Identification of multiple HTF-island associated genes in the human major histocompatibility complex class III region

Carole A.Sargent[1], Ian Dunham[2] and R.Duncan Campbell

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

[1]Present address: Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
[2]Present address: Department of Genetics, Washington University Medical School, 4566 Scott Avenue, St Louis, MO 63110, USA

Communicated by A.Williams

Chromosome walking in the major histocompatibility complex (MHC) class III region has resulted in the isolation of 541 kb of genomic DNA in two sets of overlapping cosmid clones. These two sets encompass the 340 kb separating the *C2* and *tumour necrosis factor* (*TNF*) α and β genes, except for a 22 kb gap 108 kb centromeric to the *TNF*α gene. The genomic DNA inserts have been characterized for the presence of clusters of restriction sites with CpG dinucleotides in their recognition sequence. In conjunction with pulsed field gel electrophoresis the exact sites which cleave in chromosomal DNA have been established and this has suggested the presence of a number of HTF-islands. Genomic probes flanking the HTF-islands have been hybridized to Northern blots of RNA from a number of cell lines. Transcripts ranging in size from 0.6 to 6 kb corresponding to the products of 12 novel, single copy genes have been identified. In addition the human equivalent of the murine *B144* gene was mapped ~10 kb centromeric of the *TNF*α gene. The location of so many new genes in this region raises the question as to whether they play any role in the observed HLA associations with an individual's susceptibility to develop autoimmune disease.

Key words: MHC/HTF-islands/novel genes/class III region

Introduction

The major histocompatibility complex (MHC) consists of three linked gene clusters, which in humans comprise the HLA region of chromosome 6 (Olaisen *et al.*, 1987). The class I and class II regions encode polymorphic cell surface molecules that are involved in the presentation and recognition of foreign antigens during immune regulation (Strachan, 1987; Trowsdale, 1987; Davis and Bjorkman, 1988). In the human MHC the class I and class II regions are separated by ~1000 kb of DNA termed the class III region (Carroll *et al.*, 1987; Dunham *et al.*, 1987). This segment of DNA contains a number of unrelated genes which include the loci encoding the complement components C2, Factor B and C4 (Carroll *et al.*, 1984), the enzyme cytochrome P-450 steroid 21-hydroxylase (CYP21) (Carroll *et al.*, 1985; White *et al.*, 1985), and the cytokines tumour necrosis factors (TNF) α and β (Spies *et al.*, 1986; Carroll *et al.*, 1987; Dunham *et al.*, 1987). Studies using pulsed field gel electrophoresis (PFGE) have established linkage of these genes with the flanking class I and class II loci. The class II gene *DRA* lies ~350 kb from the *CYP21B* gene, the *C2* gene is ~350 kb from the *TNF* genes, and these in turn lie ~220 kb from the class I gene *HLA-B* (Dunham *et al.*, 1987). Lately, nine new genes within the human MHC class III region have been described. The *RD* gene, so called because the predicted protein has an unusual repeating unit, lies adjacent to the *Factor B* gene (Levi-Strauss *et al.*, 1988), and a duplicated locus encoding the major heat shock protein HSP70 lies 92 kb telomeric to the *C2* gene (Sargent *et al.*, 1989). A further six genes have been positioned in a 160 kb segment of DNA around the *TNF* genes (Spies *et al.*, 1989).

The MHC is of particular interest as susceptibility to >40 autoimmune diseases is influenced by this segment of the genome (Moller, 1983; Tiwari and Terasaki, 1985; Batchelor and McMichael, 1987; Todd *et al.*, 1988b). Although direct involvement of the known MHC gene products has been proposed (Todd *et al.*, 1987, 1988a,b; Batchelor and McMichael, 1987), in many cases the molecular basis for these associations is far from clear. Population studies have suggested that important disease susceptibility factors may reside within the class III region. In order to address this issue, 541 kb of genomic DNA containing the complement and *TNF* genes have been cloned in overlapping cosmids.

To search for the presence of novel loci in the cloned DNA, the approach chosen was to define the position of CpG-rich sequences called HTF (*Hpa*II Tiny Fragment) islands, since HTF-islands are frequently associated with the 5′ ends of housekeeping, and also some tissue-specific, genes (Bird, 1986, 1987; Gardiner-Garden and Frommer, 1987). One of the characteristics of HTF-islands is that they contain a high density of non-methylated CpG dinucleotides and can be detected in chromosomal DNA as clustered sites for certain infrequently cutting restriction enzymes (Brown and Bird, 1986; Lindsay and Bird, 1987). Three of the most useful diagnostic enzymes for HTF-island mapping are *Bss*HII, *Eag*I and *Sac*II (Lindsay and Bird, 1987). Therefore, these were chosen for characterizing the class III region of the MHC for the presence of putative HTF-islands. Once the distribution of potential HTF-islands was known, flanking genomic probes were isolated from the cloned DNA to investigate whether the HTF-islands were associated with genes. The probes were hybridized to Northern blots of total RNA from a panel of tissue culture cell lines and the products of 12 novel genes were identified.

Results

Chromosome walking

Two sets of overlapping cosmid clones were isolated by chromosome walking within the class III region of the human MHC (Figure 1). The first was initiated from the

