

Molecular organization of the HLA-SB region of the human major histocompatibility complex and evidence for two SB β -chain genes

(Ia antigens/cosmid library/nucleotide sequence/ α - β linkage/allelic polymorphism)

JACK GORSKI, PIERRE ROLLINI, ERIC LONG*, AND BERNARD MACH

Department of Microbiology, University of Geneva School of Medicine, Geneva, Switzerland

Communicated by V. Prelog, February 28, 1984

ABSTRACT The class II products of the major histocompatibility complex, also called Ia antigens, are composed of two polypeptide chains, the α and β chains, both encoded within the major histocompatibility complex. In man, the class II antigens can be divided into three biochemically distinct groups called HLA-DR, HLA-DC, and HLA-SB. Our isolation of cDNA clones for the polymorphic β chain of HLA-DR and HLA-DC has allowed us to study the organization of the class II genes. Here we identify the HLA-SB β -chain gene in recombinant clones from a cosmid library generated from a consanguineous homozygous B-cell line. The SB β -chain gene is linked to the SB α -chain gene and the two genes are in opposite orientation. A second SB β -chain gene, corresponding to a new SB β II locus, has also been identified and cloned. The SB β -chain genes show much less allelic restriction site polymorphism than the genes for the β chains of HLA-DR or HLA-DC.

The class II products of the major histocompatibility complex (MHC) are responsible for the restriction of cell-cell interactions between autologous cells of the immune system as well as for allogenic stimulation of lymphocyte proliferation. These properties are the result of the extensive allelic polymorphism that characterizes class II antigens (for a review, see ref. 1). In the human MHC (HLA), the proliferative and self-restrictive properties of class II molecules have been mapped to the D region of the HLA complex (2). The biochemical equivalent of HLA-D is thought to be the HLA-DR molecules (3), which are encoded in the DR subregion. A second group of polymorphic class II antigens, HLA-DC, is also encoded in the HLA-D region (4). A third class II species, HLA-SB, has been identified on the basis of restriction of antigen presentation (5) and has recently been isolated and partially characterized (6). The SB antigen has been genetically mapped centromeric to the DR subregion (7). Each of the class II products is composed of two transmembrane polypeptide chains, the α and β chain.

A detailed understanding of the role of individual class II genes in the control of the immune response will require a knowledge of their number and genetic organization. The cloning of the cDNA for the polymorphic β chain of HLA-DR and HLA-DC (8), together with the study of the corresponding genes in genomic clones, has permitted us to identify several distinct β -chain loci within HLA-DR (ref. 9; unpublished data) and HLA-DC (10). It has thus become evident that the HLA-D region is more complex than originally thought. The ability to reprogram animal cells for the surface expression of HLA-DR through DNA-mediated cell transformation (11) suggests that it will eventually be possible to test each of the multiple class II genes individually in functional assays. Here we extend our analysis of the HLA-D region of the SB genes. An SB β -chain gene has been cloned and characterized from a consanguineous homozy-

gous B-cell line, HHK. This SB β -chain gene (SB β I) is closely linked to an SB α -chain gene. Furthermore, a second "SB-like" β -chain gene has been identified and cloned, implying the existence of an SB β II locus in the D region.

MATERIALS AND METHODS

Construction of the Cosmid Library. Details of the cosmid library will be published elsewhere. In brief, DNA from the consanguineous homozygous cell line HHK was partially digested with *Sau3A1* and fractionated on sucrose density gradients. DNA fractions of 35-40 kilobases (kb) were collected and ligated with arms of cosmid pOPF (12), which had been treated with phosphatase. The ligation mix was packaged *in vitro* (13) and introduced into *Escherichia coli* ED8767. The packaged cosmids were spread, replicated (14), and screened by colony filter hybridization (15). A DNA fragment from a DR β -chain genomic clone (unpublished data) labeled by nick-translation was used for these hybridizations. Colonies that were positive after three rounds of screening were grown and cosmid DNA was prepared.

Identification of a Possible SB β cDNA Clone. The cloning of class II β -chain cDNA clones from a DR4,w6 cell line has been described (9). The cDNA clone, shown here to be an SB β -chain cDNA (see below), was characterized as a clone from that bank which did not correspond to either DR β - or DC β -chain cDNA on the basis of restriction maps and hybridization analysis.

