MOLECULAR MAPPING OF THE HUMAN MAJOR HISTOCOMPATIBILITY COMPLEX

Ian Dunham

Trinity College, Oxford



A thesis submitted for the

Degree of Doctor of Philosophy

October 1988

MRC Immunochemistry Unit

Department of Biochemistry

University of Oxford

To Oj and Tigga, and also to Rob

for giving a home to one of them.

ACKNOWLEDGEMENTS

I would like to thank Duncan Campbell for his excellent supervision over the past three years and his unceasing encouragement to get on with job. I am also grateful to Carole Sargent for her willingness to collaborate in much of the work described here. Belated thanks should go to Prof. Porter for having enough faith in me to give me a studentship.

Thanks are also due to David Hunter and Duncan for growing cell lines, Beryl Moffat and Konrad Kölble for taking blood and to Konrad for being a general nuisance. I would like to thank Elaine Kendall and Wendy Thomson for proofreading, Ken Johnson for help in preparation of the figures. Thanks also to the many others who have provided advice or reagents.

Finally I wish to express my gratitude to Ian, Hilary and Joss for providing a home in the last few months and to all the members of the MRC Immunochemistry Unit for an enjoyable time, particularly those who have spent long Friday nights in the Lamb and Flag.

ABBREVIATIONS

The one and three letter codes for amino acids are as recommended by the IUPAC-IUB Commission on Biochemical Nomenclature (1969). The nomenclature of HLA genes and antigens is as recommended by Bodmer *et al*. (1988). However, of necessity, in some figures genes are denoted by their previous (Greek) designations. The nomenclature of complement components is that recommended by the World Health Organisation (1968; 1981). Restriction enzymes are referred to by the three letter nomenclature of Smith & Nathans (1973). The remaining abbreviations are listed in alphabetical order below, or are defined in the text when they are first used.

A_{260}, A_{600}, etc.	- Absorbance at 260, 600 nm etc.
APS	- Ammonium persulphate.
BCIG	- Bromochloroindoyl- -D-galactoside.
bp, kb, Mb	- Base pairs, kilobase pairs, megabase pairs.
BSA	- Bovine serum albumin.
cDNA	- Complementary DNA.
cM	- CentiMorgan.
CpG, GpC	- 5' CG 3' dinucleotide, etc.
DNA, RNA	- Deoxyribonucleic acid, ribonucleic acid.
DNase, RNase	- Deoxyribonuclease, ribonuclease.
DEAE	- Diethylaminoethyl.
DTT	- Dithiothreitol.
EDTA	- Ethylenediaminetetraacetic acid.
HLA	- Human leucocyte antigens.
Hepes	- N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid.

IPTG	- Isopropylthio- -D-galactoside.
LGT (HGT)	- Low (High) gelling temperature.
MHC	- Major Histocompatibility Complex.
mol. wt.	- Molecular weight.
mRNA	- Messenger RNA.
dNTP	- Deoxyribonucleoside triphosphate (A, C, G or T specified instead of N where appropriate).
ddNTP	- Dideoxynucleoside triphosphate. (N = A, G, C or T)
OFAGE	- Orthogonal field alternation gel electrophoresis.
21-OHase or 21-OH	- 21-Hydroxylase.
PBMN	- Peripheral blood mononuclear cell.
PBS	- Phosphate buffered saline.
PEG	- Polyethylene glycol.
PFGE	- Pulsed field gel electrophoresis.
p.f.u.	- Plaque forming units.
PIPES	- Piperazine-N,N'-bis[2-ethane-sulphonic acid].
rATP	- Adenosine triphosphate.
RFLP	- Restriction fragment length polymorphism.
SDS	- Sodium dodecylsulphate.
SLE	- Systemic lupus erythematosus.
TEMED	- N, N, N', N' - tetramethylethylene-diamine.
Tris	- Tris (hydroxymethyl) aminoethane.
u.v.	- Ultraviolet light.

ABSTRACT

Ian Dunham

Trinity College,
Oxford

Submitted for the
degree of D. Phil.

Michaelmas 1988

MOLECULAR MAPPING OF THE HUMAN MAJOR HISTOCOMPATIBILITY COMPLEX

1. Pulsed field gel electrophoresis (PFGE) and cosmid walking have been used to establish a molecular map of the human major histocompatibility complex (MHC). Single copy hybridisation probes were isolated from clusters of cosmid clones, one containing the C2, factor B, C4 and 21-hydroxylase genes, and the other containing the genes for tumour necrosis factors (TNF) α and β. These probes and HLA class I and class II gene probes were hybridised to Southern blots of genomic DNA which had been digested with infrequently cutting restriction endonucleases and separated by PFGE. The data obtained allowed the construction of a long range genomic restriction map, indicating the MHC spans 3800 kb. This map orients the complement gene cluster with respect to DRA, the C2 gene being telomeric to the 21-hydroxylase B gene. In addition, the positions of the TNF genes were defined. The DRA and 21-hydroxylase B genes are separated by no greater than 390 kb, while the distance between the C2 and TNFA genes is 325 kb. The HLA-B locus lies about 250-300kb telomeric of the TNFB gene.

2. The long range DNA organisation of the class II and class III regions in eight HLA homozygous cell lines has been analysed using PFGE. Comparison of the size of the BssHII restriction fragment observed for these cell lines and five individuals possessing one to three C4 genes, shows that the organisation of the C4 genes on each chromosome can be deduced from a single PFGE experiment. Outside of the C4 and 21-OHase loci the class III region shows a highly invariant structure, with no detectable differences in the amount of DNA present. Moreover the class III region is rich in CpG-islands, one of which has been characterised, and contains at least thirteen new genes. However, in the class II region, two differences between common haplotypes have been found. The DRw52-related haplotypes have the same DNA organisation. DR2 haplotypes possess 20-30 kb more DNA in the DRB region. DRw53 haplotypes have 100-130 kb more DNA than DRw52-related haplotypes in the region containing the DRB and DQA genes.

PUBLICATIONS

Dunham, I., Sargent, C. A., Trowsdale, J. & Campbell, R. D. Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis (1987) Proc. Natl. Acad. Sci. USA 84, 7237-7241. *

Rodrigues, N. R., Dunham, I., Yu, C. Y., Carroll, M. C., Porter, R. R. & Campbell, R. D. Molecular characterisation of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia (1987) EMBO J. 6, 1653-1661. *

Rodrigues, N. R., Dunham, I., Yu, C. Y., Carroll, M. C., Porter, R. R. & Campbell, R. D. Structure of the HLA-linked steroid 21-hydroxylase genes (1987) Biochem. Soc. Trans. 15, 656-657.

Dunham, I., Sargent, C. A., Trowsdale, J. & Campbell, R. D. Orientation and Location of the Complement Genes in the Human MHC by Pulsed-Field Gel Electrophoresis (1987) Complement 4, 152-153 (abstract).

Dunham, I., Sargent, C. A., Trowsdale, J. & Campbell, R. D. Mapping of the Human Major Histocompatibility Complex by Pulsed-Field Gel Electrophoresis (1988) in Immunobiology of HLA Vol. II: Immunogenetics and Histocompatibility (Spriger-Verlag, N. Y.).

Campbell, R. D., Dunham, I. & Sargent, C. A. Molecular Mapping of the HLA-Linked Complement Genes and the RCA Linkage Group (1988) Expl. Clin. Immunogenet. 5, 81-98. *

Dunham, I., Sargent, C. A., Dawkins, R. L. & Campbell, R. D. Direct Observation of the Gene Organisation of the Complement C4 and 21-Hydroxylase Loci by Pulsed-Field Gel Electrophoresis (1988) manuscript submitted.

Sargent, C. A., Dunham, I., Trowsdale, J. & Campbell, R. D. The human MHC contains genes for the major heat shock protein HSP70 (1988) manuscript submitted.

* Copies of these publications can be found at the back of this thesis.

C O N T E N T S

Page No.

Acknowledgements i
Abbreviations ii
Abstract iv
Publications v
Contents

CHAPTER 1 INTRODUCTION

1.1 THE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) 1
1.2 POLYMORPHISM IN THE MHC 5
1.2.1 Methods 5
1.2.2 The significance and evolution of polymorphism 6
1.2.3 Population genetics 7

1.3 THE PROTEINS AND GENES OF THE MHC 9
1.3.1 Class I 9
1.3.2 Class II 14
1.3.3 Class III 20

1.4 HLA AND DISEASE 30
1.5 TECHNIQUES 34
1.5.1 Cosmid walking 35
1.5.2 Pulsed Field Gel Electrophoresis (PFGE) 36

1.6 AIMS OF PROJECT 39

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS 40
2.1.1 Enzymes 40
2.1.2 Chemical reagents 41
2.1.3 Bacterial strains, plasmids, phage and yeast 42
2.1.4 Media 43
2.1.5 Radioactive nucleotides 44

Page No.

2.1.6 Nucleic acids 44

2.2 OVERNIGHT CULTURES 45

2.3 PREPARATION OF DNA 46
2.3.1 Standard techniques 46
2.3.2 Small scale isolation of plasmid and cosmid DNA 46
2.3.3 Large scale isolation of plasmid DNA 47
2.3.4 Isolation of mammalian chromosomal DNA in aqueous solution 48
2.3.5a Isolation of mammalian chromosomal DNA in agarose blocks 49
2.3.5b Purification of LGT agarose by treatment with DEAE-cellulose 51
2.3.6 Isolation of yeast chromosomal DNA in agarose blocks .. 51
2.3.7 Isolation of λ phage and λ DNA from λ cI857 S7 and preparation of λ concatemers 52
2.3.8 Isolation of bacteriophage gt10 DNA by the plate lysate method 54

2.4 PREPARATION OF RNA 55

2.5 DIGESTION OF DNA BY RESTRICTION ENDONUCLEASES 56
2.5.1 DNA in solution 56
2.5.2 DNA in agarose blocks 57

2.6 RADIOLABELLING OF DNA 58
2.6.1 "Fill in" reaction 58
2.6.2 Random hexanucleotide priming 58

2.7 GEL ELECTROPHORESIS OF NUCLEIC ACIDS AND RECOVERY OF DNA FROM GELS 59
2.7.1 Polyacrylamide gel electrophoresis 59
2.7.2 Agarose gel electrophoresis of DNA 59
2.7.3 Agarose gel electrophoresis of RNA 61
2.7.4 Pulsed field gel elctrophoresis of high mol. wt. DNA .. 61

2.8 DNA AND RNA TRANSFER BLOT ANALYSIS 64
2.8.1 Southern blot analysis of DNA fragments separated in agarose gels 64

Page No.

2.8.2 Northern blot analysis of RNA separated in agarose gels 65
2.8.3 Hybridisation of membranes 66
2.8.4 Washing, autoradiography and rehybridisation of membranes 66

2.9 SUBCLONING OF DNA FRAGMENTS INTO pATX 67
2.9.1 Preparation of E. coli NM 554 competent cells 67
2.9.2 Preparation of BamHI- and PvuII-cut and phosphatased pATX vector 68
2.9.3 Ligation and transformation 68
2.9.4 Analysis of recombinants 69

2.10 SCREENING OF cDNA LIBRARIES IN λgt10 70
2.10.1 Plating out of λgt10 clones 70
2.10.2 Screening of libraries 71
2.10.3 Rescreening of positives 71

2.11 DIDEOXY NUCLEOTIDE SEQUENCING 72
2.11.1 Preparation of JM 103 competent cells 72
2.11.2 Subcloning into SmaI-cut and phosphatased M13mp8 72
2.11.3 Preparation of single stranded template DNA 73
2.11.4 Sequenase reaction 74
2.11.5 Electrophoresis of sequencing reactions 75
2.11.6 Analysis of nucleotide sequence data 76

CHAPTER 3 A MOLECULAR MAP OF THE HUMAN MHC

3.1 INTRODUCTION 77

3.2 EXPERIMENTAL STRATEGY 81

3.3 RESULTS 83
3.3.1 Probes 83
3.3.2 Isolation of overlapping cosmid clones from the class III region 84
3.3.3 Orientation and location of the complement genes in the class III region 84
3.3.4 Mapping of the class II subregions 85

Page No.

3.3.5 Mapping of the TNFA gene in the MHC 88
3.3.6 Mapping of the class I region 90
3.3.7 Long Range Restriction Map of the MHC 91

3.4 DISCUSSION 91

CHAPTER 4 DIRECT OBSERVATION OF THE GENE ORGANISATION OF THE COMPLEMENT C4 AND 21-HYDROXYLASE LOCI BY PULSED-FIELD GEL ELECTROPHORESIS

4.1 INTRODUCTION 100

4.2 EXPERIMENTAL STRATEGY 103

4.3 RESULTS 106
4.3.1 Probes 106
4.3.2 Southern blot analysis of the cell lines for the TaqI polymorphism 106
4.3.3 Large DNA fragment RFLPs associated with the class III complement gene loci 107
4.3.4 Physical maps of the complement gene loci in the cell lines 110
4.3.5 Analysis of the large fragment RFLPs in DNA from individuals 111

4.4 DISCUSSION 114

CHAPTER 5 ANALYSIS OF THE VARIATION IN THE LONG RANGE GENOMIC ORGANISATION OF THE HLA CLASS II AND CLASS III REGIONS BY PFGE

5.1 INTRODUCTION 117

5.2 EXPERIMENTAL STRATEGY 119

5.3 RESULTS 120
5.3.1 Probes 120
5.3.2 PFGE Southern blot analysis 120

5.3.3 The class III region of the human MHC is highly invariant in structure outside the C4 and 21-OHase loci 121
5.3.4a Haplotype-specific differences in DNA organisation exist in the class II region 124
5.3.4b Mapping of the B8 DR3 haplotype (cell line 2) allows accurate location of the class II loci 126
5.3.4c The DR3, DR5 and DR6 haplotypes have the same DNA organisation 127
5.3.4d Evidence for more DNA in the DR2 and DR4 haplotypes ... 128

5.4 DISCUSSION 130

CHAPTER 6 CHARACTERISATION OF CpG-RICH ISLANDS WITHIN THE CLASS III REGION OF THE HUMAN MHC

6.1 INTRODUCTION 137

6.2 EXPERIMENTAL STRATEGY 140

6.3 RESULTS 141
6.3.1 Mapping of rare-cutting restriction enzyme clusters in the cloned cosmid DNA from the class III region 141
6.3.2 Mapping of "rare-cutter clusters" in class III region genomic DNA 142
6.3.3 Analysis of a CpG-rich island 27 kb telomeric of the C2 gene 145

6.4 DISCUSSION 154

CHAPTER 7 DISCUSSION

7.1 A MOLECULAR MAP OF THE HUMAN MHC 160

7.2 LARGE DNA FRAGMENT RFLPs TO DEFINE C4 GENE ORGANISATION 161

7.3 THE DNA ORGANISATION OF DIFFERENT HLA HAPLOTYPES 163

Page No.

7.4 THE CLASS III REGION CONTAINS MANY GENES AND IS CpG-RICH 164

7.5 THE SIGNIFICANCE FOR HLA-ASSOCIATED DISEASE 167

REFERENCES

PUBLICATIONS

CHAPTER 1

INTRODUCTION

1.1 THE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC)

The MHC is a cluster of loci coding for polymorphic proteins involved in the presentation of antigen to thymus-derived (T) lymphocytes. This definition refers specifically to the class I and class II loci (see below) since historically it was the study of the products of these genes that led to the discovery of this genetic complex. However, the term MHC will also be used here in reference to the chromosomal segment that contains not only the class I and class II genes, but in addition a number of unrelated genes. The distinction between the MHC as defined as a group of functionally related loci or as a specific chromosomal segment has been discussed previously (Klein, Figueroa & Nagy, 1983), but is not important here. The MHC has been studied most extensively in the mouse and in man, and it is worth considering the historical development of our present understanding of the MHCs in these two species.

It has long been known that tumours from a particular mouse will only grow in mice of the same inbred strain. Genetic analysis of tumour rejection between inbred strains demonstrated that one of the loci responsible was identical to a gene controlling the presence of a blood group antigen (Gorer, 1937). Subsequently the genetic organisation of the loci encoding the cell surface structures responsible for graft rejection was extensively characterised utilising alloantisera prepared by cross-immunisation between congenic, and later recombinant congenic,

mice. This work revealed that the mouse H-2 (Histocompatibility-2) locus was a complex of two genetically separable genes or gene clusters (H-2K and H-2D) each encoding H-2 antigens. At each locus multiple allelic variants are possible giving rise to the complex serology observed with alloantisera (reviewed in Klein, 1975). In addition genetic control of a quantitative variant in the serum of normal mice (Serum substance, Ss) was shown to be encoded by a locus lying within the H-2 (Shreffler, 1964). Serum substance was later identified as the mouse complement component C4 (Curman et al., 1975; Lachman et al., 1975; Meo, Krasteff & Shreffler, 1975).

Unrelated studies on the antibody response of inbred mouse strains to synthetic branched polypeptides showed that there was a high degree of polymorphism in the ability of different mouse strains to mount a response to antigenic challenge. It was demonstrated that this immune responsiveness was under genetic control and that the "immune response" (Ir) genes were very closely linked to the H-2 complex (McDevitt & Chinitz, 1969). McDevitt et al. (1972) then mapped the Ir gene controlling the response to a specific polypeptide to a distinct region, the I region, lying between H-2K and Ss. Further analysis of genes controlling either the response to other antigens or the determinants recognised by antisera raised between congenic mouse strains of different Ir phenotypes, subdivided the I region into several subregions (I-A, I-B, I-C, I-E and I-J). Later molecular genetic studies have established that only two of these subregions (I-A and I-E) are represented by genes and that the antisera defining the other regions are probably the result of complex interactions between the I-A and I-E gene products (reviewed in Klein et al., 1983; Hood, Steinmetz and Malissen, 1983; Travers & McDevitt, 1987). Thus the classical genetic studies defined the existence of a cluster of loci lying on chromosome

17 of the mouse (Klein & Figueroa, 1986) which display remarkable polymorphism and control both transplantation antigens and the immune response. The genetic structure of the H-2 complex is shown diagrammatically in Fig 1.1.

In parallel to the immunogenetic studies in the mouse, advances in clinical transplantation led to the discovery of a system analogous to H-2 in man, termed the HLA (human leukocyte antigens). Antibodies against leukocytes were described in the sera of polytransfused individuals, multiparous women and volunteers immunised with leukocytes (Bach & van Rood, 1976) and these alloantisera were shown to have many different antigenic specificities. Studies of natural recombinants in families, and also population genetics, mapped the leukocyte antigens to closely linked loci (HLA-A, HLA-B and much later HLA-C) and indicated that they were related to the alloantigens defined in tissue transplantation.

The human equivalent of the mouse I region (HLA-D) was defined as the region controlling the response in the mixed lymphocyte reaction (MLR). This test involves culturing allogeneic lymphocytes and measuring the proliferative response. Evidence from HLA-A and B-identical siblings that were recombinant for the MLR suggested that the locus determining MLR stimulation was linked to, but distinct from, the established HLA loci (Yunis & Amos, 1971). Further analysis of recombinants in families mapped HLA-D centromeric of HLA-A and HLA-B (Reinsmoen et al., 1977). Using the MLR, a series of alleles apparently at a single locus were defined (Bach & van Rood, 1976). Serological studies identified alloantisera that recognised determinants that cosegregated with the MLR-determined HLA-D specificities and these determinants became known as the HLA-DR (HLA-D related) antigens. Like the mouse I region, the D region has been further subdivided (into HLA-DR, DQ and DP subregions)

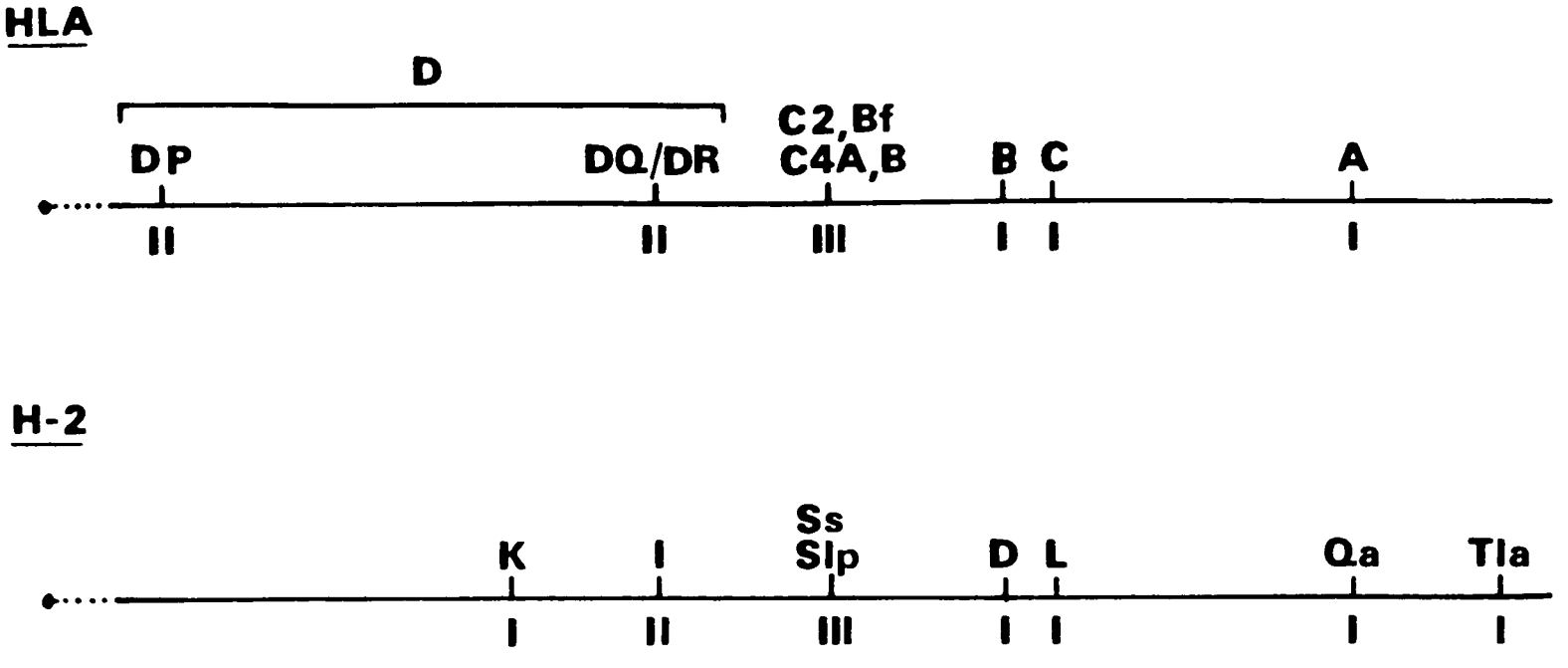


Fig 1.1 Schematic representation of the genetic organisations of the human and mouse MHCs. The centromeres are to the left. Roman numerals below each line refer to the class of molecule encoded. Ss and Slp are encoded in the murine S region.

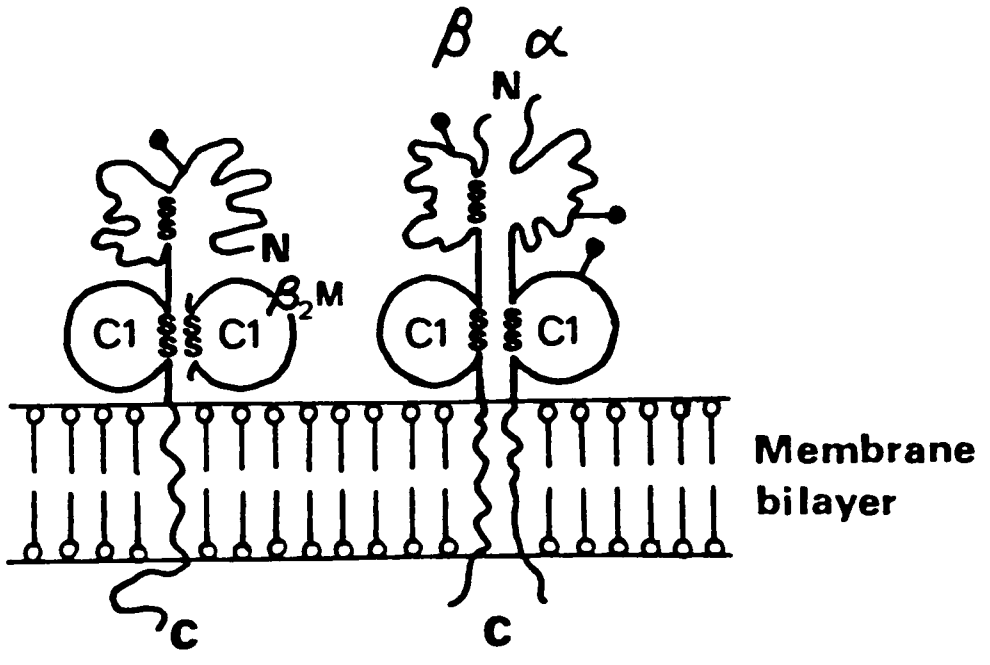


Fig 1.2 The domain organisation of the class I and class II proteins. The domains are numbered from the N terminus as described in the text. N and C indicate the amino and carboxyl terminal ends, respectively. S–S symbols represent disulphide bonds and | represents N-linked carbohydrates. C1 refers to domains with sequence and secondary structure homology with the constant domains of immunoglobulins (Williams & Barclay, 1988).

on the basis of the genetic analysis of serological phenomena and cellular typing (reviewed in Travers & McDevitt, 1987; Duquesnoy & Trucco, 1988). The DR and DQ subregions have not been separated by analysis of recombinants but the DP subregion was mapped centromeric of DR and DQ (Shaw et al., 1981).

Loci for the complement components C2, factor B and C4 (see section 1.5) have also been linked to the human MHC. Allen (1974) demonstrated that the electrophoretic polymorphism of factor B segregated with the HLA. Deficiencies of C2 and C4 in plasma linked these proteins to the MHC (Fu et al., 1974; Hauptmann et al., 1974) and this was confirmed from studies of electrophoretic variants (Alper, 1976; Teisberg et al., 1976). The electrophoretic patterns observed for C4 also indicated that there were two C4 loci, now referred to as C4A and C4B, and also led to the realisation that the Chido and Rodgers blood groups which had been linked to the HLA were distinct antigenic components of C4 (O'Neill et al., 1978a;b).

The human MHC has been mapped to the 6p21.3 band on the short arm of chromosome 6 (Francke & Pellegrino, 1977; Lamm & Olaisen, 1985). The genetic map of the region is also illustrated in Fig 1.1.

From these studies in man and mouse, it appears that the MHC linkage group has been conserved across species. Studies on the MHC in other mammalian species confirm this observation (Klein & Figueroa, 1986). A recent report on the MHC in the chicken has shown that the class I and class II loci are also linked in this species (Guillemot et al., 1988), although the relative gene organisation is considerably different to that found in mouse and man, and as yet no genes for complement components have been found in the complex.

1.2 POLYMORPHISM IN THE MHC

The most striking feature of the products of the MHC in both mouse and man is their polymorphism. Table 1.1 shows the current list of recognised HLA specificities (Bodmer et al., 1988). The nature and extent of the polymorphism of each MHC product is discussed in section 1.3. It is interesting that in addition to the HLA-A, B, C and DR loci, the C4 loci also display extensive polymorphism. At most of the polymorphic HLA loci the heterozygosity is about 90% and most of the alleles occur at a frequency of less than 0.15. In the mouse it is estimated that at least 100 alleles exist at each of the highly polymorphic loci. However in some species, for instance the Syrian hamster, there is very little MHC polymorphism (Klein & Figueroa, 1986).

1.2.1 Methods

It is worth noting the methods used to identify the polymorphic HLA alleles as these are the basis for present-day tissue typing. The initial identification of HLA polymorphism was on the basis of serological methods using alloantisera obtained as described above. Monoclonal antibodies are now widely used. Historically the MLR was the next method to be applied and this is still widely used in combination with a standard set of HLA-D homozygous-typing cells to define HLA-Dw specificities. A number of variants and refinements on the MLR have also been developed such as the primed lymphocyte test. The availabilty of T cell clones which detect single determinants on HLA molecules has increased the accuracy of the cellular typing methods. A number of biochemical methods have been applied to solubilised class I and class II molecules resulting in typing methods that are defined by the results

A	B	C	D	DR	DQ	DP
A1	B5	Cw1	Dw1	DR1	DQw1	DPw1
A2	B7	Cw2	Dw2	DR2	DQw2	DPw2
A3	B8	Cw3	Dw3	DR3	DQw3	DPw3
A9	B12	Cw4	Dw4	DR4	DQw4	DPw4
A10	B13	Cw5	Dw5	DR5	DQw5 (w1)	DPw5
A11	B14	Cw6	Dw6	DRw6	DQw6 (w1)	DPw6
Aw19	B15	Cw7	Dw7	DR7	DQw7 (w3)	
A23 (9)	B16	Cw8	Dw8	DRw8	DQw8 (w3)	
A24 (9)	B17	Cw9 (w3)	Dw9	DR9	DQw9 (w3)	
A25 (10)	B18	Cw10 (w3)	Dw10	DRw10		
A26 (10)	B21	Cw11	Dw11 (w7)	DRw11 (5)		
A28	Bw22		Dw12	DRw12 (5)		
A29 (w19)	B27		Dw13	DRw13 (w6)		
A30 (w19)	B35		Dw14	DRw14 (w6)		
A31 (w19)	B37		Dw15	DRw15 (2)		
A32 (w19)	B38 (16)		Dw16	DRw16 (2)		
Aw33 (w19)	B39 (16)		Dw17 (w7)	DRw17 (3)		
Aw34 (10)	B40		Dw18 (w6)	DRw18 (3)		
Aw36	Bw41		Dw19 (w6)			
Aw43	Bw42		Dw20	DRw52		
Aw66 (10)	B44 (12)		Dw21			
Aw68 (28)	B45 (12)		Dw22	DRw53		
Aw69 (28)	Bw46		Dw23			
Aw74 (w19)	Bw47		Dw24			
	Bw48		Dw25			
	B49 (21)		Dw26			
	Bw50 (21)					
	B51 (5)					
	Bw52 (5)					
	Bw53					
	Bw54 (w22)					
	Bw55 (w22)					
	Bw56 (w22)					
	Bw57 (17)					
	Bw58 (17)					
	Bw59					
	Bw60 (40)					
	Bw61 (40)					
	Bw62 (15)					
	Bw63 (15)					
	Bw64 (14)					
	Bw65 (14)					
	Bw67					
	Bw71 (w70)					
	Bw70					
	Bw72 (w70)					
	Bw73					
	Bw75 (15)					
	Bw76 (15)					
	Bw77 (15)					
	Bw4					
	Bw6					

The listings of broad specificities in parenthesis after a narrow specificity, e.g. HLA-A23(9) is optional. The following is a listing of these specificities which arose as clear-cut splits of other specificities.

Original broad specificities	Splits
A9	A23, A24
A10	A25, A26, Aw34, Aw66
Aw19	A29, A30, A31, A32, Aw33, Aw74
A28	Aw68, Aw69
B5	B51, Bw52
B12	B44, B45
B14	Bw64, Bw65
B15	Bw62, Bw63, Bw75, Bw76, Bw77
B16	B38, B39
B17	Bw57, Bw58
B21	B49, Bw50
B21	B49, Bw50
Bw22	Bw54, Bw55, Bw56
B40	Bw60, Bw61
Bw70	Bw71, Bw72
Cw3	Cw9, Cw10
DR2	DRw15, DRw16
DR3	DRw17, DRw18
DR5	DRw11, DRw12
DRw6	DRw13, DRw14
DQw1	DQw5, DQw6
DQw3	DQw7, DQw8, DQw9
Dw6	Dw18, Dw19
Dw7	Dw11, Dw17

The following specificities are generally agreed inclusions of HLA-B specificities Bw4 and Bw6.

Bw4: B5, B13, B17, B27, B37, B38(16), B44(12), Bw47, B49(21), B51(5), Bw52(5), Bw53, Bw57(17), Bw58(17), Bw59, Bw63(15), Bw77(15).

Bw6: B7, B8, B14, B18, Bw22, B35, B39(16), B40, Bw41, Bw42, B45(12), Bw46, Bw48, Bw50(21), Bw54(w22), Bw55(w22), Bw56(w22) Bw60(40), Bw61(40), Bw62(15), Bw64(14), Bw65(14), Bw67, Bw70, Bw71(w70), Bw72(w70), Bw73, Bw75(15), Bw76(15).

The following specificities are generally agreed to be associated with DRw52 and DRw53:

DRw52: DR3, DR5, DRw6, DRw8, DRw11(5), DRw12(5), DRw13(w6), DRw14(w6), DRw17(3), DRw18(3).

DRw53: DR4, DR7, DR9

Table 1.1 Complete listing of recognised HLA specificities (1987). Taken from Bodmer et al. (1988).

of one- or two-dimensional electrophoresis or isoelectric focussing. Recently the polymorphism of the class I and class II genes has been extensively studied by hybridising cDNA probes to Southern blots of DNA from typed individuals to establish linked RFLPs and there has been some success using RFLPs for typing (for example see Böhme et al., 1985; Bell et al., 1987a; Smeaton at al., 1987). Nucleic acid sequencing has been used to uncover the basis of the polymorphic differences between alleles (Bell et al., 1987b; Strachan, 1987) and this knowledge has been used to develop typing methods based on the polymerase chain reaction and allele-specific oligonucleotides (Saiki et al., 1986).

1.2.2 The significance and evolution of polymorphism

The reason for the high level of polymorphism in some species is still unclear. Although the MHC gene products are involved in presentation of antigen, the situation is different from the case of the T cell receptor and immunoglobulins because each individual possesses only two sets of alleles at each MHC locus. One possibility is that a given set of MHC alleles on a chromosome represents only a partial solution to the problem of dealing with all possible antigens. Each set of MHC proteins may possess a "blind spot", being unable to deal with particular foreign antigens. An individual with two different forms of allelic MHC molecules has an improved chance of mounting an immune response to a wider range of antigens. Therefore, heterozygous advantage is probably involved in selecting for and maintaining polymorphism (Klein & Figueroa, 1986). Analysis of the pattern of nucleotide substitutions in the MHC class I genes in humans and mice suggests that heterozygous advantage does play a major part in selecting for MHC polymorphism (Hughes & Nei, 1988). A second consideration is that

pathogen evolution will be guided by the need to subvert abundant allelic forms of MHC molecules so that rare alleles will tend to enjoy a selective advantage. So selection will favour multiple alleles with low individual allelic frequencies and high heterozygosities. Finally MHC polymorphism may become essential when a species expands into new territories and finds itself in new environments. It may be no coincidence that the two ubiquitous species, man and mouse, have the most polymorphic MHCs, whereas the Syrian hamster may be restricted to its small geographical area by the very limited polymorphism of its MHC.

A number of recent studies of MHC molecules at the level of nucleotide sequence indicate that the polymorphism is transpecific, i.e. it predates certain speciation events. In the mouse, generation of two groups of I-A_β alleles occurred at least 3 million years ago, before the speciation events leading to several modern Mus species (McConnell et al., 1988). Similarly allelic differences have been found that are in common between the mouse and the rat, and therefore must have arisen at least 10 million years ago (Figueroa, Gunther & Klein, 1988). Analysis of HLA-A and B nucleotide sequences compared to their analogues in the chimpanzee (ChLA) also indicates that much of the contemporary diversity in the class I genes had accumulated before the divergence of the two species 3.7-7.7 million years ago (Lawlor et al., 1988). Therefore the existence of a relatively large pool of MHC alleles may be an ancient phenomenon.

1.2.3 Population genetics

The frequencies of the HLA antigens in different ethnic populations have been extensively studied (Tiwari & Terasaki, 1985) and are known to vary considerably among different races. For instance the

frequency of HLA-B8 is high in American Caucasians (17.1%), but low in Japanese (0.2%) and Mexicans (6.0%). Each racial group has a characteristic distribution of each HLA antigen. Variation in frequency also occurs within racial groups depending on geographical area. HLA-B5 shows a south to north distribution, being highest in Middle East Caucasians and decreasing northwards.

Family studies have shown that recombination in the HLA region is rare (<1%) and thus a complete set of antigens is usually inherited as a unit. Such combinations of HLA alleles are referred to as a haplotype (Ceppellini et al., 1967). In addition population studies have shown that certain combinations of HLA alleles occur together more often than would be expected for a random mating population at Hardy-Weinberg equilibrium. For instance, in American Caucasians HLA-A1 occurs at a frequency of 0.138 and HLA-B8 at a frequency of 0.090. The observed frequency of the A1, B8 haplotype is 0.0609 which is significantly different from the expected value of 0.0124, the product of the A1 and B8 frequencies. This phenomenon is known as linkage disequilibrium. The departure from random assortment is measured by the value of the difference between the observed haplotype frequency and the expected haplotype frequency which is designated as delta (Δ) or d. In the case of A1 and B8 there is positive linkage disequilibrium and a positive Δ value (see Tiwari & Terasaki, 1985 for review). The degree of linkage disequilibrium between two alleles also depends upon the population studied (Bodmer et al., 1987). Linkage disequilibrium may be the consequence of natural selection for or against a specific gene combination or it may be due to the fact that the population has not yet reached equilibrium. For very closely linked genes the rate of approach to equilibrium is slow. The stable relationships between alleles has been taken to suggest that these combinations represent preserved

ancient or ancestral haplotypes (Dawkins et al., 1983).

1.3 THE PROTEINS AND GENES OF THE MHC

The MHC gene products can be divided into three classes depending on their structures and functions. The class I and class II molecules are integral membrane glycoproteins involved in presentation of antigen to T cells. The class III molecules include a number of proteins such as the serum complement components C2, factor B and C4, the microsomal enzyme cytochrome P450 21-hydroxylase, and the cytokines tumour necrosis factors α and β (also known as cachectin and lymphotoxin, respectively), none of which are related structurally to the class I and class II proteins. All of the genes for the class III proteins lie between the class I and class II loci in man and mouse, in a region called the class III region.

1.3.1 Class I

1.3.1a Class I molecules

The class I antigens (HLA-A, B and C, H-2K, H-2D and H-2L) were originally identified as the determinants involved in graft rejection (Gorer, 1937) and have kept the name transplantation antigens, although both class I and class II molecules act as alloantigens during transplantation. The class I antigens are expressed on the surface of virtually all nucleated somatic cells. However, in mice, in addition to the H-2K, D and L molecules, there are a number of class I-related proteins encoded by the H-2 associated Qa2,3 and Tla regions. These are less polymorphic and have a limited tissue distribution pattern (Flaherty,

1981).

The class I antigens are responsible for the MHC restriction of T cell mediated cytotoxicity to allogeneic or virus-infected target cells (Zinkernagel & Doherty, 1974; 1979). The cells involved, the cytotoxic T lymphocytes (CTLs), recognise foreign antigen only in the context of self MHC class I molecules. Foreign antigen is processed within the infected cell and the processed peptides are presented, together with the class I protein, at the surface of the cell, to be recognised by the T cell receptor. Studies of the CTL response to influenza virus infection shows that the CTLs are raised against epitopes of the influenza nucleoprotein which is not expressed at the cell surface. In addition virus-specific CTLs will lyse uninfected cells of the appropriate class I type to which peptide fragments of the viral protein have been added (Townsend et al., 1986). These observations and the demonstration that MHC molecules have affinity for peptides (Buus et al., 1987) suggest that the antigenic peptide is bound to the MHC molecule and this is confirmed by the crystal structure of the class I protein (see below).

An additional element in the recognition of class I molecules by T cells is the role of the CD4 and CD8 cell surface glycoproteins. Immature thymocytes are $CD4^+$ $CD8^+$. Mature T cells express only one of these molecules and the expression correlates strictly with the class of MHC protein with which the cell interacts. $CD8^+$ T cells are generally cytotoxic and recognise antigen in the context of class I. $CD4^+$ T cells are generally of the helper phenotype and recognise antigen plus MHC class II. The ability of the T cell receptor to distinguish between MHC classes depends on the CD4 and CD8 molecules (reviewed in Davis & Bjorkman, 1988; Janeway, 1988).

Each class I molecule consists of a polymorphic heavy chain (α) of

mol. wt. 45 000 associated with an invariant light chain, β_2-microglobulin, (Grey et al., 1973) which is encoded outside the MHC on chromosome 15 in man (Goodfellow et al., 1975) and chromosome 2 in mouse (Michaelson, 1981). Amino acid and nucleotide sequencing of purified protein and cDNA have shown that the class I heavy chain has a characteristic domain structure (Pleogh, Orr & Strominger, 1981). The extracellular portion can be divided into three domains (Fig 1.2). The N terminal α_1 and α_2 domains consist of 90 amino acids each and are both glycosylated. Secondary structure predictions suggest that they are composed primarily of β sheet, but have a short region of α helix. The α helix in the α_2 domain (positions 146-150) may be important for interaction with CTLs as substitutions in this region decrease the efficiency of CTL-mediated lysis (Krangel et al., 1983). The α_1 and α_2 domains contain the majority of CTL and monoclonal antibody determinants (Flavell et al., 1986; Stroynowski et al., 1985; Parham et al., 1988; Srivastava et al., 1985). The α_3 domain, like β_2-microglobulin, has strong sequence homology to the constant domains of immunoglobulins and thus both belong to the immunoglobulin superfamily (Williams & Barclay, 1988). The extracellular domains are anchored in the membrane by a short transmembrane segment and a cytoplasmic peptide of some 35 amino acids.

A soluble fragment of a class I molecule, HLA-A2, consisting of α_1, α_2 and α_3 domains and β_2-microglobulin generated by papain cleavage, has been crystallised and the 3-D structure solved by X-ray diffraction (Bjorkman et al., 1987a). The structure shows that the α_3 and β_2-microglobulin domains, which would lie closest to the membrane, do indeed have the characteristic β pleated sheet immunoglobulin constant region fold, but pair by a novel interaction. The α_1 and α_2 domains each consist of antiparallel β pleated sheet spanned by a long α helical region. These domains pair to form a single β sheet topped by the two α

helices, between which is formed a large 18 Å groove which could be the binding site for processed antigen. In fact the crystal structure has electron density of unknown origin, presumably an antigenic peptide, bound in this pocket. The 3-D structure suggests that it is this combination of self MHC molecule α_1 and α_2 domains with processed antigenic peptide that is recognised by the T cell receptor.

Amino acid and cDNA sequencing of different H-2 and HLA class I alleles has revealed that there is a high degree of variation at certain residues in the α_1 and α_2 domains (Srivastava et al., 1985; Strachan, 1987). It is thought that the polymorphism at these positions is responsible for the serologically defined polymorphism and could be the basis of the specificity of recognition by CTLs and the variation in responsiveness to particular foreign antigens. Examination of the locations of the polymorphic and T cell epitope residues in the 3-D structure of HLA-A2 has revealed that most of these residues are clustered either in the groove that probably forms the antigenic recognition site or on the edges of the site where they could be directly recognised by the T cell receptor (Bjorkman et al., 1987b; Parham et al., 1988; Davis & Bjorkman, 1988).

1.3.1b Class I genes

cDNA and genomic clones have been obtained for the class I transcripts and genes from both man and mouse (Hood et al., 1983; Flavell et al., 1986; Srivastava et al., 1985; Strachan, 1987). The domain structure of the class I protein is reflected by the intron/exon organisation of the genes. Class I genes cover approximately 3 kb and contain 8 exons. The first exon encodes the 5' untranslated region and a signal peptide of 24 amino acids. The α_1, α_2 and α_3 domains are each

encoded in separate exons of about 270 bp. The transmembrane region and part of the cytoplasmic tail are coded for by the fifth exon and the remaining exons encode the rest of the cytoplasmic region and the 3' untranslated sequence (Flavell et al., 1986; Srivastava et al., 1985).

The organisation of the class I genes has been well established in the mouse for both BALB/c ($H\text{-}2^d$), and C57BL/10 or B10 ($H\text{-}2^b$) haplotypes (Hood et al., 1983; Flavell et al., 1986; Müller et al., 1987b). The H-2K region contains two genes arranged head to tail and has been linked by cosmid walking to the I-A region. In BALB/c, the H-2D and H-2L regions contain four H-2D genes and one H-2L gene head to tail. The situation in B10 is not known. Overlapping cosmids link the H-2D and H-2L genes to the Qa region which contains 8 class I-like genes in BALB/c and 10 genes in B10. The Tla region contains the most class I-like genes and is located telomeric of Qa. In BALB/c it contains 18 genes on three clusters of overlapping cosmid clones while in B10 the Tla region is on a single cluster of cosmid clones containing 13 genes.

In man, although only three genetic loci have been defined serologically (HLA-A, B and C), hybridisation of class I cDNA probes to Southern blots of genomic DNA is consistent with the existence of a large number of class I-related sequences. Screening of cosmid libraries from several sources has identified many class I genes which are different on the basis of RFLPs (Malissen et al., 1982; Srivastava et al., 1985; Strachan, 1987). Srivastava et al. (1985) isolated more than 40 different genes in cosmid clones from three cell lines, 27 of which were derived from a single cell line. Several clones have been isolated that correspond to alleles of classical class I genes at the HLA-A, B and C loci (reviewed in Strachan, 1987). Attempts to cosmid walk between these loci have not been successful, but the relative positions of the genes have been studied using pulsed field gel electrophoresis (PFGE)

and these results are discussed in Chapter 3. Of the remaining non-classical class I genes, several have been shown to be functional genes (Koller et al., 1988; Srivastava et al., 1987; Paul et al., 1987) while many may be pseudogenes (Malissen et al., 1982; Srivastava et al., 1985; Strachan, 1987). One of the functional genes, HLA-E, has been mapped between B and A (Koller et al., 1988; Carroll et al., 1987) while another RS5 lies telomeric to HLA-A (Srivastava et al., 1987). Studies of deletion mutant cell lines suggests that the remainder of the class I-related sequences lie telomeric of HLA-A (Orr & DeMars, 1983). The location of the non-classical class I genes, their degree of nucleotide sequence divergence from the classical genes (Srivastava et al., 1987; Koller et al., 1988), and limited data on the expression of these gene products in lymphoid lineages (Paul et al., 1987; Koller et al., 1988) has suggested that these genes may be analogous to the murine Qa and Tla genes.

The number of class I genes present in individuals may vary considerably (Srivastava et al., 1985; 1987) as has been noted for different inbred strains of mice.

1.3.2 Class II

1.3.2a Class II molecules

The class II antigens were defined as the elements controlling the immune response in mice (I-A, I-E) and as the antigenic specificities recognised in the MLR in humans (DR, DQ, DP). Using the MLR and serological methods 26 recognised HLA-Dw specificities have been defined at present while 20 DR, 9 DQ and 6 DP specificities are recognised (Table 1.1, Bodmer et al., 1988). In the mouse it has been estimated

that at least 100 alleles may exist at the H-2A_β and H-2E_β loci.

Class II antigens are normally expressed only on B cells, antigen presenting cells and activated T cells although they may be expressed abnormally elsewhere (Pujol-Borrell et al., 1987). They are responsible for MHC restriction of T helper cells in the regulation of the humoral immune response (Shevach & Rosenthal, 1973; Benacerraf, 1981). T helper cells recognise antigen only in the context of MHC class II molecules. Intact antigen is internalised by accessory cells (dendritic cells and macrophages) and partially degraded. Peptide fragments of the antigen are then reexpressed on the cell surface in association with the class II molecule, and are recognised by the T cell receptor of $CD4^+$ T cells. A number of experiments have demonstrated that processing of antigen is both necessary and sufficient for presentation to MHC restricted T cells. Glutaraldehyde fixed accessory cells or membranes containing class II molecules will present antigenic peptides, but not native or denatured intact antigen, to primed T cells (Shimonkevitz et al., 1983; Watts et al., 1985). It has also been demonstrated that purified class II molecules have the ability to bind immunogenic peptides (Buus et al., 1987).

Class II proteins are membrane bound, glycosylated heterodimers consisting of an α chain of mol. wt. 33 000 and a β chain of mol. wt. 28 000, the differences in mol. wt. being mainly due to glycosylation differences. The amino acid sequence derived from protein sequencing and nucleotide sequencing of cDNA clones shows that both chains have a similar domain structure (Fig 1.2). The N terminal α_1 and β_1 domains are both about 90 amino acids in length and contain most of the polymorphic residues. The α_2 and β_2 domains are also about 90 amino acids long, have a disulphide bond, and show significant homology to immunoglobulin constant region domains (Hood et al., 1983; Kaufman et al., 1984). Class

II antigens are thus members of the immunoglobulin superfamily of proteins (Williams & Barclay, 1988). The α_2 and β_2 domains lie closest to the membrane. The C terminal ends of both polypeptide chains consist of a transmembrane segment and a cytoplasmic tail of between 10 and 20 amino acids.

Although a crystal structure is not yet available for a class II molecule, comparison of the amino acid sequences of many class I and II proteins suggests that they are related in domain structure. The class II molecule could have an antigen binding site composed of the α_1 and β_1 domains similar in three dimensional structure to that found in the class I protein (Brown et al., 1988). In the predicted structure the polymorphic residues would be distributed either in the binding cleft to interact with antigen or on the edge to interact with the T cell receptor.

Amino acid and cDNA sequencing of a number of class II alleles has demonstrated that the level of sequence variation differs markedly between the HLA-D region products. The DRα chain is constant from individual to individual whereas the DQα chain is extensively polymorphic. The DPα chain shows relatively little variation between alleles. In contrast, all the β chains are polymorphic, although DPβ is less variable than DRβ and DQβ (Trowsdale, 1987). Most of the variation is restricted to the N terminal domain of the chains and is located in so-called allelic hypervariable regions. Sequencing of cDNA clones obtained from homozygous-typing cells possessing the different serologically defined variants has established the basis of the variation and a base line of standard class II sequences for DQ molecules (Todd, Bell & McDevitt, 1987; Sinha et al., 1988) and for DR molecules (Gregersen et al., 1986; Bell et al., 1987b). For the DR proteins the structural basis of the major serological specificities

DR1-9 and the supertypic specificities DRw52 and DRw53, which are shared by DR3, -5 and -6 and DR4, -7 and -9, respectively, has been established (Bell et al., 1987b).

1.3.2b Class II genes

Both α and β chain genes have an intron/exon structure that reflects the domain organisation of the proteins. The α chain genes consist of five exons, namely a 5' untranslated region and signal sequence exon, α_1 and α_2 domain exons, an exon containing the transmembrane region, the cytoplasmic tail and part of the 3' untranslated region and a fifth exon containing the remainder of the 3' untranslated region. The β chain genes have a broadly similar organisation except that the cytoplasmic domain is split over the fourth, fifth and an additional sixth exon. The size of the introns differs markedly between genes (Trowsdale, 1987).

The genomic organisation of the H-2I region has been defined by cosmid walking in several mouse strains, BALB/c ($H\text{-}2^d$) (Hood et al., 1983), C57BL/10 ($H\text{-}2^b$) (Flavell et al., 1986), AKR ($H\text{-}2^k$) and B10.WR7 ($H\text{-}2^{wr7}$) (Steinmetz et al., 1984), and is similar in all cases. The I-A region is represented by ~50 kb containing one E_α and two E_β ($E_{\beta 1}$ and $E_{\beta 2}$) genes. The two β chain genes are arranged head to tail and in the opposite orientation to the E_α gene. The I-A region covers ~70 kb and contains one A_α gene 25 kb from the second E_β gene and two A_β genes ($A_{\beta 1}$ and $A_{\beta 2}$) in the opposite orientation to the A_α gene. A third A_β-related sequence ($A_{\beta 3}$), thought to be a pseudogene, lies at least 160 kb away and is linked to the H-2K region. Sequence analysis of the I region genes suggest that the E_α and E_β genes are homologous to the human genes for DRα and DRβ, respectively. The I-A region genes are homologous to

the genes of the human DC subregion, A_{α} being related to DQA, $A_{\beta 1}$ to DQB and $A_{\beta 2}$ to DOB. The $A_{\beta 3}$ gene is thought to be related to the DPB gene (Travers & McDevitt, 1987; Figueroa & Klein, 1986).

The molecular map of the mouse I region is not congruent with the classical genetic map deduced from recombination frequencies. Meiotic recombination is not random, but occurs frequently in four areas termed recombination hotspots. The regions of DNA containing the recombinational breakpoints have been cloned and sequenced and shown to contain either a CAGG tetramer tandemly repeated 16-20 times or tandemly repeated CAGA. It is believed that these sequences provide the fuctional basis for the enhancement of meiotic recombination at the hotspots (Steinmetz, Stephan & Lindahl, 1986; Uematsu et al., 1986; Kobori et al., 1986).

Analysis of cDNA and genomic clones has allowed the division of the HLA-D region into subregions, DR, DQ, DP and DO/DN each containing at least one α and β chain gene. Serologically defined products have only been detected for DR, DQ and DP. The relative order of the subregions and genes has been established by a combination of genetic data, cosmid cloning and PFGE mapping with chain specific probes, as centromere-DP-DNA-DOB-DQ-DRB-DRA-telomere (Hardy et al., 1986, see Chapter 3).

The DP subregion has been isolated in 100 kb of overlapping cosmid clones (Trowsdale et al., 1984a). There are a pair of genes for each of the α and β chains arranged in a head to head fashion in the order DPB2-DPA2-DPB1-DPA1. Sequence analysis suggests that the DPB2 and DPA2 genes are pseudogenes. The DPB1 gene contains a processed pseudogene for the ribosomal subunit protein L32 within the intron 5' to the α_1 encoding exon.

Two genes DNA and DOB are located in the DN/DO subregion. The DNA

gene was located by cross-hybridisation with a DRA cDNA probe (Trowsdale et al., 1983). DOB cDNA clones have been isolated from a lymphoblastoid cell line (Tonnelle, DeMars & Long, 1985) and the DOB gene has been mapped between the DP and DQ subregions using PFGE (Hardy et al., 1986). No genomic clones have been isolated that link the genes to one another or to any other subregion. There is no evidence for a protein product for either gene and they would be unlikely to constitute a heterodimer as they are not coordinately expressed at the RNA level (Tonnelle et al., 1985).

The DQ subregion, like DP, contains two pairs of related α and β chain genes. Genomic clones have been isolated that contain the DQA1/DQB1 and DQA2/DQB2 pairs of genes (Okada et al., 1985) arranged head to head, although these clusters have not been linked by chromosome walking. It is not known whether the DQA2 and DQB2 genes are expressed. Deletion mutant cell lines where DQA1 and DQB1, but not DQA2, are lost with the DR genes suggests that the DQA1 and DQB1 genes are closer to DR (Erlich et al., 1986).

DRA and DRB cDNA clones have been isolated and sequenced (Lee et al., 1982; Long et al., 1983; Bell et al., 1987b). The DR subregion contains one DRA gene, but a variable number of DRB genes dependent on the haplotype (Böhme et al., 1985; Bell et al., 1987a). There are two expressed β chain genes plus one or more pseudogenes in most haplotypes although the DR1 haplotype has only one expressed β chain gene (Trowsdale, 1987). The DR subregion has been cloned in overlapping cosmids from a DR4 haplotype (Spies et al., 1985) and a DR6 haplotype (Rollini, Mach & Gorski, 1985). These clones show that there are three DRB genes arranged head to head and an additional β_1 exon 15 kb 3' to the DRA gene. The DRA gene is located 90 kb from an intact DRB gene and is orientated in a tail to tail fashion (Spies et al., 1985). The DRB1

gene corresponds to the serologically detected DR type while DRB3 and DRB4 determine the DRw52 or DRw53 type, respectively (Trowsdale, 1987; Bell et al., 1987b). The DRA gene is located telomeric to the DRB genes (Hardy et al., 1986).

1.3.3 Class III

Loci for the complement components C2, factor B and C4 have been mapped into the MHC as described above. In man and mouse the genes could be mapped between the class I and class II loci (see above and Chapter 3). Subsequently a number of other genes have also been mapped into this region that is called here the class III region.

1.3.3a Class III proteins

C2, C4 and factor B

The complement pathway is a cascade of 13 plasma proteins under the control of at least 7 other plasma proteins and 15 membrane bound receptors, which forms the principal effector arm of the humoral immune response (reviewed in Reid, 1986; Campbell et al., 1988). The pathway can be activated by two distinct routes, the classical and alternative pathways (Fig 1.3). The classical pathway is activated by interaction of the first component C1 with the Fc regions of immunoglobulins in immune complexes or aggregates containing IgG or IgM. The alternative pathway is mainly activated by the polysaccharides of microbial cell walls and is thus part of innate immunity. The two pathways converge with the formation of the C5 convertase which cleaves C5 leading to assembly of the lytic complex via the terminal components of the pathway.

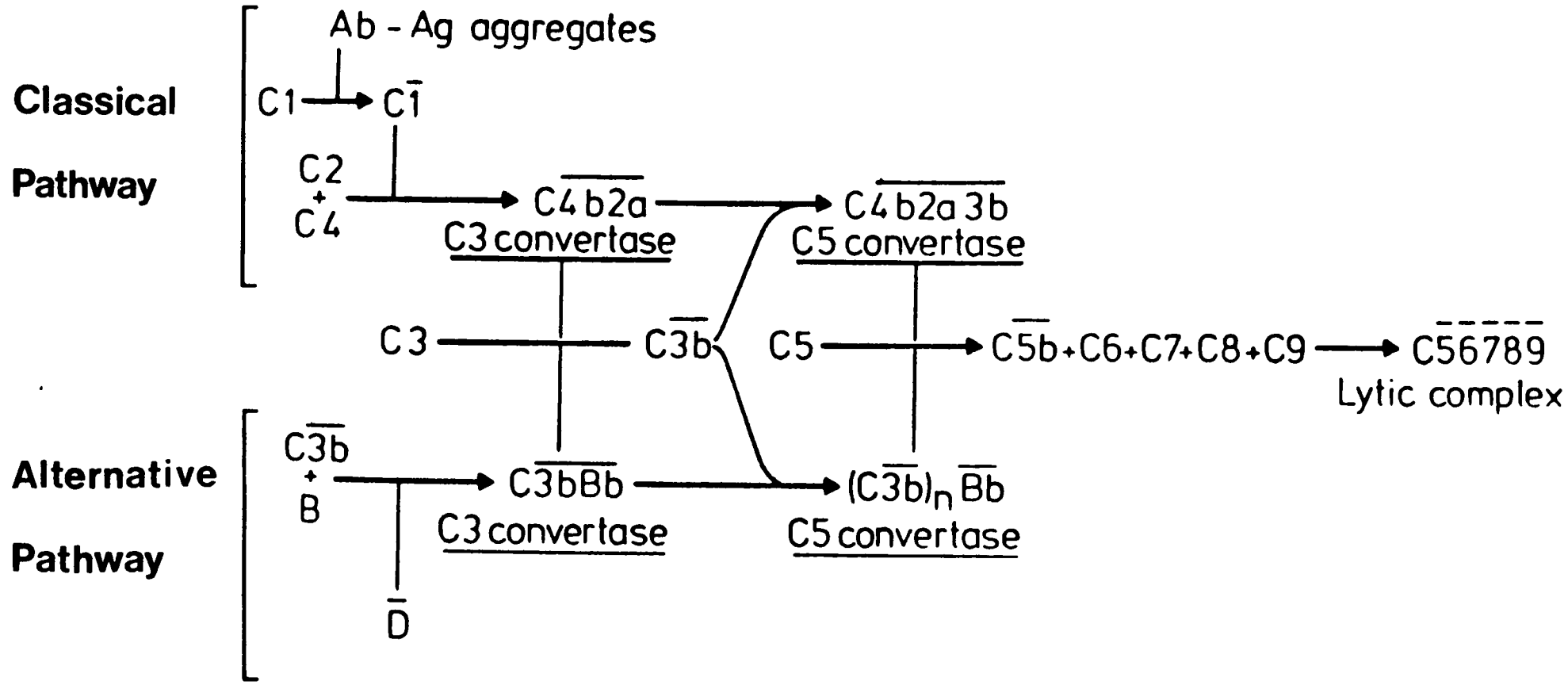


Fig 1.3 An outline of the activation of the classical and alternative pathways of the complement system. Activated components are designated by bars e.g. C1̄. The system is down-regulated mainly by the following control proteins. C1-Inhibitor (for activated C1), C4b-binding protein and factor I (for C4b), MCP, factors I and H (for C3b), S-protein (for C5b-9), CR1, DAF and factor I (for C3b and C4b). Properdin has an enhancing activity via stabilisation of the C3bBb complex (see Campbell et al., 1988). MCP and DAF stand for membrane cofactor protein and decay accelerating factor, respectively.

C2 and factor B are the serine protease components of the C3 and C5 convertases of the classical and alternative pathways, respectively. On activation of the classical pathway, C2 associates with C4b and is cleaved by a serine protease, the C1s subcomponent of the C1 complex, into a noncatalytic chain of mol. wt. 30 000 (C2b) and a 70 000 mol. wt. C terminal catalytic chain (C2a). The C4b,2a complex formed can cleave C3, the next step in the cascade. In the alternative pathway factor B associates with C3b or $C3(H_2O)$ and is cleaved by factor D, another serine protease, into Ba, the 30 000 mol. wt. noncatalytic chain, and Bb, the C terminal catalytic chain of 60 000 mol. wt., to form the C3 convertase C3b,Bb or $C3(H_2O)$,Bb. The amino acid sequence derived from cDNA sequencing has shown that C2 and factor B are very similar in gross structure and share about 35% amino acid sequence homology. This similarity reflects the homologous roles the proteins play in the two pathways of activation. The C terminal catalytic chain of both enzymes (C2a or Bb) is unusual in that it is over twice the size of the catalytic chains of most other serine proteases. The first half of the catalytic chain shares homology with the cell surface glycoprotein Mac-1, von Willebrand factor and cartilage matrix protein and may confer collagen-binding activity (Pytela, 1988). The C terminal half of the catalytic chain contains the serine protease domain. The N terminal fragments (C2b or Ba) are composed of three regions of internal homology, each about 60 amino acids in length (Morley & Campbell, 1984; Bentley, 1986). Subsequent examination of the derived amino acid sequences of a number of other complement and non-complement proteins has shown that C2 and factor B are members of a superfamily of proteins possessing the 60 amino acid repeat termed the short consensus repeat (reviewed in Reid, 1986; Reid *et al.*, 1986; Campbell *et al.*, 1988).

Polymorphism of the C2 and factor B proteins has been observed

using electrophoretic methods. The most common allelic form of C2 (gene frequency of 97% in Caucasian populations) is C2C, but rarer variants occur, C2A (<1%) and C2B (2%) (Alper, 1976). There are two major allelic forms of factor B, F and S, two less common variants, F_1 and S_1, as well as other minor variants (Alper, 1981).

C4 is the noncatalytic component of the C3 convertase of the classical pathway. C4 is activated through cleavage by C1s, which removes the C4a peptide (mol. wt. 9 000) from C4b (mol. wt. 190 000). C4b binds covalently to cell surfaces via the reactive acyl group of a glutamate residue released from a thiolester bond forming an ester or amide linkage (Law, Lichtenberg & Levine, 1979; Campbell, Dodds & Porter, 1980). C4b interacts with C2 to form the C3 convertase. More than 35 polymorphic variants of C4 have been described (Mauff et al., 1983), attributable to two isotypic loci, C4A and C4B (O'Neill et al., 1978a). The two isotypes are highly homologous differing by <1% in their amino acid sequences, but they have different activities in haemolytic assays, C4B being the more active. At neutral pH, C4A reacts exclusively with amino groups while C4B is more reactive with hydroxyl groups (Law, Dodds & Porter, 1984; Isenman & Young, 1984).

Steroid 21-hydroxylase

Steroid 21-hydroxylase (21-OHase) is a cytochrome P450 which catalyses one of the key enzymatic steps in the synthesis of glucocorticoids and mineralocorticoids in the adrenal gland. 21-OHase mediates the conversion of progesterone to 11-deoxycortisol, a precursor of cortisol. Defect in any of the enzymes of adrenal steroidogenesis leads to an accumulation of corticotropin (ACTH) since cortisol is the primary feedback inhibitor of ACTH secretion, and an excess of other

androgens. Such defects are responsible for congenital adrenal hyperplasia (CAH), 21-OHase deficiency being the most common form (New et al., 1982).

Studies of families with 21-OHase deficiency have shown that the disease is inherited as a monogenic, autosomal recessive trait and is linked to the HLA region (Dupont et al., 1977).

Tumour Necrosis Factors (TNF) α and β

TNF α and β, or cachectin and lymphotoxin, are related proteins which play a role in the destruction of virally-infected cells and tumour cells (reviewed in Old, 1985; Beutler & Cerami, 1986; Paul & Ruddle, 1988). Purified TNFα has a mol. wt. of 17 000, consists of 157 amino acids and is not glycosylated. TNFβ is a glycosylated polypeptide of 171 amino acid and 25 000 mol. wt. Both proteins aggregate into multimers. The amino acid sequences of TNFα and TNFβ have 30% homology (Gray et al., 1984; Pennica et al., 1984) and it is known that they both bind to the same receptor (Old, 1987a). The biological activities of TNFα have been most thoroughly studied. In addition to the cytotoxic effect of TNFα on certain tumour lines, it has been shown to be involved in both septic shock after bacterial infection and weight loss (cachexia) associated with chronic parasitic infections or cancer (Old, 1987b; Oliff, 1988; Cerami & Beutler, 1988). TNFα has been dirctly implicated in macrophage killing and both TNFα and TNFβ may be part of a network of lymphokines which regulate the immune response (Old, 1987a).

Heat Shock Protein (HSP70)

Heat shock proteins (HSPs) are synthesised in prokaryotic and

eukaryotic cells in response to an increase in temperature above normal physiological levels, or following exposure to a variety of toxic agents. The most abundant HSP is a protein of 70 000 mol. wt. (HSP70) which is highly conserved across species ranging from bacteria to man (Pelham, 1986). Primate cells also contain a protein (HSX70) which is closely related in sequence and antigenicity to HSP70, but is expressed at a significant basal level.

The function of HSP70 has been the subject of considerable discussion. During heat shock, HSP70 proteins concentrate in the nucleus and bind to the nuclear matrix. Subsequently they concentrate in the nucleolus apparently associated with ribonucleoprotein (Welch & Suhan, 1986). It has been proposed that HSP70 may be involved in the maintenance and stabilisation of multimeric complexes including aiding the reassembly of damaged preribosomes. HSP70 may bind to hydrophobic regions of proteins either naturally exposed or revealed as a result of heat shock denaturation to prevent inappropriate protein-protein interactions (Pelham, 1986). In unstressed cells HSP70 may be involved in protein folding and intracellular translocation of proteins (Pelham, 1988; Deshaies et al., 1988; Chirico, Waters & Blobel, 1988). In yeast, HSP70 proteins are found to accompany secreted and organelle proteins from their site of synthesis, the ribosome, to the appropriate membrane surface.

1.3.3b Class III genes

The complement gene cluster

The human genes for C2, factor B and C4 have been cloned and their structures determined (Campbell et al., 1988). The genes have been

linked together in a cluster of overlapping cosmid clones (Carroll et al., 1984a; Dunham et al., 1987). The genes are arranged head to tail within a region of less than 120 kb. The factor B and C2 genes are closely linked, lying only 421 bp apart (Wu, Morley & Campbell, 1987a). The two C4 loci, C4A and C4B, are separated by about 10 kb with the C4A gene lying 30 kb from the 3' end of the factor B gene (see Fig 4.1). Two copies of the gene for 21-OHase (21-OHaseA and 21-OHaseB) are also found in the cluster, lying 3 kb downstream of each C4 gene (Carroll, Campbell & Porter, 1985; White et al., 1985). The 21-OHaseB gene is active while the 21-OHaseA gene is a highly homologous pseudogene (Higashi et al., 1986; White, New & Dupont, 1986; Rodrigues et al., 1987). The presence of the 21-OHase gene in the class III region and the demonstration of defects or deletion of the 21-OHaseB gene in CAH patients (White, New & Dupont, 1985; Rodrigues et al., 1987) explains the HLA association of this disease. An additional gene, that would encode a novel protein with a central core consisting of the dipeptide (arginine or lysine with aspartic or glutamic acid) reiterated 21 times, called RD, has been mapped to the region between the factor B and C4A genes (Levi-Strauss et al., 1988). The RD gene is organised in the opposite transcriptional orientation to the other genes in this cluster.

The tight linkage of these genes means that they are normally inherited together as a single unit of alleles termed a complotype (Alper et al., 1983). Complotypes are designated by an abbreviated form of their factor B, C2, C4A and C4B types given in that arbitarily defined order. Thus "SCO1" refers to a complotype containing BfS, C2C, C4AQ0 and C4B1.

In the mouse, the organisation of these genes is similar to that described for man (Chaplin et al., 1983; Chaplin, 1985). A cluster of overlapping cosmid clones containing the genes for C2, factor B and C4

as well as the gene for sex-limited protein (Slp) within a region of about 180 kb has been isolated from BALB/c DNA (Chaplin et al., 1983). Slp is the other component of serum substance besides C4 and shares marked structural homology with C4, but is not functional in haemolytic assays. The C2 and factor B genes are closely linked being separated by less than 2.5 kb. Approximately 45-50 kb separates the 3' end of the factor B gene from the Slp gene which lies some 70-80 kb from the functional murine C4 gene. Thus it appears that the murine H-2 linked complement gene cluster is slighty larger than the human equivalent due to expansion of the intragenic regions. Genes for 21-OHase and RD are located in analogous positions to the human organisation (Chaplin, 1985; Levi-Strauss et al., 1988) although sequence analysis has shown that in the murine case it is the 21-OHaseA gene which is functional while the 21-OHaseB gene is a pseudogene (Chaplin et al., 1986).

In man, the C2 and factor B genes encode mRNA species of 2.9 kb and 2.6 kb respectively (Morley & Campbell, 1984; Campbell, Bentley & Morley, 1984; Bentley & Porter, 1984). cDNA clones for both messages have been isolated using oligonucleotide probes based on available protein sequence data (Bentley & Porter, 1984; Bentley, 1986; Woods et al., 1982; Campbell & Porter, 1983) and the derived amino acid sequences obtained from these clones has revealed structural homologies between the proteins (see above). The factor B gene is 6 kb in length (Campbell & Porter, 1983; Campbell et al., 1984) and is composed of 18 exons with the three regions of internal homology being encoded in separate exons. The exon organisation of the serine protease domain is typical of that of other serine proteases with the catalytic residues being split into separate exons, but there is an extra exon (exon E) which encodes a sequence unrelated to other serine proteases (Campbell et al., 1984). The C2 gene is considerably larger than the factor B gene at 17 kb

although the sizes of the mRNAs encoded are similar. The exon organisation of the C2 gene has not yet been published, but it is probable that the size difference is due to larger introns in the C2 gene. Given the close functional and structural homologies of C2 and factor B, and the close proximities of their genes, it is highly likely that the two genes arose by a duplication event of a common ancestral gene.

DNA sequence analysis of four C4A and five C4B cDNA and genomic clones (Belt et al., 1984; 1985; Yu et al., 1986) has defined nucleotide differences at 18 positions that lead to amino acid differences at 11 positions between these alleles in the polymorphic C4d fragment. These studies have also established that the A and B isotypes are determined by the presence of four specific residues within a six amino acid sequence at positions 1101-1106, PCPVLD in C4A and LSPVIH in C4B. The major Rodgers (Rg) and Chido (Ch) antigenic determinants (Rg1 and Ch1) are defined by the residues at positions 1188-1191, the Rg1 epitope being VDLL and the Ch1 epitope being ADLR (Yu et al., 1986). These isotypic and antigenic differences can also followed at the genomic level by the use of NlaIV and Eco0109 RFLPs (Yu & Campbell, 1987).

The two C4 loci are sometimes referred to as locus I and locus II, locus I being closest to the factor B gene. In general locus I encodes C4A and locus II encodes C4B. The sizes of the C4 genes can vary due to the presence or absence of a 6-7 kb intron located at the 5' end of the gene (Schneider et al., 1986; Yu et al., 1986; Palsdottir et al., 1987a). All C4A genes at locus I are 22 kb in length, whereas C4B genes may be either 22 kb or 16 kb long depending on the presence or absence of the intron. Genes that are 22 kb in length are called long genes and 16 kb genes are referred to as short genes. These differences in size can be followed using a TaqI RFLP. In addition there is variation in the

number of C4 gene present in individuals due to previous deletion or duplication of one C4 gene. This is discussed further in chapter 4. Variation in the number of C4-like genes also occurs between inbred mouse strains (Tosi et al., 1985; Müller et al., 1987b).

The 21-OHase genes are located 2-3 kb from the 3' end of each C4 gene (Carroll et al., 1985a; White et al., 1985). The 21-OHaseA and B genes can be distinguished by the size of the TaqI fragment associated with the gene. A 3.7 kb TaqI fragment segregates with the 21OHaseB gene and a 3.2 kb fragment with the 21OHaseA gene. Homozygous deletion of the 21OHaseA gene in normal individuals defines that the 21OHaseB gene is active (White et al., 1985) which is confirmed by the nucleotide sequences of the two genes (White et al., 1986; Higashi et al., 1986; Rodrigues et al., 1987). The gene sequences also show that the 21OHaseA and B genes are highly homologous and that there may be some variability in sequence between individuals (Rodrigues et al., 1987).

The TNF A and B genes

The genes for human TNFα (Shirai et al., 1985; Nedwin et al., 1985) and TNFβ (Nedwin et al., 1985) have been cloned. Both genes are about 2.5 kb long and are split into four exons, the last of which shows homology between the two genes and may encode the domain responsible for the cytolytic activity of TNF. The genes are arranged in tandem, head to tail within a 6 kb region of DNA separated by about 1 kb (Nedospasov et al., 1985). The organisation of the genes in mouse is similar (Nedospasov et al., 1986). The TNFA and B genes have been mapped to chromosome 6 in man (Nedwin et al., 1985) and chromosome 17 of the mouse (Nedospasov et al., 1986). Southern blot analysis of HLA-deletion mutants mapped the genes to within the MHC (Spies et al., 1986), while

in the mouse the genes were accurately located 70 kb from the H-2D gene within the cluster of cosmids containing the H-2D and Qa regions (Müller et al., 1987a).

Recently a gene that is expressed in B cells and macrophages, termed B144, has been mapped to the same cosmid cluster about 10 kb centromeric of the TNFA gene (Tsuge et al., 1987).

HSP70 genes

A number of human HSP70 genomic clones have been isolated (Hunt & Morimoto, 1985; Voellmy et al., 1985; Wu, Hunt & Morimoto, 1985). The genes are intronless, have sequences that are highly conserved and appear to be part of a multigene family with at least ten highly homologous genes or pseudogenes dispersed throughout the genome (Mues, Munn & Raese, 1986). Human HSP70 loci have been mapped to chromosomes 6, 14, 17 and 21 (Goate et al., 1987; Harrison et al., 1987). Recently it has been demonstrated that two tandemly arranged HSP70 genes, identical to those cloned by Hunt & Morimoto (1985), are located in the class III region of the human MHC (Sargent et al., 1988). One of the genes is known to be expressed (Wu et al., 1987b) and has been designated HSX70 (Pelham, 1986). The genes are about 12 kb apart and lie some 92 kb telomeric of the 5' end of the C2 gene.

Others

A number of other genes have been mapped to the MHC. Variation in the mobility of certain liver enzymes, which depends on alleles of the enzyme neuraminidase (Neu-1), between congenic mice and between rat strains suggests that the Neu-1 gene is linked to the MHC (Womack &

David, 1982; Figueroa et al., 1982; VandeBerg et al., 1981; Samollow et al., 1987). The observation of a rare form of sialidosis combined with CAH has suggested that the a human neuraminidase gene may also lie in the MHC (Oohira et al., 1985).

Other genetic analyses of congenic mouse strains have suggested that loci controlling the incubation period of scrapie and Creutzfeldt-Jacob disease (Kingsbury et al., 1983) and red blood cell, liver and spleen Mg^{2+} and Zn^{2+} concentrations (Henrotte, Santarroma & Pla, 1987) may lie within the H-2 region.

The locus for hereditary haemochromatosis, an iron storage disease, is thought to map close to the HLA-A locus (Edwards et al., 1986; David et al., 1987). For information on other genes that could map to the HLA region see Olaisen, Sakaguchi & Naylor (1987).

1.4 HLA AND DISEASE

An enormous number of studies comparing the frequencies of HLA antigens in patient populations with the frequencies present in control populations has shown that susceptibility to a range of diseases is associated with certain HLA haplotypes (Tiwari & Terasaki, 1985; Batchelor & McMichael, 1987). A summary of some of the strongest associations is given in Table 1.2. In general these diseases can be divided into three groups on the basis of their aetiology. The first is a group of diseases that are autoimmune in nature such as insulin dependent diabetes mellitus (IDDM), rheumatoid arthritis (RA) and ankylosing spondylitis (AS). In the second small group immunological mechanisms are not thought to be involved, for instance CAH due to 21-OHase deficiency and idiopathic haemochromatosis. Finally there is a group of diseases with extremely uncertain pathogenesis including

Table 1.2 Some examples of HLA associations with disease. Relative risks are defined as hK/Hk where h is the number of patients positive for the antigen, H the number of controls positive for the antigen, and k and K are the numbers of patients and controls negative for the antigen. Taken from Batchelor & McMichael (1987).

GROUP	DISEASE	MARKER	RELATIVE RISK
Disease not due to chronic autoimmunity	21-Hydroxylase deficiency	Bw47	15
	Idiopathic Haemochromatosis	A3	8
Diseases with autoimmunity	Ankylosing spondylitis	B27	90-350
	Rheumatoid arthritis	DR4/Dw4	4-6
	Coeliac disease	B8/DR3/DQw2	8-11
	IDDM	B8/DR3/DQw2	3-6
		B15/DR4/DQw3	2-3
Diseases of unknown pathogenesis	Narcolepsy	DR2	50-350

Table 1.3 Restriction enzymes that cleave infrequently in the mammalian genome. N indicates any nucleotide, W indicates A or T.

ENZYME	RECOGNITION SITE	ENZYME	RECOGNITION SITE
NotI	GCGGCCGC	NarI	GGCGCC
MluI	ACGCGT	SalI	GTCGAC
NruI	TCGCGA	XhoI/PaeR71	CTCGAG
PvuI	CGATCG	ClaI	ATCGAT
BssHII	GCGCGC	RsrII	CGGWCCG
SacII	CCGCGG	SmaI	CCCGGG
EagI	CGGCCG	SfiI	$GGCCN_5GGCC$
NaeI	GCCGGC		

narcolepsy and psoriasis.

Study of HLA disease associations is hampered by several factors. First it is essential that the patient group should be clinically homogeneous. This can be difficult since many of the diseases have complex aetiologies. The control group against which the patients are compared must be well matched. Of particular importance is the ethnic distribution of the patient and control populations since the frequency of HLA antigens and the patterns of linkage disequilibrium vary between the races. In some cases, such as the association of HLA-B27 with AS, the susceptibilty to disease is consistent irrespective of race suggesting a direct role for the antigen. A further problem is the linkage disequilibrium in the HLA region, so that diseases are often associated with a particular haplotype rather than with a particular HLA antigen. Indeed raised frequencies of a particular antigen in patient groups may merely reflect the extent to which the allele is distributed outside the haplotype (Dawkins et al., 1983). This leads to several possibilities. An individual HLA antigen may be responsible for the disease susceptibility, a combination of alleles at HLA loci may be important or the HLA alleles may be in linkage disequilibrium with the pathogenic locus which has not yet been identified.

The relative risks for the diseases and particular HLA antigens or haplotypes are often low (Table 1.2). Multiple genetic factors, in addition to the HLA antigens, may contribute to disease susceptibility. For instance in the nonobese diabetic (NOD) mouse, which is an experimental model for IDDM, at least three recessive genes on different chromosomes are required for the development of diabetes (Prochazka et al., 1987). It may be that environmental factors are also involved in the pathogenesis. For instance although 95% of patients with AS carry HLA-B27 compared to 40% in normals, the incidence of the disease is

discordant in monozygotic twins suggesting an environmental influence. In coeliac disease, the abnormal mucosal immune response associated with the haplotype HLA-B8, DR3, DQw2 is triggered by the presence of dietary gluten and possibly prior exposure to the intestinal adenovirus Ad-12 (reviewed in Oldstone, 1987; Shanahan & Weinstein, 1988).

Some progress has been made towards understanding the mechanisms underlying HLA disease associations. In the case of CAH due to 21-OHase deficiency, it is now clear that the gene defect is in the 21-OHaseB gene and the association of the classical (salt wasting) form of the disease with Bw47 is due to linkage disequilibrium with the deleted 21-OHaseB gene in most Bw47 haplotypes (Schneider et al., 1986).

A direct role for the class I and class II gene products has been proposed for some cases. One mechanism that could result in autoimmune phenomenon is "molecular mimicry" (Oldstone, 1987). There is now evidence that mimicry may be occurring in AS and Reiter's syndrome, where there is a strong association with B27. Both diseases have been linked epidemiologically with certain microbial infections (for instance Klebsiella pneumoniae). Sequence analysis of one of the immunodominant antigens of Klebsiella, the nitrogenase, has revealed that six consecutive residues are identical between the nitrogenase and the allelic hypervariable domain of the B27 antigen. Sera from patients reacts against a chemically synthesised peptide from this region of B27. Presentation of the homologous peptide during infection might lead to proliferation of CTLs that escape the tolerance mechanisms and cross-react with the B27 molecule (Oldstone, 1987). With respect to the generality of molecular mimicry in autoimmune disease it is interesting that an increasing number of viruses, bacteria and parasites have been shown to contain proteins homologous to host proteins (Oldstone, 1987; Goundis & Reid, 1988; Kotwal & Moss, 1988) including a gene in human

cytomegalovirus that is similar to a class I gene (Beck & Barrell, 1988).

Recent studies on IDDM have implicated specific class II sequences as the susceptibility determinants. In man, and also in the NOD mouse and the BB rat models, MHC-linked genes are required for susceptibility to IDDM (Todd et al., 1988a;b; Prochazka et al., 1987; Eisenbarth, 1986). The disease is characterised by invasion of the pancreatic islets by T and B cells. Transfer of T cells can transfer the disease from diabetic to normal animals. In Caucasians HLA-DR3 or DR4 occur in about 95% of the patients as opposed to 40% of the general population. DR2 appears to confer protection against IDDM. Todd et al. (1987) have sequenced the DRB1, DRB3, DQA and DQB gene segments that encode the polymorphic residues by using the polymerase chain reaction to amplify these sequences from cDNA prepared from patients and controls. The sequences show that there is no allele unique to IDDM. However normal DQβ chains contained Asp at position 57 whereas diabetics commonly had Ala, Val or Ser at this position. Comparison of the sequence of the I-A_β gene of the NOD mouse with the nondiabetic mouse sequence showed that the normals had Asp at position 57 but the NOD gene had Ser (Acha-Orbea & McDevitt, 1987). Position 57 of the class II molecule would be located in the antigen binding site (Brown et al., 1988). It was concluded that the amino acid residue at position 57 correlates strongly with resistance and susceptibility to IDDM.

In a similar analysis of class II sequences of normals and patients with pemphigus vulgaris, an autoimmune disease of the skin a new DQB1 allele was defined which stongly correlates with disease susceptibility on DRw6 haplotypes (Sinha et al., 1988; Scharf et al., 1988).

Bottazo et al. (1983) have noted that in autoimmune disease, the

target organ, for instance thyrocytes in Grave's disease or islet β cells in IDDM, aberrantly express class II antigen and that this may be a key part of the autoimmune process. Inappropriate expression of class II molecules is stimulated by γ interferon (IFN) (Bottazo et al., 1983) or in the case of islet cells by γIFN plus TNFα (Pujol-Borrell et al., 1987). Viral infection might trigger a local immune response leading to production of γIFN by T cells and TNFα by macrophages. Induction of class II molecule expression on the cell surface could lead to an autoimmune response.