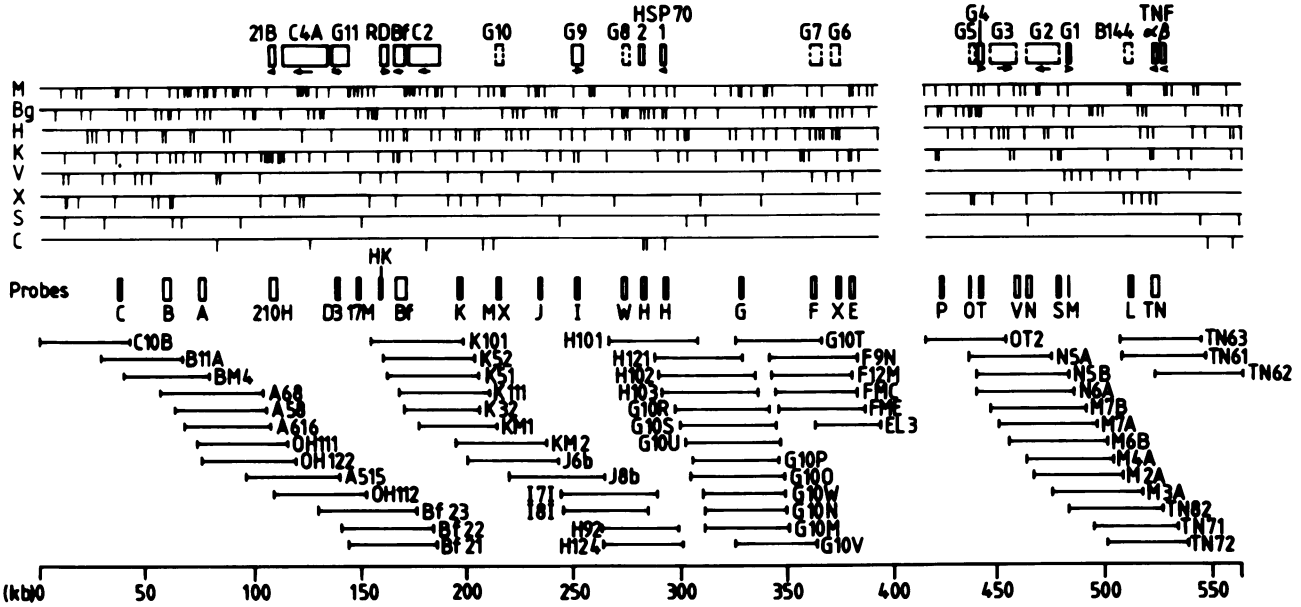

Fig. 1. Molecular map of 563 kb of the HLA class III region. Two sets of overlapping cosmid clones containing the complement−CYP21 and the TNF genes, and covering 541 kb of genomic DNA were isolated by successive rounds of screening of the cosmid libraries using the probes for *Factor B* (Bf), *CYP21* (21OH) and *TNFα*(TN), and the walking probes A, B, C, K, J, I, H, G, F, E, M, N and O. The positions of the probes used in the chromosome walking, and also in the PGFE and Northern blot analysis are indicated, and are described in detail in Materials and methods. The initial letter(s) of the designation of the cosmid refers to the probe used to isolate the cosmid. The limits of the cosmid inserts are indicated by the horizontal bars. The genomic inserts were mapped with the enzymes *Bam*HI (M), *Bgl*II (Bg), *Hin*dIII (H), *Kpn*I (K), *Eco*RV (V), *Xho*I (X), *Sal*I (S) and *Cla*I (C) in single and double digest combinations. Open boxes at the top of the figure represent the locations of genes. The direction of transcription of the genes, represented by arrows, was defined by sequence analysis of cDNA and genomic fragments. Dashed lines indicate that the limits of the gene have not been defined. Characterization of the genes and mRNA will be published elsewhere.

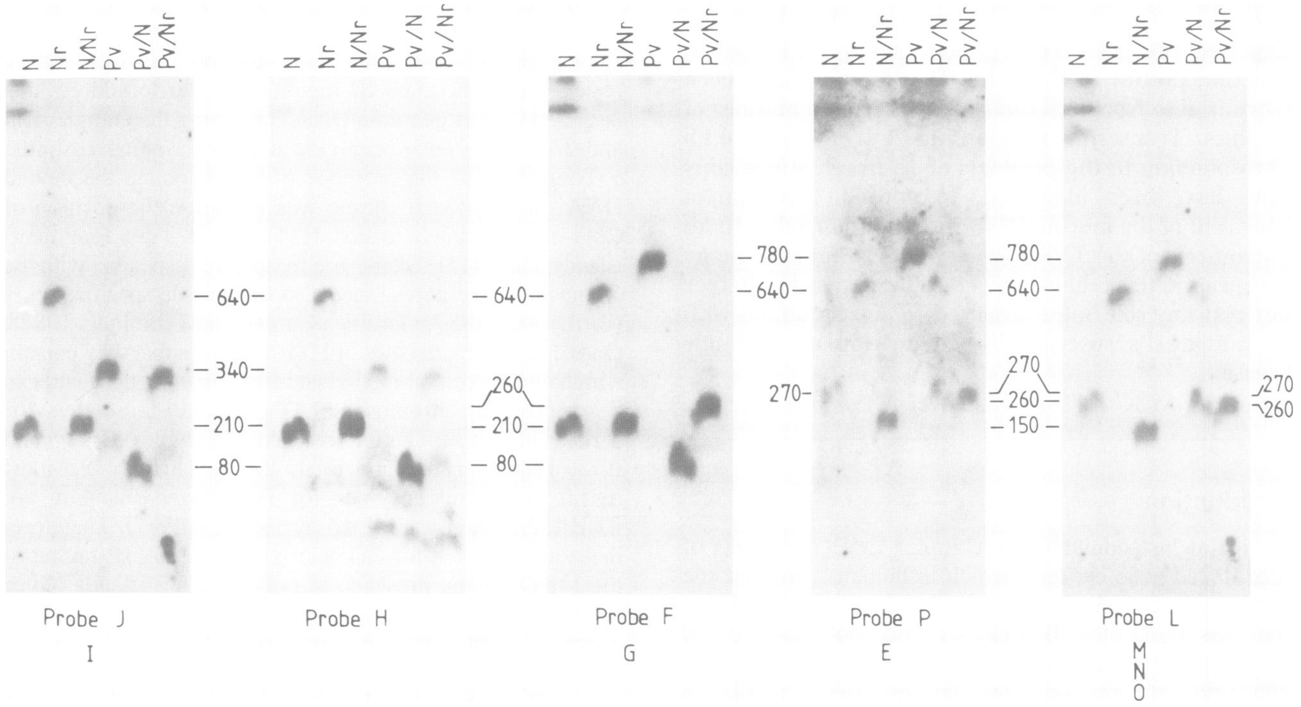

Fig. 2. Linkage of walking probes. The various genomic probes used in the chromosome walk between the *C2* and TNF genes were hybridized sequentially to a single Southern blot of genomic DNA digested with *Not*I (N), *Nru*I (Nr) and *Pvu*I (Pv) in single and double digest combinations. The digests were separated on an OFAGE system using a pulse time of 65 s. The probes used are indicated below each panel. Where two probes gave the same result only one is shown. The locations of the walking probes in the molecular map can be found in Figure 1. The numbers represent the size (in kb) of the fragments which were detected. All of the probes hybridized to a common 640 kb *Nru*I fragment.

complement−CYP21 gene cluster using probes for the *Factor B* and *CYP21* genes, and a 1.6 kb genomic fragment (probe K) situated ~6.6 kb telomeric to the *C2* gene. In total, 13 recombinants spanning 168.5 kb were recovered using these probes (Figure 1). Characterization of these cosmids showed that the gene order and physical distances of the complement and *CYP21* genes in the HLA haplotype A2 C7 B7 C2C BfS C4A3 BQO DR2 were consistent with