Restriction Mapping. Maps of the cosmids were derived from standard single, double, and triple restriction enzyme digests. Southern blot analysis was performed on the same gels to identify the fragments containing gene sequences. Orientation of the genes was determined by using restriction fragments representing the 5' and 3' portions of cDNA clones specific for HLA-DR β chain (9), HLA-DR α chain (16), and the SB β chain. Occasionally a fragment from a cosmid was subcloned to obtain a more detailed map.

Nucleotide Sequence Analysis. The chemical degradation procedure according to Maxam and Gilbert (17) and the dideoxy chain-termination method (18) after subcloning in M13 vectors (19) were used to obtain the sequence information. A sequence strategy map is given (Fig. 1).

Southern Blot Analysis. Southern blot analysis (20) was as described (21). DNA from cell lines differing in their SB specificity was a generous gift of Steve Shaw and Marius Giphart. The DNA probe used was a *Bam*HI/*Eco*RI fragment corresponding to the first domain of the SB β -chain gene and flanking intron sequences. A second probe corresponding to the next *Eco*RI fragment 5' to the first probe was used in some of the studies of restriction site polymorphism. The DNA probes were labeled by nick-translation.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); MHC, major histocompatibility complex.

*Present address: Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20205.

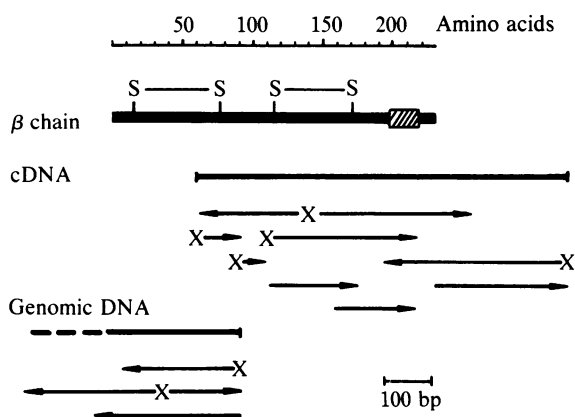


FIG. 1. Sequencing strategy for the SB β -chain gene. Sequence information was derived from a genomic subclone or from a cDNA clone. Arrows show sites used and direction of sequence. An X denotes chemical degradation; a plain arrow denotes dideoxy chain-termination. The SB β -polypeptide chain is diagrammed above. The domains are identified by the disulfide bridges and the transmembrane region is denoted by a hatched box. bp, base pairs.

RESULTS

Mapping of Cosmid Clones Containing the SB β -Chain Gene. A cosmid library of digested DNA from the homozygous B-cell line HHK was constructed by using the cosmid vector pOPF (12). The library was screened by colony filter hybridization with a probe corresponding to the first domain of an HLA-DR β -chain gene probe cloned in bacteriophage λ (unpublished data). Five of the 10 cosmids that gave a positive signal did not correspond in restriction patterns to either DR β - or DC β -chain genes and did not hybridize to 3'-untranslated probes specific for DR or DC genes. Two of these clones also showed a strong hybridization to a DR α -chain probe. The clones could be divided into two groups on the basis of their restriction pattern and strength of hybridization with the DR β -chain probe. The first group was mapped by using (i) a DR β -chain cDNA probe (9), (ii) an HLA class II β -chain cDNA probe (later identified as SB β on the basis of sequence overlap), and (iii) a DR α -chain cDNA probe (14). The map of the first set of cosmid clones is shown in Fig. 2 together with the positions of the fragments containing β - and α -chain gene sequences. The orientation of the β -chain gene (SB β I) and the α -chain gene was determined by hybridization using 5' and 3' specific probes (see *Materials and Methods*). As can be seen, the β -chain gene and one of the α -chain genes are located in a "head-to-head" orientation. A

second, more weakly hybridizing α -chain gene (α') was localized to the 3' end of the β -chain gene. The cosmids of the second group gave a clearly different restriction pattern and thus contain a related but different β -chain gene (SB β II; see below).