Complement deficiencies can also lead to HLA disease association. C4 null alleles occur at a raised frequency in systemic lupus erythematosus (SLE) patients (Fielder et al., 1983). Tissue damage in SLE is caused by failure to limit the size of immune complexes. The key role of C4 in the clearance of immune complexes suggests it may be pathologically relevant (Batchelor & McMichael, 1987). In addition drugs which induce SLE such as hydralazine have been shown to inhibit C4 (Sim, Gill & Sim 1984).

1.5 TECHNIQUES

The human MHC is estimated to cover 3-4 cM which corresponds to approximately 3-4 Mb. As described above the region is well mapped at the genetic level and covered with cloned gene probes. This makes the MHC particularly suitable for mapping using the novel technique of PFGE (Schwartz & Cantor, 1984; Carle & Olson, 1984; Hardy et al., 1986). In addition as discussed in Chapters 3 and 6, there is reason to believe that more genes may lie within the class III region. This makes the MHC a suitable testbed in which to search for coding sequences. Analysis of the human MHC may have considerable bearing on efforts to map the human

genome (see Weinberg, 1988 for instance).

1.5.1 Cosmid walking

Cosmid vectors are plasmids that contain the λ phage cos site and thus can be efficiently introduced into host bacteria through in vitro packaging (Collins & Hohn, 1978; Ish-Horowicz & Burke, 1981; Grosveld et al., 1982; Hohn & Murray, 1977). Recombinant molecules with a minimum size of 38 kb and a maximum size of 52 kb can be packaged. Due to their large cloning capacity (effectively up to 50 kb), cosmids have proven valuable cloning vectors for studies of mammalian DNA, for instance the mouse H-2 complex (Hood et al., 1983). The process of isolating sequential overlapping cosmid clones to move from one locus on a chromosome to another is called walking (Bender, Spierer & Hogness, 1983). Walking can be divided into three stages. First isolation of the initial cosmid clones, followed by characterisation of the genomic inserts by restriction enzyme mapping and Southern blotting, and finally identification of single-copy sequences from the ends of the genomic insert with which to rescreen the library and proceed in the walk.

A number of other techniques have recently been devised to facilitate progression along a chromosome. Collins & Weissman (1984) proposed that directional "jumps" could be made away from an initial probe by allowing large DNA fragments to ligate at low concentration, but in the presence of an excess of a selectable marker. The DNA circularises and the ends of the fragment can be obtained by restricting the circles and cloning the ends tagged with the marker into a suitable vector. The end product is a small clone containing two sequences which are distant from one another on the chromosome so that a jump has been made. Jumping libraries have been constructed (Poustka et al., 1987;

Collins et al., 1987), but the technique has not been widely used and suffers from the disadvantage that the DNA between the two points is not cloned.

Burke, Carle & Olson (1987) have constructed vectors based on the functional elements required for maintenance and segregation of yeast chromosomes. The vectors can accept genomic DNA fragments up to at least 500 kb and propagate as linear artificial chromomsomes, hence the acronym YAC (yeast artificial chromosome). It seems that the future of cloning large regions of DNA lies with YAC vectors although this technology is still in its infancy (see Coulson et al., 1988).

1.5.2 Pulsed Field Gel Electrophoresis (PFGE)

The ability to conveniently separate large DNA fragments has been a major advance recently in molecular genetics. Conventional agarose gel electrophoresis has an effective upper limit of separation by molecular sieving of about 50 kb. Efforts to improve resolution above this size by the use of dilute agarose gels (down to 0.035% agarose) have been hampered by poor resolution and the fragility of the gels (Fangman, 1978; Serwer, 1980; 1981). Schwartz and Cantor (1984) realised that if the DNA was subjected to alternating, approximately perpendicular electric fields a novel form of separation was achieved. Thus, they were able to separate yeast chromosomal DNAs up to an estimated size of 2000 kb and the technique of PFGE was born. Carle & Olson (1984) modified the system to produce orthogonal field alternation gel electrophoresis (OFAGE) with a more convenient elctrode configuration and were also able to separate yeast chromosomes. Since that time, a number of further modifications to the original idea have been made so that the technique is now widely available. In particular a variety of electrode

geometries, homogeneous fields, field switching regimes and angles between fields have been explored to produce linear DNA trajectories. These include a system where the two fields at 180° are alternated in either duration or strength (Carle, Frank & Olson, 1986), a vertical system with an orthogonal design (Gardiner, Laas & Patterson, 1986), systems where multiple electrode arrays are individually voltage clamped or autonomously controlled by a computer to produce homogeneous fields (Chu, Vollrath & Davis, 1986; Clark et al., 1988; Bancroft & Wolk, 1988) or systems in which the gel is rotated between equal and opposite angles in a homogeneous field (Southern et al., 1987; Serwer, 1987). PFGE has been used to separate the chromosomes of Saccharomyces cerevisiae, Trypanosomes (Van de Ploeg et al., 1984), Plasmodium (Kemp et al., 1985), Candida albicans (Snell & Wilkins, 1986) and Schizosaccharomyces pombe (Vollrath & Davis, 1987; Smith et al., 1987a; Birren et al., 1988). The upper size limit of DNA molecules that have been separated is now estimated as 10 Mb.

Parameters which affect the separation in PFGE include temperature, voltage gradient, reorientation angle and the interval between field switching (Birren et al., 1988). The most useful variable is switching interval. Briefly as the switching interval is increased, DNA molecules of increasing size are separated from the bulk of slowly moving unresolved DNA. Maximum resolution between molecules in a given size range occurs closest to the switching interval required to separate the DNA from the unresolved region (see Chapter 4). The mechanism of separation is still obscure. Cantor, Smith & Mathew (1988) argue simplistically that the time required for the DNA to reorient is dependent on the length of the DNA molecule and therefore the mobility of each molecule depends on the fraction of the switching interval remaining after reorientation. Southern et al. (1987) have proposed a

"ratcheting" model to explain the observed requirement for field angles >90°. The DNA molecules retain their extended configurations when the field direction is changed so that the trailing edge of each molecule leads off in the next direction and separation proportional to the length of the DNA is achieved. However, this model cannot explain the separation obtained when the alternate fields are oriented at 180° (Carle et al., 1986). Clearly further theoretical and physical studies of the process of migration and reorientation of DNA molecules are required (see Deutsch, 1988; Lalande et al., 1987; Holzworth et al., 1987).

In order to use PFGE for mapping of mammalian DNA, which is at least two orders of magnitude larger than, say, the yeast genome, methods are required to reproducibly fragment the DNA into sizes that can be analysed. A number of restriction endonucleases are known that cleave mammalian DNA rarely enough to produce fragments in the 50-1000 kb range (see Anand, 1986; van Ommen & Verkerk, 1986 for reviews, Table 1.3). These enzymes have either 6 or 8 bp recognition sites which contain one or more CpG dinucleotides. CpG is rare in the mammalian genome (Josse, Kaiser & Kornberg, 1961; Swartz, Trautner & Kornberg, 1962) and 60-90% of CpG is methylated. Since many of these enzymes are sensitive to cytosine methylation in their recognition sequences, cleavable sites are infrequent (see Chapter 6). Additionally unsheared genomic DNA is required for digestion and this can be conveniently prepared in agarose blocks (Anand, 1986; van Ommen & Verkerk, 1986).

Thus, restriction enzyme digestion of high mol. wt. DNA with infrequently cutting enzymes, separation of the restriction fragments by PFGE, Southern blotting and hybridisation with appropriate probes can lead to the construction of a physical map of a region in an analogous way to conventional restriction mapping. This strategy has been

successfully applied to a number of regions of mammalian chromosomes (see for instance Brown & Bird, 1986; Hardy et al., 1986; Burmeister & Lehrach, 1986; Dunham et al., 1987; Burmeister et al., 1988).

1.6 AIMS OF PROJECT

The aims of this project were:

1. To map the human MHC and to establish the exact position and orientation of the complement gene cluster within the class III region. At the same time it was aimed to map the positions of the TNF genes within the MHC to see if they were in analogous positions to those in the mouse.

2. To compare the physical structures of the MHC in a number of MHC haplotypes.

3. To use the map of the class III region to search for new genes within that region.

Details of the specific aims are given in each experimental chapter.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Enzymes

(a) Restriction enzymes

All restriction enzymes were purchased from Amersham International or GIBCO BRL except the following,

*Bss*HII, *Eag*I, *Mlu*I, *Nae*I, *Not*I, *Pae*R71	- New England Biolabs
*Cla*I	- Boehringer Mannheim
*Nru*I, *Sac*II, *Sfi*I	- Northumbria Biologicals Ltd.

(b) Others

Lysozyme (Grade I), proteinase K (Type XI), RNase A (Type III-A), DNase (Crude)	- Sigma

Alkaline phosphatase, proteinase K, T4 polynucleotide kinase	- Boehringer Mannheim
DNA polymerase I-Klenow fragment, T4 DNA ligase, Random hexanucleotide priming kit	- Amersham International
Sequenase kit	- United States Biochemical Corporation
Zymolyase-100T (from Arthrobacter luteus)	- Miles

2.1.2 Chemical reagents

Chemical reagents were obtained from BDH, Sigma, Fisons and Boehringer Mannheim except for the following:

Agar No.1, Yeast extract	- Oxoid
Agarose, type I & type VII	- Sigma
Agarose, HGT & LGT (Seaplaque)	- Seakem
Bactotryptone	- Difco
Dextran sulphate, Ficoll 400	- Pharmacia

Phosphate buffered saline (PBS) - Flow
(Dulbecco's without calcium and
magnesium, Dulbecco & Vogt, 1954)

2.1.3 Bacterial strains, plasmids, phage and yeast.

Escherichia coli strain	MC 1061	(Casdaban & Cohen, 1980)
" " "	NM 514	(Watson & Jackson, 1985)
" " "	NM 554	(Gift from Noreen Murray)
" " "	JM 103	(Messing, Crea & Seeburg, 1981)
Plasmids	pAT-X	(Gift from George Brownlee)
	pUC8	(Messing, 1983)
Phage	λgt10	(Huynh, Young & Davis, 1985)
	λcI857s7	(Maniatis, Fritsch & Sambrook, 1982)
	M13mp8	(Messing & Vieira, 1982)

Saccharomyces cerevisiae (Gift from Jill Ogden)
strain X2-180-1B

2.1.4 Media

Media were prepared as described below and autoclaved before use. Solid media was prepared by adding either 15 g/l No.1 Agar to liquid media for plates or 7 g/l agarose to liquid media for top agarose.

2 x TY : 15 g bactotryptone, 10 g yeast extract, 5 g NaCl to 1 l with water. pH adjusted to 7.4 with ∿2 ml NaOH.

L broth : 10 g bactotryptone, 5 g yeast extract, 10 g NaCl to 1 l with water.

T broth : 10 g bactotryptone, 5 g NaCl, 2.2 g $MgSO_4.7H_2O$ to 1 l with water.

SOB : 20 g bactotryptone, 5 g yeast extract, 0.5 g NaCl to 1 l with water. pH adjusted to 7.5 with KOH. Just before use add 10 ml/l sterile 1M MgCl, 1M $MgSO_4.7H_2O$.

YPD : 20 g bactotryptone, 10 g yeast extract, 20 g glucose to 1 l with water. Autoclaved for not more than 20 min to avoid degradation of the glucose.

2.1.5 Radioactive nucleotides

All radioactive nucleotides were obtained from Amersham International. [α^{32}P]-dNTPs were 10 mCi/ml, specific activity 3000 Ci/mmol. [α^{35}S]-dATP was 8-10 mCi/ml, specific activity 1200 Ci/mmol.

2.1.6 Nucleic acids

The 1 kb DNA ladder and the 0.16-1.77 kb RNA ladder were obtained from GIBCO BRL.

Plasmids containing various HLA gene probes were gifts as follows:

pDRH2	- a DRA cDNA (Lee et al., 1982),	John Trowsdale
p11-13	- a DPB genomic fragment (Trowsdale, 1987),	
p8ABC5	- a class I cDNA (Trowsdale et al., 1984b),	
pDA1α3B	- a DPA cDNA (Trowsdale, 1987).	
pDQA	- a DQA cDNA (Hardy et al., 1986),	John Bell
pDRA	- a DRB cDNA (Hardy et al., 1986).	
pC250	- HLA-C and HLA-B-specific genomic	Tom Strachan
pB250	probes (Strachan et al., 1986).	
pTNFA	- a TNFA genomic probe (Shirai et al., 1985).	Mark Bodmer
pDOB	- a DOB cDNA (Tonnelle et al.,	Eric Long

1985).

pAT-A - a C4 cDNA (Belt et al., 1984). Duncan Campbell

p21B - a 21-OHaseB genomic probe (Rodrigues et al., 1987). Nanda Rodrigues

All other probes used were genomic DNA fragments either isolated directly from cosmid clone inserts by restriction enzyme digestion and separation on LGT agarose gels (see section 2.7.2) or from subclones containing the fragment of interest. The cloned cosmid DNA was kindly provided by Carole Sargent (Dunham et al., 1987; Sargent, 1988; Sargent et al., 1988).

2.2 OVERNIGHT CULTURES

Bacteria from a glycerol stock or a single colony were used to inoculate 1-50 ml of liquid media and were incubated with shaking at 37°C overnight. Yeast cultures were incubated shaking at 30°C.

E. coli JM 103 were grown in 2 x TY broth. E. coli MC 1061 and NM 554 were grown in either L broth or SOB medium. E. coli NM 514 were grown in T broth, 0.2% maltose.

A 0.9 ml aliquot of the overnight culture was removed and 100 µl of 10 x Hogness freezing buffer (40% glycerol, 36 mM dipotassium hydrogen phosphate, 13 mM potassium dihydrogen phosphate, 20 mM sodium citrate, 10 mM $MgSO_4.7H_2O$, pH 7.4) was added to give glycerol stocks which were stored at -70°C. Glycerol stocks of yeast were prepared by adding an equal volume of glycerol to an aliquot of saturated yeast culture in YPD broth and were stored overnight.

2.3 PREPARATION OF DNA

2.3.1 Standard techniques

Ethanol precipitation- DNA was precipitated by addition of 1/10th volume of 3 M sodium acetate, pH 6.0 and 2½ volumes of ethanol and left at -70°C for 30 min or at -20°C overnight. The DNA was recovered by centrifugation at 10 000g for 10 min. The pellet was air dried and then resuspended in 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0.

Phenol extraction- Aqueous solutions were extracted by addition of an equal volume of phenol (saturated with 20 mM Tris-HCl, 1 mM EDTA, pH 7.4), then vortexed before centrifugation at 10 000g for 5 min. The aqueous layer was removed to a fresh tube and treated as appropriate.

2.3.2 Small scale isolation of plasmid and cosmid DNA (10-50 ml cultures) (Birnboim & Doly, 1979)

Cultures of *E. coli* were grown overnight to saturation in L broth containing 50 µg/ml ampicillin (for ampicillin resistant plasmids). All subsequent steps were carried out at 4°C unless otherwise stated. The cells were pelleted in an MSE 6L centrifuge at 5000 r.p.m. for 20 min and resuspended in 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0 (500 µl/10 ml culture). The cells were lysed by addition of 2 volumes of 0.2 M NaOH, 1% SDS for 5 min followed by one half volume 3 M sodium acetate, pH 4.8 for 5 min. The solution was centrifuged at 15 000 g for 20 min. The clear supernatant was taken off and the plasmid DNA

precipitated for 30 min by addition of an equal volume of isopropanol. Plasmid DNA was recovered by centrifugation at 15 000 g for 10 min, resuspended in 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0 (80 µl/10 ml culture) and treated with 50 µg/ml RNase A at 37°C for 30 min. The solution was extracted with an equal volume of phenol and precipitated with ethanol. The recovered pellet was resuspended in 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0 (50 µl/10 ml starting culture) and stored at -20°C.

2.3.3 Large scale isolation of plasmid DNA (500-1000 ml cultures) (Maniatis et al., 1982)

From an overnight culture of E. coli containing plasmid, a fresh culture was started by dilution of 1 in 100 into fresh L broth containing ampicillin (50 µg/ml).The culture was grown at 37°C to A_{600} = 0.8. Chloramphenicol was added to 200 µg/ml and incubation continued overnight. The cells were harvested and lysed as described above (section 2.3.2) and the DNA was precipitated in ethanol. The plasmid DNA pellet was resuspended in 6 ml SET buffer (0.15 M NaCl, 5 mM EDTA, 50 mM Tris-HCl, pH 8.0) and 200 µl of 0.5% ethidium bromide and 8 g of CsCl was added to the sample. The density of the solution was checked to be between ρ= 1.61-1.7 g/ml and the solution was transferred to a 13.5 ml Beckman quickseal ultracentrifuge tube, topped up with SM-caesium chloride solution and liquid paraffin, capped and sealed. The DNA was centrifuged for 16 hr at 55 000 r.p.m. and 18°C in a Ti70.1 rotor. The speed was then reduced to 45 000 r.p.m. for 1 hr to allow "relaxation" of the gradient. The plasmid band (the lower, closed circular band) was recovered by puncturing the tube with a needle just below the band and allowing the DNA to drain into a Sterilin tube.

Ethidium bromide was removed by repeated extraction with equal

volumes of butan-1-ol. The DNA was recovered by additon of 3 volumes of water and 2 volumes of ethanol and precipitation at -70°C overnight. The DNA was then reprecipitated and resuspended in 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0. The DNA concentration was estimated from the A_{260} assuming that 1.0 A_{260} unit = 50 µg/ml of double stranded DNA.

2.3.4 Isolation of mammalian chromosomal DNA in aqueous solution (Bell et al., 1981)

Chromosomal (or genomic) DNA was prepared from a HLA homozygous, EBV transformed lymphoblastoid cell line (HLA and complement typed as A2, B7, DR2, C2C, BfS, C4A3, C4BQ0) known as "Ice 5". The cells were grown and harvested by David Hunter. The cells were collected by centrifugation at 250 g for 10 min and were washed twice in PBS (isotonic phosphate buffered saline). Cell nuclei were prepared by dispersing the pelleted cells in 40 ml of lysis buffer (0.32 M sucrose, 10 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 1% Triton X-100) and spinning at 2000 r.p.m. for 15 min at 5°C in the MSE 6L centrifuge. The supernatant was discarded and the lysis procedure repeated twice. The nuclei were resuspended in 10 ml SET buffer, proteinase K added to 100-200 µg/ml and the suspension incubated at 37°C until all clumps had dissolved. The solution was then extracted with an equal volume of phenol: chloroform: isopropanol (25:24:1) and once with an equal volume chloroform: isopropanol (24:1). After centrifugation at 2000 r.p.m. for 15 min at 4°C the organic layer was removed using a syringe, leaving the aqueous layer and the interface. The aqueous layer was removed to a clean tube and the genomic DNA precipitated with 2.5 volumes of chilled ethanol. The DNA precipitate was spooled out with a sealed pasteur pipette and allowed to resuspend overnight at 4°C in 200 µl 10 mM Tris-HCl, 1 mM

EDTA, pH 7.4. It was then dialysed against 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.4 overnight at 4°C and stored at 4°C. The DNA concentration was determined from the A_{260} as above.

2.3.5a Isolation of mammalian chromosomal DNA in agarose blocks (van Ommen & Verkerk, 1986; Dunham et al., 1987)

High mol. wt. DNA was isolated in agarose blocks both from HLA homozygous, EBV transformed lymphoblastoid cell lines (Ice 5 grown and harvested by David Hunter, others were gifts from Roger Dawkins and were grown and harvested by Duncan Campbell) and from peripheral blood mononuclear cells (PBMNs).

Tissue culture cells were washed in PBS, counted using a haemocytometer, washed again in PBS and resuspended at 4×10^7 cells/ml in PBS.

Freshly drawn blood (10-20 ml) was immedietely made 5 mM in EDTA. Blood was kindly taken by either Beryl Moffat or Konrad Kölble. The blood was spun at 1200 g in a bench centrifuge for 10 min, the top EDTA-plasma layer recovered and stored in 0.5 ml aliquots at -20°C. The remaining cells were resuspended in 20 ml PBS, layered onto 10 ml HISTOPAQUE-1077 (Sigma) in a Falcon tube and centrifuged at 400g for exactly 30 min at room temperature. The opaque interface containing the PBMNs was collected into a fresh Falcon tube. The cells were washed three times in PBS containing 1 mM EDTA and were recovered by centrifugation at 250 g for 10 min. The PBMNs were counted using a haemocytometer and resuspended in PBS at 4×10^7 cells/ml.

Cells at 4×10^7 cells/ml were warmed to 37°C and were mixed with an equal volume of LGT agarose that had been held at 40°C after melting. Either 2% Sigma Type VII low-gelling temperature (LGT) agarose that had

been treated with DEAE-cellulose (Jackson & Cook, 1985, see section 2.3.5b below for method) or 2% Seaplaque LGT agarose was used. Both types of agarose allow the high mol. wt. DNA prepared to be restricted easily and without degradation. The cell-agarose mixture was then dispensed into a perspex mould containing slots of dimensions 9 mm x 7 mm x 2 mm (made in the Biochemistry Dept. workshop), as illustrated (Fig. 2.1), on ice. The blocks were allowed to gel for 10 min.

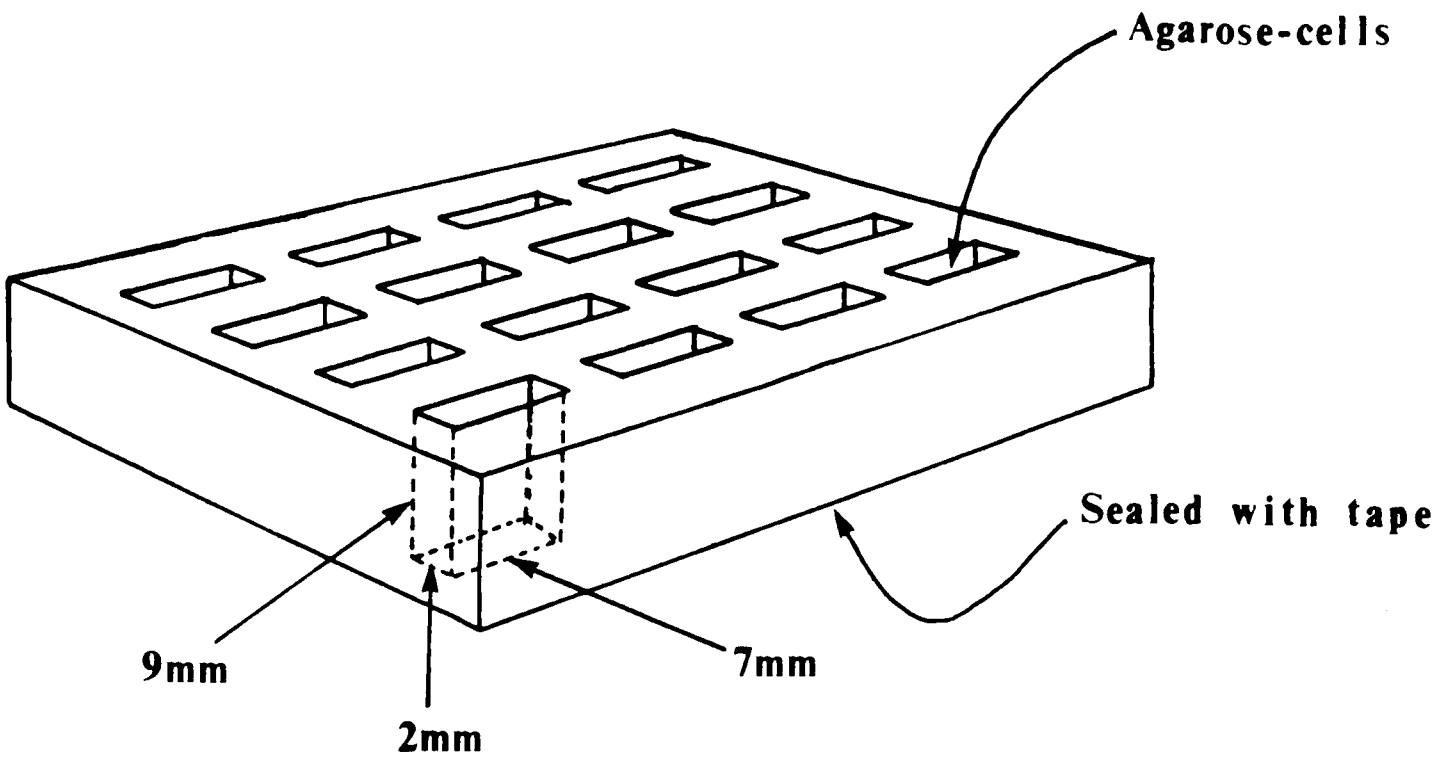


Fig 2.1 Diagrammatic representation of the perspex block former. The agarose-cell mixture was pipetted molten into the slots which had been sealed as shown with Tucktape. The dimensions of the slots are as shown in millimeters.

The backing tape was removed and the blocks knocked into a Falcon tube containing 10 ml NDS (0.5 M EDTA, 10 mM Tris-HCl, pH 9.5, 1% Sarkosyl) made 1 mg/ml with proteinase K, using a bent pasteur pipette. Blocks were incubated overnight at 50°C in this solution and then again in fresh NDS-proteinase K solution for a further 2 days at 50°C. Finally the blocks were washed twice in 10 ml NDS and were stored in 10 ml NDS at 4°C for up to 2 years.

2.3.5b Purification of LGT agarose by treatment with DEAE-cellulose (Jackson & Cook, 1985)

This method can be used to remove negatively charged inhibitors of restriction enzymes from LGT agaroses. LGT agarose was dissolved by boiling in 1 x PBS and held at 50°C. The agarose was then mixed with 5ml of packed DE-52 (Whatman) that had been previously equilibrated in 5 x PBS. After 30 min at 50°C the DE-52 was removed by centrifugation at 3000 g for 10 min and the treatment repeated with fresh DE-52. Finally the DE-52 was removed by 3 successive spins at 3000 g for 10 min, decanting the molten agarose to a fresh Falcon tube and holding at 50°C to prevent gelling between each spin. The agarose was stored at 4°C.

2.3.6 Isolation of yeast chromosomal DNA in agarose blocks (van Ommen & Verkerk, 1986)

Saccharomyces cerevisiae strain X2-180-1B was grown on a YPD plate for 2 days at 30°C. A single colony was picked off and used to inoculate a 50 ml overnight culture which was grown shaking at 30°C. The overnight culture was diluted 1 in 10 into 500 ml of fresh YPD and incubation was continued until A_{660} = 1.0. The cells were collected by centrifugation at 3000 g for 10 min at 18°C. The cells were washed in PBS and then resuspended in four times the cell pellet volume of PBS. The yeast cell suspension was mixed with an equal volume of molten 2% LGT agarose containing 20 mM DTT and 10 µg/ml Zymolyase-100T at 50°C and blocks formed as described above (section 2.3.5a).

The yeast blocks were incubated at 37°C for 90 min in an equal volume of PBS containing 20 mM DTT and 10 µg/ml Zymolyase-100T to form spheroplasts. The blocks were then treated with NDS and proteinase K as

for human genomic DNA (section 2.3.5a). Yeast chromosomes in blocks were stored at 4°C in NDS for up to 1 year.

This method gives DNA at a very high concentration in the blocks and very thin slices were loaded onto PFGE.

2.3.7 Isolation of λ phage and λ DNA from λcI857 S7 and preparation of λ concatemers (Maniatis et al., 1982; Anand, 1986; Waterbury & Lane, 1987)

Two methods of preparing concatemers as markers for PFGE were used, one utilising λ phage and the other using λ DNA. λcI857 S7 is a temperature sensitive, lysogenic phage that can be induced to enter lytic growth by transiently raising the temperature of the growing bacterial culture. This can be used as a convenient source of both packaged phage and DNA (Maniatis et al., 1982).

Lysogenic bacteria were streaked out onto each of two T plates and one was grown at 30°C and the other at 42°C. Colonies should only form on the plate at 30°C. A single colony was used to inoculate a 10 ml overnight starter culture in T broth at 30°C. This was diluted 1 in 50 into 500 ml of T broth and the culture was grown at 30°C until A_{600} = 0.5. The phage was induced by warming the cells at 42°C for 15 min and the incubation was continued at 37°C for a further 2 hr. At this stage a 1 ml sample was checked for evidence of phage growth by adding a few drops of chloroform to the cells and incubating for a further 5 min at 37°C. The suspension went clear if the lysogen was properly induced.

The cells were collected by centrifugation at 4000 g for 10 min at 4°C and resuspended in 5 ml SM buffer (100 mM NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris-HCl, pH 7.5, 2% gelatin). 10 drops of chloroform were added to the cells, the mixture was vortexed well and was allowed to stand at

room temperature for 30 min. DNase and RNase A were added at 0.2 µg/ml each and the suspension further incubated for 1 hr. Cell debris was removed by centrifugation at 12 000 g for 15 min and the supernatant was extracted with an equal volume of chloroform: isopropanol (24:1).

The volume of the phage suspension was measured and 0.75 g/ml solid caesium chloride was added. The phage was transferred to a 13.5 ml Beckman quickseal ultracentrifuge tube, topped up with SM to which 0.75 g/ml caesium chloride had been added and the tube was capped and sealed. The solution was spun at 55 000 r.p.m. in a Ti70.1 rotor overnight. The phage band was eluted through the side of the tube using a needle and dialysed against 1 l 10 mM NaCl, 10 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.0 overnight. At this stage the phage was taken to make concatemers in agarose blocks.

To prepare phage DNA, the phage suspension was made to 20 mM EDTA, 50 µg/ml proteinase K, 0.5% SDS and incubated at 65°C for 1 hr. The DNA solution was extracted once with an equal volume of phenol and once with an equal volume of phenol: chloroform: isopropanol (25:24:1) and was then dialysed three times against 1 l of 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0 overnight at 4°C. The DNA concentration was estimated from the A_{260}.

To prepare concatemers of λ from intact phage (Waterbury & Lane, 1987) the phage (DNA) concentration was estimated by lysis of phage in 20 mM EDTA, 0.5% SDS and observation of ethidium bromide fluorescence after electrophoresis on a 0.7% agarose gel. The phage was diluted in PBS and mixed with an equal volume of 2% LGT agarose to give a final DNA concentration of about 0.3 µg/µl. Blocks were formed as described above and were incubated in NDS and proteinase K at 50°C overnight. The blocks were stored at 4°C and sliced thinly (≤1 mm) to be loaded onto PFGE gels.

To prepare concatemers of λ DNA in solution (Anand, 1986), the DNA

was incubated at 200 µg/ml in 2 x SSC (1 x SSC = 150 mM NaCl, 15 mM sodium citrate, pH 7.0), 5% glycerol and bromophenol blue at 37°C for 30 min, followed by room temperature for 18-24 hr. Concatemers can be stored at 4°C for a year and over, and 5 µl is sufficient for one marker track on PFGE.

Both methods of preparing concatemers give suitable markers. The liquid solutions can be loaded into thin slots to give tight bands on PFGE. The agarose blocks are generally easier to handle if a large number of other DNA samples in agarose are being processed. The size of one monomer unit of λ is accurately known from sequencing, so the ladder of concatemers gives useful size markers every 48.5 kb in PFGE.

The λ DNA prepared by the above method can be restricted with HindIII to give size markers in the range 23-0.6 kb which along with the BRL 1 kb DNA ladder were used for short switching interval PFGE runs.

2.3.8 Isolation of bacteriophage λgt10 DNA by the plate lysate method (Maniatis et al., 1982; Helms et al., 1985)

Approx. 10^5 p.f.u. of a purified stock of bacteriophage (see section 2.10) were incubated with 200 µl of a fresh overnight culture of E. coli NM514 at 37°C for 15 min. This was mixed with 7 ml T-top agarose, poured onto 14 cm T agar plates and incubated at 37°C overnight to give confluent plaques. 10 ml SM buffer was added to the plates, the plates were sealed and stored overnight at 4°C with occasional shaking. Alternatively, the plates were incubated shaking at room temperature for 3 hr.

The SM lysate was harvested and the plates washed with 5 ml of SM. A few drops of chloroform were added to the pooled SM lysate. After vortexing briefly, the lysate was centrifuged for 10 min at 4000 g and

4°C. The supernatant was then transferred to a 38.5 ml Beckman quickseal ultracentrifuge tube which was topped up with SM and spun at 35 000 r.p.m. in a Ti60 rotor for 90 min at 18°C. The pelleted phage was resuspended in 3 ml/plate 10 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$ by either shaking overnight at 4°C or at 37°C for 2 hr.

The phage solution was layered onto a 2.5 ml DE52 column that had been poured in a 10 ml disposable pipette and equilibrated with 2 column volumes of 10 mM Tris-HCl, pH 8.0. The column was then washed with 5 ml of 10 mM Tris-HCl, pH8.0, 10 mM magnesium acetate, followed by 0.8 ml of 10 mM Tris-HCl, pH8.0, 60 mM magnesium acetate. Finally the phage was eluted by washing with 1.0 ml of 10 mM Tris-HCl, pH 8.0, 60 mM magnesium acetate and the eluate was collected. The column eluate was made to 0.5% SDS and 40 µg/ml proteinase K and stood at room temperature for 5 min. Phage proteins were removed by addition of 1/10 volume 3 M potassium acetate, heating at 88°C for 20 min and centrifugation at 10 000 g for 10 min.

The λ DNA was precipitated with an equal volume of isopropanol at -20°C for 1 hr and resuspended in 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA at a concentration of 1 µg/µl. This protocol gives an average yield of 10 µg/plate.

2.4 PREPARATION OF RNA (Chirgwin et al., 1979)

All glassware and disposable equipment was sterilised before use by autoclaving. Gloves were worn for all steps involving RNA in order to prevent degradation of the RNA by nucleases.

Tissue culture cells were harvested and counted by Beryl Moffat, and were washed twice in PBS. The cells were resuspended in 3 ml of fresh 4 M guanidium isothiocyanate, 0.1 M β-mercaptoethanol solution by

vortexing. To this solution was added 3 ml of 5.7 M caesium chloride, 10 mM EDTA, 25 mM sodium citrate, pH 7.0 and this was layered onto 3 ml of the caesium chloride solution in a 13.5 ml Beckman quickseal ultracentrifuge tube which was topped up with paraffin. This was centrifuged at 35 000 r.p.m. for 18 hr at 18°C in a Ti70.1 rotor. The RNA pellet was resuspended in 400 µl of 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0, 1% SDS, precipitated once in 0.3 M sodium acetate with 2½ volumes of ethanol at -20°C for 2 hr and finally resuspended in 200 µl 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0, 1% SDS.

The concentration of RNA was determined from the A_{260} assuming that a RNA solution of 40 µg/ml has an A_{260} = 1.0. RNA was stored aliquoted in 1 mg-20 µg aliquots as an ethanol precipitate.

2.5 DIGESTION OF DNA BY RESTRICTION ENDONUCLEASES

2.5.1 DNA in solution

Restriction digests of DNA in solution were performed using the conditions recommended by the supplier. Spermidine at 5 mM was included in all buffers containing NaCl concentrations ≧50 mM. Analytical digests used 50 ng-1 µg DNA in a reaction volume of 10-20 µl. Preparative digests used 5-20 µg DNA in a reaction volume of 50-100 µl. Digests were usually performed with a 5 fold excess of enzyme (where 1 unit of enzyme is that amount stated by the supplier to completely digest 1 µg of λ DNA in 1 hr at the appropriate temperature). Genomic DNA was digested for 16-24 hr, all other DNA samples were digested to completion in 2 hr.

2.5.2 DNA in agarose blocks (Anand, 1986)

One third of an agarose block (3 mm x 7 mm x 2 mm) containing ~5-10 µg DNA was sliced from the originally prepared blocks (section 2.3.5a) for each digest. Blocks were washed three times in 10 ml of 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0, 0.1 mM phenylmethylsulphonyl fluoride for 30 min at 4°C and were then equilibrated with 0.5 ml restriction buffer without BSA, spermidine and DTT. Restriction enzyme digestion was carried out according to the suppliers recommendations in 100 µl of restriction buffer containing 500 µg/ml BSA with 20 units of enzyme for 3 hr unless otherwise stated. For TaqI digestion was at 55°C for 16 hr with 50 units of enzyme. For some enzymes (e.g. BssHII) it is possible to use less enzyme to obtain complete digestion and this was titrated for individual batches of enzyme and agarose blocks. All buffers containing ≧50 mM NaCl also contained 5 mM spermidine. The presence of at least 500 µg/ml BSA in the reaction buffer was found to be important to ensure complete digestion in agarose blocks (Parker & Seed, 1980).

Double digestions were carried out as follows. The first enzyme digestion was carried out as above and then the blocks were equilibrated in 0.5 ml of the second enzyme restriction buffer without BSA, spermidine and DTT. Digestion was then continued with the second enzyme.

After digestion, blocks were either placed at 4°C for at least 10 min, sliced in half and loaded into PFGE gel slots, or stored for up to 6 months in NDS at 4°C. Some digests were treated at 50°C in 0.5 ml NDS containing 500 µg/ml proteinase K for 1 hr before loading.

2.6 RADIOLABELLING OF DNA

2.6.1 "Fill in" reaction

DNA fragments were labelled with [α^{32}P]-dNTPs by the "fill in" reaction using the Klenow fragment of DNA polymerase I. Restriction digests were labelled directly by adding 4 units/50 µl reaction volume of DNA polymerase I (Klenow fragment), 5 µCi of [α^{32}P]-dNTP and the three complementary dNTPs to a final concentration of 20 µM. For preparation of fragments for blunt end ligation, 1/10 of the digested DNA was end-filled as described and the rest was filled in with all 4 (cold) dNTPs under the same conditions. The two reactions were then combined and the DNA was separated by polyacrylamide gel electrophoresis.

2.6.2 Random hexanucleotide priming (Feinberg & Vogelstein, 1984)

DNA probes (10-20 ng) were labelled by random hexanucleotide priming using the Amersham "Multipriming kit" using 20 µCi [α^{32}P]-dCTP for 3 hr at room temperature. Specifically, the probe DNA solution was diluted in water to a volume of 31 µl and heated at 95°C for 5 min. After cooling the probe solution on ice for 10 min, 10 µl of buffer (250 µM dATP, 250 µM dGTP, 250 µM dTTP, 200 mM Tris-HCl, pH 7.8, 20 mM $MgCl_2$, 1 mM β-mercaptoethanol) was added, followed by 5 µl of primer solution (random hexadeoxyribonucleotides in aqueous solution including 4 mg/ml BSA). 20 µCi [α^{32}P]-dCTP and 2 units of DNA polymerase I were added, the reaction mixed and left for 3 hr. Labelled DNA was separated from free nucleotides by gel filtration on a 2 ml Sephadex G100 column. Probes were labelled to specific activities of $>10^8$ c.p.m./µg.

Probes were denatured by boiling for 5 min and snap cooling on ice for 5 min before use in DNA-DNA filter hybridisation.

2.7 GEL ELECTROPHORESIS OF NUCLEIC ACIDS AND RECOVERY OF DNA FROM GELS

2.7.1 Polyacrylamide gel electrophoresis (Sanger & Coulson, 1978; Maxam & Gilbert, 1980)

DNA was electrophoresed on 40 cm x 20 cm 4% or 6% polyacrylamide gels at 30 mA in 1 x TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) for 1½-4 hr. Denaturing gels also contained 7 M urea. DNA was loaded in glycerol/dyes which were 50% glycerol, 5 mM EDTA, 0.1% xylene cyanol and bromophenol blue for native gels and 60% formamide, 35 mM EDTA, 0.1% xylene cyanol and bromophenol blue for denaturing gels.

Gels were covered with Saran wrap (Dow chemicals) and autoradiographed using X-OMAT S X-ray film (Kodak) with a phosphotungstate intensifying screen (Cronex Lightning Plus, Dupont) at room temperature or -70°C (Laskey & Mills, 1977). The fims were sensitised by pre-flashing before use and were developed in an Kodak MEI X-OMAT automatic processor.

DNA was recovered from polyacrylamide gels by excising the gel slice containing the fragment of interest and eluting the DNA from the gel slice overnight in 2 M ammonium acetate at 37°C, followed by three sequential precipitations with ethanol. The DNA was finally resuspended in 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0.

2.7.2 Agarose gel electrophoresis of DNA (Southern, 1979)

For analytical purposes, DNA was routinely electrophoresed in 30

ml 0.7% HGT agarose submarine minigels using 50 ml 1 x TBE containing 5 µl 0.5% ethidium bromide at 30-60 mA for 1-3 hr. For preparative electrophoresis, restriction enzyme digested DNA was electrophoresed in 1.0% LGT agarose minigels either at 2-10 mA overnight or at 30 mA for 3 hr at 4°C. DNA samples were mixed with 1/10th volume glycerol/dye mix before loading.

Restriction enzyme digests of genomic DNA were run in 0.7-1.5% HGT agarose slab gels using the BRL gel apparatus. The running buffer was 1 x TBE and electrophoresis was overnight at 40-50 mA. Solution DNA samples were mixed with 1/10th volume glycerol/dye mix before loading. Digested DNA samples in agarose blocks were loaded directly into the wells formed in the gel as desribed in section 2.7.4. Gels were stained after electrophoresis in 1 x TBE containing 200 µl 0.5% ethidium bromide for ½ hr and then destained in water for upto 1 hr.

Ethidium bromide stained DNA was visualised under 300 nm ultra-violet light (Sharp, Sugden & Sambrook, 1973) using a Fotodyne transilluminator and photographed using a Polaroid MP4 camera fitted with a red filter and Polaroid 667 film.

DNA was eluted from LGT agarose gels by the method of Weislander (1979). A slice of agarose containing the desired DNA was cut out of the gel, weighed and 3 ml of water/g of agarose was added. The agarose was melted by heating at 65°C for 5 min. At room temperature the melted gel slice was extracted once with phenol, once with phenol: chloroform: isopropanol (25:24:1) and once with chloroform: isopropanol (24:1). The DNA was recovered by ethanol precipitation and resuspended in 10 mM Tris, 0.1 mM EDTA, pH 8.0. This DNA was then suitable for labelling by hexanucleotide priming or for ligation into the appropriate cut and phosphatased vector.

2.7.3 Agarose gel electrophoresis of RNA (Lehrach et al., 1977; Fourney et al., 1988)

RNA samples were pelleted after storage as an ethanol precipitate and dissolved in 25 mM EDTA, 0.1% SDS. To 10 µg total RNA in a volume of 5 µl was added 25 µl of RNA loading buffer (53% deionised formamide, 1 x MOPS (0.02 M [3-(N-morpholino) propanesulphonic acid], 5 mM sodium acetate, 1 mM EDTA, pH 7.0), 6.2% (2.2 M) formaldehyde, 7% glycerol, 0.5% bromophenol blue). Samples were heated at 65°C for 15 min, 2 µl of 0.5% ethidium bromide solution was added and samples were loaded on the gel. The markers used were 250 ng λ DNA digested with HindIII per track which were treated as described above.

RNA was eletrophoresed in 1% agarose gels containing 1.8% (0.66 M) deionised formaldehyde in 1 x MOPS as running buffer overnight at 40 mA.

The RNA could be visualised directly and photographed on the transilluminator as described above.

2.7.4 Pulsed field gel elctrophoresis of high mol. wt. DNA (Brown & Bird, 1986; Southern et al., 1987)

Agarose blocks containing DNA that had been digested with restriction enzymes (section 2.5.2) were sliced in half (1 mm x 3 mm x 7 mm) with a sterile razor blade and loaded directly into the wells of PFGE gels. Molecular weight markers were either intact yeast chromosomes (section 2.3.6), concatemers of λcI857 S7 DNA (section 2.3.7), λ DNA digested with HindIII or the BRL 1 kb DNA ladder.

Two PFGE gel systems were used. In both systems buffer was circulated through an external cooling device to maintain constant temperature.

1. Initially an orthogonal field alternation gel electrophoresis (OFAGE) system was used. The gel box and switching interval timer (described in Brown & Bird, 1986) were built by Duncan Fletcher of the MRC Mammalian Genome Unit and the apparatus is illustrated schematically in Fig 2.2(a). Digested DNA samples were electrophoresed on 20 cm x 20 cm x 0.5 cm 1.5% agarose (Sigma Type I) gels in 4 l of 0.5 x TAE (20 mM Tris acetate, 1 mM EDTA, pH 8.5) for 22 hr at 12°C and 330 V with switching intervals of 5-70 s. The gels were poured with wells of dimensions 14 mm x 4 mm x 1½ mm and the two halves of a digested block were loaded side by side in these wells to counteract the focusing-in effect characteristic of this gel system.

2. Later the crossed field or "Waltzer" gel system was used (described in Southern et al., 1987). The gel box was built by Bob Chick in the Biochemistry Dept. workshop and electrical components were wired in by Martin Johnson. The apparatus is also represented schematically in Fig. 2.2(b). Digested DNA samples were electrophoresed on 22 cm diameter x 0.5 cm 1.5% agarose (Sigma Type I) circular gels in 5 l of 0.5 x TAE for 30-36 hr at 18°C and 150 V with switching intervals of 7.5-70 s. The gels were poured with wells of dimensions 7 mm x 4 mm x 2 mm and half a digested block was loaded per track. The other half of the digested block was stored as described in section 2.5.2.

PFGE gels were stained in 500 ml 0.5 x TAE containing 200 µl ethidium bromide for ½ hr and then destained for up to 1 hr in water. Gels were photographed using a Fotodyne transilluminator, a Polaroid MP4 camera and Polaroid 667 film through a red filter.

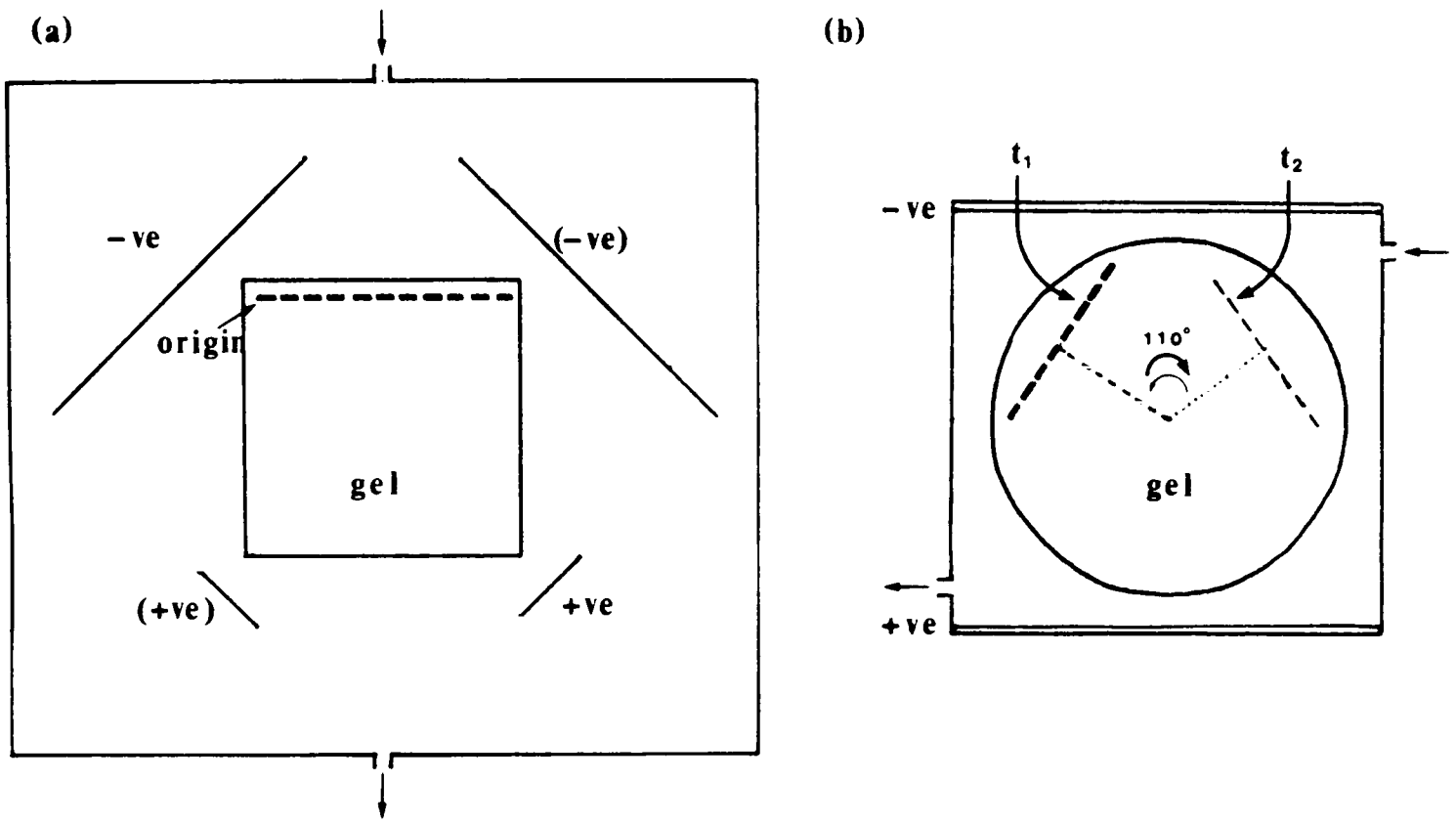


Fig 2.2 Schematic representation of the PFGE apparatuses. In both cases the DNA samples are loaded in the wells indicated by the bold broken line. The direction of migration of the DNA is down the page. Arrows indicate the inlet and outlet points for recirculation of the electrophoresis buffer through a thermostatic cooler. (a) The OFAGE box built at the MRC Mammalian Genome Unit. The anodes and cathodes are indicated by the +ve and -ve signs. The polarity after switching is shown in brackets. (b) The "Waltzer" apparatus. The two alternate positions of the sample wells during the run are indicated by t_1 and t_2. (Adapted from Southern et al., 1987)

2.8 DNA AND RNA TRANSFER BLOT ANALYSIS

2.8.1 Southern blot analysis of DNA fragments separated in agarose gels (Southern, 1979; Wahl, Stern & Stark, 1979)

After staining with ethidium bromide and photography, gels were treated as follows.

1. Twice in 0.25 M HCl for 20 min at room temperature. This step depurinates the DNA (Wahl et al., 1979) and was omitted for fragments below 4 kb.

2. Once in 0.4 M NaOH, 0.6 M NaCl for 30 min at room temperature.

3. Once in 0.5 M Tris-HCl, 1.5 M NaCl, pH 7.4 for 30 min at room temperature.

The gels were then blotted by capillary transfer in 10 x SSC onto either nylon-backed nitrocellulose (Hybond-C Extra, Amersham) or nylon (GeneScreenPlus, New England Nuclear) membranes using sponges to increase capillary action. Gels were blotted overnight for genomic DNA samples or for 2 hr for cloned DNA samples.

Hybond-C Extra membranes were pretreated by wetting in water. After blotting, these membranes were baked for 2 hr at 80°C before hybridisation.

GeneScreenPlus membranes were wetted briefly in water and then in 10 x SSC for 30 min and placed with side B (see manufacturers instructions) next to the gel. After blotting, GeneScreenPlus membranes were treated for 60 seconds in 0.4 M NaOH to ensure complete

denaturation of the immobilised DNA and then for 60 seconds in 0.2 M Tris, pH 7.4, 2 x SSC. The DNA was crosslinked onto the nylon membrane by exposure of the DNA side of the membrane to u.v. at 254 nm for 4 min (CAMAG Universal u.v. lamp held ~15 cm from the gel).

Alternatively, DNA was blotted onto nylon membranes (Hybond-N, Amersham) by alkaline transfer (Reed & Mann, 1986). Prior to blotting the gels were treated as follows.

1. Twice in 0.25 M HCl for 20 min.

2. Once in 0.5 M NaOH, 1.5 M NaCl for 30 min.

3. Once in 0.25 M NaOH, 1.5 M NaCl for 15 min.

The gel was then blotted onto Hybond-N, which requires no pre-wetting, overnight in 0.25 M NaOH, 1.5 M NaCl. After blotting the DNA was fixed onto the membrane by air drying for 10 min.

The procedures for hybridisation of membranes are described in section 2.8.3.

2.8.2 Northern blot analysis of RNA separated in agarose gels (Lehrach et al., 1977; Fourney et al., 1988)

After electrophoresis and staining the gel was soaked for 10 min in 0.05 M NaOH, 1 x SSC and then washed twice in 10 x SSC for 20 min. Capillary blot transfer was then set up in 10 x SSC overnight, transferring the RNA onto Hybond-C Extra which had been prewetted in 10 x SSC.

Following transfer, the RNA was fixed to the membrane by baking

for 2 hr at 80°C.

2.8.3 Hybridisation of membranes (Bernards & Flavell, 1980)

Filters were prewashed in 1 M NaCl, 1 mM EDTA, 50 mM Tris-HCl, pH 8.0, 0.1% SDS for at least 30 min at 42°C. Prehybridisation was carried out at 42°C in 50% deionised formamide, 1 M NaCl, 50 mM Tris-HCl, pH 7.4, 0.2% Ficoll 400, 0.2% BSA, 0.2% polyvinylpyrrolidone, 0.1% sodium pyrophosphate, 10% dextran sulphate, 200 µg/ml sonicated salmon sperm DNA, 0.1% SDS (hybridisation buffer) for at least 2 hr. The hybridisation buffer was boiled for 5 min and then cooled to 42°C before use. For GeneScreenPlus membranes the hybridisation buffer contained 1% SDS and the filters were prehybridised overnight on the first occasion of use and then for at least 2 hr on subsequent occasions.

Filters were hybridised at 42°C for 16 hr in the same buffer to which had been added [^{32}P]-labelled probe at 0.5-1 x 10^6 c.p.m./ml (1-2ng/ml). All probes had been labelled by random hexanucleotide priming (section 2.6.2) and were denatured by heating at 95°C for 5 min followed by snap cooling on ice.

When hybridising multiple blots with the same probe, the filters were placed in the same hybridisation bag, ensuring that hybridisation buffer penetrated between each membrane.

2.8.4 Washing, autoradiography and rehybridisation of membranes

After hybridisation, filters were washed twice at room temperature for 10 min in 2 x SSC, then twice in 0.2 x SSC, 0.1% SDS at 65°C for 30 min. Finally filters were rinsed in 0.2 x SSC at room temperature for 10 min. For GeneScreenPlus membranes, the 65°C washing solution contained

1% SDS.

Membranes were drained of excess washing solution and were wrapped in Saran wrap. It is essential that GeneScreenPlus membranes are not allowed to dry as this will result in irreversible binding of probe. Filters were autoradiographed using X-OMAT S X-ray film between two intensifying screens at -70°C for between 1 and 15 days (Laskey & Mills, 1977). The films were sensitised by pre-flashing before use and were developed in a Kodak MEI X-OMAT automatic processor.

Filters were stripped of probe by washing in 2 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1% SDS (1% SDS for GeneScreenPlus) at 80°C for up to 2 hr. The efficiency of this stripping procedure was checked by autoradiography.

Stripped filters were either hybridised with the next probe or were stored dry at room temperature.

2.9 SUBCLONING OF DNA FRAGMENTS INTO pATX

2.9.1 Preparation of E. coli NM 554 competent cells

A 10 ml culture of NM 554 was grown overnight in SOB medium from a single colony. 200 µl of this overnight was used to inoculate a 50 ml starter culture which was grown to A_{600} = 0.2. This culture was diluted 1 in 20 into 1 l of SOB and grown to A_{600} = 0.3.

The cells were chilled in an ice bath for 10 min and then collected by centrifugation at 6000 r.p.m. in an MSE 6L centrifuge for 5 min at 4°C. The pelleted cells were gently resuspended in 400 ml of transformation buffer (50 mM $CaCl_2$, 10 mM PIPES, pH 6.6, 15% glycerol) and placed at 4°C for 20 min. The cells were collected again as above. The bacteria were finally resuspended in 50 ml of transformation buffer,

divided into 0.5 ml aliquots and snap frozen on dry ice. Competent cells were stored at -70°C.

The transformation efficiency of these cells with pUC8 was 10^8/µg.

2.9.2 Preparation of BamHI- and PvuII-cut and phosphatased pATX vector

The plasmid vector pATX (10µg) was restricted to completion with 20 units of either BamHI or PvuII for 2 hr at 37°C and then treated with 5 units calf intestinal alkaline phosphatase for 30 min at 37°C. The reaction was stopped by addition of EDTA to 0.25 M and heating for 15 min at 68°C. The DNA was purified by phenol extraction and ethanol precipitation. The vector preparation was tested for

1. background of uncut vector,
2. its ability to religate to itself before phosphatasing,
3. efficiency of phosphatasing as judged by abolition of self-religation,
4. its ability to accept suitable fragments in test ligations (see section 2.9.3).

2.9.3 Ligation and transformation

Insert DNA was prepared by restriction endonuclease digestion, "filling-in" if necessary and separation on either LGT agarose or polyacrylamide gels (section 2.7). Fragments were ligated by mixing 5 ng cut and phosphatased vector with ~5 ng insert DNA fragments in 50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 20 mM DTT, 50 µg/ml BSA, 1 mM spermidine, 1 mM rATP and 1 unit of T4 DNA ligase in a 10 µl reaction volume. The reaction was incubated at room temperature for 5 hr.

Before transformation frozen competent cells (section 2.9.1) were thawed on ice. 100 µl of competent cells were added to each ligation reaction. The mixture was incubated on ice for 10 min and then heat shocked for 5 min at 37°C. An equal volume of L broth was added to the mixture which was incubated at 37°C for 40 min. The cells were plated onto L agar plates containing 50 µg/ml ampicillin and grown overnight at 37°C.

2.9.4 Analysis of recombinants

Recombinants were analysed by two methods.

1. Individual recombinants were picked off and plasmid DNA was prepared from 10 ml of an overnight culture (in L broth containing ampicillin) by the method described in section 2.3.2. DNA inserts were characterised by restriction enzyme digestion, electrophoresis on agarose minigels (section 2.7.2) and Southern blotting.

2. Alternatively the recombinants were screened for the presence of a particular DNA insert in the following way. Dry nitrocellulose filters were placed onto the bacterial colonies and were keyed asymmetrically with ink. The filters were removed and placed, bacterial colony side up, on a fresh L plate containing ampicillin and were incubated at 37°C for 5-8 hr. The master plate was allowed to recover at 37°C for 2 hr. The colonies on the nitrocellulose were lysed by placing the filters, colony side up, successively on Whatman 3MM paper soaked in the following solutions for 4 min,

1. 10% SDS,

2. 0.5 M NaOH,

3. 1 M Tris-HCl, pH 7.4.

The filters were then washed in 0.5 M Tris-HCl, pH 7.4, 1.5 M NaCl for 15 min and in 2 x SSC for 15 min. The nitrocellulose was baked at 80°C for 2 hr and hybridised with a labelled DNA probe for the fragment of interest, washed and autoradiographed as described in sections 2.8.3 and 2.8.4.

2.10 SCREENING OF cDNA LIBRARIES IN λgt10 (Huynh et al., 1985)

Two cDNA libraries were screened. The HL60 library was constructed by Dimitrios Goundis from total RNA extracted from HL60 cells that had been treated with cycloheximide (Goundis, 1988). The second cDNA library was constructed by Alex Law from EBV-transformed lymphoblastoid cell line total RNA. 3×10^5 clones were screened from the HL60 cDNA library and $\sim 3 \times 10^6$ clones from the lymphoblastoid cell line cDNA library.

2.10.1 Plating out of λgt10 clones

The HL60 library was plated at 5×10^4 recombinants/plate. For each plate an aliquot of the packaged library was incubated with 200 µl of a fresh overnight culture of *E. coli* NM 514 at 37°C for 15 min. This was mixed with 7ml T-top agarose held at 50°C, poured onto 14 cm T agar plates and incubated at 37°C for 8 hr. The lymphoblastoid cell line cDNA library was provided as DNA on nitrocellulose filters for screening.

2.10.2 Screening of libraries

The plates were allowed to dry for 45 min in a laminar flow hood at room temperature. Dry 132 mm nitrocellulose filters were placed in direct contact with the plaques for 1 min and keyed with ink asymmetrically to mark both filter and plate. The filters were removed and placed phage side up on Whatman 3MM paper. A second set of replicas were made by placing nitrocellulose circles on the plates for 90 s and marking the filters in the same places as the first set of replicas. The library master plates were then sealed and stored at 4°C.

Both sets of replica filters were then treated sequentially on Whatman 3MM paper soaked in the following solutions,

1. 0.5 M NaOH, 0.5 M NaCl, for 3 min.
2. 0.5 M Tris-HCl, pH 7.4, 1.5 M NaCl, twice for 3 min.

Filters were washed in 6 x SSC for 5 min and then baked at 80°C for 2 hr. The filters were hybridised with radiolabelled probe as described in sections 2.8.3 and 2.8.4.

2.10.3 Rescreening of positives

Autoradiographs were examined for plaques that hybridised in duplicate (i.e. on both replica filters). The positions of putative positives were aligned with the master plates and an agar plug of diameter 3-4 mm was taken from the area of the positive plaque using a 1 ml Gilson pipette tip from which the end had been cut. The plug was placed in 1 ml of SM buffer containing 1% chloroform (v/v) and stood at 4°C for at least 3 hr.

10 µl of the positive phage stock was diluted 1 in 100 into 1 ml SM buffer and 100 µl, 25 µl and 5 µl were incubated with 200 µl of a fresh NM 514 overnight culture for 15 min at 37°C. The mixture was added to 3 ml of T-top agarose held at 50°C and poured onto a 9 cm T plate. The plate was incubated at 37°C for 8 hr, and plaque lifts were taken and screened as above (section 2.10.2).

This process of rescreening was repeated until a single duplicating positive plaque could be picked off a plate. The plaque was placed in 1 ml of SM, 1% chloroform and this purified stock of positive infectious bacteriophage was used to isolate phage DNA containing the cDNA insert by the plate lysate method described in section 2.3.8. The DNA could then be analysed by restriction enzyme mapping and nucleotide sequencing.

2.11 DIDEOXY NUCLEOTIDE SEQUENCING (Sanger, Nicklen & Coulson, 1977)

2.11.1 Preparation of JM 103 competent cells

100 µl of a fresh overnight culture of JM 103 in 2 x TY broth was diluted into 10 ml of 2 x TY and grown to an A_{600} = 0.2-0.4. The cells were pelleted by centrifugation at 2000 g and 4°C for 10 min. The cells were resuspended in 5 ml of 50 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.4 and placed on ice for 20 min. The cells were pelleted again and resuspended in 1.5 ml of the same buffer and stored at 4°C.

2.11.2 Subcloning into SmaI-cut and phosphatased M13mp8

The SmaI-cut and phosphatased M13mp8 vector was supplied by Amersham International.

DNA fragments were prepared by restriction enzyme digestion, "filling-in", and purification on LGT agarose or 4% polyacrylamide gels (section 2.7). ∿5-25 ng of DNA fragment was ligated into the M13 vector as described in section 2.9.3.

100 µl of the JM 103 competent cells were added to each ligation reaction and the mixture was incubated on ice for 40 min, then at 42°C for 5 min. The transformed cells were added to 3 ml molten L agarose held at 45°C containing 25 µl of 2.5% aqueous solution of isopropylthiogalactoside (IPTG), 25 µl of bromochloroindolylgalactoside (BCIG) and 200 µl of a fresh JM 103 overnight culture. This mixture was immediately poured onto a 9 cm 2 x TY plate which was incubated overnight at 37°C. Recombinant clones are unable to express β-galactosidase activity due to insertion of the cloned fragment into the SmaI site within the lac gene and are hence white. Non-recombinants do express β-galactosidase which is induced by IPTG and can hydrolyse BCIG to produce a blue indolyl dye. Therefore recombinant plaques are coloured blue.

2.11.3 Preparation of single stranded template DNA (Sanger et al., 1980)

Agar plugs of white plaques were grown for 5 hr at 37°C in 1.5 ml of a 1 in 100 dilution of a JM 103 overnight culture in 2 x TY broth. The culture was centrifuged at 10 000g for 5 min, transferred to a fresh Eppendorf tube and respun for 10 min. The supernatant was transferred to a fresh tube and the phage was precipitated by addition of 200 µl of 2.5 M NaCl, 20% PEG (polyethylene glycol), mixing and standing on ice for 30 min. The phage was pelleted at 10 000 g for 10 min and resuspended in 200 µl 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0.

The phage was extracted with an equal volume of phenol, ethanol

precipitated and finally resuspended in 30 µl 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0.

2.11.4 Sequenase reaction

Both orientations of insert DNA were identified either by sequencing or by "T-tracking", that is performing only the T reaction of the sequencing protocol to obtain a diagnostic fingerprint of each clone.

The sequenase kit (United States Biochemical Corporation) was used to determine the nucleotide sequence of single stranded M13mp8 clones. The Sequenase enzyme is a modified form of T7 DNA polymerase (Tabor & Richardson, 1987) which is superior to the Klenow fragment of E. coli DNA polymerase I in dideoxy sequencing reactions.

The kit includes;

1. Sequenase buffer- 200 mM Tris-HCl, pH 7.5, 100 mM MgCl, 250 mM NaCl.
2. 0.5 pmol/µl universal M13 primer (5'-GTAAAACGACGGCCAGT-3')
3. dGTP labelling mix- 7.5 µM dGTP, 7.5 µM dCTP, 7.5 µM dTTP.
4. Termination mixes in 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.4.

ddGTP mix : 80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP, 50 mM NaCl.

ddATP mix : 80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddATP, 50 mM NaCl.

ddTTP mix : 80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM

ddTTP, 50 mM NaCl.

ddCTP mix : 80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddCTP, 50 mM NaCl.

All sequencing reactions were performed in 1.5 ml Eppendorf tubes. 7 µl of single stranded DNA (1-2 µg) was mixed with 2 µl of Sequenase buffer and 1 µl of universal primer, heated at 65°C for 2 min in a water bath which was then allowed to cool to 35°C over about 30 min. To the annealed primer-template mix was added 1 µl 0.1 M DTT, 2 µl labelling mix diluted 1 in 10 in water, 1 µl [α^{35}S]-dATP and 2 µl of Sequenase enzyme diluted 1 in 8 in ice cold 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.4. This mixture was incubated at room temperature for 5 min. This reaction labels the nascent extended strand at limiting concentrations of dNTPs.

3.5 µl of the labelling reaction was transferred into the four termination tubes each of which contained 2.5 µl of the appropriate ddNTP mix. The termination reactions were incubated at 37°C for 5 min, then stopped by addition of 4 µl of STOP solution containing 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol. The samples were heated at 75°C for 2 min just before loading on the sequencing gel.

2.11.5 Electrophoresis of sequencing reactions

The samples were run on either 6% denaturing polyacrylamide gels at 30 mA for 2.5-5 hr or buffer gradient polyacrylamide gels at 30 mA for 80 min (Biggin *et al.*, 1983). The back plate of both types of gel was treated with 2.5 ml ethanol containing 70 µl 10% acetic acid and 7 µl 7.5% "Bind Silane" (LKB) (Garoff & Ansorge, 1981) to bind the gel

covalently to the plate. After electrophoresis the gels adhering to the silane-treated plate were fixed by treatment in 10% acetic acid, 10% methanol for 15 min and then washed in water for 15 min to remove the urea. The gel was dried down onto the plate by baking at 110°C for 30 min. The gels were then autoradiographed at room temperature for 1-5 days without intensifying screens or preflashing the film.

Buffer gradient gels were made as follows. 7 ml of 2.5 x TBE mix (6% acrylamide, 2.5 x TBE, 7 M urea) was mixed with 70 µl 10% ammonium persulphate solution and 14 µl TEMED (N,N,N',N'-tetramethylethylenediamine). 35 ml 0.5 x TBE mix (6% acrylamide, 0.5 x TBE, 7 M urea, 5% sucrose, 0.05% bromophenol blue) was mixed with 140 µl 10% ammonium persulphate solution and 70 µl TEMED. 4 ml of the 0.5 x TBE mix was taken up into a 10 ml pipette, followed by 7 ml of the 2.5 x TBE mix. A gradient was formed by the introduction of 2-3 air bubbles into the pipette and through the interface. The mixture was pipetted between the gel plates and the remaining space was filled with 0.5 x TBE mix. These gels were run with 0.5 x TBE in the top buffer reservoir of the electrophoresis chamber and 1 x TBE in the bottom.

2.11.6 Analysis of nucleotide sequence data (Staden, 1982a; 1982b; 1984)

Nucleotide sequence data was analysed using the Staden software available on the Oxford University Vax cluster network. In particular the general sequence analysis programme, ANALYSEQ, and the sequencing database programmes were used to manipulate sequence data and to search for significant features. The FASTP and FASTN programmes (Lipman & Pearson, 1985) were used to search the available NBRF protein and EMBL databases. Further information on usage of software can be obtained in Chapter 6.

CHAPTER 3

A MOLECULAR MAP OF THE HUMAN MHC

3.1 INTRODUCTION

The human MHC consists of three major linked gene clusters. The class I and class II regions each encode highly polymorphic families of cell surface glycoproteins involved in immune regulation (see Chapter 1). The class I region contains at least 17, and possibly as many as 40, highly related genes (Strachan, 1987; Srivastava et al., 1985) which include those encoding the classical transplantation antigens (HLA-A, B and C). The class II region (HLA-D) is arranged into subregions DP, DQ and DR, each containing at least one A and B pair of genes encoding the α and β polypeptide chains (Hardy et al., 1986). A number of other class II-related sequences (including DO, DN and DV) have also been defined within the HLA-D region (Trowsdale, 1987; Inoko et al., 1988). The class III loci include genes encoding the serum complement proteins C2, Factor B and C4, as well as two copies of the 21-OHase gene which are closely linked to the two C4 loci, C4A and C4B (Carroll et al., 1984a; Carroll et al., 1985a).

The possibility that a number of other genes may reside within the MHC has been raised recently by the finding that the genes for the tumour necrosis factors (TNF) α and β (TNFA and TNFB) are linked to the HLA (Spies et al., 1986) and have been mapped close to the H-2D region in mouse (Müller et al., 1987a). Subsequently, a gene encoding a macrophage and B cell specific cDNA called B144 was shown to lie ~10 kb from the TNFA gene in mouse (Tsuge et al., 1987). Similarly a gene

predicted to encode a protein with an unusual periodic structure, the RD gene, has been mapped to a position between the C4 and factor B genes in both mouse and man (Levi-Strauss et al., 1988). A number of other loci have been linked to the MHC (Lamm & Olaisen, 1985; Olaisen et al., 1987) including the Neu-1 (neuraminidase-1) gene which maps to the H-2S region of the mouse (Figueroa et al., 1982). Susceptibility to a number of diseases has also been linked to certain HLA haplotypes (see Chapter 1). Although direct involvement of the known MHC gene products has been proposed (Batchelor & McMichael, 1987; Todd et al., 1988a), the molecular basis for these associations is far from clear.

The genetic structure of the human MHC has been intensively studied and the genetic map is shown in Fig 3.1. Analysis of recombinant HLA haplotypes in family studies has established that the class I loci are telomeric to the class II genes (Reinsmoen et al., 1977). Within the class II region the DP subregion maps centromeric to DQ and DR (Shaw et al., 1981). Located between the HLA-DR and HLA-B loci are the class III genes (Lamm & Olaisen, 1985), but the orientation of these genes relative to the class I and class II genes has not been determined due to the absence of informative recombination events within the region. Attempts have been made to establish the orientation of the class III genes by scoring of the number of putative recombinations that would have had to occur to produce modern haplotypes from presumed ancestral haplotypes. However these analyses have lead to conflicting results (Wilton & Charlton, 1986; Marshall et al., 1984; Abbal et al., 1987). The position of the complement genes has been the subject of several studies of families showing recombination events between the HLA-DR and HLA-B loci. In some (Raum et al., 1979; 1981; Robinson et al., 1985; Yunis et al., 1985) cases the complement allotypes under consideration segregated more frequently with HLA-DR suggesting close linkage to this

locus. However other studies provided conflicting results (reviewed in Robson & Lamm, 1984; Lamm & Olaisen, 1985). The genetic distances presented in Fig 3.1 for this region were obtained from combining the linkage data from several family studies of allotypic variants (Robson & Lamm, 1984) and correspond most accurately to the observed physical distances. Whitehead et al. (1985) have used hybridisation of a C4 gene probe to DNA from cell lines with deletions in the MHC to try to define the positions of the complement genes. Their results located the complement gene cluster between HLA-B and HLA-DR, possibly closer to HLA-B.

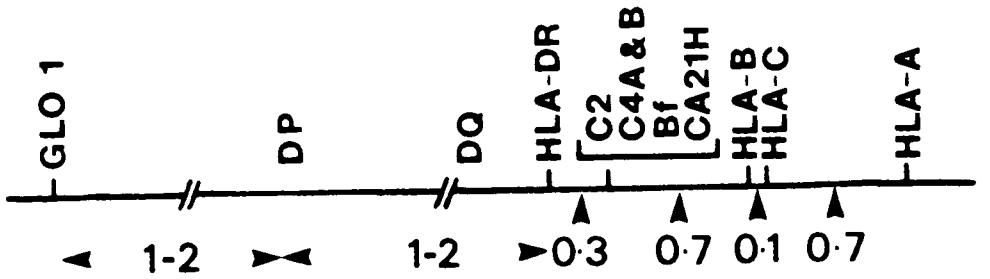


Fig 3.1 Genetic map of the human MHC. The numbers below the line are the distances in cM for the gaps shown. GLO 1 is the glyoxalase-1 locus. CA21H stands for congenital adrenal hyperplasia due to 21-OHase deficiency. Taken from Robson & Lamm (1984).

A physical map of the human MHC would clarify the gene organisation in this complex region and would facilitate the identification of novel genes within the HLA region. Such a physical linkage map might also play an important part in understanding the basis of HLA-disease associations.

Clusters of overlapping cosmid clones have been isolated from the subregions of the class II loci (reviewed in Trowsdale, 1987) and from the class III region (Carroll et al., 1984a) but have not yet been linked by mapping or walking procedures. Recently the technique of pulsed field gel electrophoresis (PFGE) (Schwartz & Cantor, 1984; Carle

& Olson, 1984) in combination with restriction enzymes that cut rarely in the mammalian genome, has been applied to long-range restriction site mapping of the MHC in man (Hardy et al., 1986; Ragoussis et al., 1986; Lawrance et al., 1987) and mouse (Müller et al., 1987b), as well as other regions of mammalian genomes (see Chapter 1). Using PFGE Hardy et al. (1986) determined a physical map for the human class II region and established the order of the HLA-D subregions. Physical linkage between the human class II and class III loci has also been demonstrated using this technique (Ragoussis et al., 1986; Lawrance et al., 1987) but the relative position and orientation of these genes was not established. A complete map of the human MHC has not been achieved, however, due to the absence of informative probes. In mouse the orientation and molecular map position of the complement gene cluster has been established although the class II and class III genes were linked by only one restriction fragment and neither were linked to the major class I region (Müller et al., 1987b).

A number of questions, therefore, remain to be answered about the organisation of the human MHC.

1. What is the orientation of the complement gene cluster within the MHC?

2. What is the position of the complement genes relative to the HLA-B and DRA loci?

3. What is the exact position of the genes for TNFα and β within the human MHC and is it analagous to that in the mouse?

Making use of PFGE together with restriction enzymes that cut

rarely in mammalian genomes and Southern blotting (see Chapter 1), it was aimed to answer some of these questions by constructing a physical map of the human MHC. At the same time a parallel mapping project involving cosmid walking from the complement gene cluster was being undertaken by Carole Sargent (Sargent, 1988). Information from the cosmid clones obtained in that work was invaluable to the production of the map described here.

3.2 EXPERIMENTAL STRATEGY

Genetic data has suggested that the human MHC covers 3-4 cM. This size and the large number of well characterised loci within the region mean that the MHC is a suitable system for the application of large DNA restriction fragment mapping. The recent advances in this area have already been explained (Chapter 1).

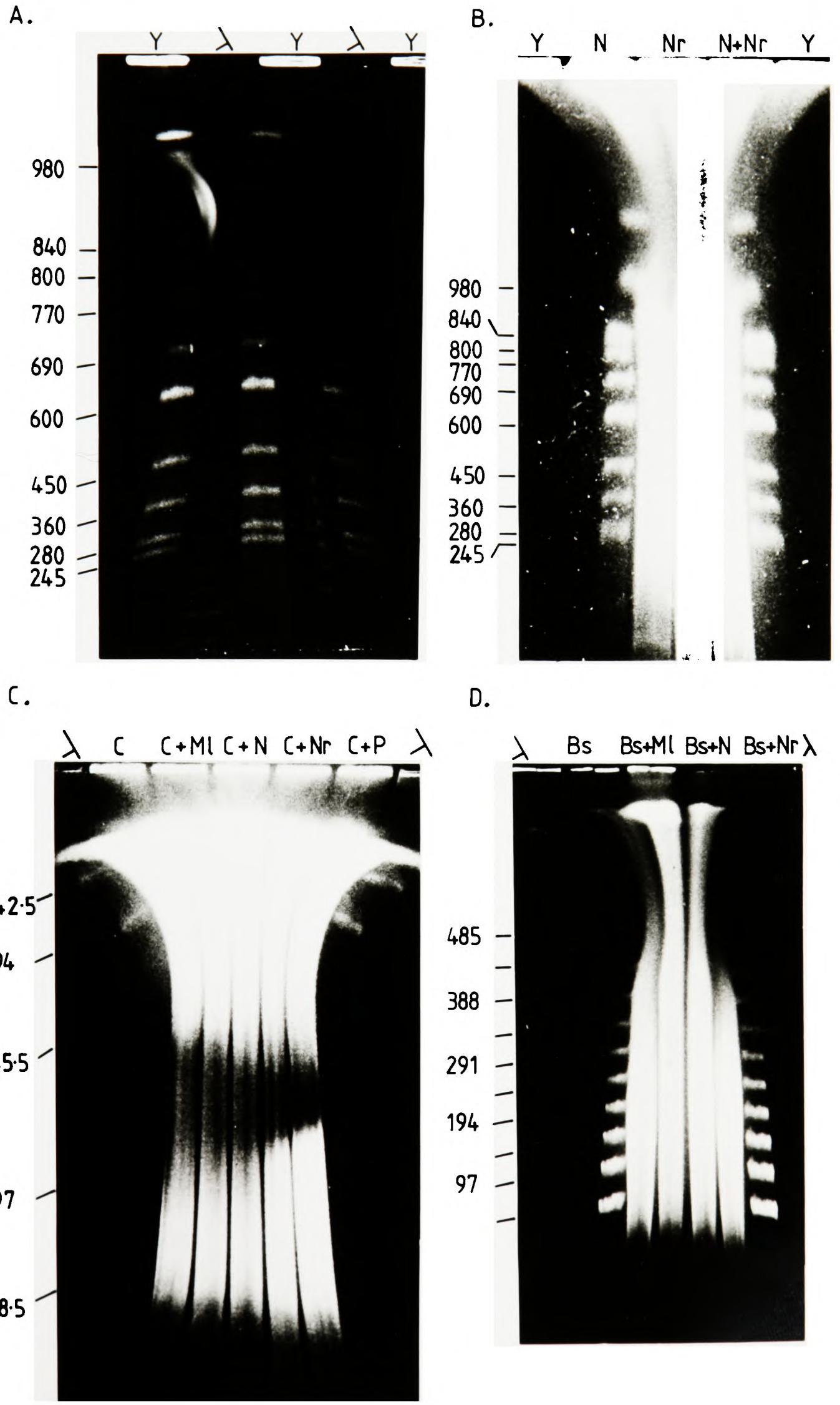
In order to construct a restriction map of the MHC in man, genomic DNA from a HLA homozygous cell line (Ice 5, HLA type-A2, B7, DR2, BfS, C2C, C4A3, C4BQ0) was used to minimise mapping problems caused by possible haplotype-specific RFLPs. The cell line used has only single C4 and 21-OHase loci on each copy of chromosome 6 (Carroll et al., 1985b; Yu & Campbell, 1987) which also serves to simplify restriction site mapping. High molecular weight DNA in agarose blocks was prepared from this cell line as described in section 2.3.5a. This DNA was digested with a range of restriction enzymes that cleave rarely in the human genome, alone or in double-digest combinations. The DNA fragments were then separated by PFGE at an appropriate switching interval and transferred to GeneScreenPlus membranes. These filters were hybridised successively with a panel of [^{32}P]-labelled cDNA and genomic probes specific for the class I, class II and class III regions of the MHC (see

section 3.3). The linkage of probes on one DNA fragment was established by observation of hybridisation of both probes to the same DNA restriction fragment when the same PFGE filter was used.

The possibility that the observed hybridisation to a common fragment was due to comigration of two different fragments was reduced by the use of data from a range of restriction enzyme digests mapped relative to one another by double digestion and by separating the same DNA samples at different switching intervals. Construction of a physical map of the MHC utilised this single and double digestion data, the known genetic data and information on the positions of probes and restriction sites from cosmid cloning.

The PFGE system used has been described in section 2.7.4. It was an OFAGE system which provides reasonable separation of DNA fragments in the 50-1000 kb size range but does suffer from a "focusing-in" effect. This means that only the central lanes of the gel can be used with any reliability and that the comparison of DNA fragment sizes in different lanes requires caution. Fig 3.2 illustrates these points with some examples of separations of DNA samples obtained with the OFAGE system. Similarly comparison of sizes obtained on different gel runs are difficult because both the amount of DNA loaded and differences in the gel running conditions, particularly the switching interval, can affect the apparent fragment size (Michiels, Burmeister & Lehrach, 1987; Bernards et al., 1986). Data obtained subsequent to that described here using the crossed field gel system of Southern et al. (1987) suggest that fragment sizes estimated with these gels were generally too large (Chapters 4 & 5).

Fig 3.2 Examples of separations obtained on the OFAGE system. Fragment sizes are given on the left in kb. A. Yeast chromosomes (Y) and concatemers (λ) separated at a 50 s switching interval. B. NotI (N) and NruI (Nr) single and double digests of genomic DNA. The switching interval was 65 s. C. ClaI (C) single and double digests of genomic DNA separated at a 7 s switching interval. Other enzymes are MluI (Ml) and PvuI (P). D. BssHII (Bs) single and double digests of genomic DNA separated at 30 s switching interval. All other experimental details are given in Chapter 2.



3.3 RESULTS

3.3.1 Probes

Most of the probes used have been described in section 2.1.6. Probe J is a 0.9 kb BamHI/XhoI genomic fragment isolated from a cosmid DNA insert and located ∿47 kb from the transcriptional start site of the C2 gene. Probe L is a 1.4 kb BamHI genomic fragment isolated from a cosmid clone containing the genes for TNFα and β and lies ∿10 kb from the 3' end of the TNFA gene (Fig. 3.3).

The authenticity of class I and class II probes was checked by hybridisation to the correct bands on Southern blots of genomic DNA digested with BamHI, BglII and KpnI (results not shown). It should be noted that because of the relatedness of the sequences (Strachan, 1987; Trowsdale, 1987) cross-hybridisation is observed between some of the loci. The DPA cDNA cross-hybridises to a limited extent with DRA and DQA sequences. All the class II B gene probes hybridise to other B gene sequences due to regions of related sequence particularly in the third exon. The HLA-B and -C "specific" probes also cross-hybridise with each other because of some related sequences in the introns from which they are derived (Strachan et al., 1986). Cross-hybridisation can be minimised by the use of stringent washing conditions and in general this problem was not serious because the "real" bands were significantly more intense than those due to cross-hybridisation. The general class I probe hybridises to all class I-related sequences including the HLA-B and -C genes by virtue of the sequences encoding the β_2-microglobulin binding domain which are highly conserved between class I genes.

