which segments of the subclones contained the respective exons (Figure 1C). No detailed sequence information on the SB β 2 gene is available as yet.

Characterization of the SB α 1 gene

The subclone p412-2, containing the 3.8-kb *Bam*HI fragment, and the overlapping subclone p412-1 containing a 4.8-kb *Hind*III fragment were isolated. The two subclones, both hybridizing to the DR α probe, were restriction mapped and sequenced (Figure 1C). The nucleotide sequence determined was compared with the sequence of an SB α cDNA clone (Auffray *et al.*, 1984). Thereby, three exons were identified, two of which encode the first and second domains, respectively, whereas the third exon contains information for the transmembrane and cytoplasmic segments as well as the

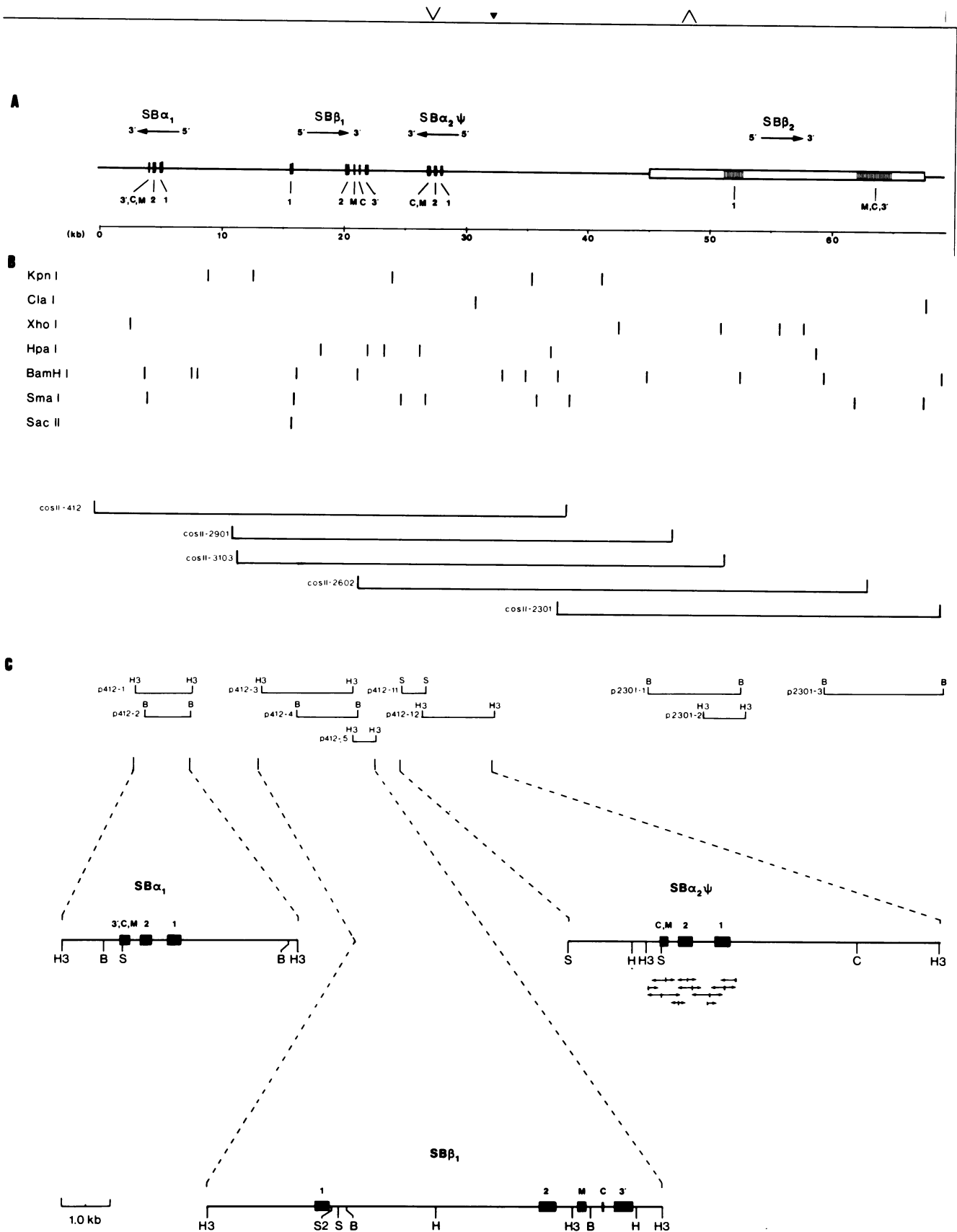


Fig. 1. (A) Molecular map of the HLA-SB region. Exons are shown as filled boxes. The SB β_2 gene is depicted as an open box with shadowed areas indicating the approximate location of segments encoding the first domain, the transmembrane, the cytoplasmic and the 3'-untranslated portions, respectively. The arrows show the orientation of the genes. (B) Restriction maps of cosmid clones. Small fragments of <1 kb may have escaped detection under the conditions used. cosII-412 was cloned in the vector pHEP, all other cosmid clones in pNNL. (C) Restriction maps of the subclones used to establish the organization of the SB genes. K = *Kpn*I, C = *Cla*I, X = *Xho*I, H = *Hpa*I, B = *Bam*HI, S = *Sma*I, S2 = *Sac*II and H3 = *Hind*III.

first few bases of the 3'-untranslated region of the α gene (Figure 2B). Although the sequenced portions of the exons displayed a few nucleotide differences as compared with the SB α cDNA clone, the predicted amino acid sequences were identical. The exon corresponding to the distal portion of the 3'-untranslated region of the SB α cDNA clone is not present in the isolated subclones. Thus, the size of the intron interrupting the 3'-untranslated region must exceed 1.4 kb. The overall organization of the SB α gene is similar to that of the DR α (Korman *et al.*, 1982; Lee *et al.*, 1982) and the DX α genes (Auffray *et al.*, 1984) and all splice junctions conform to the GT-AG rule (Breathnach and Chambon, 1981).