Sequence Analysis of the SB β -Chain Gene. A *Pst* I-*Sau*3A1 fragment corresponding to the first domain of the β -chain gene described in Fig. 2 was subcloned in M13 and sequenced. From this sequence analysis, the β -chain gene was identified as an SB β -chain gene on the basis of the homology of its deduced amino acid sequence with the partial NH₂-terminal protein sequence of the SB β polypeptide (6). At six of the seven positions determined by amino acid sequence, the sequence from the cosmid clone was identical. In one position the protein sequence showed a tyrosine, whereas the sequence of the gene coded for phenylalanine. In parallel, a class II β -chain cDNA corresponding to neither DR β - nor DC β -chain cDNA was also sequenced. It revealed a large sequence overlap, which permitted us to conclude that the cDNA sequence was also that of an SB β -chain gene, very likely corresponding to an allele of the SB β gene (see *Discussion*). These data are presented in the form of an overlapping sequence of the genomic and cDNA clones (Fig. 3). The positions previously determined by amino acid sequencing (6) are marked by asterisks below the nucleotide sequence.

Analysis of the SB β -Chain Gene by Southern Blot Hybridization. Southern blot hybridization allows a comparison of the restriction fragments containing the SB β -chain genes, both in the cloned cosmids and in the cellular DNA from which the clones were derived. A DNA fragment corresponding to the first domain of the SB β -chain gene was used as the probe. DNA from the cosmids corresponding to each of the two different SB β -chain genes and DNA from the B-cell line (HHK) used to construct the library were digested with *Hind*III, blotted, and probed with the SB β -chain first domain probe. Two strongly hybridizing bands were seen in the HHK cellular DNA, each corresponding to a band present in one of the SB β -chain cosmids (Fig. 4). Hybridization was strongest to the genomic band homologous to the SB β -chain gene from which the probe was derived. The band corresponding to the second cosmid group was weaker. At longer exposures faint bands corresponding to DR β -chain genes could be discerned.

In the case of HLA-DR and DC β -chain genes, an extensive restriction site polymorphism can be demonstrated by Southern blot hybridization (21). We have analyzed the SB β -chain gene with the same procedure. Genomic DNA from a number of individuals differing in their SB typing but iden-

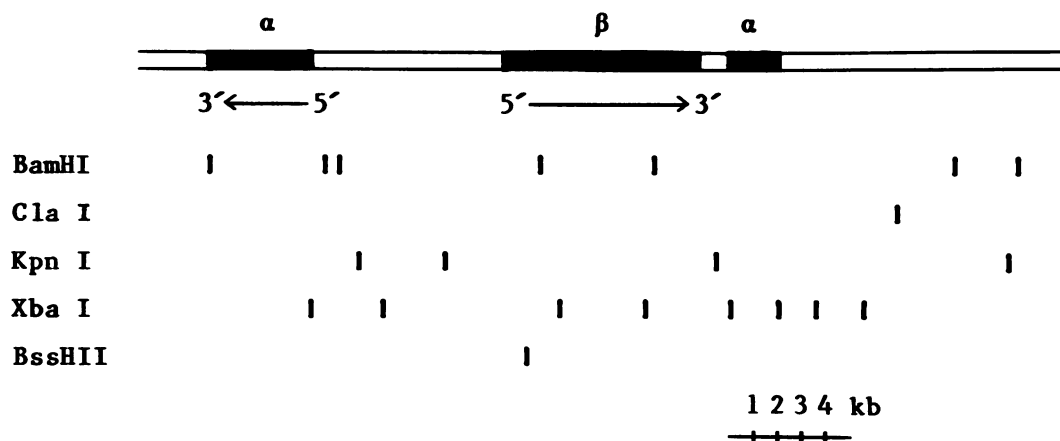


FIG. 2. Molecular map of the SB region. The restriction map is derived from three overlapping cosmids. Boxes indicate DNA fragments hybridizing to either α - or β -chain probes. Arrows below the genes indicate direction of transcription.