Fig 3.3 Molecular map of the human MHC. A. Overlapping cosmid clones from the complement/21-OHase and TNF regions. The locations of genes are shown by filled boxes (C2 gene, C2; Factor B gene, Bf; C4A gene, C4; 21-OHaseB gene, 21OH) and the 5' 3' orientation indicated by horizontal arrows. Vertical arrows indicate the positions of the probes used in PFGE analysis. The positions of infrequently cutting restriction endonuclease sites known to cut in genomic DNA are indicated by arrow heads, other sites by vertical bars. Restriction enzyme sites are designated as follows: C, ClaI; Ml, MluI; N, NotI, Nr, NruI; Pa PaeR7I; Pv, PvuI; S, SalI. B. Restriction map obtained by Southern blot analysis of genomic DNA separated by PFGE after single and double digests. Filled boxes indicate the exact position of genes, open boxes indicate the limits of the region hybridising to the probes used (see text). The cross-hatched box indicates the probable position of the DRA gene (DR_α, see Text). For convenience the 950 kb NotI fragment that hybridised with the B and C locus-specific probes is not to scale. A, the 675 kb and 400 kb NotI fragments that hybridised with the class I cDNA probe (Fig 3.4) and lie telomeric to the C locus are not shown. In some cases partial digests were obtained, as represented by two sets of data for some enzymes.*

A

Probes: 21-OH, J, L TNFα

Genes: 21-OH C4, Bf C2, TNFα TNFβ

Cosmids

Restriction sites in cosmids: SS C S C S C Nr C C N S Nr S C S 10kb
Nr Pv Ml Ml Ml Pa C S

B

100kb

Subregions and genes: DP DQ DR 21-OH C2 TNF B C A*

Not I

Nru I

Pvu I

Mlu I

Pae R7 I

Sal I

Bss H II

Cla I

3.3.2 Isolation of overlapping cosmid clones from the class III region

Approximately 230 kb of genomic DNA encompassing the C2, Factor B, C4 and 21-OHase B genes was isolated in a series of overlapping cosmid clones (Fig 3.3A) by Carole Sargent. This cluster of genomic clones was mapped with restriction enzymes having 6 or 8 bp long recognition sites and containing one or two CpG dinucleotides. Since CpG is known to be under-represented in the human genome (Bird, 1986) these enzymes are consequently of use in construction of long-range restriction maps in conjunction with PFGE (Brown & Bird, 1986). Examination of these restriction site data (Fig 3.3A) revealed a striking cluster of sites for infrequently cutting enzymes (3 MluI, 2 ClaI, 1 NotI and 1 NruI,) in a 25 kb region located about 25 kb from the transcriptional start site of the C2 gene. A single copy genomic hybridisation probe (probe J, Fig 3.3A) was isolated from the cosmid DNA distal to this cluster of restriction sites and the C2 gene.

More detailed mapping of the cosmid clones obtained here for other rare-cutting restriction enzymes revealed other sites arranged in clusters (see Chapter 6). In particular there is a dense cluster of sites around the NotI site identified in Fig 3.3A.

3.3.3 Orientation and location of the complement genes in the class III region

Hybridisation of NotI-digested DNA with DQA, DRA and 21-OHase probes identified a common 980 kb NotI fragment that links the DQ and DR subregions with the complement loci (Fig 3.4). A linking NotI fragment of similar size has previously been observed by others (Ragoussis et al., 1986; Lawrance et al., 1987). This fragment was cut by NruI to

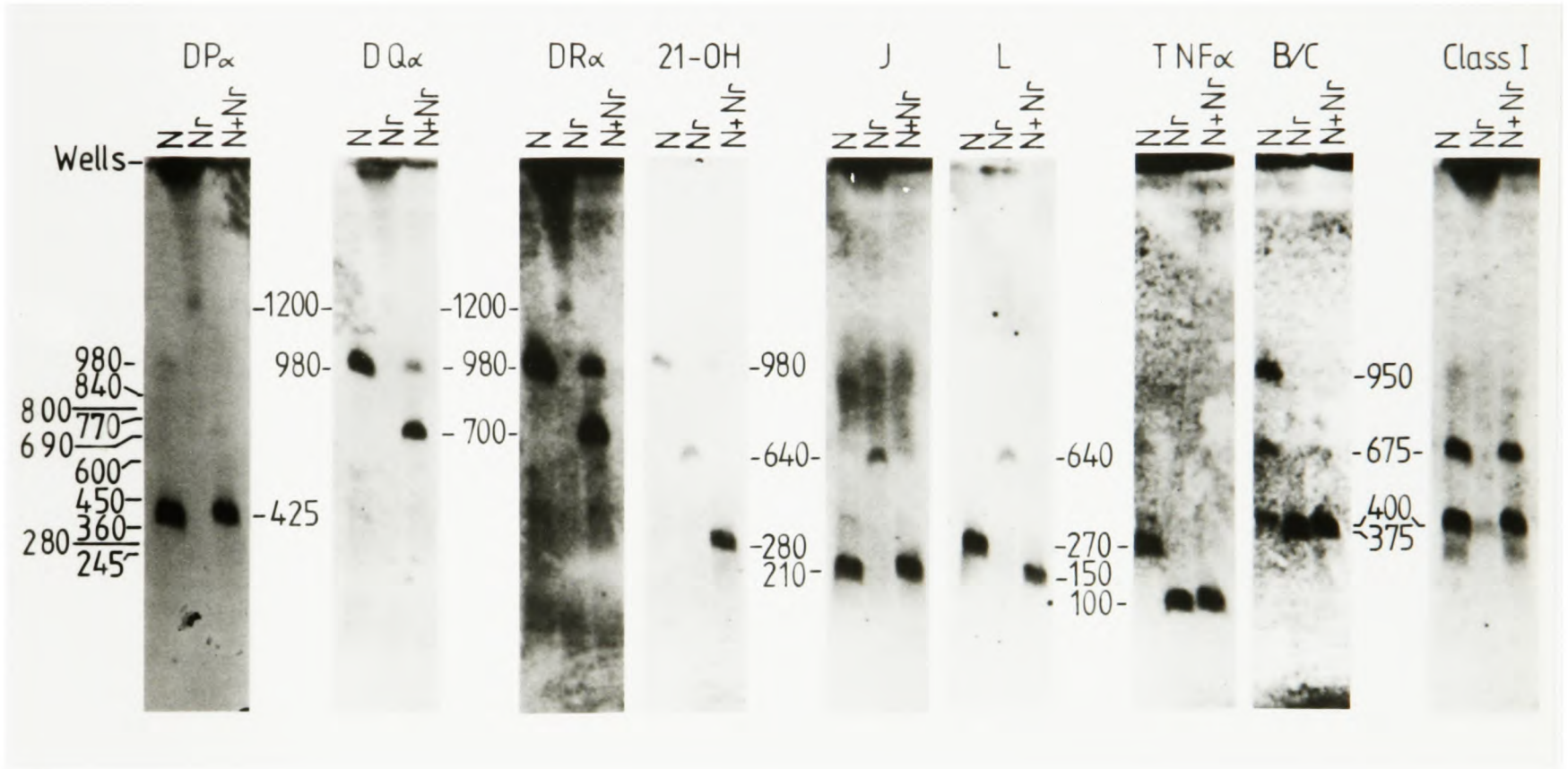


Fig 3.4 Southern blot analysis of genomic DNA digests separated by PFGE. Fragments generated by NotI (N), NruI (Nr), or NotI+NruI (N+Nr) were separated using a 65s pulse interval. All autoradiograms were obtained from a single filter hybridised sequentially with the probes shown. B/C indicates that the B and C locus-specific probes gave the same hybridisation pattern and only the B locus probe result is shown. Yeast chromosomes were run as size markers and are indicated on the left. Fragment sizes are given in kb. The size of the class II-hybridising NruI fragment is approximate.

yield a fragment of 700 kb with the DRA and DQA probes and a fragment of 280 kb with the 21-OHase probe (Fig 3.4). Probe J from the class III cosmid cluster, however, hybridised to a NotI fragment of 210 kb. Since the exact position of the NotI site is known from the cosmid map (Fig 3.3A) these data establish the orientation of the complement genes relative to the class II genes, the C2 gene being telomeric to the 21-OHase B gene which is closest to the DRA locus.

Hybridisation of the 21-OHase and J probes to PvuI single and double digests (Fig 3.5) was consistent with the map shown in Fig 3.3B. Both probes hybridised to a 340 kb PvuI fragment. In the PvuI/NotI double digest the 21-OHase probe hybridised to a 270 kb fragment while probe J hybridised to a 80 kb fragment consistent with the location of a PvuI site ∿100 kb from the 5' end of the C2 gene. Later experiments revealed the presence of two PvuI sites separated by only 11 kb at this point (Sargent et al., 1988). The minimum distance between the complement gene cluster and the DRA gene was estimated from data provided using SalI (Fig 3.6). In the SalI digest the 21-OHase probe hybridised to a fragment of 280 kb and a partially digested fragment of 490 kb. The NotI/SalI double digest positioned a SalI site 85 kb telomeric to the NotI site and this was confirmed by hybridisation of the blot with probe J. Hybridisation of the same blot with the DRA probe revealed no common fragments. Thus the minimum distance between the 21-OHaseB gene and the DR subregion is 300 kb.

3.3.4 Mapping of the class II subregions

Hybridisation of class II probes with DNA singly and doubly digested with a number of enzymes allowed the limits of the class II subregions, DP, DQ and DR, to be defined (see Table 3.1 for summary of

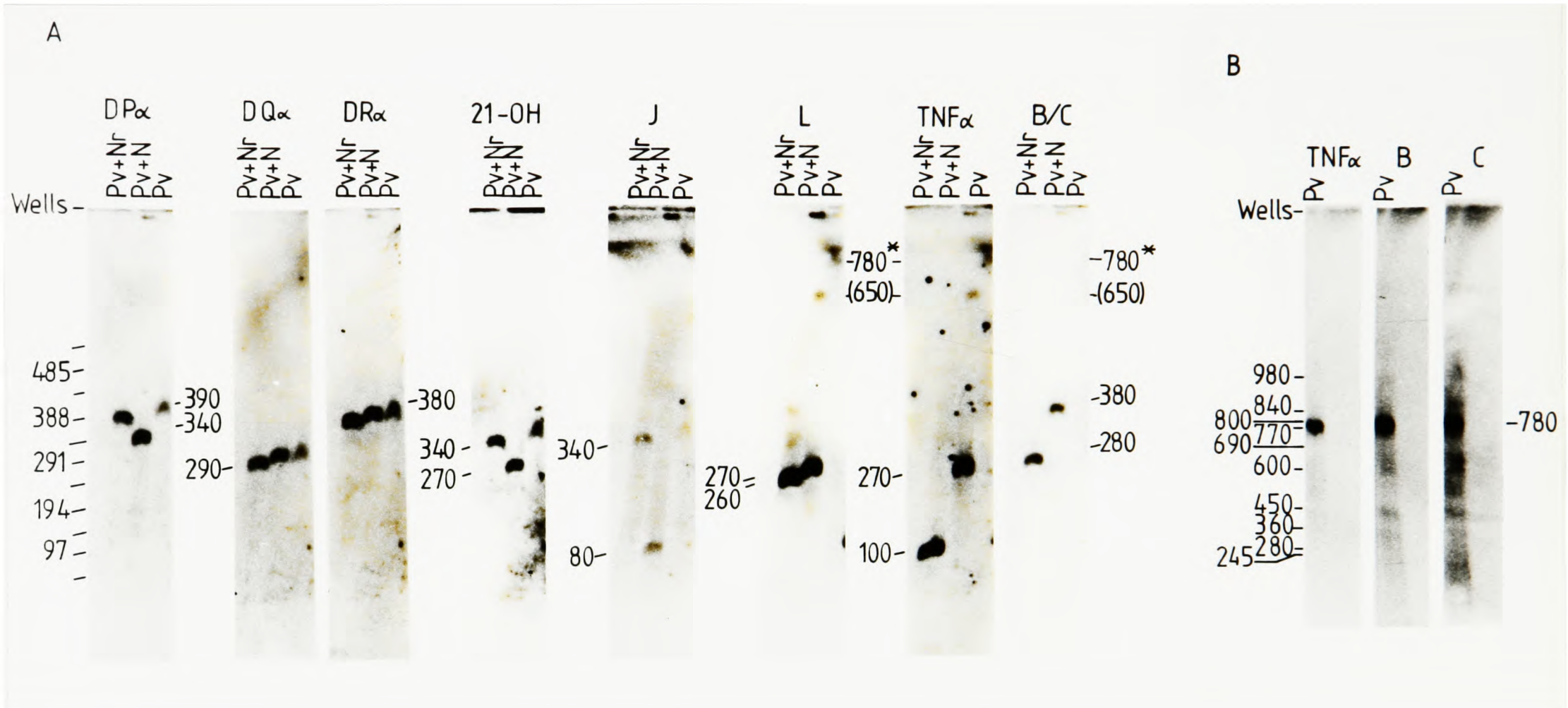


*Fig 3.5 Southern blot analysis of PvuI-digested DNA separated by PFGE. A. Sequential hybridisation of the probes shown to a single Southern blot of PvuI (Pv), PvuI+NotI (Pv+N) and PvuI+NruI (Pv+Nr) digested DNA separated using a 30s pulse interval. Concatemers of λ cI857 S7 DNA were run as size markers and are indicated on the left. Fragment sizes are given in kb. Partial digestion products are enclosed in brackets. *The 780 kb PvuI fragment is completely resolved in B. B. Sequential hybridisation of TNFA (TNFα), B-specific (B) and C-specific (C) probes to PvuI-digested DNA resolved using a 60s switching interval. Yeast chromosomes were run as size markers and are indicated on the left.*

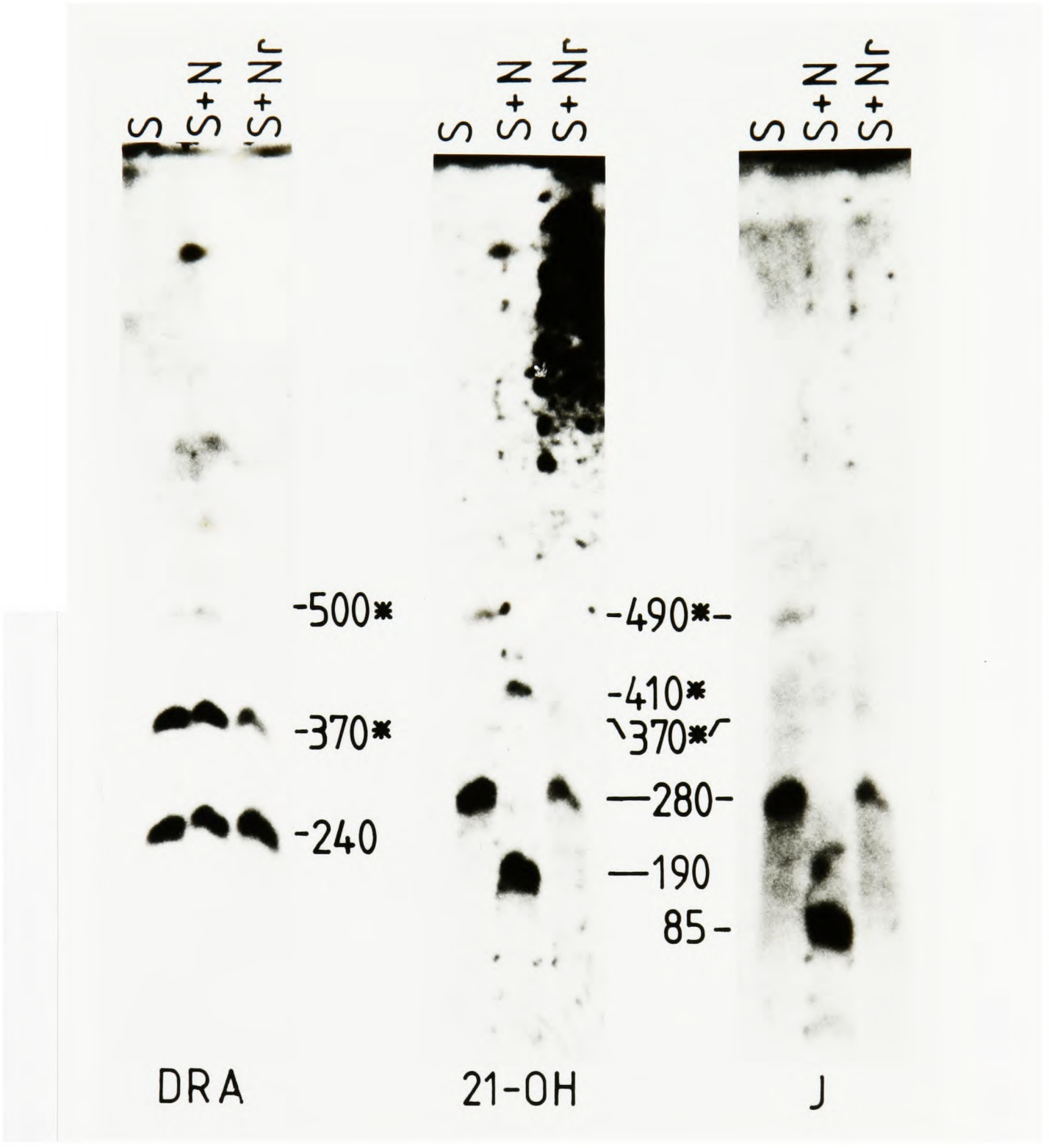


*Fig 3.6 Southern blot analysis of SalI digested DNA separated by PFGE. Sequential hybridisation of the probes DRA, 21-OHase (21-OH) and J to a single Southern blot of SalI (S), SalI+NotI (S+N) and SalI+NruI (S+Nr) digested DNA separated using a 35s switching interval. Concatemers of λ cI857 S7 DNA were run as size markers. Fragment sizes are given in kb. Fragments indicated by *, partial digest products (see Table 3.1 and text).*

*Table 3.1 Size of restriction fragments (kb) observed after hybridisation of a panel of MHC-region probes to Southern blots of restriction enzyme digests of the DNA. For the DP and DR subregions, where different fragments were observed with A and B gene probes, these are indicated by A and B. Multiple sets of data indicate that several fragments were observed and the major mapped fragments are given. Where it is known that the higher mol. wt. fragment is the product of partial digestion this is indicated by *. Fragments that link probes within a digest are underlined. - indicates that no fragment was observed using PFGE, ND indicates not determined.*

	PROBES									
	Class II			Class III				Class I		
	DP	DQ	DR	21OHase	J	L	TNF	HLA-B	HLA-C	HLA-ABC
ENZYME										
NotI	425	980	980	980	210	270	270	950	950	675 400
NruI	>1200	>1200	>1200	640	640	640	100	375	375	-
NotI/NruI	425	700	700	280	210	150	100	375	375	675 400
PvuI	390	290	380	340	340	780	780	780	780	-
PvuI/NotI	340	290	380	270	80	270 650*	270 650*	380 650*	380 650*	ND
PvuI/NruI	390	290	380	340	340	260	100	280	280	ND
MluI	B90 A440* 290	550 440* 150	550	225	270	190	190	360 180 130	360 180	ND
MluI/NotI	B90 A245	550 210* 150	550	225	210	190	190	360 180 130	360 180	ND
MluI/NruI	B90 A440* 290	490 440* 150	490	225	270	90	100	240 180 130	240 180	ND

Continued overleaf....

Table 3.1 continued

	DP	DQ	DR	21OHase	J	L	TNF	HLA-B	HLA-C	HLA-ABC
MluI/PvuI	B90 A290	150	380	225	85	ND	190	360 180 130	ND	ND
SalI	240 380*	260 500*	240 500*	280 490*	280 490*	220	220	420	420	ND
SalI/NotI	240 380*	260 500*	240 500*	190 410*	85	220	220	390	390	ND
SalI/NruI	240 380*	260 500*	240 500*	280 370*	280 370*	190	30	290	290	ND
BssHII	240	190 480	480	100	150	ND	340	ND	ND	ND
BssHII/MluI	130	150 480	480	100	80	ND	ND	ND	ND	ND
ClaI	230 280*	110 140	B80 A110 240*	140	85	ND	165	140 195*	40 80*	ND
ClaI/MluI	B90 A95 140	100 140	BND A110	ND	80	ND	120	120	20	ND
ClaI/NruI	230 280*	110 140	A110 BND	ND	85	ND	30	140 195*	40 80*	ND
ClaI/SalI	ND	ND	B80 A60 110	ND	ND	ND	ND	ND	ND	ND
PaeR71	ND	ND	270	ND	280	-	290 490*	290 490*	ND	ND
PaeR71/NotI	ND	ND	270	ND	210	-	95	180	ND	ND
PaeR71/NruI	ND	ND	270	ND	280	-	95	80	ND	ND
PaeR71/MluI	ND	ND	270	ND	270	-	95	20	ND	ND

fragment sizes). Although both DQA and DRA probes hybridised to a common 980 kb NotI fragment, the DPA and DPB probes recognised a separate 425 kb NotI fragment. It was already known from genetic data that the DP subregion lies centromeric to DQ and DR (Shaw et al., 1981) and this had been subsequently confirmed by PFGE (Hardy et al., 1986). The DP subregion NotI fragment must therefore lie centromeric of DR and DQ. This was confirmed by the identification of a common very large (>1200 kb) NruI fragment that links all the class II probes but does not include the class III genes. Double digestion with NotI/NruI produced a 700 kb fragment that hybridised with the DQA and DRA probes and a 425 kb fragment that hybridised with the DPA probe. This is consistent with the linkage of the class II subregions as shown (Fig 3.3B).

Hybridisation of class II probes to PvuI single and double digests showed that each subregion lies on a different PvuI fragment (Fig 3.5). The PvuI/NotI digest probed with DPA mapped the 390 kb DP region-containing PvuI fragment relative to a NotI site which was subsequently shown to be the telomeric site from the results of MluI and MluI/PvuI digests probed with the DPA and DPB probes (Fig 3.7, Table 3.1, see below).

Hybridisation of the DR probe to MluI digested DNA identified a 550 kb fragment. This MluI fragment was cleaved by NruI to 490 kb consistent with the 550 kb MluI fragment being adjacent to the MluI fragment that contains the complement gene cluster (Fig 3.7). The DQA probe also hybridised to this 550 kb MluI fragment. In addition the DQA probe recognised 150 kb and 440 kb MluI fragments. Hybridisation of the DPA probe to the MluI digest suggested that the 440 kb fragment which was in common between the DPA and DQA probes was the product of partial digestion. The 150 kb MluI fragment lies adjacent to the 550 kb MluI fragment that includes the DR subregion and there must be DQA sequences

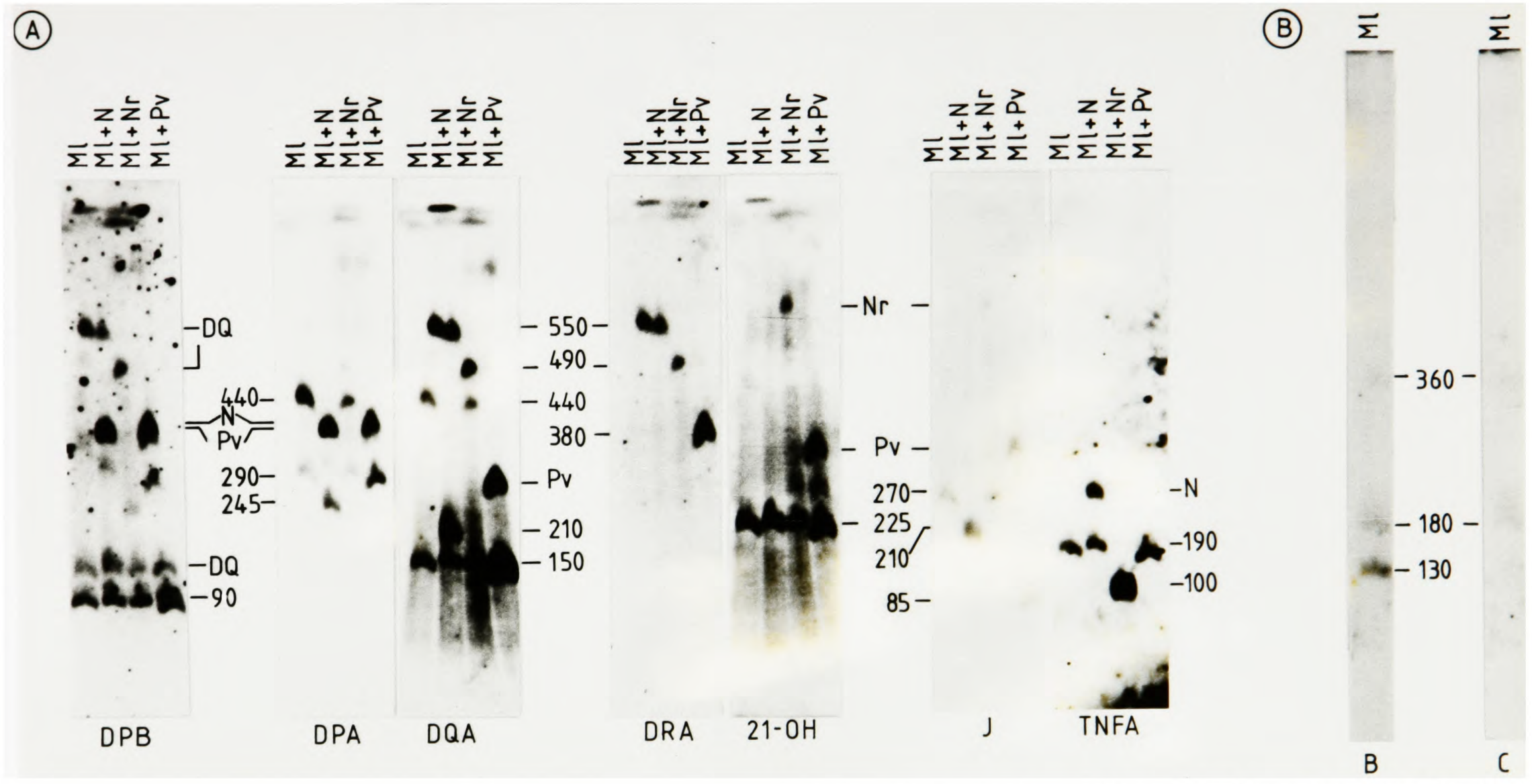


Fig 3.7 Southern blot analysis of MluI digested DNA separated by PFGE. A. Sequential hybridisation of the probes shown beneath each panel to a single Southern blot of MluI (Ml), MluI+NotI (Ml+N), MluI+NruI (Ml+Nr) and MluI+PvuI (Ml+Pv) digested DNA separated using a 30s switching interval. Concatemers of λ cI857 S7 DNA were run as size markers. Fragment sizes are given in kb. Fragments indicated with letters representing restriction enzymes are the single digest products of the enzyme other than MluI. In the DPB panel fragments indicated by DQ are thought to be due to cross-hybridisation with DQB genes. B. Sequential hybridisation of B-specific (B) and C-specific (C) probes to MluI-digested DNA resolved at a 30 s switching interval. Markers were concatemers of λ cI857 S7 DNA.

divided by the cleaved MluI site. A similar situation occurs with the DQA probe hybridised to BssHII-digested DNA (Table 3.1, Fig 3.3A). This is probably because the DQA probe hybridises with both the DQA gene and the highly related DQA2 sequence.

The DPA and DPB probes lie on separate sets of MluI fragments, the DPA probe hybridising to 290 kb and 440 kb MluI fragments while the DPB probe hybridised to a 90 kb fragment. Cosmid cloning studies suggest the presence of a pair of DPA and DPB genes and a pair of related pseudogenes in the DP region arranged in the order DPB2, DPA2, DPB1, DPA1 from the centromere (see Trowsdale, 1987 for review). The fact that the DPA and DPB probes hybridise to different MluI fragments may be because the A genes are less related than the B genes. At high washing stringency the DPA cDNA probe may hybridise only to the expressed DPA1 gene. In addition there is a cluster of rare-cutting sites at the 5' (telomeric) end of the DPB1 gene including a MluI site (John Trowsdale, personal communication) so that both DPB genes are in the same MluI fragment which also contains the DPA2 gene. The 290 kb MluI fragment that hybridised with the DPA probe is cleaved by NotI to a 245 kb fragment. Since the DPA probe is known to be linked to the DQA probe on a 440 kb partial fragment whereas the DPB MluI fragment is not, the NotI site that is cleaved must lie telomeric to DP. The 440 kb MluI fragment is also cleaved by PvuI (Fig 3.7, Table 3.1) and therefore as mentioned above the PvuI fragment that contains DP must lie as shown in Fig 3.3B.

Further data obtained from single and double digests as shown in Table 3.1 were consistent with the location of the class II subregions as shown in Fig 3.3B. The map of the class II region shows a number of striking similarities to that published by Hardy et al. (1986), and confirms the organisation of the class II subregions. Although the order of the A and B genes within a subregion was not determined, the approx-

imate position of the DRA gene was defined from the result of ClaI and ClaI/SalI digests. A 290 kb partial ClaI fragment which hybridised with the DRA probe, but not the DRB probe, was cut down by SalI to 60 kb (Fig 3.3B). Assuming the position of the SalI site that is cutting is defined by the 490 kb partial seen with the 21-OHase probe, this would suggest that the DRA gene lies 300 kb to 360 kb from the 21-OHaseB gene. Superimposing the restriction map of Hardy et al. (1986) onto the map described here suggests that this location is accurate (see section 3.4). Results from a cell line that carries the DR3 haplotype (see Chapter 5) have allowed the location of the DRA and DRB genes to be further defined and suggest that the distance between DRA and 21-OHaseB is 330-390 kb.

3.3.5 Mapping of the TNFA gene in the MHC

Recently the genes for TNFα and β have been linked to the human MHC (Spies et al., 1986), and in mouse they have been mapped about 70 kb from the H-2D region, the murine equivalent of HLA-B (Müller et al., 1987a). Hybridisation of the TNFA probe to NotI and NruI digested DNA failed to reveal any bands in common with the other MHC region probes (Fig 3.4). However, in the PvuI digest a 780 kb fragment was identified with the TNFA probe which also hybridised with the HLA-B and -C locus probes (Fig 3.5).

To establish the position of the TNFA gene relative to the known MHC loci, a cluster of overlapping cosmid clones covering ∿82 kb of genomic DNA surrounding the TNFA gene was isolated and characterised by Carole Sargent (Fig 3.3A). The position of the TNFA gene was located by restriction enzyme mapping and Southern blot analysis. The position of the TNFB gene was inferred by comparison with the previously published

map (Nedospasov et al., 1985). These cosmids were mapped for infrequently cutting restriction enzymes. A single NruI site was found 2.5 kb from the 3' end of the TNFA gene (Fig 3.3A). A single copy hybridisation probe (Probe L, Fig 3.3A) was isolated from the cloned cosmid DNA distal to the NruI site and the TNFA gene. When hybridised to NruI digested genomic DNA this probe detected the same 640 kb NruI fragment that contains the complement genes (Fig 3.4). This result confirmed the linkage of the complement genes and the TNFA gene and, from the sizes of the NotI/NruI fragments detected with probes J and L and the known positions of these sites relative to the genes, established that the C2 gene lies 390 kb centromeric to the TNFA gene. In addition the cosmid restriction map data indicates that the TNFB gene lies telomeric to the TNFA gene.

Further single and double digests were carried out to estimate the distance between the TNFA and HLA-B loci. A 290 kb PaeR7I fragment was found to hybridise with both probes (Fig 3.8A), but not probe L. Restriction mapping of the cosmid DNA for PaeR7I revealed 5 sites as shown in Fig 3.3A, one of which lies within the TNFA gene. To establish which of these sites are cut in genomic DNA, a single PaeR7I digest and double digests were resolved by conventional gel electrophoresis and blotted onto nitrocellulose. Hybridisation of the TNFA probe revealed a 3.9 kb PaeR7I fragment produced by complete cleavage at the sites within the TNFA gene and immediately adjacent (Fig 3.8B). Hence, the TNFA probe hybridised to both a 290 kb fragment resolvable by PFGE and a 3.9 kb fragment which can only be seen on standard agarose gels. Probe L hybridised to a series of fragments consistent with partial cleavage at the remaining three PaeR7I sites (Fig 3.8B). No PaeR7I fragment can be seen with probe L by PFGE analysis. The double digests with BamHI and EcoRV confirm this observation and generate the appropriate fragments

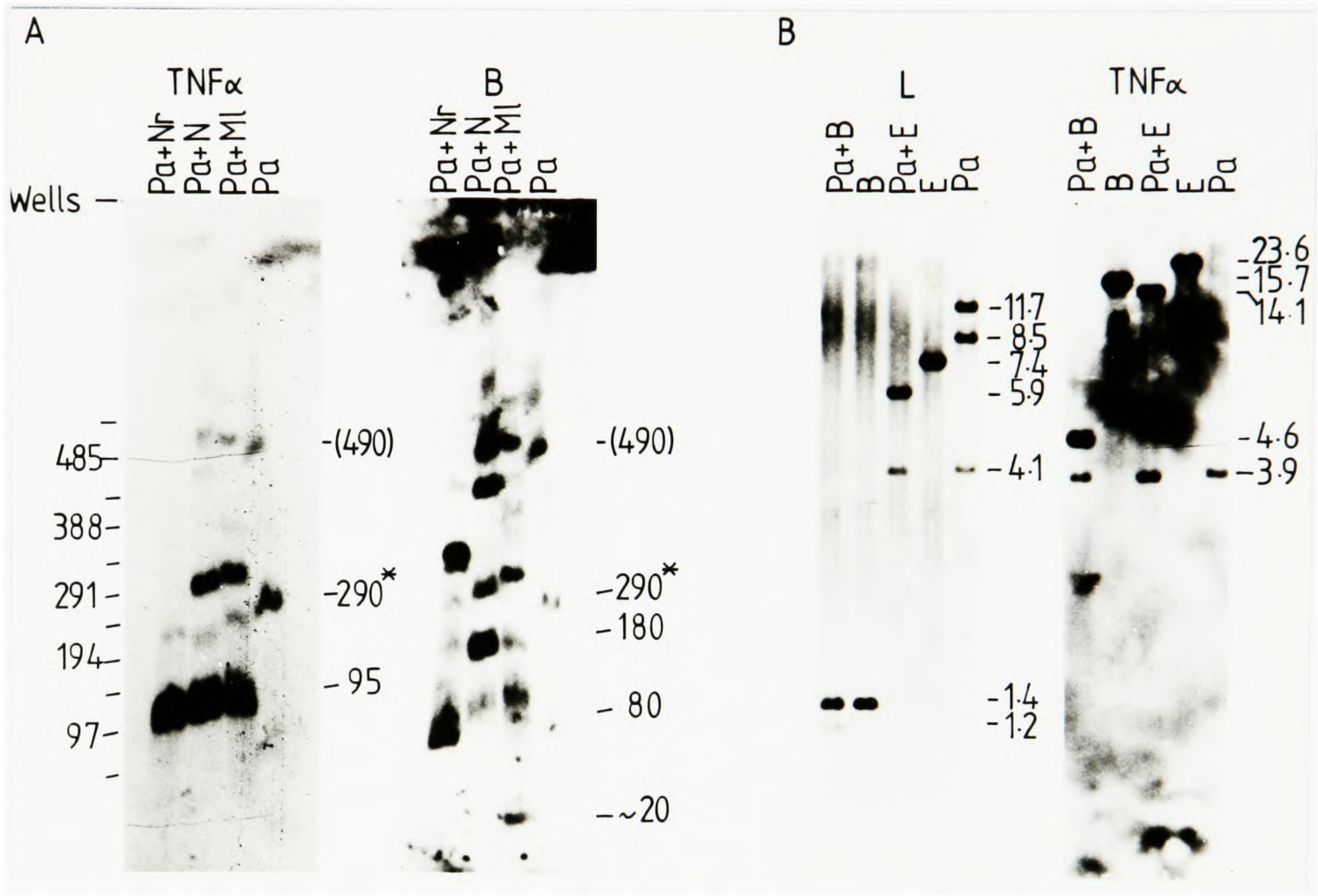


*Fig 3.8 Southern blot analysis of PaeR7I-digested DNA. A. Hybridisation of TNFA (TNFα) and B-specific (B) probes to PaeR7I (Pa) single and double digests separated using a 30s pulse interval. Other restriction enzymes are designated as before. Positions of λ DNA concatemers are indicated on the left. Fragment sizes are in kb. A partial PaeR7I digestion product is in brackets. *The 290kb PaeR7I fragment has been sized on other blots and here shows slightly anomalous mobility. Unmarked fragments are the products of partial digestion which have not been mapped. B. Southern blot analysis of PaeR7I sites in genomic DNA around the TNFA gene. Genomic DNA was digested with single and double digest combinations of PaeR7I, EcoRV (E) and BamHI (B), separated on a conventional 0.7% agarose gel and hybridised with probe L and the TNFA (TNFα) probe. The three bands in the PaeR7I lane with probe L are due to partial cleavage at two of the five PaeR7I sites mapped in cosmid DNA in this region (Fig 3.3A). The TNFA probe hybridised to a 3.9 kb PaeR7I fragment due to cleavage at the PaeR7I site within the TNFA gene and also to the 290 kb fragment which can only be resolved by PFGE (see panel A). Fragment sizes are in kb.*

predicted from the restriction map of the cloned DNA (Nedospasov et al., 1985; Sargent, 1988).

The results from a series of double digests of PaeR7I with NotI, NruI and MluI (Fig 3.8A) allowed the map in Fig 3.3B to be constructed showing that the TNFA and B genes lie about 250 kb centromeric to the HLA-B gene.

3.3.6 Mapping of the class I region

The B and C locus-specific probes hybridised to the same fragments in NotI, PvuI and SalI restriction digests. However, on the basis of hybridisation to MluI, ClaI and MluI/ClaI digests, the B and C loci are separated by at least 80 kb (Table 3.1). Both B- and C-specific probes hybridised to 360 kb and 180 kb MluI fragments (Fig 3.7). The B-specific probe also hybridised to a 130 kb MluI fragment.

Hybridisation of the general class I probe to PFGE blots did not reveal any stongly hybridising fragments that linked the HLA-B and C probes with the HLA-A region (Fig 3.3A, Table 3.1). The general class I probe hybridised only weakly to the NotI, NruI and NotI/NruI bands containing the HLA-B and C genes (Fig 3.4). This is probably due to two factors. First, the probe is derived from a HLA-A or HLA-A-related cDNA and hybridises more strongly to the A-like sequences which are more closely related to each other than to the B and C genes (Jordan et al., 1985). Second, it is likely that a large number of class I-like genes are located telomeric to HLA-A (Orr & DeMars, 1983) and, hence, the more intense bands may contain many class I-like sequences.

From sequence studies of HLA class I genes it is known that the 5' ends of class I genes are relatively rich in CpG dinucleotides (Tykocinski & Max, 1984; Pontarotti et al., 1988) and these sequences

are unmethylated for the HLA-A, B, C and some other genes (Pontarotti et al., 1988). Therefore it is likely that the HLA-A gene is separated from the B and C genes by at least one cleavable cluster of sites for rare-cutting restriction enzymes and this may explain why they cannot be linked (Chimini et al., 1988; Pontarotti et al., 1988).

3.3.7 Long Range Restriction Map of the MHC

The complete linkage map (Fig 3.3B) between the individual loci can be derived from the NotI, NruI, PvuI and MluI fragments shown in Figs 3.4, 3.5 and 3.7. The restriction enzyme digestion data presented here is sufficient to provide overlapping fragments from DP to HLA-C. The map is, therefore, completely internally consistent. For instance because there are fragments that link the complement genes to the TNF genes and the TNF genes to the class I genes, the orientation of the complement gene cluster is confirmed.

The estimated maximum size of the human MHC from the limits of the DP region to the HLA-A hybridizing NotI fragments (Fig 3.3) is approximately 3800 kb from the sum of the NotI fragments that hybridised to the MHC probes. This assumes that the NotI fragments that contain the HLA-A gene are immediately adjacent to the HLA-B NotI fragment. This estimate would be in good agreement with the estimated size of the human MHC from the genetic data (see section 3.4).

3.4 DISCUSSION

Pulsed field gel electrophoresis, together with restriction enzymes that cut genomic DNA infrequently and Southern blotting, have been used to produce a long range restriction map of the human MHC

(Dunham et al., 1987). The orientation of the complement and 21-OHase loci relative to the class I and class II loci has been established, the 21-OHaseB gene being telomeric to the C2 gene. The distance between the 21-OHaseB gene and the DRA locus is 300-360 kb, while that between the C2 gene and the HLA-B locus is ∿650 kb. The genes for TNFα and β have been located precisely within the MHC, the TNFA gene lying 390 kb telomeric of the C2 gene and ∿250 kb centromeric of HLA-B.

The total size of MHC hybridising fragments is ∿3800 kb which represents about 1/750 of the human genome. This is in good agreement with the estimate of 3-4 cM from recombinational events in family studies. Moreover the physical distances between the complement loci and the flanking HLA-DR region and HLA-B are very much like the genetic distances from the linkage studies discussed in Robson & Lamm (1984, Fig. 3.1 and Table 3.2). Therefore there appears to be no evidence for altered recombination rates in this region and the accuracy of these genetic estimates appears to be very high in this region.

Table 3.2 Comparison of predicted genetic and physical distances.

Region	*Genetic distance (cM)*	*Physical distance (kb)*
HLA-D	*1-2*	*1000*
C4 to HLA-DR	*0.3*	*300-360*
C2 to HLA-B	*0.7*	*650*
HLA-B to HLA-C	*0.1*	*>80*
HLA-C to HLA-A	*0.7*	*>600*

The data confirms the organisation of the class II region proposed previously (Hardy et al., 1986). Comparison of the restriction fragment map obtained by Hardy et al. (1986) with that presented here allows the

proposed organisation of the class II genes to be superimposed approximately on to the subregion map (Fig 3.9). The combined data suggests that the DRA gene lies within the limits predicted by the results presented here. However Hardy et al. (1986) estimated that the total size of the class II region is 1100 kb whereas the data presented here suggest a figure of 1000 kb. In Chapter 5 results are described which show that the DNA organisation of the DR2 (Ice 5) and DR4 (Hardy et al. 1986) haplotypes in the class II region differ by the presence of an extra 100 kb in the DRB to DQ region in the DR4 haplotype. This may explain the difference in the estimates of the total size of the class II region. Although the DR4 DQ and DR subregion organisations can be superimposed on the DR2 subregion intervals this is possible because the subregion limits defined here are necessarily quite large (230 kb for the DQ subregion, 190 kb for DR).

The estimated size of the class II region by PFGE corresponds with the lowermost limits of the genetic distance estimates. It is known that a high frequency of recombination occurs between the DP and DQ/DR genes (Termijtelen et al., 1983) whereas no recombination has yet been reported between DQ and DR which are in very strong linkage disequilibrium. The PFGE maps suggest that the physical distances between the subregions are similar. Therefore it is possible that, as has been previously proposed, a recombinational hotspot may exist between DP and DQ (Bodmer & Bodmer, 1984) and that this may increase the observed genetic distances in the HLA-D region.

Since the completion of this work a number of other partial maps of the HLA region have been published (Carroll et al., 1987; Inoko & Trowsdale, 1987; Ragoussis et al., 1988; Pontarotti et al., 1988; Tokunaga et al., 1988; Inoko et al., 1988). By and large these maps and the other previously published maps (Hardy et al., 1986; Ragoussis et

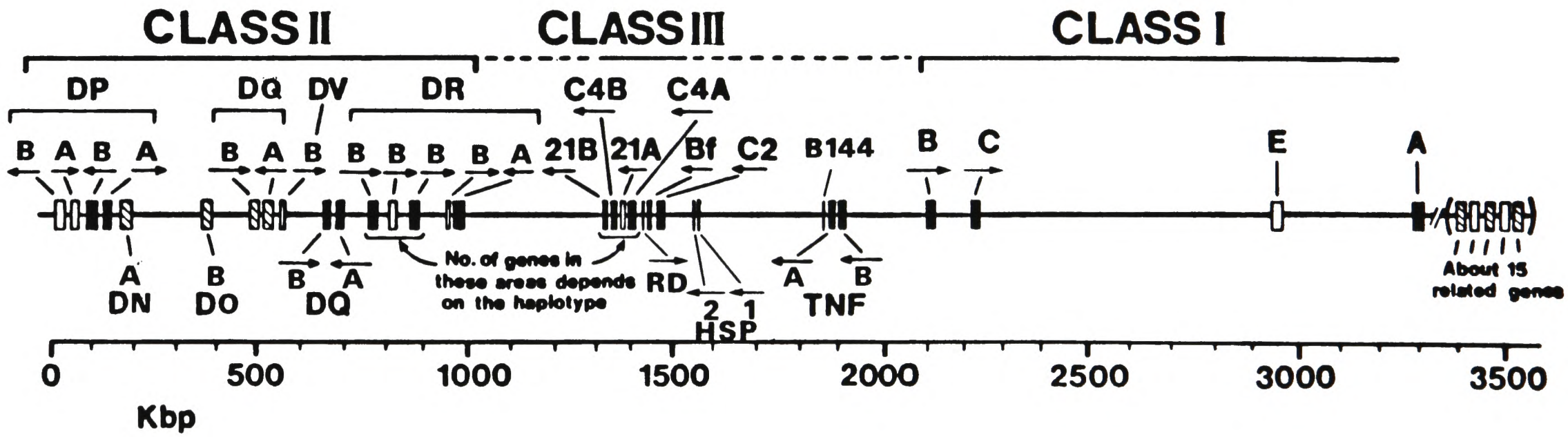


Fig 3.9 Complete updated map of the human MHC. The centromere is to the left. Horizontal arrows indicate the transcriptional orientation of the genes where this is known. Closed boxes are expressed genes, open boxes are pseudogenes and cross-hatched boxes are genes of unknown status. Regions and subregions are indicated above the genes. Data on the locations of various genes and subregions are taken from references cited in Chapters 1 and 3. The positior of the human B144 gene is taken from Sargent (1988). The dimensions of the map are as established in this work.

al., 1986; Lawrance et al., 1987) are in broad agreement with the map presented here although there is considerable variation in the sizes estimated for certain fragments. Ragoussis et al. (1986), Lawrance et al. (1987), Carroll et al. (1987) and Tokunaga et al. (1988) all demonstrate linkage of the DQ, DR and complement genes on the basis of the large NotI fragment. However, none of these groups were able to accurately position the class II subregions or the complement gene cluster and could not orientate the complement genes. Therefore there has been no independent confirmation of the organisation presented here for the human class III region. Lawrance et al. (1987), Carroll et al. (1987) and Tokunaga et al. (1988) all present restriction fragment maps of the class II and class III regions but because they lacked sufficient probes in the class III region they could not place these fragments accurately. In the case of Lawrance et al. (1987) the two restriction enzymes used were not mapped relative to each other, so that their "map" is in fact a collection of HLA-hybridising restriction fragment sizes. Hence they were unable to predict the gap between the complement loci and the HLA-B locus. Tokunaga et al. (1988) present a map that relies heavily on the previously published data (uncredited).

Carroll et al. (1987) and Ragoussis et al. (1988) provide convincing confirmation that the genes for TNFα and β are located telomeric to the C2 gene. Carroll et al. (1987) and Inoko & Trowsdale (1987) also locate the TNFA and B genes relative to the HLA-B gene and this result has subsequently been confirmed by cosmid cloning (Spies et al. unpublished data, referenced in Carroll et al. (1987)). Both Carroll et al. (1987) and Lawrance et al. (1987) have located the positions of different new class I genes. However as yet the HLA-B and C genes have not been linked to the HLA-A gene.

Differences in observed fragment sizes can be explained by

variation in the PFGE systems used, by haplotype-specific RFLPs, or by insertions or deletions of DNA (see Chapter 5). There could also be methylation differences between cell lines leading to some sites not being recognised, as some enzymes appear to be sensitive to cytosine methylation at CpG dinucleotides in their recognition sequences (Brown & Bird, 1986; Lindsay & Bird, 1987). For instance, although a NruI site centromeric to the TNF genes was observed to be cleaved in two cases (Dunham et al., 1987; Carroll et al., 1987) it was not seen in another case (Ragoussis et al., 1988).

The conservation of rare-cutting restriction sites between different cell lines in this very polymorphic region of the genome is striking and may be the product of a non-random distribution of nucleotides, C+G rich regions being found in clusters. For instance, 25 kb telomeric of the C2 gene is a cluster of at least 7 rare-cutting sites within 25 kb. Some of these sites (two out of three MluI sites and the NruI site) are not recognised in genomic DNA, possibly due to cytosine methylation, but a NotI and MluI site within 1 kb of each other are both cleaved (see also Chapter 6). Other possible clusters of rare-cutting sites include a region 100 kb telomeric of the TNF genes. Such apparent non-random distribution of infrequently cutting sites in the mammalian genome has been observed previously (Brown & Bird, 1986; Smith et al., 1987b) and may represent so-called HTF Islands (HpaII Tiny Fragments; Bird, 1986). The investigation of the organisation of different HLA haplotypes and the extent of large DNA fragment RFLPs is carried further in Chapters 4 and 5. An analysis of the the CpG clusters within the class III region of the MHC is presented in Chapter 6.

A recent study of the organisation of the class I genes using cosmid cloning and PFGE (Pontarotti et al., 1988) has revealed that the HLA-B and C genes are separated by 130 kb. The MluI fragment

organisation they observed is slightly different to that proposed here because the C locus probe hybridised to a 170 kb MluI fragment (sized in this work at 180 kb) while the B locus probes revealed 130 kb (3') and 40 kb (5') fragments. This discrepancy can be explained if the 180 kb MluI fragment that was observed with the B-specific probe (from the 3' end of the gene) was the product of partial digestion and comigrates with the C locus MluI fragment (Fig 3.10) or if cross-hybridisation of the B probe with the C-containing fragment was occurring. However this does not affect the interpretation of the other data.

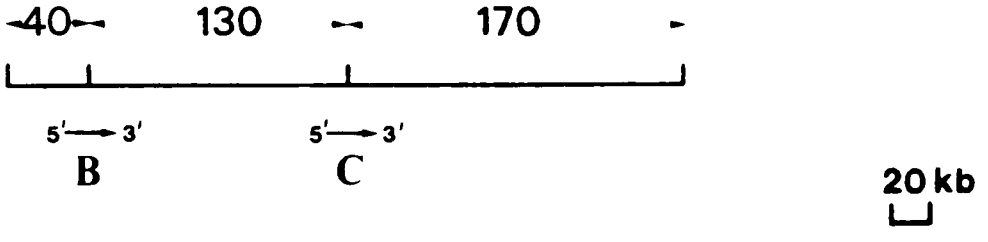


Fig 3.10 MluI fragments in the HLA-E and C region (from Pontarotti et al., 1988). The positions of the B and C genes and their transcriptional organisation are indicated relative to the centromere to the left. MluI sites are indicated by vertical lines, the numbers are distances in kb.

The organisation of the human MHC is very similar to that recently determined in the mouse (Müller et al., 1987b; Fig 3.11). The orientation of the class III genes is the same, and the proposed distances between the class III and class II loci, and between class III and TNF are comparable (Table 3.3) given the accuracy of the PFGE technique. However the distance between the factor B and the C4-like Slp gene is ∿50 kb compared to 30 kb for the equivalent distance in man, and the two C4 analogues in the mouse are separated by ∿80 kb (Chaplin, 1985). The results presented here suggest that the TNF genes are ∿250 kb from the HLA-B region compared to about 70 kb for the analogous gene in the murine MHC. It should also be pointed out that the human and mouse

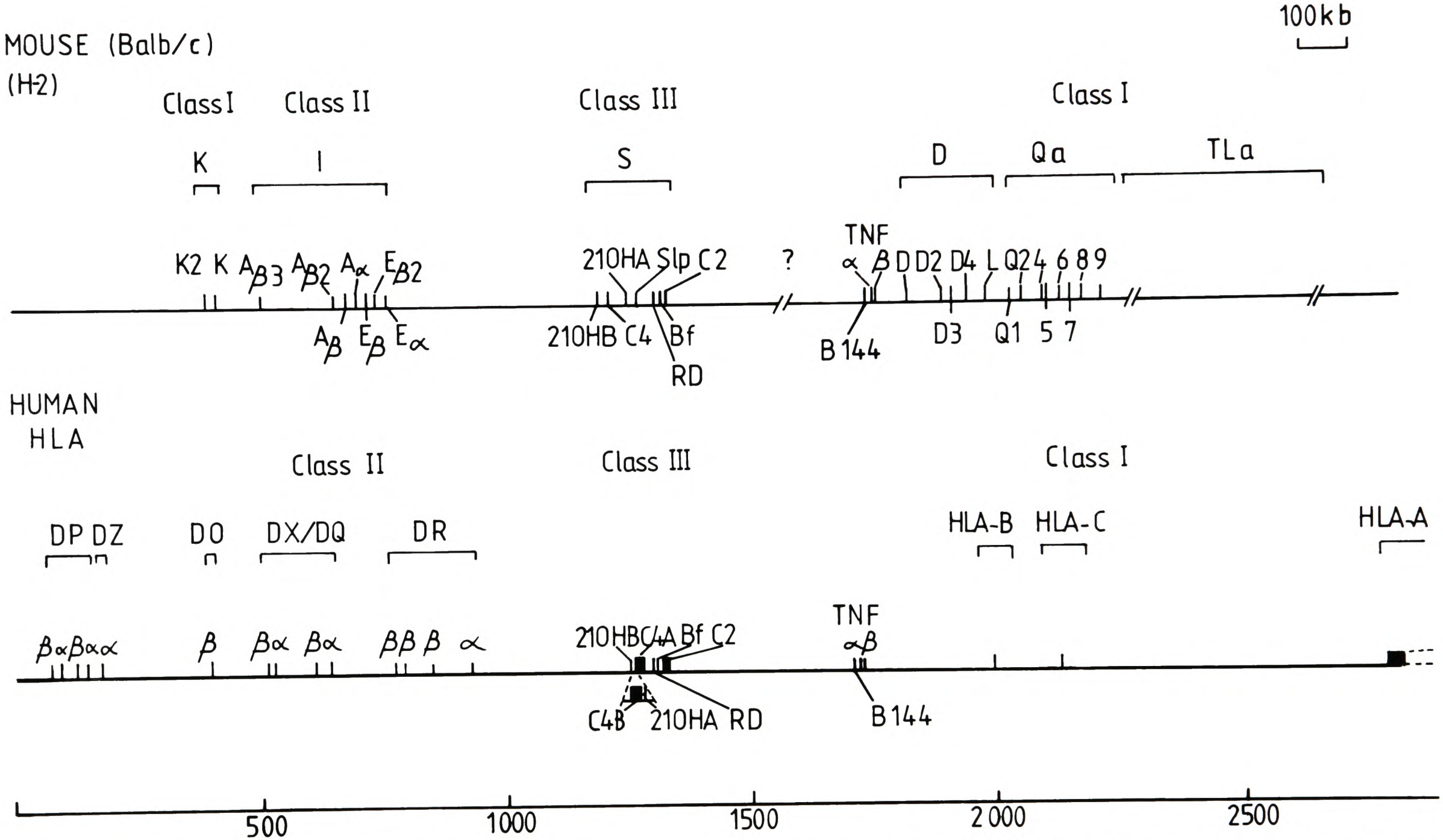


Fig 3.11 Comparison of the maps of the human and mouse MHCs. The centromere is to the right in both maps. The position of 21-OHaseA and C4B genes in normal haplotypes is shown by the inset. The distance between HLA-C and A has been reduced for convenience as in Fig 3.3. Data were taken from Hardy et al. (1986), Müller et al. (1987b) and this work.

MHC maps differ in that the H-2K locus is separated from the other class I genes at the other side of the class II region. It is possible that the difference in distance between the TNF and class I genes in the two species is associated with the genetic event that separated the mouse class I loci. The estimated total size of the human MHC is twice as large as the murine MHC (4000 kb vs 2000 kb), but most of this difference is made up of fragments which hybridise to general class I gene probes and the map of the murine class I region is not complete in the Qa/Tla region lying telomeric to H-2L.

However it does appear that the class II region in man is expanded somewhat relative to the mouse whereas the organisation of the central class III region of the MHC is conserved between the two species. This may be due to differences in the duplication processes that have occurred to expand the class II regions in mouse and man. The map order of the homologous loci is the same in man and mouse (Hardy et al. 1986). However the BALB/c mouse class II region contains only 7 genes for the α and β chains of the class II molecule whereas the human has at least 16 class II-related genes although not all of these are known to be expressed (Trowsdale, 1987; Inoko et al., 1988). Therefore it seems likely that the human class II region is larger than that in the BALB/c mouse because more duplication events have occured. A scheme has been proposed for the comparative evolutions of the class II regions in mouse and man based on the relatedness of deduced amino acid sequence of the class II molecules in the two species (Figueroa & Klein, 1986). This suggests that both regions originated from a common progenitor class II region with a series of duplications leading to the human structure, while several deletions produced the mouse class II organisation. The differences in the class I regions are more tentative because in neither species is the region completely characterised. The estimated number of

class I-like genes in man is possibly as many as 40 (Srivastava et al., 1985), but they have not been located precisely. In the mouse 31 class I genes have been located telomeric of the TNF genes.

Table 3.3 Comparison of the physical distances in the human and murine MHCs.

Region (murine equivalent)	*Distance (kb)*	
	Man	*Mouse*
DRA to 21OHaseB (E to 21OHaseB)	*300-360*	*430*
C2 to TNFA	*390*	*>420*
TNFA to HLA-B (TNFA to H-2D)	*250*	*70*
Class II	*1000*	*<400*
Class I (not H-2K)	*>1000*	*>800*

The conservation of the central class III region of the MHC between the two species may be significant. A considerable amount of data (see Chapter 5 and 6) suggests that the class III region of the human MHC may contain a number of new "housekeeping" genes. It is likely that these genes are also present in the mouse class III region and the similarity in organisation of the complex in this region has been maintained because the structure cannot be disrupted without deleterious effects. In contrast the class I and II regions, and to a lesser extent the C4 locus, have been shown to be more fluid being altered by duplication and deletion both within a species and between species (see Chapters 1 & 5).

The physical linkage map of the human MHC presented here will play an important part in understanding HLA disease associations. The distances between the DRA and 21-OHase genes, and between the C2 and TNFA genes, are sufficiently large to accommodate a number of as yet

unidentified genes. The fact that the size and organisation of this region appears to be conserved between mouse and man also points towards the presence of sequences that are important lying in these gaps. Further mapping of disease-affected haplotypes and the isolation of a complete set of overlapping cosmids covering the DRA to HLA-B region should permit the characterisation of such genes. The demonstration of the precise position of the TNFA and B genes within the human MHC is also of major interest because of the role of TNFα and β as mediators of a number of responses that may play a role in autoimmune disease (see Chapter 7).

CHAPTER 4

DIRECT OBSERVATION OF THE GENE ORGANISATION OF THE COMPLEMENT C4 AND 21-HYDROXYLASE LOCI BY PULSED-FIELD GEL ELECTROPHORESIS

4.1 INTRODUCTION

The class III region of the human MHC between HLA-B and HLA-DR includes genes encoding the complement components C2, factor B, C4A and C4B and the cytochrome P450, steroid 21-hydroxylase (21-OHase) (Lamm & Olaisen, 1985; Dunham et al., 1987; see Chapter 1). The C2 and factor B genes are separated by 421 bp (Wu et al., 1987a) and lie 30 kb from the C4A locus which in turn is ~10 kb from the C4B locus (Carroll et al., 1984a). One of the two copies of the 21-OHase gene, 21-OHaseA and 21-OHaseB, lies 3 kb downstream of each C4 gene (Carroll et al., 1985a; Fig 4.1). In man, the 21-OHaseB gene is active while the 21-OHaseA gene is a highly homologous pseudogene (Higashi et al., 1986; White et al., 1986; Rodrigues et al., 1987). Recently an additional gene, called RD, has been mapped very close to the 3' end of the factor B gene (Levi-Strauss et al., 1988).

Although the two C4 isotypes C4A and C4B are highly homologous and differ by <1% in their derived amino acid sequences, C4 is highly polymorphic (Mauff et al., 1983). The sequences of four C4A and five C4B cDNA and genomic clones have established the pattern of polymorphism in the C4d fragment of the α chain of C4 and have provided a structural basis for the observed functional and serological differences between the isotypes (Belt et al., 1984; 1985; Yu et al., 1986).

There is also heterogeneity in C4 gene size (Schneider et al.,

1986; Yu et al., 1986; Palsdottir et al., 1987a). All C4A genes studied are 22 kb in size (long C4 gene). C4B genes may be either 22 kb or 16 kb (short C4 gene) due to the presence or absence of a 6-7 kb intron about 2.5 kb from the 5' end of the gene.

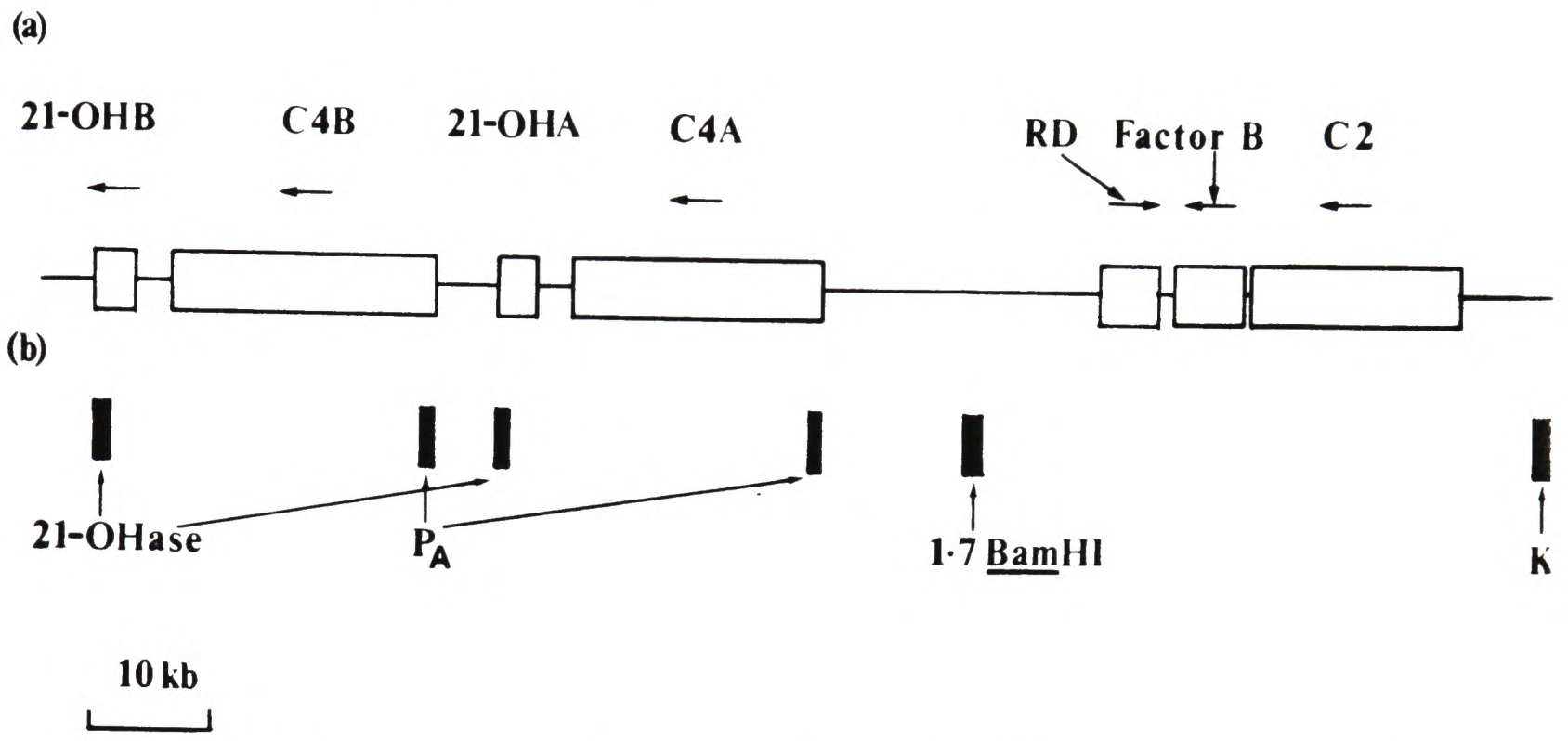


Fig 4.1 Schematic representation of the HLA class III region complement and 21-OHase loci. (a) The organisation of the C2, Factor B, RD, C4 and 21-OHase (21-OH) genes. The horizontal arrows denote the direction of transcription 5'→3'. (b) The positions of the sequences hybridising to the probes used is shown by the solid boxes.

In addition, variation in the number of copies of C4 genes present on individual chromosomes has been observed. Null alleles (designated by QO for quantity zero) are defined by the absence of detectable C4 isotype in the serum. Null alleles occur at a high frequency at either locus and gene frequencies of 5-15% for C4AQO alleles and 10-20% for C4BQO alleles have been estimated (Schendel et al., 1984; Partanen & Koskimies, 1986). C4 null alleles also occur with increased frequency in some HLA-associated diseases such as SLE (Fielder et al., 1983). About half of these null alleles are due to deletion of the C4 gene together with the flanking 21-OHase gene (Carroll et al., 1985b; Schneider et al., 1986). However some null alleles are the result of conversion of the gene at the "nonexpressed" locus to the same isotype as the gene at

the expressed locus (Palsdottir et al., 1987b; Yu & Campbell, 1987). Because of the size of the C4 genes analysis of the extent of the deletions has required cosmid cloning studies. It would be useful to be able to follow the gene organisation at the C4 locus in uncloned DNA in order to rapidly subdivide the null alleles with respect to disease susceptibility. Information about the population genetics of the C4 deletions would possibly lead to further understanding of the evolution of this locus.

A duplication of C4B as C4B1, C4B2 has been inferred at the protein level (Raum et al., 1984). Duplication of C4B genes on the same extended haplotype B14 DR1 has been confirmed by cosmid cloning (Carroll et al., 1984b). The frequency of these duplications has been estimated from phenotyping as 1-2% but Schneider et al. (1986) have obtained evidence that this number may be increased by the existence of more homoduplicated haplotypes than were previously identified.

A number of RFLPs are available to follow the differences at the human C4 locus in the population (Yu & Campbell, 1987; see Table 4.1). However these methods suffer from two main drawbacks. First, it is necessary to compare intensities of the bands revealed by autoradiography to deduce the gene copy number and comparison between different samples is difficult. Second, family studies are necessary to imply which genes lie on the same chromosome.

Having constructed a physical linkage map of the human MHC (Chapter 3, Dunham et al., 1987), it was of interest to define whether the documented differences at the C4 and 21-OHase loci could be observed between uncloned DNAs using PFGE. Müller et al. (1987b) have previously demonstrated that in mouse strains containing at least three Slp genes, the size of a PvuI fragment observed by hybridisation with S region probes was increased by over 150 kb compared to the BALB/c mouse which

	Restriction Enzyme	Probe	Band Size	Remarks	References
1) Long/short C4 genes (i.e. presence/absence of a 6-7 kb intron)	BamHI	5' cDNA (P_A)	4.8 kb 3.3 kb	Long gene+ Short gene*	Carroll et al. (1985a;c) Yu et al. (1986)
	KpnI	C4d genomic (or pALU-7)	7.5 kb 8.5 kb	Long gene Short gene	
2) Nature of Locus and probably presence of the 6-7 kb intron	KpnI	5' cDNA (P_A)	12 kb 3.5 kb	Locus I, or I/II recombinant Locus II	Carroll et al. (1985c)
	TaqI	5' cDNA (P_A)	7.0 kb 6.0 kb 5.4 kb 6.4 kb	Locus I Locus II, long Locus II, short Locus II, short of SC01 complotype (possibly I/II recombinant)	Schneider et al. (1986) Yu & Campbell (1987)
3) C4A or C4B gene	N1aIV	C4d genomic	(276+191)bp 467 bp	C4A gene C4B gene	Yu et al. (1986) Yu & Campbell (1987)
4) Rg1/Ch1 determinants	Eco0109	C4d genomic	565 bp 458 bp	expresses Rg1 expresses Ch1	Yu et al. (1986) Yu & Campbell (1987)

*Table 4.1 A summary of diagnostic RFLPs for human C4 genes. + indicates a long gene, 22 kb in size, with a 6-7 kb intron. * indicates a short gene, 16 kb in size, without a 6-7 kb intron. Rg1 and Ch1 refer to the major Rodgers and Chido blood group antigenic determinants. Taken from Yu & Campbell (1987).*

possesses a single *S1p* gene. It was also possible that other changes in the DNA organisation between haplotypes exist around the complement gene loci in the HLA. In view of the possibility that other genes might reside close to the already characterised loci, as demonstrated by the identification of the RD locus between C4 and factor B (Levi-Strauss *et al.*, 1988; see Chapter 1), the existence of deletions within this region of the MHC might be relevant to the basis of certain HLA-linked diseases. Also such studies would give some idea as to how fluid this region of the genome is within the population. Therefore, PFGE studies have been extended to compare the DNA organisation of different HLA haplotypes using PFGE and homozygous-typing lymphoblastoid cell lines around the complement gene loci.

4.2 EXPERIMENTAL STRATEGY

In order to be able to compare accurately the sizes of large DNA fragments, it was necessary to have a PFGE gel system that had straight lanes and which enabled a large number of samples to be electrophoresed on the same gel. The OFAGE system that has been described in Chapter 3 did not fit these criteria because of the strong curvature of the outer tracks which prevented use of all but the centremost lanes. A number of PFGE designs have been described which give straight lanes (Carle *et al.*, 1986; Chu *et al.*, 1986; Gardiner *et al.*, 1986; Southern *et al.*, 1987; Bancroft & Wolk, 1988; Clarke *et al.*, 1988). The system used here is that of Southern *et al.* (1987) and has been described in Chapter 2. Figs 4.2 and 4.4a illustrate typical results using the "Waltzer" PFGE box showing that both a large number of samples can be loaded and the lanes are straight. Fig 4.4c shows a plot of mobility against mol. wt. for the λ concatemers separated on the PFGE gel in Fig 4.4a. The plot

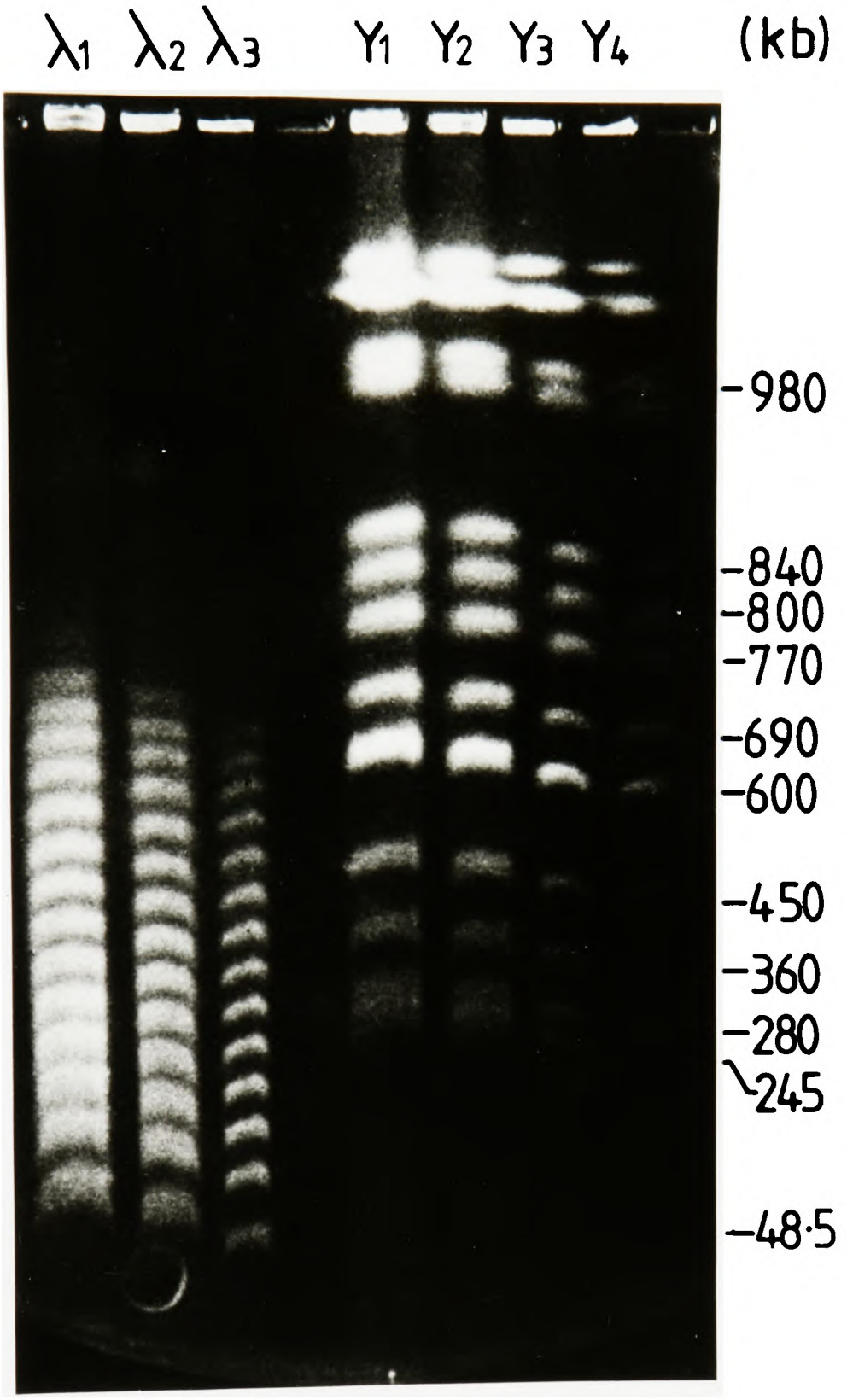


Fig 4.2 Separation of yeast chromosomes and λ cI857 S7 concatemers on the "Waltzer" PFGE system. 12 μg (λ_1), 6 μg (λ_2) and 3 μg (λ_3) of λ concatemer DNA and 1, 1/2, 1/4 and 1/8 (Y_1, Y_2, Y_3, Y_4, respectively) of a yeast chromosome block were electrophoresed as described in Chapter 3 at a switching interval of 55 s. The figure shows that straight DNA trajectories are obtained and that overloading of the gel can lead to variations in fragment size estimates, in this case of 50 kb or greater (compare λ_1 and λ_3). Fragment sizes are indicated for the lowest DNA concentrations.

illustrates the characteristic pattern of separation obtained with PFGE gels (Southern et al., 1987; Vollrath & Davis, 1987; Birren et al., 1988). There is a region of good resolution where mobility is linearly related to mol. wt. which in this case is up to the ninth λ concatemer. Increasing the length of the switching interval extends this region of linear separation to higher mol. wt. but decreases the resolution. There is then a short region of maximum resolution where separation of the markers is increased (for this gel between the ninth and thirteenth step of the ladder). Finally at high mol. wt. there is a region where little or no separation of DNA is obtained. The mol. wt. at which resolution is lost again is dependent upon the switching interval. In order to obtain accurate estimates of the sizes of various DNA fragments samples were separated at a switching interval so as to optimally resolve the DNA fragment size of interest.

To observe differences between the DNA organisation in several common HLA haplotypes, high mol. wt. DNA in agarose blocks was prepared from both the HLA homozygous lymphoblastoid cell line, Ice 5, (see Chapter 3, known as cell line 1 here) and from a series of homozygous-typing lymphoblastoid cell lines, provided by Roger Dawkins, whose HLA types are detailed in Table 4.2 (cell lines 2-8). The DNA was digested with restriction enzymes that cut infrequently in the mammalian genome and were known to give DNA fragments of an informative size around the complement gene cluster in the HLA class III region (Dunham et al., 1987; Chapter 3). DNA from each of the cell lines digested with a single restriction enzyme was then separated by PFGE at an appropriate switching interval. The PFGE gels were depurinated, transferred to nylon membranes and then hybridised with genomic and cDNA probes specific for the complement gene cluster (see Chapter 2). The restriction enzyme fragment sizes with the different haplotype cell lines were observed.

It is important to note that the amount of DNA loaded on a PFGE gel affects the migration of the DNA relative both to other genomic DNA samples and to the DNA size markers (Michiels et al., 1987; Fig 4.2) and that DNA fragment sizing can be inconsistent between different gel runs and switching intervals (Bernards et al., 1986). These problems can be overcome by loading equivalent amounts of DNA for each sample and by comparing the sizes of restriction fragments for different cell lines on the same gel.

Since a large region of the HLA class III region has now been cloned in cosmids (Dunham et al., 1988; Sargent 1988), there are available a number of internal size markers for the prototype cell line 1 (Ice 5). The exact size of certain genomic DNA fragments produced with specific enzyme and probe combinations is known by comparison with the cloned cosmid DNA. For instance the MluI fragment that hybridised with the 21-OHase probe is estimated to be 200 kb. The size expected size based on the cosmid map is 205 kb. It is also interesting to note that the restriction fragment sizes observed with the "Waltzer" PFGE apparatus were generally smaller than with the OFAGE system (Chapter 3, Dunham et al., 1987) and were found to match more accurately with the sizes predicted from the cloned DNA.

By utilisation of large DNA fragment RFLPs it was possible to observe directly the variations in size of the C4 loci in different haplotypes. Specifically, the size of the diagnostic BssHII or SacII restriction fragment observed with a C4 or 21-OHase specific DNA probe indicates both the number of C4 genes present on a chromosome and their size (C4 long or short). This technique can be applied to peripheral blood lymphocyte DNA isolated from individuals and together with the previously established RFLPs (Schneider et al., 1986; Yu & Campbell, 1987) allows a complete definition of the C4 gene organisation of an

individual.

In addition the limit of the disruption of the DNA in the deleted haplotypes can be defined to lie within the SacII fragment that contains the C4 and 21-OHase genes. No other differences were observed in the immediate vicinity of the HLA-linked complement genes between the cell lines studied (see Chapter 5 for a more long-range analysis).

4.3 RESULTS

4.3.1 Probes

The positions of the probes used in the analysis relative to the complement gene loci are shown in Fig 4.1. The human C4 5' cDNA probe, P_A, is a 476 bp BamHI-KpnI restriction fragment from the full length C4 cDNA clone, pAT-A (Belt et al., 1984). The human 21-OHase probe is a 1.2 kb ClaI-PvuII genomic DNA restriction fragment isolated from a cloned 21-OHaseB gene (Rodrigues et al., 1987). The other probes were a 1.6 kb HindIII/BamHI genomic DNA fragment located about 6 kb 5' of the C2 gene known as probe K and a 1.7 kb BamHI genomic DNA fragment lying ~10 kb 5' of the C4 gene which were both derived from cloned cosmid inserts as described in section 2.1.6.

4.3.2 Southern blot analysis of the cell lines for the TaqI polymorphism

Hybridisation of the P_A and 21-OHase probes to Southern blots of genomic DNA digested with TaqI allows characterisation of the C4 and 21-OHase loci (Schneider et al., 1986; Yu et al., 1986; White et al., 1984; Table 4.1). The basis of this diagnostic RFLP is illustrated in Fig 4.3a. Therefore high mol. wt. DNA in agarose blocks from each cell

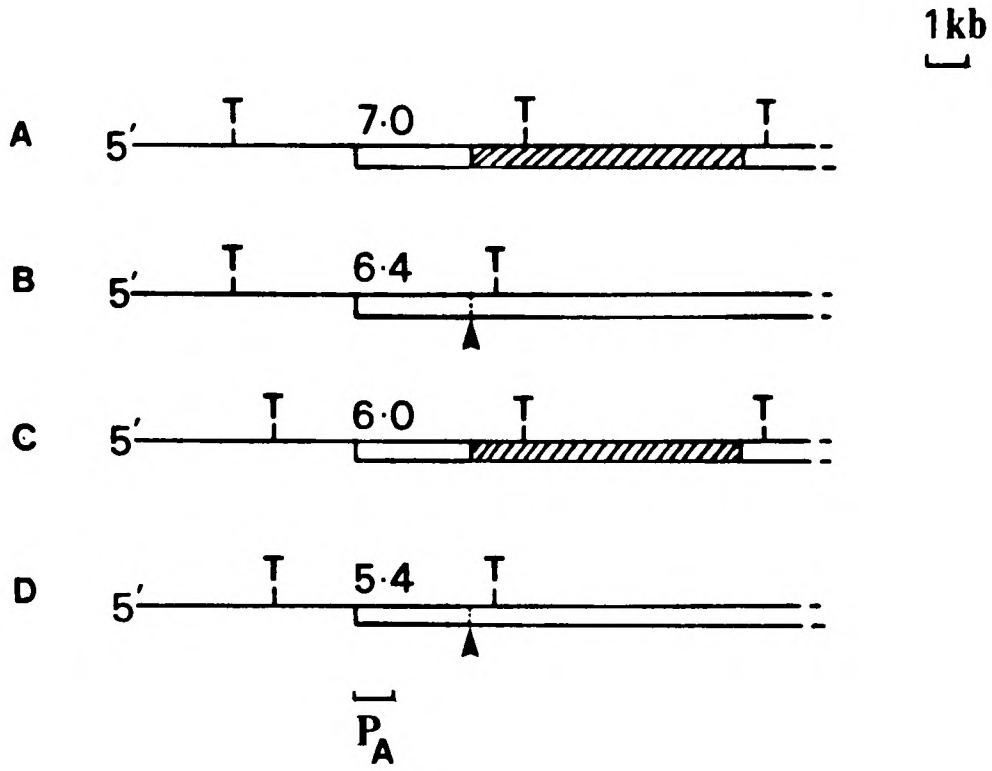
line was digested with TaqI and the digested DNA separated on a conventional 0.7% agarose gel as described in section 2.7.2 and Fig 4.3. A Southern blot of the separated TaqI digested DNA was hybridised with probe P_A and the 21-OHase probe (Fig 4.3b).

Using TaqI cell lines 1 (Ice 5) and 4 had a single 7.0 kb fragment with P_A and a 3.7 kb fragment with 21-OHase. Thus, these cell lines have a single long C4 gene (which is C4A from the complotyping) and a single 21-OHaseB gene on both chromosomes. The TaqI results for cell lines 2 and 3 showed a 6.4 kb fragment with P_A and a 3.7 kb band with 21-OHase indicating that they possess a single short C4 gene (the recombinant gene characteristic of the SC01 complotype) at locus I/II and a 21-OHaseB gene. However, cell lines 5 and 8 showed two TaqI fragments of 7 kb and 5.4 kb with P_A and 3.2 kb and 3.7 kb fragments with the 21-OHase probe. This suggests that cell lines 5 and 8 have one long C4 gene, one short C4 gene and a copy of both 21-OHaseA and 21-OHaseB genes. Cell lines 6 and 7 had TaqI bands characteristic of two long C4 genes and a copy of each 21-OHase gene.

4.3.3 Large DNA fragment RFLPs associated with the class III complement gene loci

The results of hybridisation of MluI-digested DNA from each of the homozygous typing cell lines with the 21-OHase probe are shown in Fig 4.4 and Table 4.2. In cell lines 1-4 a single 200 kb MluI fragment was seen, in cell lines 5 and 8 a 225 kb MluI fragment hybridised with the probe, while cell lines 6 and 7 gave a 230 kb MluI fragment. Thus there appears to be a 25-30 kb difference in the size of the MluI fragment that encompasses the complement genes between cell lines 1-4 and 5-8. Analysis of the cell line DNAs for the TaqI polymorphism with P_A and the

a.



b.