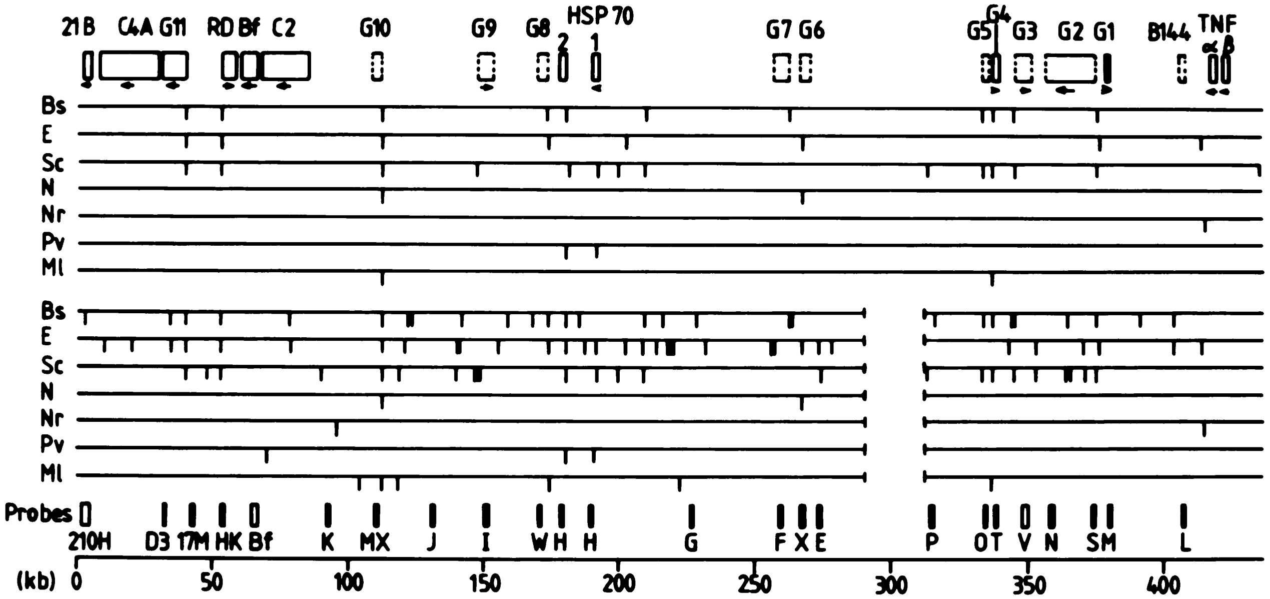

Fig. 3. Molecular maps illustrating the locations of sites for infrequently cutting restriction enzymes cleaved in (**A**) genomic DNA and (**B**) cosmid DNA. (**A**) The probes shown at the bottom of the figure were hybridized to Southern blots of genomic DNA that had been digested with *Bss*HII (Bs), *Eag*I (E), *Sac*II (Sc), *Not*I (N), *Nru*I (Nr), *Pvu*I (Pv) and *Mlu*I (Ml) and separated on a crossed field gel electrophoresis apparatus (Waltzer) using a switching interval appropriate for the fragments of interest. The autoradiographs of Southern blots of *Bss*HII- and *Sac*II-digested DNA after hybridization with the various probes are shown in Figure 4. (**B**) Cosmid DNA was digested with the infrequently cutting restriction enzymes singly or in double digests with *Bam*HI or *Bgl*II, and the rare enzyme sites were placed on the molecular map on the basis of the double digest products. Many of the recognition sequences constitute clusters of sites within 1–2 kb of DNA, which also cleave in chromosomal DNA, suggesting the presence of a number of HTF-islands between the *C4* and TNF genes.

the established map (Carroll *et al.*, 1984), allowing for the deletion which encompasses the *CYP21A* and *C4B* genes in this haplotype (Yu and Campbell, 1987). Successive rounds of screening of the cosmid libraries with the walking probes A, B and C increased the extension to ~102 kb from the 3′ end of the *CYP21B* gene, while a chromosome walk in the opposite direction using the probes J, I, H, G and F increased the extension to ~207 kb from the 5′ end of the *C2* gene (Figure 1). The total amount of cloned DNA in the 45 cosmids encompassing the *CYP21B* and complement genes was ~393 kb.

The second chromosome walk was initiated by screening the cosmid libraries with a genomic probe for TNFα. The nine positives isolated spanned ~82.5 kb around the *TNFα* locus, extending ~40 kb from both the 3′ and 5′ ends of the gene (Figure 1). The position of the *TNFβ* locus was deduced from comparison of the restriction data with the published maps of Nedospasov *et al.* (1985, 1986). The sites for the restriction enzymes *Bam*HI, *Hin*dIII, *Kpn*I and *Xho*I were identical to those mapped in the published bacteriophage λ genomic clones containing the *TNFα* and *TNFβ* genes, and TNFβ was therefore concluded to lie telomeric to the TNFα locus (Figure 1). A chromosome walk from the centromeric endpoint of the cloned region towards the cosmid cluster containing the complement–CYP21B genes using the probes M, N and O increased the extension to ~108 kb from the 3′ end of the *TNFα* gene (Figure 1). In total the 16 recombinants in the TNF cosmid cluster encompassed 148 kb of DNA.

Linkage of the cosmid clusters by PFGE

All of the walking probes isolated from the genomic inserts were linked to pre-existing class III markers by PFGE (Figure 2). Each was hybridized to a Southern blot of chromosomal DNA digested with *Not*I, *Nru*I and *Pvu*I in single and double digest combinations. The blot was stripped between each probing as described in Materials and methods. Hybridization to common restriction fragments confirmed that the cosmid genomic inserts were all derived from contiguous segments of the class III region of the MHC. All of the walking probes hybridized to a common 640 kb *Nru*I fragment (Figure 2) which also contains the complement genes (Dunham *et al.*, 1987). In the *Not*I digest probes J, I, H, F and G hybridized to a 210 kb fragment, while probes E, P, O, N, M and L hybridized to a 270 kb fragment in common with the TNFα probe (Figure 2). This is consistent with the mapping of two *Not*I sites in the cloned DNA, one ~25 kb telomeric of the *C2* gene and the second ~150 kb centromeric of the *TNFα* gene (Figure 3). In the *Pvu*I digest probes F, G, P, E, L, M, N and O hybridized to a 780 kb fragment which also contains the TNF genes and extends into the class I region (Dunham *et al.*, 1987), while probes J and I hybridized to the same 340 kb fragment that contains the complement genes (Figure 3). Probe H, however, hybridizes to both fragments as it is derived from a duplicated region encoding the major heat shock protein HSP70 (Sargent *et al.*, 1989). The results from all the single and double digests agree with the previously published map of this region (Dunham *et al.*, 1987; Sargent *et al.*, 1989). Thus using the orthogonal field gel electrophoresis (OFAGE) system, the gap between the *C2* and *TNFα* loci was sized at ~390 kb. However, the restriction fragment sizes observed with the crossed field gel electrophoresis (Waltzer) system of Southern *et al.* (1987) were found to be consistently ~10–20% smaller than those observed using the OFAGE system and were found to match more closely with the sizes of the fragments predicted from the cloned DNA. The results from the Waltzer system suggested that the distance between the *C2* and *TNFα* genes was closer to 340 kb (Figure 3). Of this ~315 kb was already cloned in