SB α 2 is a pseudogene

The second sequence, weakly hybridizing with a DR α probe was cloned as a 6.0-kb *Hind*III fragment (p412-12) and as an overlapping 1.9-kb *Sma*I fragment (p412-11) (Figure 1C). Sequence analysis of the cloned fragments revealed the presence of the exons corresponding to the first domain, the second domain and the transmembrane and the cytoplasmic portions of a class II α gene (Figure 3). The exons of the gene are 75% homologous to an SB α cDNA clone (Auffray *et al.*, 1984). Thus, the SB α 2 gene has diverged considerably from the gene corresponding to the cDNA clone. A closer inspection of the sequence revealed that SB α 2 is probably a pseudogene by several criteria. (i) Two single-base deletions, one in the first domain exon and one in the transmembrane exon cause translational frame shifts. (ii) The donor splice junction sequence after the first domain exon deviates from the GT-AG rule (Breathnach and Chambon, 1981). This mutated splice site is located in the 110-bp sequence which has been duplicated (underlined in Figure 3). This duplication probably arose after the SB α 2 gene was inactivated since both copies of the sequence are identical and contain the frame-shift mutation and the changed splice junction sequence. (iii) A premature stop codon is present in the SB α 2 sequence which would shorten the cytoplasmic tail of a putative gene product by four amino acid residues. (iv) Codon 159 in the SB α 2 gene (Figure 3), a cysteine codon in the SB α cDNA clone, has been mutated to a leucine codon. However, a compensatory mutation has changed position 157 to a cysteine codon.

The first two of the changes described above involve major mutations, each of which by itself would probably be sufficient to inactivate the gene. However, splicing at a donor site 26 bp 5' of the one used in all other α genes would give rise to a protein with a first domain 8 amino acids shorter. The third and fourth criteria distinguish SB α 2 from other SB α sequences, but would not necessarily make SB α 2 a pseudogene. Whether a shortened cytoplasmic tail and a shortened disulfide-loop will affect the structure of a putative SB α 2 gene product, were it derived from an active gene, is a matter of conjecture only.