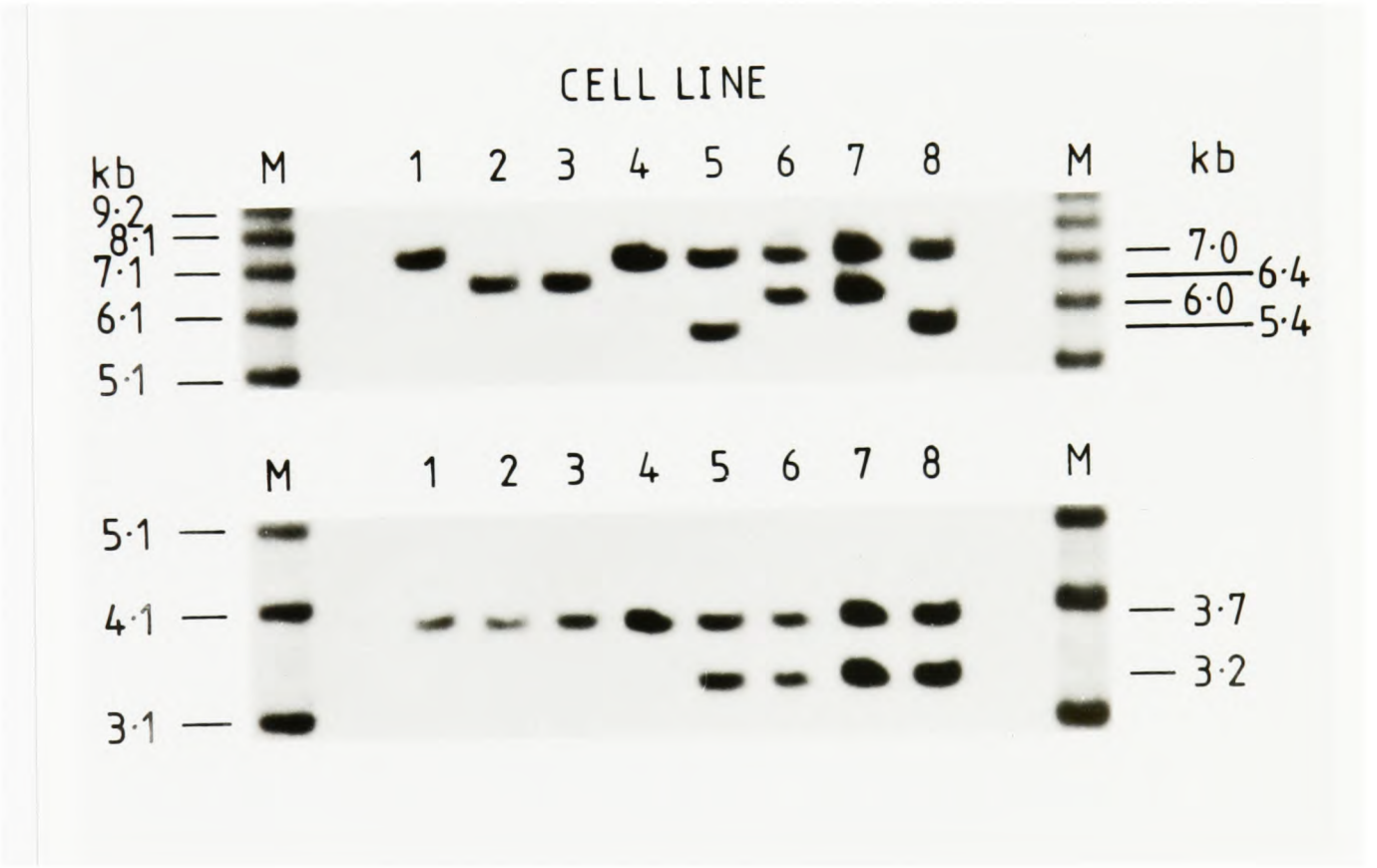


Fig 4.3 TaqI polymorphism analysis of cell line DNA. a. The basis of the TaqI polymorphism. TaqI (T) restriction maps of the 5' ends of C4 genes are shown as is the position of probe P_A. The 6-7 kb intron containing a TaqI site is shown cross-hatched. A is a long gene at locus I. B is a short gene, a locus I/II recombinant. C is a long gene at locus II. D is a short gene at locus II. The sizes of the TaqI fragments identified in each case by P_A is shown above the line in kb. b. Southern blot analysis of TaqI digests of genomic DNA from cell lines 1-8. The top panel shows the results of hybridisation of P_A and the bottom panel shows the results of hybridisation of a 21-OHase probe to the same Southern blot of TaqI digests separated by electrophoresis on a 0.7% agarose gel. The sizes of the TaqI fragments observed are shown on the right. Markers (M) were the BRL 1 kb DNA ladder.

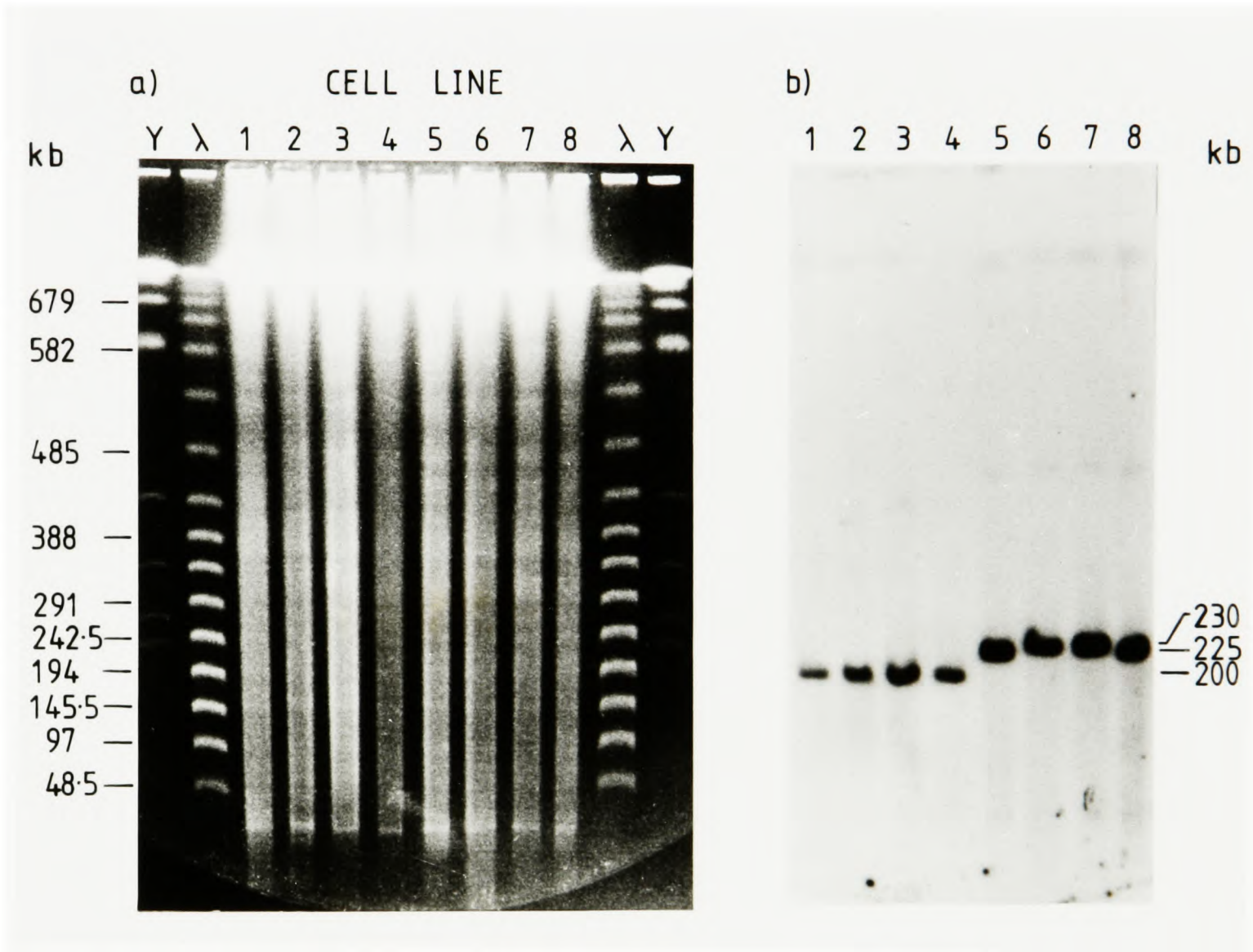


Fig 4.4 MluI restriction fragment patterns of cell lines 1-8. Panel (a) is the ethidium bromide stained gel showing separation of MluI digested cell line DNA by PFGE at a 30 s switching interval for 30 hr. Comigrating markers are yeast chromosomes (Y) and λ concatemers (λ). Panel (b) shows the result after hybridisation of the Southern blot of this gel with the 21-OHase probe. Fragment sizes are indicated on the outside of each panel.

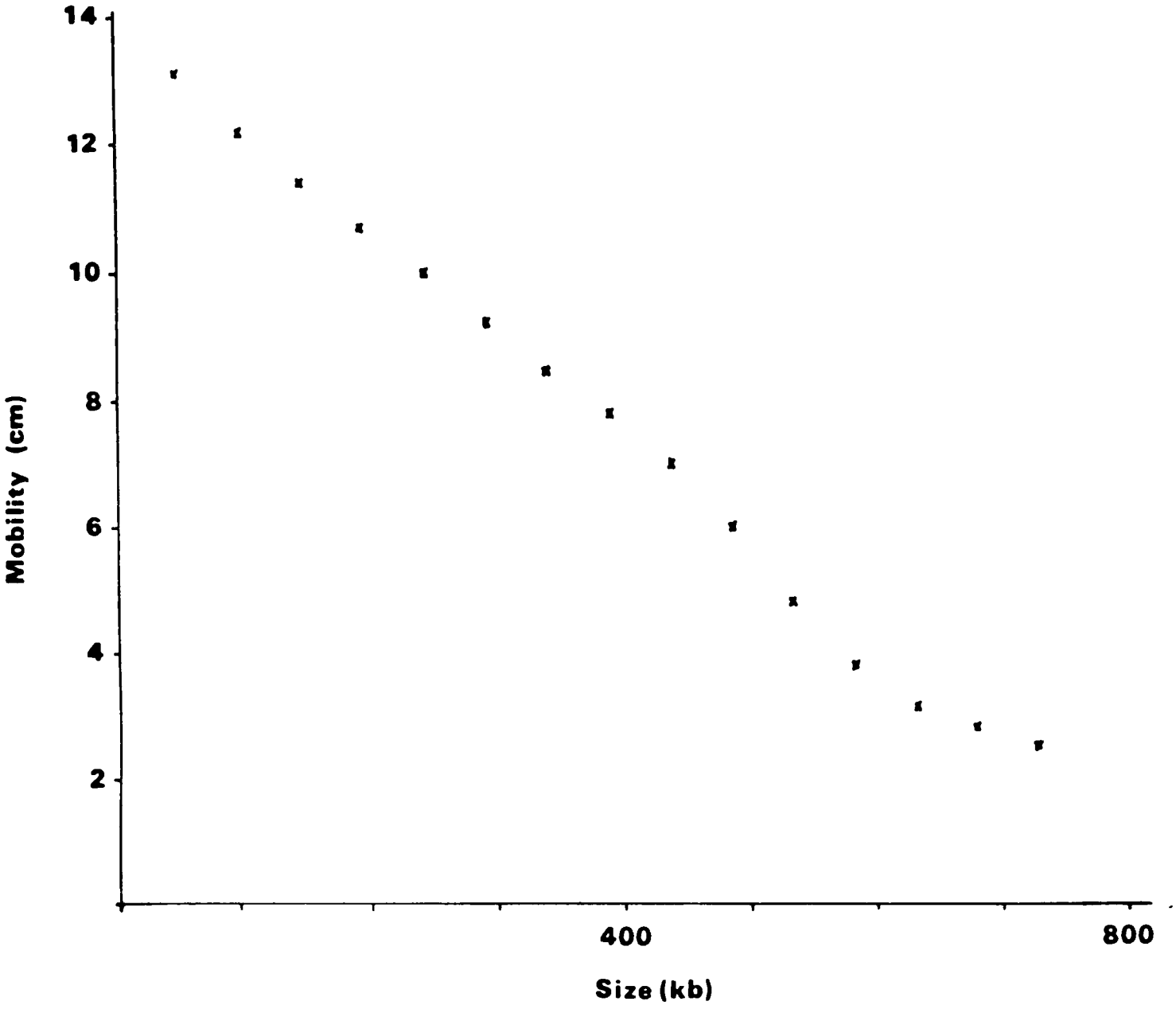


Fig 4.4c Plot of mobility against mol. wt. for the λ concatemers separated on the PFGE gel shown in Fig 4.4a. Crosses mark the points for each concatemer.

			RESTRICTION FRAGMENT SIZE (kb)									
		Enzyme	MluI	BssHII			SacII			TaqI		
Cell line	HLA Type	Probe	21OH/ P_A	21OH/ P_A	1.7 BamHI	K	21OH/ P_A	1.7 BamHI	K	P_A	21OH	C4 Genotype
1	B7 C4A3 C4BQ0 DR2		200	80	12	55	40	12	55 25	7.0	3.7	long locus I
2	B8 C4AQ0 C4B1 DR3		200	70	12	55	33	12	55 25	6.4	3.7	short locus I/II recombinant
3	B8 C4AQ0 C4B1 DR3		200	70	12	55	33	12	55 25	6.4	3.7	short locus I/II recombinant
4	B18 C4A3 C4BQ0 DR3		200	80	12	55	40	12	55 25	7.0	3.7	long locus I
5	B18 C4A4 C4B2 DR2		225	105	12	55	65	12	55 25	7.0 5.4	3.7 3.2	long locus I + short locus II
6	B44 C4A3 C4BQ0 DR4		230	115	12	55	70	12	55 25	7.0 6.0	3.7 3.2	long locus I + long locus II
7	B35 C4A3 C4B1 DR5		230	115	12	55	70	12	55 25	7.0 6.0	3.7 3.2	long locus I + long locus II
8	B55 C4A4 C4B5 DR6		225	105	12	55	65	12	55 25	7.0 5.4	3.7 3.2	long locus I + short locus II

Table 4.2 Restriction fragment sizes observed with cell line DNA (kb). C4 genotypes were established as described in the text. 21OH refers to the 21-OHase probe, 1.7 BamHI refers to the 1.7 kb BamHI probe.

21-OHase probe suggested the basis for the differences in size observed using MluI (Fig 4.3 and Table 4.2). Since the size of a C4-21-OHase gene unit is ∿23 kb when the C4 gene is short and ∿30 kb when there is a long C4 gene, the observed difference in size of the complement genes MluI DNA restriction fragment appears to be due to the possession of two C4 genes in cell lines 5-8, one of which is deleted in cell lines 1-4.

Indeed, the MluI fragment for cell lines 6 and 7 was ∿5 kb larger than for cell lines 5 and 8 which corresponds to the presence of the 6-7 kb intron (Schneider et al., 1986) at the second C4 locus in cell lines 6 and 7 but not in cell lines 5 and 8, allowing for the limits of the size resolution of this PFGE gel. However, the equivalent difference due to presence or absence of this intron could not be detected between cell lines 1 and 4 and cell lines 2 and 3. Hybridisation of other probes to the same Southern blot of MluI-digested genomic DNA (see Chapter 5 Fig 5.3) suggested that the mobility of the DNA fragments for cell lines 2 and 3 was slightly retarded in the PFGE gel and hence the size difference due to the absence of the 6-7 kb intron has been obscured.

In order to confirm these observations, high mol. wt. DNA from each of the cell lines was digested with BssHII (Fig 4.5 and Table 4.2) and SacII (Table 4.2). The digested DNA was separated by PFGE at a switching interval of 7.5 s to obtain optimal resolution in the 10-200 kb size range. Fig 4.5 shows the results of hybridisation of probes from the complement gene cluster of a Southern blot of BssHII-digested DNA. Hybridisation of the 1.6 kb HindIII/BamHI fragment (K) telomeric to C2 with the BssHII digested cell line DNA revealed a common 55 kb fragment for all the cell lines. The size of this fragment is different to that which had been observed previously (Dunham et al., 1987; Chapter 3) where only a larger 150 kb fragment was observed most likely due to partial digestion.

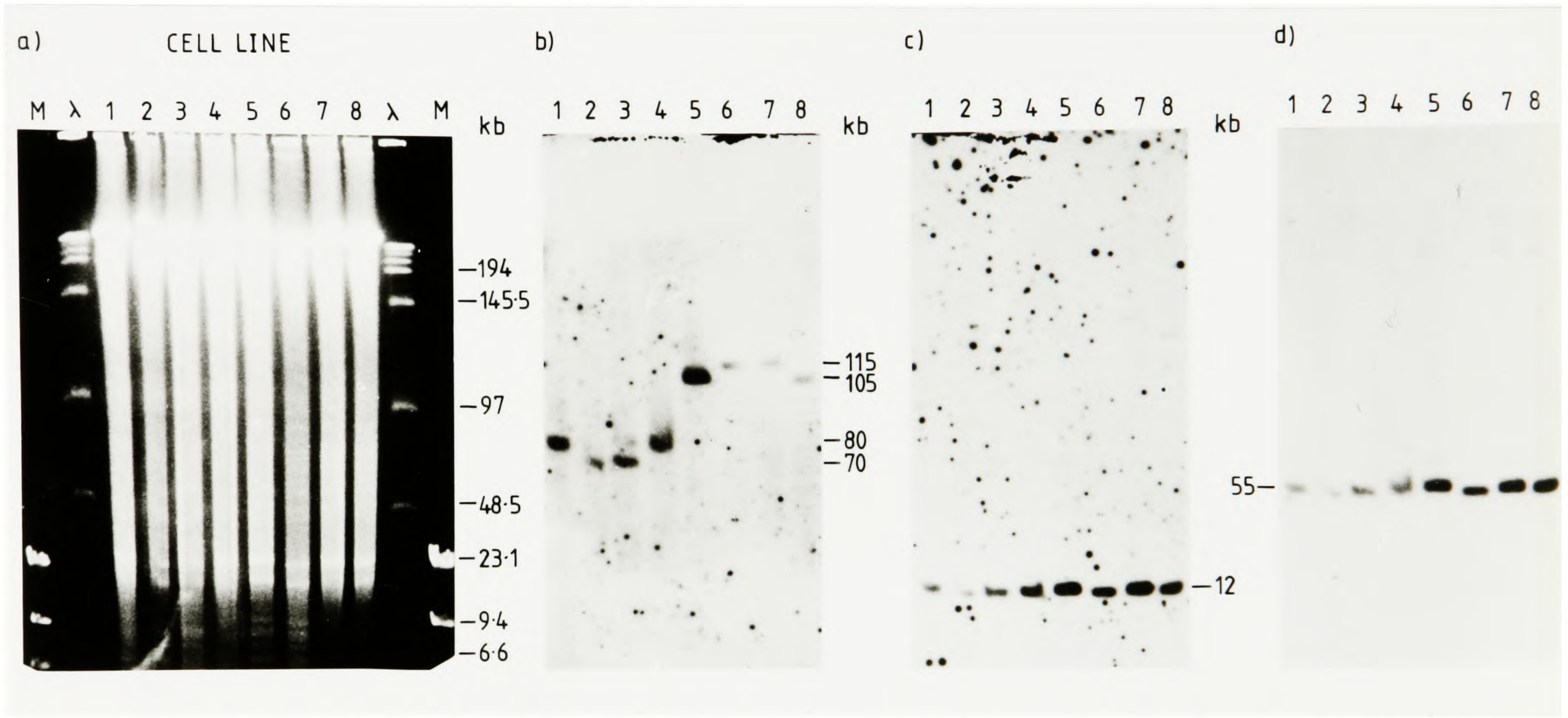


Fig 4.5 Southern blot analysis of BssHII cell line genomic DNA separated by PFGE. a) Ethidium bromide stained PFGE gel of BssHII digested cell line DNA after separation at a switching interval of 7.5 s. DNA samples were digested with 10 uits of BssHII for 3 h at 50°C. The markers are λ DNA concatemers (λ) and λ DNA digested with HindIII (M). Panels b) to d) show the results of successive hybridisation of 21-OHase (b), 1.7 kb BamHI (c) and K (d) probes with a Southern blot of the gel shown in a). Numbers above tracks refer to the cell lines (see Table 4.2). Fragment sizes are shown in kb.

Hybridisation of the same blot with a 1.7 kb BamHI fragment lying 10 kb upstream of the C4A gene identified a 12 kb BssHII fragment also in all the cell lines. Hence there appears to be no difference between these cell lines in ∿67 kb of DNA containing the C2 and factor B genes, within the larger MluI fragment. Hybridisation of these probes to SacII digested DNA confirmed this result (Table 4.2), probe K hybridising with 55 kb and 25 kb SacII fragments, while the 1.7 kb BamHI probe gave a 12 kb SacII band. The 25 kb SacII fragment observed with probe K is the result of partial digestion of the 55 kb fragment at a SacII site 5' of the C2 gene (see Fig 4.6).

However, when either the 21-OHase or P_A probes were hybridised to the Southern blot of BssHII digested DNA, the sizes of the fragments observed were different between cell lines depending on the organisation of the C4 loci. Cell lines 2 and 3, which possess a single short C4B gene and a 21-OHaseB gene, gave a 70 kb BssHII fragment with the 21-OHase probe. Cell line 3 also showed some hybridisation to a larger BssHII fragment of 82 kb which also hybridised with the 1.7 kb BamHI probe and must therefore be the product of partial digestion. For cell lines 1 and 4, which instead have a long C4A gene, the 21-OHase probe hybridised to an 80 kb BssHII fragment. Thus using PFGE at this resolution the difference in size due to the 6-7 kb intron can be readily detected. Similarly in the cell lines with one long and one short C4 gene (5 and 8) the BssHII fragment was ∿10 kb shorter than in those with two long C4 genes (6 and 7). In addition the difference between the BssHII fragment sizes for haplotypes with one C4 gene or two C4 genes corresponded to the size of the C4-21-OHase unit that is deleted or present. The size of the BssHII fragment that hybridises to P_A or 21-OHase is directly related to the amount of DNA that is present at the C4 and 21-OHase locus.

The results using BssHII were completely reflected using SacII (Table 4.2). With P_A or the 21-OHase probe a smaller SacII fragment (70 kb for two long C4 genes) was observed than with BssHII, but the size differences between the haplotypes were again fully accounted for by the number and type of C4 gene present. Therefore the extent of the differences between these cell lines due to deletion of C4 and 21-OHase genes is completely defined to be within the limit of the SacII fragment that hybridises to P_A or 21-OHase. It is also apparent that the size differences between these haplotypes that were observed in the complement gene MluI fragment can be completely accounted for by the difference in the size of the BssHII fragment that contains the C4 genes. In these haplotypes at least it seems that there are no other changes in the amount of DNA present in the region surrounding the C4 genes as defined by the MluI fragment.

4.3.4 Physical maps of the complement gene loci in the cell lines

The region of genomic DNA contained in the 200 kb MluI fragment that hybridised with the 21-OHase probe has been cloned in a series of overlapping cosmid clones from a cosmid library prepared from DNA of cell line 1 (Dunham et al., 1987; Sargent, 1988). These cosmid clones were mapped for rare cutting restriction endonucleases and the data are presented in Fig 4.6a. The restriction enzyme sites that are cleaved in the genomic DNA of cell line 1 were identified using a series of probes derived from the cosmid clone inserts (see Chapter 6) and these data are also shown in Fig 4.6. It is obvious that some sites that are present in the cell line DNA are not cleaved by the restriction enzymes in genomic DNA. This is presumably because the sites are methylated in the genomic DNA as both BssHII and SacII are sensitive to methylation at CpG

Fig 4.6 Molecular maps of the complement gene cluster in eight HLA homozygous cell lines. (a) Restriction map of cloned cosmid DNA containing the complement gene cluster. Genes are marked by open boxes. The positions of probes is indicated by vertical lines at the top. From left to right the probes are 21-OHase, P_A, 1.7 kb BamHI, K. Vertical lines above the line are BamHI sites, below the line rare cutter sites as follows: BssHII, B; SacII, S; MluI, M; NruI, Nr; PvuI, P. (b) Sites for infrequently cutting enzymes that are cleaved in genomic DNA of the cell lines. Vertical lines mark the positions of sites which are designated as in (a). Dotted sites are cleaved partially.

(a)

10kb

21-OHB C4A RD Bf C2

BS S B BBS S BS P B S Nr M MBS

(b)

CELL LINE

1, 4

21-OHB C4A RD Bf C2

M B S B BS BS S MBS

2, 3

21-OHB C4B RD Bf C2

M B S B BS BS S MBS

5, 8

21-OHB C4B 21-OHA C4A RD Bf C2

M B S B B BS BS S MBS

6, 7

21-OHB C4B 21-OHA C4A RD Bf C2

M B S B B BS BS S MBS

dinucleotides (Lindsay and Bird, 1987). Taking this information together with the MluI, BssHII, SacII and TaqI data present above it is possible to construct genomic restriction maps for the cell lines analysed (Fig 4.6b) depicting the differences in size and number of C4 and 21-OHase genes present. It is not possible to say whether sites that are not restricted in cell line 1 genomic DNA are present in the other cell lines.

4.3.5 Analysis of the large fragment RFLPs in DNA from individuals

The results presented above suggested that it might be possible to apply the BssHII and SacII RFLPs to the analysis of the C4 loci of different individuals. To this end, high mol. wt. DNA from 5 healthy C4- and, in 3 cases, HLA-typed individuals was prepared in agarose blocks from PBMNs. One possible pitfall of this approach might be that the pattern of methylation and restriction enzyme sites in PBMNs may be different from that observed for the EBV-transformed lymphoblastoid cell lines. The DNAs were digested with TaqI, BssHII and SacII and separated on conventional 0.7% agarose gels or PFGE gels as appropriate. Southern blots of the separated, digested DNAs were hybridised with P_A and 21-OHase probes. The results are shown in Fig 4.7 (TaqI) and Fig 4.8 (BssHII and SacII) and are summarised in Table 4.3.

Individual A had a single BssHII fragment of 115 kb identified by hybridisation with the 21-OHase probe migrating in the same position as the BssHII fragment for cell line 7 which was also electrophoresed on this gel as a control. This band is characteristic of a chromosome possessing two long C4 genes and the information from the TaqI digests confirms the presence of long C4 genes at locus I and locus II (Fig 4.7) and one copy of each of the 21-OHaseA and B genes. Similarly this

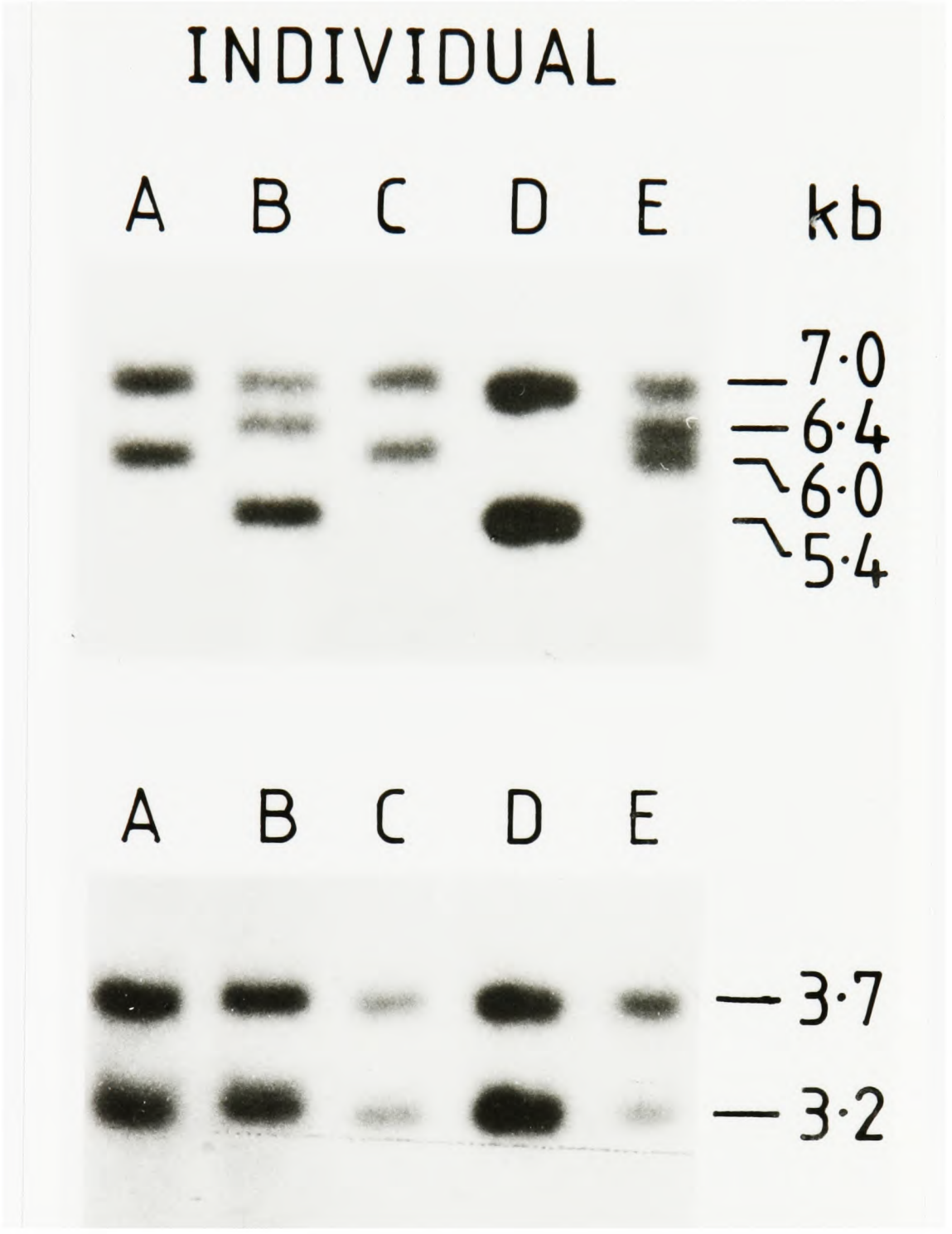


Fig 4.7 TaqI polymorphism analysis of individuals. The top panel shows the results of hybridisation of P_A to a Southern blot of TaqI digests of genomic DNA from individuals A-E separated on a 0.7% agarose submarine gel. The bottom panel show the results of hybridisation of a 21-OHase probe to the same Southern blot. The sizes of restriction fragments observed are indicated on the right.

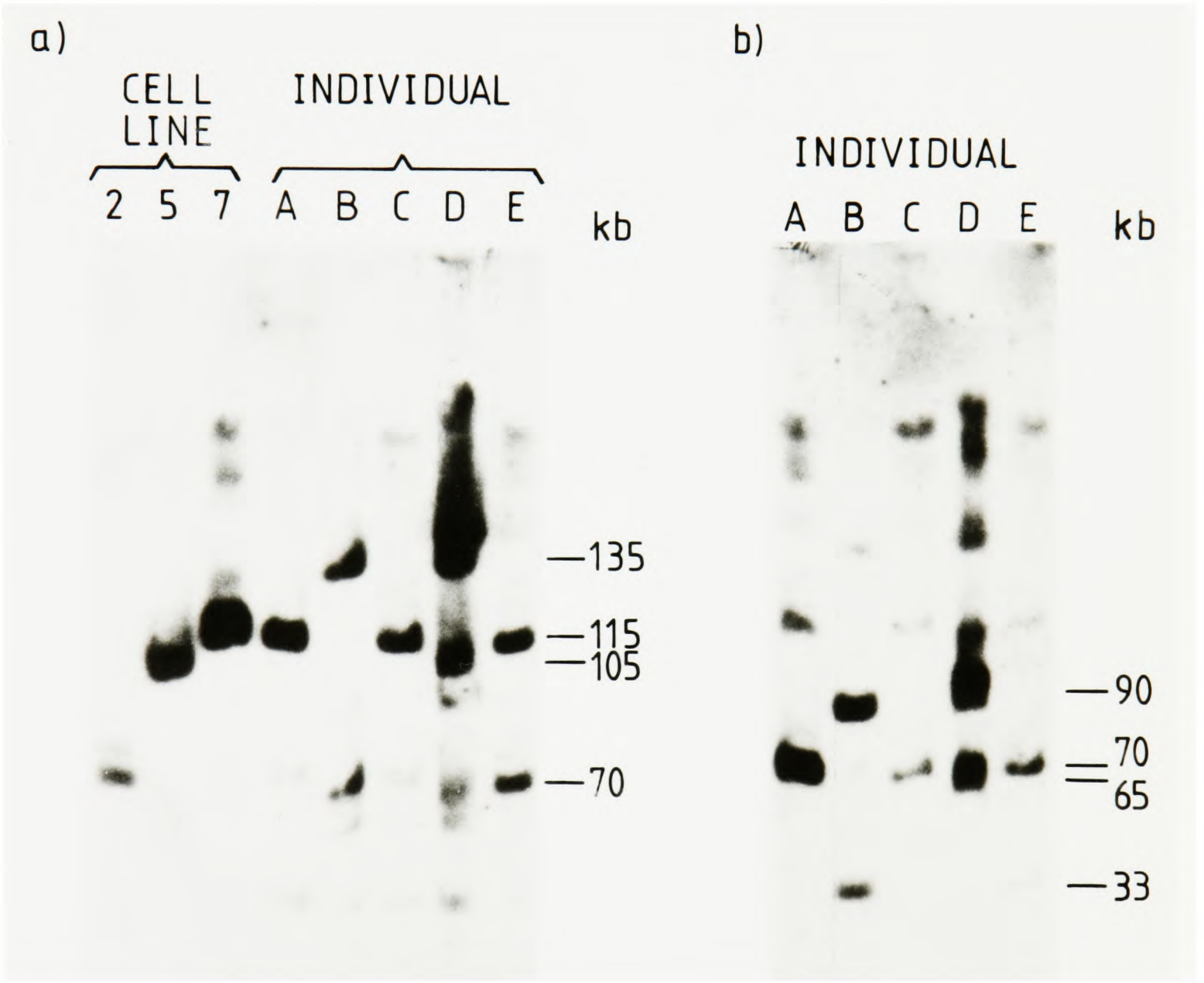


Fig 4.8 Genomic Southern blot analysis of BssHII and SacII digests of genomic DNA isolated from PBMNs of individuals A-E. DNA samples were digested with 15 units of BssHII and 40 units of SacII for 3 h at the appropriate temperature and the digests were separated by PFGE at a 7.5 s switching interval and Southern blotted. The results of hybridisation of the 21-OHase probe with a Southern blot of BssHII, a) and SacII, b) digests are shown. BssHII digests of cell lines 2, 5 and 7 were also separated on this PFGE gel. Fragment sizes are indicated in kb on the right of each panel. Unmarked fragments are the result of partial digestion.

Individual	HLA/complement type	Enzyme	BssHII	SacII	TaqI		C4 Genotype
		Probe	P_A/21OHase	P_A/21OHase	P_A	21OHase	
A	B7 C4A3 C4B1 DR5		115	70	7.0	3.7	2 x long locus I & II
	B44 C4A3 C4BQ0 DR6				6.0	3.2	
B	B8 C4AQ0 C4B1 DR3		135	90	7.0	3.7	1 x Short locus I/II recombinant
	B14 C4A2 C4B1,2 DR13		70	33	6.4	3.2	
					2 x 5.4		1 x long locus I + short locus II & III
C	B45 C4A3 C4BQ0 DR11		115	70	7.0	3.7	2 x long locus I + II
	B18 C4A3 C4B1 DR4				6.0	3.2	
D	C4A3 C4A4		135	90	2 x 7.0	2 x 3.7	1 x long locus I + short locus II + III
	C4B1 C4B2		105	65	3 x 5.4	3 x 3.2	
							1 x long locus I + short locus II
E	C4A3 C4AQ0		115	70	7.0	2 x 3.7	1 x long locus I + II
	C4B1 C4B1		70	33	6.4	3.2	1 x short locus I/II recombinant
					6.0		

RESTRICTION FRAGMENT SIZE (kb)

Table 4.3 Restriction enzyme fragment sizes observed with probes P_A and 21-OHase using DNA from 5 individuals. The C4 genotype has been deduced from the characteristic RFLPs (see text).

individual gave a SacII fragment of 70 kb which confirms the gene organisation. Therefore individual A has two long C4 genes on each copy of chromosome 6.

Individual B possessed a 70 kb BssHII fragment characteristic of a short C4 gene. This corresponds in size to the 70 kb band of cell line 2 which was also electrophoresed on this gel as a standard. In addition this individual has a novel 135 kb BssHII fragment. Examination of the TaqI restriction enzyme data for this individual showed the presence of a 6.4 kb TaqI fragment which is characteristic of a short C4 gene on the SCOI complotype (Schneider et al., 1986; Yu et al., 1986). This TaqI band corresponds to the 70 kb BssHII fragment which also goes with this complotype. Comparison of the other band intensities in the TaqI digest for individual B showed the presence of a single long C4 gene (7.0 kb TaqI fragment) and 2 short C4 genes (5.4 kb TaqI fragment) at locus II. Since only one C4 gene lies on the 70 kb BssHII fragment it can be deduced that the 135 kb BssHII fragment contains one long C4 gene and two short C4 genes. The size of the fragment is consistent with this. The SacII results also indicate one chromosome with one short C4 gene (33 kb) and a novel larger band of 90 kb which would be consistent with three C4 genes, one long and two short, on the other chromosome. A similar gene organisation with duplicated C4B genes on one haplotype has been observed previously (Carroll et al., 1984b).

Individual C was essentially the same as individual A with respect to the fragments seen with BssHII, SacII and TaqI and therefore has the same C4 gene organisation although residing on a different HLA haplotype (Table 4.3).

Individual D had a different combination of BssHII fragments revealed with probe P_A. The 105 kb fragment contains one long C4 gene and one short C4 gene on one chromosome while the 135 kb band again

represents a chromosome with three C4 genes, one long and two short. The SacII fragments for individual D were 90 kb and 65 kb which is consistent with the presence of two C4 genes on one chromosome and three on the other. Examination of the results from the TaqI digest confirmed that this is the case. With P_A, the band intensities suggested the presence of two long C4 genes (7.0 kb TaqI fragment) and 3 copies of the short C4 gene (5.4 kb TaqI fragment). Similarly there are two 21-OHaseB genes and three 21-OHaseA genes. Therefore one chromosome has one long C4 gene and two short C4 genes with two 21-OHase A genes and one 21-OHase B gene. Schneider et al. (1986) have previously observed this organisation with homoduplication of the C4B gene on a haplotype that was C4A3 C4B1,1. The C4-typing for individual D suggests she possesses the same haplotype. In this case it is clearly demonstrated that since the size of the BssHII fragment is altered with the gene copy number present in an individual rather than the band intensity, it is easy to identify the number of C4 genes present. In addition the size of the BssHII fragment is also diagnostic of the C4 gene organisation on an individual chromosome.

Individual E had BssHII fragments consistent with two long C4 genes on one chromosome and one short C4 gene on the other SCOI haplotype and this was confirmed by the TaqI digests and the SacII results.

In Fig 4.8a it can also be seen that in addition to the series of strongly hybridizing bands revealed with P_A, a number of less intense BssHII fragments were observed. These additional bands were not previously observed in the cell line DNAs (Fig 4.5) but were observed in the cell line DNAs digested in parallel with the individuals DNA possibly because more enzyme was used (Fig 4.8a). They are probably the result of partial digestion of BssHII sites, perhaps due to partial

methylation at these sites. In particular, it appears that the BssHII sites that lie in the 3' end of the 21-OHase genes are being partially digested. This phenomenon does not affect the information obtained from the digestion because the informative bands are still the major products. Such faintly hybridising bands were not seen in the SacII digests (Fig 4.8b) presumably because as in Ice 5 (cell line 1) there are no additional SacII sites lying between the two that are cleaved to give the diagnostic fragment. Therefore it appears that the BssHII and SacII polymorphisms can be applied to DNA isolated from PBMNs and are diagnostic of the organisation of the C4 loci on each chromosome.

4.4 DISCUSSION

Using PFGE, large DNA fragment RFLPs for the enzymes BssHII, MluI and SacII have been observed at the complement gene loci in the HLA region. The size of the observed fragment with P_A or 21-OHase probes for these enzymes is directly related to the number and length of C4 genes present in the DNA samples analysed (Tables 4.2 and 4.3). These RFLPs, in particular those with BssHII and SacII can be used to directly observe the C4 gene organisation on both chromosomes using PBMN high mol. wt. DNA isolated from the whole blood of an individual. In combination with the previously described polymorphism for TaqI (Schneider et al., 1986), these RFLPs can give a complete picture of the C4 gene organisation, and by implication the 21-OHase gene organisation, for an individual without the need for family studies or DNA cloning. Since the size of the BssHII fragment observed is altered with gene copy number present on a chromosome, it is easy to identify the number of C4 genes without the need to interpret band intensities.

In addition it has been demonstrated that in the cell lines

studied, representing 7 different haplotypes (cell lines 2 and 3 are the same haplotype although they are derived from different sources), the only alterations in the DNA content at the 2-5 kb resolution of these PFGE gels are the deletions/duplications associated with the C4 and 21-OHase loci. The extent of these differences between the haplotypes studied are confined to the SacII fragment which contains the C4 and 21-OHase genes. This SacII fragment starts ∿10 kb 5' to the first C4 gene and ends 8 kb from the 3' end of the last 21-OHase gene (Fig 4.6). No other differences in the DNA content of the haplotypes studied over the extent of the MluI fragment that contains all the HLA class III complement genes were detected using PFGE. It seems that the rest of the DNA organisation around the C4 genes is relatively stable between haplotypes at least at the level detectable by PFGE. This observation may be of considerable significance. It is highly likely that a number of as yet uncharacterised genes lie close to the complement-21-OHase gene cluster (Chapter 6, Sargent, 1988). The existence of the RD gene between the C4 and factor B genes in mouse and man lends further support to this (Levi-Strauss et al., 1988). The existence of deletions and duplications at the C4 and 21-OHase loci might be explained by unequal crossover between sister chromatids (Carroll et al., 1985b) due to alignment of the highly related sequences at the A and B loci of C4 or 21-OHase. The absence of other disruptions around these genes might be explained by the absence of suitable duplicated sequences in the region or because of the presence of essential coding sequences which cannot be lost without deleterious effects.

The RFLPs described here will be invaluable in screening individuals for their C4 gene organisation. It is already known that C4 null alleles occur with increased frequency in certain HLA-associated diseases (Fielder et al., 1983) and this method in combination with the

other available RFLPs should give a rapid screening procedure for splitting deleted from non-deleted null alleles. It is also interesting to note that of the five individuals that were analysed (Table 4.3), there were two chromosomes with deleted C4BQ0 alleles and two chromosomes with non-deleted C4 null alleles. In addition there were two examples of chromosomes with three C4 genes present, both with one long C4 gene and two short C4 genes. Although there have been previous reports of three C4 genes on one chromosome (Carroll et al., 1984b; Schneider et al., 1986), it is still not clear how frequently these duplications occur in the population. Using the described polymorphisms it should be possible to assess the frequency of these deletions and duplications.

Both BssHII and SacII give the same information as to the organisation of the C4 gene loci. The fragments produced with SacII are 40-45 kb smaller than those observed with BssHII using the P_A or 21-OHase probes and this may make it easier to resolve size differences between different gene organisations. Also, since there are no uncleaved SacII sites within the observed fragments, at least from our observations of cell line 1, no smaller partial fragments are produced. However, routinely BssHII gives complete digestion with less units of enzyme, whereas SacII can give very partial digestion, and BssHII has given more consistent results.

Finally it may be possible to make use of these RFLPs along with isotype and Rg/Ch-specific oligonucleotides to give a complete C4 genotype for an individual with a single PFGE gel. The procedure of Wallace (Thein & Wallace, 1986) for hybridisation of oligonucleotides to dried down gel wafers could be usefully applied to PFGE gels for this purpose.

CHAPTER 5

ANALYSIS OF THE VARIATION IN THE LONG RANGE GENOMIC ORGANISATION OF THE HLA CLASS II AND CLASS III REGIONS BY PFGE

5.1 INTRODUCTION

The genomic structures of the three major linked gene clusters of the human MHC have been studied using PFGE. The class I loci (HLA-A, B, C and E) cover >1.5 Mb (Lawrance et al., 1986; Dunham et al., 1987; Carroll et al., 1987) at the telomeric end of the complex. As many as 20-40 other highly related sequences are also located close by. The class II loci (HLA-D) are found in a ∿1.1 Mb region at the centromeric end of the MHC arranged in subregions, DP, DQ and DR (Hardy et al., 1986), each subregion containing at least one A/B pair of genes. In addition a number of class II pseudogenes are also located in the HLA-D region.

Lying between the class I and class II regions are a number of unrelated genes which together comprise the class III genes (Lamm & Olaisen, 1985; Dunham et al., 1987; Sargent, 1988). Genes encoding factor B, C2, C4A, C4B, 21-OHase and the novel protein RD lie in a cluster (Carroll et al., 1984a; 1985a; Levi-Strauss et al., 1988) at least 300 kb telomeric of the DRA gene (Dunham et al., 1987; Chapter 3) and orientated with the C2 gene being telomeric to the 21-OHaseB gene. A further 92 kb telomeric of the C2 gene lie duplicated copies of a gene for the major heat shock protein, HSP70 (Sargent et al., 1988) at least one of which is functional. In addition the genes for the lymphokines, tumour necrosis factors α and β are located ∿390 kb telomeric of the C2 gene and ∿250 kb centromeric of HLA-B (Dunham et al., 1987; Chapter 3).

A number of the genes of the human MHC exhibit extensive polymorphism with multiple allelic variants present in the population. The HLA-A, B and C products exhibit extensive polymorphism, there being over 87 recognised class I specificities (Bodmer et al., 1988), and their genes show a number of RFLPs (Strachan, 1987). Of the class II region products all the chains are polymorphic except the DR_{α} chain. The DRB1 gene shows the most variation with over 20 alleles identified (Bell et al., 1987a). The products of the complement C4 loci also show exceptional variability (Mauff et al., 1983). C2 and factor B show limited polymorphism (Alper, 1981). There is also some evidence suggesting allelic variation in the DNA sequence of the 21-OHase genes (Rodrigues et al., 1987) although this has not been firmly established.

In addition there exists variation in the number of copies of some MHC genes present between different haplotypes. For instance, the number of DRB genes may vary from possibly one to four depending on the DR specificity (Böhme et al., 1985; Bell et al., 1987a). Variations in the number of C4 and 21-OHase genes present on different haplotypes have been extensively documented, the number of genes present varying from one to at least three (Carroll et al., 1984b; 1985b; Schneider et al., 1986; Chapter 4). Although limited long range variability around the class I loci has been demonstrated (Chimini et al., 1988), it is unclear whether there is variation in the size or organisation of the human MHC, particularly in the extensive regions that do not encode known protein products.

Susceptibility to a number of diseases has been shown to be associated with the possession of certain HLA haplotypes. Contributory roles have been suggested for many of the known MHC gene products in disease aetiology (Todd et al., 1988; Batchelor and McMichael, 1987; Jacob & McDevitt, 1988). An alternative explanation is that HLA alleles

may be in linkage disequilibrium with another allele which is of pathogenic significance. It has recently become obvious that there may be several other loci within the human MHC particularly within the region between the class II and class I genes (Levi Strauss et al., 1988; Dunham et al., 1987; Sargent, 1988). Therefore it is important to understand whether there are significant differences in the organisation of the human MHC between haplotypes in order to determine what, if any, influence these might have on disease susceptibility. Therefore, PFGE was used to compare the structures of the HLA class II and class III regions in the eight HLA homozygous cell lines used in Chapter 4.

5.2 EXPERIMENTAL STRATEGY

In order to assess whether gross alterations in the structure of the human MHC genomic organisation exist, the class II and class III regions in eight HLA homozygous cell lines representing seven common HLA haplotypes were compared by PFGE using essentially the same strategy as described in Chapter 4. High mol. wt. DNA from one HLA homozygous, consanguineous lymphoblastoid cell line (cell line 1) and seven HLA homozygous-typing cell lines (cell lines 2-8) was prepared in agarose blocks. The HLA types of the cell lines are given in Table 4.2. DNA from each cell line was digested with the infrequently cutting restriction enzymes MluI, PvuI, NotI, NruI, BssHII, PvuI+MluI and NotI+NruI and the digested samples were separated on PFGE gels at an appropriate switching interval for the fragment sizes of interest.

Restriction enzyme fragments separated by PFGE were transferred to GeneScreenplus membranes and these filters were hybridised successively with a range of probes known to give complete coverage of the class II and class III regions of the human MHC (Dunham et al., 1987). No attempt

was made to study the class I loci as the physical map in that region is still incomplete. The sizes of restriction fragments produced with each probe were compared between cell lines and differences observed were analysed for consistency with all the enzymes used. In addition, the information for each individual cell line using different rarely cutting enzymes was pooled from different gel runs to produce physical maps of the class II and class III regions of each cell line which were internally consistent for each single and double digest. The maps produced for each cell line were then compared to identify similarities and differences.

5.3 RESULTS

5.3.1 Probes

The positions of the probes used are illustrated in Figs 5.1 and 5.2. A number of probes were isolated from the two clusters of overlapping cosmid clones covering 541 kb of the class III region (Sargent, 1988; see section 2.1.6).

5.3.2 PFGE Southern blot analysis

The sizes of restriction fragments observed for each cell line and enzyme for the class III and class II regions are presented in Tables 5.1 and 5.2, respectively. In general the fragment sizes observed using the "waltzer" PFGE system were smaller than observed with the OFAGE gels and represent a more accurate size estimate as discussed in Chapter 4. Therefore the physical maps of the different cell lines in the class III and class II regions shown in Figs 5.1 and 5.2 have somewhat smaller

*Table 5.1 Sizes of restriction fragments (kb) observed after hybridisation of MHC class III region probes to Southern blots of restriction enzyme digests of cell line genomic DNA. Full HLA types of the cell lines are given in Table 4.2. Where two probes gave the same result the fragment sizes are given on the same line. Multiple sets of data indicate that several fragments were observed for the same probe. * indicates fragments that are the products of partial digestion. ND indicates not determined. +, BssHII results for cell line 1 are also given in Table 6.2, the differences in the size estimates are due to the different gel runs used.*

ENZYME	PROBE	CELL LINE							
		1	2	3	4	5	6	7	8
	C4 type	A3 BQ0	AQ0 B1	AQ0 B1	A3 BQ0	A4 B?	A3 BQ0	A3 B1	A4 B5
MluI	21-OHase	200	200	200	200	225	230	230	225
	J	240	240	240	240	240	240	240	240
	TNFA	145	145	145	145	145	145	145	145
PvuI	21-OHase/J	310	310	310	310	340	340	340	340
	TNFA	700	700	700	700	700	700	700	700
MluI+PvuI	21-OHase	200	200	200	200	230	230	230	230
	J	60	60	60	60	60	60	60	60
	TNFA	145	145	145	145	145	145	145	145
NotI	21-OHase	950	920	920	920	980	>1050	950	950
	J	150	150	150	150	150	150	150	150
	TNFA	220	220	220	220	220	220	220	220
NruI	21-OHase	550	550	550	550	580	580	580	580
	J	ND	ND	ND	ND	ND	ND	ND	ND
	TNFA	80	80	80	80	80	80	80	80
NotI+NruI	21-OHase	250	250	250	250	280	280	280	280
	J	150	150	150	150	150	150	150	150
	L	140	140	140	140	140	140	140	140
	TNFA	80	80	80	80	80	80	80	80
BssHII	C	30	38*	38*	30	30	30	38*	38*
			30	30				30	30
	B	8	38*	38*	8	8	8	38*	38*
			8	8				8	8
	21-OHase	80	70	70	80	105	115	115	105
	1.7 kb BamHI	12	12	12	12	12	12	12	12
	K	55	55	55	55	55	55	55	55
	J	62	62	62	62	62	62	62	62
	HSP	26	26	26	26	26	26	26	26
		7	7	7	7	7	7	7	7
	F/G	55	55	55	55	55	55	55	55
	E	70	70	70	55	70*	70*	70	70*
						55	55		55
	F	70	70	70	16	70*	70*	70	70*
						16	16		16
	N	31	31	31	31	31	31	31	31
	TNFA	280	280	280	280	280	280	280	280

Table 5.2 Sizes of restriction fragments (kb) observed after hybridisation of MHC class II region probes to Southern blots of restriction enzyme digests of cell line genomic DNA. Full HLA types of the cell lines are given in Table 4.2. Where two probes gave the same result, the fragment sizes are given together. Multiple sets of data indicate that several fragments were observed for the same probe. In the PvuI+MluI digests the MluI digestion was partial and hence fragments that are the result of single digestion with PvuI were also observed. ND indicates not determined.

ENZYME	PROBE	CELL LINE							
		1	2	3	4	5	6	7	8
	DR type	2	3	3	3	2	4	5	6
MluI	DRA	540	650	650	650	540	>750	650	520
	DRB	540	650	650	650	540	>750	650	520
	DQA	540	650	650	650	540	>750	650	520
		125				410			125
						320			
	DOB	ND	ND	ND	ND	ND	ND	ND	ND
	DPA	280	280	280	280	410	90	280	280
			90	140		90		90	90
				90					
	DPB	110	110	110	110	110	110	110	110
PvuI	DRA	350	390	330	390	350	730	390	390
			330						
			290						
	DRB	350	390, 60	330	390	350	730	390	390
		275	330, 40	200	200	275		200	200
		125	200	60		125			
	DQA	275	200	200	200	275	730	200	200
		125				125			125
									70
	DOB/DPA/DPB	370	370	370	370	370	370	370	370
PvuI+MluI	DRA	350	390	330	390	350	>700	390	390
			330						
			290						
	DRB	350	390, 60	330	390	350	>700	390	390
		275	330, 40	200	200	275		200	200
		125	200	60		125			
	DQA	275	200	200	200	275	>700	200	200
		125				125			125
									70
	DOB	370	370	370, 140	370	370	370	370	370
		280	280	280	280	280	280	280	280
			170	170		170	170	170	170
	DPA	370	370	370, 140	370	370	370	370	370
		280	280	280	280	280	280	280	280
			90	90		90	90	90	90
	DPB	370	370	370	370	370	370	370	370
		90	90	90	90	90	90	90	90
					80				
NotI	DRA/DQA/DOB	950	920	920	920	980	>1050	950	950
	DRB	ND	ND	ND	ND	ND	ND	ND	ND
	DPA	340	340	340	340	340	340	340	340
	DPB	ND	ND	ND	ND	ND	ND	ND	ND
NruI	DRA	>1200	>1200	>1200	>1200	>1200	>1200	>1200	>1200
NotI+NruI	DRA/DQA	690	660	660	660	690	770	660	660
	DPA	340	340	340	340	340	340	340	340
BssHII	DRA	490	460	460	460	340	270	270	270
			270	270	270	270			
	DRB	490	460	460	460	340	490	360	370
			200	200	200	230			
						160			
	DQA	490	460	460	460	230	490	360	370
		200	200	200	200	200			
						160			
	DOB	200	200	200	200	200	490	360	370
	DPA/B	310	310	310	310	310	310	310	310
		230	230	230	230	230	230	230	230
								115	

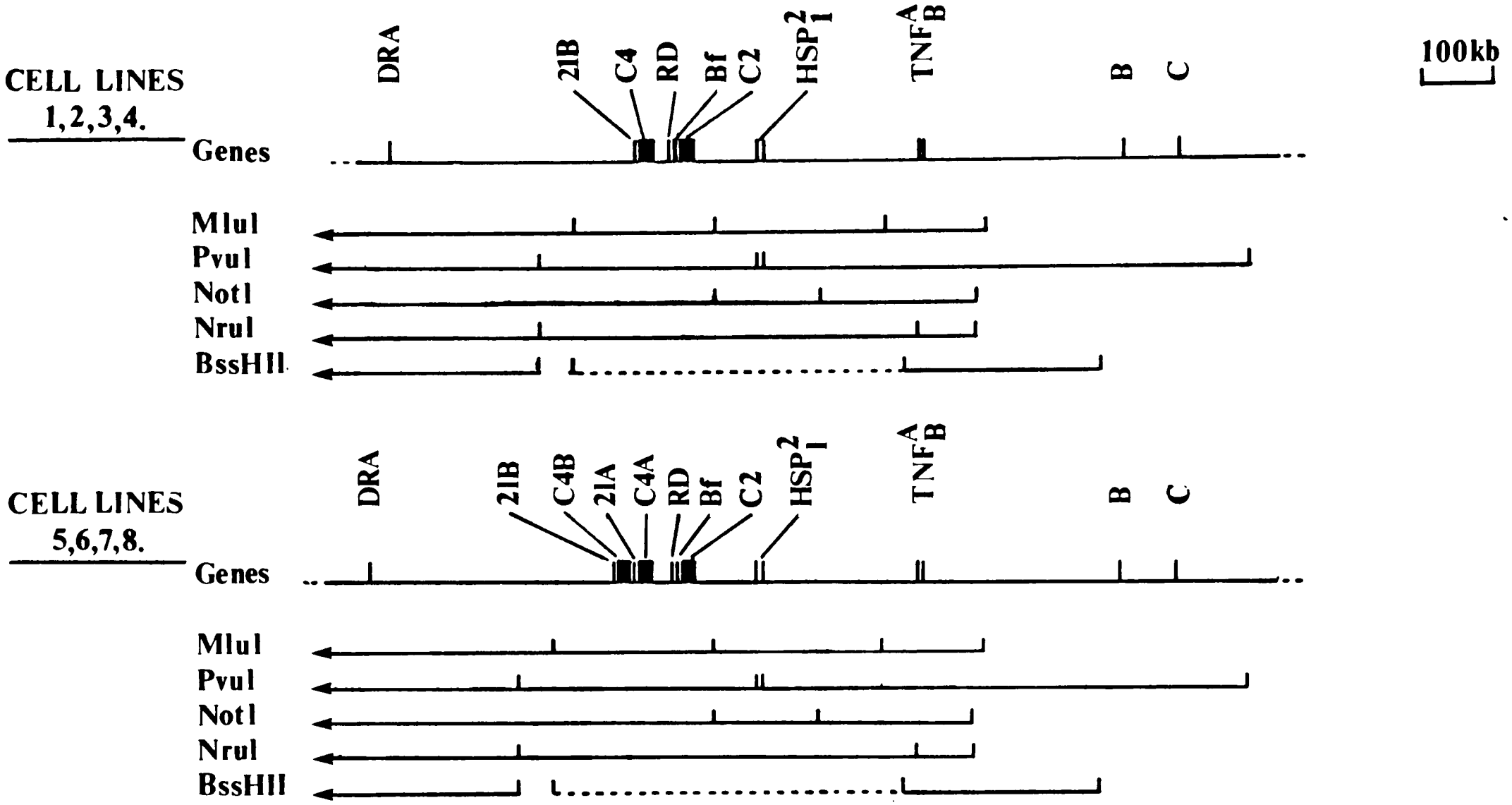


Fig 5.1a Long range maps of the class III region of the MHC in eight HLA homozygous cell lines. The centromere is to the left. At this scale the haplotypes can be split into those with one C4 gene and those with two C4 genes. No other difference in the amount of DNA present was detected. The top line of each map shows the positions of known genes in the class III region. The restriction fragments observed are below the gene map, vertical lines indicating restriction sites, arrowheads indicate that fragments extend into the class II region. The dotted line for BssHII shows the region covered by the BssHII mapping presented in Fig 5.1b. Note that a PvuI site is cleaved within each copy of the duplicated HSP70 genes to give an 11 kb fragment (Sargent et al., 1988).

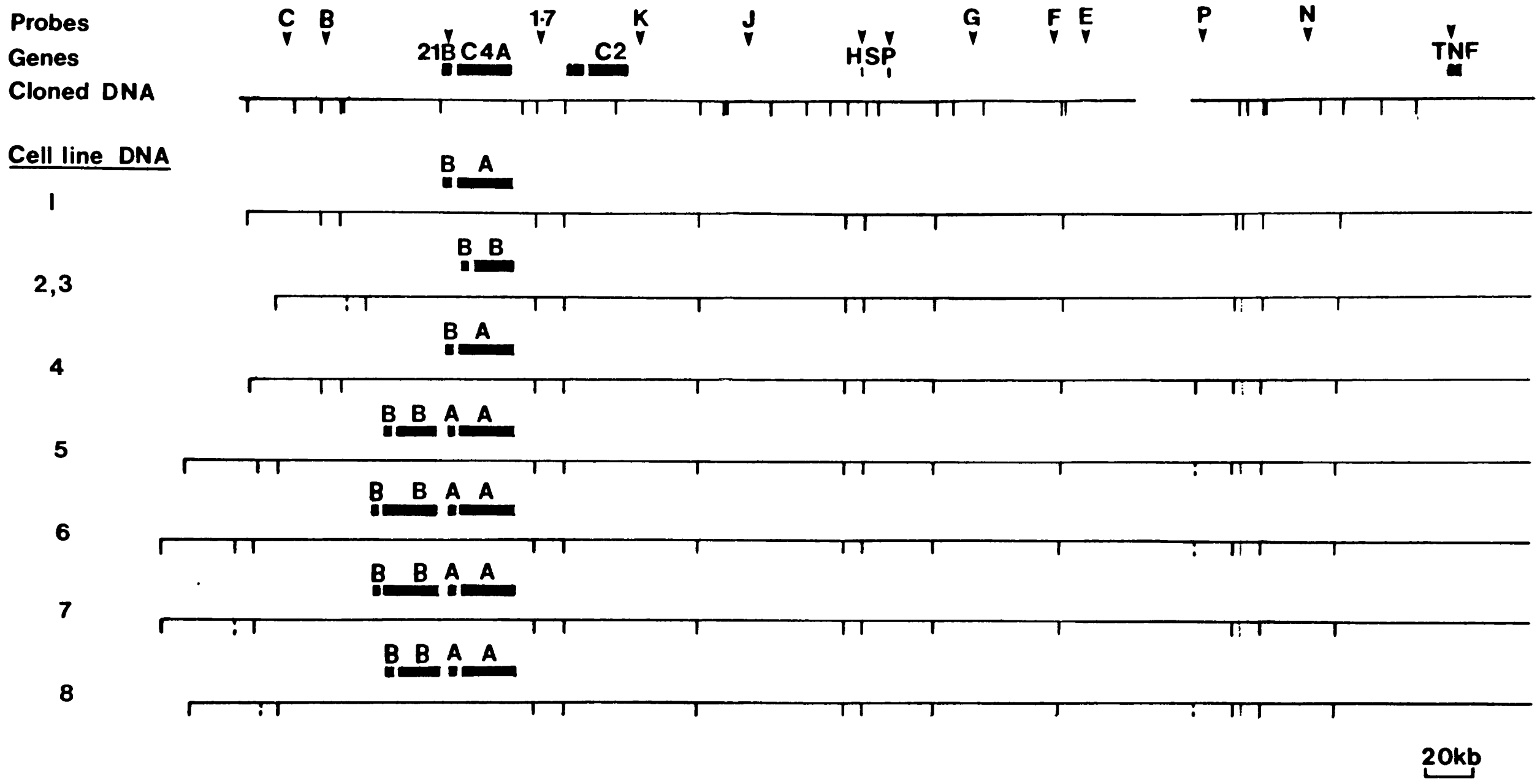


Fig 5.1b BssHII restriction sites cleaved in class III region genomic DNA from eight HLA homozygous cell lines. The positions of the probes used in the analysis are marked by arrowheads. The top line shows the BssHII sites mapped in cloned cosmid DNA from cell line 1 up to the TNFA and B genes, the positions of the sites are indicated by vertical lines. Below this are genomic BssHII restriction fragment maps for each cell line. Filled boxes are genes. Only the positions of the C4 and 21-OHase genes are illustrated for the genomic maps, the type of gene being indicated by the size of the box for C4 genes and by A or B as appropriate.

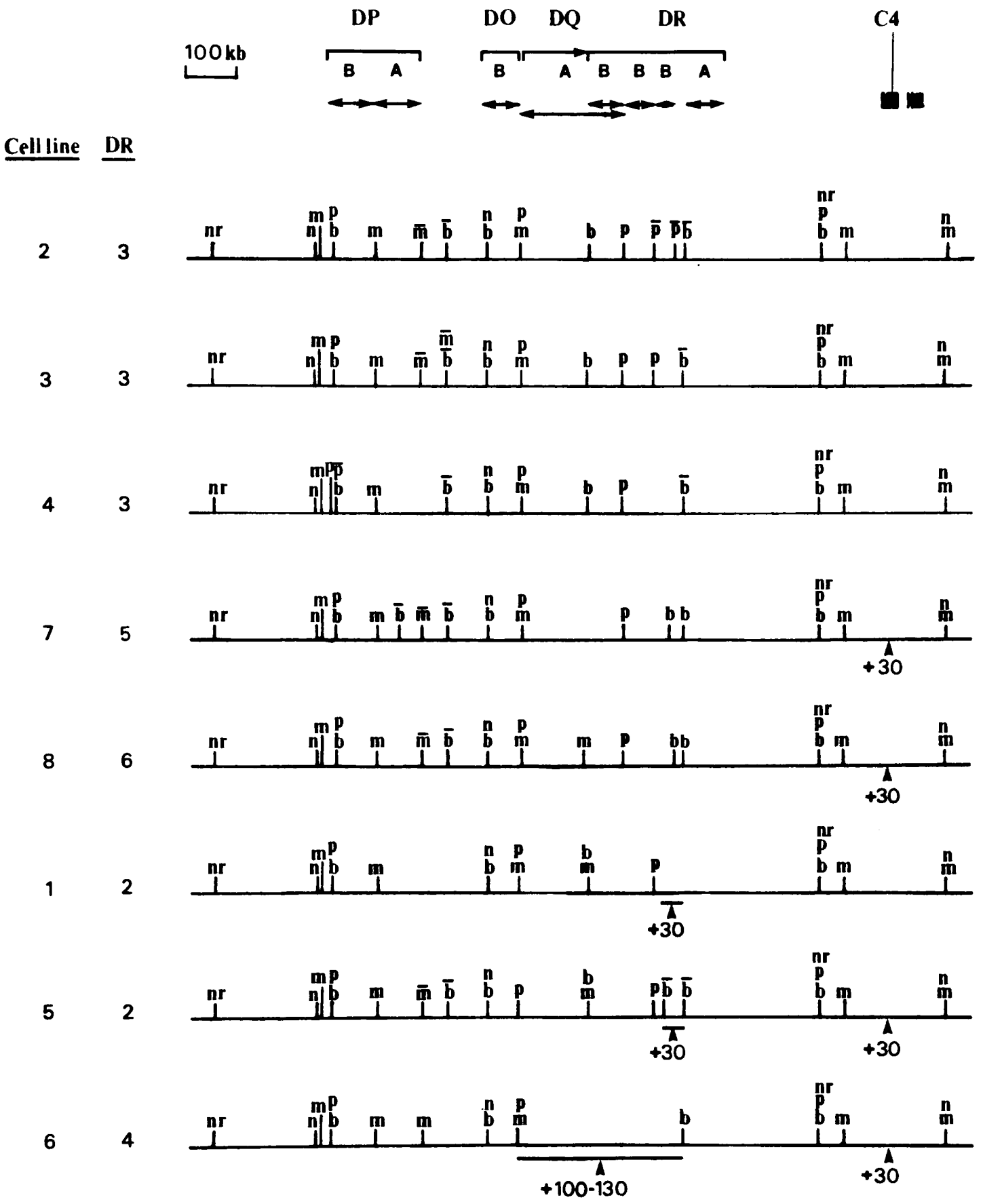


Fig 5.2 Comparison of the molecular maps of the class II regions of eight HLA homozygous cell lines. The centromere is to the left. The positions of the limits of the class II subregions are shown at the top. The limits of the regions that could contain class II genes are indicated by arrows. The location of the DRA gene is limited by the position of a BssHII site and the distance from the nearest DRB gene as found in cosmid clones (Spies et al., 1985). Positions of restriction sites are indicated by vertical lines, enzymes are, n, NotI; m, MluI; p, PvuI; nr, NruI; b, BssHII. The maps for the cell lines have been aligned by allowing for the extra DNA present in certain haplotypes. The positions and amounts of DNA (kb) that have been removed are marked by arrowheads and numbers beneath the lines.

distances than predicted in Chapter 3. The physical distances in the class III region based on these results are estimated as 325 kb between the C2 and TNFA genes, 250-300 kb between the TNF and HLA-B genes and <390 kb between the 21-OHaseB and DRA genes (see below). A comparison of the restriction fragment sizes observed with the two systems is shown in Table 5.3. Seven of the cell lines used here are serologically homozygous and may exhibit heterozygosity at some loci. Therefore it is impossible to distinguish the case where partial digestion is occurring at a restriction enzyme site on both chromosomes from the situation where a restriction enzyme site is absent or uncleavable on one chromosome but not the other. Also it has recently been demonstrated that homologous chromosomes can be differently methylated in the same tissue (Silva & White, 1988). It is well established that rarely cutting restriction endonucleases are sensitive to cytosine methylation (Lindsay & Bird, 1987). Therefore, the maps presented here must be regarded as composite maps of both copies of the MHC within a cell line. This may or may not be the same as the map of a single haplotype.

5.3.3 The class III region of the human MHC is highly invariant in structure outside the C4 and 21-OHase loci

Hybridisation of MluI-digested cell line DNA with 21-OHase, J and TNFA probes demonstrated that the structure of 585 kb of the class III region is constant between the 8 cell lines except for the deletion of one C4 and 21-OHase gene in four cases (Fig 5.3). The TNFA and J probes hybridised to MluI fragments of 145 kb and 240 kb, respectively, in all cell lines. The 21-OHase probe revealed the 25-30 kb difference in the size of the MluI fragment that encompasses the complement genes between cell lines 1-4 and cell lines 5-8 (see Chapter 4).

Table 5.3 Comparison of some restriction fragment sizes (kb) observed for cell line 1 (Ice 5) with the two PFGE systems.

ENZYME	PROBE	FRAGMENT SIZE (kb)	
		OFAGE	"WALTZER"
NotI	DPA	425	340
	DRA/21-OHase	980	950
	J	210	150
	TNFA	270	220
NruI	21-OHase	640	550
	TNFA	100	80
PvuI	DPA	390	370
	DQA	290	275
	DRA	380	350
	21-OHase/J	340	310
	TNFA	780	700
MluI	DPB	90	110
	DPA	290	280
	DQA	550	540
		150	125
	DRA	550	540
	21-OHase	225	200
	J	270	240
	TNFA	190	145
BssHII	DPA	240	230
	DQA	480	490
		190	200
	DRA	480	490
	21-OHase	100	80
	TNFA	340	280

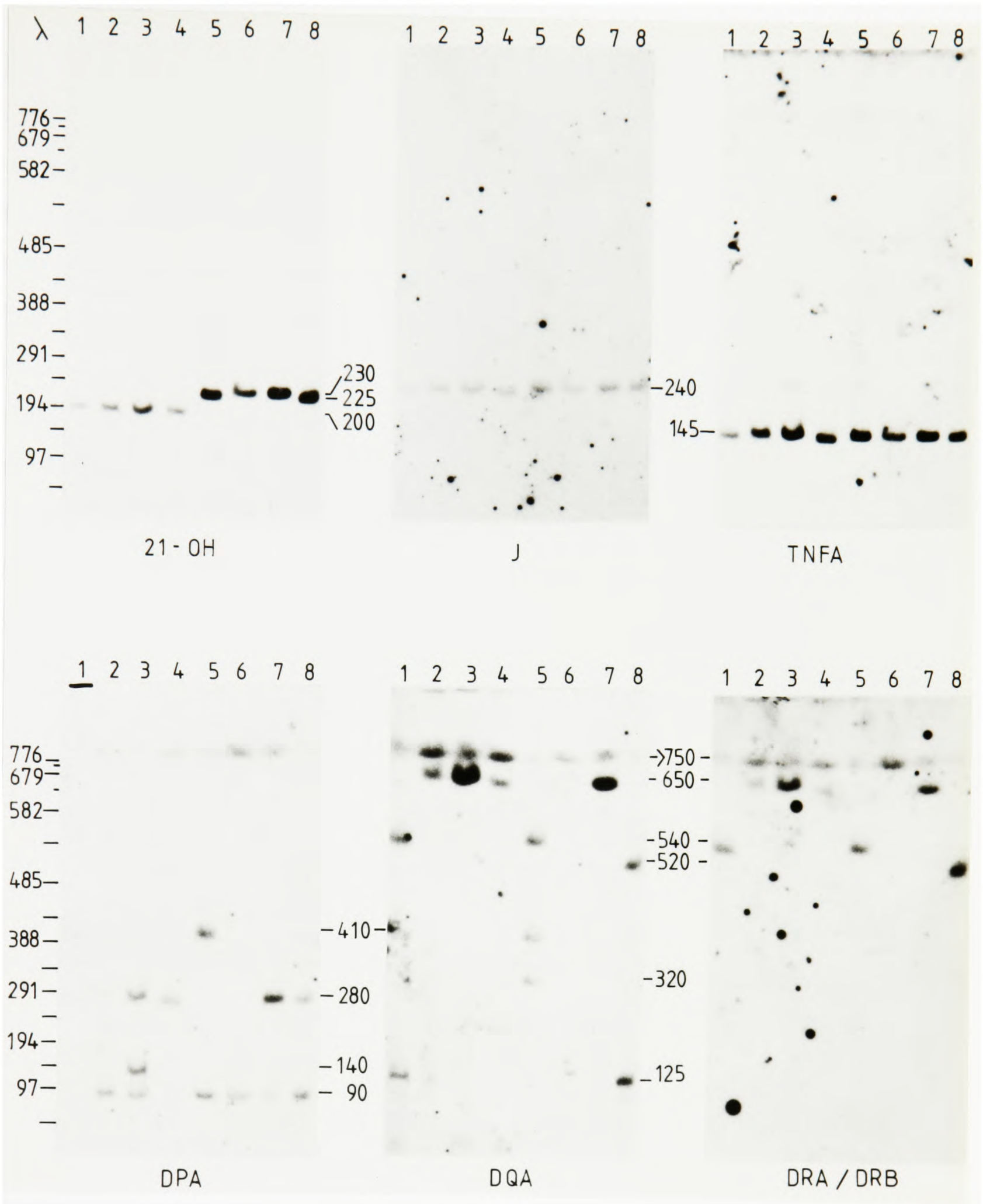


Fig 5.3 Genomic Southern blot analysis of MluI-digested DNA from eight HLA homozygous cell lines. MluI-digested DNA was separated by PFGE at a 30 s switching interval. The panels show the results of sequential hybridisation of the same Southern blot filter with the probes indicated beneath. Numbers above each autoradiograph refer to the cell lines. The positions of λ concatemer markers are on the left. Fragment sizes are in kb.

The fragments observed with the class III region probes using PvuI, NruI, NotI and the double digests (Table 5.1) confirmed that within the limits of resolution of these PFGE gels, estimated to be ±5% of the fragment size, the DNA organisation of the class III region of the human MHC is constant in these 8 cell lines representing 7 common haplotypes except for the known alterations at the C4 and 21-OHase loci. This is illustrated in the maps presented in Fig 5.1a. The region covered by the fragments analysed extends from the HLA-B and C loci which are contained within the 700 kb PvuI fragment that hybridised to the TNFA probe (Dunham et al., 1987; Chapter 3) to ~140 kb centromeric of the 21-OHaseB gene. The results at the telomeric end of the class III region were further confirmed by the observation of a 280 kb BssHII fragment that hybridised to the TNFA probe in all cell lines and which extends towards the HLA-B locus. It is interesting to note that with the class III region probes no RFLPs were observed between the haplotypes studied with the rare cutting enzymes illustrated in Fig 5.1a.

In order to confirm these observations at finer resolution the cell lines were analysed further with BssHII. 541 kb of DNA from the class III region of cell line 1 was cloned in two clusters of overlapping cosmid clones and these cosmid DNA inserts were mapped with BssHII (Fig 5.1b). Extensive mapping using PFGE and probes from across the cloned region has determined that BssHII gives a number of restriction fragments across the class III region which would be suitable for the detection of small (2-5 kb) differences between cell lines. Comparison of these data with the cosmid BssHII map revealed which BssHII sites were cleaved in cell line 1 genomic DNA (Chapter 6, Fig 5.1b). Not all BssHII sites identified in the cloned cosmid DNA were cleaved in cell line 1 genomic DNA presumably because uncleavable sites are methylated at CpG dinucleotides (Lindsay & Bird, 1987). Therefore DNA in agarose

blocks from each cell line was digested with BssHII and separated by PFGE at a 7.5 s switching interval to optimise resolution in the 5-150 kb size range. The separated DNA was transferred to GeneScreenplus membranes and hybridised sequentially with probes that cover the extent of the class III region cloned in cosmids. The results obtained are shown in Fig 5.4 and the results are summarised in Table 5.1.

Fig 5.4 shows that hybridisation of probes covering the DNA from just 5' of the C4 gene to probe F defined identically size BssHII fragments for all cell lines. Since the positions of the BssHII sites that are restricted in cell line 1 and the positions of the probes are known it can be concluded that the same sites are cutting in the other cell lines over ∿220 kb of class III DNA. The size of the BssHII fragment that hybridised with the 21-OHase probe is entirely determined by the number and size of the C4 and 21-OHase genes present on the particular haplotype as described in Chapter 4 and it can be concluded that the same BssHII site 56 kb centromeric of the 3' end of the most centromeric C4 gene is restricted in each cell line. Centromeric of this BssHII site the restriction fragment patterns are split into two types. Cell lines 1, 4, 5 and 6 gave an 8 kb fragment with probe B and a 30 kb fragment with probe C. This indicates that in these cell lines the site that lies between the two probes was cleaved completely. In cell lines 2, 3, 7 and 8, an additional BssHII fragment of 38 kb was observed with both probes equivalent to the sum of the digestion products observed to hybridise to each probe individually (Table 5.1). There are several possible explanations for this. First there could be heterozygosity within the cell lines at the BssHII site that lies between the two probes so that the site is not present on one chromosome in some of the cell lines. Second, the apparent partial digestion could be generated by partial methylation at the BssHII site within the cell population.

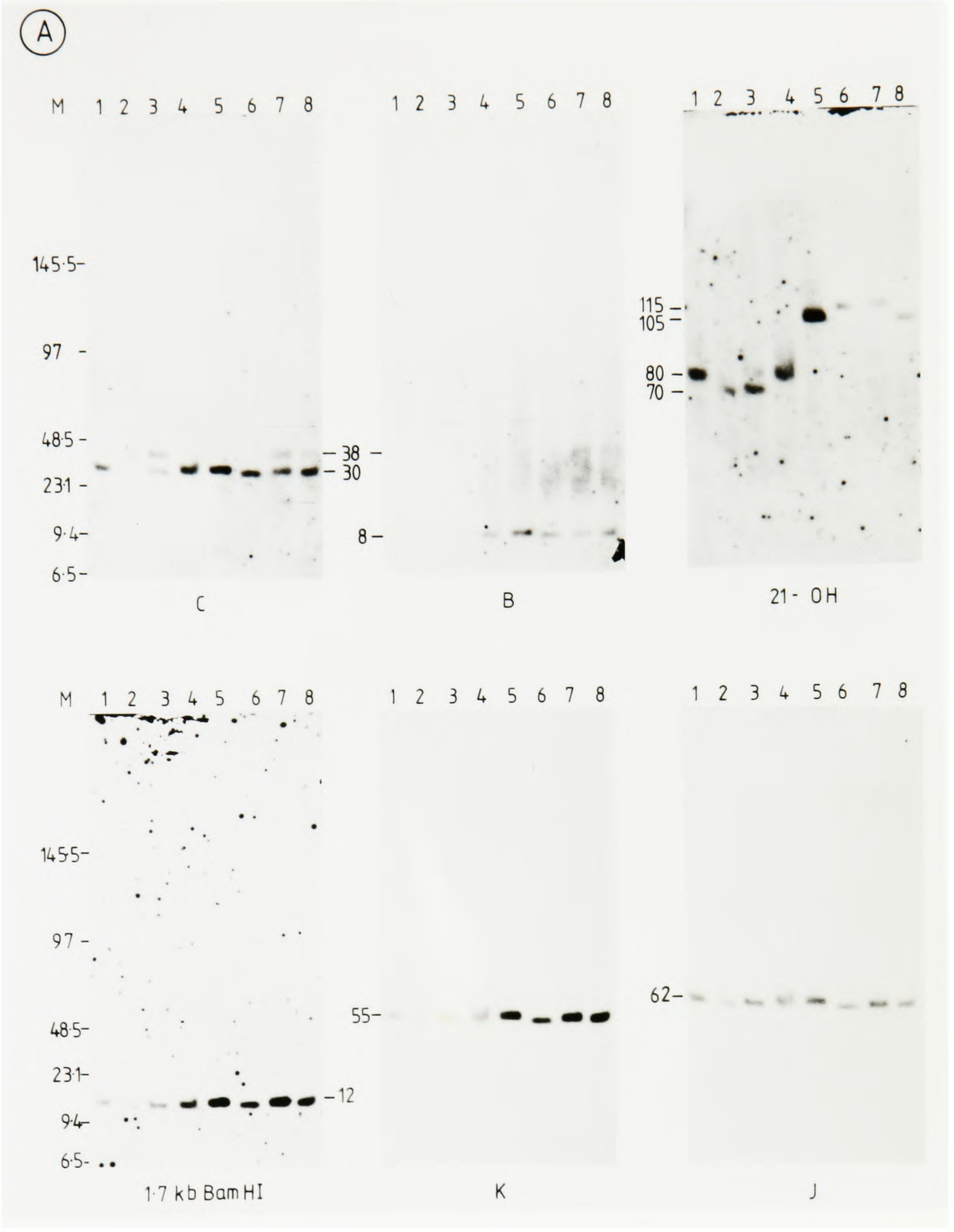


Fig 5.4 Genomic Southern blot analysis of BssHII-digested DNA from the eight cell lines. (A) and (B) BssHII-digested DNA was separated by PFGE at a 7.5 s switching interval. The results of hybridisation of the probes shown to a Southern blot of the separated digests are shown. The positions of λ concatemers and λ/HindIII markers are shown on the left (M). (C) The result of hybridisation of the TNFA probe to a Southern blot of BssHII-digested DNA separated by PFGE at a 25 s switching interval. The positions of λ concatemers are shown on the left. 1-8 refer to the cell lines. Fragment sizes are given in kb. Continued overleaf.....