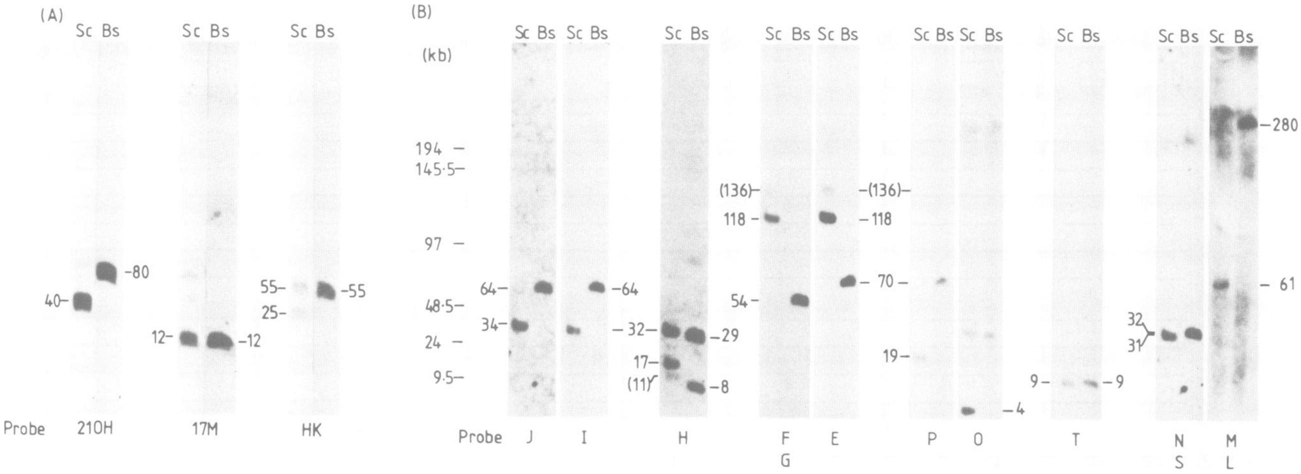

Fig. 4. Genomic Southern blot analysis of *Bss*HII (Bs) and *Sac*II (Sc) digested DNA separated by crossed field gel electrophoresis (Waltzer) at a 7.5 s switching interval. (**A**) represents results obtained using probes from around the complement genes, while (**B**) shows results using probes from between the *C2* and TNF genes. The probes used are indicated below each panel. Where two probes gave the same result only one is shown. The location of the probes in the molecular map can be found in Figure 1. The numbers represent the size (in kb) of the fragments detected. The positions of co-migrating markers (concatemers of λ DNA and *Hin*dIII digest of λ DNA) in (**B**) are shown at the left. Fragments in brackets are the minor products of partial digestion at *Sac*II sites. Probe H, in addition to being duplicated, is from the dispersed multicopy *HSP70* gene family and hence hybridized to several fragments. For probe O the *Bss*HII fragment (estimated from digests of the cloned DNA at ~3.5 kb) has been electrophoresed off the gel. The results from this analysis were used to construct the molecular map shown in Figure 3.

the two sets of overlapping cosmids obtained by chromosome walking. In order to have an accurate estimate of the distance between the endpoints of the cloned regions, the sites for the enzymes *Bss*HII, *Sac*II and *Eag*I were mapped at both the chromosomal level and within the cosmid inserts (Figure 3). Probes E and P from adjacent ends of the cosmid clusters were hybridized to Southern blots from PFGE of DNA cleaved with these enzymes (Figure 4). The lengths of the restriction fragments spanning the uncloned portion were calculated as 70 kb for *Bss*HII, 110 kb for *Eag*I and 118 kb for *Sac*II (Figure 4). As both the cleavage points for each observed fragment could be positioned in the cloned DNA by comparison with blots hybridized to other class III region probes, the gap between the two cosmid clusters was found to be ~22 kb.

Mapping HTF-islands

In order to define the location of CpG-rich islands between the *C4* and *TNFα* genes, the cosmid clones were characterized for the presence of clusters of infrequently cutting restriction enzyme sites (Figure 3). In particular the cosmids were mapped for the endonucleases *Bss*HII, *Eag*I and *Sac*II which have 6 bp recognition sequences comprised entirely of C and G and containing two CpG dinucleotides, and which are often located in HTF-islands (Lindsay and Bird, 1987). A surprisingly large number of sites were found for these endonucleases within the cloned DNA; at least 28 sites for *Bss*HII, 33 for *Eag*I and 24 for *Sac*II (Figure 3B). Furthermore many of these sites appeared to be clustered, with two or more sites occurring over a region of 1 kb, or less, of DNA. To determine the methylation status of these sites in chromosomal DNA, single copy sequences from the cloned region were hybridized to Southern blots from PFGE. Successive probings of the same blots allowed the endpoints of the observed restriction fragments detected for *Bss*HII and *Sac*II (Figure 4), and also *Eag*I, to be accurately positioned on the molecular map (Figure 3). This analysis revealed that 18 clusters of sites, or single sites, between the *C4* and *TNFα* genes, could be restricted in chromosomal DNA, and were therefore unmethylated. No recognition sites for *Bss*HII, *Sac*II and *Eag*I were found to cleave in the segment of chromosomal DNA corresponding to the 22 kb gap between the two sets of overlapping cosmids (Figure 3).

Potential HTF-island structures are here defined as regions of DNA up to 2 kb in size containing two or more sites for infrequently cutting endonucleases shown to be cleaved at a chromosomal level, or clustered sites for a single enzyme mapped in the cloned DNA at least one of which cleaved in chromosomal DNA. By these criteria the region between the *TNFα* and *C2* genes contains 11 potential HTF-islands. The remaining sites which cleave in chromosomal DNA correspond to a single site for that enzyme in the cloned DNA. We cannot rule out the possibility that these also constitute clusters of sites for a single enzyme, but which lie very close together in the cloned DNA.

Neither the complement nor *CYP21B* genes are associated with an HTF-island (which is consistent with the tissue-specific expression of these genes). However, two potential islands were mapped between the *Factor B* and *C4A* genes. These are characterized by the presence of *Bss*HII, *Sac*II and *Eag*I sites, all of which cleave in chromosomal DNA (Figures 3 and 4A).

Searching for HTF-island-associated transcripts between the C4 and TNFα genes

Once the positions of potential HTF-islands were established, unique DNA sequences were chosen to allow a more detailed characterization of the class III region for the presence of novel genes. Probes adjacent to or including the CpG-rich sequences (Figure 3), with the exception of one CpG-rich sequence 19 kb telomeric of the *HSP70-1* gene (where no suitable genomic fragment was available), were isolated. These were hybridized with Southern blots of genomic DNA to confirm that they were non-repetitive. Probes which were moderately repetitive were pre-annealed with total genomic DNA as described by Sealey *et al.* (1985), prior to further hybridization analysis. All of the probes, with the exception of probe H (Sargent *et al.*, 1989), gave hybridization patterns