The SB α 2 and β 2 genes are oriented in opposite directions

The subclones p412-1, -2, -3, -4, -5, -11, and -12 all contain restriction sites that have been mapped in the cosmid clones (Figure 1). Accordingly, the orientations of the SB β 1, SB α 1 and SB α 2 genes could be determined relative to each other. The orientation of the SB β 2 gene is evident from the hybridization patterns of the subclones containing this gene using the probes corresponding to the 5' end and the 3' end of the SB β 1 gene. Thus, it can be concluded that the two β genes are oriented in one direction and the two α genes in the other. This seems to be a general feature of class II genes, as

the same organization has earlier been noted for the murine class II α and β genes (Steinmetz *et al.*, 1982).

The fact that SB α and β genes are transcribed from different strands and the observation that they occur pairwise may suggest that during evolution a chromosomal segment encoding an α and a β gene duplicated to give rise to the SB region. Whether duplications of pairs of genes have played a role in generating the DR and DC loci is currently under examination. The isolation of three cosmid clones containing SB α and β genes from two different genomic libraries, has been reported. One of these libraries is HLA-heterozygous and the other is untyped (Trowsdale *et al.*, 1984). A partial map of the SB region has also been published by Gorski *et al.* (1984). The molecular map of the SB region deduced from these cosmids is similar to the one established in this communication. The distances between the genes as well as the location of restriction sites are remarkably identical in the two maps, suggesting that the SB region is highly conserved among different haplotypes.

Materials and methods

Isolation and characterization of genomic clones

Two genomic libraries were constructed from DNA isolated from an HLA-homozygous human individual typed to be Dw4/DR4 and SB4 (Jørgensen *et al.*, 1974 and F. Bach, personal communication), using the cosmid vectors pHEP and pNNL (Grosveld *et al.*, 1982). The selection of the HLA-homozygous donor and the construction of the cosmid libraries will be described elsewhere.

A total of 900 000 colonies were screened with the 790-bp *Sac*I-*Hind*III fragment of the DR β cDNA clone pII- β -3 (Gustafsson *et al.*, 1984b). Several of the isolated clones were found to hybridize strongly to the 600-bp *Rsa*I-*Hpa*I fragment of the SB β cDNA clone pII- β -7 (Gustafsson *et al.*, 1984a).

In Southern blot analyses of the cosmid clones the following probes were used: (a) a 600-bp *Rsa*I-*Hpa*I fragment of pII- β -7 which encodes the last third of the first domain, the second domain, and the transmembrane and cytoplasmic segments of an SB β chain; (b) a 490-bp *Pst*I fragment of pII- α -1 (Larhammar *et al.*, 1982b) encoding the first and second domains of the DR α chain; (c) a 600-bp *Eco*RI-*Ava*I fragment of pDA α 13B (J. Young, personal communication) encoding the first, the second, the membrane-spanning and the cytoplasmic domains of an SB α chain; (d) total vector pNNL; (e) a 680-bp *Acc*I fragment of subclone p412-3, which corresponds to part of the intron 5' of the first domain exon of the SB β 1 gene; (f) a 240-bp *Nar*I-*Sac*II fragment of subclone p412-1, corresponding to the first domain of the SB β 1 gene; (g) a 1255-bp *Hind*III-*Hpa*I fragment of subclone p412-3, which corresponds to the membrane coding part, the cytoplasmic and the 3'-untranslated regions of the SB β 1 gene. Probes were labelled by nick-translation (Rigby *et al.*, 1977).

The SB α sequences were detected by the DR α cDNA probe using low stringency washes (2 x SSC at 68°C). Library hybridizations were washed at a stringency of 1 x SSC at 68°C. All other hybridizations were washed in 0.2 x SSC at 68°C. (1 x SSC equals 0.15 M sodium chloride and 0.015 M sodium citrate, pH 7.0.)

DNA sequence determination

Nucleotide sequences were determined with the chemical degradation procedure (Maxam and Gilbert, 1980) and with a modification of the dideoxy chain termination procedure (Sanger *et al.*, 1977) using exonuclease III and synthetic oligonucleotide primers.

Acknowledgements

We wish to thank Ms. A. Morén and P. Agren for skilful technical assistance. Drs. John Young and John Trowsdale kindly provided the cDNA clone pDA α 13B. This work was supported by grants from the Swedish Cancer Society, King Gustav V:s 80-years' Fund, and Marcus Borgström's Fund.