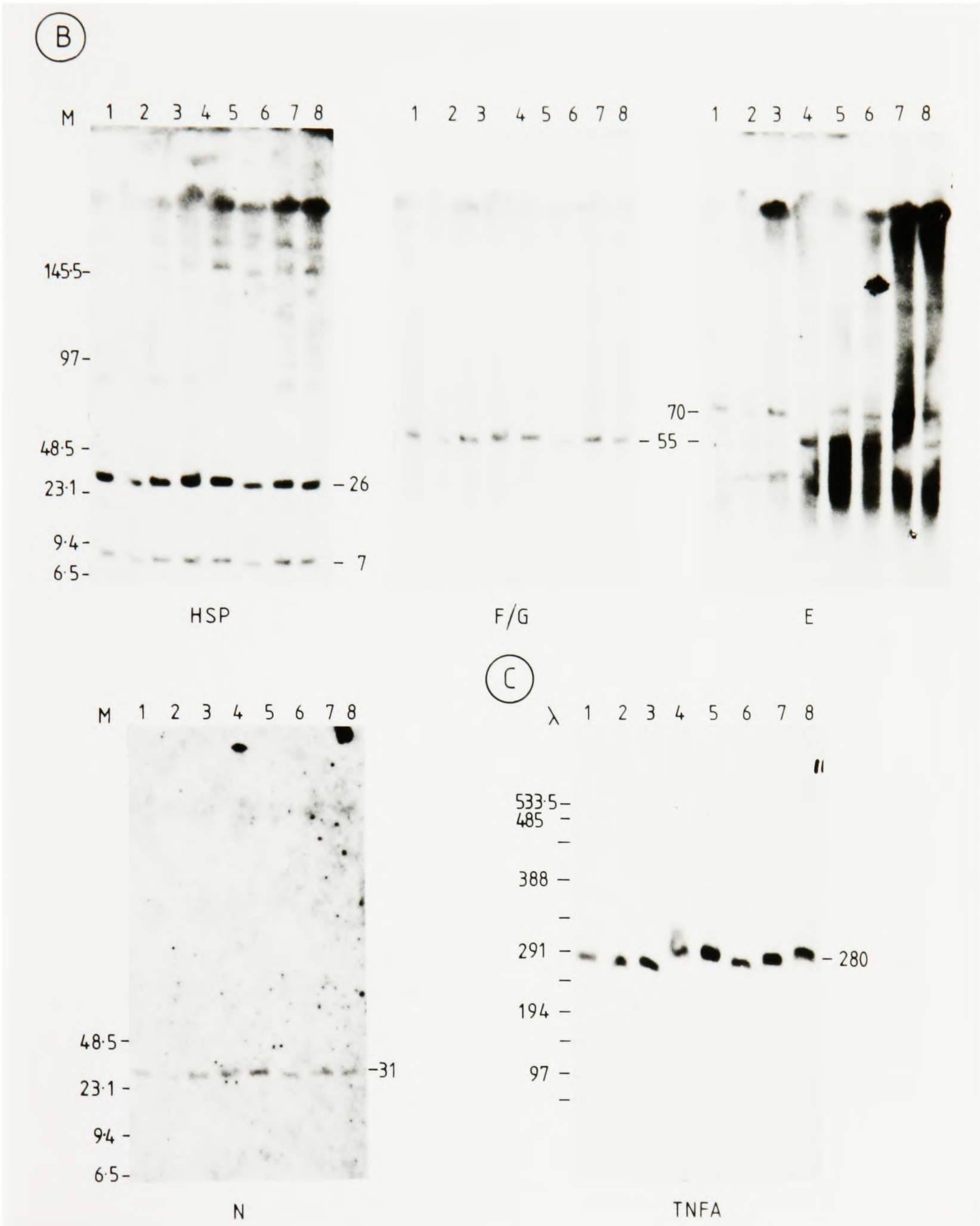


Fig 5.4 Continued.....

Finally some other property of the DNA preparations of the refractory cell lines may make the BssHII site difficult to restrict. A similar situation was observed with probe E. In cell lines 1, 2, 3 and 7, a 70 kb BssHII fragment was shown to hybridise to this probe. In cell lines 5, 6 and 8 an additional 55 kb fragment was also seen which was the only observed product in cell line 4. Probe P hybridised to the 70 kb fragment in cell lines 1, 2, 3 and 7 and to the expected products in the other cell lines. The BssHI site that is cleaved in the cell line 4 and lies ∿100 kb from the 3' end of the TNFA gene is known to be present in all the cell line 1 cosmid clones for this region but is not cleaved in genomic DNA. It therefore seems likely that in the cell lines other than cell line 4 this BssHII site is subject to partial or complete methylation that makes it refractory to cleavage.

Regardless of these minor differences at two BssHII sites the overall conclusion from these results is that over approximately 450 kb of the human MHC class III region the DNA content does not vary between the different haplotypes studied except at the C4 and 21-OHase loci.

5.3.4a Haplotype-specific differences in DNA organisation exist in the class II region

Hybridisation of NotI-digested cell line DNA with DOB, DQA, DRA or 21-OHase probes revealed differences in the fragment sizes observed (Fig 5.5). Some of the differences observed can be accounted for by the number of C4 gene units present in each cell line. For instance cell lines 2, 3 and 4 which have single C4 and 21-OHase genes each possessed a 920 kb NotI fragment that hybridised with these probes. Cell lines 7 and 8 which have two C4 and two 21-OHase genes both possessed a 950 kb NotI fragment. The 30 kb difference in the size of the NotI fragment is

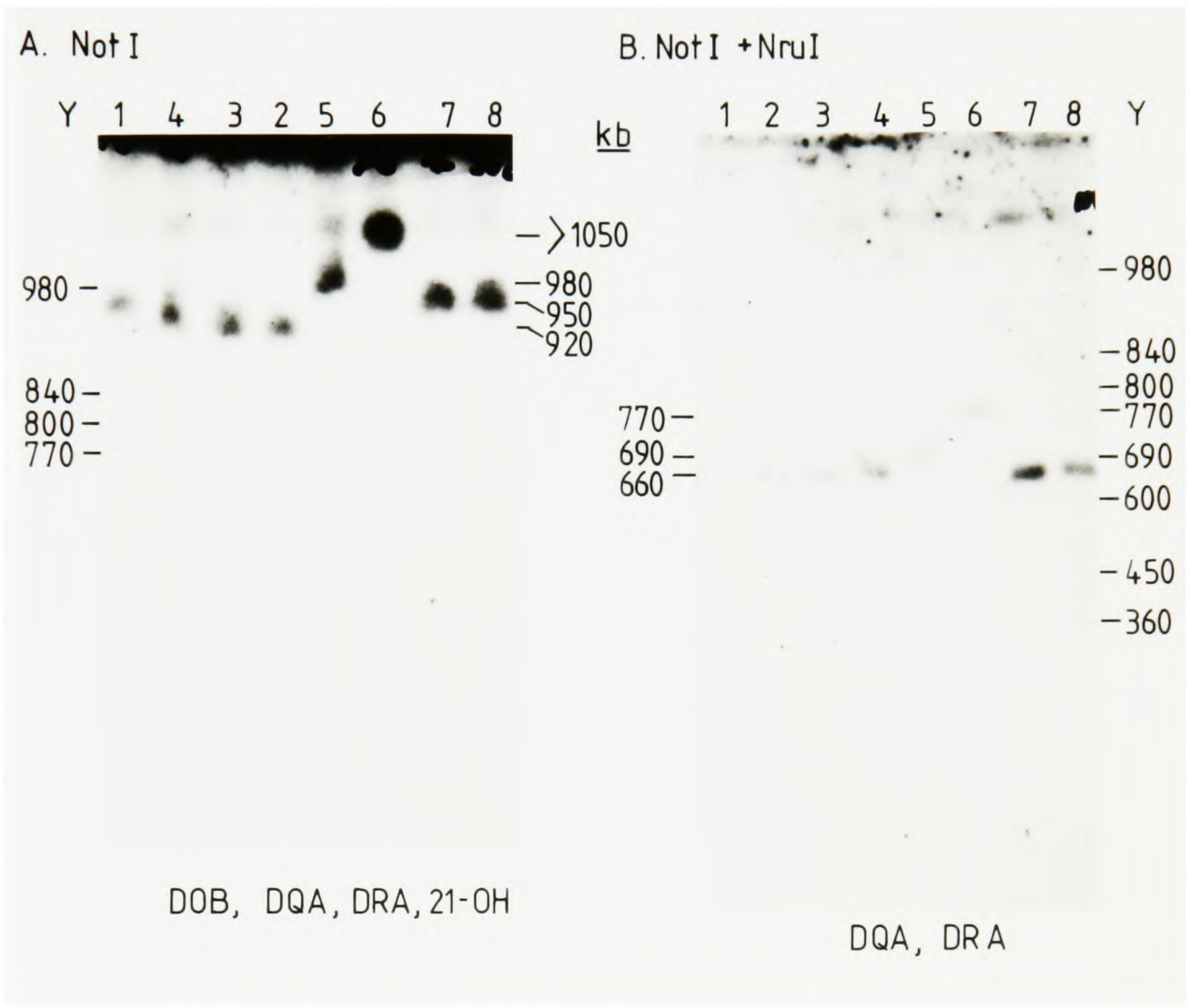


Fig 5.5 Genomic Southern blot analysis of NotI and NotI+NruI-digested DNA from the cell lines. A. Representative autoradiograph showing the result of hybridisation of the probes shown to NotI genomic DNA digests separated at a 58.5 s switching interval by PFGE. B. Representative autoradiograph showing the result of hybridisation of the probes shown to NotI+NruI genomic DNA digests separated at a 53.5 s switching interval. Y indicates the positions of some of the yeast chromosome markers. Fragment sizes are given in kb.

completely due to the additional 30 kb provided by the second C4 and 21-OHase genes present in cell lines 7 and 8. Similarly cell line 5 which has two C4 and 21-OHase genes had a 980 kb NotI fragment while the other DR2 haplotype, cell line 1, with single C4 and 21-OHase genes had a 950 kb NotI fragment. Hybridisation of the DRA probe with a NotI/NruI digest showed that when the region of the NotI fragment containing the C4 genes was removed, there were three sets of restriction fragment sizes. Cell lines 2, 3, 4, 7 and 8 had a 660 kb NotI/NruI fragment, cell lines 1 and 5 had a 690 kb fragment and cell line 6 had a 770 kb fragment (Fig 5.5). Therefore, there appeared to be three sets of sizes for the DRA and DQA hybridising region. Cell lines 2, 3, 4, 7 and 8 were placed in one group and cell lines 1 and 5 in another. The DR4 haplotype in cell line 6 has a NotI/NruI fragment at least 80 kb larger than that of cell line 5 and 110 kb larger than cell line 2. Hybridisation of NotI digested DNA with the DPA probe showed that each cell line has identically sized 340 kb NotI fragments in the DP region. Since previous mapping data (Dunham et al., 1987; Chapter 3) and the results of partial NotI digestions (not shown) had shown that the DPA NotI fragment is immediately adjacent to the NotI fragment that contains the DOB, DQA, DRA and 21-OHase genes, it is highly unlikely that the difference in size observed with the DRA probe in the different cell lines is due to the polymorphic location of the NotI sites giving rise to the observed restriction fragments. Rather there must be differences in the amounts of DNA present within the NotI fragment that hybridised to DOB, DQA, DRA and 21-OHase probes in the different cell lines. Since the class III region which is accessible to cloned probes has already been excluded as the source of this variation (see above) the class II region was analysed using a number of other restriction enzymes (Figs 5.2, 5.3, 5.6 and 5.7, Table 5.2).

5.3.4b Mapping of the B8 DR3 haplotype (cell line 2) allows accurate location of the class II loci

It is constructive to consider the mapping results from cell line 2 alone initially, since these data allow placement of many of the class II genes. Hybridisation of DRA, DRB and DQA probes to MluI-digested DNA identified a common 650 kb fragment containing these genes (Fig 5.3). In a PvuI digest (Fig 5.6) the DRA probe hybridised to 330 kb and 390 kb fragments, and to a much fainter 290 kb fragment. The DRB probe also hybridised to the 330 kb and 390 kb PvuI fragments, but also to 60 kb and 200 kb fragments as well as to a very much fainter 40 kb PvuI fragment. Therefore the DRA and one DRB gene sequence lie on a common 330 kb PvuI fragment and are separated by a very partially cleaved PvuI site to give rise to the 290 kb and 40 kb fragments seen with DRA and DRB, respectively. The 60 kb PvuI fragment seen with the DRB probe is presumably a product of partial digestion of the 390 kb PvuI fragment to yield the 330 kb and 60 kb fragments. The 200 kb PvuI fragment was also seen on hybridisation of PvuI digested DNA with DQA locating DRB and DQA sequences within 200 kb of each other. Hybridisation of DRA and DRB to BssHII digested DNA identified a common 460 kb fragment (Fig 5.7). This 460 kb BssHII fragment was also partially digested to separate the DRA and DRB sequences in 270 kb and 200 kb fragments, respectively. The telomeric end points of both the BssHII and PvuI fragments can be tentatively placed at ∿130 kb from the 3' end of the C4 gene on the basis of previous mapping data (Dunham et al., 1987) and the internal consistency of the maps. This would suggest that the maximum distance between the DRA and 21-OHaseB genes is 390 kb. Additionally at least 3 DRB gene sequences must be present on this haplotype split by PvuI sites as indicated in Fig 5.2. Comparison of these data with cloned cosmid DNA

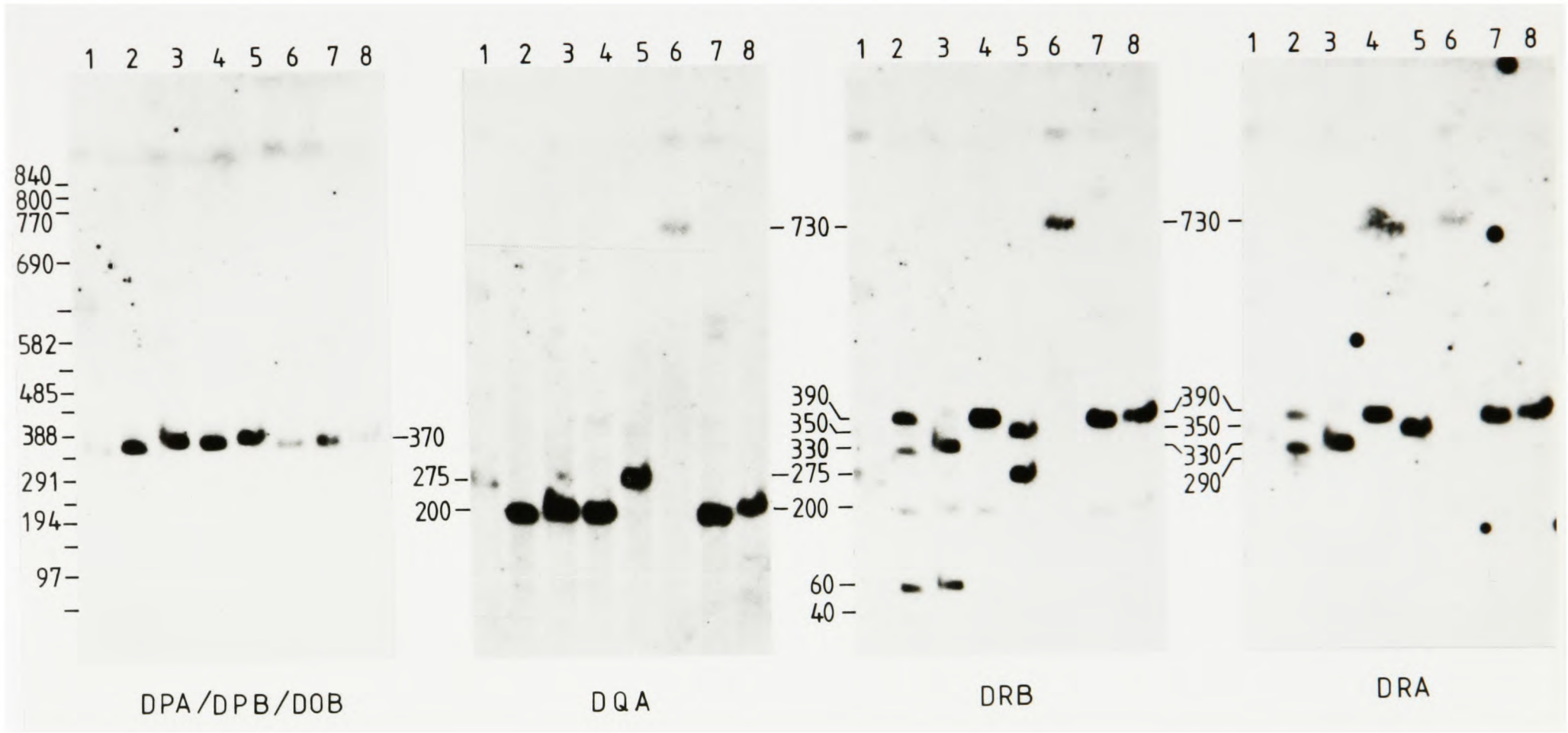


Fig 5.6 Genomic Southern blot analysis of PvuI-digested DNA from cell lines 1-8 separated by PFGE at a 43.5 s switching interval. Panels show the autoradiographs of successive hybridisation of the same filter with the probes indicated. The positions of yeast chromosome and λ concatemer markers are shown on the left. Fragment sizes are in kb.

covering this region in DR3 and DR2 haplotypes that have been mapped for PvuI and BssHII (Rollini et al., 1985; Inoko et al., 1988) confirms these conclusions and excludes the possibility that these sites are contained within the DRB genes.

The position of the DQA gene can only be limited to within the 200 kb PvuI fragment to which the DQA probe hybridised. The DOB gene must lie within a 50 kb segment of DNA bounded on the centromeric side by NotI and BssHII sites and on the telomeric side by MluI and PvuI sites. The DPA probe hybridised to MluI fragments of both 90 kb and 280 kb. The 280 kb fragment was in common with the DOB probe (from PvuI/MluI double digest data) which also hybridised to a 170 kb MluI fragment. This suggests that the two sequences on the common 280 kb fragment are separated by a partially digested MluI site. The DPB probe hybridised to a separate 110 kb MluI fragment. The probable explanation for the DPA and DPB probes hybridising with separate MluI fragments has been discussed in Chapter 3.

5.3.4c The DR3, DR5 and DR6 haplotypes have the same DNA organisation

The data from all the single and double digests with all the probes used were consistent with the map shown in Fig 5.2 for cell line 2. The information obtained for the other cell lines was compared with this map and the results for cell lines 3, 4, 7 and 8 were found to be consistent with this DNA organisation. It was found, however, that the restriction patterns were not identical. This was particularly surprising for cell lines 2, 3 and 4 which all possess the DR3 haplotypes, and in the case of 2 and 3 on the same extended haplotype (B8 DR3 SCO1). This could be due either to the presence or absence of restriction enzyme sites or to methylation differences at the

recognition sequences. An example is the fragment pattern when cell line DNA digested with PvuI was hybridised with DR region probes (Fig 5.6). The situation in cell line 2 has been described above. In cell lines 4, 7 and 8 the DRA probe hybridised only to the 390 kb PvuI fragment whereas the DRB probe hybridised to both the 390 kb fragment and a 200 kb fragment in common with DQA. This is consistent with the two partially cleaved PvuI sites in cell line 2 not being restricted at all in cell lines 4, 7 and 8. In cell line 3 the DRA probe hybridised only to a 330 kb PvuI fragment while the DRB probe hybridised to PvuI fragments of 330 kb, 200 kb and 60 kb. This suggests that the PvuI site responsible for producing the 390 kb partial digest product in cell line 2, was restricted to completion in cell line 3. Further analysis of related haplotypes would be necessary to determine the significance of these observations.

In a similar way to that described above it was possible to rationalise all the data for all the restriction enzymes presented in Table 5.2 for cell lines 2, 3, 4, 7 and 8, providing strong evidence for conservation of the DNA structure in the class II region in these haplotypes (DR3, 5 and 6).

5.3.4d Evidence for more DNA in the DR2 and DR4 haplotypes

For the two DR2 haplotypes, cell lines 1 and 5, it had previously been noted that the DRA hybridising NotI fragments were ~30 kb larger than was observed with cell lines 2, 3, 4, 7 and 8 when the number of C4 genes present was taken into account. Similarly the DR4 cell line, cell line 6, had a NotI fragment that was at least 70 kb larger than the other cell lines. It also was observed that the restriction fragment sizes and maps for these cell lines with all enzymes using the DOB, DPA

and DPB probes were consistent with those seen with the other cell lines. Therefore it can be concluded that cell lines 1 and 5 possess ∿20-30 kb more DNA and cell line 6 ∿100 kb more DNA than the other cell lines in the region between DOB and the limits of the class III region analysed above.

Hybridisation of the DRA probe to BssHII-digested DNA identified a 270 kb fragment (in some cases a partial product) in common between all cell lines except cell line 1 (Fig 5.7). This fragment extends from between the DRA and DRB genes towards the class III region, ending adjacent to the NruI site which defined the limits of the class III structure analysed. Therefore, the differences in DNA content between the cell lines must lie between the DOB and DRA genes. For cell lines 2 and 5 hybridisation of DRB to PvuI-digested DNA identified 350 kb and 275 kb PvuI fragments compared to the 330 kb, 390 kb and 200 kb fragments observed in cell line 3 (Fig 5.6). This difference can be explained by the presence of ∿30 kb more DNA present within the limits bounded by the BssHII site and PvuI site in the DRB gene region in cell line 5 as indicated on Fig 5.2. Allowing for this extra DNA the maps of cell lines 1 and 5 can be completely reconciled with those of cell line 2, 3, 4, 7 and 8. By analogy with cell line 5, the position of the presumed difference in DNA organisation in cell line 1 is also indicated in Fig 5.2.

In the case of cell line 6 carrying the DR4 haplotype it can be concluded that there is 100-130 kb more DNA present in the region including the DRB and DQA genes. However because of the lack of suitable cleavable restriction sites in this cell line the position of the extra DNA cannot be defined further than within the ∿450 kb region bounded by the BssHII site centromeric of the DRA gene and the PvuI and MluI sites centromeric to the DQA gene. Accounting for this major difference in DNA

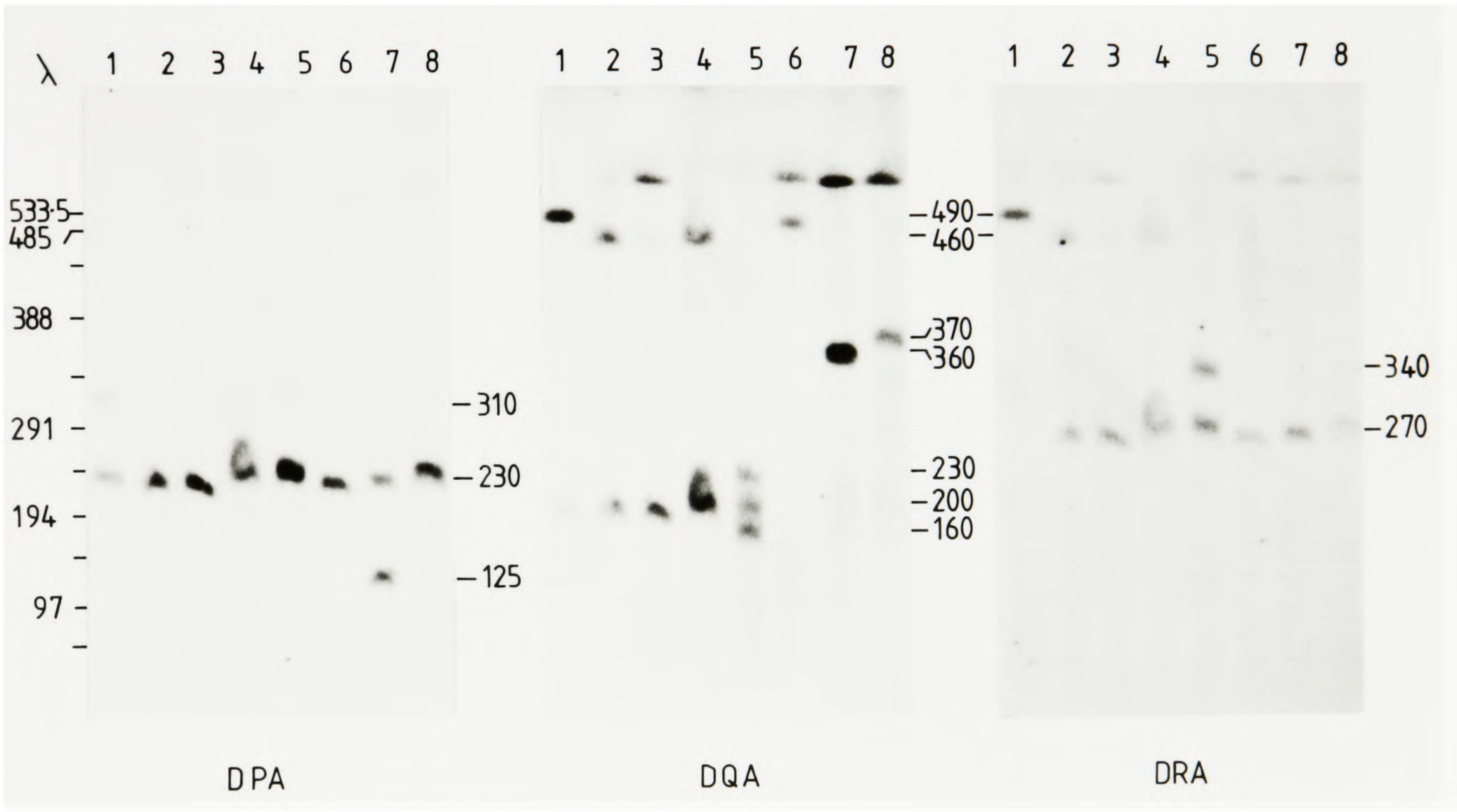


Fig 5.7 Genomic Southern blot analysis of BssHII-digested DNA of cell lines 1-8 separated by PFGE at a 25 s switching interval. The panels show autoradiographs after hybridisation of the same Southern blot filter sequentially with the probes shown. The positions of the λ concatemer markers are indicated on the left, fragment sizes are in kb.

structure allows the map of the DR4 haplotype (cell line 6) to be reconciled with the other cell line maps (Fig 5.2).

5.4 DISCUSSION

Using PFGE, rarely cutting restriction endonucleases and Southern blotting, the DNA organisation of the class II and class III regions of 8 HLA homozygous cell lines possessing 7 common HLA haplotypes has been compared. Analysis of the human MHC class III region with a range of rare-cutting restriction enzymes suggested that the DNA organisation from the DRA gene to the HLA-B and C genes is remarkably conserved between these haplotypes. Further observations at finer resolution indicated that over a 450 kb region of class III DNA including the genes for C2, factor B, C4, 21-OHase, RD and HSP70 there are no differences in the DNA organisation between haplotypes apart from the known differences in C4 and 21-OHase gene number (Carroll et al., 1985b; Schneider et al., 1986). The resolution of the PFGE technique in the fragment size range studied was sufficient to be be able to accurately detect fragment size differences of 2 kb and over.

Comparative analysis of restriction site mapping data in the human MHC class II region suggested that the DNA organisation ir cell lines possessing DR3, DR5 and DR6 haplotypes was the same. However, in two cell lines possessing the DR2 haplotype evidence was obtained for the presence of an extra 20-30 kb of DNA in the DRB region when compared to the DR 3, 5 and 6 haplotypes. Similarly the DR4 haplotype appeared to have a large additional segment of DNA (100-130 kb) compared to the DR3, 5 and 6 haplotypes. Because of the large size of the informative restriction enzyme fragments in the DR4 haplotype, the location of the difference could only be refined to within a ∿450 kb region containing

the DQA and DRB genes. These haplotype-specific differences could have some bearing both on the analysis of disease susceptibility and on the ability of chromosomes bearing different HLA haplotypes to recombine within the DQ/DR subregions.

The apparent conservation of the DNA structure in the class III region DNA outside of the C4 and 21-OHase loci is striking. It is also striking that of the 5 rare-cutting restriction enzymes studied over a region covering >1Mb of class III DNA, only one example of a possible RFLP was found (the BssHII fragment size observed with probe E or P). Evidence from cell line 1 where the BssHII site in question was found to be present in cloned DNA but not cleaved in genomic DNA suggest that this difference is due to methylation rather than presence or absence of the BssHII site. Other workers have indicated that there may be "deletions" of DNA (10-20 kb) within the class III region in B8 DR3 and B18 DR3 haplotypes, contained within the PvuI fragment that hybridised to the 21-OHase probe (Tokunaga et al., 1988). Similarly they predicted a 40 kb deletion telomeric to the TNFA and B genes. The data presented here including two cell lines with the B8 DR3 haplotype and one B18 DR3 haplotype do not support these conclusions. It must be assumed that the deletions observed by these authors were specific to their cell lines or that they have inadequately considered the problems of accurate comparative sizing of restriction fragments from different DNA samples on PFGE gels.

The apparent conservation of DNA organisation and restriction sites within the class III region in contrast to the frequency of deletion and duplication at the C4 and 21-OHase loci (Carroll et al., 1985b) is at first site contradictory. However, the absence of RFLPs for the rare-cutting restriction enzymes is not surprising. It is known that many of the rare-cutting restriction enzyme sites that are cleaved in

genomic DNA are located in unmethylated, CpG-rich islands since cleavage by these enzymes is sensitive to methylation at the CpG dinucleotides within their recognition sites (Brown & Bird, 1986). Data from comparison of restriction sites present in cloned cosmid DNA with those cleaved in genomic DNA suggests that nearly all the sites for rarely cutting enzymes that are cleaved in cell line 1, and therefore are useful for the mapping presented here, are located in such islands (Sargent, 1988; Chapter 6). Since they are unmethylated it is unlikely that the endonuclease recognition sequences will be lost, because it is thought that the principle mutational force on CpG pairs is deamination of the methylated cytosine (Bird, 1986; Cooper & Yousouffian, 1988). Therefore the frequency of differences in the location of rare-cutting restriction enzyme sites in a region such as this which is rich in CpG islands will probably be low. However, fragment size differences due to insertion or deletions of DNA such as for the C4 and 21-OHase genes will be observed (see Chapter 4). This situation is in contrast to the case of enzymes such as TaqI which contain CpG in their recognition sites but are not affected by cytosine methylation. These enzymes cut genomic DNA frequently and are likely to detect single base changes frequently because of the high frequency of mutation at methylated CpG (Barker, Schäfer & White, 1984; Cooper et al., 1985).

Many of the CpG-rich islands found in the cloned class III cosmid DNA are, as expected (Lindsay & Bird, 1987; Bird, 1986), associated with RNA transcripts that are present in all the tissues that have been examined so far (Sargent, 1988). The implies the existence of many "housekeeping" genes (Bird, 1986) across the class III region of the human MHC. The presence of these genes would have considerable influence on the stability of the class III DNA structure since recombination within these putative genes or deletion of them would be likely to be

deleterious. These two observations of stability in the class III DNA structure and the presence of multiple "housekeeping" coding sequences within this region are intrinsically linked. Reduced recombination within the class III region because of its deleterious nature might explain the observations of extended haplotypes and linkage disequilibrium within the MHC. Although there is considerable data on recombination events within the class III region from family studies, the results have been contradictory (Robson & Lamm, 1984; Lamm & Olaisen, 1985). No recombination events that separate the complement genes have been observed. The availability of cloned DNA from across the class III region should enable recombinations that have occurred in families to be studied and the precise location of the event to be determined.

Obviously the C4 and 21-OHase loci are exceptions to the genomic stability of the class III region and this may be due to the presence of a large region (~30 kb) of highly homologous, tandemly repeated sequence at these loci.

In the class II region the results indicate that the DR2 and DR4 haplotypes are different from the DR3, DR5 and DR6 haplotypes due to the possession of ~30 kb and 100-130 kb of extra DNA, respectively. In addition there are a number of variations in the positions of cleavable restriction sites between even the related cell lines which might be due to methylation or DNA sequence differences. The possibility that the differences in class II gene structure were caused by the process of immortalisation of the cell lines can be excluded for two reasons. First, in the case of the DR2 haplotype, two independently derived cell lines possess the extra 20-30 kb of DNA. Second, in a number of other studies of the class II gene organisation a number of DR haplotypes have been studied. Tokunaga *et al*. (1988) have compared restriction fragment

sizes by PFGE between haplotypes possessing the DR2 and DR3 specificities and found evidence for more DNA (they estimate 40-50 kb) in the DR2 haplotype. Inoko et al. (1988) have also analysed a DR2 haplotype and the results are consistent with the organisation presented here. Hardy et al. (1986) established the first map of the class II region of the human MHC using a cell line with the DR4 haplotype and the results presented here with a different DR4 haplotype-containing cell line are in agreement with their map. In addition, a number of groups have studied the DR7 haplotype (Tokunaga et al., 1988; Inoko et al., 1988) which is related to the DR4 haplotype by virtue of sharing the DRw53 supertypic specificity (Böhme et al., 1985; Gregersen et al., 1986; Bell et al., 1987a) and have obtained results consistent with 100-150 kb more DNA present in the region containing the DRB and DQA genes. Therefore, there appears to be consistency in the observations of the amount of DNA present in cell lines possessing the same haplotype. These observations also suggest that the inconsistencies between published maps of the human MHC class II region are due to haplotype specific differences as well as to differences in the estimations of the size of restriction fragments.

The nature of the alterations in DNA organisation observed between haplotypes is open to question. The DR3, DR5 and DR6 haplotypes have previously been shown to be related in the DRB region since they share common RFLPs (Böhme et al., 1985; Bell et al., 1987a) whereas the DR2 and DR4 haplotypes have distinct DRB restriction fragment patterns. It is therefore possible that there could be size variation between the haplotypes because of differences in the DRB gene organisation or because of the presence of different numbers of duplicated DRB genes. This would seem particularly likely in the DR2 case where the difference of 20-30 kb is localized to the DRB gene region. Efforts have been made

to estimate the number of DRB genes present on the different DR haplotypes, but because of the extensive similarity between the genes, two or more genes may reside on restriction fragments of the same length and therefore cannot be resolved electrophoretically (Böhme et al., 1985; Bell et al., 1987a). Böhme et al. (1985) have estimated that there are at least three DRB genes on each of the haplotypes DR2-DR7 while the DR4 cell line was shown to possess four DRB gene sequences by cosmid cloning (Spies et al., 1985; Böhme et al., 1985). However, further comparison of cloned cosmids covering the DR region from different haplotypes will be necessary to confirm how the DNA organisation varies.

A further possibility for the differences is that there are variations in the amount of uncharted DNA present in the HLA-D region between haplotypes. This would seem likely in the case of the DR4 haplotype because an alteration of 100-130 kb could not be caused by the presence of an extra DRB gene and the organisation of the DRB genes in the cloned cosmid DNA is not consistent with such an insertion (Spies et al., 1985; Hardy et al., 1986). The results described here place the extra DNA in the region containing the DQA and DRB genes. Comparison of these results with the map of a DR4 haplotype constructed by Hardy et al. (1986) suggests the extra DNA must lie between the DQA and DRA genes. In a related DR7 cell line, Inoko et al. (1988) predict the presence of extra DNA between DQA and DRB1 when compared to a DR2 haplotype. Therefore, there appears to be 100-130 kb of DNA which does not contain class II-like sequences inserted between DQA and DRB1 in the DR4 (and DR7) haplotypes. Whether this DNA contains any expressed sequences can only be determined by cloning of this segment of DNA.

The observation of these differences may be of significance for the maintenance of linkage disequilibrium within the DR and DQ subregions and for the lack of recombination events between DR and DQ

(see Chapter 1). It appears that supertypic DR groups (e.g. DR3, DR5 and DR6; DR4 and DR7) have maintained the same long range genomic organisation. It seems reasonable to propose that these organisations have been maintained because the differences in size of the HLA-D region favour recombination between related haplotypes rather than between haplotypes with different structures. This may go some way to preserving ancestral haplotypes in the HLA-DQ and DR subregions.

The long range structure of the HLA class I region in different haplotypes has also been examined using PFGE (Chimini et al., 1988) and very little structural polymorphism was observed. This suggests that both the class III region and the regions around the B, C and A loci are relatively constant in their genomic organisation between haplotypes.

The data that has been presented here should provide a significant resource in the study of HLA disease association. It should now be possible to compare the organisation of uncloned DNA in the human MHC between patients and controls for affected haplotypes, in addition to comparison of known loci (Todd et al., 1988a). Observation of deletions or insertions may give some indication of whether linked, as yet unidentified, genes play any role in HLA-associated diseases.

CHAPTER 6

CHARACTERISATION OF CpG-RICH ISLANDS WITHIN THE CLASS III REGION OF THE HUMAN MHC

6.1 INTRODUCTION

Several lines of evidence suggest that as yet undiscovered genes may lie in the human MHC class III region. First, PFGE mapping studies show that there are large regions of uncharted DNA lying between the class II and class I regions and the complement gene cluster (See Chapter 3). The sizes of these DNA tracts are constant between different HLA haplotypes and also are apparently conserved between mouse and man.

Second, a number of precedents for new genes within the class III region have recently been described. Genes for the cytochrome P-450 steroid 21-hydroxylase were mapped to the cosmid cluster containing C2, C4 and factor B (Carroll et al., 1985a). The genes encoding tumour necrosis factors α and β have been mapped to the MHCs of man (Spies et al., 1986) and mouse (Müller et al., 1987a) and were positioned between the complement gene cluster and the class I genes (Dunham et al., 1987; Carroll et al., 1987; Inoko & Trowsdale, 1987; Ragoussis et al., 1988; Müller et al., 1987a). Levi-Strauss et al. (1988) have located a new gene which they call RD between the C4 and factor B genes. Tsuge et al. (1987) have mapped a B cell and macrophage specific transcript close to the TNFA gene. Sargent et al. (1988) have mapped genes for the major heat shock protein, HSP70, 92 kb telomeric of the C2 gene.

Third, although direct roles for known MHC genes have been proposed for the observed association of susceptibility to certain diseases with particular MHC haplotypes, the relationships are still

uncertain. It may be that other gene products within the MHC play significant roles in at least some diseases. Finally, the PFGE map of the human MHC suggests there may be a number of clusters of sites for rare-cutting restriction enzymes. As discussed below these clusters may be CpG-rich "islands" which are frequently associated with genes.

A number of methods have been described that allow the detection of coding sequences within cloned DNA. Unique sequence probes from cosmid clones could be used to probe Northern blots. Alternatively, [^{32}P]-labelled cDNA has been hybridised to cosmid clones to identify putative coding sequences (Trowsdale et al., 1984a; Guillemot et al., 1988) but this method appeared to favour identification of low-copy repeated elements. In certain cases it may be possible to genetically map cloned transcripts to a region of interest and then identify the precise position by hybridisation to cloned genomic DNA or PFGE blots. This approach has proved successful in the case of the B144 cDNA clone (Tsuge et al., 1987) and the TNF genes (Müller et al., 1987a; Dunham et al., 1987).

A more generally applicable strategy has been described by Monaco et al. (1986). In this method it is assumed that coding sequences are likely to be evolutionarily conserved and therefore single-copy probes that contain exons should hybridise at high stringency to DNA from other animal species on Southern blots ("zoo blots"). This approach has been successfully used to locate the coding region of the Duchenne muscular dystrophy gene (Monaco et al., 1986) and to map the novel RD gene into the complement gene cluster in the mouse and human MHCs (Levi-Strauss et al., 1988).

When a large region of cloned DNA is to be screened for coding sequences, it is useful to be able to rapidly identify the positions of candidate sequences. Bird (1987) has proposed that the positions of many

genes could be identified by making use of the presence of CpG-rich islands at their 5' ends. In bulk vertebrate DNA the frequency of the dinucleotide CpG is suppressed to 0.2-0.25 of the frequency expected from the base composition (Josse et al., 1961; Swartz et al., 1962) and 60-90% of the CpG is methylated at the 5 position on the cytosine ring. However, a discrete fraction of the genome (1%), which is extensively cleaved with the methylation sensitive enzyme HpaII (recognition sequence CCGG), contains 15% of the total number of CpG dinucleotides and is unmethylated (Cooper et al., 1983). The sequences from this "HTF" fraction belong to islands of DNA, 500 bp to at least 2 kb long, with a C+G content of >50% and no CpG suppression so that the dinucleotide occurs at the frequency expected from the base composition (Bird et al., 1985). The majority of island sequences are "unique" by the criterion of DNA reassociation.

Sequence analysis of a number of genes has shown that CpG-rich sequences occur at the 5' ends of all "housekeeping" and several tissue specific genes (Tykocinski & Max, 1984; Gardiner-Garden & Frommer, 1987), but only in a few cases has the methylation status of these genes been determined (see Bird, 1986 for review). A number of examples of CpG-rich islands at the 3' ends of genes have been described (Toniolo et al., 1984; Fischel-Ghodsian, Nicholls & Higgs, 1987a), but it now seems likely that these islands are associated with the 5' ends of previously unidentified genes (Bird, 1987). It has been estimated that there may be 30 000 such islands in the mammalian genome and equating this with the number of genes suggest that a high proportion of islands will be gene associated (Brown & Bird, 1986). Support for this proposition comes from analysis of randomly picked islands in the mouse (Lavia, Macleod & Bird, 1987) and from the human X chromosome (Lindsay & Bird, 1987), with five of the seven islands studied detecting transcripts.

In order to detect the positions of islands within the genome, Brown & Bird (1986) reasoned that restriction enzymes that have CpG dinucleotides within their recognition sites and that are methylation sensitive should cleave genomic DNA predominantly at sites that lie in islands. The sites for these enzymes are not randomly distributed but are clustered at CpG-rich islands which show no CpG suppression and are unmethylated. Lindsay & Bird (1987) further argued that enzymes whose recognition sites comprise only C and G and contain two CpGs should preferentially cleave in islands even in cloned DNA where the advantage of methylation is lost. Most islands should contain sites for one or more of these enzymes. Therefore the positions of coding sequences as marked by islands should be detected by use of these enzymes. Support for this proposal has come from the analysis of randomly selected clones (Lindsay & Bird, 1987) and from more directed searches for genes in specific chromosomal locations (Estivill et al., 1987; Rappold et al., 1987).

6.2 EXPERIMENTAL STRATEGY

Carole Sargent has cloned 541 kb of genomic DNA from the class III region of the human MHC in two clusters of overlapping cosmid clones (Sargent, 1988). In order to identify CpG-rich islands within this cluster of cosmids two complementary approaches were adopted. The cosmid clones were mapped for the restriction enzymes BssHII, EagI and SacII which have 6 bp recognition sequences comprised entirely of C and G and contain two CpGs. 74% of the sites for these enzymes should occur in islands (Lindsay & Bird, 1987). In addition the positions of sites that are cleaved in genomic DNA were determined by separating restricted genomic DNA on short switching interval PFGE gels, transferring the DNA

to nylon membranes and hybridising with a series of probes whose positions in the cloned DNA was known. From a combination of the information from the cosmid restriction map, the location of the probe and the size of the observed genomic restriction fragment, the position of sites that are unmethylated could be determined. In this way clusters of sites for these enzymes in the cloned DNA were identified and the methylation status of the putative islands determined in genomic DNA from lymphoblastoid cells.

One of the identified islands has been further characterised by sequence analysis and has been tested for the presence of associated transcripts.

6.3 RESULTS

6.3.1 Mapping of rare-cutting restriction enzyme clusters in the cloned cosmid DNA from the class III region

Sites for the enzymes BssHII, EagI and SacII and other rare-cutting enzymes were mapped by Carole Sargent in the cosmid inserts from the two clusters of overlapping clones covering 541 kb of the class III region. Each enzyme was used singly and in double digest combination with BamHI and BglII, the positions of the sites for which were already known. A surprisingly large number of sites were found for BssHII, EagI and SacII within the cloned DNA as illustrated in Fig 6.1 and Table 6.1. At least 33 BssHII, 43 EagI and 31 SacII sites were identified in 541 kb, occurring much more frequently than predicted from calculations based on the bulk DNA dinucleotide frequencies (Dramanac et al., 1986). Furthermore many of these sites were clustered, with two or more sites occurring within less than 1 kb (Fig 6.1).

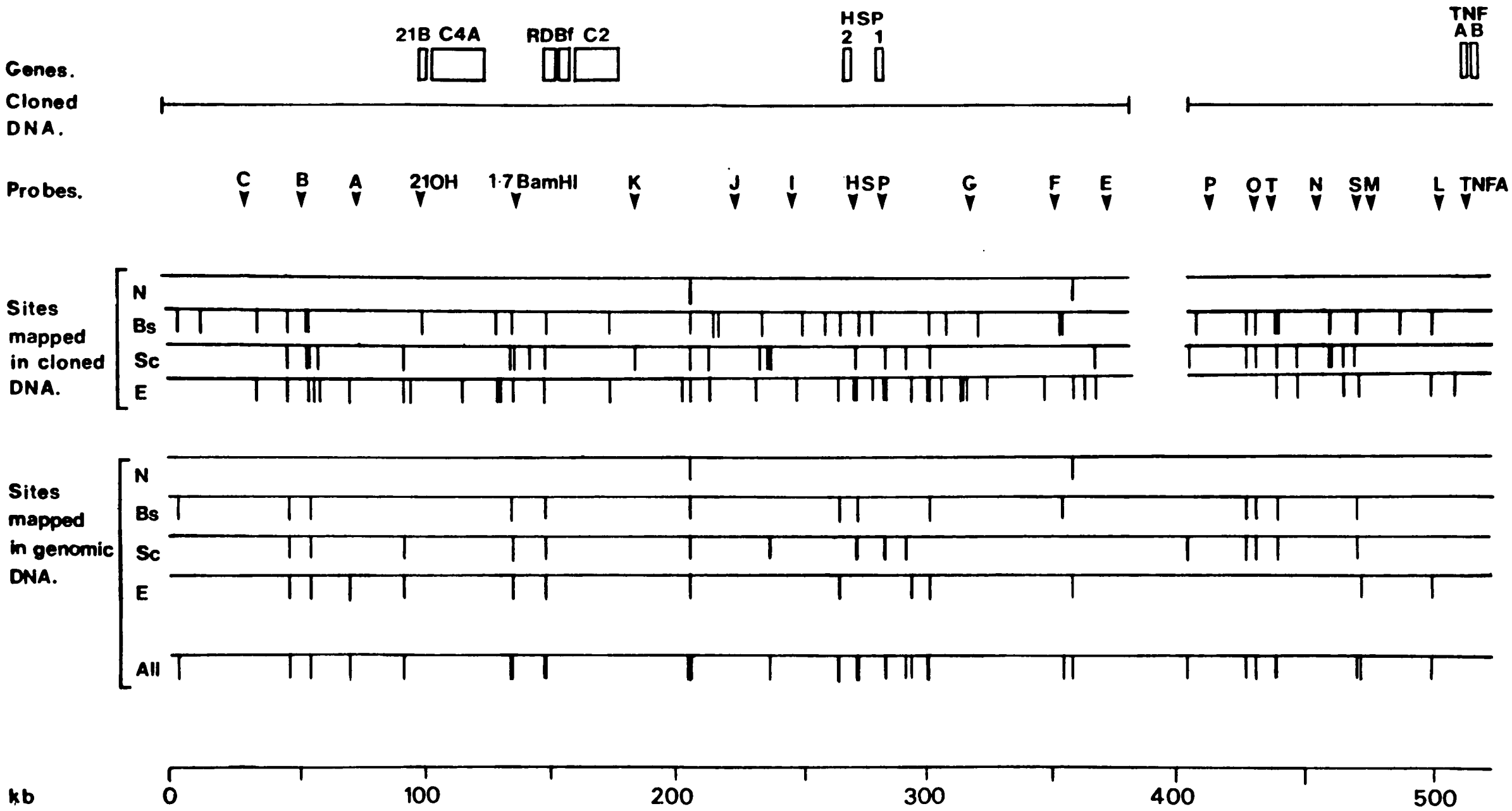


Fig 6.1 Rare cutting restriction enzyme sites within the class III cloned cosmid DNA up to the TNF genes. Open boxes are genes. The top line gives the extent of the cloned cosmid DNA. The positions of probes used in the analysis is shown by arrowheads. Below this are shown the positions of sites for NotI (N), BssHII (Bs), SacII (Sc) and EagI (E) mapped in the cosmid DNA. The bottom map shows the positions of sites that are cut in genomic DNA from the lymphoblastoid cell line Ice 5. "All" indicates the positions of all sites that are cleaved in genomic DNA. Vertical lines mark each restriction site.

Enzyme	*Recognition sequence*	*N*	*F(kb)*	*E(kb)*	*E/F*
NotI	*GCGGCCGC*	*2*	*270.5*	*3 000*	*11.1*
NruI	*TCGCGA*	*3*	*180.3*	*100.5*	*0.6*
PvuI	*CGATCG*	*3*	*180.3*	*126.5*	*0.7*
MluI	*ACGCGT*	*8*	*67.6*	*132.0*	*2.0*
BssHII	*GCGCGC*	*33*	*16.4*	*189.0*	*11.5*
EagII	*CGGCCG*	*43*	*12.6*	*148.5*	*11.8*
SacII	*CCGCGG*	*31*	*17.5*	*149.0*	*8.5*

Table 6.1 Analysis of the frequency of rare-cutting restriction enzyme sites in 541 kb of cloned cosmid DNA from the class III region (Taken from Sargent, 1988). N = The number of sites mapped in the cloned DNA on the basis of restriction enzyme digestion. There may be more sites present as small fragments (less than 0.4 kb) were not detected. F = the mean distance between sites i.e. 541/N where 541 kb = total length of cloned DNA. E = The expected average distance between sites from the bulk DNA dinucleotide frequencies (Dramanac et al., 1986).

6.3.2 Mapping of "rare-cutter clusters" in class III region genomic DNA

The enzymes used above should cut predominantly in CpG-rich islands (Lindsay & Bird, 1987). However, since so many sites were found and to simplify the analysis, sites that could be restricted in lymphoblastoid cell line genomic DNA, and were therefore unmethylated, were identified. Genomic DNA in agarose blocks from the lymphoblastoid cell line, Ice 5, was digested with BssHII, EagI and SacII and the digests were separated on PFGE ("Waltzer") gels at a 7.5 s switching interval in order to optimise separation in the 10-200 kb size range. The separated DNA was transferred by alkaline blotting onto Hybond N membranes and then hybridised with probes isolated from the cloned cosmid DNA covering the class III region (the probes are detailed in Fig 6.1 and Table 6.2). Successive probing of the same blots combined with the information on the exact position of the probe allowed the end points of observed fragments to be accurately positioned relative to the map of the cosmids. In some cases because restriction enzyme sites lie close together it was not possible to distinguish which site was cleaved. Although probes were not available for some small fragments,

restriction at two close sites could be inferred from the sizes of the fragments observed with adjacent probes.

Figs 6.2 and 6.3 show the results of sequential probing of Southern blots of BssHII and SacII digested DNA. The sizes of the BssHII and SacII fragments and also the EagI restriction fragments observed are shown in Table 6.2. The map of the sites cleaved in genomic DNA derived from these data is shown in Fig 6.1. In the complement gene cluster (Fig 6.2), probe K and the 1.7 kb BamHI probe hybridised to the same sized fragments with each enzyme, 55 kb and 12 kb respectively. The 21-OHase probe hybridised to 40 kb EagI and SacII fragments, but to an 80 kb BssHII fragment. From the sizes of the fragments observed with each enzyme, the positions of the probes and inspection of the cosmid map, it can be deduced that there are two clusters of unmethylated rare-cutter sites lying between the C4 and factor B genes. One cluster is located about 10 kb 5' to the C4A gene and the other is 6-7 kb 3' of the factor B gene. Another striking cluster of sites including MluI and NotI sites that are known to cut in genomic DNA (see Chapter 3) lies 27 kb telomeric of the C2 gene. Centromeric of the 21-OHase gene, about 8 kb from its 3' end, lie cuttable EagI and SacII sites, while the BssHII site that is cleaved is located in another cluster of sites 80 kb from the 21-OHase gene.

Interestingly, although there are two clusters of unmethylated rare-cutting sites, and therefore two CpG-rich islands, within the complement gene cluster neither island appears to be associated with the complement or 21-OHase genes. Analysis of the distribution of CpG and GpC dinucleotides for the sequenced regions of the tissue specific C2, C4, factor B and 21-OHase genes including the 5' ends showed that these genes are typical "CpG-depleted" genes (Gardiner-Garden & Frommer, 1987) and possess no CpG-rich islands (results not shown). Therefore, these

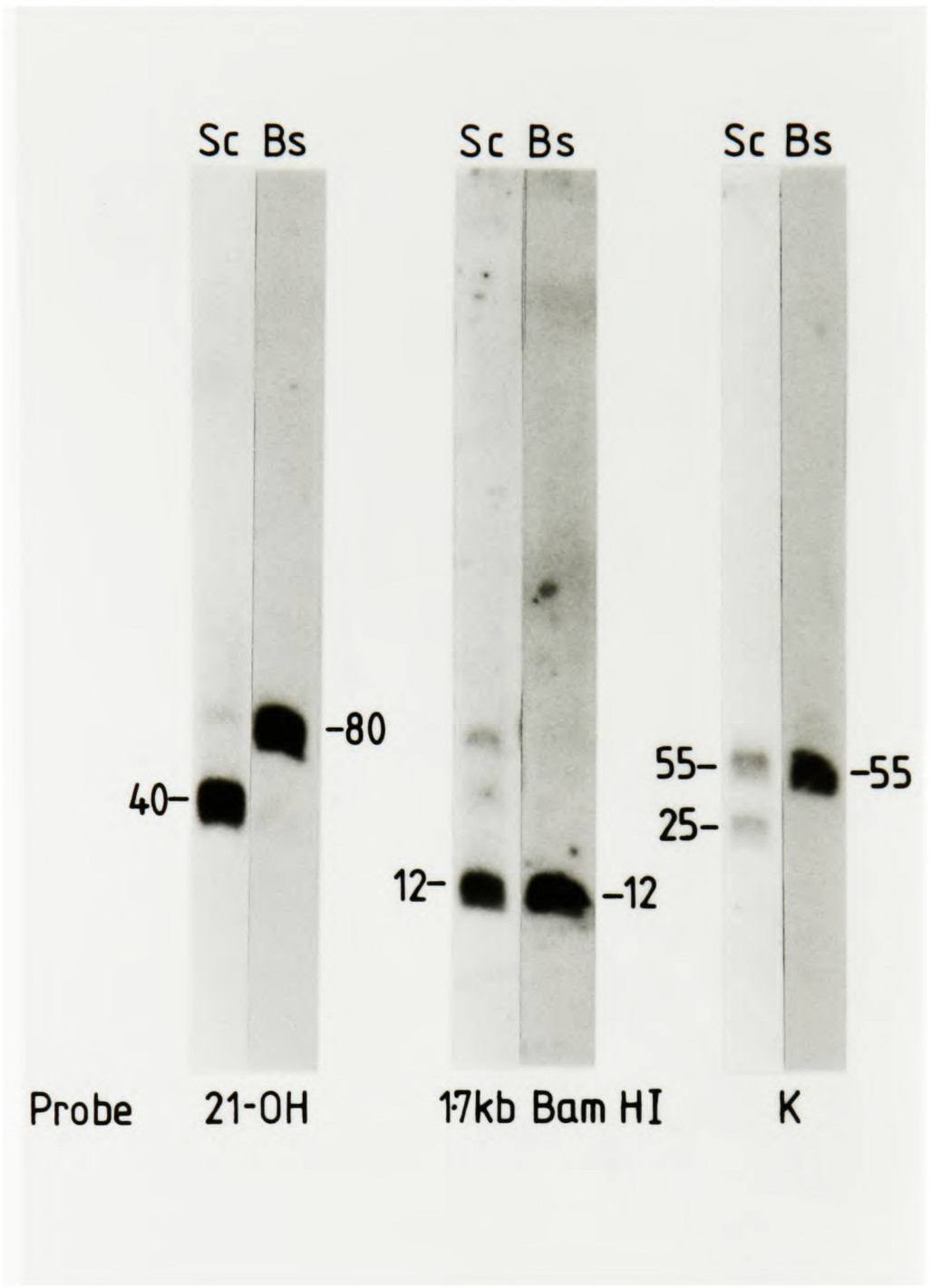


Fig 6.2 Genomic Southern blot analysis of BssHII (Bs) and SacII (Sc) digests of Ice 5 high mol. wt. DNA separated by PFGE at a 7.5 s switching interval. Each panel shows the autoradiograph of the same filter hybridised with the probes indicated below the panels. Fragment sizes are indicated in kb.

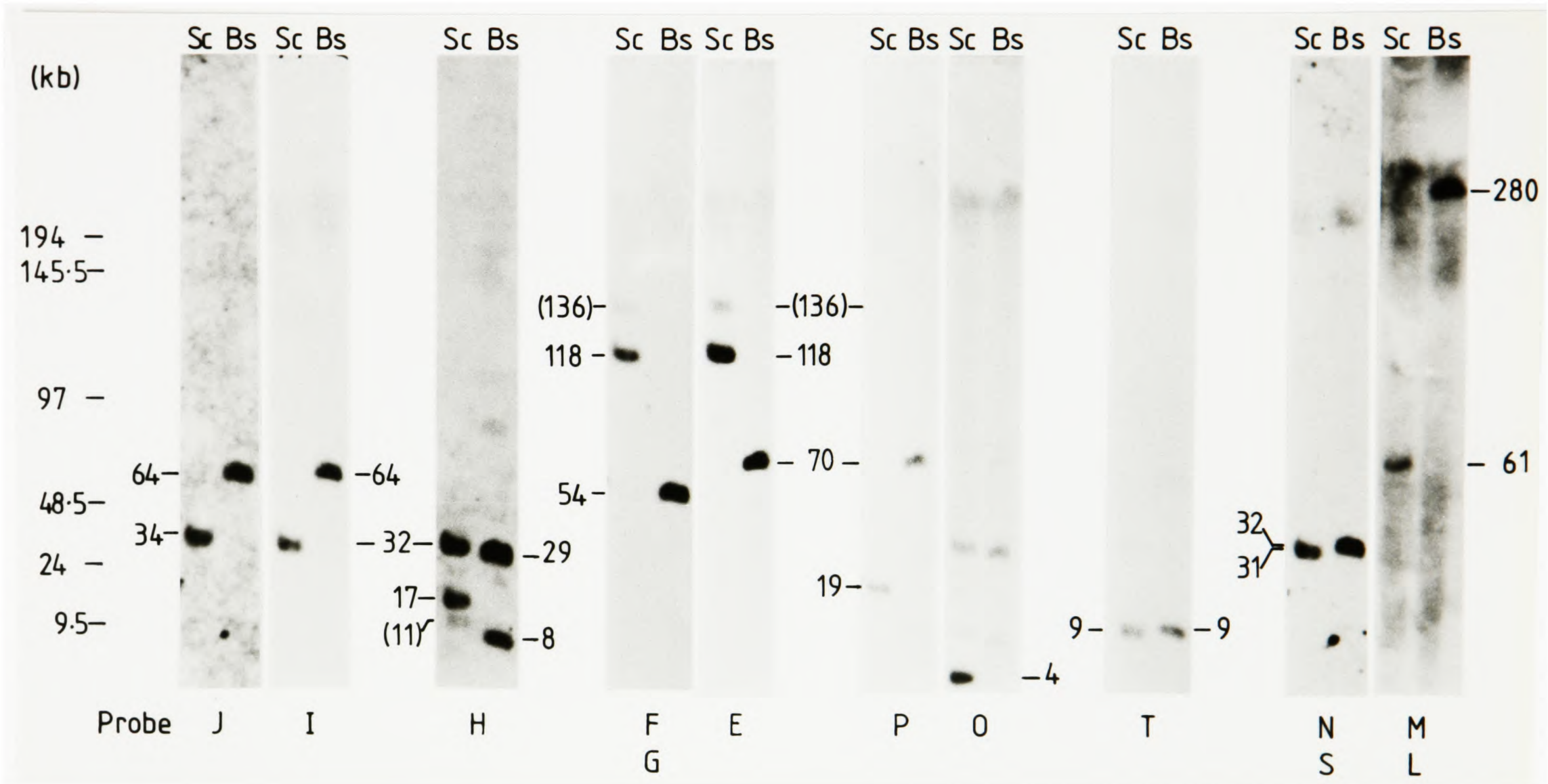


Fig 6.3 Southern blot analysis of BssHII (Bs) and SacII (Sc) digested DNA separated by PFGE at a 7.5 s switching interval. The probes used are indicated below each panel. Where two probes gave the same result only one is shown. The positions of comigrating markers are shown at the left. Sizes of fragment are indicated in kb. Fragments in brackets are the minor products of partial digestion at SacII sites. Probe H is from the duplicated HSP genes and hence hybridised to several fragments. Since there also other highly homologous copies of HSP genes in the genome faint fragments are also seen for probe H. For probe O the BssHII fragment has been electrophoresed off the gel. The unmarked fragments in the O panel cannot be mapped to the class III region and must result from homology to some other sequence in the genome.

Table 6.2 Sizes of restriction enzyme fragments observed (kb) after hybridisation of MHC class III region probes to Southern blots of PFGE separated digested DNA. The size predicted for each fragment from inspection of the cosmid map is recorded under "cosmid". ND, not determined, -, no fragment observed, NA, not applicable as part of the fragment is not cloned.

	ENZYME					
	BssHII		*SacII*		*EagI*	
PROBE	*PFGE*	*Cosmid*	*PFGE*	*Cosmid*	*PFGE*	*Cosmid*
C	30	32	ND	NA	ND	NA
B	8	8	8	8	8	8
A	80	82	35	38	20, 35	22, 38
21-OHase	80	82	40	45	40	45
1.7 BamHI	12	13	12	13	12	13
K	55	58	55	58	55	58
J	64	62	34	35	64	62
I	64	62	32	32	64	62
H	29, 8	29, 7	32, 17 (11)	32, 19 (11)	28 (35)	29 (35)
G/F	54	53	118 (136)	114 (133)	60	59
E	70	NA	118 (136)	114 (133)	110	NA
P	70	NA	19 (136)	19 (133)	110	NA
O	-	3.5	<6.6	4.5	110	NA
T	9	8	9	8	110	NA
N/S	32	31	31	31	110	NA
M	280	NA	61	NA	35	38
L	280	NA	61	NA	35	38

genes would not have been detected by a CpG-rich island search of this region. The islands located between C4 and factor B are, in fact, associated with genes. The island 6-7 kb 3' of the factor B gene marks the 5' end of the human RD gene (Levi-Strauss et al., 1988; Sargent, 1988) while the island 10 kb 5' of C4 is at the 5' end of a new gene for which a cDNA has been cloned and sequenced (Sargent, 1988). Both these genes are expressed in all tissues or cell lines that have been examined so far and so probably fall into the category of "housekeeping" genes (Levi-Strauss et al., 1988; Sargent, 1988).

In a similar way the results presented in Fig 6.3 and Table 6.2 give rise to the map of rare-cutter sites shown in Fig 6.1. The size of the gap between the cosmid cluster containing the complement genes and the cosmid cluster containing the TNF genes was estimated on the basis of hybridisation of probes E and P to BssHII and EagI digested DNA. Both probes hybridised to a 70 kb BssHII and a 110 kb EagI fragment. This further demonstrates the linkage of the two cosmid clusters (see Chapter 3) and since the end points of these fragments in each cluster is known the distance separating the cloned DNA is 22 kb.

In Table 6.2, the sizes of the restriction fragments observed on the PFGE blots are compared with the predicted size of the fragments derived from the cosmid restriction map. In general the size estimates from PFGE are remarkably close to the predicted value. Only in the case of the EagI and SacII fragments that hybridised with the 21-OHase probe do the figures differ by more than 4%. It has not been possible to explain the discrepancy for these two fragments. However, overall, it seems that at the switching intervals used the PFGE gels give accurate estimation of fragment size.

Inspection of the map in Fig 6.1 reveals the presence of several putative CpG-rich islands between the C2 and TNF genes. Unique sequence

probes from the immediate vicinity of most of these islands have been hybridised to Northern blots of total RNA from several cell lines and transcripts ranging from 0.6-6 kb have been detected (Sargent, 1988). cDNA clones for a number of these transcripts have been isolated and it appears that this region of DNA contains a number of new genes (Carole Sargent & Duncan Campbell, personal communication). The strategy of identifying CpG-rich islands as gene markers has proved very successful in this region. The implications of these findings are discussed further in section 6.4.

6.3.3 Analysis of a CpG-rich island 27 kb telomeric of the C2 gene

A striking cluster of sites for rare-cutting restriction endonucleases lies 27 kb telomeric of the 5' end of the C2 gene. Sites for NotI (and therefore also EagI), MluI, BssHII and SacII that were cleavable in genomic DNA were found to be located in a single 1.6 kb BamHI fragment in the insert DNA of cosmid M2 (Dunham et al., 1987; gift from Carole Sargent). In order to investigate this region and confirm its identity as a CpG-rich island, the 1.6 kb BamHI fragment was purified from BamHI-digested cosmid M2 DNA by LGT agarose gel electrophoresis (section 2.7.2) and subcloned into BamHI-cut and phosphatased pATX (section 2.9) to give pB16.1. A 1.3 kb BamHI/XhoI fragment located immediately centromeric of the 1.6 kb BamHI fragment was also purified from a BamHI/XhoI digest of cosmid M2 that had been filled-in with dNTPs and was subcloned by blunt-end ligation into PvuII-cut and phosphatased pATX to give construct p1.3MX. The relative positions of these fragments in the class III region and an infrequently cutting restriction enzyme map are shown in Fig 6.4a.

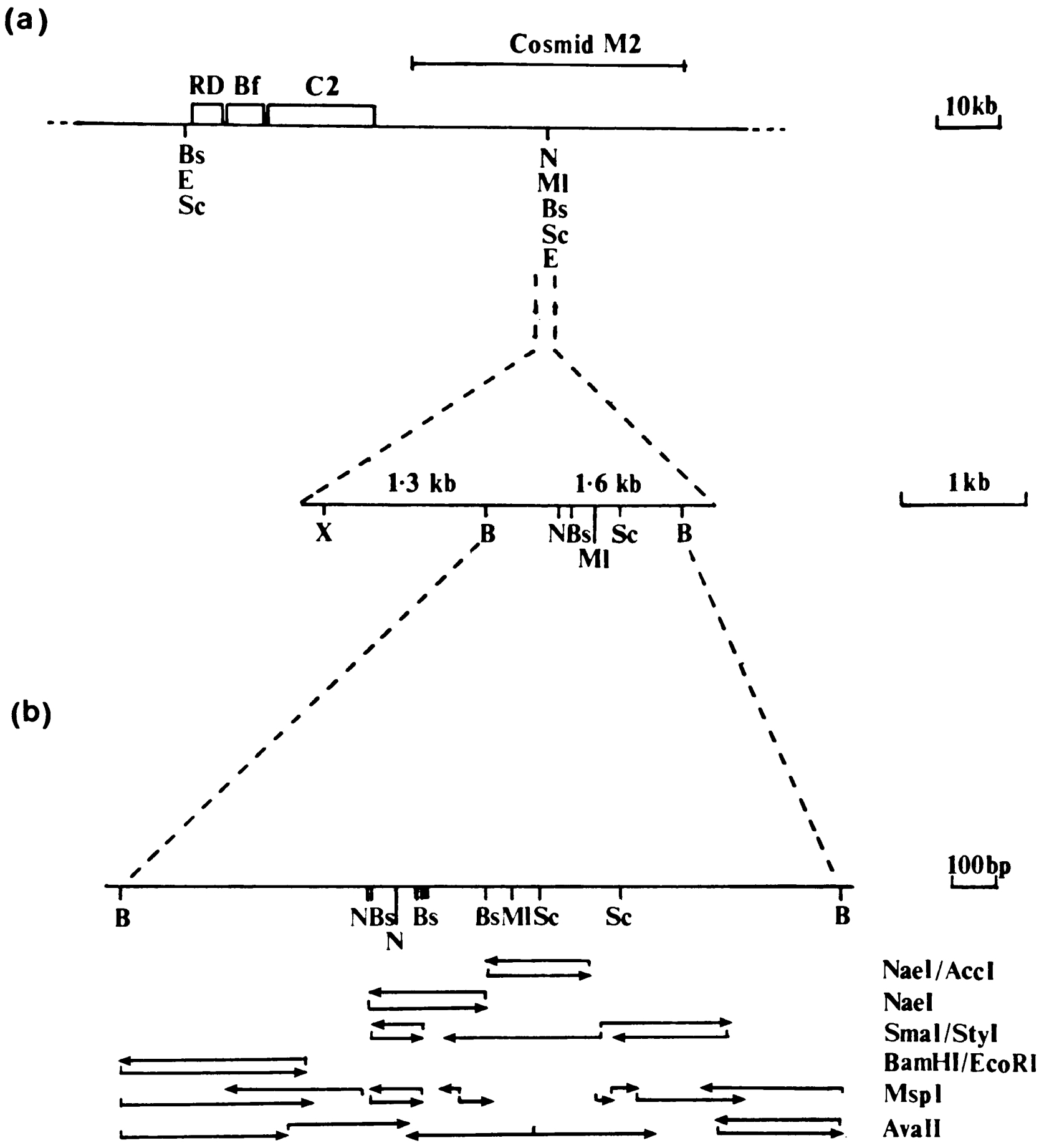


Fig 6.4 (a) The location of the 1.6 kb BamHI (marked 1.6 kb) and 1.3 kb BamHI/XhoI (1.3 kb) fragments relative to the genes in the complement gene cluster. Open boxes are genes, vertical lines are restriction sites (NotI, N; MluI, Ml; BssHII, Bs; SacII, Sc; EagI, E; BamHI, B; XhoI, X). The bar above the line indicates the region of DNA contained in the insert of cosmid M2. (b) Sequencing strategy for the 1.6 kb BamHI fragment. The sequencing strategy is shown beneath the restriction map of the fragment for rare cutters derived from the sequence, enzymes as in (a). Arrows indicate the extent and the orientation in which the fragments designated on the right were sequenced by the dideoxy chain termination method.

6.3.3a Analysis of the extent of CpG dinucleotide methylation

To assess the extent of methylation at CpG dinucleotides in the 1.6 kb BamHI fragment at least in lymphoblastoid cell line DNA, genomic DNA was digested with BamHI in double digest combination with NotI, MluI, HpaII or MspI. Digestion by both NotI and MluI is sensitive to methylation at the CpG dinucleotides in their recognition sites (Lindsay & Bird, 1987). HpaII digests CCGG, but not $C^{m}CGG$ (Bird & Southern, 1978) while its isoschizomer, MspI, will cleave at both these sequences. The digested DNA was separated on a 1.5% agarose gel and transferred onto a GeneScreenplus membrane as described in sections 2.7.2 and 2.8. The blot was hybridised with the 1.6 kb BamHI fragment prepared from subclone pB16.1 and the result is shown in Fig 6.5. In the BamHI alone digest the probe hybridised to a single 1.6 kb BamHI fragment showing that the probe recognises itself and is a single copy sequence. Two fragments were identified in the BamHI/MluI digest, one of 740 bp and one of 860 bp. The single MluI site mapped in the 1.6 kb BamHI fragment is restricted and therefore unmethylated. The probe hybridised to fragments of 980 bp and 540 bp in the BamHI/NotI digest. The sum of the BamHI/NotI fragment sizes is 1520 bp compared to 1600 bp for the sum of the BamHI/MluI fragments. This suggests that in the BamHI/NotI digest a fragment of 80 bp has been lost from the gel during electrophoresis or transfer. In fact the sequence of the 1.6 kb BamHI fragment (see below) revealed that there are two NotI sites separated by about 70 bp. Therefore both sites are unmethylated and are cleaved in genomic DNA. These sites are so close together that they act as a single point in PFGE and most conventional Southern blot analysis.

The probe recognised 540 bp and 460 bp fragments in both BamHI/HpaII and BamHI/MspI digests demonstrating that the CpG dinucleo-

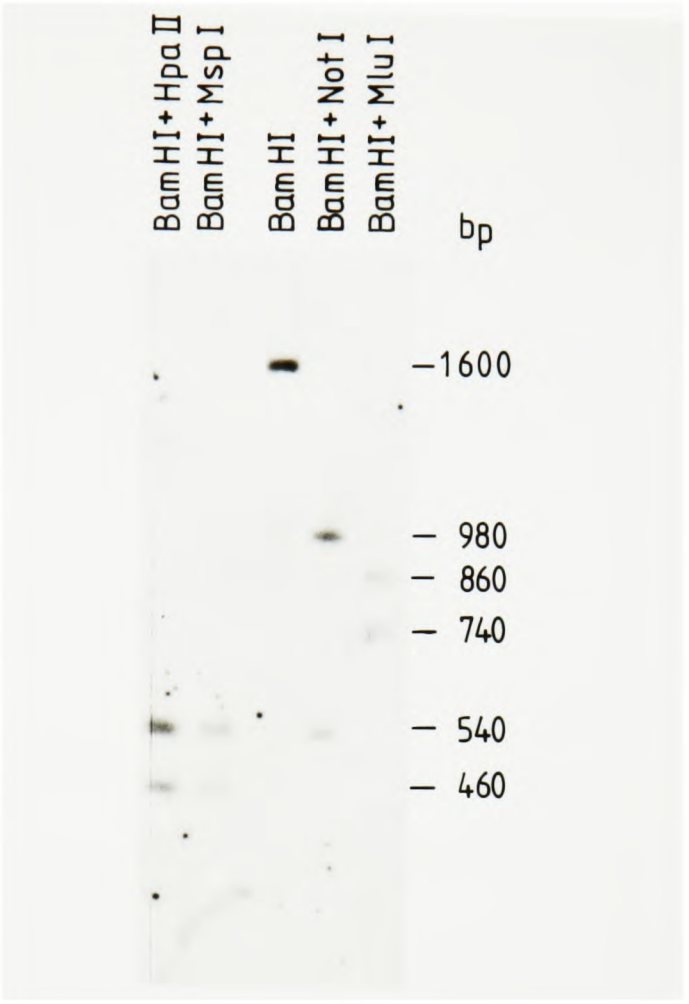


Fig 6.5 Genomic Southern blot analysis of the extent of methylation of sites for some enzymes with CpG in their recognition sequences. Digests of Ice 5 DNA with the enzymes indicated were separated by electrophoresis in a 1.5% agarose gel. The Southern blot of this gel was probed with the 1.6 kb BamHI fragment. The sizes of fragments observed after autoradiography are indicated in bp.

tides in all the HpaII/MspI sites in this fragment are unmethylated. The sum of these fragment sizes shows that approximately 600 bp of the 1.6 kb BamHI fragment is composed of very small HpaII fragments (HpaII tiny fragments or "HTF", Cooper et al., 1983) that are lost during electrophoresis or Southern transfer. The region appears to be very rich in unmethylated sites for restriction enzymes with CpG in their recognition sequence and has the characteristics of an HTF island (Bird et al. 1985)., Whether the sequence is unmethylated in tissues other than lymphoblastoid cells has not been investigated.

6.3.3b Nucleotide sequence analysis

The CpG-rich nature of this island was further investigated by nucleotide sequence analysis. In collaboration with Elaine Kendall, a series of overlapping restriction fragments from the 1.6 kb BamHI fragment were subcloned into SmaI-cut and phosphatased M13mp8 (Messing & Vieira, 1982) and sequenced by the dideoxy chain termination method (Sanger et al., 1977). The sequencing strategy is shown in Fig 6.4b. The fragment was sequenced completely on one strand and 97% on the other strand. The resulting nucleotide sequence is shown in Fig 6.6.

Bird (1986) has proposed that a sequence should be regarded as "HTF-like" if the C+G content is greater than 50% and if the frequency of CpG roughly equals the frequency of GpC i.e. there is no CpG suppression. Gardiner-Garden & Frommer (1987) regard sequences with a C+G content greater than 50% and a ratio of observed CpG frequency to CpG frequency expected from the C+G content greater than 0.6 as CpG-rich. The sequence of the 1.6 kb BamHI fragment has a total C+G content of 66.6% which is well above the average for bulk DNA. The restriction map for enzymes with CpG in their recognition sites derived from the

```
GATCCCTAGT GCGTGTGAAG GTGTAACACG CCTGATATAA ACAGAAGAGA CATTTAGCCA CAGGTTTTCA GAAGGGCACG CTGCTCCAGG GGACATGAAG CAGCATGCTA GAACCAGCCA
      10         20         30         40         50         60         70         80         90        100        110        120

AGAATAAAGG TGTCCTCCGC GTCCCTATCG CCCCTTCGTG CTCGTTAGAA AGAGCAAAAA GGCGAGGCCA AAACTGAGGC CAAGCCTTCT ATCTGGTAAC CCAAAAACTA GAAAGGCAGA
     130        140        150        160        170        180        190        200        210        220        230        240

AAAGCAACCA GAGGCGGGAA AAGCCCAACC AGGCCAGAGG AAATACAGCC GCTGCCGAGA GCCGAGAAGT TCAGAAGCTC CGCTTGTTAC CTAAAACCAG ATGGGGACAG AGACGCCAGA
     250        260        270        280        290        300        310        320        330        340        350        360

CCTGGGCCCA GGCGTGTCTC GCCCTCGGTC CCTGCAGAAC TGAGGCCAGG AATGACAGGA ATTCTCCCTT GCAGCTCCCA GCTTAGGACA GTCACAGCCA TCTCCGAGGC CGCTCTCAGG
     370        380        390        400        410        420        430        440        450        460        470        480

GGAAAATGGC CAAAAAGCAT GTAGCGGCTC CTTTAAGGCT CCTCCTCGCC CTCCCTCCCT CGCCCACGTG ACCCGGGGCG GCCGCGCGCC GGCTCGGCCC CCAGCGCAAG CGGCGATGGC
     490        500        510        520        530        540        550        560        570        580        590        600

GGCGGCGGCG GGAGCTGCAG CGGCGGCGGC CGCCGAGGTG CGGAAGGGGA GGGGGAGAGG CGGGTGCATG CCCGCGCGCG CGCCCGGGGG AAGCGCGCGC CCGTGCAATG CTCCGGGGGT
     610        620        630        640        650        660        670        680        690        700        710        720

GCAACGGGGC CGGGGGGCCC GGGCGGGACT TGCAGCAATC CAGGGGGCCA GCACCGGGGG CGGATGAGGG GACCAGGGTG TGGGAGGTGG GGGGAGTGGG CACGGAGGCG CGCGTGCAGC
     730        740        750        760        770        780        790        800        810        820        830        840

CGCCGGCTGC CCCTGAGTGG GCGACGGGCG CCGGGGCCCC AGGAGCACGC GTGAGGGGCA CGAGGGGCCG TCCACCGGGC TGGGGGGCAG CGGGTGCTGG AGGGGTTGGT GCCCGGGGTC
     850        860        870        880        890        900        910        920        930        940        950        960

CCCGCGGCCT GGGGGACAAA GGGGGAGCGG CGCGTGCAGC CGGGAGGAGG GGGCGGGGCC GGGGCGCGGA GGCCCCGCCC CCTCCCCCTC CTCTTCTCTC GGCCCCAGAG ATGCGGGGTC
     970        980        990       1000       1010       1020       1030       1040       1050       1060       1070       1080

TACCGAGAGG GAGGGGGTTG ATGCGGGCCC GGGGGAGGGG TCGTGCGGCC CCTCCGGGCA GCCGAGGCCG CGGAAGGGGG GGGCCCCACA GAGGAAGAGG TAGGCCCCGG AGCCTACTCT
    1090       1100       1110       1120       1130       1140       1150       1160       1170       1180       1190       1200

CTCTTCCCAG GGCCCAGGCA TCCTGGACCC CCCAACTCTC TACTGGGCTG ACCAGCCCTC CTGTCCCTTG TCTCCCCTCC CAGGGGGAGG CCCCCGCTGA GATGGGGGCG CTGCTGCTGG
    1210       1220       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320

AGAAGGAAAC CAGAGGAGCC ACCGAGAGAG GTGAGTGCAG CTAAAACGGG CCCCAATTGG ACCGTCACCT TCCCTAAACA CAGGATCTCC TGAGCCCTTA GCTCTGAACT TTTTCTCTTC
    1330       1340       1350       1360       1370       1380       1390       1400       1410       1420       1430       1440

CTCTACAGTT CATGGCTCTT TGGGGGACAC CCCTCGTAGT GAAGAAACCC TGCCCAAGGC CACCCCCGAC TCCCTGGAGC CTGCTGGCCC CTCATCTCCA GCCTCTGTCA CTGTCACTGT
    1450       1460       1470       1480       1490       1500       1510       1520       1530       1540       1550       1560

TGGTGATGAG GGGGCTGACA CCCCTGTAGG GGCTACACCA CTCATTGGGG ATGAATCTGA GAATCTTGAG GGAGATGGGG ACCTCCGTGG GGGCCGGATC C
    1570       1580       1590       1600       1610       1620       1630       1640       1650       1660
```

Fig 6.6 The nucleotide sequence of the 1.6 kb BamHI fragment. The sequence is written 5'→3' going telomeric.

sequence reflects the CpG-rich nature (Fig 6.7d). For instance there are 17 HpaII sites, 16 of these closely packed in a central region of 650 bp, no two separated by more than 120 bp. This explains the previous observation of ∿600 bp of "missing" DNA in the BamHI/HpaII genomic digest. Similarly there are two NotI sites separated by 70 bp, seven BssHII sites including two clusters of overlapping sites, and two SacII sites. The frequencies of these restriction sites as compared to the frequencies predicted for an "average" island are summarised in Table 6.3. For all the enzymes except NarI and PvuI the frequencies observed in this 1.6 kb fragment are greater than predicted. It is also interesting to note that all the CpG-enzyme sites except a single HpaII site are clustered in a central 650 bp core of the sequence (Fig 6.7d) which reflects the distribution of CpG dinucleotides in the sequence as described below.

Enzyme	*Site*	*No. CpGs*	*Predicted sites/island**	*Observed sites/ island+*
HpaII	*CCGG*	*1*	*11.10*	*17*
HhaI	*GCGC*	*1*	*11.10*	*16*
SmaI	*CCCGGG*	*1*	*1.20*	*5*
NaeI	*GCCGGC*	*1*	*1.20*	*2*
NarI	*GGCGCC*	*1*	*1.20*	*1*
BssHII	*GCGCGC*	*2*	*1.20*	*7*
SacII	*CCGCGG*	*2*	*1.20*	*2*
EagI	*CGGCCG*	*2*	*1.20*	*2*
MluI	*ACGCGT*	*2*	*0.34*	*1*
PvuI	*CGATCG*	*2*	*0.34*	*0*
NotI	*GCGGCCGC*	*2*	*0.12*	*2*

*Table 6.3 Comparison of observed and predicted frequencies for infrequently cutting restriction endonucleases in a CpG-rich island. *Taken from Lindsay & Bird (1987). +From the sequence of the 1.6 kb BamHI fragment.*

The distribution of nucleotides and dinucleotides within the 1.6 kb fragment sequence was then further examined (Fig 6.7). Fig 6.7a shows

a plot of the C+G content per 200 bp calculated from a 200 bp window moved 50 bp at a time. Over the whole of the 1.6 kb BamHI fragment the C+G content never drops below 50%. The central 650 bp core of the sequence is extremely C+G rich (>70%) reaching a maximum of 80%. Fig 6.7b shows a plot of the CpG and GpC doublet frequencies across the sequence calculated as above. The position of each CpG and GpC in the sequence is indicated in Fig 6.7c. In the central region of the island the CpG frequency is almost equal to the frequency of GpC suggesting a lack of CpG suppression, but towards the outer limits of the sequence the CpG frequency drops well below that of GpC. The extent of CpG suppression within the fragment can be seen more clearly in Fig 6.7a. The ratio of observed to expected (obs/exp) frequency of CpG taking into account the C+G content is plotted relative to position for 200 bp windows across the sequence. At the telomeric side of the island the obs/exp value drops to 0.2 which is the level expected for bulk DNA with CpG suppression. However in the central island core, the obs/exp ratio for CpG frequency is above 0.6 as expected for a CpG-rich island and even reaches 1.0 showing that there is no CpG suppression. On the centromeric side the obs/exp CpG frequency ratio drops but does not reach the bulk DNA level suggesting that the island structure may extend out beyond the available sequence in this direction. It can also be seen that in each plot the central 650 bp core of the island has two parts, the plots having a major peak over 200 bp at around nucleotide 650 in the sequence and a second minor satellite peak at nucleotide 900. This double peak distribution of CpG can also be seen in a number of other island sequences (Lavia et al., 1987; Bird et al., 1987; Fischel-Ghodsian et al., 1987a) and may be a common feature of an island sequence.