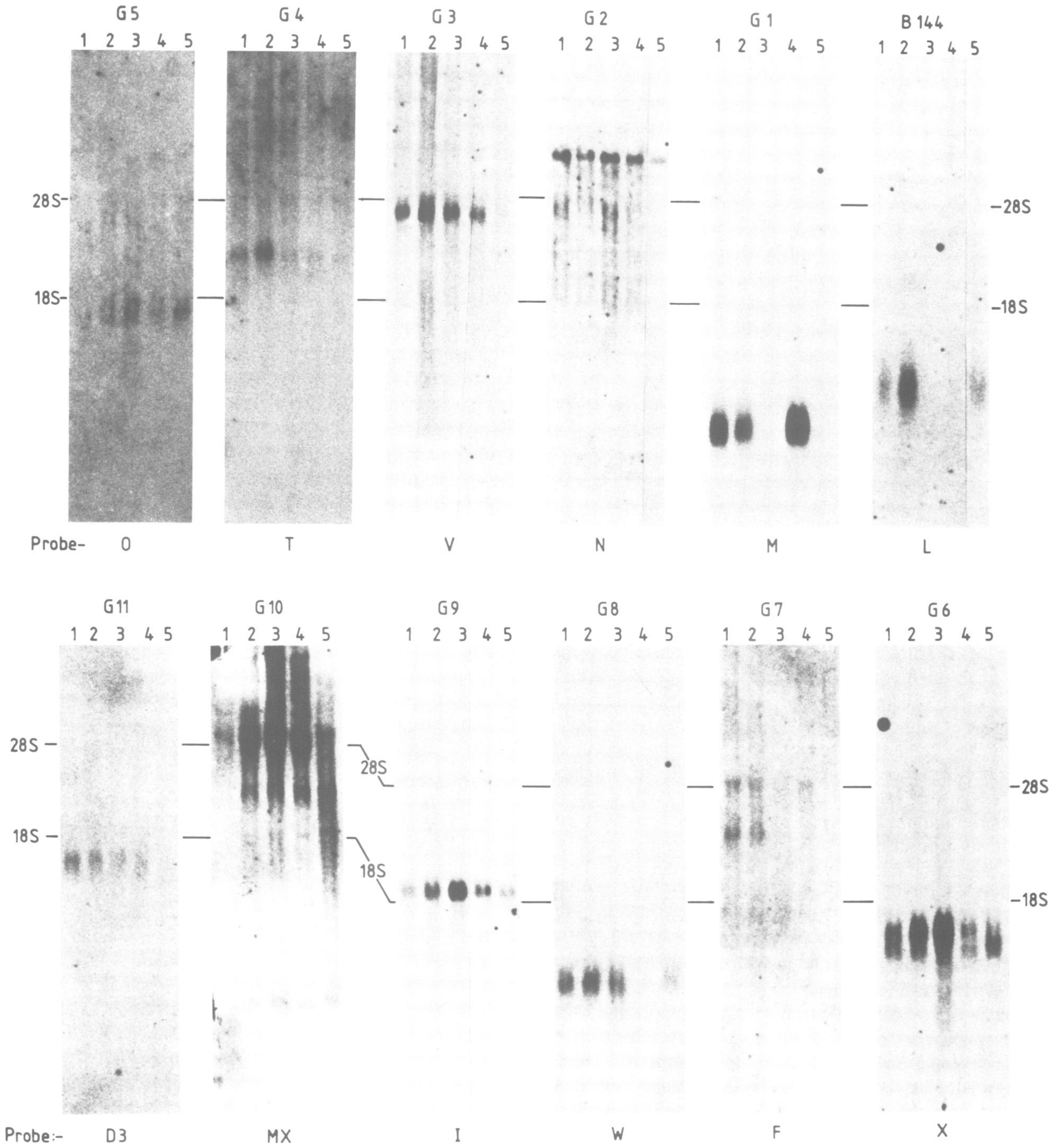

Fig. 5. Northern blot analysis for HTF-island-associated transcripts. Total cytoplasmic RNA (20 μg) derived from the cell lines (1) U937, (2) U937 stimulated with PMA, (3) HepG2, (4) Molt4 and (5) Raji were fractionated on 1% agarose−formaldehyde denaturing gels, transferred onto nitrocellulose and hybridized with the probes shown below each panel. The locations of the probes in the molecular map used in this analysis can be found in Figures 1 and 3. The position of migration of the 28S and 18S RNA is shown. A summary of the transcripts detected, which range in size from 0.6 kb (G1) to 6.0 kb (G2), and the cell lines which express them, can be found in Table I.

consistent with single-copy sequences. To search for putative coding regions the genomic fragments were hybridized to Southern blots of *Bam*HI-digested DNA from man, mouse, cat, sheep and shark. As coding sequences ought to be more highly conserved than non-coding regions, they should cross-hybridize between related species (Monaco *et al.*, 1986). Most of the genomic fragments associated with HTF-islands were found to detect restriction fragments in the other mammalian DNA samples following washing under high stringency conditions (65°C, 0.2 × SSC, 0.1% SDS) (data not shown). These fragments were chosen to study the HTF-islands between the *C4* and *TNFα* loci by Northern blot analysis, to determine whether they were indicative of novel genes with expressed transcripts (Figure 5). Each probe was used to screen total RNA isolated from a panel of tissue culture cell lines representing monocyte (U937), macrophage [U937 stimulated with phorbol 12-myristate 13-acetate (PMA)], liver (HepG2), T lymphocyte (Molt4) and B lymphocyte (Raji) lineages.

The mRNA species detected ranged from 0.6 to 6 kb in

Table I. Distribution of transcripts mapped within the cloned portion of the class III region of the human MHC

Gene	Probe	Transcript (kb)	Cell line[a]				
			1	2	3	4	5
B144	L	0.8	+	+	−	−	+
G1	M	0.6	+	+	−	+	−
G2	N	6.0	+	+	+	+	+
G3	V	3.8	+	+	+	+	+
G4	T	2.5	+	+	+	+	+
G5	O	1.6	+	+	+	+	+
G6	X	1.4, 1.5	+	+	+	+	+
G7	F	3.0	+	+	−	−	−
HSP70-1	−	2.4	N.D.				
HSP70-2	−	N.D.	N.D.				
G8	W	1.0	+	+	+	+	+
G9	I	1.9	+	+	+	+	+
G10	MX	2.6	+	+	+	+	+
RD	HK	1.6	+	+	+	+	+
G11	D3	1.4	+	+	+	+	+

The five cell lines were analysed by Northern blot analysis, and are scored for expression (+) of the mRNA species.

[a]The cell lines analysed were (1) U937, (2) U937-PMA, (3) HepG2, (4) Molt4, (5) Raji.

N.D. The expression of this transcript has not been determined in this study.

length (Figure 5, Table I), and represent the products of 12 putative loci (Figure 1). Most of the novel transcripts were expressed in all of the cell lines (Figure 5, Table I) and also HeLa cells (data not shown), as would be predicted from previous studies of HTF-island-associated transcripts. However, the quantity of some of the transcripts appeared to differ between the cell lines, even though similar amounts of total RNA were loaded in each track. This could suggest that although the transcripts represent the products of ubiquitously transcribed genes, controls over the level of mRNA expression may vary according to the cell type. Two exceptions to this general observation were the 0.6 kb transcripts designated G1, which is only seen in the cell lines Molt4, U937 and U937-PMA, and the 3 kb transcript designated G7, which is only seen in the cell lines U937 and U937-PMA (Figure 5).