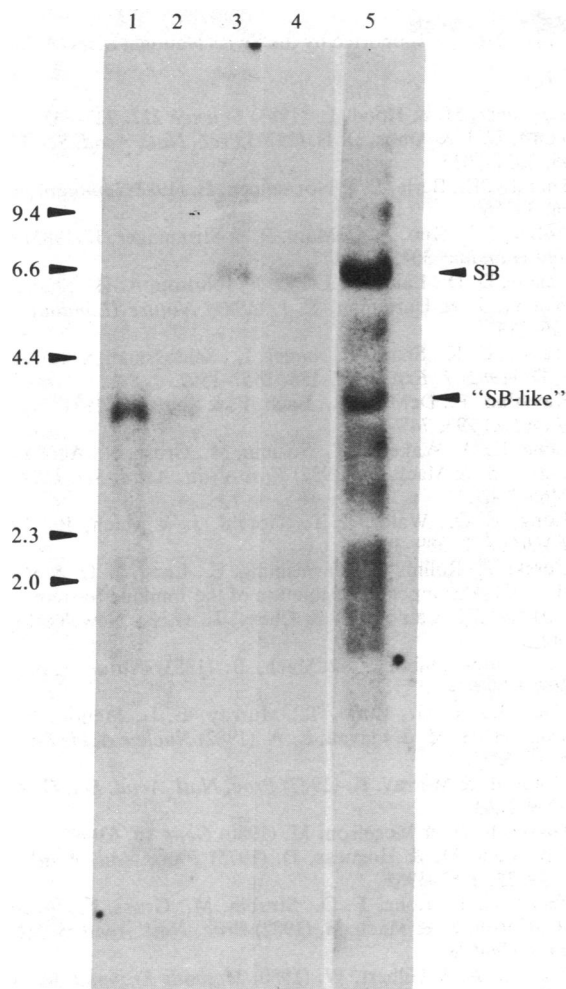


FIG. 4. Southern blot hybridization of recombinant cosmids and of cellular DNA digested with *Hind*III. Lanes 1 and 2, two cosmids of the SB-like groups (SB β II) representing separate isolates; lane 3, the cosmid from which the SB β I gene sequence was derived; lanes 4 and 5, cellular DNA (HHK), with lane 5 being a longer exposure. The markers on the left are phage λ digested with *Hind*III.

point mutation as a mechanism of divergence in the generation of SB β -chain polymorphism.

As detailed elsewhere (27), sequence comparisons between the different subregions of HLA have shown that the β -chain gene of SB is equally related to DR β and DC β (75% homology). The murine homologues of DR β and DC β are I-E β and I-A β , respectively, whereas the best hybridization to SB β was shown by a second I-E β gene (β^2 in ref. 28). Furthermore, the E β^2 gene is not restricted to a second do-