It is apparent from these plots and the limited data on methyl-

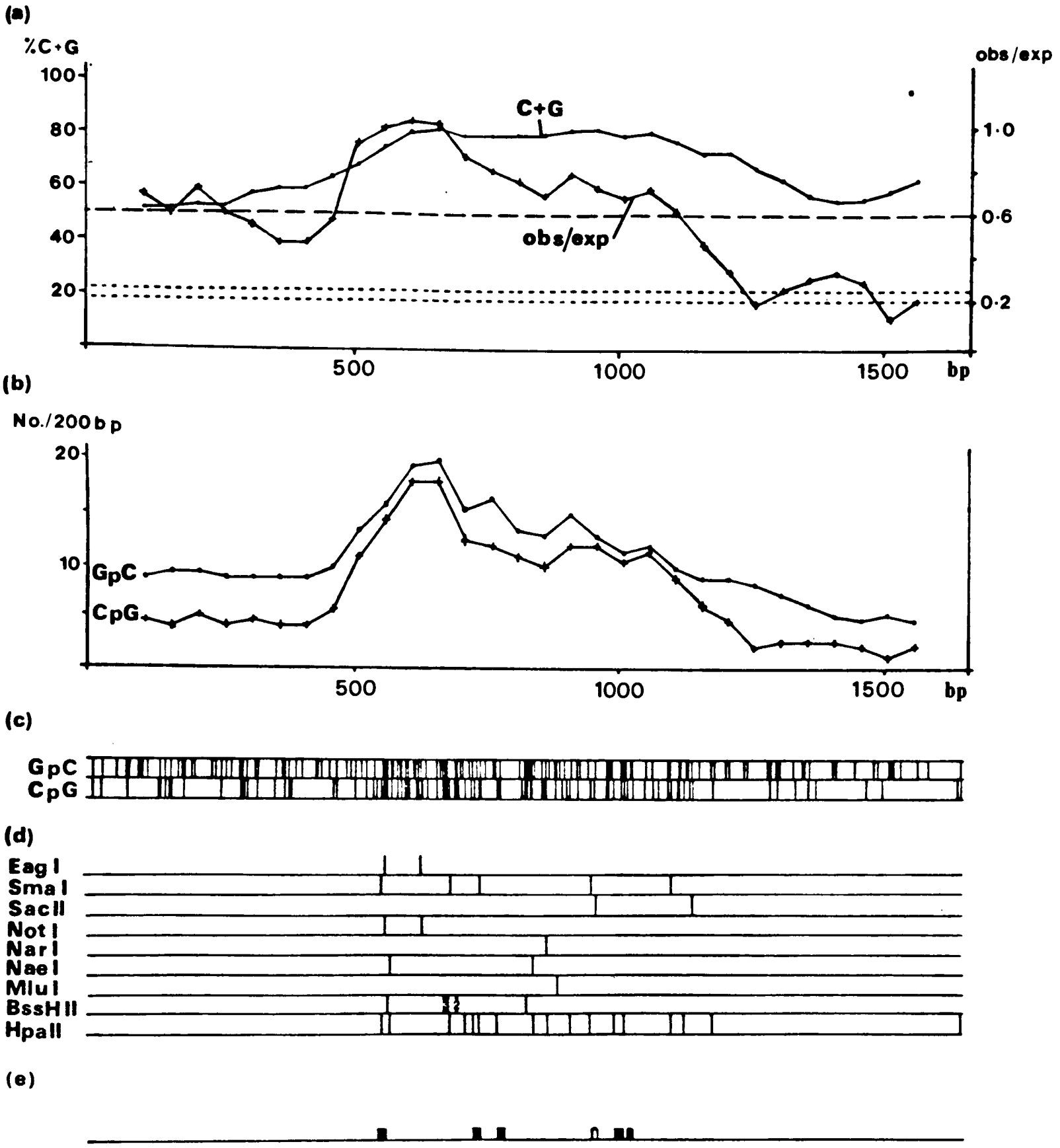


Fig 6.7 Analysis of the distribution of CpG within the sequence of the 1.6 kb BamHI fragment. The centromere is to the left. (a) %C+G and obs/exp ratio against nucleotide position. %C+G is calculated for a 200 bp window at 50 bp intervals across the sequence. The obs/exp ratio for CpG is calculated for 200 bp windows as obs/exp= (Number of CpG/ number of C x number of G) x N, where N is the number of nucleotides considered. The single dashed line shows the values of %C+G and obs/exp for an island. The double dashed lines show the obs/exp ratio for bulk DNA. (b) Plot of number of CpG and GpC per 200 bp moving window against nucleotide position. (c) The position of every CpG and GpC dinucleotide in the sequence is indicated by a vertical line. (d) Potential restriction sites for enzymes with CpG in their recognition sequence are indicated by vertical lines. Numbers mark multiple overlapping BssHII sites. (e) Closed boxes mark the positions of potential Sp1 binding sites, open box an AP2 binding site.

ation of CpG doublets that this sequence has all the characteristics of a CpG-rich or HTF island.

6.3.3c Does this CpG-rich island mark a gene?

CpG-rich islands mark the 5' ends of all sequenced housekeeping and some tissue specific genes (see section 6.1). Having established that the sequence lying 27 kb telomeric of the C2 gene was indeed an island it was obvious to ask if it was associated with a coding region.

Coding sequences are often highly conserved between species and several genes have been identified by the "zoo-blot" stategy. Therefore the 1.6 kb BamHI and 1.3 kb BamHI/XhoI fragments were hybridised with Southern blots of BamHI digested DNA from a number of animal species (kindly provided by Carole Sargent) and the blots were washed at high stringency (0.2 x SSC, 65°C, Fig 6.8). The 1.6 kb BamHI probe hybridised to itself in the human DNA (Fig 6.8). The 1.3 kb BamHI/XhoI probe hybridised to the 3.7 kb BamHI fragment from which the probe was derived in the human DNA, but also faintly to a 15.1 kb BamHI fragment of unknown origin. Although this probe had previously seemed unique in genomic Southern blot analysis it must be related to another sequence in the human genome. Both probes cross-hybridised with BamHI restriction fragments in mouse, cat, sheep and rabbit DNA at this stringency, but not to chicken or shark DNA. This implies that these sequences are evolutionarily conserved across these species and therefore may contain coding sequences. Probe K which lies some 21 kb centromeric of the island did not hybridise to any of the animal DNAs, providing a convenient negative control as a sequence that is apparently not conserved. Interestingly, although the 1.6 kb BamHI and 1.3 kb BamHI/XhoI probes were hybridised to different Southern blots, the sizes

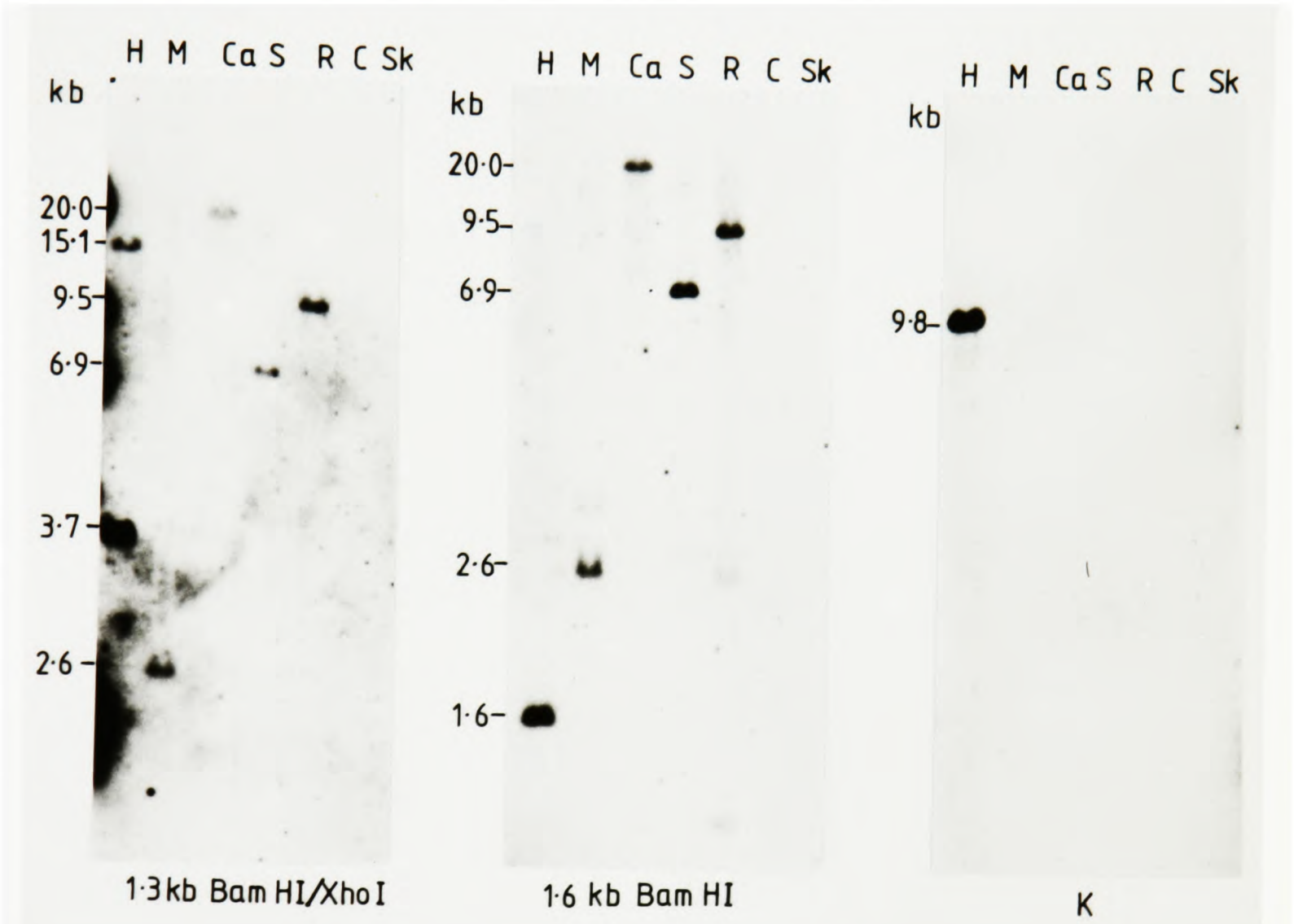


Fig 6.8 Hybridisation of genomic DNA probes to Southern blots of genomic DNA from various animals. BamHI digests of the various DNAs were separated by electrophoresis in a 0.7% agarose gel. DNAs are human (H), mouse (M), cat (Ca), sheep (S), rabbit (R), chicken (C) and shark (Sk). Southern blots of these gels were hybridised with the probes as indicated below the panels and washed at 0.2 x SSC, 0.1% SDS at 65°C.

of the major BamHI fragments revealed in all four animal DNAs were the same for both probes except that the 1.6 kb BamHI probe also hybridised faintly to additional fragments in mouse and rabbit DNA. This implies that the linkage of the conserved sequence in each human BamHI fragment is also conserved across these species.

Both probes were then used in Northern blot analysis of total RNA from a number of cell lines as shown in Fig 6.9. Although it is possible that a transcript may extend telomeric of the 1.6 kb BamHI fragment no suitable single copy probe could be obtained from this side of the island. The probes hybridised strongly to a region that coincides with the position of the 28S rRNA and less strongly with the 18S rRNA. It was assumed that this represented spurious homology as it is known that the high C+G content of island sequences makes cross-hybridisation with rRNA a frequent occurrence (Lindsay & Bird, 1987). The 1.3 kb BamHI/XhoI probe hybridised to a 2.6 kb transcript. The 1.6 kb BamHI probe also hybridised faintly to this transcript as well as to transcripts of 1.0 kb and 1.4 kb. Therefore it appears that this island is associated with transcripts that are present in each of the cell lines tested. It should be noted that these signals were faint and poorly reproducible and it is possible that the transcripts are present at low copy number or are unstable.

Both probes have been used to screen two cDNA libraries as detailed in section 2.10. Six rescreening positive clones were obtained from the HL60 cDNA library when screened with the 1.6 kb BamHI probe which contained inserts varying in size between 0.5-1.3 kb. However on the basis of hybridisation of the cloned cDNA inserts to class III region-containing cosmid DNA and to genomic DNA on Southern blots, the cDNAs were not derived from either cosmid M2 or the cloned class III region (results not shown). From the pattern of fragments observed in

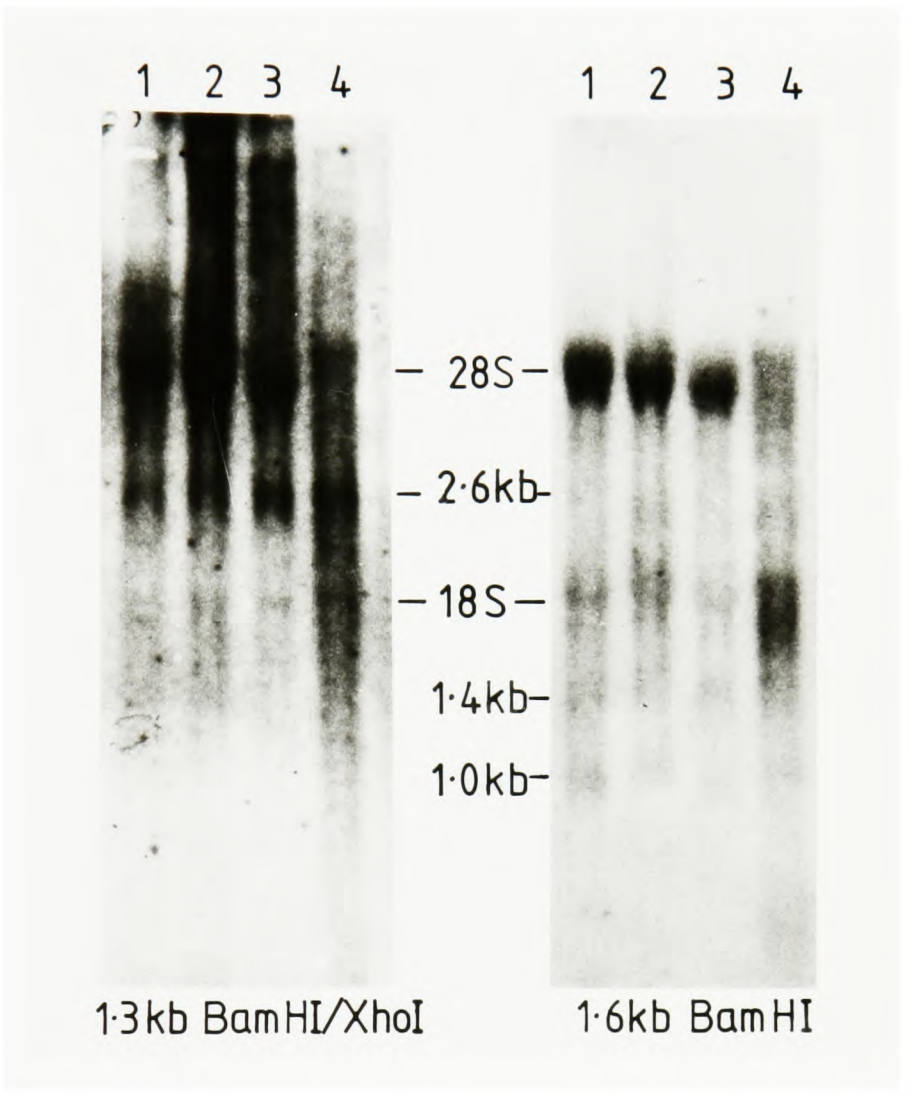


Fig 6.9 Northern blot analysis of RNA from several cell lines with the 1.3 kb BamHI/XhoI and 1.6 kb BamHI probes. Conditions were as described in Chapter 2. The tracks are RNA from 1, U937; 2, Raji; 3, Daudi; 4, HepG2. The positions of the 28S and 18S rRNAs are indicated.

genomic Southern blot analysis and Northern blot analysis, four of the clones appeared to be highly homologous to the ferritin H chain multigene family (Jain et al., 1985; Boyd et al., 1985). The 1.6 kb BamHI fragment may be hybridising to ferritin related cDNA clones because of the high G+C content of both sequences. Comparison of the 1.6 kb BamHI fragment nucleotide sequence with the sequences of a number of published ferritin H chain cDNA and gene sequences (Hentze et al., 1986; Constanzo et al., 1986; Boyd et al., 1985) did not reveal extensive homology. Moreover the RNA species observed in Fig 6.9 do not correspond in size to the ferritin mRNA which is 1.1-1.2 kb (Boyd et al., 1985). These cDNA clones were not analysed further.

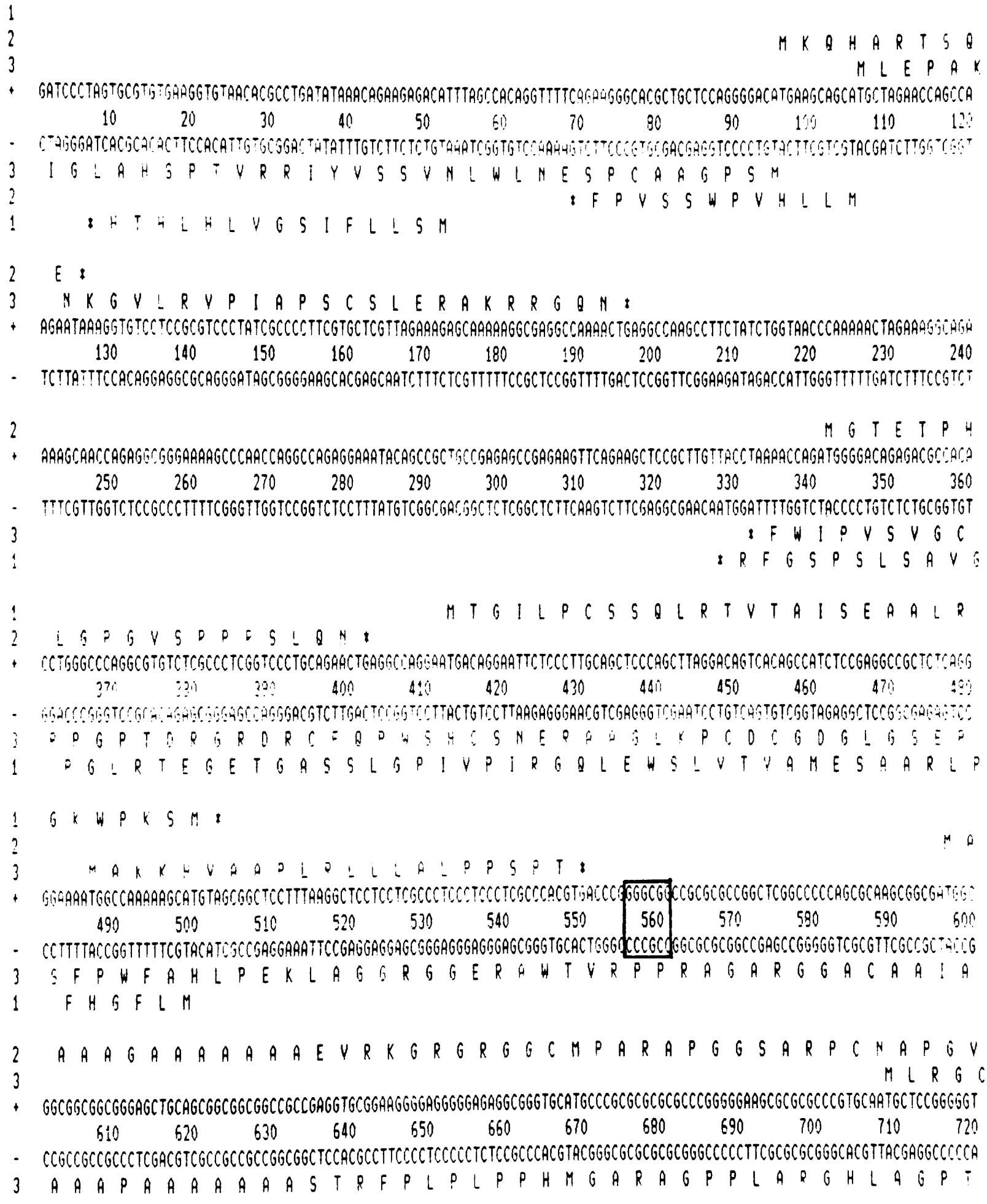
Screening of a lymphoblastoid cell line library with the 1.3 kb BamHI/XhoI fragment identified a single rescreening positive clone from 3×10^6 clones. This cDNA clone contained a 300 bp insert which on the basis of hybridisation to genomic DNA and cosmid insert DNA appeared to be derived from the 1.3 kb BamHI/XhoI fragment in cosmid M2 (results not shown). Preliminary nucleotide sequence analysis of this cDNA clone and of the 1.3 kb BamHI/XhoI fragment suggested that this was indeed the case but time has not allowed further characterisation.

In the absence of any suitable cDNA sequence it was decided to examine the DNA sequence of the 1.6 kb BamHI fragment for evidence of potential coding sequences. Segments of the sequence 400 bp in length were used in the FASTN programme to screen the EMBL nucleotide sequence database (version 14). No homologies were observed that gave any clue as to the nature of the putative gene. Translation of the DNA sequence into protein sequence reveals a number of possible open reading frames. The derived amino acid sequences from each reading frame were used in 100 amino acid stretches to screen the NBRF protein database (version 17) using the FASTP programme (Lipmann & Pearson 1985). Again no significant

homologies with known sequences were observed. Assuming that the island marks the 5' end of a gene, several open reading frames that begin with methionines can be identified and these are shown in Fig 6.10. However the significance of these is limited without a fully characterised cDNA clone.

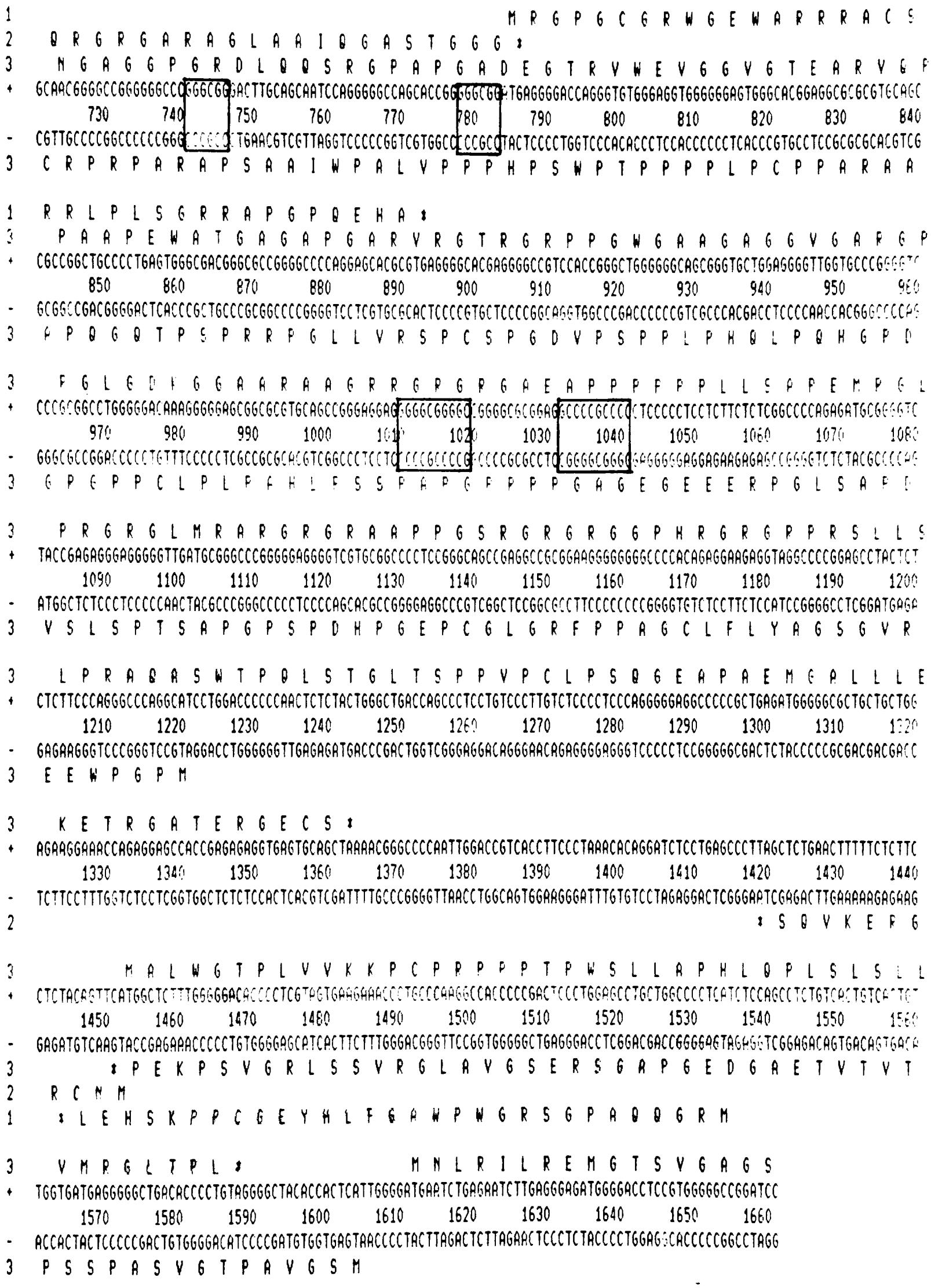
Bird (1986) has suggested that islands might bind ubiquitous transcription factors and Gardiner-Garden & Frommer (1987) have noted that potential binding sequences for the transcription factor Sp1 frequently occur upstream of the transcription start sites of genes with CpG-rich islands at their 5' ends. Therefore the genomic DNA sequence of the 1.6 kb BamHI fragment was searched for the presence of the potential transcription factor binding sites detailed by Jones, Rigby & Ziff (1988). One potential binding site for transcription factor AP2 (CCCCAGGC) and six sites for the Sp1 consensus, GGGCGG, were observed. These sites are clustered in the central CpG-rich core of the island as shown in Fig 6.7e. Transcription factor Sp1 has been implicated in promoter function for several cellular genes, all of which have CpG island sequences around their promoter (Kadonaga, Jones & Tjian, 1986; Dynan, 1986). Three of the Sp1 binding sites match with the more extensive Sp1 consensus sequence GGGGCGGGGC (Briggs et al., 1986) and these occur in an inverted repeat region of 46 bp around nucleotide 1026 in the sequence (Fig 6.11). Lavia et al. (1986) have previously observed two copies of the Sp1 consensus sequence closely spaced in an island that is associated with divergent transcripts. Conserved CpG-rich islands at the 5' end of the nucleolin genes of rat and hamster also contain tandemly arranged Sp1 binding sites about 150 bp 5' to the start site of transcription (Bourbon, Prudhomme & Amalric, 1988). The observation of these potential binding sites within the fragment may be fortuitous due to the high C+G content (see Gardiner-Garden & Frommer,

Fig 6.10 Open reading frames (ORF) starting with methionines in the sequence of the 1.6 kb BamHI fragment. The derived protein sequence for each ORF is shown above or below the sequence for the + or - strands, respectively. Numbers at the beginning of each line refer to the reading frame in which the ORF is found, reading frame 1 starting at the first 5' nucleotide. Boxes indicate the positions of potential Sp1 binding sites.



Contiued overleaf.....

Fig 6.10 continued.



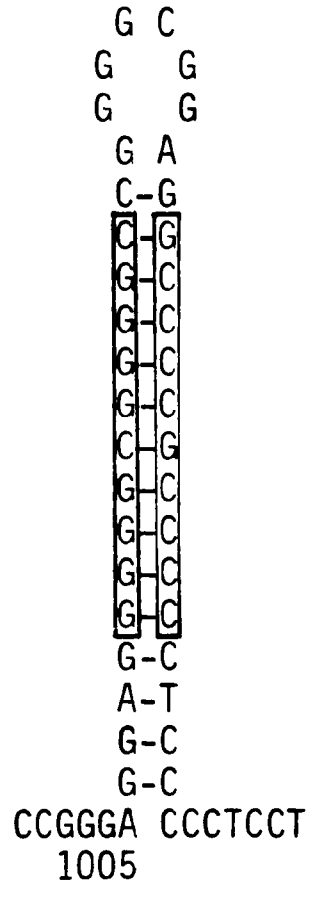


Fig 6.11 Possible stem-loop structure that could be formed by an inverted repeat sequence starting at nucleotide 1005 in the sequence of the 1.6 kb BamHI fragment. Boxes mark the positions of two Sp1 binding sites.

1987 for discussion) and the presence of Sp1 binding sites does not necessarily mean that Sp1 will bind (Kadonaga et al., 1986). However these results do lend support to the idea that a gene may be associated with this island.

6.4 DISCUSSION

The DNA of the class III region of the human MHC that has been studied here contains a large number of CpG-rich islands. Examination of the frequency of sites for rarely cutting restriction endonucleases within 541 kb of cloned cosmid DNA compared to the frequency of these sites predicted from bulk dinucleotide frequencies (Table 6.1) revealed that these sites occur more often than was expected. In particular sites for NotI, BssHII, SacII and EagI are up to eleven times overrepresented. In contrast PvuI and NruI which both contain A and T in their recognition sequence as well as C and G occur at just below the expected frequency. A similar analysis of cosmids picked at random from human chromosome 3 also showed an overabundance of rare-cutting sites (Smith et al., 1987b) 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 and this was also observed by Smith et al. (1987b). The apparent excess of sites can be explained by a highly non-random distribution of nucleotides and dinucleotides in the genome both at the local level of CpG-rich islands and on a wider scale as discussed below. The relative paucity of sites for CpG-enzymes with A/T in their recognition sites as well as C/G may also reflect clustering of CpG in regions that are C+G rich.

The analysis of sites that are cleavable in genomic DNA has

revealed a high concentration of putative CpG-rich islands within the class III region of the MHC (Figs 6.1, 6.12). The islands occur at well above the frequency of one per 100 kb predicted (Brown & Bird, 1986). Similar high densities of islands have also been observed in other regions of the mammalian genome, for instance around the human α globin locus (Fischel-Ghodsian et al., 1987b), around the human G6PD locus (Toniolo et al., 1984) and in the mouse surfeit locus (Williams et al., 1988). Such clustering may be a common feature of the organisation of islands.

The islands that were detected in this study have been tested for the presence of associated transcripts (Sargent, 1988). Fig 6.12 shows the positions of the putative genes within the cloned MHC class III region as defined by the detection of transcripts by genomic hybridisation probes. Many of the newly detected genes are constitutively expressed in all tissues tested so far (G2 is an exception being expressed in monocyte, macrophage and T cell lines, but not liver or B cell lineages) and thus may be regarded as housekeeping. This is in contrast with the previously known genes which are expressed mainly in a tissue specific manner.

These findings raise several points. First, the use of CpG-rich islands to detect genes is a convenient and successful strategy. In this case mostly housekeeping genes have been detected. As mentioned previously the complement and 21-OHase genes would not have been detected by this search. It is not yet clear how biased this method is towards housekeeping genes and it may be that other tissue specific genes that do not possess associated islands have not been detected. Second, the class III region contains a large number of previously undefined loci and there are probably more genes in the cloned and uncloned regions. This finding also illustrates that many genes in a

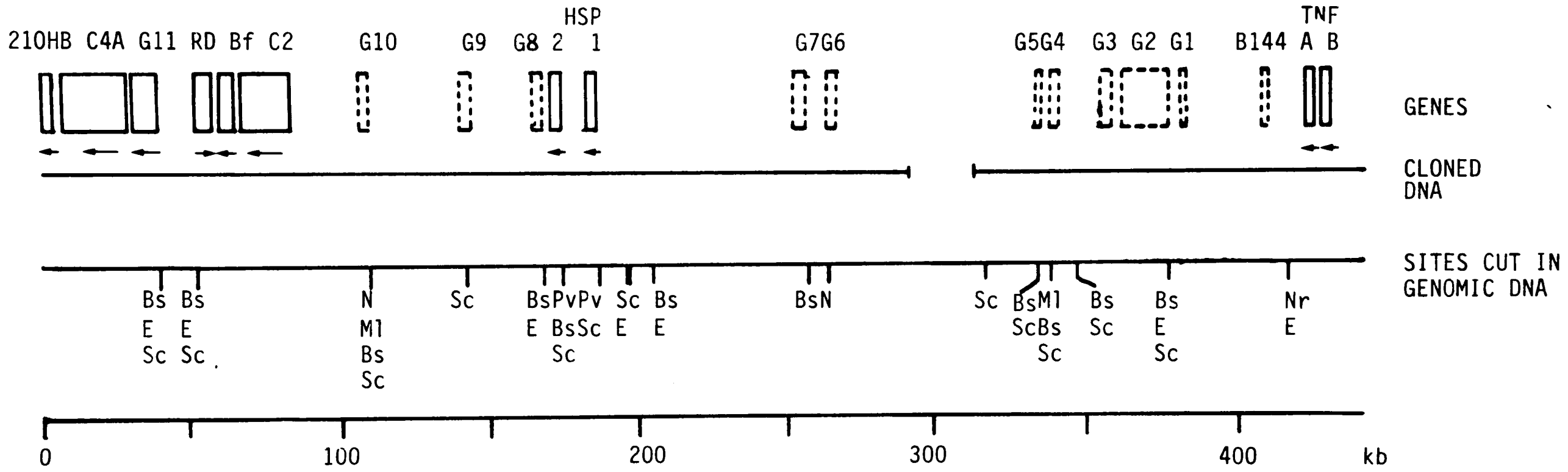


Fig 6.12 The locations of putative genes in the human MHC class III region. Open boxes are genes whose extents have been accurately determined from cDNA and genomic clones. Arrows indicate the direction of transcription. Dotted open boxes are regions for which transcripts have been detected. The positions of rare cutting restriction sites that are cleaved in Ice 5 genomic DNA are indicated to show the positions of CpG-rich islands. Restriction enzymes are BssHII, Bs; EagI, E; SacII, Sc; NotI, N; MluI, Ml; NruI, Nr; PvuI, Pv.

region may go unobserved until a systematic effort is made to detect them. At this point in the genetic analysis of the human genome our concept of gene organisation has been biased by the understandable tendency to identify the genes whose products are known, thereby overlooking the vast majority of genes. The density of genes in this region appears to be very high. In 541 kb of cloned DNA from the class III region twenty distinct expressed genes have so far been identified, that is one every 27 kb. A simplistic calculation based on 50 000 expressed gene products for mammals and 3×10^6 kb of DNA in the mammalian genome suggests an average gene density of one gene every 60 kb. As it is expected that more genes may lie in this cloned region, there may well be a non-random distribution of genes in the genome as well as a non-random distribution of nucleotides. As discussed in Chapter 5, this high density of mostly housekeeping genes may have some influence on the observed genetic phenomena within the MHC.

Regions of tightly packed gene clusters may be more common than has been previously thought. Several regions have been identified with clustered CpG islands and genes as discussed above. A number of other observations on genome organisation may be pertinent at this point. Density gradient centrifugation of vertebrate nuclear DNA results in separation into fractions of varying base compositions (Bernardi et al., 1985). It has been proposed that the genome of warm blooded vertebrates is a mosaic composed of large regions ($>>$200 kb) called isochores which are fairly homogeneous in base composition and belong to a small number of classes distinguished by their C+G content. The shearing of these isochores during preparation of genomic DNA generates the observed separable fractions, each of which has a different base content. Analysis of the distribution of genes within these fractions has shown that genes are preferentially distributed in the heavy C+G rich

isochores and that the C+G content of the genes reflects the C+G content of the component in which they are observed. For instance the α globin gene cluster which has a high C+G content and multiple CpG islands is located in the heavy isochore.

Studies of chromosome banding by staining with Giemsa (G) or Quinacrine (Q) have shown that the G/Q bands correspond to regions that are A+T rich and replicate late in the S phase of the cell cycle. The Giemsa light (R) bands correspond to early replicating C+G rich regions (Comings, 1978; Holmquist et al., 1982). Housekeeping genes are preferentially located in early replicating DNA (Goldman et al., 1984) which reflects the distribution of these genes predominantly in the C+G rich isochores. 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 in the chromosome bands. In this respect it is interesting that the MHC lies in 6p21 which is a Giemsa light band. The class III region is evidently C+G rich and contains a number of presumed housekeeping genes. CpG-rich islands have also been detected within the sequences of class I and II genes (Tykocinski & Max, 1984; Gardiner-Garden & Frommer, 1987) and unmethylated CpG islands have been shown to be present in the class I region (Pontarotti et al., 1988). 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. Other regions of the genome may be depleted in genes and contain only tissue specific genes. It is noteworthy however that the MHC contains tissue specific genes both with and without islands dispersed among the C+G rich housekeeping clusters.

It has also been observed that the distribution of specific repeat elements within the genome correlates with the chromosome banding patterns. The Alu family (56% G+C) dominates in R bands and the L1

family (58% A+T) follows the G/Q banding (Korenberg & Rykowski, 1988). The Alu family has been implicated as a mediator of unequal homologous and non-homologous exchange from studies of deletions in the α and β globin gene clusters (Nicholls et al., 1987; Henthorn et al., 1986). The location of the MHC in an Alu-rich band suggests that such a mechanism could also be involved in the generation of the known deletions and duplications within the MHC.

The analysis of an isolated CpG-rich island from the MHC has confirmed its nature on the basis of clustering of unmethylated sites for enzymes with CpG doublets in their recognition sites and nucleotide sequence analysis. The island is unmethylated in lymphoblastoid cells. Although no other tissue has been tested, it is predicted that HTF-like sequences will be unmethylated at least in germ line cells (Bird, 1986). The sequence of the island shows a high C+G content and a lack of CpG suppression. The island also contains recognition sites for the transcription factor Sp1 which may be typical of CpG island gene promoters (Gardiner-Garden & Frommer, 1987; Dynan, 1986). However, although the region appears to be a typical CpG-rich island and RNA transcripts are associated with it, isolating a cDNA clone from the putative gene has proved difficult. This may be due to several factors. The mRNA may be present at very low levels or may be unstable. This problem is compounded by the fact that although apparently unique in the genome on the basis of Southern blots, the 1.6 kb BamHI fragment cross-hybridises with other sequences in cDNA library screens. A more extensive analysis of RNAs from other tissues might reveal a better source of the mRNA from which to prepare cDNA libraries. Additionally it may be possible that the genomic probes used do not contain much exonic sequence and therefore do not readily detect the mRNA species. Since the cDNA libraries screened were prepared by oligo dT priming of the reverse

transcriptase which creates a bias in favour of 3' end sequences in the library, and the probes are presumably close to the 5' end of the transcript(s), this may also explain why clones have been difficult to obtain. Probes derived further away from the island might prove more profitable.

CHAPTER 7

DISCUSSION

7.1 A MOLECULAR MAP OF THE HUMAN MHC

Data from PFGE genomic Southern blot analysis using HLA gene probes and single-copy probes isolated from genomic clones obtained during cosmid walking have been used to establish a molecular map of the human MHC. The maximum size of the region estimated from the sum of the restriction fragment sizes observed with all MHC probes is 3800 kb. The position and orientation of the cosmid cluster containing the genes for C2, Factor B, C4 and steroid 21-hydroxylase was not previously known. The maps presented in Chapters 3 and 5 indicate that these genes are orientated with the C2 gene telomeric of the 21-OHase genes. The distance between the the most centromeric 21-OHase gene and the most telomeric of the class II genes (DRA) is no greater than 390 kb. Genes for TNF α and β have been located in the human MHC 325 kb telomeric of the C2 gene. The distance between the TNF genes and the HLA-B gene is 250-300 kb. Comparison of the restriction fragment maps of cosmid clones containing the TNFA and TNFB genes with published restriction maps of TNF genomic clones (Nedospasov et al., 1985) indicates that this gene cluster is orientated with the TNFA gene centromeric of the TNFB gene.

The positions of many of the genes in the class II subregions have been established from the analysis of a cell line carrying the B8, DR3, SCO1 haplotype and from comparison of the organisation of the six cell lines carrying other haplotypes. These data confirm the class II subregion organisation established by Hardy et al. (1986).

The precise organistion of the class I region between the HLA-C and A genes is still unclear as these genes have not been linked on a common restriction fragment or by chromosome walking. Given that there are many class I-like genes located telomeric of the HLA-A gene which hinders cosmid walking and that the relative positioning of the restriction sites in the region has meant that PFGE mapping has proved uninformative (Dunham et al., 1987; Carroll et al., 1987; Pontarotti et al., 1988), it is likely that this issue will only be resolved by chromosome walking with YAC vectors (Burke et al., 1987). The HLA-E gene which is located centromeric of HLA-A (Koller et al., 1988; Carroll et al., 1987) might prove to be a useful "landmark" in such a chromosome walk. The nature of the DNA between HLA-C and A may also be of interest. Since it is known that most class I-like sequences lie telomeric to the A gene (Orr & DeMars, 1983), it seems unlikely that many more class I genes will be found in this region, which covers at least 600 kb. Whether most of the region is noncoding or contains other genes unrelated to class I will have important implications for the understanding of the events that have occurred in the evolution of the class I region.

7.2 LARGE DNA FRAGMENT RFLPs TO DEFINE C4 GENE ORGANISATION

Comparison of the sizes of BssHII and SacII restriction fragments observed by hybridisation of a C4 or 21-OHase gene probe to the digested genomic DNA of eight cell lines and five individuals possessing, variously, one to three C4 genes, has shown that the organisation of the C4 genes on each chromosome can be deduced from a single PFGE experiment. The size of the fragment observed is directly related to the number and length of the C4 genes present on either haplotype in the DNA

sample.

This observation has significant implications for C4 genotyping. First, application of these RFLPs offers considerable advantages over previous methods. These RFLPs monitor the complete gene organisation whereas other conventional RFLPs see only a small portion of the gene (see Fig 4.3a for instance). It is possible that rare, unusual recombinants could be identified using the large RFLPs. There is no need to interpret fragment intensities to deduce the gene number. It is probable that the number of duplications (i.e. three C4 genes on one chromosome) in the population has been previously underestimated due to this difficulty. Similarly since each BssHII or SacII fragment contains all the C4 genes on one chromosome, there is no need for family studies or deduction of probable complotypes from the allotyping data and assumptions on likely ancestral haplotypes to infer the C4 gene organisation on each chromosome. Second, since C4 null alleles are known to be associated with certain HLA-linked diseases (Fielder et al., 1983; Hauptmann et al., 1986; Batchelor & McMichael, 1987) this method combined with previously described RFLPs (Yu & Campbell, 1987) will enable rapid screening of individuals to split deleted and non-deleted null alleles.

As discussed in Chapter 5, large DNA fragment RFLPs caused by loss of restriction enzyme sites may be rare because of the way sites for rare cutting enzymes are distributed in unmethylated CpG-rich islands. However, RFLPs such as this which are caused by large changes in the DNA content at a particular locus due to deletion or duplication may be more common. Indeed several examples are already present in the literature (Maeda et al., 1986; Vollrath, Nathans & Davis, 1988; den Dunnen et al., 1987).

7.3 THE DNA ORGANISATION OF DIFFERENT HLA HAPLOTYPES

Analysis of the long range DNA organisation of the class II and class III regions in eight homozygous cell lines using infrequently cutting enzymes and PFGE has revealed a number of interesting features of the MHC region. The class III region shows a highly invariant structure outside of the C4 and 21-OHase loci, with no detectable differences in the amount of DNA present. The estimated resolution of the technique over ~450 kb of the class III region would enable size differences down to 2 kb to be detected. This constant organisation is at first surprising, but since it is now known that the class III region contains a high density of genes, many of which probably have housekeeping functions, then the observation makes sense.

In the class II region, although of necessity the analysis was at lower resolution, two differences between common haplotypes have been found. The DRw52 supertypically related DR3, DR5 and DR6 haplotypes have the same DNA organisation. DR2 haplotypes possess 20-30 kb more DNA which may be due to differences in the organisation of the DRB gene region. The haplotypes that are related by sharing the DRw53 supertypic specificity (i.e. DR4 and DR7) have 100-130 kb more DNA than the DRw52-related haplotypes in the region containing the DRB and DQA genes. It is not clear what form this extra DNA takes. These findings may have implications for the patterns of recombination between different haplotypes and the strong linkage disequilibrium observed between DQ and DR. Recombination between DR and DQ has not yet been observed. In addition these observations have confirmed the evolutionary relatedness of haplotypes sharing the DRw52 or DRw53 supertypic specificities, since their genomic organisations appear to be conserved.

The implications of these differences between haplotypes will not

become clear until the class II DNA organisation in a large number of individuals has been determined. However, it is possible that there may be other disease-specific differences in the class II region which are responsible for certain HLA associated diseases. For instance almost all patients with narcolepsy carry DR2. It may be worth comparing the long range DNA structure between patients and controls. The maps and fragment sizes established here should provide a standard against which to compare future results.

7.4 THE CLASS III REGION CONTAINS MANY GENES AND IS CpG-RICH

Using single-copy genomic hybridisation probes isolated from two clusters of cosmid clones covering 541 kb of DNA, the positions of unmethylated sites for rare-cutter restriction enzymes have been mapped. The class III region of the human MHC contains a number of clusters of such sites (Fig 6.1). These clusters, and also some single sites, probably represent CpG-rich or HTF islands. A limited analysis of CpG methylation and the nucleotide sequence of 1.6 kb of DNA surrounding one cluster of rare-cutter sites has confirmed its HTF island nature. It has subsequently been shown that many of these islands are associated with phylogenetically conserved sequences as judged by hybridisation to Southern blots of animal DNA. Hybridisation of suitable probes to Northern blots has shown that many of the islands are associated with mRNA transcripts. In some cases cDNAs for these transcripts have been cloned. Thirteen putative genes have been identified in this way in the class III region in addition to the human homologues to the murine RD and B144 loci. Two of the loci have been identified as genes for the heat shock protein, HSP70, at least one of which is functional.

These observations have considerable implications for our

understanding of genome organisation and the structure of the MHC.

First, the class III region has a very high density of genes. As discussed in Chapter 6 it may be that many genes in the genome are organised into densely packed clusters. In particular housekeeping genes may be arranged in regions rich in CpG islands and C+G, such as the class III region. The α globin gene cluster is also associated with a number of CpG islands in a C+G rich region (Fischel-Ghodsian et al., 1987a;b). These clusters might be contained in the C+G-rich heavy isochores (Bernardi et al., 1985). Other regions of the genome may contain fewer genes which are tissue specific. Until other regions are analysed for the presence of genes this question of genome organisation will remain unanswered.

Second, it is known from "zoo blots" that the sequences in the class III region are conserved across mammalian species (Sargent, 1988). In the case of the HSP70 genes, the sequence conservation extends to the shark and even to Drosophila. None of the class III gene sequences so far determined is related to the class I and class II antigen genes. It is also known that in man and mouse the main class I encoding region of the MHC is separated from the class II region by the class III region. However, this is not the case in the chicken (Guillemot et al., 1988). If it is assumed that the class I and class II genes arose from an ancient duplication of a single gene, the question arises as to how the class III genes came to be located in the MHC. One possibility is that an inversion with one end point within the MHC brought the array of class III genes between the class I and class II genes which had previously been adjacent (Klein & Figueroa, 1986). Another is that the class III genes were initially on another chromosome and a crossover event has inserted them into the MHC. A third possibility is that many insertion events have occurred into the middle of the MHC to bring in

unrelated genes. It should be possible to make use of the cross-species hybridisation of the class III region probes in PFGE analysis to provide some information relevant to this issue. For instance the linkage of the class III sequences to each other and also to the class I and class II genes in different species could be examined.

It is interesting that the MHC contains a number of duplicated loci. In some cases the duplicated sequences diverge to produce different gene products. For instance, this could be said to be true of the initial duplication that lead to the class I and class II genes. The gene products now have different tissue distributions and different although related functions. Similarly the genes for C2 and factor B probably arose from a duplication of a common ancestral locus (Morley & Campbell, 1984). Divergence has produced genes for two proteins that share structural and functional homologies, but have distinct roles. The TNF genes may also have arisen by duplication of an ancestral gene, since they share sequence homology, are closely linked, and their products have analogous functions. However, the proteins are synthesised by different tissues. In other cases the sequences of duplicated loci in the MHC have remained very close. This is true of the genes for C4 which share >99% homology. The proteins are extremely similar in function although the C4 isotypes are distinct in their haemolytic activities (Law et al., 1984; Isenman & Young, 1984). The HSP70 genes are duplicated and highly homologous, although it is not known whether both copies are functional. Finally, in some cases duplication followed by some divergence has generated a series of pseudogenes that are highly homologous to the functional genes. For instance at least one DRB gene (DRB2) in each haplotype is not expressed. Similarly in the DQ and DP subregions one pair of A and B genes is not functional (Trowsdale, 1987). The 21-OHaseA gene has also been demonstrated to be a pseudogene

(Higashi et al., 1986; White et al., 1986; Rodrigues et al., 1987). It will be interesting to see if there are any other duplicated loci to be found in the MHC.

7.5 THE SIGNIFICANCE FOR HLA-ASSOCIATED DISEASE

The linkage disequilibrium in the MHC means that any loci found within the class III region are candidate "disease genes". At the present little is known about many of the new loci identified and it is impossible to assess what relevance they might have. However, two of the new genes could be of some significance.

The linkage of the TNFA and TNFB genes to the human MHC is intriguing. It is known that TNFα is the agent responsible for the wasting observed during chronic parasitic infection or cancer, and it is also involved in toxic shock following bacterial infection. Jacob & McDevitt (1988) have shown that in the (NZWxNZB)F_1 hybrid mouse, which develops a severe autoimmune disease similar to SLE in humans due in part to a dominant NZW MHC gene, a RFLP in the TNFA gene correlates with low levels of TNFα produced by NZW mice and that replacement therapy with recombinant TNFα can delay the onset of the disease. TNFα is a protein with several activities that might be involved in autoimmune disease. For instance TNF-α augments MHC antigen expression in some tissues (Collins et al., 1986; Pujol-Borrell et al., 1987), which may lead to enhanced cytolytic damage mediated by class I antigens, or presentation of autoantigens by epithelial cells aberrantly expressing class II antigens (Pujol-Borrell et al., 1987). It has been suggested that such a mechanism could occur after an initial stimulus, such as a viral infection, has led to local production of TNFα by macrophages. TNFα has also been found in the serum of patients with rheumatoid

arthritis (Teppo & Maury, 1987), one of the classical HLA-associated diseases. TNFα is also known to stimulate collagenase production in synovial cells and bone resorption (Oliff, 1988) which could both be involved in inflammatory disease of the joints such as ankylosing spondylitis. It is possible that there could be variations in the levels or activity of TNFα between indivivduals and the gene responsible would be in linkage disequilibrium with certain HLA haplotypes and could provide susceptibility to certain diseases given the appropriate environmental stimuli. However the evidence for such a mechanism remains poor.

The discovery of genes for the major heat shock protein, HSP70, in the class III region suggests another mechanism that might be involved in HLA-associated autoimmune diseases, molecular mimicry. HSPs from a wide range of species share a high degree of sequence homology. One of the genes (also known as HSX70) mapped into the MHC is known to be functional (Wu et al., 1987b) and is expressed at a relatively high basal level. A number of pathogenic organisms present HSP70 analogues as antigens during infection. For instance, immunisation of rats with Mycobacterium tuberculosis causes adjuvant arthritis, an animal model of rheumatoid arthritis. One of the major T cell epitopes from this model has been cloned and has been shown to be derived from a 65 000 mol. wt. HSP (van Eden et al., 1988). Similarly analysis of T cell clones indicates that many CD4+ lymphocytes in mice innoculated with M. tuberculosis are directed against the 65 000 mol. wt. protein and a 71 000 mol. wt. HSP70 analogue (Young et al., 1988). Exposure to the stress protein analogues probably leads to long term immunity to related proteins of other pathogens, but it could also present a particular challenge to the immune system in distinguishing self from non-self. If there are allelic sequence differences in the MHC-linked HSP70 genes

this could determine whether an autoimmune outcome occurs. Persistent cross-reactivity of T cells with host HSPs caused by prior infection with such a pathogen would increase damage to tissues during inflammation because of the increased production of HSPs. It is worth noting that susceptibility to tuberculosis and leprosy which are both caused by mycobacteria have been shown to have HLA-associations (Lamb & Rees, 1988).

During recovery from physiological stress, HSP70 proteins migrate to the nucleolus where they become associated with ribonucleoproteins (RNPs; Welch & Suhan, 1986). HSP70 could play an important role in the reassembly of damaged preribosomes and other RNPs. Autoantibodies against nuclear RNPs are commonly found in the sera of patients with SLE, mixed connective tissue disease, Scleroderma, and polymyositis (Tan, 1982). In one recent SLE study a high proportion of patients were found to have antibodies to HSP70 (Minota & Winfield, 1988). Although these antibodies could be raised against host cell intracellular proteins because of extensive cell stress, death and lysis during the course of the disease and therefore are a symptom rather than a causative agent, it is worth considering the possibility that allelic sequence differences might impair the activity of the HSP70 protein and contribute to susceptibility. Limited sequencing data suggest there may be some variation in the MHC-linked genes.

Although the function of the other genes located in the class III region is not known, it remains possible that one or a combination of several might have some role in HLA disease associations. In addition it is likely that other genes remain to be found in the region. Further analysis of the products of class III region genes as well as of the known products of the class I and II loci should lead to an understanding of the mechanisms of these intriguing associations.

REFERENCES

Abbal, M., Moennarid, C., Cambon-Thomsen, A., Tkaczuk, J., Ohayon, E. & Mauff, G. (1987) Immunogenetics 26, 320-322.

Acha-Orbea, H. & McDevitt, H. O. (1987) Proc. Natl. Acad. Sci. USA 84, 2435-2439.

Allen, F. H. (1974) Vox Sang. 27, 382-384.

Alper, C. A. (1976) J. Exp. Med. 144, 1111-1115.

Alper, C. A. (1981) in The Role of the Major Histocompatibility Complex in Immunobiology pp 173-220 (Garland, N. Y.) Dorf, M. D. ed.

Alper, C. A., Raum, A., Karp, S., Awdeh, Z. L. & Yunis, E. J. (1983) Vox. Sang. 45, 62-67.

Anand, R. (1986) Trends in Genet. 2, 278-283.

Bach, F. H. & van Rood, J. J. (1976) N. Engl. J. Med. 295, 806-813 (I), 872-878 (II), 927-936 (III).

Bancroft, I. & Wolk, C. P. (1988) Nucl. Acids Res. 16, 7405-7418.

Barker, D., Schafer, M. & White, R. (1984) Cell 36, 131-138.

Batchelor, J. R. & McMichael, A. J. (1987) Brit. Med. Bull. 43, 156-183.

Beck, S. & Barrell, B. G. (1988) Nature 331, 269-272.

Bell G. I., Karam, J. H. & Rutter, W. J. (1981) Proc. Natl. Acad. Sci. USA 78, 5759-5763.

Bell, J. I., Denney, D., Jr., MacMurray, A., Foster, L., Watling, D. & McDevitt, H. O. (1987a) J. Immunol 139, 562-573.

Bell, J. I., Denney, D., Jr., Foster, L., Belt, T., Todd, J. A. & McDevitt, H. O. (1987b) Proc. Natl. Acad. Sci. USA 84, 6234-6238.

Belt, K. T., Carroll, M. C. & Porter, R. R. (1984) Cell 36, 907-914.

Belt, K. T., Yu, C. Y., Carroll, M. C. & Porter, R. R. (1985) Immunogenetics 21, 173-180.

Benacerraf, B. (1981) Science 212, 1229-1238.

Bender, W., Spierer, P. & Hogness, D. S. (1983) J. Mol. Biol. 168, 17-33.

Bentley, D. R. (1986) Biochem. J. 239, 339-345.

Bentley, D. R. & Porter, R. R. (1984) Proc. Natl. Acad. Sci. USA 81, 1212-1215.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. & Rodier, F. (1985) Science 228, 953-958.

Bernards, A, Kooter, J. M., Michels, P. A. M., Moberts, R. M. P. & Borst, P.(1986) Gene 42, 313-322.

Bernards, R. & Flavell, R. A. (1980) Nucl. Acids Res. 8, 1521-1534.

Beutler, B. & Cerami, A. (1986) Nature 320, 584-588.

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.

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

Bird, A. P. (1987) Trends in Genet. 3, 342-347.

Bird, A. P. & Southern, E. M. (1978) J. Mol. Biol. 118, 27-47.

Bird, A., Taggart, M., Frommer, M., Miller, O. J. & MacLeod, D. (1985) Cell 40, 91-99.

Bird, A. P., Taggart, M. H., Nicholls, R. D. & Higgs, D. R. (1987) EMBO J. 6, 999-1004.

Birnboim, H. C. & Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.

Birren, B. W., Lai, E., Clark, S. M., Hood, L. & Simon, M. I. (1988) Nucl. Acids Res. 16, 7563-7582.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987a) Nature, 329, 506-512.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987b) Nature, 329, 512-518.

Bodmer, J. G., Kennedy, L. J., Lindsay, J. & Wasik, A. M. (1987) Brit. Med. Bull. 43, 94-121.

Bodmer, J. & Bodmer, W. (1984) Immunol. Today 5, 251-254.

Bodmer, W. F., Albert, E., Bodmer, J. G., Dupont, B., Mach, B., Mayr, W., Sasazuki, T., Schreuder, G. M. T., Svejgaard, A. & Terasaki, P. I. (1988) Bulletin of W.H.O. (in press).

Böhme, J., Andersson, M., Andersson, G., Möller, E., Peterson, P. A. & Rask, L. (1985) J. Immunol. 135, 2149-2155.

Bottazo, G. F., Pujol-Borrell, R., Hanafusa, T. & Feldmann, M. (1983) Lancet ii, 1115-1119.

Boyd, D., Vecoli, C., Belcher, D. M., Jain, S. K. & Drysdale, J. W. (1985) J. Biol. Chem. 260, 11755-11761.

Bourbon, H-M., Prudhomme, M. & Amalric, F. (1988) Gene 68, 73-84.

Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. (1986) Science 234, 47-52.

Brown, J. H., Jardetsky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J. & Wiley, D. C. (1988) Nature 332, 845-850.

Brown, W. R. A. & Bird, A. P. (1986) Nature 322, 477-481.

Burke, D. T., Carle, G. F. & Olson, M. V. (1987) Science 236, 806-812.

Burmeister, M. & Lehrach, H. (1986) Nature 324, 582-585.

Burmeister, M., Monaco, A. P., Gillard, E. F., van Ommen, G-J. B., Affara, A., Ferguson-Smith, M. A., Kunkel, L. M. & Lehrach, H. (1988) Genomics 2, 189-202.

Buus, S., Sette, A., Colon, S. M., Miles, C. & Grey, H. M. (1987) Science 235, 1353-1358.

Campbell, R. D., Dodds, A. W. & Porter, R. R. (1980) Biochem. J. 189, 67-80.

Campbell, R. D. & Porter, R. R. (1983) Proc. Natl. Acad. Sci. USA 80, 4464-4468.

Campbell, R. D., Bentley, D. R. & Morley, B. J. (1984) Phil. Trans. R. Soc. Lond. B306, 367-378.

Campbell, R. D., Law, S. K. A., Reid, K. B. M. & Sim, R. B. (1988) Ann. Rev. Immunol. 6, 161-195.

Cantor, C. R., Smith, C. L. & Mathew, M. K. (1988) Ann. Rev. Biophys. Biophys. Chem. 17, 287-304.

Carle, G. F. & Olson, M. V. (1984) Nucl. Acids Res. 12, 5647-5664.

Carle, G. F., Frank, M. & Olson, M. V. (1986) Science 232, 65-68.

Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984a) Nature 307, 237-241.

Carroll, M. C., Belt, K. T., Palsdottir, A. & Porter, R. R., (1984b) Phil. Trans. R. Soc. Lond. B306, 379-388.

Carroll, M. C., Campbell, R. D. & Porter, R. R. (1985a) Proc. Natl. Acad. Sci. USA 82, 521-525.

Carroll, M. C., Palsdottir, A., Belt, K. T. & Porter, R. R., (1985b) EMBO J. 4, 2547-2552.

Carroll, M. C., Belt, K. T., Palsdottir, A. & Yu, Y. (1985c) Immunol. Rev. 87, 39-60.

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

Casadaban, M. J. & Cohen, S. N. (1980) J. Mol. Biol. 138, 179-207.

Ceppellini, R., Curtoni, E. S., Mattiuz, P. L., Miggiano, V., Scudeller, G. & Serra, A. (1967) in Histocompatibility Testing 1967, pp 149-187 (Munksgaard, Copenhagen) Curtoni, E. S., Mattiuz, P. L. & Tosi, R. M. (eds.).

Cerami, A. & Beutler, B. (1988) Immunol. Today 9, 28-31.

Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. R. & Seidman, J. G. (1983) Proc. Natl. Acad. Sci. USA 80, 6947-6951.

Chaplin, D. D. (1985) Imm. Rev. 87, 61-80.

Chaplin, D. D., Galbraith, L. J., Seidman, J. G., White, P. C. & Parker, K. L. (1986) Proc. Natl. Acad. Sci. USA 83, 9601-9605.

Chimini, G., Pontarotti, P., Nguyen, C., Toubert, A., Boretto, J. & Jordan, B. R. (1988) EMBO J. 7, 395-400.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochem. 18, 5294-5299.

Chirico, W. J., Waters, M. G. & Blobel, G. (1988) Nature 332, 805-810.

Chu, G., Vollrath, D. & Davis, R. W. (1986) Science 234, 1582-1585.

Clark, S. M., Lai, E., Birren, B. W. & Hood, L. (1988) Science 241, 1203-1205.

Collins, F. S. & Weissman, S. M. (1984) Proc. Natl. Acad. Sci. USA 81, 6812-6816.

Collins, F. S., Drumm, M. L., Cole, J. L., Lockwood, W. K., Vande Woude, G. F. & Iannuzzi, M. C. (1987) Science 235, 1046-1049.

Collins, J. & Hohn, B. (1978) Proc. Natl. Acad. Sci. USA 75, 4242-4246.

Collins, T., Lapierre, L. A., Fiers, W., Strominger, J. L. & Pober, J. S. (1986) Proc. Natl. Acad. Sci. USA 83, 446-450.

Comings, D. E. (1978) Ann. Rev. Genet. 12, 25-46.

Constanzo, F., Colombo, M., Staempfli, S., Santoro, C., Marone, M., Frank, R., Delius, H. & Cortese, R. (1986) Nucl. Acids Res. 14, 721-736.

Cooper, D. N., Taggart, M. H. & Bird, A. P. (1983) Nucl. Acids Res. 11, 647-658.

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. & Schmidtke, J. (1985) Hum. Genet. 69, 201-205.

Cooper, D. N. & Youssoufian, H. (1988) Hum. Genet. 78, 151-155.

Coulson, A., Waterston, R., Kiff, J., Sulston, J. & Kohara, Y. (1988) Nature 335, 184-186.

Curman, B., Östberg, L., Sandberg, L., Malmheden-Eriksson, I., Stalenheim, G., Rask, L. & Peterson, P. A. (1975) Nature 258, 243-245.

Davis, M. M. & Bjorkman, P. J. (1988) Nature 334, 395-402.

David, V., Paul, P., Yaounq, J., Blayau, M., Fauchet, R., Cohen, D., Le Gall, J. Y. & Simon, M. (1987) Cytogenet. Cell Genet. 46, 604 (Abstract).

Dawkins, R. L., Christiansen, F. T., Kay, P. H., Garlepp, M., McCluskey, J., Hollingsworth, P. N. & Zilko, P. J. (1983) Immunol. Rev. 70, 5-22.

den Dunnen, J. T., Bakker, E., Klein Breteler, E. G., Pearson, P. L. & van Ommen, G. J. B. (1987) Nature 329, 640-642.

Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. (1988) Nature 332, 800-805.

Deutsch, J. M. (1988) Science 240, 922-924.

Dramanac, R., Petrovic, N., Glisin, V. & Crkvenjakov, R. (1986) Nucl. Acids Res. 14, 4691-4692.

Dulbecco, R. & Vogt, M. (1954) J. Exp. Med. 99, 167-182.

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

Dupont, B., Oberfield, S. E., Smithwick, E. M., Lee, T. D. & Levine, L. S. (1977) Lancet ii, 1309-1312.

Duquesnoy, R. J. & Trucco, M. (1988) Crit. Rev. Immunol. 8, 103-145.

Dynan, W. S. (1986) Trends in Genet. 2, 196-197.

Edwards, C. Q., Griffen, L. M., Dadone, M.M., Skolnick, M. H. & Kushner, J. P. (1986) Am. J. Hum. Genet. 38, 805-811.

Eisenbarth, G. S. (1986) N. Engl. J. Med. 314, 1360-1368.

Erlich, H., Lee, J. S., Petersen, J. L., Bugawan, T. & DeMars, R. (1986) Hum. Immunol. 16, 205-219.

Estivill, X., Farrall, M., Scambler, P. J., Bell, G. M., Hawley, K. M. F., Lench, N. J., Bates, G. P., Kruyer, H. C., Frederick, P. A., Stanier, P., Watson, E. K., Williamson, R. & Wainwright, B. J. (1987) Nature 326, 840-845.

Fangman, W. L. (1978) Nucl. Acids Res. 5, 653-665.

Feinberg, A. P. & Vogelstein, B. (1984) Anal. Biochem. 137, 266-267.

Fielder, A. H. L., Walport, M. J., Batchelor, J. R., Rynes, R. I., Black, C. M., Dodi, I. A. & Hughes, G. R. U. (1983) Br. Med. J. 286, 425-428.

Figueroa, F. & Klein, D., Tewarson, S. & Klein, J. (1982) J. Immunol. 129, 2089-2093.

Figueroa, F. & Klein, J. (1986) Immunol. Today 7, 78-81.

Figueroa, F., GÜnther, E. & Klein, J. (1988) Nature 335, 265-267.

Fischel-Ghodsian, N., Nicholls, R. D. & Higgs, D. R. (1987a) Nucl. Acids Res. 15, 9215-9225.

Fischel-Ghodsian, N., Nicholls, R. D. & Higgs, D. R. (1987b) Nucl. Acids Res. 15, 6197-6207.

Flaherty, L. (1981) in The Role of the Major Histocompatibility Complex in Immunobiology pp 33-57 (Garland, N. Y.) Dorf, M. D. (ed.).

Flavell, R. A., Allen, H., Burkly, L. C., Sherman, D. H., Waneck, G. L. & Widera, G. (1986) Science 233, 437-443.

Fourney, R. M., Miyakoshi, J., Day, R. S., III & Paterson, M. C. (1988) Focus 10, 5-7 (Bethesda Research Laboratories).

Francke, U. & Pellegrino, M. A. (1977) Proc. Natl. Acad. Sci. USA 74, 1147-1151.

Fu, S. M., Kunkel, H. G., Brusman, H. P., Allen, F. H. & Fotino, M. (1974) J. Exp. Med. 140, 1108-1111.

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

Gardiner, K., Laas, W. & Patterson, D. (1986) Som. Cell Mol. Genet. 12, 185-195.

Garoff, H. & Ansorge, W. (1981) Anal. Biochem. 115, 450-457.

Goate, A. M., Cooper, D. N., Hall, C., Leung, T. K. C., Solomon, E. & Lim, L. (1987) Hum. Genet. 75, 123-128.

Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A. & Nag, A. (1984) Science 224, 686-692.

Goodfellow, P. N., Jones, E. A., van Heyningen, V., Solomon, E., Bobrow, M., Miggiano, V. & Bodmer, W. F. (1975) Nature 254, 267-269.

Gorer, P. A. (1937) J. Pathol. Bacteriol. 44, 691-697.

Goundis, D. (1988) D. Phil. Thesis, Oxford (Submitted).

Goundis, D. & Reid, K. B. M. (1988) Nature 335, 82-85.

Gray, P. W., Aggarwal, B. B., Benton, C. V., Bringman, T. S., Henzel, W. J., Jarrett, J. A., Leung, D. W., Moffat, B., Ng, P., Svedersky, L.P., Palladino, M. A. & Nedwin, G. E. (1984) Nature 312, 721-724.

Gregersen, P. K., Moriuchi, T., Karr, R. W., Obata, F., Moriuchi, J., Maccari, J., Goldberg, D., Winchester, R. J. & Silver, J. (1986) Proc. Natl. Acad. Sci. USA 83, 9149-9153.

Grey, H. M., Kubo, R. T., Colon, S. M., Poulik, M. D., Cresswell, P., Springer, T., Turner, M. & Strominger, J. L. (1973) J. Exp. Med. 138, 1608-1612.

Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) Nucl. Acids Res. 10, 6715-6732.

Guillemot, F., Billault, A., Pourquie, O., Behar, G., Chausse, A-M., Zoorob, R., Kreibich, G. & Auffray, C. (1988) EMBO J. 7, 2775-2785.

Hardy, D. A., Bell, J. I., Long, E. O., Lindsten, T. & McDevitt, H. O. (1986) Nature 323, 453-455.

Harrison, G. S., Drabkin, H. A., Kao, F-T., Hartz, J., Hart, I. M., Chu, E. H. Y., Wu, B. J. & Morimoto, R. I. (1987) Som. Cell Mol. Genet. 13, 119-130.

Hauptmann, G., Grosshans, E., Heid, E., Mayer, S. & Basset, A. (1974) Annls. Derm. Syph. 101, 479-496.

Hauptmann, G., Goetz, J., Uring-Lambert, B. & Grosshauss, E. (1986) Prog. Allergy 39, 232-249.

Helms, C., Graham, M. Y., Dutchik, J. E. & Olson, M. V. (1985) DNA 4, 39-49.

Henrotte, J. G., Santarromana, M. & Pla, M. (1987) Immunogenetics 25, 408-410.

Henthorn, P. S., Mager, D. L., Huisman, T. H. J. & Smithies, O. (1986) Proc. Natl. Acad. Sci. USA 83, 5194-5198.

Hentze, M. W., Keim, S., Papadopoulos, P., O'Brien, S., Modi, W., Drysdale, J., Leonard, W. J., Harford, J. B. & Klausner, R. D. (1986) Proc. Natl. Acad. Sci. USA 83, 7226-7230.

Higashi, Y., Yoshioka, H., Yamane, M., Gotoh, O. & Fujii-Kuriyami, Y. (1986) Proc. Natl. Acad. Sci. USA 83, 2841-2845.

Hohn, B. & Murray, K. (1977) Proc. Natl. Acad. Sci. USA 74, 3259-3263.

Holmquist, G., Gray, M., Porter, T. & Jordan, J. (1982) Cell 31, 121-129.

Holzworth, G., McKee, C. B., Steiger, S., & Crater, G. (1987) Nucl. Acids Res. 15, 10031-10044.

Hood, L., Steinmetz, M. & Malissen, B. (1983) Ann. Rev, Immunol. 1, 529-568.

Hughes, A. L. & Nei, M. (1988) Nature 335, 167-170.

Hunt, C. & Morimoto, R. I. (1985) Proc. Natl. Acad. Sci. USA 82, 6455-6459.

Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in DNA cloning: a practical approach pp 49-78 (IRL Press, Oxford) Glover, D. M. (ed.).

Inoko, H. & Trowsdale, J. (1987) Nucl. Acids Res. 15, 8957-8962.

Inoko, H., Ando, A., Kawai, J., Trowsdale, J. & Tsuji, K. (1988) in press.

Isenman, D. E. & Young, J. R. (1984) J. Immunol. 132, 3019-3027.

Ish-Horowicz, D. & Burke, J. F. (1981) Nucl. Acids Res. 9, 2989-2998.

IUPAC-IUB Commission on Biochemical Nomenclature (1969) Biochem. J. 113, 1-4.

Jacob, C. O. & McDevitt, H. O. (1988) Nature 331, 356-358.

Jackson, D. A. & Cook, P. R. (1985) EMBO J. 4, 913-918.

Jain, S. K., Barrett, K. J., Boyd, D., Favreau, M. F., Crampton, J. & Drysdale, J. W. (1985) J. Biol. Chem. 260, 11762-11768.

Janeway, C. A., Jr. (1988) Nature 335, 208-210.

Jones, N. C., Rigby, P. W. J. & Ziff, E. B. (1988) Genes & Development 2, 267-281.

Jordan, B. R., Caillol, D., Damotte, M., Delovitch, T., Ferrier, P., Kahn-Perles, B., Kourilsky, F., Layet, C., Le Bouteiller, P., Lemonnier, F. A., Malissen, M., N'Guyen, C., Sire, J., Sodoyer, R., Strachan, T. & Trucy, J. (1985) Immunol. Rev. 84, 73-92.

Josse, J., Kaiser, A. D. & Kornberg, A. (1961) J. Biol. Chem. 236, 864-875.

Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) Trends in Biochem. Sci. 11, 20-23.

Kaufman, J. F., Auffray, C., Korman, A. J., Shakelford, D. A. & Strominger, J. (1984) Cell 36, 1-13.

Kemp, D. J., Corcoran, L. M., Coppel, R. L., Stahl, H. D., Bianco, A. E., Brown, G. V. & Anders, R. F. (1985) Nature 315, 347-350.

Kingsbury, D. T., Kasper, K. C., Stites, D. P., Watson, J. D., Hogan, R. N. & Prusiner, S. B. (1983) J. Immunol. 131, 491-496.

Klein, J. (1975) Biology of the Mouse Histocompatibility-2 Complex (Springer, N. Y.).

Klein, J. & Figueroa, F. (1986) CRC Crit. Rev. Immunol. 6, 295-386.

Klein, J., Figueroa, F. & Nagy, Z. A. (1983) Ann. Rev. Immunol. 1, 119-142.

Kobori, J. A., Strauss, E., Minard, K. & Hood, L. (1986) Science 234, 173-179.

Koller, B. H., Geraghty, D. E., Shimizu, Y., DeMars, R. & Orr, H. T. (1988) J. Immunol 141, 897-904.

Korenberg, J. R. & Ryowski, M. C. (1988) Cell 53, 391-400.

Kotwal, G. J. & Moss, B. (1988) Nature 335, 176-178.

Krangel, M. S., Biddison, W. E. & Strominger, J. L. (1983) J. Immunol. 130, 1856-1862.

Lachman, P. J., Grennan, D., Martin, A. & Demant, P. (1975) Nature 258, 242-243.

Lalande, M., Noolandi, J., Turmel, C., Rousseau, J. & Slater, G. W. (1987) Proc. Natl. Acad. Sci. USA 84, 8011-8015.

Lamb, J. R. & Rees, A. D. M. (1988) Brit. Med. Bull. 44, 600-610.

Lamm, L. U. & Olaisen, B. (1985) Cytogenet. Cell Genet. 40, 128-155.

Laskey, R. A. & Mills, A. D. (1977) FEBS Lett. 82, 314-316.

Lavia, P., Macleod, D. & Bird, A. (1987) EMBO J. 6, 2773-2779.

Law, S. K., Lichtenberg, N. A. & Levine R. P. (1979) J. Immunol. 123, 1388-1394.

Law, S. K. A., Dodds, A. W. & Porter, R. R. (1984) EMBO J. 3, 1819-1823.

Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P. & Parham, P. (1988) Nature 335, 268-271.

Lawrance, S. K., Smith, C. L., Srivastava, R. Cantor, C. R. & Weissman, S. M. (1987) Science 235, 1387-1390.

Lee, J. S., Trowsdale, J., Travers, P. J., Carey, J., Grosveld, F., Jenkins, J. & Bodmer, W. F. (1982) Nature 299, 750-752.

Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) Biochemistry 16, 4743-4751.

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

Lindsay, S. & Bird, A. P. (1987) Nature 327, 336-338.

Lipman, D. J. & Pearson, W. R. (1985) Science 227, 1435-1441.

Long, E. O., Wake, C. T., Gorski, J. & Mach, B. (1983) EMBO J. 2, 389-394.

Maeda, N., McEvoy, S. M., Harris, H. F., Huisma, T. H. J. & Smithies, O. (1986) Proc. Natl. Acad. Sci. USA 83, 7395-7399.

Malissen, M., Damotte, M., Birnbaum, D., Trucy, J. & Jordan, B. R. (1982) Gene 20, 485-489.

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

Marshall, W. H., Neugebauer, M., Baur, M. P. & Albert, E. D. (1984) in Histocompatibility Testing 1984 pp 313-316 (Springer-Verlag, Berlin).

Mauff, G., Alper, C. A., Awdeh, Z., Batchelor, J. R., Betrams, J., Bruun-Petersen, G., Dawkins, R. L., Démant, P., Edwards, J., Grosse-Wilde, G., Hauptmann, G., Klouda, P., Lamm, L., Mollenhauer, E., Nerl, C., Olaisen, B., O'Neill, G., Rittner, C., Roos, M. H., Skanes, V., Teisberg, P. & Wells, L. (1983) Immunobiology 164, 184-191.

Maxam, A. M. & Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.

McConnell, T. J., Talbot, W. S., McIndoe, R. A. & Wakeland, E. K. (1988) Nature 332, 651-654.

McDevitt, H. O. & Chinitz, A. (1969) Science 163, 1207-1208.

McDevitt, H. O., Deak, B. D., Shreffler, D. C., Klein, J., Stimpfling, J. H. & Snell, G. (1972) J. Exp. Med. 135, 1259-1278.

Meo, T., Krasteff, T. & Shreffler, D. C. (1975) Proc. Natl. Acad. Sci. USA 72, 4536-4540.

Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucl. Acids Res. 9, 309-321.

Messing, J. & Vieira, J. (1982) Gene 19, 269-276.

Messing, J. (1983) Meth. Enzymol. 101, 20-78.

Michaelson, J. (1981) Immunogenetics 13, 167-171.

Michiels, F. M., Burmeister, M. & Lehrach, H.(1987) Science 236, 1305-1308.

Minota, S. & Winfield, J. B. (1988) Arthrit. Rheum. 31, S13.

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

Morley, B. J. & Campbell, R. D. (1984) EMBO J. 3, 153-157.

Mues, G. I., Munn, T. Z. & Raese, J. D. (1986) J. Biol. Chem. 261, 874-877.

Müller, U., Jongeneel, C. V., Nedospasov, S. A., Lindahl, K. F. & Steimetz, M. (1987a) Nature 325, 265-267.

Müller, U., Stephan, D., Philippsen, P. & Steinmetz, M. (1987b) EMBO J. 6, 369-373.

Nedospasov, S. A., Shakov, A. N., Turetskaya, R. L., Mett, V. A., Georgiev, G. P., Dobrynin, V. N. & Korobko, V. G. (1985) Dokl. Akad. Nauk SSSR 285, 1487-1490.

Nedospasov, S. A., Hirt, B., Shakov, A. N., Dobrynin, V. N., Kawashima, E., Acolla, R. S. & Jongeneel, C. V. (1986) Nucl. Acids Res. 14, 7713-7725.

Nedwin, G. E., Naylor, S. L., Sakaguchi, A. Y., Smith, D., Jarret-Nedwin, J., Pennica, D., Goeddel, D. V. & Gray, P. W. (1985) Nucl. Acids Res. 13, 6361-6373.

New, M. I., Dupont, B., Grumback, K. & Levine, L. S. (1982) in The Metabolic Basis of Inherited Disease pp 973-1000 (McGraw-Hill, N. Y.) Stanbury, J. B., Wyngarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. eds.

Nicholls, R. D., Fischel-Ghodsian, N. & Higgs, D. R. (1987) Cell 49, 369-378.

O'Neill, G. J., Yang, S. Y. & Dupont, B. (1978a) Proc. Natl. Acad. Sci. USA 75, 5165-5169.

O'Neill, G. J., Yang, S. Y., Tegoli, J., Berger, R. & Dupont, B. (1978b) Nature 273, 668-670.

Okada, K., Boss, J. M., Prentice, H., Spies, T., Mengler, R. A., Auffray, C., Lillie, J., Grossberger, D. & Strominger, J. L. (1985) Proc. Natl. Acad. Sci. USA 82, 3410-3414.

Olaisen, B., Sakaguchi, A. Y. & Naylor, S. L. (1987) Cytogenet. Cell Genet. 46, 147-169.

Old, L. J. (1985) Science 230, 630-632.

Old, L. J. (1987a) Nature 326, 330-331.

Old, L. J. (1987b) Nature 330, 602-603.

Oldstone, M. B. A. (1987) Cell 50, 819-820.

Oliff, A. (1988) Cell 54, 141-142.

Oohira, T., Nagata, N., Akaboshi, I., Matsuda, I. & Naito, S. (1985) Hum. Genet. 70, 341-343.

Orr, H. T. & DeMars, R.(1983) Nature 302, 534-536.

Palsdottir, A., Fossdal, R., Arnason, A., Edwards, J. H. & Jensson, O. (1987a) Immunogenetics 25, 299-304.

Palsdottir, A., Arnason, A., Fossdal, R. & Jenson, O. (1987b) Hum. Genet. 76, 220-224.

Parham, P., Lomen, C. E., Lawlor, D. A., Ways, J. P., Holmes, N., Coppin, H. L., Salter, R. D., Wan, A. M. & Ennis, P. D. (1988) Proc. Natl. Acad. Sci. USA 85, 4005-4009.

Parker, R. C. & Seed, B. (1980) Meth. Enzymol. 65, 358-363.

Partanen, J. & Koskimies, S. (1986) Hum. Hered. 36, 269-275.

Paul, N. L. & Ruddle, N. H. (1988) Ann. Rev. Immunol. 6, 407-438.

Paul, P., Fauchet, R., Boscher, M. Y., Sayagh, B., Masset, M., Medrignac, G., Dausset, J. & Cohen, D. (1987) Proc. Natl. Acad. Sci. USA 84, 2872-2876.

Pelham, H. R. B. (1986) Cell 46, 959-961.

Pelham, H. R. B. (1988) Nature 332, 776-777.

Pennica, D., Nedwin, G. E., Hayflick, J. S., Seeburg, P. H., Derynck, R., Palladino, M. A., Kohr, W. J., Aggarwal, B. B. & Goeddel, D. V. (1984) Nature 312, 724-729.

Pleogh, H. L., Orr, H. T. & Strominger, J. L. (1981) Cell 24, 287-299.

Pontarotti, P., Chimini, G., Nguyen, C., Boretto, J. & Jordan, B. R. (1988) Nucl. Acids Res. 16, 6767-6778.

Poustka, A., Pohl, T. M., Barlow, D. P., Frischauf, A-M, & Lehrach, H. (1987) Nature 325, 353-355.

Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. (1987) Science 237, 286-289.

Pujol-Borrell, R., Todd, I., Doshi, M., Bottazo, G. F., Sutton, R., Gray, D., Adolf, G. R. & Feldmann, M. (1987) Nature 326, 304-306.

Pytela, R. (1988) EMBO J. 7, 1371-1378.

Ragoussis, J., van der Bliek, A., Trowsdale, J. & Ziegler, A. (1986) FEBS Lett. 204, 1-4.

Ragoussis, J., Bloemer, K., Weiss, E. H. & Ziegler, A. (1988) Immunogenetics 27, 66-69.

Rappold, G. A., Stubbs, L., Labeit, S., Crkvenjakov, R. B. & Lehrach, H. (1987) EMBO J. 6, 1975-1980.

Raum, D., Glass, D., Carpenter, C. B., Schur, P. H. & Alper, C. A. (1979) Am. J. Hum. Genet. 31, 35-42.

Raum, D., Awdeh, Z. L., Glass, D., Yunis, E. & Alper, C. A. (1981) Immunogenetics 12, 473-483.