Probe X, a 2.4 kb *Hin*dIII fragment which spans the *Not*I site centromeric to *TNFα*, was found to hybridize to two mRNA species in each of the cell lines (Figure 5). These were estimated at 1.4 and 1.5 kb in all but the B cell line. Here, the upper band of the doublet was slightly smaller, at ~1.45 kb in length. The doublet may infer that the transcripts are the products of alternative splicing of a single precursor mRNA, or that the HTF-island lies adjacent to two loci, perhaps arranged 5′ to 5′ (Lavia *et al.*, 1987). The slightly smaller mRNA observed in Raji cells could reflect a cell-specific difference in the processing of one of these putative gene products.

The two HTF-islands between the *C4A* and *Factor B* genes were both found to be associated with transcripts (Table I). The HK probe hybridized to a mRNA species of ~1.6 kb in all the cell lines. The size of the mRNA and the location of the probe is consistent with this HTF-island being associated with the *RD* gene (Levi-Strauss *et al.*, 1988). Probe D3 hybridized to a 1.4 kb mRNA in all the cell lines (Figure 5) and this corresponds to the transcript of a novel gene (*G11*) lying adjacent to the *C4A* gene (Figure 3).

Recently a new gene called *B144* has been mapped ~10 kb centromeric to the mouse *TNFα* gene in the H-2 complex (Tsuge *et al.*, 1987). Situated ~10 kb centromeric of the human *TNFα* gene is a unique 1.4 kb *Bam*HI fragment (probe L—Figure 1). This fragment was found to cross-hybridize strongly to sequences in mammalian DNA, but not to shark. On subsequent Northern blot hybridization a 0.8 kb mRNA was observed in U937 and PMA-stimulated U937 cells as well as faintly in Raji cells (Figure 5). No mRNA was evident in the HepG2 or Molt4 cell lines. The size of the transcript and chromosomal location of the gene are both consistent with this being the human equivalent of the murine *B144* locus.

Of the 13 class III region transcripts described here, one can be assigned to the *RD* gene (Levi-Strauss *et al.*, 1988), and another to the *B144* gene (Tsuge *et al.*, 1987; Spies *et al.*, 1989). A further two loci have been identified as members of the heat shock HSP70 family (Sargent *et al.*, 1989). At least one of the MHC-linked *HSP70* loci, here designated *HSP70-1*, is known to be expressed (Wu *et al.*, 1985). The functional status of the second gene has not been ascertained. The remaining class III region transcripts, G1−G11, are of unknown structure and function. Three, *G2, G3* and *G5*, are probably equivalent to the *Bat-2, Bat-3* and *Bat-4* genes described by Spies *et al.* (1989). Full-length cDNA clones corresponding to G11, G9, G8, G6, G4, G1 and partial cDNA clones for G10 and G2 have been isolated. It is hoped that characterization of these cDNA clones and the corresponding genes will help to define salient features of these transcripts and the protein products which might help to determine their roles *in vivo*.

Discussion

Two sets of overlapping cosmid clones have been isolated from the class III region of the human MHC. Together, the 61 genomic inserts have been characterized to produce a detailed molecular map spanning 563 kb of DNA, allowing for a gap of 22 kb as calculated from PFGE experiments. This represents just over 50% of the estimated 1000 kb which constitutes the region between the flanking class I and class II genes. When the 220 kb of cloned DNA between the *TNFβ* and *HLA-B* genes isolated by Spies *et al.* (1989) is also taken into account, this means that 783 kb of the class III region has been characterized in overlapping cosmid clones. Only the ~250 kb gap between the end of cosmid C10B and the *DRA* gene remains to be isolated.

Chromosomal DNA corresponding to the cloned portion of the class III region has been characterized by PFGE with the rarely cutting endonucleases *Bss*HII, *Eag*I and *Sac*II to define the presence of putative HTF-islands. In conjunction with restriction mapping of the cosmid genomic inserts the exact endpoints of the chromosomal fragments have been localized on the molecular map. The large number of recognition sequences for infrequently cutting endonucleases at the cloned level is similarly reflected by a high frequency of sites cleaved at the chromosomal level, and may be indicative of an overall C+G-rich character for this part of the MHC. Comparison of the average distances between the sites seen in the cosmid genomic inserts with the expected frequencies estimated by Drmanac *et al.* (1986) shows that *Not*I, *Bss*HII, *Sac*II and *Eag*I are up to 11 times more common. In contrast, restriction sites for *Pvu*I, *Nru*I and *Mlu*I (all of which contain A and T in addition to C and G in their recognition sequences) occur at closer to the expected frequencies. A similar analysis of cosmids picked at random from human chromosome 3 also showed an overabundance of rare-cutting sites (Smith *et al.*, 1987) and the human α globin locus shows a high frequency of sites for 'rare-cutters' (Fischel-Ghodsian *et al.*, 1987a,b). In addition the sites for these enzymes are not randomly distributed, but are clustered; this was also observed by Smith *et al.* (1987). In the region between *C4* and *TNFα*, only 40.7% of *Bss*HII, 25.8% of *Eag*I and 54.2% of *Sac*II sites were cut, as determined from PFGE analysis. However, the sites that are cut also lie predominantly in clusters. Subsequent characterization of the potential islands and their putative gene association suggests that the majority of sites which are unmethylated are close to coding regions.

Altogether 15 class III region genes have been defined between the *C4* and *TNFα* genes. Two correspond to the *RD* and *B144* genes, two represent members of the *HSP70* gene family, but the remaining 11 are of unknown function. The approach used here to define these novel loci limits the number of potential new genes detected to a minimum. The mapping of HTF-islands is biased towards the identification of ubiquitously expressed mRNAs, as previous studies have shown that HTF-islands are normally associated with housekeeping genes (Bird, 1987). In particular, none of the tissue-specific genes previously described in the class III region is found adjacent to an HTF-island and they would not have been detected in this study. If other tissue-specific genes do reside in this portion of chromosome 6, then the true gene density may eventually prove to be much higher. An additional method which could help to establish the total number of loci would be the hybridization of whole cosmid DNA onto panels of mRNA isolated from a wider range of tissues.

Studies of chromosome structure by staining with Giemsa or quinacrine have shown that the Q/G dark bands correspond to regions which are A+T rich, whereas the Giemsa light, or R, bands correspond to regions which are C+G rich (Comings, 1978; Holmquist *et al.*, 1982). By analysing the distribution of genes within C+G or A+T rich isochores separated by caesium chloride density gradient centrifugation, most housekeeping genes have been localized to those which are C+G rich (Goldman *et al.*, 1984). Since the estimated size of the isochores is comparable with that of chromosome bands (Bernardi *et al.*, 1985; Holmquist *et al.*, 1982), it has been suggested that isochores represent the DNA segments present within these bands. In this respect it is interesting that the MHC lies in 6p21 which is a Giemsa light band. Other properties of such segments of the genome include a high Alu content, early DNA replication in the synthetic phase of the cell cycle, and a high proportion of housekeeping to tissue-specific genes (Comings, 1978; Goldman *et al.*, 1984; Bernardi *et al.*, 1985; Korenberg and Ryowski, 1988). The class III region is evidently C+G rich and contains a number of presumed housekeeping genes, as indicated by the large number of HTF-islands and ubiquitously expressed transcripts mapped between *C4* and *TNFα*. CpG-rich islands have also been detected in association with the class I and II genes (Tykocinski and Max, 1984; Gardiner-Garden and Frommer, 1987) and unmethylated CpG islands have been shown to be present in the class I region (Pontarotti *et al.*, 1988). Thus, the class III region, and perhaps the whole of the MHC, could represent a C+G rich structure typical of gene clusters within the genome.