References

- Auffray, C., Korman, A.J., Roux-Dosseto, M., Bono, R. and Strominger, J.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6337-6341.
- Auffray, C., Lillie, J.W., Arnot, D., Grossberger, D., Kappes, D. and Strominger, J. (1984) *Nature*, **308**, 327-333.
- Benacerraf, B. (1981) *Science (Wash.)*, **212**, 1229.
- Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.

- Cresswell,P. and Geier,S.S. (1975) *Nature*, **257**, 147-149.
- Cullen,S.E., David,C.S., Shreffler,D.C. and Nathenson,S.G. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 648-652.
- Gorski,J., Rollini,P., Long,E. and Mach,B. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3934-3938.
- Grosveld,F.G., Lund,T., Mellor,A.L., Dahl,H.H.M., Bullman,H., Bud,H. and Flavell,R.A. (1982) *Nucleic Acids Res.*, **10**, 6715-6732.
- Gustafsson,K., Emmoth,E., Widmark,E., Böhme,J., Peterson,P.A. and Rask,L. (1984a) *Nature*, **309**, 76-78.
- Gustafsson,K., Wiman,K., Emmoth,E., Larhammar,D., Böhme,J., Hyldig-Nielsen,J.J., Peterson,P.A. and Rask,L. (1984b) *EMBO J.*, **3**, 1655-1661.
- Hurley,C.K., Shaw,S., Nadler,L., Schlossman,S. and Capra,J.D. (1982) *J. Exp. Med.*, **156**, 1557-1562.
- Jørgensen,F., Lamm,L.U. and Kissmeyer-Nielsen,F. (1974) *Tissue Antigens*, **4**, 419-428.
- Klareskog,L., Sandberg-Trägårdh,L., Rask,L., Lindblom,J.B., Curman,B. and Peterson,P.A. (1977) *Nature*, **265**, 248-251.
- Korman,A.J., Auffray,C., Schamboeck,A. and Storminger,J.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6013-6017.
- Larhammar,D., Schenning,L., Gustafsson,K., Wiman,K., Claesson,L., Rask,L. and Peterson,P.A. (1982a) *Proc. Natl. Acad. Sci. USA*, **79**, 3687-3691.
- Larhammar,D., Gustafsson,K., Claesson,L., Bill,P., Wiman,K., Schenning,L., Sundelin,J., Widmark,E., Peterson,P.A. and Rask,L. (1982b) *Cell*, **30**, 153-161.
- Larhammar,D., Hyldig-Nielsen,J.J., Servenius,B., Andersson,G., Rask,L. and Peterson,P.A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7313-7317.
- Lee,J.S., Trowsdale,J., Travers,P.J., Carey,J., Grosveld,F., Jenkins,J. and Bodmer,W.F. (1982) *Nature*, **299**, 750-752.
- Long,E.O., Wake,C.T., Gorski,J. and Mach,B. (1983) *EMBO J.*, **2**, 389-394.
- Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
- Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
- Roux-Dosseto,M., Auffray,C., Lillie,J.W., Boss,J.M., Cohen,D., DeMars,R., Mawas,C., Seidman,J.G. and Strominger,J.L. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6036-6040.
- Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
- Schenning,L., Larhammar,D., Bill,P., Wiman,K., Jonsson,A.-K., Rask,L. and Peterson,P.A. (1984) *EMBO J.*, **3**, 447-452.
- Shaw,S., Johnson,A.H. and Shearer,G.M. (1980) *J. Exp. Med.*, **152**, 565-580.
- Steinmetz,M. and Hood,L. (1983) *Science (Wash.)*, **222**, 727-733.
- Steinmetz,M., Minard,K., Horvath,S., McNicholas,J., Frelinger,J., Wake,C., Long,E., Mach,B. and Hood,L. (1982) *Nature*, **300**, 35-42.
- Trowsdale,J., Kelly,A., Lee,J., Carson,S., Austin,P. and Travers,P. (1984) *Cell*, **38**, 241-249.

Received on 16 August 1984; revised on 11 October 1984