Raum, D., Awdeh, Z. L., Anderson, J., Strong, L., Grandos, J., Teran, L., Giblett, E., Yunis, E. & Alper, C. A. (1984) Am. J. Hum. Genet. 36, 72-79.

Reed, K. C. & Mann, D. A. (1986) Nucl. Acids Res. 13, 7207-7221.

Reid, K. B. M. (1986) Essays in Biochem. 22, 27-68.

Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986) Immunol. Today 7, 230-234.

Reinsmoen, N. L., Friend, P. S., Miller, W. V., Burgdorf, A., Giblett, E.R. & Yunis, E. J. (1977) Nature 267, 276-278.

Robinson, M. A., Carroll, M. C., Johnson, A. H., Hartzman, R. J., Belt, K. T. & Kindt, T. J. (1985) Immunogenetics 21, 143-152.

Robson, E. B. & Lamm, L. U. (1984) Cytogenet. Cell Genet. 37, 47-70.

Rodrigues, N. R., Dunham, I., Yu, C. Y., Carroll, M. C., Porter, R. R. & Campbell, R. D. (1987) EMBO J. 6, 1653-1661.

Rollini, P., Mach, B. & Gorski, J. (1985) Proc. Natl. Acad. Sci. USA 82, 7197-7201.

Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1986) Nature 324, 163-166.

Samallow, P. B., Ford, A. L., Kunz, H. W. & Gill, T. J., III, (1987) Immunogenet. 26, 188-189.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sanger, F. & Coulson, A. R. (1978) FEBS Lett. 87, 107-110.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) J. Mol. Biol. 143, 161-178.

Sargent, C. A. (1988) D. Phil. Thesis, Oxford (submitted).

Sargent, C. A., Dunham, I., Trowsdale, J. & Campbell, R. D. (1988) Manuscript submitted.

Scharf, S. J., Friedmann, A., Brautbar, C., Szafer, F., Steinman, L., Horn, G., Gyllenstern, U. & Erlich, H. A. (1988) Proc. Natl. Acad. Sci. USA 85, 3504-3508.

Schendel, D. J., O'Neill, G. J., & Wank, R. (1984) Immunogenetics 20, 23-31.

Schneider, P. M., Carroll, M. C. , Alper, C. A., Rittner, C., Whitehead, A. S., Yunis, E. J. & Colten, H. R. (1986) J. Clin. Invest. 78, 650-657.

Schwartz, D. C. & Cantor, C. R. (1984) Cell 37, 67-75.

Serwer, P. (1980) Biochemistry 19, 3001-3004.

Serwer, P. (1981) Anal. Biochem. 112, 351-356.

Serwer, P. (1987) Electrophoresis 8, 301-304.

Shanahan, F. & Weinstein, W. M. (1988) N. Engl. J. Med. 319, 782-783.

Sharpe, P. A., Sugden, B. & Sambrook, J. (1973) Biochemistry 12 3055-3063.

Shaw, S., Kavathas, P., Pollack, M. S., Charmot, D. & Mawas, C. (1981) Nature 293, 745-747.

Shevach, E. M. & Rosenthal, A. S. (1973) J. Exp. Med. 138, 1213-1229.

Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. (1983) J. Exp. Med. 158, 303-316.

Shirai, T., Yamagushi, H., Ito, H., Todd, C. W. & Wallace, R. B. (1985) Nature 313, 803-806.

Shreffler, D. C. (1964) Genetics 49, 973-978.

Silva, A. J. & White, R. (1988) Cell 54, 145-152.

Sim, E., Gill, E. W. & Sim, R. B. (1984) Lancet ii, 422-424.

Sinha, A. A., Brautbar, C., Szafer, F., Friedmann, A., Tzfoni, E., Todd, J. A., Steinman, L. & McDevitt, H. O. (1988) Science 239, 1026-1029.

Smeaton, I., Summers, C. W., Harris, R. & Strachan, T. (1987) Immunogenetics 25, 179-183.

Smith, C. L., Matsumoto, T., Niwa, O., Klco, S., Fan, J-B., Yanagida, M. & Cantor, C. R. (1987a) Nucl. Acids Res. 15, 4481-4489.

Smith, D. I., Golembieski, W., Gilbert, J. D., Kizyma, L. & Miller, O. J. (1987b) Nucl. Acids Res. 15, 1173-1184.

Smith, H. O. & Nathans, D. (1973) J. Mol. Biol. 81, 419-423.

Snell, R. G. & Wilkins, R. J. (1986) Nucl. Acids Res. 14, 4401-4406.

Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.

Southern, E. M. (1979) Meth. Enzymol. 68, 152-176.

Southern, E. M., Anand, R., Brown, W. R. A. & Fletcher, D. S. (1987) Nucl. Acids Res. 15, 5925-5943.

Spies, T., Sorrentino, R., Boss, J. M., Okada, K. & Strominger, J. L. (1985) Proc. Natl. Acad. Sci. USA 82, 5165-5169.

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

Srivastava, R., Duceman, B. W., Biro, P. A., Sood, A. K. & Weissman, S. M. (1985) Immunol. Rev. 84, 93-121.

Srivastava, R., Chorney, M. J., Lawrance, S. K., Pan, J., Smith, Z., Smith, C. L. & Weissman S. M. (1987) Proc. Natl. Acad. Sci. USA 84, 4224-4228.

Staden, R. (1982a) Nucl. Acids Res. 10, 2951-2961.

Staden, R. (1982b) Nucl. Acids Res. 10, 4731-4751.

Staden, R. (1984) Nucl. Acids Res. 12, 521-538.

Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R. A., Dastoornikoo, G. R., Stephan, D., Gibb, E. & Romaniuk, R. (1984) EMBO J. 3, 2995-3003.

Steinmetz, M., Stephan, D. & Lindahl, K. F. (1986) Cell 44, 895-904.

Strachan, T., Dodge, A. B., Smillie, P., Dyer, P. A., Sodoyer, R., Jordan, B. R. & Harris, R. (1986) Immunogenetics 23, 115-120.

Strachan, T. (1987) Brit. Med. Bull. 43, 1-14.

Stroyanowski, I., Clark, S., Henderson, L. A., Hood, L., McMillan, M. & Forman, J. (1985) J. Immunol 135, 2160-2166.

Swartz, M. N., Trautner, T. A. & Kornberg, A. (1962) J. Biol. Chem. 237, 1961-1967.

Tabor, S. & Richardson, C. C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771.

Tan E. M. (1982) Adv. Immunol. 33, 167-240.

Teisberg, P., Alesson, I., Olaisen, B., Gedde-Dahl, T., Jr. & Thorsby, E.
(1976) Nature 264, 253-254.

Teppo, A. M. & Maury, C. P. J.. (1987) Clin. Chem. 33, 2024-2027.

Termijtelen, A., Meera Khan, P., Shaw, S. & van Rood, J. J. (1983) Immunogenetics 18, 503-512.

Thein, S. L. & Wallace, R. B. (1986) in Human Genetic Disease; A Practical Approach pp 33-50 (IRL Press, Oxford) Davies, K. E. (ed.).

Tiwari, J. L. & Terasaki, P. I. (1985) HLA and Disease Associations (Springer, N. Y.).

Todd, J. A., Bell, J. I. & McDevitt, H. O. (1987) Nature 329, 599-604.

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. & McDevitt, H. O. (1988a) Science 240, 1003-1009.

Todd, J. A., Bell, J. I. & McDevitt, H. O. (1988b) Trends in Genet. 4, 129-134.

Tokunaga, K., Saueracker, G., Kay, P. H., Christiansen, F. T., Anand, R. & Dawkins, R. L. (1988) J. Exp. Med. 168, 933-940.

Toniolo, D., D'Urso, M., Martini, G., Persico, M., Tufano, V., Battistuzzi, G. & Luzzatto, L. (1984) EMBO J. 3, 1987-1995.

Tonnelle, C., DeMars, R. & Long, E. O. (1985) EMBO J. 4, 2839-2847.

Tosi, M., Lévi-Strauss, M., Georgatsou, E., Amor, M. & Meo, T. (1985) Immunol. Rev. 87, 151-183.

Townsend, A. R. M., Rothbard, J., Gotch, F. M., Bahadur, G., Wraith, D. & McMichael, A. J. (1986) Cell 44, 959-968.

Travers, P. & McDevitt, H. O. (1987) in The Antigens, Vol. VII pp 147-211 (Academic Press, N. Y.) Sela, M. ed.

Trowsdale, J., Lee, J., Carey, J., Grosveld, F., Bodmer, J. & Bodmer, W. F. (1983) Proc. Natl. Acad. Sci. USA 80, 1972-1976.

Trowsdale, J., Kelly, A., Lee, J., Carson, S., Austin, P. & Travers, P. (1984a) Cell 38, 241-249.

Trowsdale, J., Lee, J., Kelly, A., Carey, J., Jenkins, J., Travers, P. & Bodmer, W. F. (1984b) Mol. Biol. Med. 2, 53-61.

Trowsdale, J. (1987) Brit. Med. Bull. 43, 15-36.

Tsuge, I., Shen, F-W., Steinmetz, M. & Boyse, E. A. (1987) Immunogenetics 26, 378-380.

Tykocinski, M. L. & Max, E. E. (1984) Nucl. Acids Res. 12, 4385-4396.

Uematsu, Y., Keifer, H., Schulze, R., Fischer-Lindahl, K. & Steinmetz, M. (1986) EMBO J. 5, 2123-2129.

Van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R. & Borst, P. (1984) Cell 37, 77-84.

van Eden, W., Thole, J. E. R., van der Zee, R., Noordzij, A., van Embden, J. D. A., Hensen, E. J. & Cohen, I. R. (1988) Nature 331, 171-173.

VandeBerg, J. L., Bittner, G. N., Meyer, G. S., Kunz, H. W. & Gill, T. J. (1981) J. Immunogenet. 8, 239-242.

van Ommen, G. J. B. & Verkerk, J. M. H. (1986) in Human Genetic Disease: A Practical Approach pp 113-133 (IRL Press, Oxford) Davies, K. E. (ed.).

Voellmy, R., Ahmed, A., Schiller, P., Bromley, P. & Rungger, D. (1985) Proc. Natl. Acad. Sci. USA 82, 4949-4953.

Vollrath, D. & Davis, R. W. (1987) Nucl. Acids Res. 15, 7865-7876.

Vollrath, D., Nathans, J. & Davis, R. W. (1988) Science 240, 1669-1672.

Wahl, G. M., Stern, M. & Stark, G. R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.

Waterbury, P. G. & Lane, M. J. (1987) Nucl. Acids Res. 15, 3930.

Watson, C. J. & Jackson, J. F. (1985) in DNA cloning: a practical approach pp 79-88 (IRL Press, Oxford) Glover, D. M. (ed.).

Watts, J. H., Gariépy, J., Schoolnik, G. K. & McConnell, H. N. (1985) Proc. Natl. Acad. Sci. USA 82, 5480-5484.

Weinberg, R. A. (1988) Bioessays 9, 91-92.

Weislander, L. (1979) Anal. Biochem. 98, 305-309.

Welch, W. J. & Suhan, J. P. (1986) J. Cell Biol. 103, 2035-2052.

White, P. C., New, M. I. & Dupont, B. (1984) Proc. Natl. Acad. Sci. USA 81, 1986-1990.

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

White, P. C., New, M. I. & Dupont, B. (1986) Proc. Natl. Acad. Sci. USA 83, 5111-5115.

Whitehead, A. S., Colten, H. R., Chang, C. C. & DeMars, R. (1985) J. Immunol. 134, 641-643.

Wiliams, A. F. & Barclay, A. N. (1988) Ann. Rev. Immunol. 6, 381-405.

Williams, T., Yon, J., Huxley, C. & Fried, M. (1988) Proc. Natl. Acad. Sci. USA 85, 3527-3530.

Wilton, A. N. & Charlton, B. (1986) Immunogenetics 24, 79-83.

Womack, J. F. & David, C. S. (1982) Immunogenet. 16, 177-180.

Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, H. R. (1982) Proc. Natl. Acad. Sci. USA 79, 5661-5665.

World Health Organisation (1968) W. H. O. Bull. 39, 935-938.

World Health Organisation (1981) W. H. O. Bull. 59, 489-491.

Wu, B., Hunt, C. & Morimoto, R. (1985) Mol. Cell Biol. 5, 330-341.

Wu, B. J., Williams, G. T. & Morimoto, R. I. (1987b) Proc. Natl. Acad. Sci. USA 84, 2203-2207.

Wu, L-C., Morley, B. J. & Campbell, R. D. (1987a) Cell 48, 331-342.

Young, D., Lathigra, R., Hendrix, R., Sweetster, D. & Young, R. A. (1988) Proc. Natl. Acad. Sci. USA 85, 4267-4270.

Yu, C. Y., Belt, K. T., Giles, C. M., Campbell, R. D. & Porter, R. R. (1986) EMBO J. 5, 2873-2881.

Yu, C. Y. & Campbell, R. D. (1987) Immunogenetics 25, 383-390.

Yunis, E. J. & Amos, D. B. (1971) Proc. Natl. Acad. Sci. USA 68, 3031-3035.

Yunis, E. J., Awdeh, Z., Johnson, A., Suciu-Foca, N., Robinson, M. A., Hartzman, R., Raum, D., Fleischnick, E. & Alper, C. A. (1985) Immunogenetics 21, 25-31.

Zinkernagel, R. M. & Doherty, P. C. (1974) Nature 248, 701-702.

Zinkernagel, R. M. & Doherty, P. C. (1979) Adv. Immunol. 27, 51-177.

Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 7237–7241, October 1987
Immunology

Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis

(HLA/gene map/tumor necrosis factor/cosmid cloning)

IAN DUNHAM*, CAROLE A. SARGENT*, JOHN TROWSDALE†, AND R. DUNCAN CAMPBELL*

*Medical Research Council Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, United Kingdom; and †Imperial Cancer Research Fund, Lincoln's Inn Fields, London, WC2A 3PX, United Kingdom

Communicated by J. H. Humphrey, June 23, 1987 (received for review June 4, 1987)

ABSTRACT **Pulsed-field gel electrophoresis and "cosmid walking" have been used to establish a molecular map of the human major histocompatibility complex (MHC). We have isolated ≈230 kilobases (kb) of genomic DNA in overlapping cosmid clones covering the genes for the second and fourth components of complement (C2 and C4, respectively), factor B, and steroid 21-hydroxylase, and ≈82 kb of genomic DNA surrounding the genes for the tumor necrosis factors α and β. Single-copy hybridization probes isolated from these cosmid clusters and probes for the known MHC gene loci were hybridized to Southern blots of genomic DNA that had been digested with infrequently cutting restriction endonucleases and separated on pulsed-field gels. The data obtained allowed the construction of a long-range genomic restriction map and indicated that the MHC spans 3800 kb. This map orients the MHC class III gene cluster with respect to the DR subregion; the C2 gene is on the telomeric side of the 21-hydroxylase B gene. In addition we have defined the positions of the genes for the tumor necrosis factors α and β in the human MHC. Genes for the α chain of DR and 21-hydroxylase B are separated by at least 300 kb, while the distance between the genes for C2 and tumor necrosis factor α is 390 kb. The HLA-B locus lies ≈250 kb on the telomeric side of the tumor necrosis factor genes.**

The human major histocompatibility complex (MHC) is located on the short arm of chromosome 6 in the distal portion of the 6p21.3 band (1). It consists of three major linked gene clusters. The class I and class II regions each encode highly polymorphic families of cell-surface glycoproteins involved in immune regulation. The class I loci consist of at least 17 highly related genes (2) that include those encoding the classical transplantation antigens (HLA-A, -B, and -C). The class II region (HLA-D) is arranged into four subregions DP, DZ/DO, DQ, and DR, each containing at least one α- and β-chain pair of genes (3). The class III region contains the genes encoding the serum complement proteins factor B and the second and fourth components of complement (C2 and C4, respectively), as well as two copies of the steroid 21-hydroxylase (21-OHase) genes that are closely linked to the two C4 loci, *C4A* and *C4B* (4, 5). Analysis of recombinant HLA haplotypes in family studies has established that the class I loci are telomeric to the class II genes. Within the class II region the DP subregion maps on the centromeric side of DQ and DR. Located between the *HLA-DR* and *HLA-B* loci are the class III genes, but the orientation and distance of these genes relative to the class I and class II genes have not been determined. The possibility that a number of other genes may reside within the MHC was exemplified by the finding that the genes for the tumor necrosis factors α and β (TNFα and TNFβ, respectively) are linked to the HLA genes (6) and map close to the *H-2D* region in mouse (7).

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.

Clusters of overlapping cosmid clones have been isolated from the subregions of the class II loci (3) and from the class III region (4) but have not yet been linked. The technique of pulsed-field gel electrophoresis (PFGE) (8, 9), which allows the resolution of DNA fragments >2 megabases, in combination with restriction enzymes that cut rarely in the mammalian genome, has been used for long-range restriction-site mapping of the MHC in man (10–12) and in mouse (13). Using PFGE, Hardy *et al.* (10) determined a map for the human class II region and established the order of the HLA-D subregions. Physical linkage between the human class II and class III loci has also been demonstrated using this technique (11, 12). In mouse the orientation and molecular map position of the complement gene cluster have been established (13).

We have used PFGE, restriction enzymes that cut genomic DNA infrequently, and Southern blotting to derive a map of ≈4000 kb of DNA encompassing the human MHC. Making use of single-copy hybridization probes generated by "cosmid walking," we have orientated the class III loci with respect to the gene for the α chain of DR (DRα), the C2 gene being on the telomeric side of the 21-OHase B gene. In addition, we have defined the position of the TNFα and, by inference, TNFβ genes in the MHC. Our mapping indicates that the DRα and 21-OHase B genes are separated by at least 300 kb, whereas the distance between the C2 gene and the TNFα gene is 390 kb. In addition the HLA-B locus lies 250 kb from the telomeric side of the TNF genes.

MATERIALS AND METHODS

Preparation of DNA. Genomic DNA for PFGE analysis and for construction of cosmid libraries was prepared from a HLA-homozygous lymphoblastoid cell line, HLA and complement typed as, A2, B7, DR2, C2 C, Bf S, C4A 3, C4B QO. For PFGE, high molecular weight genomic DNA in agarose blocks was prepared essentially as described by van Ommen and Verkerk (14), except that 2% (wt/vol) low-geling temperature agarose (Sigma) that had been treated with DEAE-cellulose (15) was mixed 1:1 (vol/vol) with the cells suspended at 2×10^7 cells per ml of isotonic phosphate-buffered saline. High molecular weight genomic DNA for the construction of cosmid libraries was prepared as described by Bell *et al.* (16).

Construction of Cosmid Libraries and Cosmid Walking. Cosmid libraries were prepared according to the method of Steinmetz *et al.* (17). Insert fragments of 35–50 kb, generated from 300 μg of genomic DNA by partial digestion with *Mbo* I, were ligated into the *Bam*HI-digested cosmid vector pDVcos, and the DNA was packaged according to the method of Sternberg *et al.* (18), prior to transduction of *Escherichia coli* strain NM554. Approximately 10^6 colonies

Abbreviations: MHC, major histocompatibility complex; C2 and C4, second and fourth components of complement, respectively; 21-OHase, 21-hydroxylase; TNFα and -β, tumor necrosis factors α and β, respectively; PFGE, pulsed-field gel electrophoresis; DRα, DQα, and DPα, HLA-DR, -DQ, and -DP α chains.

were screened by hybridization according to Grosveld *et al.* (19). Cosmid DNA inserts were characterized by restriction enzyme mapping and Southern blotting (20), and regions of nonrepetitive DNA were detected by probing Southern blots with radiolabeled genomic DNA. Linkage of single-copy sequences with preexisting probes was confirmed by hybridization to common fragments on standard genomic Southern blots (20) and on Southern blots from PFGE.

Restriction Enzyme Digests and PFGE Analysis. Agarose blocks containing ≈5 μg of DNA were washed three times in 10 ml of 10 mM Tris·HCl, pH 8.0/0.1 mM EDTA/0.1 mM phenylmethylsulfonyl fluoride for 30 min at 4°C and were then equilibrated with restriction buffer (22) at 4°C for 30 min prior to digestion. Restriction enzyme digestions were carried out according to the supplier's recommendations with 20 units of enzyme for 3 hr. All buffers containing NaCl concentrations ≥50 mM also contained 5 mM spermidine. After digestion, the blocks were incubated in 0.5 M EDTA, 10 mM Tris·HCl, 1% lauroyl sarcosine (pH 9.5), and proteinase K at 1 mg/ml for 2 hr at 50°C prior to loading.

The PFGE apparatus used was provided by the MRC Mammalian Genome Unit (Edinburgh) (21). Digested samples were electrophoresed on 20 cm × 20 cm × 0.5 cm 1.5% agarose (Sigma) gels in 20 mM Tris acetate/1 mM EDTA, pH 8.5, for 22 hr at 12°C and 330 V with pulse intervals up to 70 s. Molecular weight markers were intact yeast chromosomes (*Saccharomyces cerevisiae* strain X2180-1B) prepared in agarose blocks (14) or concatemers of λ *c*I857S7 DNA (22). After ethidium bromide staining and depurination with 0.25 M HCl, DNA fragments were blotted onto nylon membranes (GeneScreen*Plus*, New England Nuclear). Hybridization and high-stringency washing (30 mM NaCl/3 mM sodium citrate/1% $NaDodSO_4$, 65°C) of filters was carried out following standard procedures (4). Autoradiography was between two intensifying screens at −70°C for 1–5 days. Filters were stripped of probe by washing twice in 2 mM Tris·HCl, pH 7.4/1 mM EDTA/1% $NaDodSO_4$ at 80°C for 1 hr.

Probes. The probes used were a DRα cDNA (23), a DQα-chain (DQα) cDNA (10), a DPα-chain (DPα) cDNA (3), a DRβ-chain (DRβ) cDNA (10), a DPβ-chain (DPβ) genomic probe (3), a 21-OHase genomic probe (24), a class I cDNA (25), HLA-B (B250) and HLA-C (C250)-specific genomic probes (26), factor B cDNA (4), C4 cDNA (4), and a TNFα genomic probe (a gift from Mark Bodmer) (27). Probe J, a 0.9-kb *Bam*HI–*Xho* I genomic fragment, and probe L, a 1.4-kb *Bam*HI genomic fragment, were isolated from cosmid clones as described above. All probes were labeled by random hexanucleotide priming (28).

RESULTS

Isolation of Overlapping Cosmid Clones from the Class III Region. To construct a restriction map of the MHC in man, genomic DNA from a HLA-homozygous cell line (HLA type—A2, B7, DR2, C2 C, Bf S, C4A 3, C4B QO) was used to minimize mapping problems caused by possible haplotype-specific restriction fragment length polymorphisms. The cell line used has only single C4 and 21-OHase loci on each copy of chromosome 6 (29, 30), which also serves to simplify restriction site mapping. Approximately 230 kb of genomic DNA encompassing the C2, factor B, C4, and 21-OHase B genes was isolated in a series of overlapping cosmid clones (Fig. 1*A*). This cluster of genomic clones was mapped with restriction enzymes having 6- or 8-base-pair recognition sites and containing one or two CpG dinucleotides. Since CpG is known to be underrepresented in the human genome (31), these enzymes are consequently of use in construction of

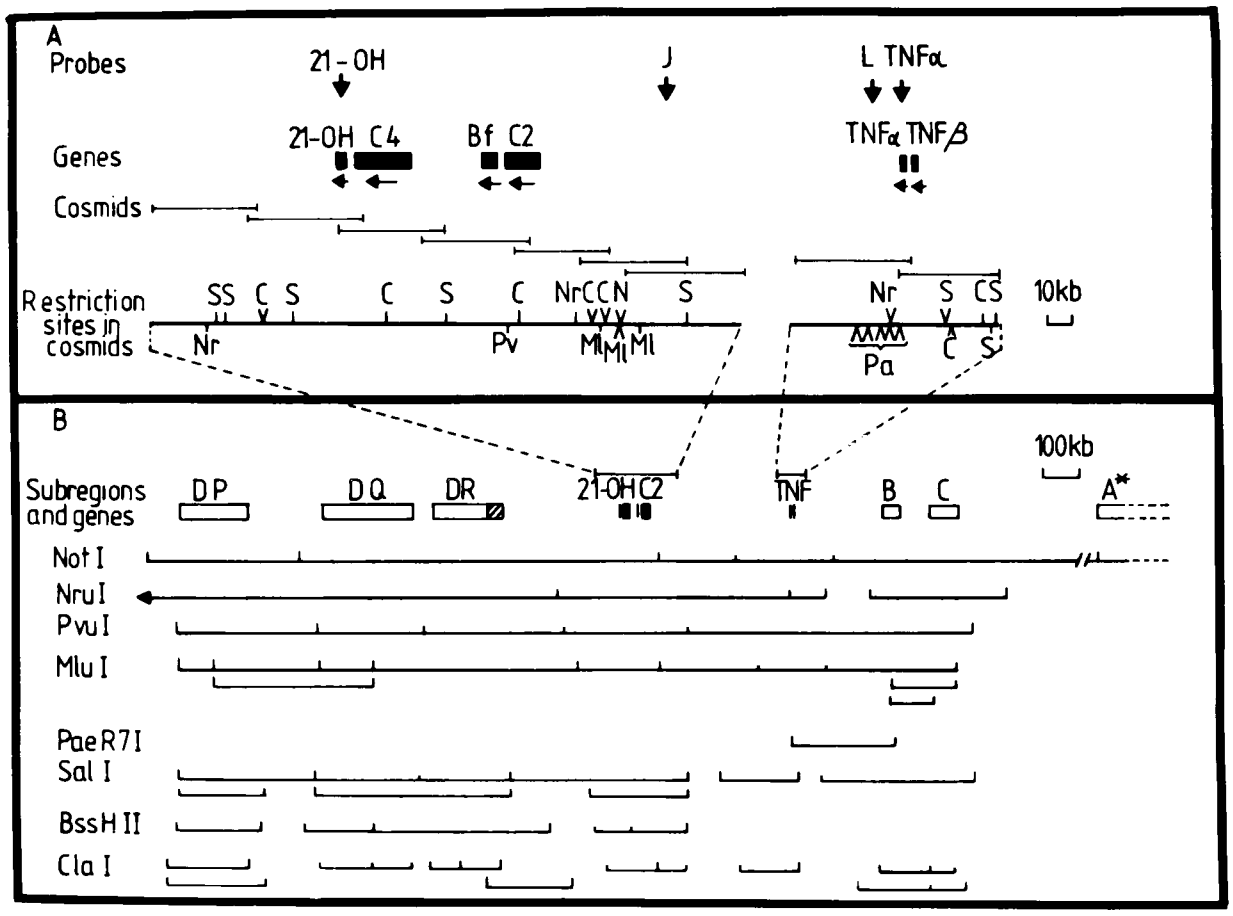


FIG. 1. Molecular map of the human MHC. (*A*) Overlapping cosmid clones from the complement/21-OHase and TNF regions. The locations of genes are shown by solid boxes (Bf, factor B; 21-OH, 21-OHase), and the 5′ → 3′ orientation is indicated by horizontal arrows. Vertical arrows indicate the positions of the probes used in PFGE analysis. The positions of infrequently cutting restriction endonuclease sites known to cut in genomic DNA are indicated by arrowheads; other sites are indicated by vertical bars. C, *Cla* I; Ml, *Mlu* I; N, *Not* I; Nr, *Nru* I; Pa, *Pae*R7I; Pv, *Pvu* I; S, *Sal* I. (*B*) Restriction map obtained by Southern blot analysis of genomic DNA separated by PFGE after single and double digests. Solid boxes, exact position of genes; open boxes, limits of the region hybridizing to the probes used; cross-hatched box, probable position of the DRα gene. For convenience the 950-kb *Not* I fragment that hybridizes with the B- and C-locus-specific probes is not to scale, and A*, the 675-kb and 400-kb *Not* I fragments that hybridize with the class I cDNA probe (see Fig. 2) and lie on the telomeric side of the C locus, are not shown. In some cases partial digests were obtained, as represented by two sets of data for some enzymes.

long-range restriction maps in conjunction with PFGE (14, 22). Examination of these restriction site data (Fig. 1*A*) reveals a striking cluster of sites for infrequently cutting enzymes (three *Mlu* I, two *Cla* I, one *Not* I, and one *Nru* I) in a 25-kb region located ≈25 kb from the transcriptional start site of the C2 gene. A single-copy genomic hybridization probe (probe J, Fig. 1*A*) was isolated from the cosmid DNA distal to this cluster of restriction sites and the C2 gene.

To establish a long-range restriction map of the human MHC, high molecular weight HLA-homozygous cell line DNA in agarose blocks was digested with a range of suitable restriction enzymes, alone or in double-digest combinations. The DNA fragments were separated by PFGE and transferred to nylon membranes. These filters were hybridized successively with a panel of ^{32}P-labeled cDNA and genomic probes specific for the class I, class II, and class III regions of the MHC (see Fig. 1).

Orientation of the Complement Genes in the Class III Region. Hybridization of *Not* I-digested DNA with DQα, DRα, and 21-OHase probes identified a common 980-kb *Not* I fragment that links the DQ and DR subregions with the complement loci (refs. 11 and 12 and Fig. 2). This fragment is digested by *Nru* I to yield a fragment of 700 kb with the DRα probe and a fragment of 280 kb with the 21-OHase probe (Fig. 2). Probe J from the class III cosmid cluster, however, hybridizes to a *Not* I fragment of 210 kb. Since the exact position of the *Not* I site is known from the cosmid map (Fig. 1*A*), these data establish the orientation of complement genes relative to the class II genes; the C2 gene is on the telomeric side of the 21-OHase B gene closest to the DR locus.

Hybridization of probes DQα, DRα, 21-OHase, and J to *Pvu* I, *Mlu* I, and *Sal* I digests is consistent with the map shown in Fig. 1*B*. In the *Sal* I digest the 21-OHase probe hybridizes to a fragment of 280 kb and a partial fragment of 490 kb (data not shown). The *Not* I/*Sal* I double digest positions a *Sal* I site 85 kb on the telomeric side of the *Not* I site, and this was confirmed by hybridization of the blot with probe J. Hybridization of the same blot with the DRα probe revealed no common fragments. Thus, the minimum distance between the 21-OHase B and the DR subregion is 300 kb.

Further single- and double-digestion data allowed the limits of the DP, DQ, and DR subregions to be defined (Fig. 1*B*). The map of the class II region shows a number of striking similarities to that published by Hardy *et al.* (10) and confirms the organization of the class II subregions. Although the order of the α and β genes within a subregion was not determined, the approximate position of the DRα gene was defined from the result of *Cla* I and *Cla* I/*Sal* I digests. A 290-kb partial *Cla* I fragment, which hybridized with the DRα probe but not the DRβ probe, was reduced by *Sal* I to 60 kb (Fig. 1*B*). This would suggest that the DRα gene lies 300–360 kb from the 21-OHase B gene.

Mapping of the TNFα Gene in the MHC. The genes for TNFα and -β have been linked to the human MHC (6), and in mouse they have been mapped ≈70 kb from the H-2D region, the murine equivalent of HLA-B (7). Hybridization of the TNFα probe to *Not* I- and *Nru* I-digested DNA failed to reveal any bands in common with the other probes (Fig. 2). However, in the *Pvu* I digest, a 780-kb fragment that also hybridized with the HLA-B and -C locus probes was identified with the TNFα probe (Fig. 3*B*).

To establish the position of the TNFα gene relative to the known MHC loci, a cluster of overlapping cosmid clones covering ≈82 kb of genomic DNA surrounding the TNFα gene was isolated and characterized (Fig. 1*A*). The position of the TNFα gene was located by restriction mapping and Southern blot analysis. The position of the TNFβ gene was inferred by comparison of the map with that of Nedospasov *et al.* (32). These cosmids were mapped for infrequently cutting restriction enzymes. A single *Nru* I site was found 2.5 kb from the 5′ end of the TNFα gene (Fig. 1*A*). A single-copy hybridization probe (probe L, Fig. 1*A*) was isolated from the cloned cosmid DNA distal to the *Nru* I site and the TNFα gene. When hybridized to *Nru* I-digested genomic DNA, this probe detected the same 640-kb *Nru* I fragment that contains the complement genes (Fig. 2). This result confirmed the linkage of the complement genes and the TNFα gene and established that the C2 gene lies 390 kb on the centromeric side of the TNFα gene. This also indicates that the TNFβ gene lies on the telomeric side of the TNFα gene.

Further single and double digests were carried out to estimate the distance between the TNFα and *HLA-B* loci. A 290-kb *Pae*R7I fragment was found to hybridize with both probes (Fig. 4*A*), but not probe L. The results from a series of double digests (Fig. 4*A*) allowed the map in Fig. 1*B* to be constructed showing that the TNFα and -β genes lie ≈250 kb on the centromeric side of the class I genes. Restriction mapping of the cosmid DNA for *Pae*R7I revealed five sites (Fig. 1*A*), one of which lies within the TNFα gene. To

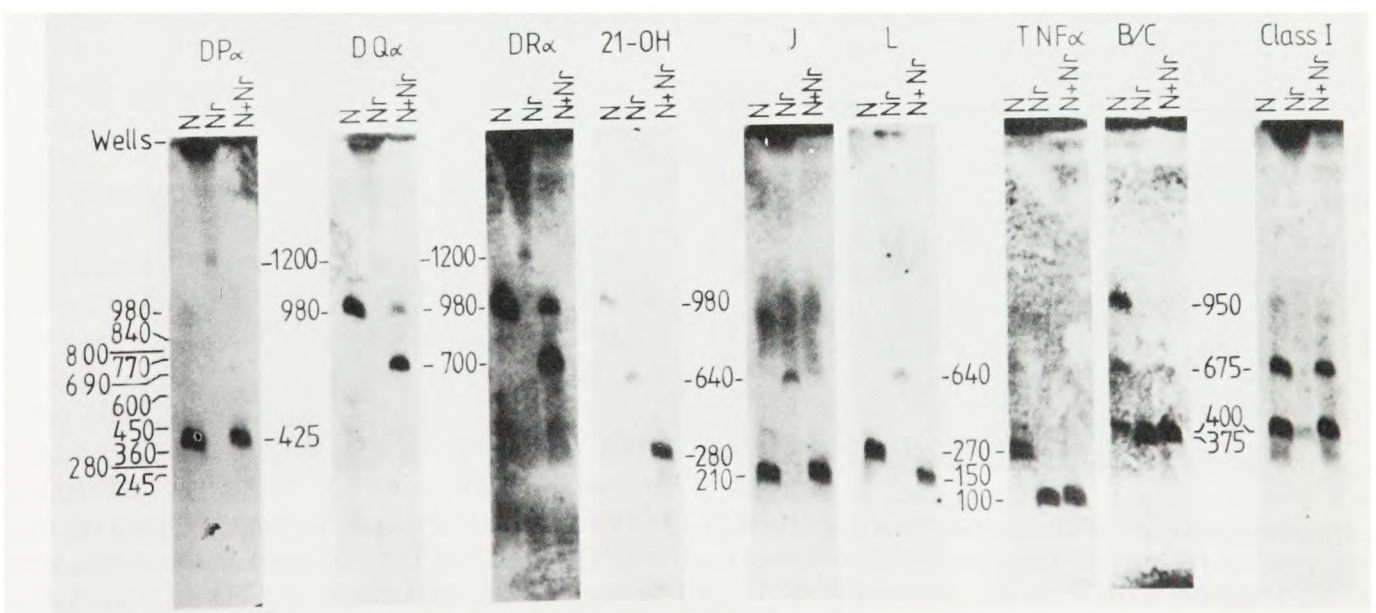


FIG. 2. Southern blot analysis of genomic DNA digests separated by PFGE. Fragments generated by *Not* I (N), *Nru* I (Nr), or *Not* I/*Nru* I (N + Nr) were separated using a 65-s pulse interval. All autoradiograms were obtained from a single filter hybridized sequentially with the probes shown. B/C indicates that the B- and C-locus-specific probes gave the same hybridization pattern, and only the B-locus-probe result is shown. Yeast chromosomes were electrophoresed as size markers and are indicated on the left. Fragment sizes are given in kb. The size of the class II-hybridizing *Nru* I fragment is approximate.

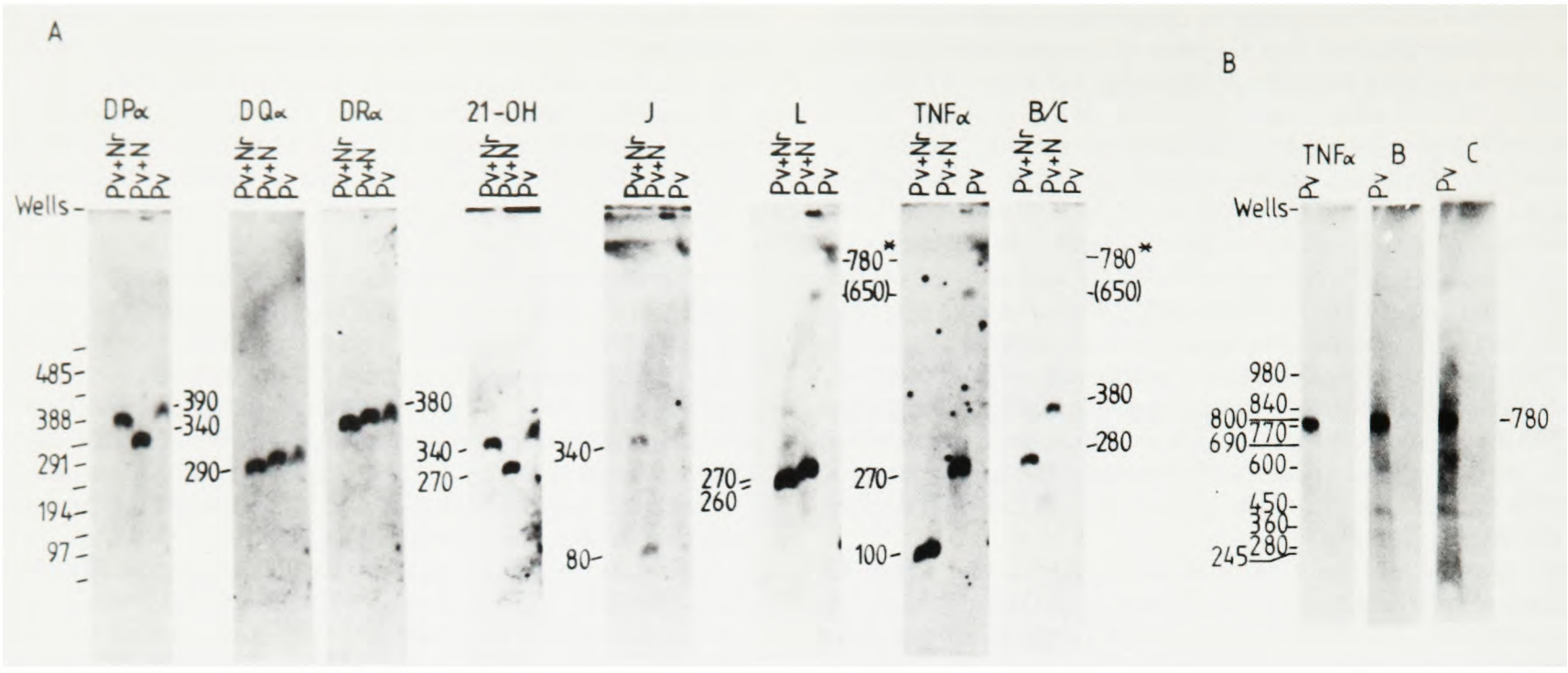


FIG. 3. Southern blot analysis of *Pvu* I-digested DNA separated by PFGE. (*A*) Sequential hybridization of the probes shown to a single Southern blot of DNA digested with *Pvu* I (Pv), *Pvu* I/*Not* I (Pv + N), or *Pvu* I/*Nru* I (Pv + Nr) and separated using a 30-s pulse interval. Concatemers of λ *c*I857S7 DNA were run as size markers and are indicated on the left. Fragment sizes are given in kb. Partial digestion products are enclosed in brackets. *, The 780-kb *Pvu* I fragment is completely resolved in *B*. (*B*) Sequential hybridization of TNFα, B-specific (lane B), and C-specific (lane C) probes to *Pvu* I-digested DNA resolved using a 60-s pulse interval. Yeast chromosomes were electrophoresed as size markers and are indicated on the left.

establish which of these sites are cut in genomic DNA, a single *Pae*R7I digest and double digests were resolved by conventional gel electrophoresis and blotted onto nitrocellulose. Hybridization of the TNFα probe revealed a 3.9-kb *Pae*R7I fragment produced by complete cleavage at the sites within the TNFα gene and immediately adjacent (Fig. 4*B*). Hence, TNFα hybridizes to a 290-kb fragment resolvable by PFGE and a 3.9-kb fragment that can only be seen on standard agarose gels. Probe L hybridizes to a series of fragments consistent with partial cleavage at the remaining three *Pae*R7I sites (Fig. 4*B*). The double digests with *Bam*HI and *Eco*RV confirm this observation and generate the appropriate fragments predicted from the restriction map of the cloned DNA (ref. 32; C.A.S. and R.D.C., unpublished data).

B- and C-locus-specific probes hybridized to the same fragments in *Not* I, *Pvu* I, and *Sal* I digests. However, on the basis of hybridization to *Mlu* I, *Cla* I, and *Mlu* I/*Cla* I digests, the B and C loci are separated by at least 80 kb (Fig. 1*B*).

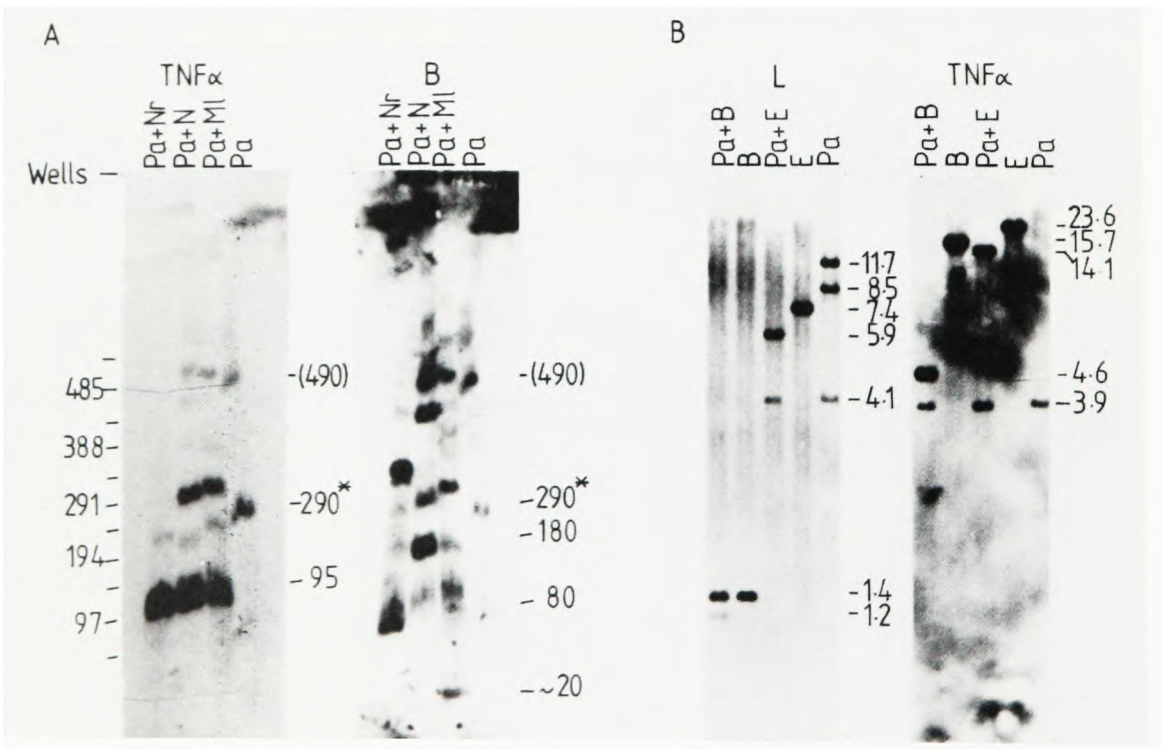


FIG. 4. Southern blot analysis of *Pae*R7I-digested DNA. (*A*) Hybridization of TNFα- and B-specific probes to *Pae*R7I (Pa) single and double digests separated using a 30-s pulse interval. Other restriction enzymes are *Not* I (N), *Nru* I (Nr), and *Mlu* I (Ml). Positions of λ DNA concatemers are indicated on the left. Fragment sizes are in kb. A partial *Pae*R7I digestion product is in brackets. *, The 290-kb *Pae*R7I fragment has been sized on other blots and here shows slightly anomalous mobility. Unmarked fragments are the products of partial digestion, which have not been mapped. (*B*) Southern blot analysis of *Pae*R7I sites in genomic DNA around the TNFα gene. Genomic DNA was digested with single- and double-digest combinations of *Pae*R7I (Pa), *Eco*RV (E), and *Bam*HI (B), separated on a conventional agarose gel, and hybridized with probe L and the TNFα probe. The three bands in the *Pae*R7I lane with probe L are due to partial cleavage at two of the five *Pae*R7I sites mapped in cosmid DNA in this region (Fig. 1*A*). The TNFα probe hybridizes to a 3.9-kb *Pae*R7I fragment due to cleavage at the *Pae*R7I site within the TNFα gene and also to the 290-kb fragment that can only be resolved by PFGE (see *A*). Fragment sizes are in kb.

DISCUSSION

PFGE, restriction enzymes that cut genomic DNA infrequently, and Southern blotting have been used to produce a long-range restriction map of the human MHC. We have established the orientation of the complement and 21-OHase gene loci relative to the class I and class II loci, the 21-OHase B gene being on the centromeric side of the C2 gene. The distance between the 21-OHase B gene and the DRα locus is 300–360 kb, while that between the C2 gene and the *HLA-B* locus is ≈650 kb. We have located precisely the genes for TNFα and -β within the MHC; the TNFα gene lies 390 kb from the C2 gene. The total size of MHC hybridizing fragments is ≈3800 kb, which represents ≈1/750 of the human genome.

Our data confirm the organization of the class II region proposed by Hardy *et al.* (10). Differences in observed fragment sizes can be explained by haplotype-specific restriction fragment length polymorphism or by insertions or deletions of DNA. There could also be methylation differences between cell lines leading to some sites not being recognized, as some enzymes with CpG in their recognition sequences appear to be sensitive to cytosine methylation (21). However, the conservation of rare-cutting restriction sites between cell lines (refs. 10–12 and this paper) in this very polymorphic region of the genome is striking and may be the product of a nonrandom distribution of nucleotides, C+G-rich regions being found in clusters. For instance, 25 kb on the telomeric side of the C2 gene is a cluster of at least seven rare-cutting sites within 25 kb. Some of these sites (two out of three *Mlu* I sites and the *Nru* I site) are not recognized in genomic DNA, possibly due to cytosine methylation, but a *Not* I and a *Mlu* I site within 1 kb of each other are both cleaved. The apparent nonrandom distribution of infrequently cutting sites in the mammalian genome has been observed (21, 33) and may represent so-called HTF Islands (*Hpa* II Tiny Fragments) (21, 31).

The organization of the human MHC is very similar to that determined in the mouse (7). The orientation of the class III genes is the same, and the proposed distances between the class III and class II loci and between class III and TNF are comparable. However, our results suggest that the TNF genes are ≈250 kb from the HLA-B region compared to ≈70 kb for the analogous genes in the murine MHC. It should be pointed out that the human and mouse MHC maps differ in that the *H-2K* locus is separated from the other class I genes at the other side of the class II region. It is possible that the difference in distance between TNF and class I in the two species is associated with the genetic event that separated the mouse class I loci. The estimated total size of the human MHC is twice as large as the murine MHC (4000 kb vs. 2000 kb), but most of this difference is made up of fragments that hybridize to general class I gene probes.

We believe that the physical linkage map of the human MHC presented here will play an important part in understanding HLA disease associations. The distances between the DRα and 21-OHase genes and between the C2 and TNFα genes are sufficiently large to accommodate a number of as yet unidentified genes. Further mapping of disease-affected haplotypes and the isolation of a complete set of overlapping cosmids covering the DRα to HLA-B region should permit the characterization of such genes. The demonstration of the precise position of the TNFα and -β genes within the human MHC is of major interest because TNFα and -β are mediators of a number of responses that may play a role in autoimmune disease. For instance, TNFα augments MHC antigen expression in some tissues (34, 35), which may lead to enhanced cytolytic damage mediated by class I antigens or presentation of autoantigens by epithelial cells aberrantly expressing class II antigens (35).

We sincerely thank Tom Strachan, John Bell, and Mark Bodmer for the class I, class II, and TNFα probes, respectively; Jasper Rees for pDVcos; Ian Hunter for growing the lymphoblastoid cell line; Ken Reid, Dick Cotton, Ed Southern, and Rakesh Anand for critical review of the manuscript; and Ms. C. Brooks and Mr. F. Caddick for secretarial and photographic help. I.D. and C.A.S. hold Science and Engineering Research Council and Medical Research Council Studentships, respectively. R.D.C. acknowledges the advice and encouragement he received from the late Prof. R. R. Porter who would have seen the result of the first PFGE blot in September 1985, but for his untimely death.

1. Lamm, L. U. & Olaisen, B. (1985) *Cytogenet. Cell Genet.* **40,** 128–155.
2. Strachan, T. (1987) *Br. Med. Bull.* **43,** 1–14.
3. Trowsdale, J. (1987) *Br. Med. Bull.* **43,** 15–36.
4. Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984) *Nature (London)* **307,** 237–241.
5. Carroll, M. C., Campbell, R. D. & Porter, R. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 521–525.
6. Spies, T., Morton, C. C., Nedospasov, S. A., Fiers, W., Pious, D. & Strominger, J. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8699–8702.
7. Müller, U., Jongeneel, C. V., Nedospasov, S. A., Lindahl, K. F. & Steinmetz, M. (1987) *Nature (London)* **325,** 265–267.
8. Schwartz, D. C. & Cantor, C. R. (1984) *Cell* **37,** 67–75.
9. Carle, G. F. & Olson, M. V. (1984) *Nucleic Acids Res.* **12,** 5647–5664.
10. Hardy, D. A., Bell, J. I., Long, E. O., Lindsten, T. & McDevitt, H. O. (1986) *Nature (London)* **323,** 453–455.
11. Ragoussis, J., van der Bliek, A., Trowsdale, J. & Ziegler, A. (1986) *FEBS Lett.* **204,** 1–4.
12. Lawrance, S. K., Smith, C. L., Srivastava, R., Cantor, C. R. & Weissman, S. M. (1987) *Science* **235,** 1387–1390.
13. Müller, U., Stephan, D., Phillippsen, P. & Steinmetz, M. (1987) *EMBO J.* **6,** 369–373.
14. van Ommen, G. J. B. & Verkerk, J. M. H. (1986) in *Human Genetic Disease, A Practical Approach*, ed. Davies, K. E. (IRL, Oxford), pp. 113–133.
15. Jackson, D. A. & Cook, P. R. (1985) *EMBO J.* **4,** 913–918.
16. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5758–5763.
17. Steinmetz, M., Stephan, D., Dastoornikoo, G. R., Gibb, E. & Romaniuk, R. (1986) in *Immunological Methods*, eds. Lefkovitz, I. & Pernis, B. (Academic, New York), Vol. 3, pp. 1–19.
18. Sternberg, M., Tiemeier, D. & Enquist, L. (1977) *Gene* **1,** 225–280.
19. 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.
20. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
21. Brown, W. R. A. & Bird, A. P. (1986) *Nature (London)* **322,** 477–481.
22. Anand, R. (1986) *Trends Genet.* **2,** 278–283.
23. Lee, J. S., Trowsdale, J., Travers, P. J., Carey, J., Grosveld, F. G., Jenkins, J. & Bodmer, W. F. (1982) *Nature (London)* **299,** 750–752.
24. Rodrigues, N. R., Dunham, I., Yu, C. Y., Carroll, M. C., Porter, R. R. & Campbell, R. D. (1987) *EMBO J.* **6,** 1653–1661.
25. Trowsdale, J., Lee, J., Kelly, A., Carey, J., Jenkins, J., Travers, P. & Bodmer, W. F. (1984) *Mol. Biol. Med.* **2,** 53–61.
26. Strachan, T., Dodge, A. B., Smillie, D., Dyer, P. A., Sodoyer, R., Jordan, B. R. & Harris, R. (1986) *Immunogenetics* **23,** 115–120.
27. Shirai, T., Yamaguchi, H., Ito, H., Todd, C. W. & Wallace, B. (1985) *Nature (London)* **313,** 803–806.
28. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
29. Carroll, M. C., Palsdottir, A., Belt, K. T. & Porter, R. R. (1985) *EMBO J.* **4,** 2547–2552.
30. Yu, C. Y. & Campbell, R. D. (1987) *Immunogenetics* **25,** 383–390.
31. Bird, A. P. (1986) *Nature (London)* **321,** 209–213.
32. Nedospasov, S. A., Shakhov, A. N., Turetskaya, R. L., Mett, V. A., Georgiev, G. P., Dobryrin, V. N. & Korobko, V. G. (1985) *Dokl. Acad. Nauk. SSSR* **285,** 1487–1490.
33. Smith, D. I., Golembieski, W., Gilbert, J. D., Kizyma, L. & Miller, O. J. (1987) *Nucleic Acids Res.* **15,** 1173–1184.
34. Collins, T., Lapierre, L. A., Fiers, W., Strominger, J. L. & Pober, J. S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 446–450.
35. Pujol-Borrell, R., Todd, I., Doshi, M., Bottazo, G. F., Sutton, R., Gray, D., Adolf, G. R. & Feldmann, M. (1987) *Nature (London)* **326,** 304–306.

The EMBO Journal vol.6 no.6 pp.1653–1661, 1987

Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia

Nanda R.Rodrigues, Ian Dunham, C.Yung Yu, Michael C.Carroll[1], Rodney R.Porter[2] and R.Duncan Campbell

Medical Research Council Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK, and [1]Department of Pediatrics, Harvard Medical School, Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

[2]Deceased

Communicated by A.F.Williams

21-Hydroxylase deficiency which causes congenital adrenal hyperplasia is one of the most common defects of adrenal steroidogenesis. There are two 21-hydroxylase genes in man, A and B, and these have been mapped to the HLA class III region. Only the 21-hydroxylase B gene is thought to be active. To understand the molecular basis of congenital adrenal hyperplasia in a patient with the salt-wasting form of the disease, we cloned and characterized his single 21-hydroxylase B gene. The nucleotide sequence of this gene and a 21-hydroxylase B gene from a normal individual have been determined. Comparison of the two sequences has revealed 11 nucleotide alterations, of which two are in the 5′ flanking region, four are in introns, one is in the 3′ untranslated region and four are in exons. Two of the differences in exons cause codon changes, with Ser-269 and Asn-494 in the normal 21-hydroxylase B gene being converted to Thr and Ser, respectively. These amino acid substitutions may give an insight into those residues necessary for 21-hydroxylase enzymatic activity. We have also confirmed that the 21-hydroxylase A gene is a pseudogene due to three deleterious mutations in the exons. In addition, comparison of the 21-hydroxylase B gene sequence with other published sequences indicates that this microsomal cytochrome P-450 may be polymorphic.

Key words: 21-hydroxylase /congenital adrenal hyperplasia/gene structure/HLA/polymorphism

Introduction

Of the hydroxylases participating in the biosynthesis of adrenal steroids, the enzyme steroid 21-hydroxylase [21-OHase; steroid, hydrogen-donor: oxygen oxidoreductase (21-hydroxylating) EC 1.14.99.19] is the most extensively studied.

21-Hydroxylation was the first enzymatic activity ascribed to any cytochrome P-450 (Cooper *et al.*, 1965) and is one of the key enzymatic steps in the synthesis of glucocorticoids and mineralocorticoids. 21-OHase mediates the conversion of progesterone to 11-deoxycortisol, a precursor of cortisol (Finkelstein and Shaefer, 1979). Since cortisol is the primary feedback inhibitor of corticotropin (ACTH) secretion in man, impairment of cortisol synthesis due to deficiency of any of the enzymes involved in adrenal steroidogenesis results in elevated levels of ACTH causing congenital adrenal hyperplasia (CAH) (New *et al.*, 1982). CAH due to defective 21-hydroxylation is the most common of these inborn errors of metabolism affecting ~1 in 5000 to 1 in 15 000 births in a severe form and 0.3% in a milder nonclassical form in the general Caucasian population (Speiser *et al.*, 1985).

21-OHase deficiency is linked to the major histocompatibility complex (HLA) on the short arm of human chromosome 6 and is inherited as a monogenic autosomal recessive trait (Dupont *et al.*, 1977). There are two 21-OHase genes in man (White *et al.*, 1984a) and they have been mapped to the HLA class III region together with the genes encoding the serum complement proteins C2, Factor B, C4A and C4B (Carroll *et al.*, 1984b). Detailed restriction mapping of cosmid clones and of uncloned genomic DNA has localized the two 21-OHase genes, 21-OHase A and 21-OHase B, to 6-kb regions flanking the 3′ end of the C4A and the C4B genes, respectively (Carroll *et al.*, 1985a; White *et al.*, 1985). It has been proposed that only one of the two 21-OHase genes encodes an active enzyme. Individuals with the HLA haplotype A3, B47, DR7, C4A1 BQO and suffering from CAH were found to have a homozygous deletion of the 21-OHase B gene, while individuals with homozygous deletion of the 21-OHase A gene are hormonally normal (White *et al.*, 1985). Recent structural analysis of both genes has supported this observation and has shown that the 21-OHase B gene appears to be the active gene, while the 21-OHase A gene is a highly homologous pseudogene (Higashi *et al.*, 1986; White *et al.*, 1986).

Four distinct clinical manifestations of 21-OHase deficiency have been identified; simple virilizing, salt wasting, late onset and cryptic (New and Levine, 1984). Simple virilizing due to excessive production of fetal adrenal androgens and salt wasting due to aldosterone deficiency are the classical forms of 21-OHase deficiency and have shown significant association with HLA B47 and to a lesser extent with HLA B51/5, B53 and B60. Linkage disequilibrium has also been reported for the nonclassical forms of 21-OHase deficiency with a significantly increased frequency of the HLA haplotype B14, DR1 in both the late onset and cryptic forms. In a recent study it was reported that in the majority of patients with classical CAH, the defect was not due to gene deletion (Rumsby *et al.*, 1986). The 21-OHase B gene was present on at least one chromosome, but obviously nonfunctional, or the expressed product partly or wholly inactive.

Due to the high incidence of this serious inborn error of metabolism it is clearly essential to define molecular markers for the presence of particular disease genes which can be used in prenatal diagnosis, and be useful in characterization, prognosis and treatment. In order to achieve this goal and to understand the molecular basis of CAH in a patient with the salt-wasting form of the disease we cloned and characterized his single 21-OHase B gene. We present here the nucleotide sequence of this gene and compare it with the sequence of a 21-OHase B gene from a normal individual and with a 21-OHase A gene. The defective 21-OHase gene shows only two amino acid substitutions and they may give an important insight into those residues necessary for 21-OHase enzymatic activity. In addition, comparison of the 21-OHase B gene from the normal individual with other published sequences indicates that this microsomal cytochrome P-450 may be polymorphic.

© IRL Press Limited, Oxford, England

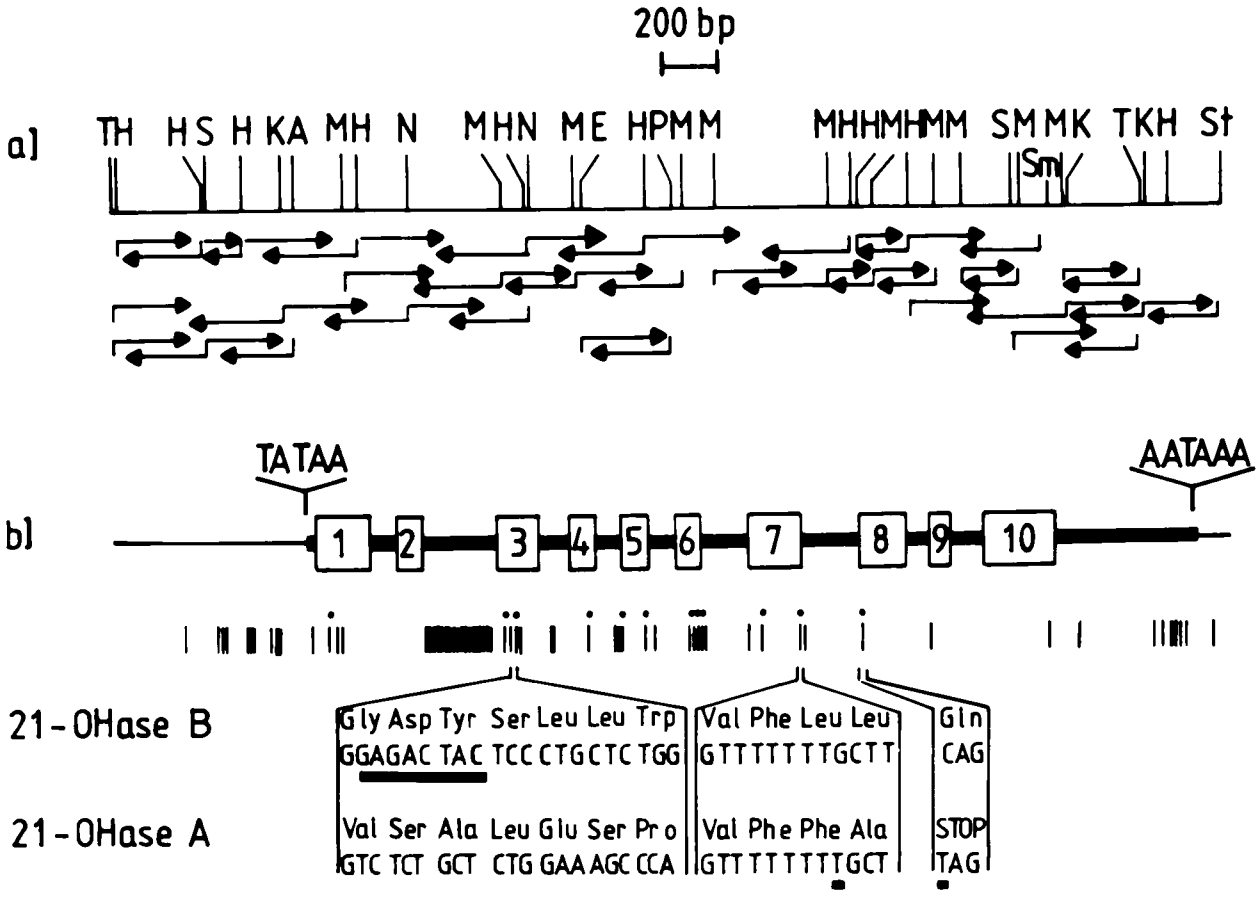


Fig. 1. Sequencing strategy and gene structure of the human 21-OHase B gene. **(a)** The sequencing strategy for the 21-OHase B gene is shown beneath the corresponding restriction map. The various DNA fragments were cloned into M13mp8 or mp9 and sequenced by the chain termination method. The arrows indicate the extent and the orientation in which each fragment was sequenced. The sequencing strategy for the 21-OHase A gene and the 21-OHase B gene from the patient with CAH was essentially the same. The restriction enzyme sites shown are: A, *Acc*I; E, *Eco*RI; H, *Hin*fI; K, *Kpn*I; M, *Msp*I; N, *Nco*I; P, *Pvu*II; S, *Stu*I; Sm, *Sma*I; St, *Sty*I; T, *Taq*I. **(b)** Intron−exon structure of the 21-OHase gene. The extent of the gene from the putative transcription start site to the polyadenylation site is illustrated by the thick line. Exons are shown boxed and are numbered 1−10. The position of the TATAA box and the polyadenylation signal AATAAA are shown. The positions at which the 21-OHase A gene differs from the 21-OHase B gene are marked by bars beneath the gene structure. (˙) over a bar indicates a codon change due to nucleotide alteration. The three deleterious mutations in the 21-OHase A gene are illustrated below the corresponding bars. The 8-bp deletion and the T insertion, which cause frameshifts, and the C → T transition, which generates a stop codon, are underlined.

```
TCGACAGCTAGATTTCCAGGCTGGAATCCTGCCCTCCACAACATGCGAACAATACCCGTGTTGCATATAGAGCATGGCTGTGAAGAGTTGAGTGAGTGCCCACAAAGCACTTAGAGCAGT -625

GTCTGGTACATGCTATTACTCCGCAGCGGGAAACCACTTCCTCCTTTGTCTTCTGGGCACTTTTGTGAGTGAAAGGAGGCACTAATAACAATCACACTGGGATACCTGTATATACTGGAA -505
                                                                    G

TGCCCCAGGCAAACCAGGCTTAAACTGTATTACTCTATCTGTAGCTTAAACTAACAAACAA-CCCACACAAATCACATTTTGTTCTTCAGGCGATTCAGGAAGGCCTATTAGGCAGGGAC -386
                                                              A

TGCCATTTTCTCTCTGAGACAAACATCATGCCAGTAAACTGGCCCACGGTGGGGTGGCAGAGGGAGAGGGCCCAGGTGGGGGCGGACACTATTGCCTGCACAGGTGATGTGGAACCAGAA -266
                                                                    C          CC         G   ^
                                                                                              T
AGCTGACTCTGGATGCAGGAAAAAGGTCAGGGTTGCATTTCCCTTCCTTGCTTCTTGATGGGTGATCAATTTTTTT-GAAATACGGACGTCCCAAGGCCAATGAGACTGGTGTCATTCCA -147
                                                              C          T         T

GAAAAGGGCCACTCTGTGGGCGGGTCGGTGGGAGGGTACCTGAAGGTGGGGTCAAGGGAGGCCCCAAAACAGTCTACACAGCAGGAGGGATGGCTGGGGCTCTTGAGCTATAAGTGGCAC -27
                      T           A C      G

                 1 M  L  L  L  G  L  L  L  L  L  P  L  L  A  G  A  R  L  L  W  N  W  W  K  L  R  S  L  H  L  P
CTCAGGGCCCTGACGGGCGTCTCGCCATGCTGCTCCTGGGCCTGCTGCTGCTGCTGCCCCTGCTGGCTGGCGCCCGCCTGCTGTGGAACTGGTGGAAGCTCCGGAGCCTCCACCTCCCGC 94
                      T                                                                                                 T*

P  L  A  P  G  F  L  H  L  L  Q  P  D  L  P  I  Y  L  L  G  L  T  Q  K  F  G  P  I  Y  R  L  H  L  G  L  Q  67
CTCTTGCCCCGGGCTTCTTGCACTTGCTGCAGCCCGACCTCCCAATCTATCTGCTTGGCCTGACTCAGAAATTCGGGCCCATCTACAGGCTCCACCTTGGGCTGCAAGGTGAGAGGCTGA 214
                        C                    C
                        C                    C

                                                                                  68 D  V  V  V  L  N  S  K  R  T  I  E
TCTCGCTCTGGCCCTCACCATAGGAGGGGCGGAGGTGACGGAGAGGGTCCTCTCTCCGCTGACGCTGCTTTGGCTGTCTCCCAGATGTGGTGGTGCTGAACTCCAAGAGGACCATTGAG 334

 E  A  M  V  K  K  W  A  D  F  A  G  R  P  E  P  L  T  97
GAAGCCATGGTCAAAAAGTGGGCAGACTTTGCTGGCAGACCTGAGCCACTTACCTGTAAGGGCTGGGGGCATTTTTTCTTTCTTAAAAAAATTTTTTTTT---AAGAGATGGGTTCTTGC 451
                                                                                C                C          TGTT       G

TATGCTGCCCAGGCTGGTCTTAAATTCCTAGTCTCAAATGATCCTCCCACCTCAGCCTCAAGTGTGAGCCACCTTTGGGGCATCCCCAATCC---AGGTCCCTGGAAGCTCTTGGGGGGG 568
   T              G      G      G              G          G          C    T          TCC    A                  -
```

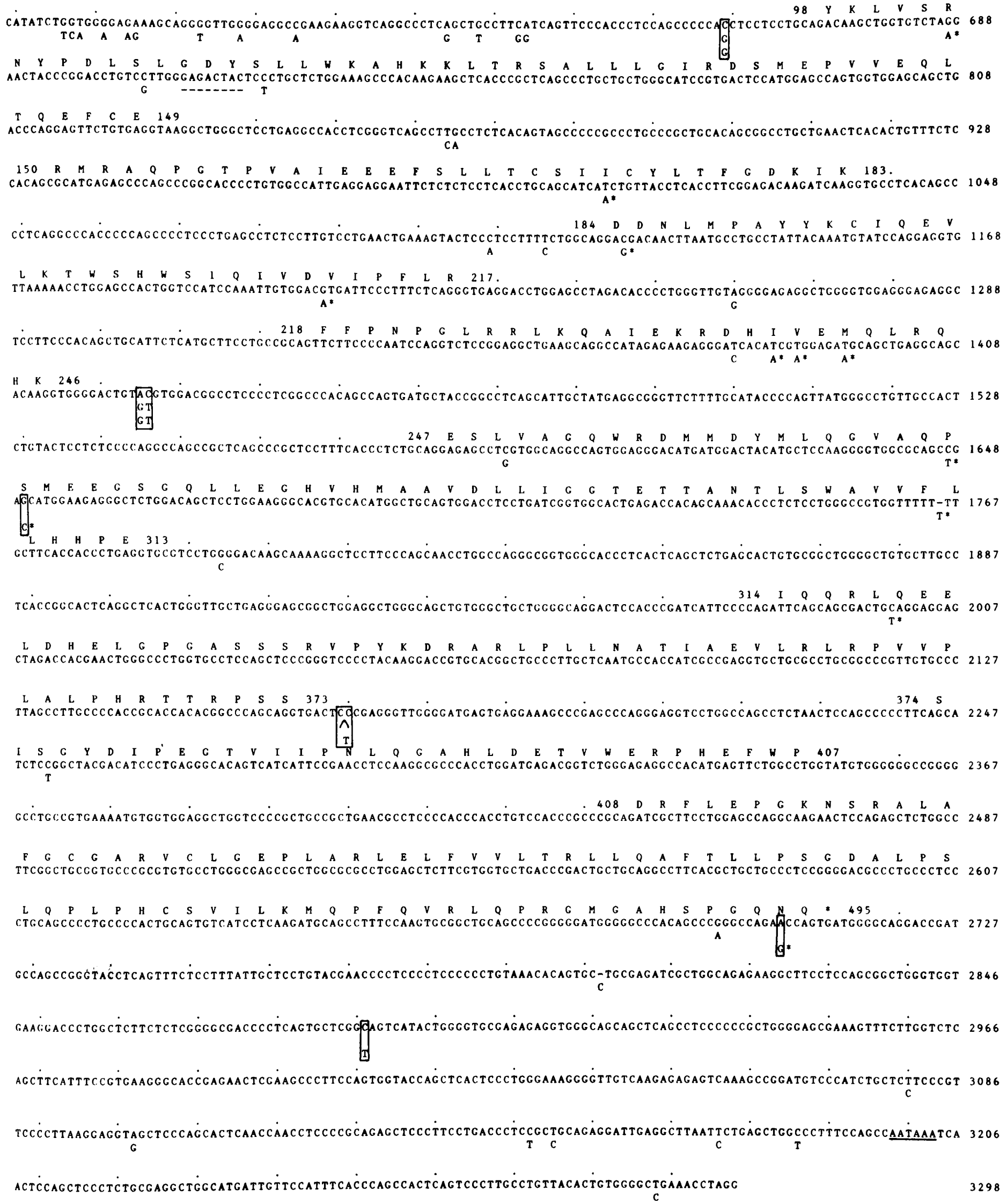


Fig. 2. Comparison of the nucleotide sequences of a normal 21-OHase B gene (first line) and a 21-OHase A gene (second line) with the 21-OHase B gene from the patient with CAH (third line). The full nucleotide sequence of the normal 21-OHase B gene and its predicted amino acid sequence are shown. The differences in the 21-OHase A gene (second line) and the 21-OHase B gene from the patient with CAH (third line) from the normal 21-OHase B gene sequence are indicated under the corresponding nucleotides of the B gene. Dashes are introduced in the sequence to maximize homology between the 21-OHase A and the normal 21-OHase B genes. (∧) indicates an insertion mutation in the 21-OHase gene from individual II.1. Those nucleotides which differ between the two B gene sequences are shown boxed. Single nucleotide alterations which cause a codon change are marked with an asterisk. The putative TATAA transcription signal and the polyadenylation signal (AATAAA) are underlined.

Results

Cloning and structure of 21-OHase genes

The two 21-OHase genes were isolated from two diferent genomic libraries. The 21-OHase B gene was identified on the basis of the 3.7-kb *Taq*I fragment which is characteristic of all 21-OHase B genes (White *et al.*, 1984a) and was isolated from cosmid KEM-I which contained a C4B gene (Carroll *et al.*, 1985b). Similarly the 21-OHase A gene was identified on the basis of the 3.2-kb *Taq*I fragment and was isolated from cosmid 1E3 containing a C4A gene (Carroll *et al.*, 1984a). A *Bgl*II/*Bam*HI fragment, 5.5 kb in length covering the entire 21-OHase gene, was then subcloned from both cosmids into the plasmid vector pAT153/*Pvu*II/8. Restriction fragments ranging from 0.15 to 0.75 kb were further subcloned into M13mp8 and M13mp9 and subsequently sequenced by the chain termination method. The strategy for determining the sequence of both genes is illustrated in Figure 1a.

The gene structure was initially deduced by comparison with the bovine 21-OHase cDNA sequence (Yoshioka *et al.*, 1986) which shows 70% homology with the human 21-OHase amino acid sequence and was further confirmed with the published human 21-OHase cDNA sequence (White *et al.*, 1986). The complete nucleotide sequence of both 21-OHase genes as shown in Figure 2 indicates that they are 3.3 kb in length and are split into 10 exons by nine introns. They are highly conserved and share 98% homology in the nucleotide sequence. The possible transcription initiation signal TATAA (Nussinov, 1986) is located in the 5′ region 38 bp upstream of the putative ATG initiation codon in both genes. S1 nuclease mapping has suggested that the major transcription start site is located at nucleotide position −9 which is 25 bp downstream from the TATAA sequence (Higashi *et al.*, 1986). The polyadenylation signal AATAAA lies 486 bp away from the termination codon in the 10th exon. cDNA sequence has shown it to be 21 bp upstream of the poly(A) tail (White *et al.*, 1986). From these data it appears that the 21-OHase gene codes for a mRNA of 2 kb and this is similar to that detected by Northern blot analysis of adrenal RNA preparations (Carroll *et al.*, 1985a; White *et al.*, 1985). In addition it is clear that the 21-OHase gene encodes a protein which contains 495 amino acids. This predicts that the mol. wt of human 21-OHase is 55 829, which is similar to that of porcine (54 000) (Biggin *et al.*, 1983) and bovine (52 000) (White *et al.*, 1984b) 21-OHase respectively. Comparison of the nucleotide sequences of the 21-OHase A and B genes reveals that there are 90 differences and these are depicted in Figures 1b and 2. Thirteen differences lie in the 5′ flanking region upstream of the ATG initiation codon, 42 lie in introns and eight lie in the 3′ untranslated region. Twenty-seven differences are in the exons, of which eight are silent and 19 cause codon changes. Nine of the codon changes would have an unpredictable effect on the gene product while the remaining differences are quite drastic. These include an 8-bp deletion in the third exon of the 21-OHase A gene at codons 111−113 causing a shift in the reading frame as shown in Figure 1b. This shift introduces a stop signal (TGA) between codons at position 133 and 134. Another frameshift occurs in the seventh exon of the 21-OHase A gene at codon 307 due to a 1-bp insertion and a third nonsense mutation occurs in the eighth exon in codon 319 due to a C → T transition (Figures 1b and 2). The changes in reading frame and the introduction of inphase premature termination codons show that the 21-OHase A gene would encode a truncated and presumably nonfunctional protein. Northern blot analysis of adrenal RNA using oligonucleotides specific for the 21-OHase A and B genes has shown that only the 21-OHase B gene probe hybridized to an mRNA species (Higashi *et al.*, 1986). Thus the 21-OHase A gene is clearly a pseudogene and non-expression of the putative A gene product may be due to lack of transcription of the gene, incorrect processing or instability of the mRNA.

Southern blot analysis of a patient with CAH

In patients with classical and nonclassical 21-OHase deficiencies, certain HLA haplotypes appear with either a significantly increased or decreased frequency relative to their frequency in the general population. Thus there is genetic linkage disequilibrium between the 21-OHase deficiency and specific HLA alleles (New *et al.*, 1982). Classical CAH is most commonly associated with the HLA haplotype B47 (Dupont *et al.*, 1980) where the relative risk if 15.4. In this haplotype, there is a deletion of the C4B gene extending to the 21-OHase B gene (Carroll *et al.*, 1985b). However, in a study of 20 patients with classical CAH and random haplotypes, only in one instance was the defect due to homozygous deletion of the 21-OHase B gene (Rumsby *et al.*, 1986). A patient (individual II.1 in Figure 3) with the HLA haplotypes A2 B44 C5 DR6 C4 A3 BQO/A3 B47 (C6) DR7 C4 A1 BQO has the salt-wasting form of CAH. In order to characterize the basis for the defect, genomic DNA from individual II.1 and his parents (family Q) was digested with the restriction enzymes *Taq*I and *Kpn*I and the Southern blots were probed with 21-OHase specific probe PA (see Materials and methods). Two fragments at 3.7 and 3.2 kb were seen in the *Taq*I digest, and at 2.9 and 4 kb in the *Kpn*I digest (Figure 3). These correspond to the 21-OHase B and 21-OHase A genes respectively, from previous reports (White *et al.*, 1984a) and the nucleotide sequence of the two genes. The gene dosage was determined by scanning densitometry of the autoradiograph and is listed in Table I. Individual II.1 has a single 21-OHase B gene determined on the basis of the intensity of the 3.7-kb *Taq*I and 2.9-kb *Kpn*I fragments and two 21-OHase A genes. Figure 4 shows a restriction map of the haplotypes of individual II.1 constructed on the basis of the Southern blot analysis. He has inherited a single 21-OHase A gene on the A3 B47 (C6) DR7 C4 A1 BQO haplotype from his mother and one 21-OHase A gene and his single 21-OHase B gene on the A2 B44 C5 DR6 C4 A3 BQO haplotype from his father. Since individual II.1 has CAH, his single 21-OHase B gene is defective. This gene was characterized further by direct cloning and sequence analysis.

Cloning of the 21-OHase B gene from individual II.1

A λ genomic library constructed using DNA isolated from individual II.1 (Yu *et al.*, 1986) was screened with 21-OHase-specific probe PA. DNA from 22 positive clones was digested with the restriction enzymes *Taq*I and *Kpn*I. The clone 15λ21B containing the 21-OHase B gene was identified on the basis of the 3.7-kb *Taq*I and 2.9-kb *Kpn*I fragments. Since it had initially been established that individual II.1 has only one 21-OHase B gene (Figure 3), the clone was further characterized. A 5.5-kb *Bgl*II/*Bam*HI fragment from clone 15λ21B was subcloned into the plasmid vector pAT153/*Pvu*II/8. Further subcloning was carried out into M13mp9 and the sequencing strategy was similar to that of the normal 21-OHase B gene as shown in Figure 1.

Comparison of the DNA sequences of the two 21-OHase B genes revealed 11 nucleotide alterations which are shown in Figures 2 and 5. Of these differences two are in the 5′ flanking region upstream of the ATG initiation codon and are due to a G → A transition at nucleotide −551 and a T insertion between

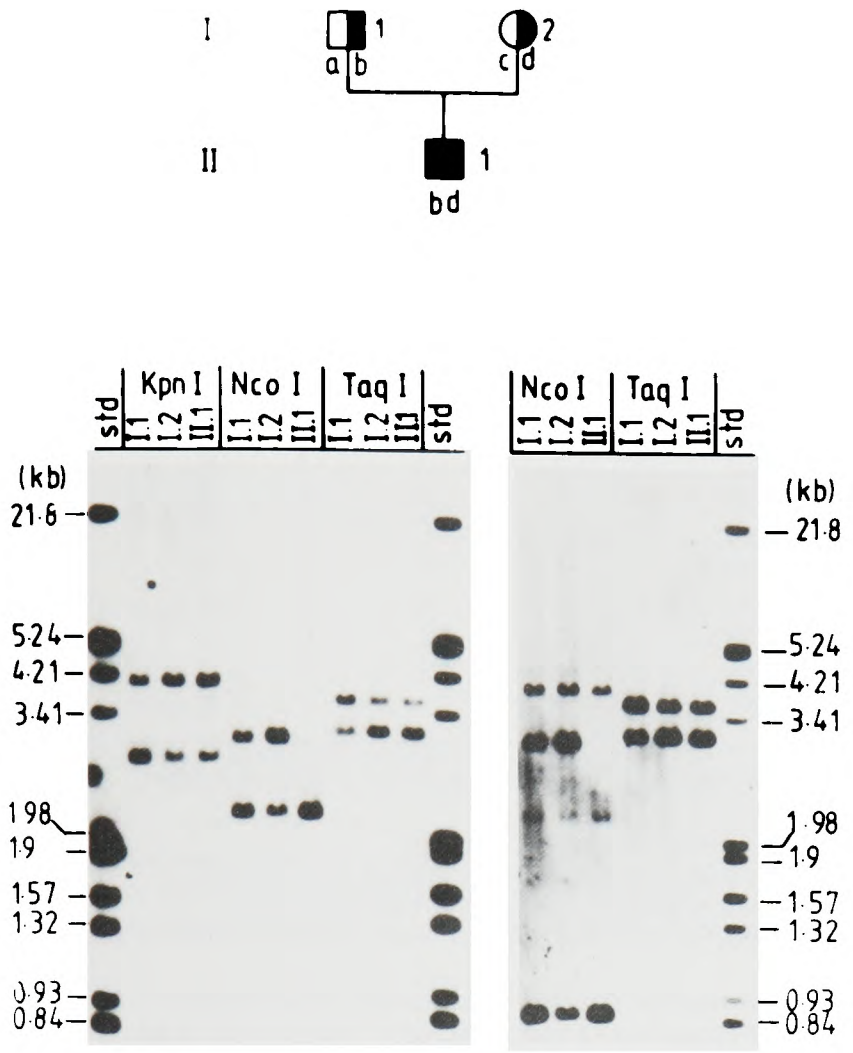


Fig. 3. Southern blot analysis of genomic DNA from a patient with CAH and his parents (family Q) with the enzymes *Taq*I, *Kpn*I and *Nco*I. The pedigree of family Q is shown at the top, and the HLA/complement haplotypes are given in Table I. ■, Male; ○, female; full shaded, homozygous for 21-OHase deficiency. The bottom left panel shows the Southern blot hybridized with probe PA. The 3.7- and 3.2-kb fragments in the *Taq*I digest and the 2.9- and 4-kb fragments in the *Kpn*I digest correspond to the 21-OHase B and 21-OHase A genes respectively. The gene dosage shown in Table I was deduced from the intensity of the bands by scanning of the autoradiograph. In the *Nco*I digest individual II.1 has a single band at 2.3 kb while his parents have bands at 2.3 and 3.1 kb. The bottom right panel shows the same blot hybridized with probe H. The *Taq*I and *Kpn*I (not shown) patterns remain unchanged. However, in the *Nco*I digest two new fragments at 4 and 0.85 kb light up. See text for discussion.