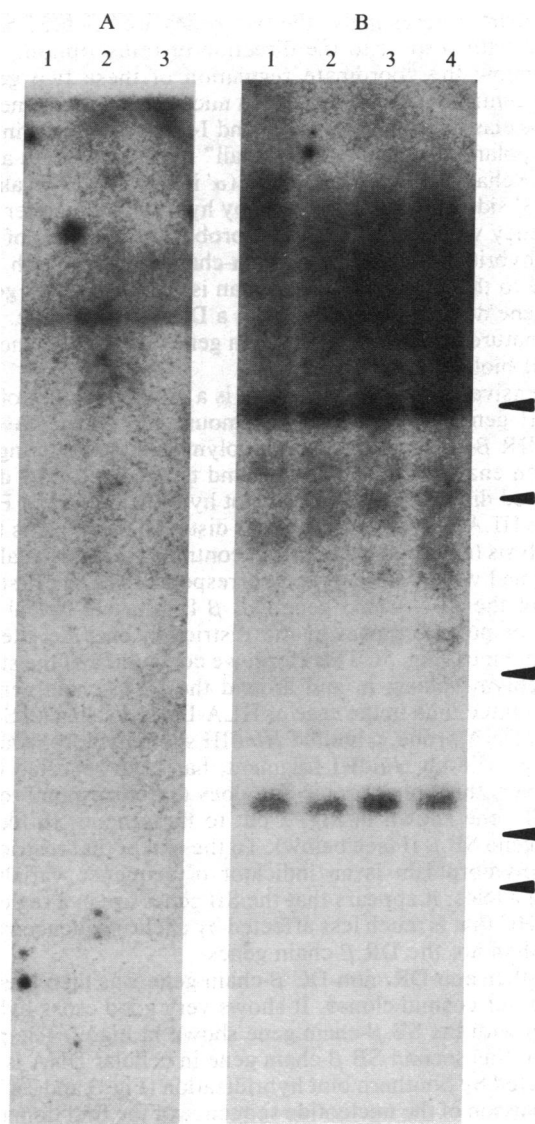


FIG. 5. Southern blot analysis of DNA from DR identical individuals varying in their SB typing. (A) DNA digested with *Bgl* I. The SB typing is as follows: lane 1, SB 1,4; lane 2, SB 2,4; lane 3, SB 3,4. (B) DNA digested with *Pvu* II. The SB typing is as follows: lane 1, SB 1,4; lane 2, SB 1,2; lane 3, SB 2,4; lane 4, SB 3,4. The arrowheads denote the positions of migration of phage λ markers after *Hind*III digestion.

main as originally thought but seems to include a β -chain first domain as well (27).

Overlapping cosmid clones have allowed us to link the SB α - and β -chain genes. The SB α - and β -chain genes are about

		66												78	
		LEU	GLU	GLU	LYS	ARG	ALA	VAL	PRO	ASP	ARG	MET	CYS	ARG	
SB β	1	CTG	GAG	GAG	AAG	CGG	GCA	CTG	CCG	GAC	AGG	ATG	TGC	AGA	
	2	---	---	---	---	---	---	---	---	---	---	G-A	---	---	
DR β	1	---	---	C--	---	---	--C	GC-	GT-	---	--CC	TAC	---	---	
	2	---	---	C--	---	---	--GC	GA-	GT-	---	--AT	TAC	---	---	
DC β	1	--A	---	-G-	GCC	---	--G	TC-	GT-	---	TC-	GTG	---	---	
	2	---	---	AG-	--A	---	--G	GC-	GT-	---	---	GTG	---	---	

FIG. 6. Comparison of SB β -chain gene sequences with DR β - and DC β -chain gene sequences. SB β 1 is from the cosmid sequence and SB β 2 is from the cDNA sequences shown in Fig. 3. DR β 1 (9) and DR β 2 (unpublished data) were determined in our laboratory. DC β 1 (24) and DC β 2 (25) are from published sequences.

10 kb apart. Interestingly, the two genes are oriented head-to-head with respect to the direction of transcription. This could allow the coordinate regulation of these two genes from a central promoter region. In mice the α and β genes of the two class II subregions, I-E and I-A, are oriented in opposite polarities but in a "tail-to-tail" manner (28). An additional α -chain gene or pseudogene (α' in Fig. 1) was localized to the 3' side of the β -chain gene by hybridization under low stringency with the DR α cDNA probe. On the basis of this weak hybridization this second α -chain gene is much less related to the DR α -chain gene than is the SB α -chain gene. This gene does not hybridize with a DC α -chain probe. The exact nature of this second α -chain gene and its relevance to class II biology is unclear.

Extensive allelic polymorphism is a striking feature of the class II genes in both man and mouse (1). In the case of HLA-DR β -chain genes, allelic polymorphism affecting restriction enzyme sites in and around the genes can be demonstrated directly by Southern blot hybridization (21). Each known HLA-DR haplotype can be distinguished by this type of analysis (unpublished data). In contrast, the same analysis performed with a DNA probe corresponding to the first domain of the SB β -chain gene (SB β I) failed to reveal evidence of polymorphism in the restriction enzyme sites of four enzymes (Fig. 5). Therefore, we conclude that the structural polymorphism in and around the SB β -chain gene is more limited than in the case of HLA-DR. By using an SB β -chain cDNA probe, a limited *Hind*III site polymorphism, involving a 5.8-kb *Hind*III fragment, has been reported (26). However, this polymorphic site does not correspond to the SB β I gene shown in Fig. 2 but to the second SB-like β -chain gene SB β II (see below). To the extent that restriction site polymorphism is an indicator of sequence variability among alleles, it appears that the SB genes are in a region of the MHC that is much less affected by allelic sequence variations than are the DR β -chain genes.