The class I and class II gene products, and the complement proteins and *TNFα* and *β*, are involved in the immune response. HSP70 may also play some role in this respect as it has recently been found that a peptide-binding protein, which could be involved in antigen presentation, is a member of the 70-kd heat shock protein family (Crump *et al.*, 1989). Whether the products of the newly described class III genes also participate in the immune response remains to be elucidated. In addition it will be of major interest to determine whether one or a combination of several of the gene products play any role in an individual's susceptibility to develop autoimmune disease.

Materials and methods

Preparation of cosmid libraries

Cosmid libraries were prepared after the method of Steinmetz *et al.* (1986). High mol. wt DNA was isolated from an Epstein–Barr virus-transformed lymphoblastoid cell line HLA typed as A2, C7, B7, C2C, BfS, C4A3, C4BQO, DR2 and partially digested with *Mbo*I before treatment with calf intestinal phosphatase (Boehringer) as described by Ish-Horowitz and Burke (1981). The digested DNA was size-fractionated on a 20–45% sucrose density gradient, and the 35–50 kb fraction ligated into the *Bam*HI cloning site of the cosmid vector pDVcos (Knott *et al.*, 1988). The ligations were packaged using packaging extracts prepared from *Escherichia coli* BHB 2980 and BHB 2988 (Grosveld *et al.*, 1982) prior to transduction into *E. coli* strain NM554 (a gift from N.Murray). Approximately 1.5 × 10^6 independent recombinants were constructed. A library of 1.3 × 10^5 recombinants was constructed using genomic DNA partially digested with *Bam*HI. Cosmid EL3 was isolated from a cosmid library prepared in a lorist 6 vector from *Hin*dIII-digested genomic DNA (a gift from T.Rabbitts, Cambridge) using probe E.

For the chromosome walking, unique sequence regions at the end of the cosmid inserts were identified by hybridization of radiolabelled genomic DNA to Southern blots of the restriction digests. These were recovered from LGT agarose gels and used to re-screen the library filters.

Characterization of cosmid DNA

Cosmid DNA was prepared using the alkaline lysis method of Birnboim and Doly (1979). The inserts were characterized using the restriction endonucleases *Bam*HI, *Bgl*II, *Cla*I, *Eco*RV, *Hin*dIII, *Kpn*I (or its isoschizimer *Asp*718), *Sal*I, *Xho*I and the infrequently cutting enzymes *Bss*HII, *Eag*I, *Mlu*I, *Not*I, *Nru*I, *Pvu*I, *Sac*II according to the suppliers' recommendations. Fragments were separated on 0.8% agarose gels and blotted onto nitrocellulose for analysis by hybridization with appropriate probes.

Analysis by PFGE

Southern blots for the linkage of walking probes were prepared by OFAGE as described previously (Dunham *et al.*, 1987). Blots for the analysis of HTF-island distribution were prepared from gels run using the crossed field gel electrophoresis (Waltzer) method of Southern *et al.* (1987). Fractionated DNA was transferred onto nylon membranes (Genescreen plus; New England Nuclear or Hybond-N, Amersham) as described (Dunham *et al.*, 1989). Following hybridization, high stringency washes were performed at 65°C in 0.2 × SSC, 1% SDS for 1 h prior to autoradiography at −70°C between two intensifying screens. Between rounds of re-hybridization, blots were stripped by washing at 80°C in 2 mM Tris–HCl, pH 7.4, 1 mM EDTA, 0.1% SDS for 1 h.

Preparation of RNA, fractionation and Northern blot analysis

Total RNA was extracted by guanidinium isothiocyanate lysis and caesium chloride ultracentrifugation from 2–9 × 10^7 cells grown in tissue culture (Chirgwin *et al.*, 1979; Maniatis *et al.*, 1982). 20 μg samples were fractionated in formaldehyde denaturing gels, and transferred onto nitrocellulose after the method of Fourney *et al.* (1988). Northern blots were hybridized at 42°C in 50% formamide, 5 × Denhardts, 10% dextran sulphate, 1 M NaCl, 50 mM Tris, pH 7.4, 0.1% SDS followed by high stringency washing at 65°C in 0.2 × SSC, 0.1% SDS for 1 h. Autoradiography was carried out at −70°C between two intensifying screens for between 12 h and 10 days.

Probes

Probes for *CYP21*, *C4*, *Factor B*, *TNFα* and probes H, K, J and L were as described previously (Morley and Campbell, 1984; Dunham *et al.*, 1987, 1989; Sargent *et al.*, 1989). All other single copy genomic probes shown in Figure 1 were isolated from cosmid DNA inserts using the appropriate restriction enzyme digests and are as follows: C, 2.2 kb *Nco*I−*Bgl*II fragment; B, 4 kb *Bam*HI fragment; A, 1.4 kb *Bam*HI−*Bgl*II fragment; D3, 0.5 kb *Nhe*I−*Bst*EII fragment; 17M, 1.7 kb *Bam*HI fragment; HK, 1.4 kb *Hin*dIII−*Kpn*I fragment; MX, 1.3 kb *Bam*HI−*Xho*I fragment; I, 1.6 kb *Kpn*I fragment; W, 0.9 kb *Bgl*II fragment; G, 0.6 kb *Bam*HI fragment; F, 1.3 kb *Bgl*II fragment; X, 2.4 kb *Hin*dIII fragment; E, 1.25 kb *Bam*HI−*Bgl*II fragment; P, 0.5 kb *Kpn*I−*Bgl*I fragment; O, 0.6 kb *Bgl*II fragment; T, 0.8 kb *Bgl*II fragment; V, 2.8 kb *Bam*HI fragment; N, 1.1 kb *Bam*HI−*Bgl*II fragment; S, 1.1 kb *Kpn*I fragment; M, 0.3 kb *Bam*HI−*Hin*dIII fragment. All probes were radiolabelled using the random hexanucleotide priming method of Feinberg and Vogelstein (1984).

Acknowledgements

We wish to thank Ken Reid, Wendy Thomson, Elaine Kendall and Caroline Milner for critical review of the manuscript, Michael Anderson for the *Bam*HI cosmid library, Terry Rabbitts for the *Hin*dIII cosmid library, George Brownlee and Jasper Rees for pDVcos, Noreen Murray for NM554, Beryl Moffat and David Hunter for growing the tissue culture cells, and Sandra Smith for excellent technical assistance. C.A.S. and I.D. held Medical Research Council and Science and Engineering Research Council Studentships, respectively.

References

Batchelor,J.R. and McMichael,A. (1987) *Br. Med. Bull.*, **43**, 156−183.

Bernardi,G., Olofsson,B., Filipski,J., Zerial,M., Salinas,J., Cuny,G., Meunier-Rotival,M. and Rodier,F. (1985) *Science*, **228**, 953−957.