Another non-DR, non-DC β -chain gene was identified on two of our cosmid clones. It shows very good cross-hybridization with the SB β -chain gene shown in Fig. 2. The presence of this second SB β -chain gene in cellular DNA is corroborated by Southern blot hybridization (Fig. 3 and ref. 27). Comparison of the nucleotide sequence of the first domain of this second SB β -chain gene to that of SB β I, DR β , and DC β (10) shows unequivocally that this second SB β -chain gene belongs to the SB family. Consequently, we refer to this second SB gene as SB β II. The existence of this gene expands the number of SB-related β -chain loci to two. Following the discovery of multiple nonallelic β -chain genes in the case of HLA-DR (9) and HLA-DC (10), this finding of two loci for SB β -chains illustrates the unexpected genetic complexity of human class II genes. The correlation between the phenotypic effects attributed to that region of the MHC and each of the different class II genes has become a real challenge. The cloning and identification of the SB β genes allows us now to study the expression and function of these genes by DNA-mediated cell transformation and thus to approach the question of their individual role in controlling the immune response.

We thank Marius Giphart and Steve Shaw for cell lines and DNA, Jack Strominger and colleagues for making available a preprint of their work, and Chantal Mattmann for excellent technical assist-

ance. This work was supported by the Swiss National Science Foundation.

- Steinmetz, M. & Hood, L. (1983) *Science* **222**, 727-733.
- Yunis, E. J. & Amos, D. B. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 3031-3035.
- Thorsby, E., Berle, E. & Nousiainen, H. (1982) *Immunol. Rev.* **66**, 39-56.
- Auffrey, C., Kuo, J., DeMars, R. & Strominger, J. (1983) *Nature (London)* **304**, 174-177.
- Eckels, D. D., Lake, P., Lamb, J., Johnson, A. H., Shaw, S., Woody, J. & Hartzman, R. J. (1983) *Nature (London)* **301**, 716-718.
- Hurley, C. K., Shaw, S., Nadler, L., Schlossman, S. & Capra, J. D. (1982) *J. Exp. Med.* **156**, 1557-1562.
- Kavathas, P., DeMars, R., Bach, F. & Shaw, S. (1981) *Nature (London)* **293**, 747-749.
- Long, E. O., Wake, C. T., Strubin, M., Gross, N., Accolla, R., Carrel, S. & Mach, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7465-7469.
- Long, E. O., Wake, C. T., Gorski, J. & Mach, B. (1983) *EMBO J.* **2**, 389-394.
- Gorski, J., Rollini, P., Kawashima, E., Long, E. O. & Mach, B., UCLA Symposia: Regulation of the Immune System, eds. Serkarz, E., Cantor, H. & Chess, L. (Liss, New York), in press.
- Rabourdin-Combe, C. & Mach, B. (1983) *Nature (London)* **303**, 670-674.
- Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10**, 6715-6732.
- Hohn, B. & Murray, K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3259-3263.
- Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63-67.
- Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961-3965.
- Wake, C. T., Long, E. O., Strubin, M., Gross, N., Accolla, R., Carrel, S. & Mach, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6979-6983.
- Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- Gronenborn, B. & Messing, J. (1978) *Nature (London)* **272**, 375-377.
- Southern, E. (1975) *J. Mol. Biol.* **98**, 503-517.
- Wake, C. T., Long, E. O. & Mach, B. (1982) *Nature (London)* **300**, 372-374.
- Auffrey, C., Lillie, J. W., Arnot, D., Grossberger, D., Kappes, D. & Strominger, J. L. (1984) *Nature (London)* **308**, 327-333.
- Trowsdale, J., Lee, J., Carey, J., Grosveld, F., Bodmer, J. & Bodmer, W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1972-1976.
- Larhammar, D., Hyldig-Nielsen, J. J., Serenius, B., Andersson, G., Rask, L. & Peterson, P. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7313-7317.
- Larhammar, D., Hammerling, U., Denaro, M., Lund, T., Flavell, R. A., Rask, L. & Peterson, P. A. (1983) *Cell* **34**, 179-188.
- Roux-Dosseto, M., Auffrey, C., Lillie, J. W., Boss, J. M., Cohen, D., DeMars, R., Mawas, C., Seidman, J. G. & Strominger, J. L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6036-6040.
- Long, E. O., Gorski, J. & Mach, B., *Nature (London)*, in press.
- Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Frelinger, J., Wake, C., Long, E., Mach, B. & Hood, L. (1982) *Nature (London)* **300**, 35-42.