Bird,A.P. (1986) *Nature*, **321**, 209−213.

Bird,A.P. (1987) *Trends Genet.*, **3**, 342−347.

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513−1523.

Brown,W.R.A. and Bird,A.P. (1986) *Nature*, **322**, 477−481.

Carroll,M.C., Campbell,R.D., Bentley,D.R. and Porter,R.R. (1984) *Nature*, **307**, 237−241.

Carroll,M.C., Campbell,R.D. and Porter,R.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 521−525.

Carroll,M.C., Katzman,P., Alicot,E.M., Koller,B.H., Geraghty,D.E., Orr,H.T., Strominger,J.L. and Spies,T. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8535−8539.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294−5299.

Comings,D.E. (1978) *Annu. Rev. Genet.*, **12**, 25−46.

Crump,B., van Buskirk,A., Margoliash,E. and Pierce,S. (1989) *Faseb J.*, **2**, 538a.

Davis,M.M. and Bjorkman,P.J. (1988) *Nature*, **334**, 395−402.

Drmanac,R., Petrovic,N., Glisin,V. and Crkvenjokov,R. (1986) *Nucleic Acids Res.*, **14**, 4691−4692.

Dunham,I., Sargent,C.A., Trowsdale,J. and Campbell,R.D. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7237−7241.

Dunham,I., Sargent,C.A., Dawkins,R.L. and Campbell,R.D. (1989) *J. Exp. Med.*, **169**, 1803−1816.

Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266−267.

Fischel-Ghodsian,N., Nicholls,R.D. and Higgs,D.R. (1987a) *Nucleic Acids Res.*, **15**, 9215−9225.

Fischel-Ghodsian,N., Nicholls,R.D. and Higgs,D.R. (1987b) *Nucleic Acids Res.*, **15**, 6197−6207.

Fourney,R.M., Miyakoshi,J., Day,R.S. and Paterson,M.C. (1988) *Focus*, **10**, 5−7.

Gardiner-Garden,M. and Frommer,M. (1987) *J. Mol. Biol.*, **196**, 261−282.

Goldman,M.A., Holmquist,G.P., Gray,M.C., Caston,L.A. and Abhijit,N. (1984) *Science*, **223**, 686−692.

Grosveld,F.G., Lund,T., Murray,E.J., Mellor,A.L., Dahl,H.H.M. and Flavell,R.A. (1982) *Nucleic Acids Res.*, **10**, 6715−6732.

Holmquist,G., Gray,M., Porter,T. and Jordan,J. (1982) *Cell*, **31**, 121−129.

Ish-Horowitz,D. and Burke,J.F. (1981) *Nucleic Acids Res.*, **9**, 2989−2998.

Knott,V., Rees,D.J.G., Cheng,Z. and Brownlee,G.G. (1988) *Nucleic Acids Res.*, **16**, 2601−2612.

Korenberg,J.R. and Ryowski,M.C. (1988) *Cell*, **53**, 391−400.

Lavia,P., MacLeod,D. and Bird,A. (1987) *EMBO J.*, **6**, 2773−2779.

Levi-Strauss,M., Carroll,M.C., Steinmetz,M. and Meo,T. (1988) *Science*, **240**, 201−204.

Lindsay,S. and Bird,A.P. (1987) *Nature*, **327**, 336−338.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moller,G. ed. (1983) *Immunology Review*, Vol. 70, pp. 5−218.

Monaco,A.P., Neve,R.L., Colletti-Feener,C., Bertelson,C.T., Kurnit,D.M. and Kunkel,L.M. (1986) *Nature*, **323**, 646−650.

Morley,B.J. and Campbell,R.D. (1984) *EMBO J.*, **3**, 153−157.

Nedospasov,S.A., Shakov,A.N., Turelskaya,R.L., Mett,V.A., Georgiev, G.P., Dobrynin,V.N. and Korobko,V.G. (1985) *Dokl. Acad. Nauk. SSSR*, **285**, 1487−1490.

Nedospasov,S.A. *et al.* (1986) *Cold Spring Harbor Symp. Quant. Biol.*, **L1**, 611−624.

Olaisen,B., Sakaguchi,A.Y. and Naylor,S.L. (1987) *Cytogenet. Cell Genet.*, **46**, 147−169.

Pontarotti,P., Chimini,G., Nguyen,C., Boretto,J. and Jordan,B.R. (1988) *Nucleic Acids Res.*, **16**, 6767−6778.

Sargent,C.A., Dunham,I., Trowsdale,J. and Campbell,R.D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1968−1972.

Sealey,P.G., Whittaker,P.A. and Southern,E.M. (1985) *Nucleic Acids Res.*, **13**, 1905−1922.

Smith,D.I., Golembieski,W., Gilbert,J.D., Kizyma,L. and Miller,O.J. (1987) *Nucleic Acids Res.*, **15**, 1178−1184.

Southern,E.M., Anand,R., Brown,W.R.A. and Fletcher,D.S. (1987) *Nucleic Acids Res.*, **15**, 5925−5943.

Spies,T., Morton,C.C., Nedospasov,S.A., Fiers,W., Pious,D. and Strominger,J.L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8699−8702.

Spies,T., Blanck,G., Bresnahan,M., Sands,J. and Strominger,J.L. (1989) *Science*, **243**, 214−217.

Steinmetz,M., Stephan,D., Dostoornikoo,G.R., Gibb,E. and Romaniuk,R. (1986) *Immunol. Methods*, **3**, 1−19.

Strachan,T. (1987) *Br. Med. Bull.*, **43**, 1−14.

Tiwari,J.L. and Terasaki,P.I. (1985) In *HLA and Disease Association.* Springer-Verlag, New York.

Todd,J.A., Bell,J.I. and McDevitt,H.O. (1987) *Nature*, **329**, 599−604.

Todd,J.A., Bell,J.I. and McDevitt,H.O. (1988a) *Trends Genet.*, **4**, 129−134.

Todd,J.A., Acha-Orbea,H., Bell,J.I., Chao,N., Fronek,Z., Jacob,C.O., McDermott,M., Sinha,A.A., Timmerman,L., Steinman,L. and McDevitt,H.O. (1988b) *Science*, **240**, 1003−1009.

Trowsdale,J. (1987) *Br. Med. Bull.*, **43**, 15−36.

Tsuge,I., Shen,F.-W., Steinmetz,M. and Boyse,E.A. (1987) *Immunogenetics*, **26**, 378−386.

Tykocinski,M.L. and Max,E.E. (1984) *Nucleic Acids Res.*, **12**, 4385−4396.

White,P.C., Grossberger,D., Onufer,B.J., Chaplin,D.D., New,M.I., Dupont,B. and Strominger,J.L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1089−1093.

Wu,B., Hunt,C. and Morimoto,R. (1985) *Mol. Cell. Biol.*, **5**, 330−341.

Yu,C.Y. and Campbell,R.D. (1987) *Immunogenetics*, **25**, 383−390.

Received on March 30, 1989; revised on April 28, 1989