Table I. Haplotypes of the members of family Q

HLA/complement haplotypes							Gene dose	
							21-OHase A	21-OHase B
I.1	a	A28	B51	C-	DR5	C4A3BQO	–	1
	b	A2	B44	C5	DR6	C4A3BQO	1	1
I.2	c	A2	B62	C3	DR4	C4A3B3	1	1
	d	A3	B47	(C6)	DR7	C4A1BQO	1	–
II.1	b	A2	B44	C5	DR6	C4A3BQO	1	1
	d	A3	B47	(C6)	DR7	C4A1BQO	1	–

nucleotides −281 and −280 (Figure 2). Four differences are in introns, one is in the 3′ untranslated region and four are in exons. Five of the differences (two in exon 1, one in intron 2 and two in intron 6) change the nucleotides at those positions to that of the 21-OHase A gene sequence (Figure 2). Two of the four differences in the exons (a T → C transition in codon 40 and an A → C transition in codon 46) are silent mutations while the other two cause codon changes (Figures 2 and 5). A transversion from G → C in the seventh exon at codon 269 causes

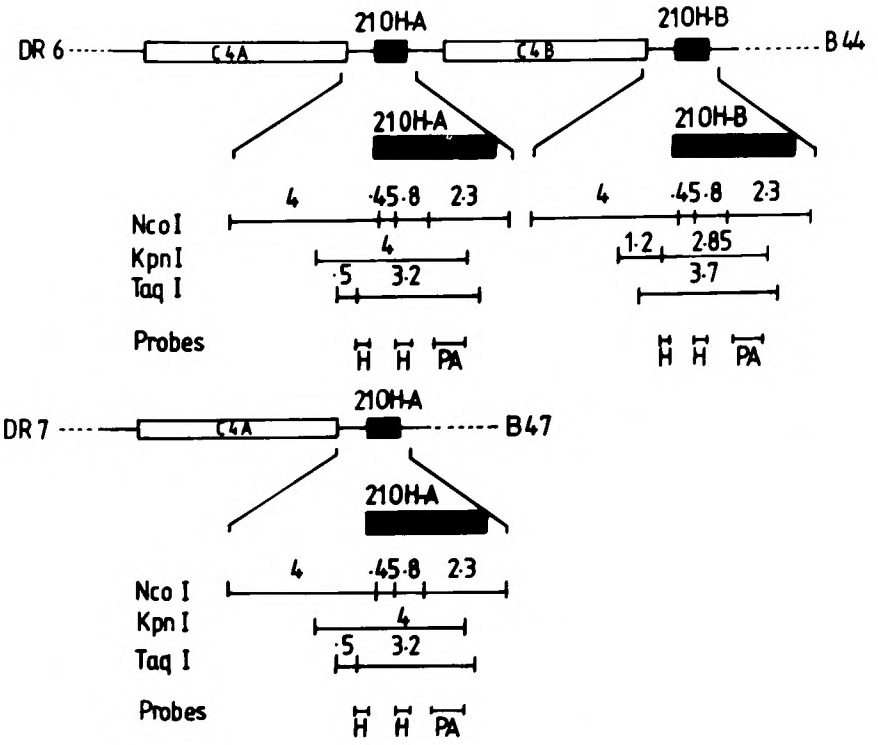


Fig. 4. Restriction maps of the 21-OHase genes on the B44 and B47 haplotypes of individual II.1 are shown for enzymes *Nco*I, *Kpn*I and *Taq*I. The 4-kb *Kpn*I and 3.2-kb *Taq*I bands correspond to the 21-OHase A gene and the 2.85-kb *Kpn*I and 3.7-kb *Taq*I bands correspond to the 21-OHase B gene. The polymorphic *Nco*I site which cleaves the 3.1-kb fragment into two fragments of 2.3 and 0.85 kb (Figure 3) is present in all the three 21-OHase genes of individual II.1. The probes used are indicated below the maps.

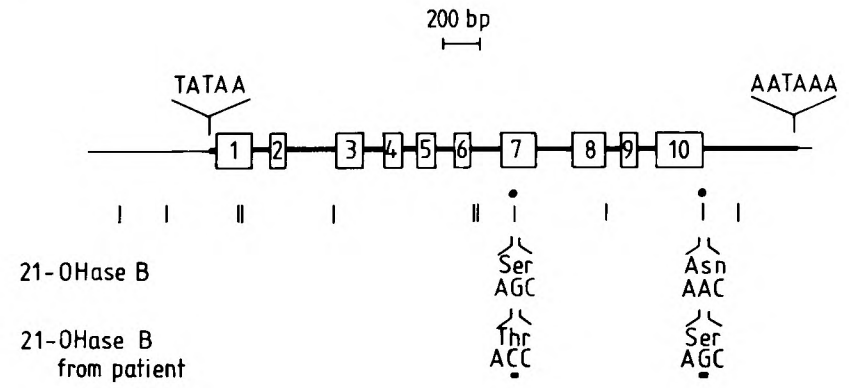


Fig. 5. A schematic comparison of the 21-OHase B gene from the patient with CAH with a normal 21-OHase B gene. The intron−exon structure is taken from Figure 2. The positions of the nucleotide differences between the two genes are indicated by bars. The two differences which cause a codon change are marked by (˙) above the bar. A G → C transversion changes Ser-269 to Thr, and an A → G transition changes Asn-494 to Ser.

a Ser to change to Thr and an Asn is changed to Ser by an A → G transition in the 10th exon at codon 494 near the COOH terminus. None of the above alterations cause a change in reading frame. The G → C transversion in the seventh exon generates an extra *Nco*I site. To confirm the *Nco*I site at the genomic level, DNA from individual II.1 and his parents was digested with this restriction enzyme and probed with probes PA and H respectively. The *Nco*I digest pattern with probe PA is shown in Figure 3. Individual II.1 has only one fragment at 2.3 kb while his parents have fragments of 3.1 and 2.3 kb. In order to map precisely the position of the *Nco*I fragments within the gene, the blot was reprobed with probe H which is 5′ to probe PA. Individual II.1 showed two fragments of 4 and 0.85 kb, while his parents each showed three fragments of 4, 3.1 and 0.85 kb (Figure 3). Hybridization of the blot with a probe specific for the 3′ end of C4 revealed that the 4-kb *Nco*I fragment extends into the 3′ region of the neighbouring C4 gene. Based on these results and scanning of the autoradiographs to determine band intensities

Table II. Differences in the published 21-OHase nucleotide sequences

No.	Position	B.1	Higashi *et al.* B.2	White *et al.* B.3	A.1	Higashi *et al.* A.2	White *et al.* A.3
1	−356	G	T		G	T	
2	−352	G	–		G	G	
3	−335	G	–		G	–	
4	−282	G	–		G	–	
5	−95	T	T	C	T	T	T
6	−8	G	G	G	G	G	T
7	−4	C	C	C	T	T	C
8	28	*C	–	–	C	C	C
9	29	*T	–	–	T	T	T
10	30	*G	–	–	G	G	G
11	118	T	T	C	C	C	C
12	138	A	A	C	C	C	C
13	274	C	C	C	C	C	–
14	422	A	C	C	C	C	C
15	456	C	T	C	T	T	T
16	532	C	C	C	C	C	G
17	533	A	A	A	T	T	CT
18	562	G	G	CG	G	G	G
19	568	G	–	G	–	–	–
20	576	G	G	G	T	T	G
21	577	G	G	G	C	C	T
22	578	T	T	T	A	A	C
23	579	G	G	G	G	G	A
24	581	G	G	G	A	A	G
25	584	G	G	G	A	A	G
26	585	A	A	A	G	G	A
27	587	A	A	A	A	A	AGA
28	589	C	C	C	C	C	A
29	590	A	A	T	A	A	T
30	597	G	G	G	G	G	A
31	598	G	G	G	A	A	G
32	604	C	C	C	C	C	A
33	605	C	C	C	A	A	C
34	653	C	C	C	C	C	G
35	659	C	C	A	G	G	G
36	687	*G	A	A	A	A	A
37	701	C	C	T	C	C	T
38	864	T	C	C	C	C	C
39	865	G	G	G	A	G	G
40	892	C	–	C	C	–	C
41	911	G	G	GG	G	G	G
42	1208	G	G	G	A	G	G
43	1260	A	A	A	G	A	A
44	1423	G	G	C	G	G	C
45	1425	A	A	A	G	G	A
46	1426	C	C	C	T	T	C
47	1647	C	C	C	T	C	C
48	1688	G	G	G	G	T	T
49	1895	C	C	CC	C	C	CC
50	1902	G	G	G	G	G	GC
51	1911	G	G	G	G	G	C
52	2113	C	C	C	C	T	T
53	2252	C	C	T	T	C	T
54	2361	G	–	G	G	–	G
55	2375	G	G	G	G	A	G
56	2504	G	C	G	G	C	G
57	2505	*C	G	C	C	G	C
58	2697	G	A	A	A	A	A
59	2764	T	C	C	T	C	C
60	2801	C	C	C	CC	C	C
61	2803	G	G	GG	G	G	GG
62	2911	G	G	G	G	G	–
63	2913	G	G	–	G	G	–
64	2914	A	A	–	A	A	–
65	2939	C	C	–	C	C	–
66	2940	C	C	–	C	C	–
67	3119	C	C	–	C	C	–
68	3120	C	C	–	C	C	–
69	3121	A	A	–	A	A	–
70	3122	A	A	–	A	A	–
71	3150	C	C	C	T	T	C
72	3177	T	T	T	C	C	T
73	3187	C	C	C	T	T	C
74	3197	C	C	T	C	C	T
75	3198	C	C	T	C	C	T
76	3288	T	T		C	T	

B and A correspond to the 21-OHase B and 21-OHase A genes respectively.
B.1, A.1: the sequences reported here.
B.2, A.2: the sequences reported by Higashi *et al.* (1986).
B.3, A.3: the sequences reported by White *et al.* (1986).
A blank space indicates that sequence data were not available at those positions.
A dash (−) indicates absence of a nucleotide at that position.
An asterisk (*) indicates a codon change.

we constructed the *Nco*I restriction map shown in Figure 4. This confirmed the polymorphic *Nco*I site within the seventh exon of the 21-OHase B gene in individual II.1. There has been no previous report of this *Nco*I polymorphism. Nucleotide sequences from random unrelated individuals do not show this site in either of their 21-OHase genes (Higashi *et al.*, 1986; White *et al.*, 1986). In this family, however, both affected haplotypes exhibit this additional *Nco*I site. From our results, it is evident that individual II.1 has this site in all his 21-OHase genes suggesting that it is not restricted to being only in the 21-OHase B gene.

Discussion

Comparison of the nucleotide sequences of the 21-OHase A and 21-OHase B genes shows that they are highly conserved (Figure 2) and share 98% sequence homology in the exons and 96% homology in the introns. Of the two genes, however, only the 21-OHase B gene encodes an active protein. The 21-OHase A gene appears to be non-functional as individuals with a homozygous deletion of this gene synthesize cortisol normally, whereas individuals with a homozygous deletion of the 21-OHase B gene have classical CAH (White *et al.*, 1984a). The sequence analysis confirms this observation. The 8-bp deletion in the third exon, the 1-bp insertion in the seventh exon and the C → T transition in the eighth exon of the 21-OHase A gene generate frameshift and/or nonsense mutation which would prevent synthesis of a functional protein. This conclusion has also been reached by Higashi *et al.* (1986) and White *et al.* (1986). That the 21-OHase A gene is a pseudogene is further substantiated by the absence of detectable 21-OHase A mRNA in an adrenal RNA preparation using oligonucleotides specific for the A and B gene sequences (Higashi *et al.*, 1986).

In the majority of individuals 21-OHase deficiency is not due to gene deletion (Rumsby *et al.*, 1986). In our study individual II.1, who has the salt-wasting form of CAH, is heterozygous for

Table III. Comparison of the amino acid sequences of human 21-OHase at **(A)** the NH_2 terminus and **(B)** residues 261–275 with other 21-OHase sequences

(A)

	1				5					10					15
Porcine 21-OHase (Yuan *et al.*, 1983)	M	V	L	V	W	L	L	L	L	T	T	L	K	A	G
Murine 21-OHase (Chaplin *et al.*, 1986)	M	L	L	P	G	L	L	L	L	L	L	L	L	A	G
Bovine 21-OHase (Chung *et al.*, 1986)	M	V	L	A	G	L	L	L	L	L	T	L	L	A	G
Human 21-OHase B (this work)	M	L	L	L	G	L	L	L	L	L	P	L	L	A	G
Human 21-OHase B (Higashi *et al.*, 1986; White *et al.*, 1986)	M	L	L	L	G	–	L †	L	L	L	P	L	L	A	G

(B)

Porcine 21-OHase		M	L	Q	E	A	G	R	Q	R	V	E	E	G	Q	G	
Bovine 21-OHase	259	M	L	Q	G	V	G	R	Q	R	V	E	E	G	P	G	273
Human 21-OHase B	261	M	L	Q	G	V	A	Q	P	S	M	E	E	G	S	G	275
Human 21-OHase B from CAH patient	261	M	L	Q	G	V	A	Q	P	T *	M	E	E	G	S	G	275

A dash (–) indicates absence of a codon.
† indicates the position at which the 21-OHase cDNA sequence begins.
* indicates the Ser to Thr change in the 21-OHase B gene from the CAH patient.

the HLA haplotypes B47 DR7/B44 DR6 and has a single 21-OHase B gene associated with the B44 DR6 haplotype. The nucleotide sequence of this gene reported here reveals 11 differences from the normal 21-OHase B gene (Figure 2). Two of these differences result in codon changes with Ser-269 and Asn-494 in the normal gene being converted to Thr and Ser respectively in the mutant gene. The Ser to Thr change at codon 269 also introduces a new *Nco*I restriction site into the gene. However, this *Nco*I polymorphism is not restricted to the 21-OHase B gene as it is also found in the 21-OHase A genes of the B44 and B47 haplotypes in this family, though it is absent in all the other reported 21-OHase sequences. Five of the other differences correspond to the nucleotide sequence of the 21-OHase A gene at those positions. These observations suggest partial gene conversion has taken place between a 21-OHase A gene and the 21-OHase B gene. Sequence analysis of the adjoining C4 gene suggests that it too has been converted from C4B to C4A (Yu and Campbell, 1987; C.Y.Yu, N.R.Rodrigues, S.Cross and R.D. Campbell, in preparation). Gene conversion has been proposed to be the cause of CAH in some patients (Donohoue *et al.*, 1986) although in individual II.1, the 21-OHase A-like changes are all silent mutations. It is clear, however, that 21-OHase deficiency in some cases will be due to the introduction of the deleterious mutations found in the 21-OHase A gene into the 21-OHase B gene presumably by gene conversion-like events.

The precise effect of the defect in individual II.1 is yet to be determined. The two nucleotide alterations in the 5′ flanking region, or even those differences in the introns, may lead to no or low expression of the gene. The gene may be fully expressed, but the protein product, due to the two amino acid substitutions, may be inactive. The active site of a cytochrome P-450 contains iron protoporphyrin IX presumably in a hydrophobic cleft in the protein (White and Coon, 1980). The heme, which is not covalently bound to the protein, is always penta- or hexa-coordinated, four of the ligands provided by the N atoms of the porphyrin ring and the fifth ligand is believed to be a specific cysteine residue of the polypeptide chain (Hahn *et al.*, 1982). This cysteine is conserved in all cytochrome P-450 enzymes (Morohashi *et al.*, 1984), corresponding to Cys-429 in 21-OHase. The sixth co-ordination position of the iron is occupied by an easily exchangeable oxygen ligand either from water or the hydroxyl group of a Ser or Thr residue (Ullrich *et al.*, 1979). If the Ser at codon position 269 is indeed involved in the sixth co-ordination position, a change in it would affect the functioning of the protein, although a Thr could provide the necessary group. A comparison with the bovine 21-OHase gene shows that at that position there is no homology between the two sequences although the flanking regions are very homologous (Table IIIB). It remains to be seen whether individuals having a single 21-OHase B gene with the *Nco*I polymorphism are normal. In that case, the second amino acid change from Asn to Ser at the carboxy terminus would be solely responsible for the loss of enzymatic activity. Although there is no conservation of sequence in that region of the gene with 21-OHase sequences from other species, it might be affecting the functional site of the protein in some way. The other differences do not show any evident effect on the enzyme function and obviously more data from different individuals with CAH will provide a clue to their relevance in the gene function.

The presence of allelic variants of 21-OHase at a single locus is strongly suggested by the four different clinical manifestations of CAH. The differential linkage disequilibrium of classical and non-classical CAH to different haplotypes seems to indicate that the less severe form of CAH is an allelic variant of the classical form. The 21-OHase genes lie between a region of the HLA which is highly polymorphic. They lie immediately 3′ to the C4A and C4B genes which show a high degree of polymorphism (Yu *et al.*, 1986). More than 40 alleles have been demonstrated in the two classes by differences in electrophoretic mobilities (Mauff *et al.*, 1983), serological typing (Roos *et al.*, 1984; Giles *et al.*, 1984; Giles, 1987) and direct DNA sequencing (Belt *et al.*, 1984, 1985; Yu *et al.*, 1986). It is quite possible that the 21-OHase genes might exhibit a similar polymorphism. Being a microsomal protein, isolation of 21-OHase is difficult and polymorphism can only be inferred from DNA sequencing data. A comparison of

21-OHase gene sequences from other groups (Higashi *et al.*, 1986; White *et al.*, 1986) with the sequences reported here suggests polymorphism in 21-OHase genes. As shown in Table II, we found 76 differences in the sequences we compared. Most of the differences may represent accumulation of point mutations in the pseudogene. Of the 15 differences in the 21-OHase B gene sequence which lie in exons, five cause a codon change. The Leu at codon 6 (CTG) in our sequence is absent in the sequences from the other two groups. Unfortunately the published cDNA sequence of 21-OHase ends at codon 7 (Table IIIA) and it has not been possible to define whether this amino acid is present in the expressed sequence. However, a comparison of 21-OHase sequences from other species with the human 21- OHase B sequence reveals a Leu at the same position (Table IIIA). This may indicate that there is polymorphism in the size of 21-OHase with either 494 or 495 amino acids. At codon 103 (AGG) we found an Arg residue in the 21-OHase B gene sequence while a Lys (AAG) was indicated from the other two published sequences. At codon 427 (CGC) an Arg residue in our sequence and that of White *et al.* (1986) differs from that of Higashi *et al.* (1986) who identified a Pro residue (CCG). These observations suggest polymorphism in the 21-OHase genes and in the expressed product.

In conclusion, we have determined the structure of the single 21-OHase B gene from an individual with classical CAH. There are two amino acid substitutions in this 21-OHase gene and these may give an insight into which amino acids are important for the function of 21-OHase. We have suggested that the 21-OHase genes are polymorphic. The varying severity of the disease may be due to a heterogeneous collection of mutations within the gene. To have a better understanding of the implication of polymorphism in the 21-OHase genes and to define the molecular basis of this very common genetic disorder in CAH patients, it is clear that further characterization of genes from different individuals is required.

Materials and methods

Preparation and analysis of genomic DNA

Genomic DNA was extracted from whole blood or from EBV-transformed cell lines according to Bell *et al.* (1981). DNA (5 μg) was digested to completion for 24 h with 20 U of the restriction enzymes *Taq*I, *Kpn*I and *Nco*I (Amersham and BRL) respectively, according to the manufacturers' instructions and electrophoresed on 0.7% agarose gels. The gels were treated as described by Wahl *et al.* (1979) before transfer to nitrocellulose filters by the procedure of Southern (1975). After baking for 2 h at 80°C the filters were pre-washed in 1 M NaCl/50 mM Tris–HCl pH 7.5/5 mM EDTA/0.1% $NaDodSO_4$ for 1 h and then prehybridized at 42°C for 6–16 h (Maniatis *et al.*, 1982). Hybridization was carried out at 42°C for 48 h. The probes were labelled with ^{32}P by nick translation (Rigby *et al.*, 1977) or by the random priming method (Feinberg and Vogelstein, 1984). Probe H consists of the 410- and 430-bp *Hin*fI fragments from the 21-OHase gene as shown in Figure 1. They span the 1st, 3rd, 4th and 5th exons. Probe PA is an 850-bp *Pvu*II/Asp-718 fragment from the 3′ end of the gene and covers the 8th, 9th and 10th exons. After hybridization, the blots were washed, dried and autoradiographed as described by Maniatis *et al.* (1982).

Preparation and analysis of cloned DNA

Cosmid and plasmid DNA was extracted from bacterial colonies using the standard alkaline $NaDodSO_4$ method of Birnboim and Doly (1979). DNA from λ clones was prepared using the plate lysate method (Maniatis *et al.*, 1982). Detailed restriction maps were constructed after single and double digests and Southern blot analysis of the cloned DNA (Southern, 1975). Appropriate restriction fragments were purified by separation on 0.8% low-gelling-temperature agarose (Sigma Type VII) by the method of Feinberg and Vogelstein (1983) or by elution from 4% polyacrylamide gels (Maxam and Gilbert, 1980). Preliminary subcloning was done by blunt-end ligation into pAT153/*Pvu*II/8 (Anson *et al.*, 1984) which had been digested previously with *Pvu*II and treated with alkaline phosphatase. Colonies containing recombinant plasmids were identified by colony hybridization. Further subcloning was done into the bacteriophage vectors M13mp8 and M13mp9 (Messing and Vieira, 1982) using standard methods (Maniatis *et al.*, 1982). The complete genomic DNA sequences of the normal 21-OHase B gene, the 21-OHase A gene and the 21-OHase B gene from the patient with CAH (individual II.1) were obtained by Sanger's dideoxy chain termination method (Sanger *et al.*, 1977). The sequencing strategy is that shown in Figure 1. All restriction sites were overlapped at least once, and the DNA fragments were sequenced twice and in both orientations.

Acknowledgements

We thank Dick Cotton, Ken Reid and Bob Sim for critical review of the manuscript. We are grateful to Ms Carolyn Brooks and Mr Frank Caddick for their help in preparing the manuscript, and to Ms Sandra Smith for her excellent technical assistance. We deeply regret the tragic death of the beloved Professor R.R.Porter and would like to use this occasion to pay our respects to the Professor for his valuable contribution to this work. N.R. is a Rhodes Scholar, I.D. holds an SERC studentship and C.Y.Y. was a Commonwealth Scholar.

References

Anson,D.S., Choo,K.H., Rees,D.J.G., Giannelli,F., Gould,K., Huddleston,J.A. and Brownlee,G.G. (1984) *EMBO J.*, **3**, 1053–1060.

Bell,G.I., Karam,J.H. and Rutter,W.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5759–5763.

Belt,K.T., Carroll,M.C. and Porter,R.R. (1984) *Cell*, **36**, 907–914.

Belt,K.T., Yu,C.Y., Carroll,M.C. and Porter,R.R. (1985) *Immunogenetics*, **21**, 173–180.

Biggin,M.D., Gibson,T.J. and Hong,G.F (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.

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

Carroll,M.C., Belt,K.T., Palsdottir,A. and Porter,R.R. (1984a) *Phil. Trans. Soc. Lond., B.*, **306**, 379–388.

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

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

Carroll,M.C., Palsdottir,A., Belt,K.T. and Porter,R.R. (1985b) *EMBO J.*, **4**, 2547–2552.

Chaplin,D.D., Galbraith,L.J., Seidman,J.G., White,P.C. and Parker,K.L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9601–9605.

Chung,B., Matteson,K.J. and Miller,W.L. (1985) *Proc. Natl. Acad. Sci. USA*, **83**, 4243–4247.

Cooper,D.Y., Levin,S., Narasimhulu,S., Rosenthal,O. and Estabrook,R.W. (1965) *Science*, **147**, 400–402.

Donohoue,P.A., Van Dop,C., McLean,R.H., White,P.C., Jospe,N. and Migeon, C.J. (1986) *J. Clin. Endocrinol. Metab.*, **62**, 995–1002.

Dupont,B., Oberfield,S.E., Smithwick,E.M., Lee,T.D. and Levine,L.S. (1977) *Lancet*, **ii**, 1309–1311.

Dupont,B., Pollack,M.S., Levine,L.S., O'Neill,G.J., Hawkins,B.R. and New, M.I. (1980) In Terasaki,P.I. (ed.), *Histocompatibility Testing*. UCLA Press, pp. 693–706.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

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

Finkelstein,M. and Shaefer,J.M. (1979) *Physiol. Rev.*, **59**, 353–406.

Giles,C.M. (1987) *Human Immunol.*, in press.

Giles,C.M., Batchelor,J.R., Dodi,I.A., Fielder,A.H.L., Rittner,C., Mauff,G., Bender,K., Levene,C., Schreuder,G.M.T. and Well,L.J. (1984) *J. Immunogenet.*, **11**, 305–317.

Hahn,J.E., Hodgson,K.O., Anderson,L.A. and Dawson,J.H. (1982) *J. Biol. Chem.*, **257**, 10934–10941.

Higashi,Y., Yoshioka,H., Yamane,M., Gotoh,O. and Fujii-Kuriyama,Y. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2841–2845.

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

Mauff,G., Alper,C.A., Awdeh,Z., Batchelor,J.R., Bertrams,J., Bruun-Petersen, G., Dawkins,R.L., Demant,P., Edwards,J., Grosse-Wilde,G., Hauptmann,G., Klouda,P., Lamm,L., Mollenhauer,E., Nerl,C., Olaissen,B., O'Neill,G., Rittner,C., Roos,M.H., Skanes,V., Teisberg,P. and Wells,L. (1983) *Immunobiol.*, **164**, 184–191.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.

Morohashi,K., Fujii-Kuriyama,Y., Okada,Y., Sogawa,K., Hirose,T., Inayama,S. and Omura,T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4647–4651.

New,M.I. and Levine,L.S. (1984) *Annu. Rev. Med.*, **35**, 649–663.

New,M.I., Dupont,B., Grumback,K. and Levine,L.S. (1982) In Stanbury,J.B., Wyngarden,J.B., Fredrickson,D.S., Goldstein,J.L. and Brown,M.S. (eds), *Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, pp. 973–1000.

Nussinov,R. (1986) *Biochim. Biophys. Acta*, **866**, 93–108.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Roos,M.H., Giles,C.M., Demant,P., Mollenhauer,E. and Rittner,C. (1984) *J. Immunol.*, **133**, 2634–2640.
Ruff,H.H., Wende,P. and Ullrich,V (1979) *J. Inorg. Biochem.*, **11**, 189–204.
Rumsby,G., Carroll,M.C., Porter,R.R., Grant,D.B. and Hjelm,M. (1986) *J. Med. Genet.*, **23**, 204–209.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5468.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Speiser,P.W., Dupont,B., Rubenstein,P., Piazza,A., Kastelan,A. and New,M.I. (1985) *Am. J. Hum. Genet.*, **37**, 650–667.
Ullrich,V., Sakurai,H. and Ruff,H.H. (1979) *Acta Biol. Med. Ger.*, **39**, 287–297.
Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3686–3687.
White,R.E. and Coon,M.J. (1980) *Annu. Rev. Biochem.*, **49**, 315–356.
White,P.C., New,M.I. and Dupont,B. (1984a) *Proc. Natl. Acad. Sci. USA*, **81**, 7505–7509.
White,P.C., New,M.I. and Dupont,B. (1984b) *Proc. Natl. Acad. Sci. USA*, **81**, 1986–1990.
White,P.C., Grossberg,D., Onufer,B.J., Chaplin,D.D. and New,M.I. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1089–1093.
White,P.C., New,M.I. and Dupont,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5111–5115.
Yoshioka,H., Morohashi,K., Sogawa,K., Yamane,M., Kominami,S., Takemori, S., Okada,Y., Omura,T. and Fujii-Kuriyama,Y. (1986) *J. Biol. Chem.*, **261**, 4106–4109.
Yu,C.Y. and Campbell,R.D. (1987) *Immunogenetics*, in press.
Yu,C.Y., Belt,K.T., Giles,C.M., Campbell,R.D. and Porter,R.R. (1986) *EMBO J.*, **5**, 2873–2881.
Yuan,P., Nakajin,S., Haniu,M., Shinoda,M., Hall,P.F. and Shively,J.E. (1983) *Biochemistry*, **22**, 143–149.

Received on February 26, 1987

Experimental and Clinical Immunogenetics
Editor-in-Chief: K. Bauer, Heidelberg

Reprint
Publisher: S. Karger AG, Basel
Printed in Switzerland

Genetics of Complement
Expl clin. Immunogenet. *5:* 81–98 (1988)

© 1988 S. Karger AG, Basel
0254-9670/88/0053-0081$2.75/0

Molecular Mapping of the HLA-Linked Complement Genes and the RCA Linkage Group

R.D. Campbell, I. Dunham, C.A. Sargent

MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, UK

Key Words. HLA-linked complement genes · RCA linkage group · Complement · Molecular mapping · Genes

Abstract. Phenotypic genetics have established linkage of the genes encoding proteins involved in the activation of the complement component C3. C2, factor B and C4, three of the structural components of the classical and alternative pathway C3 convertases, are encoded by genes which have been mapped to the class III region of the major histocompatibility complex (MHC) on human chromosome 6. The regulatory proteins factor H, C4BP, CR1, CR2 and DAF, which are involved in the control of C3 convertase activity, are encoded by closely linked genes, termed the regulators of complement activation (RCA) linkage group, that have been mapped to human chromosome 1. cDNA clones for all these proteins have been isolated, and this has made it possible to investigate the organization and structure of the MHC class III genes and the genes in the RCA linkage group. This short review summarizes some of the main features which have emerged from recent cloning work.

Introduction

The application of recombinant DNA techniques to the study of components of the complement system over the last few years has resulted in the cloning of cDNA and in most cases genomic DNA encoding these proteins [Campbell et al., 1988]. In many cases the chromosomal assignment has also been established [Reid, 1985]. These cloning studies have provided information which emphasizes that the members of the complement system can be divided into families of structurally and functionally related proteins, many of which are encoded by closely linked genes on the same chromosome.

Activation of the complement system occurs via two pathways, the classical and alternative, with the major component C3 playing a central role in each [Reid and Porter, 1981; Müller-Eberhard and Miescher, 1985; Reid, 1986]. In the classical pathway, activation is triggered primarily by the binding of C1 to IgG, or IgM, in immune complexes [Schumaker et al., 1987]. This results in the generation of proteolytic ac-

tivity in $C\overline{1}$ which cleaves and activates C4 to C4a and C4b. C2 associates with C4b and is cleaved by $C\overline{1}$ to C2a and C2b, thus yielding the C3 convertase $C\overline{4}b\overline{2}a$ which cleaves and activates C3. Activation of the alternative pathway can be mediated by a wide range of substances such as high molecular weight polysaccharides found in microorganisms, and also by complexes of IgG, IgA and IgE. Efficient activators are considered to possess sites for C3b where it is protected from control by factor I and its cofactors. Factor B associates with C3b, or $C3(H_2O)$ (a C3b-like form of C3), and is cleaved by factor $\overline{D}$ to Ba and Bb, thus yielding the C3 convertase $C\overline{3}b\overline{B}b$. This complex enzyme is stabilized by properdin and this causes potentiation of alternative pathway activation. The binding of C3b to $C\overline{4}b\overline{2}a$, or additional molecules of C3b to $C\overline{3}b\overline{B}b$, results in the generation of C5 convertase which activates C5 and initiates the self-assembly of the terminal components C5b to C9 involved in membrane lysis [Müller-Eberhard, 1986].

Control of the complement system is mediated in a number of different ways to prevent damage to the animal's own tissue. The C3/C5 convertases have a short half-life due to dissociation of C2a and Bb. In addition, the activity of the C3/C5 convertases is regulated by a number of proteins found in plasma, and by membrane-bound proteins and receptors [Holers et al., 1985; Sim et al., 1986; Reid et al., 1986; Kristensen et al., 1987]. These regulatory proteins include the plasma proteins factor H and C4b-binding protein (C4BP), and the membrane bound

Table I. Proteins involved in the activation of C3 and C5 [adapted from Law and Reid, 1988]

Generation of C3/C5 convertase

	MW kd	Plasma concentration		Chromosome location of gene	Enzymic site in activated form (+) (and natural substrate split)
		μg/ml	μ*M*		
Classical pathway					
C1-C1q[a]	462	80	0.17	1	–
C1r	83	50	0.30	12	+ (C1r, C1s)
C1s	83	50	0.30	12	+ (C4, C2)
C4	205	600	3.00	6	–
C2	102	20	0.20	6	+ (C3, C5)
C3	185	1,300	7.02	19	–
Alternative pathway					
Factor D	24	1	0.04	n.k.	+ (B)
Factor B	92	210	2.20	6	+ (C3, C5)
C3	185	1,300	7.02	19	–

[a] C1q is composed of 18 chains (6A+6B+6C). The genes for the A and B chains are on chromosome 1. The gene for the C chain has not yet been mapped.

Table 1 (continued)

Plasma proteins involved in control of C3/C5 convertase

Protein	MW kd	Plasma concentration		Specificity	Chromosome location of gene	Role
		µg/ml	µM			
C4-binding protein	500	250	0.45	C4b	1	accelerates decay of $C\overline{4}b\overline{2}a$ and acts as cofactor in the cleavage of C4b by factor I
Factor H	150	480	3.20	C3b	1	accelerates decay of $C\overline{3}b\overline{B}b$ and acts as cofactor in the cleavage of C3b by factor I
Factor I	88	35	0.39	C4b, C3b	4	protease which inactivates C4b and C3b with the aid of cofactors C4BP, H, CR1 and MCP
Properdin	220	20	0.09	C3bBb	X	positive regulator of the alternative pathway which stabilizes the C3/C5 convertases

Membrane proteins involved in control of C3/C5 convertase

Membrane molecule	MW kb	Fragment specificity	Chromosome location of gene	Principal roles	Major human cell types positive[b]
Complement receptor type 1	type D 250 type B 220 type A 190 type C 160 (four structural allotypes)	C3b, C4b	1	regulation of C3b breakdown, binding of immune complexes to erythrocytes, phagocytosis, accelerates decay of C3/C5 convertases	E, B, G, M
Complement receptor type 2	145	C3d, C3dg, iC3b	1	regulation of B cell functions, Epstein-Barr virus receptor	B
Membrane cofactor protein	45–70	C3b, C4b	n.k.	regulation of C3b breakdown	B, T, N, M
Decay-accelerating factor	70	C4b2a, C3bBb	1	accelerates decay of C3/C5 convertases	E, L, P

[b] Human cell types: B, B lymphocytes; E, erythrocytes; G, granulocytes; L, leucocytes; M, monocytes; N, neutrophils; P, platelets.

n.k. = not known.

molecules complement receptor type 1 (CR1), complement receptor type 2 (CR2), decay accelerating factor (DAF) and probably also membrane cofactor protein (MCP). A summary of some of the features of the proteins involved in the generation of the C3/C5 convertases and in the control of C3/C5 convertase activity can be found in table I.

Investigation of the phenotypic genetics of complement components [Hobart, 1984] and of inherited deficiencies of the components [Lachmann, 1984] have established linkage of the genes encoding proteins involved in the activation of C3. Three of the structural components of the C3 convertases, C2, factor B and C4, are encoded by genes which have been mapped to the class III region of the major histocompatibility complex (MHC) on the short arm of human chromosome 6 [Alper, 1981; Campbell et al., 1986]. Phenotypic genetics have also established linkage of the genes encoding regulatory proteins involved in the control of C3 convertase activity [Rodriguez de Cordoba et al., 1985]. This linkage group, termed the regulators of complement activation (RCA) linkage group, maps to human chromosome 1 and includes the genes for factor H, C4BP, CR1, CR2 and DAF.

Several comprehensive reviews dealing with the complement system [Müller-Eberhard and Miescher, 1985; Reid, 1986], the genetics of components of complement [Reid, 1985; Campbell et al., 1986, 1988], and the polymorphism of C2, factor B [Campbell, 1987] and C4 [Carroll and Alper, 1987] have been published recently. This short review will summarize some of the main features which have emerged from recent cloning work. It covers the organization and structure of the MHC class III genes and the genes in the RCA linkage group.

MHC-Linked Complement Genes

The human MHC is located on the short arm of chromosome 6 in the distal portion of the 6p21.3 band [Lamm and Olaisen, 1985]. It consists of three major linked gene clusters. The class I loci (HLA-A, -B, -C) encode cell surface glycoproteins found on almost all nucleated cells [Strachan, 1987], whereas the class II loci (HLA-DP, -DQ, -DR) encode cell surface glycoproteins found principally on B lymphocytes, macrophages and activated T cells [Trowsdale, 1987]. Both class I and class II antigens are highly polymorphic and act as restriction elements in the recognition and interaction of regulatory and effector T lymphocytes with their target cells. Analysis of recombinant MHC haplotypes in family studies has established that the class I loci are telomeric to the class II loci. Within the class II region the DP subregion maps on the centromeric side of the DQ and DR subregions, while the order of the loci within the class I region is HLA-B, HLA-C, HLA-A, telomere.

The first evidence for the existence of complement genes within the MHC was provided by Allen [1974] who demonstrated in family studies that the electrophoretic polymorphism of factor B [Alper et al., 1972] segregated with the MHC. The linkage of C2 and of C4 to the MHC followed from the description of deficiencies of these proteins in plasma [Fu et al., 1974; Hauptmann et al., 1974; Rittner et al., 1975], and from the demonstration of electrophoretic variants [Alper, 1976; Hobart and Lachmann, 1976; Meo et al., 1977; Teisberg et al., 1976, 1977]. However, in order to explain the electrophoretic patterns observed for C4, O'Neill et al. [1978a, b] proposed a two locus model for C4, now referred to as C4A and C4B. Studies

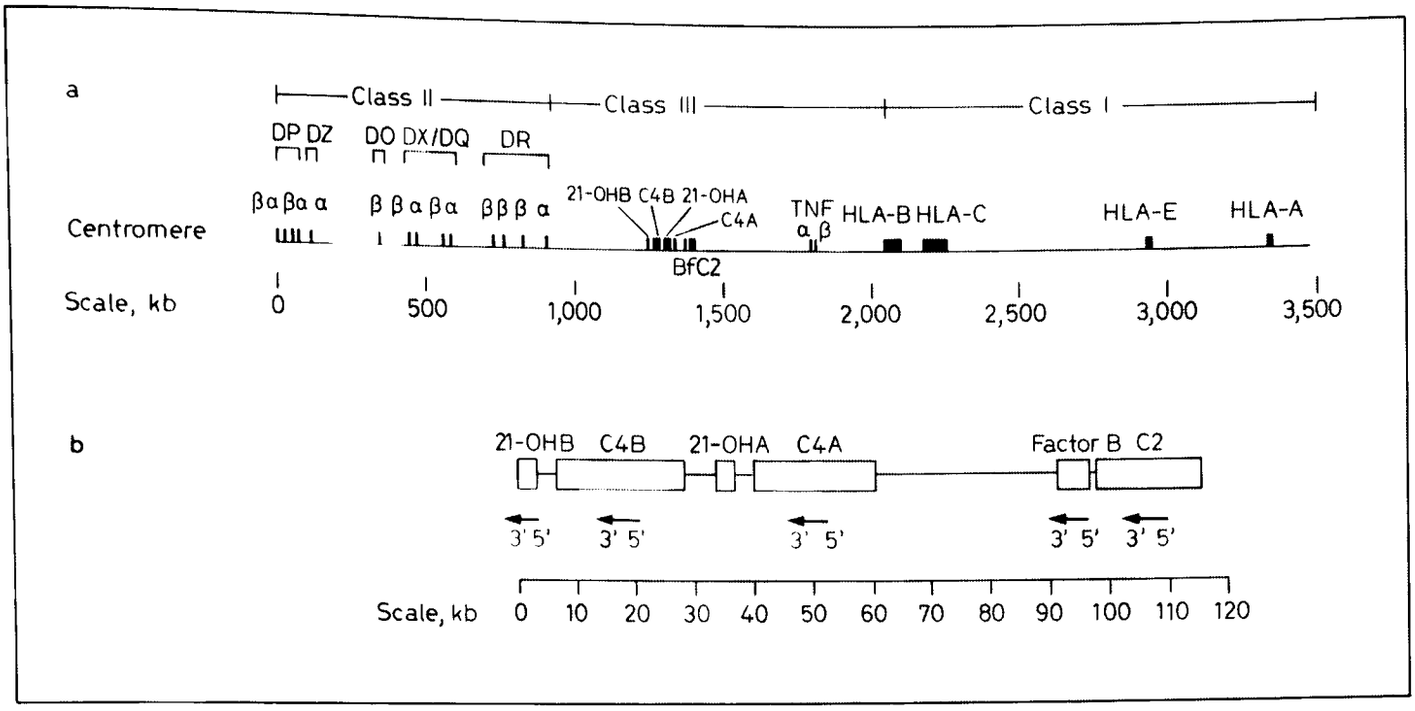


Fig. 1. Molecular map of the human MHC. **a** The molecular map has been derived from PFGE and cosmid cloning experiments which can be found in Hardy et al. [1986], Carroll et al. [1987b] and Dunham et al. [1987]. **b** Expanded map of the complement and 21-OHase (21-OH) loci in the class III region. The solid arrows denote the direction of transcription of the genes.

of the inheritance of the polymorphic variants of C2, factor B and C4, and various recombinant MHC haplotypes have placed the structural genes for these proteins (termed the class III genes) between the HLA-B and HLA-DR loci [Olaisen et al., 1983; Lamm and Olaisen, 1985]. However, despite a large number of studies [Alper et al., 1983] no recombination events have been observed between the genes, and this was taken to suggest that they were closely linked. Further, it was not possible to define the order of the genes relative to one another. The cloning of cDNA for C2, factor B and C4 has made it possible to study the genetic organization of the corresponding genes.

Molecular Cloning of C2, Factor B and C4

The isolation of cDNA clones for C2, factor B and C4 has been achieved through the screening of liver cDNA libraries with oligonucleotides synthesized on the basis of the known protein sequence. These cDNA clones have been used to screen cosmid libraries of human genomic DNA to isolate clusters of overlapping cosmid clones containing the corresponding genes [Carroll et al., 1984b]. Characterization of the cosmids revealed that the C2 and factor B genes were closely linked, and it has since been shown that the transcription start-point of the factor B gene lies only 421 bp from the polyadenylation site of the C2 gene [Wu et al., 1987]. Approximately 30 kb from the factor B gene are the C4 loci which are separated by ~ 10 kb of DNA [Carroll et al., 1984b]. Analysis of the cosmid clones using C4A and C4B class-specific synthetic oligonucleotides showed that the gene encoding C4A lay closer to the factor B gene (fig. 1) [Carroll et al., 1984a].

Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase (21-OHase) deficiency is an autosomal-recessive genetic trait. Close linkage of 21-OHase deficiency with HLA was demonstrated by Dupont et al. [1977]. Additional studies of families with intra-HLA recombinant haplotypes confirmed this linkage and suggested that the 21-OHase deficiency gene was located between HLA-B and HLA-DR [Dupont et al., 1980]. Detailed restriction mapping of the cosmids containing the complement genes revealed the presence of two 21-hydroxylase genes [Carroll et al., 1985a; White et al., 1985]. The genes were mapped to within 3 kb of the 3′ end of each C4 gene (fig. 1). However, only the gene lying 3′ to the C4B locus, termed the 21-OHase B gene, appears to be important in steroid biogenesis in the adrenal gland. Homozygous deletion of the 21-OHase B gene has been described in some individuals suffering from CAH [White et al., 1984]. In addition, DNA sequencing of both genes has established that, although 97% homologous, the 21-OHase A gene is a pseudogene due to deleterious mutations in the coding sequence [Higashi et al., 1986; White et al., 1986; Rodrigues et al., 1987].

The complete primary sequences of C2 [Bentley, 1986] and factor B [Morley and Campbell, 1984; Mole et al., 1984] have been derived from cDNA sequencing. Comparison of the cDNA and amino acid sequences indicates that they share 42% nucleotide and 39% amino acid identity. C2 and factor B perform analogous functions in the classical and alternative pathway C3 convertases. The overall similarity in structure and function of the two proteins indicates that the C2 and factor B genes arose by duplication of an ancestral gene, and the two loci have been maintained in exceptionally close linkage. Determination of the factor B gene structure has shown that it is 6 kb in length [Campbell et al., 1984] and this is substantially shorter than the C2 gene which spans 18 kb of DNA [Bentley et al., 1985]. Both genes, however, encode mRNA species of similar size. The factor B mRNA is ~2.6 kb and the C2 mRNA is ~2.9 kb. Thus the major difference in size of the two genes must be due to the C2 gene having longer introns.

The complete primary sequence of C4A has been derived from cDNA sequencing [Belt et al., 1984]. Comparison with C4B cDNA has revealed that the two C4 isotypes are >99% homologous. Of the 14 differences which were observed, 12 are clustered in a region of the molecule derived from proteolytic cleavage of C4b by factor I and called C4d. These differences cause 9 amino acid substitutions. Comparison of the sequences of 9 different C4A and C4B cDNA and genomic clones in the C4d region has established the pattern of polymorphism in the C4d fragment [Belt et al., 1985; Yu et al., 1986] and has provided a structural basis for the observed serological [see Giles, this volume] and functional differences [Law et al., 1984; Isenman and Young, 1984] between the two C4 isotypes.

The two C4 loci probably arose through a single duplication event which encompassed ~28 kb of DNA. They are often referred to as locus I and locus II. In general, locus I encodes C4A and locus II encodes C4B, though exceptions to this are rather common. Although both are transcribed into mRNA species of ~5.5 kb, a difference in the size of the loci has been observed. Long C4 genes are 22 kb in length while short C4 genes are 16 kb in length. Of the C4A genes

analysed to date all are ~ 22 kb in length [Yu et al., 1986; Schneider et al., 1986; Palsdottir et al., 1987b]. One exception to this has been suggested by Giles et al. [1987] who found a rare haplotype in a French family which expresses two C4A allotypes. Examination of the C4 genes revealed that a short gene at the second C4 locus encoded a C4A protein [see Giles, this volume]. However, no short C4A gene has yet been defined at locus 1. On the other hand, C4B genes can be 22 kb or 16 kb in length due to the presence or absence of a 6–7 kb intron about 2.5 kb from the 5′ end of the gene [Yu et al., 1986; Schneider et al., 1986; Palsdottir et al., 1987b]. Long C4B genes included C4B1 genes on most C4A3 C4B1 haplotypes and also C4BQ0 alleles. An estimate of the frequency of C4B loci with the 6.5 kb intron, based on haplotype frequencies, suggests they are twice as frequent as C4B loci without the intron. The nature of the 6–7 kb intron is not known though it has been suggested that it could be a member of the long interspersed sequences, LINE or LI [Yu et al., 1986]. These sequences are repeated about 10^4 times in the haploid genome and have homology to retroviral reverse transcriptase (retroposons) [Singer and Skowronski, 1985]. Although 80% of haplotypes carry two C4 loci, differences have also been observed in the number of C4 genes present. This was originally demonstrated at the protein level where duplication of C4B as C4B1, C4B2 was found on the extended haplotype B14 DR1 [Raum et al., 1984; Rittner et al., 1984; Uring-Lambert et al., 1984]. The gene frequency has been estimated at 1–2%, though this may be an underestimate as three new kinds of C4 gene duplication were identified in family studies by combined protein and RFLP analysis [Schneider et al., 1986]. Duplication of C4 has also been observed from cosmid cloning where three C4 genes, one C4A and two C4B, were found on one haplotype [Carroll et al., 1984a]. Haplotypes with a single C4 gene have also been characterized and these will be dealt with later in relation to C4 null alleles.

Genetic studies in man have demonstrated deficiency of C2, C4 and in a few rare cases heterozygous deficiency of factor B. By far the most prevalent disorder is deficiency of C2. The incidence of the homozygous deficiency is 1 in 10,000 [Agnello, 1978]. The C2 deficiency allele is usually found in a specific HLA extended haplotype HLA B18 C4A4 C4B2 BfS C2QO DR2 [Awdeh et al., 1981], suggesting that most C2-deficient patients will have the same mutation. No major gene deletion or rearrangement of the C2 gene has been observed by genomic Southern blot analysis in individuals deficient in C2 [Cole et al., 1985]. A defect at the level of transcription of the gene or post-transcriptional processing of C2 mRNA is the most likely reason as no detectable C2 mRNA could be found by Northern blot analysis of peripheral blood monocyte RNA preparations from C2-deficient individuals.

Complete deficiency of C4 in man is a rare disorder, and only 16 cases have been reported [Hauptmann et al., 1986]. Homozygous C4 deficiency is associated with different HLA haplotypes suggesting that different mutations may have occurred. Three complete C4-deficient patients have been analysed by Southern blot analysis [Hauptmann et al., 1987]. All 3 patients had C4A genes present, 2 showed deletion of C4B genes, while the third appeared normal at the DNA level with two C4 loci on each haplotype. This clearly demonstrates that the complete absence of C4 in plasma is due to abnormal-

ities in transcription or translation of the genes, in addition to single gene deletions.

Occurrence of null alleles in the population at either locus defined by the absence of C4A or C4B in plasma is much more common. Gene frequencies of 5–15% for C4AQO alleles and 10–20% for C4BQO alleles [Schendel et al., 1984; Partanen and Koskimies, 1986] have been estimated. About half of the null alleles are due to deletion of the gene usually together with the flanking 21-OHase gene [Carroll et al., 1985b; Schneider et al., 1986]. These single C4 gene haplotypes can be of three types: long, expressing C4A protein; long, expressing C4B protein; or short, expressing C4B protein. In other situations where the null allele is present, defects in transcription or translation as discussed above will be the cause of the absence of C4 protein. However, another possibility is that some of the genes are transcriptionally active, but encode products similar to the adjacent locus. This may be the case for the C4BQO allele on the HLA haplotype B44 C2C BfS C4A3 C4BQO DR6. Yu and Campbell [1987] were able to define an RFLP using NlaIV which distinguishes alleles encoding the C4A or C4B isotypes. It was found that the C4BQO allele had an RFLP pattern consistent with it encoding a C4A isotype. It was suggested on the basis of the phenotypes expressed by the individuals concerned that the C4BQO allele probably encoded another C4A3 allotype.

Haplotypes expressing two different C4A allotypes, e.g. C4A3 C4A2 or C4A5 C4A2 with a null allele at the C4B locus, have been found in several studies [Raum et al., 1984; Uring-Lambert et al., 1984]. It was suggested that they may be due to unequal crossover between homologous chromosomes, resulting in duplication of the C4A locus and deletion of the C4B locus [Rittner et al., 1984]. In a detailed Southern blot analysis of such haplotypes by Palsdottir et al. [1987a] it was found that both loci are present as normal, but that the second (or C4B) locus encodes a C4A protein. However, in another example of a C4A3 A2 BQO haplotype studied by Giles et al. [1987] the presence of three C4 loci has been postulated to account for the intensities of fragments seen in Southern blot analysis.

Molecular Mapping of the HLA Class III Region

A number of studies have been arried out to try and define the position and orientation of the complement genes in the class III region, but the results have been contradictory [see Lamm and Olaisen, 1985]. Whitehead et al. [1985] using deletion mutant cell lines and a C4 probe were able to physically map the C4, and thus the C2 and factor B genes between HLA-B and HLA-DR, and suggested on the basis of the number of cell lines which lost or retained the C4 loci that the complement cluster may be closer to HLA-B than HLA-DR. However, in two recent studies in families with recombinant HLA haplotypes the complement loci segregated more frequently with HLA-DR suggesting that they may be closer to the class II region than the class I region [Robinson et al., 1985; Yunis et al., 1985]. Although clusters of overlapping clones have been isolated from the subregions of the class II region [Trowsdale, 1987] and from the class III [Carroll et al., 1984b; Dunham et al., 1987] and class I [Strachan, 1987] regions, they have not yet been linked by chromosomal walking procedures. Recent advances, however, in the separation of large DNA frag-

ments by gel electrophoresis have made it possible to establish the physical linkage of loci separated by hundreds of kilobases. Pulsed field gel electrophoresis (PFGE), originally developed by Schwartz and Cantor [1984], can be used to separate DNA fragments up to 9,000 kb [Anand, 1986]. Such large DNA fragments can be generated from mammalian genomic DNA by restriction enzymes that cut rarely in the genome. Suitable rare-cutting enzymes are those with 6- or 8-base pair recognition sites that contain one or more CpG dinucleotides since CpG is known to be underrepresented in bulk mammalian DNA [Bird, 1986]. In addition, many enzymes with CpG in their recognition sequence appear to be sensitive to cytosine methylation [Brown and Bird, 1986], and this decreases the frequency at which these enzymes cut in the genome. Thus large DNA fragments generated with these enzymes can be separated by PFGE, and Southern blot hybridization analysis can be performed with probes from the area of interest to construct long-range genomic restriction maps.

Recently these procedures have been applied to the molecular mapping of the human MHC [Ragoussis et al., 1986; Hardy et al., 1986; Lawrance et al., 1987; Carroll et al., 1987b, Dunham et al., 1987]. Physical linkage of the class II and class III loci was established in a NotI fragment of ~980 kb [Dunham, et al., 1987]. This fragment is digested with NruI to yield a fragment of 700 kb that hybridizes with the DQα and DRα probes, and a fragment of 280 kb which contains the complement loci. It should be noted that the size of the NotI fragment has been estimated by others to be 920–1,000 kb. The difference in the estimation of the size of this fragment probably reflects the use of different apparatus in the pulsed field work. However, three different HLA haplotypes have also been studied and it remains possible that actual differences in fragment size exist due to RFLPs and/or insertions or deletions of DNA.

In order to orientate the complement loci it was necessary to isolate further overlapping cosmid clones as none of the restriction enzymes used in the PFGE analysis were found to cut between these genes. A series of overlapping cosmids extending from the C2 gene were mapped for NotI, MluI, NruI, and ClaI and this revealed a cluster of sites (three MluI, two ClaI, one NruI and one NotI) in a 25 kb region ~25 kb from the transcriptional start site of the C2 gene [Dunham et al., 1987]. A new single-copy hybridization probe distal to this cluster of sites and the C2 gene was isolated. This probe hybridizes to a NotI fragment of ~210 kb and not the 980 kb NotI fragment containing the complement and DRα loci. Since the exact position of the NotI site that is cleaved in genomic DNA is known from the cosmid map, this established that the C2 gene lies telomeric to the 21-OHase B gene [Dunham et al., 1987]. The results of further single and double digests established that the class III region spans ~1,100 kb and that the C2 gene lies ~650 kb from the class I region, while the 21-OHase B gene lies ~300–360 kb from the class II region (fig. 1). The physical size of the class III region is similar to that estimated from genetic recombination data (1 cM) assuming 1 cM = 1,000 kb.

Analysis of deletion mutant cell lines has recently established linkage of the genes encoding the cytokines tumour necrosis factors α and β to the MHC [Spies et al., 1986], but it was not possible to determine their exact location. A number of studies using PFGE have established that the TNFα and β genes

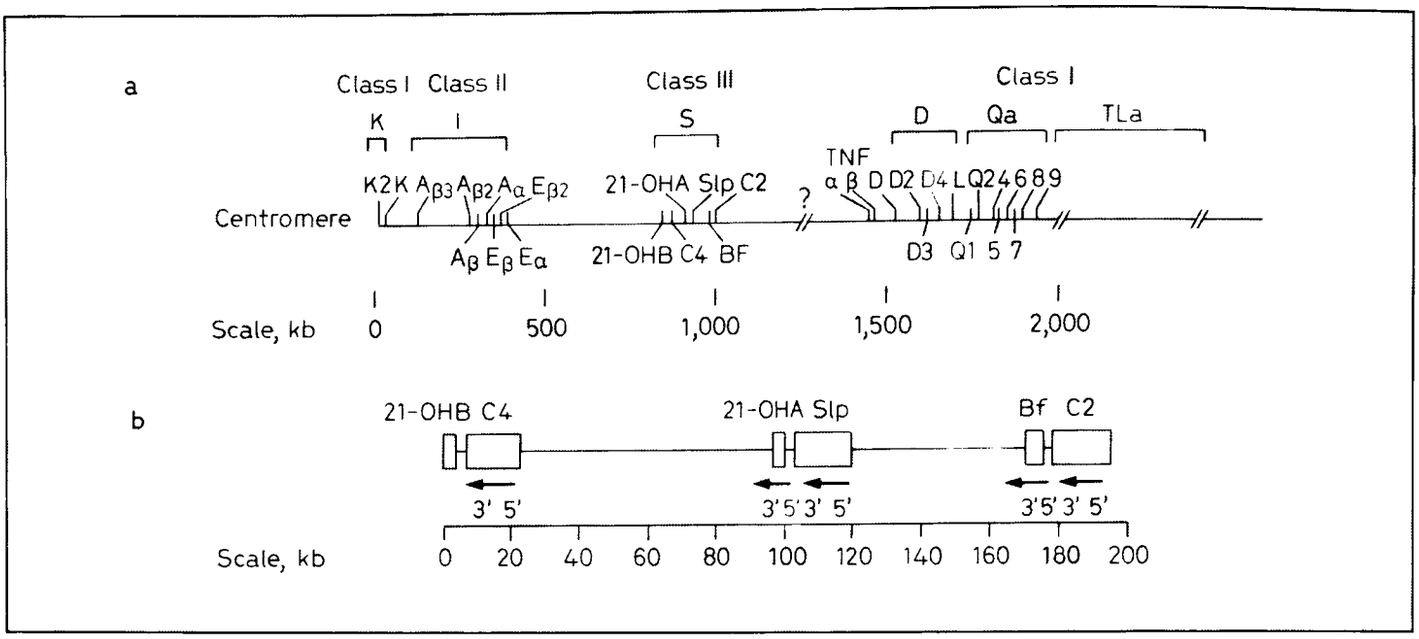


Fig. 2. Molecular map of the mouse H-2 region. **a** The molecular map has been derived from PFGE and cosmid cloning experiments which can be found in Chaplin [1985] and Müller et al. [1987b]. **b** Expanded map of the complement and 21-OHase (21-OH) loci in the S region. The solid arrows denote the direction of transcription of the genes.

are linked to the HLA-B locus [Carroll et al., 1987b; Dunham et al., 1987; Inoko and Trowsdale, 1987; Ragoussis et al., 1988], and lie between HLA-B and the C2 gene (fig. 1). Estimates of ~250 kb have been suggested for the distance between the TNF genes and HLA-B, and these are in reasonable agreement with the 210 kb which has been established through cosmid walking [cited in Carroll et al., 1987b].

Further analysis by PFGE using a number of class I and class II probes has established a long-range map of the MHC (fig. 1). This indicates that the MHC could span 3–4 Mbp of DNA [Carroll et al., 1987b; Dunham et al., 1987], which represents about 1/750 of the human genome. This is in good agreement with the overall size of the MHC of 3–4 cM which has been suggested from studies of recombination events in families and population data on linkage disequilibrium [Lamm and Olaisen, 1985].

The linkage of the C2, factor B and C4 genes to the human MHC appears not to be fortuitous, as they have been mapped to the MHC of a number of other animal species including frog. Of these, the best characterized is the mouse H-2 complex where an analysis by cosmid cloning [Chaplin, 1985] and PFGE [Müller et al., 1987b] has revealed that the organization and orientation of the complement loci is the same as that in man (fig. 2). The single C2 and factor B genes were found to be closely linked [Chaplin, 1985] and to lie ~50 kb from two C4-like genes separated by ~80 kb. It was later established that the gene closest to factor B encoded Slp [Chaplin et al., 1986]. Two genes encoding 21-OHase were also placed immediately 3′ of the Slp and C4 genes, respectively; in contrast to the situation in humans, however, it is the 21-OHase A gene that is important in steroid biogenesis [Chaplin et al., 1986].

Estimates based on PFGE placed the C4 gene ~420 kb from the Eα gene, and the C2 gene at least 470 kb from H-2D [Müller et al., 1987b] (fig. 2). Although genes for TNFα and β were shown to lie between the C2 and H-2D genes [Müller et al., 1987a], they are only 70 kb from the mouse class I gene compared with ~250 kb for the analogous genes in the human MHC.

One of the interesting features of the genes in the class III region is that they are organized in pairs of related genes and have the same 5′ to 3′ orientation. The C2 and factor B genes probably arose through duplication of an ancestral gene. However, the marked divergence of the intron lengths between the two genes and the fact that the two gene products share only 39% sequence identity suggests that the duplication occurred a considerable time ago. The TNFα and β genes also probably arose through duplication. The two genes are only 1.1 kb apart, and they are of similar size. However, the two genes differ significantly in the 5′-untranslated sequence and in the region encoding the signal peptide. In addition, they share only 35% sequence homology which again suggests that the duplication occurred a considerable time ago.

The finding of two 21-OHase genes closely linked to the two C4 genes suggests that the ancestral 21-OHase gene became linked to the ancestral C4 gene prior to duplication. Comparison of the C4A and C4B coding sequence has shown that they differ by less than 1%. The 21-OHase genes are also highly homologous and differ by only about 3%. The strong homology of these two sets of genes suggests that they have only recently duplicated. The duplication may have occurred after mammalian speciation, as only single C4 genes, and presumably also single 21-OHase genes, have been observed in the guinea pig and the Syrian hamster.

Based on the density of genes which have been identified in the cloned portions of the MHC it is highly likely that other genes will be defined in the large regions of DNA separating the class I, II and III loci. One such gene may be neuraminidase which has been mapped to the S region of the mouse H-2 complex, and may also be linked to the human MHC. Given that duplication is a general theme of the evolution of the genes in class III region it is likely that other duplicated loci will be located in this region.

RCA Linkage Group

Control of complement activation is exerted at a number of levels to ensure that overactivation of the system does not occur. One of the levels of control is the presence in plasma and on cell membranes of a number of control 'cofactor proteins' [Holers et al., 1985]. These cofactors form noncovalent complexes with the activation fragments of C3 or C4 (C3b or C4b) such that C3b or C4b in the complex can be cleaved by the serine protease factor I. Once C3b or C4b have been proteolytically cleaved by factor I they are unable to interact with factor B or C2, respectively, to form functional C3 convertase. The proteins which interact with C3b or C4b either to cause dissociation of the C3 convertase and/or to mediate breakdown of C3b or C4b include factor H, C4BP, CR1, CR2, MCP and DAF (table I).

The linkage of the genes encoding factor H, C4BP and CR1 was established by the definition of genetic polymorphism in these proteins, which has been demonstrated by

electrophoretic techniques. Three forms of C4BP have been identified by IEF of neuramidase-treated EDTA plasma and family studies have indicated the Mendelian inheritance of the forms [Rodriguez de Cordoba et al., 1983; 1984; Rodriguez de Cordoba and Rubinstein, 1987]. The results were taken to suggest the presence of three alleles at a single autosomal locus. The same technique has also demonstrated five factor H variants [Rodriguez de Cordoba and Rubinstein, 1984, 1987], which were shown to be inherited in Mendelian fashion and to be alleles at a single locus. The CR1 locus is also polymorphic and determines at least four alleles [Dykman et al., 1983, 1985; Wong et al., 1983]. However, CR1 exhibits an unusual form of polymorphism in which allotypic variants that vary in molecular weight have been identified on human erythrocytes and leukocytes by SDS-PAGE. Analysis of the genetic variants of C4BP and CR1 in three pedigrees informative for the segregation indicated no recombination suggesting close linkage of the loci [Rodriguez de Cordoba et al., 1984]. Family segregation data of the genetic variants indicated that the factor H gene was linked to the C4BP and CR1 genes [Rodriguez de Cordoba et al., 1985]. Subsequently the gene encoding DAF was also mapped close to the factor H, CR1 and C4BP genes [Rey-Campos et al., 1987], and studies using cDNA probes have shown that the genes lie on human chromosome 1.

Molecular Cloning of Factor H, C4BP, CR1, CR2 and DAF

cDNA clones for factor H, C4BP, CR1, CR2 and DAF have been reported. In most cases they were isolated from the appropriate cDNA libraries using mixed synthetic oligonucleotides based on the available protein sequence. In one instance, a CR2 cDNA was isolated using the CR1 cDNA under less stringent hybridization conditions [Weis et al., 1986], illustrating the high degree of homology between these proteins. Complete derived amino acid sequences are now available for all these proteins. Although they differ markedly in size, their structures are highly homologous and contain contiguous units each of 60 amino acids long [Reid et al., 1986]. These repeating units, termed short consensus repeats (SCR), are based on a framework of 4 cysteine residues, together with highly conserved tryptophan, glycine and proline residues. Further, these repeat units are homologous to repeat units found in C2 and factor B, and also in a number of noncomplement proteins, illustrating their widespread distribution. A fuller description of the SCR and the proteins which contain it can be found in the article by Bentley [this volume].

The cDNA probes for C4BP, factor H, CR1 and DAF have been used in genomic Southern blot analysis to estimate the size of the genes, and have also been used to isolate the C4BP and CR1 genes from lambda genomic libraries. In man a single C4BP gene of ~30 kb has been reported [Lintin and Reid, 1986; Lintin et al., 1987], and also a single DAF gene of ~35 kb [Stafford et al., 1987]. Preliminary analysis of the factor H gene suggests that it is likely to be 80–110 kb in length [McAleer et al., 1987], and a related gene or pseudogene may also be present. The CR1 gene spans about 140 kb of DNA [Wong et al., 1987], and restriction enzyme mapping and Southern blot analysis of overlapping genomic clones have determined the basis of the size polymorphism observed between the CR1-A and CR1-B allelic variants. The CR1 gene is composed of homologous

genomic segments which encode long homologous repeats (LHR). The LHR is composed of 7 SCRs and large portions of the LHRs show up to 99% identity [Klickstein et al., 1987a, b]. The CR1-B (or S) allele contains 5 distinct segments of 15–25 kb, each encoding a LHR, while the CR1-A (or F) allele contains 4 segments [Wong et al., 1987]. It appears that one of the LHR genomic segments has duplicated in the CR1-A allele to generate the CR1-B allele. The other allelic size variants of CR1 are also probably due to the loss or gain of coding sequence of 1 LHR.

Molecular Mapping of the RCA Locus

Linkage analysis of allotypes of C4BP, factor H and CR1 indicated that they are closely clustered. Analysis of human-mouse somatic cell hybrids using cDNA probes for C4BP and factor H indicate that the genes are located on chromosome 1 [Reid et al., 1986]. In situ hybridization studies have placed the CR1 and CR2 genes on the long arm of chromosome 1 in the q32 region [Weiss et al., 1987].

Recently the technique of PFGE has been used to determine the organization of these genes in man [Carroll et al., 1987a; Rey-Campos et al., 1988]. Four of the 5 genes were found to be clustered in a common NotI/NruI fragment of ~950 kb. Further analysis using NotI/MluI double digestion indicated that the genes could be split into two clusters, one containing the CR1 and CR2 genes in a 450-kb fragment, while the second contains the DAF and C4BP genes in a 500-kb fragment. It was also suggested that the CR1 gene is at least 150 kb in length and lies within 50 kb of the CR2 gene. No information is yet available on the position of the factor H gene relative to the CR1, CR2, DAF and C4BP genes. Although on chromosome 1 the genetic data of Rodriguez de Cordoba et al. [1985, 1986] suggest that it could lie some 5–10 cM from this cluster of genes.

References

Agnello, V.: Complement deficiency states. Medicine *57:* 1–23 (1978).

Allen, F.H.: Linkage of HL-A and GBG. Vox Sang. *27:* 382–384 (1974).

Alper, C.A.: Inherited structural polymorphism in human C2: evidence for genetic linkage between C2 and Bf. J. exp. Med. *144:* 1111–1115 (1976).

Alper, C.A.: Complement and the MHC; in Dorf, The role of the major histocompatibility complex in immunobiology, pp. 173–220 (Garland, New York 1981).

Alper, C.A.; Boenisch, T.; Watson, L.: Genetic polymorphism in human glycine-rich beta-glycoprotein. J. exp. Med. *135:* 68–80 (1972).

Alper, C.A.; Raum, D.; Karp, S.; Awdeh, Z.L.; Yunis, E.T.: Serum complement 'supergenes' of the major histocompatibility complex in man (complotypes). Vox Sang. *45:* 62–67 (1983).

Anand, R.: Pulsed field gel electrophoresis: a technique for fractionating large DNA molecules. Trends Genet. *2:* 278–283 (1986).

Awdeh, D.Z.; Raum, D.D.; Glass, D.; Agnello, V.; Schur, P.; Johnston, R.B.; Gelfand, E.W.; Ballow, M.; Yunis, E.; Alper, C.A.: Complement human histocompatibility antigen haplotypes in C2 deficiency. J. clin. Invest. *67:* 581–583 (1981).

Belt, K.T.; Carroll, M.C.; Porter, R.R.: The structural basis of the multiple forms of human complement component C4. Cell *36:* 907–914 (1984).

Belt, K.T.; Yu, C.Y.; Carroll, M.C.; Porter, R.R.: Polymorphism of human complement C4. Immunogenetics *21:* 173–180 (1985).

Bentley, D.R.: Primary structure of human complement component C2: homology to two unrelated protein families. Biochem. J. *239:* 339–345 (1986).

Bentley, D.R.; Campbell, R.D.; Cross, S.J.: DNA polymorphism of the C2 locus,. Immunogenetics *22:* 377–390 (1985).

Bird, A.P.: CpG-rich islands and the function of DNA methylation. Nature, Lond. *321:* 209–213 (1986).

Brown, W.; Bird, A.P.: Long-range restriction site mapping of mammalian genomic DNA. Nature, Lond. *322:* 477–481 (1986).

Campbell, R.D.: Molecular genetics of C2 and factor B. Br. med. Bull. *43:* 37–49 (1987).

Campbell, R.D.; Bentley, D.R.; Morley, B.J.: The factor B and C2 genes. Phil. Trans. R. Soc. *306:* 367–378 (1984).

Campbell, R.D.; Carroll, M.C.; Porter, R.R.: The molecular genetics of components of complement. Adv. Immunol. *38:* 203–244 (1986).

Campbell, R.D.; Law, S.K.A.; Reid, K.B.M.: Sim, R.B.: Structure, organisation and regulation of the complement genes. A. Rev. Immunol. *6:* 161–196 (1988).

Carroll, M.C.; Alicot, E.A.; Katzman, P.; Klickstein, L.B.; Fearon, D.T.: Organisation of the genes encoding CR1, CR2, DAF and C4bp in the RCA locus on human chromosome 1. Complement *4:* 141 (1987a).

Carroll, M.C.; Alper, C.A.: Polymorphism and molecular genetics of human C4. Br. med. Bull. *43:* 50–65 (1987).

Carroll, M.C.; Belt, K.T.; Palsdottir, A.; Porter, R.R.: Structure and organistion of C4 genes. Phil. Trans. R. Soc. *306:* 379–388 (1984a).

Carroll, M.C.; Campbell, R.D.; Bentley, D.R.; Porter, R.R.: A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. Nature, Lond. *307:* 237–241 (1984b).

Carroll, M.C.; Campbell, R.D.; Porter, R.R.: The mapping of 21-hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. Proc. natn. Acad. Sci. USA *82:* 521–525 (1985a).

Carroll, M.C.; Katzman, P.; Alicot, E.M.; Koller, B.H.; Geraghty, D.E.; Orr, H.T.; Strominger, J.L.; Spies, T.: A linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. Proc. natn. Acad. Sci. USA *84:* 8535–8539 (1987b).

Carroll, M.C.; Palsdottir, A.; Belt, K.T.; Porter, R.R.: Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. Eur. molec. Biol. Org. J. *4:* 2547–2552 (1985b).

Chaplin, D.D.: Molecular organisation and in vitro expression of murine class III genes. Immunol. Rev. *87:* 61–80 (1985).

Chaplin, D.D.; Galbraith, L.J.; Seidman, J.G.; White, P.C.; Parker, K.L.: Nucleotide sequence analysis of murine 21-hydroxylase genes: mutations affecting gene expression, Proc. natn. Acad. Sci. USA *83:* 9601–9605 (1986).

Cole, F.S.; Whitehead, A.S.; Auerbach, H.S.; Lint, T.; Zeity, H.J.; Kilbridge, P.; Colten, H.R.: The molecular basis for genetic deficiency of the second component of human complement. New Engl. J. Med. *313:* 11–16 (1985).

Dunham, I.; Sargent, C.A.; Trowsdale, J.; Campbell, R.D.: Molecular map of the human major histocompatibility complex by pulsed field gel electrophoresis. Proc. natn. Acad. Sci. USA *84:* 7237–7241 (1987).

Dupont, B.; Oberfield, S.E.; Smithwick, E.M.; Lee, T.D.; Levine, L.S.: Close genetic linkage between HLA and congenital adrenal hyperplasia (21-hydroxylase deficiency). Lancet *ii:* 1309 (1977).

Dupont, B.; Pollack, M.S.; Levine, L.S.; O'Neill, G.J.; Hawkins, B.R.; New, M.I.: Congenital adrenal hyperplasia. Joint report from the Eighth International Histocompatibility Workshop; in Terasaki, Histocompatibility testing, pp. 693–706 (UCLA Press, Los Angeles 1980).

Dykman, T.R.; Cole, J.L.; Iida, K.; Atkinson, J.P.: Polymorphism of human erythrocyte C3b/C4b receptor. Proc. natn. Acad. Sci. USA *80:* 1698–1702 (1983).

Dykman, T.R.; Hatch, J.A.; Aqua, M.S.; Atkinson, J.P.: Polymorphism of the C3b/C4b receptor (CR1): characterisation of a fourth allele. J. Immun. *134:* 1787–1789 (1985).

Fu, S.M.; Kunkel, H.G.; Brusman, H.P.; Allen, F.H.; Fotino, M.: Evidence for linkage between HL-A histocompatibility genes and those involved in the synthesis of the second component of complement. J. exp. Med. *140:* 1108–1110 (1974).

Giles, C.M.; Uring-Lambert, B.; Boksch, W.; Braun, M.; Goetz, J.; Neumann, R.; Mauff, G.; Hauptmann, G.: The study of a French family with two duplicated C4A haplotypes. Hum. Genet. *77:* 359–365 (1987).

Hardy, D.A.; Bell, J.I.; Long, E.O.; Lindsten, T.; McDevitt, H.O.: Mapping of the class II region of the human major histocompatibility complex by pulsed-field gel electrophoresis. Nature, Lond. *323:* 453–455 (1986).

Hauptmann, G.; Goetz, J.; Uring-Lambert, B.; Grosshans, E.: C4 deficiency. Prog. Allergy, vol. 39, pp. 232–249 (Karger, Basel 1986).

Hauptmann, G.; Grosshans, E.; Heid, E.; Mayer, S.; Basset, A.: Systemic lupus erythematosus and hereditary complement deficiency: a case with total C4 defect. Annls. Derm. Syph *101:* 479–496 (1974).

Hauptmann, G.; Uring-Lambert, B.; Vegnaduzzi-Lamouche, N.; Mascart-Lemone, F.: RFLP studies of 3 complete C4-deficient patients. Complement *4:* 166 (1987).

Higashi, Y.; Yoshioka, H.; Yamane, M.; Gotoh, O.; Fujii-Kuriyama, Y.: Complete nucleotide sequence of two steorid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. Proc. natn. Acad. Sci. USA *83:* 2841–2845 (1986).

Hobart, M.: Phenotypic genetics of complement components. Phil. Trans. R. Soc. *306:* 325–331 (1984).

Hobart, M.J.; Lachmann, P.J.: Allotypes of complement components in man. Transplant. Rev. *32:* 26–42 (1976).

Holers, V.M.; Cole, J.L.; Lublin, D.M.; Seya, T.; Atkinson, J.P.: Human C3b- and C4b-regulatory proteins: a new multi-gene family. Immunol. Today *6:* 188–192 (1985).

Holers, V.M.; Chaplin, D.D.; Leykam, J.F.; Gruner, B.A.; Kumar, V.; Atkinson, J.P.: Human CR1 mRNA polymorphism correlates with the CR1 allelic molecular weight polymorphism. Proc. natn. Acad. Sci. USA *84:* 2459–2463 (1987).

Inoko, H.; Trowsdale, J.: Linkage of the TNF genes to the HLA-B locus. Nucl. Acids Res. *15:* 8957–8962 (1987).

Isenman, D.E.; Young, J.R.: The molecular basis for the difference in immune hemolysis activity of the Chido and Rodgers isotype of human complement compoment C4. J. Immun. *132:* 3019–3027 (1984).

Klickstein, L.B.; Rabson, L.D.; Wong, W.W.; Smith, J.A.; Fearon, D.T.: CR1 5′ cDNA sequences contain a fourth LHR and identify a new B cell-specific mRNA. Complement *4:* 180 (1987a).

Klickstein, L.B.; Wong, W.W.; Smith, J:A.; Weiss, J.H.; Wilson, J.G.; Fearon, D.T.: Human C3b/C4b receptor (CR1): Demonstration of long homologous repeating domains. J. exp. Med. *165:* 1095–1112 (1987b).

Kristensen, T.; D'Eustachio, P.; Ogata, R.T.; Chung, L.P.; Reid, K.B.M.; Tack, B.F.: The superfamily of C3b/C4b-binding proteins. Fed. Proc. *46:* 2463–2469 (1987).

Lachmann, P.J.: Inherited complement deficiencies. Phil. Trans. R. Soc. *306:* 419–430 (1984).

Lamm, L.U.; Olaisen, B.: Report of the committee on the genetic constitution of chromosome 6. Cytogenet. Cell. Genet. *40:* 128–155 (1985).

Law, S.K.A.; Dodds, A.W.; Porter, R.R.: A comparison of the properties of the two classes, C4A and C4B, of the human complement component C4. Eur. molec. Biol. Org. J. *30:* 1819–1823 (1984).

Law, S.K.A.; Reid, K.B.M.: Complement; in Focus series (IRL Press, Oxford, in press 1988).

Lawrance, S.K.; Smith, C.L.; Srivastava, R.; Cantor, C.R.; Weissman, S.M.: Megabase-scale mapping of the HLA gene complex by pulsed field gel electrophoresis. Science *235:* 1387–1390 (1987).

Lintin, S.J.; Lewin, A.; Reid, K.B.M.: Studies on the structure of the human C4b-binding protein gene. Complement *4:* 186 (1987).

Lintin, S.J.; Reid, J.B.M.: Studies on the structure of the human C4b-binding protein gene. FEBS Lett. *204:* 77–81 (1986).

McAleer, M.A.; Hauptmann, G.; Brai, M.; Misiano, G.; Sim, R.B.: Restriction fragment length studies for factor H. Complement *4:* 191 (1987).

Meo, T.; Atkinson, J.P.; Bernoco, M.; Bernoco, D.; Ceppellini, R.: Structural heterogeneity of C2 complement protein and its genetic variants in man: A new polymorphism of the HLA region. Proc. natn. Acad. Sci. USA *74:* 1672–1675 (1977).

Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.: Complete primary structure of the zymogen of human complement factor B. J. biol. Chem. *259:* 3407–3412 (1984).

Morley, B.J.; Campbell, R.D.: Internal homologies of the Ba fragment from human complement component factor B, a class III MHC antigen. Eur. molec. Biol. Org. J. *3:* 153–157 (1984).

Müller, U.; Jongeneel, C.V.; Nedospasov, S.A.; Lindahl, K.F.; Steinmetz, M.: Tumour necrosis factor and lymphotoxin genes map close to H-2D in the mouse major histocompatibility complex. Nature, Lond. *325:* 265–267 (1987a).

Müller, U.; Stephan, D.; Phillippsen, P.; Steinmetz, M.: Orientation and molecular map position of the complement genes in the mouse MHC. Eur. molec. Biol. Org. J. *6:* 369–373 (1987b).

Müller-Eberhard, H.J.: The membrane attack com-

plex of complement. Annu. Rev. Immunol. *4:* 503–528 (1986).

Müller-Eberhard, H.J.; Miescher, P.A.: Complement (Springer, Heidelberg 1985).

Olaisen, B.; Teisberg, P.; Jonassen, R.; Thorsby, E.; Gedde-Dahl, T.: Gene order and gene distances in the HLA region studied by the haplotype method. Ann. hum. Genet. *47:* 285–292 (1983).

O'Neill, G.J.; Yang, S.Y.; Dupont, B.: Two HLA-linked loci controlling the fourth component of human complement. Proc. natn Acad. Sci USA *75:* 5165–5169 (1978a).

O'Neill, G.J.; Yang, S.Y.; Tegoli, J.; Berger, R.; Dupont, B.: Chido and Rodgers blood groups are distinct antigenic components of human complement C4. Nature, Lond. *273:* 668–670 (1978b).

Palsdottir, A.; Arnason, A.; Fossdal, R.; Jensson, O.: Gene organisation of haplotypes expressing two different C4A allotypes. Hum. Genet. *76:* 220–224 (1987a).

Palsdottir, A.; Fossdal, R.; Arnason, A.; Edwards, J.H.; Jensson, O.: Heterogeneity of human C4 gene size. Immunogenetics *25:* 299–304 (1987b).

Partanen, J.; Koskimies, S.: Human MHC class III genes, Bf and C4. Polymorphism, complotypes and association with MHC class I genes in the Finnish population. Hum. Hered. *36:* 269–275 (1986).

Ragoussis, J.; Bliek, A. van der; Trowsdale, J.; Ziegler, A.: Mapping of HLA genes using pulsed-field gradient electrophoresis. FEBS Lett. *204:* 1–4 (1986).

Raum, D.; Awdeh, S.L.; Anderson, J.; Strong, L.; Granados, J.; Pevan, L.; Giblett, E.; Yunis, E.J.; Alper, C.A.: Human C4 haplotypes with duplicated C4A or C4B. Am. J. hum. Genet. *36:* 72–79 (1984).

Reid, K.B.M.: Application of molecular cloning studies on the complement system. Immunology *55:* 185–196 (1985).

Reid, K.B.M.: Activation and control of the complement systems. Essays Biochem. *22:* 27–68 (1986).

Reid, K.B.M.; Bentley, D.R.; Campbell, R.D.; Chung, L.P.; Sim, R.B.; Kristensen, T.; Tack, B.F.: Complement system proteins which interact with C3b or C4b. Immunol. Today *7:* 230–233 (1986).

Reid, K.B.M.; Porter, R.R.: The proteolytic activation systems of complement. A. Rev. Biochem. *50:* 433–464 (1981).

Rey-Campos, J.; Rubinstein, P.; Rodriguez de Cordoba, S.: Mapping of DAF to the RCA gene cluster in humans. Complement *4:* 217 (1987).

Rey-Campos, J.; Rubinstein, P.; Rodriguez de Cordoba, S.: A physical map of the human regulator of complement activation gene cluster linking the complement genes CR1, CR2, DAF and C4BP. J. exp. Med. *167:* 664–669 (1988).

Rittner, C.; Giles, C.M.; Roos, M.L.H.; Demant, P.; Mollenhauer, E.: Genetics of human C4 polymorphism: detection and segregation of rare and duplicated haplotypes. Immunogenetics *19:* 321–323 (1984).

Rittner, C.; Hauptmann, G.; Gross-Wilde, H.; Grosshans, E.; Tongio, M.M.; Mayer, S.: Linkage between HL-A (major histocompatibility complex) and genes controlling synthesis of the fourth component of complement. in Histocompatibility testing, pp. 945–953 (Munksgaard, Copenhagen 1975).

Robinson, M.A.; Carroll, M.C.; Johnson, A.H.; Harzman, R.J.; Belt, K.T.; Kindt, T.J.: Localisation of C4 genes within the HLA complex by molecular genotyping. Immunogenetics *21:* 143–152 (1985).

Rodrigues, N.R.; Dunham, I.; Yu, C.Y.; Carroll, M.C.; Porter, R.R.; Campbell, R.D.: Molecular characterisation of the HLA-linked steroid 21-hydroxylase β gene from an individual with congenital adrenal hyperplasia Eur. molec. Biol. Org. J. *6:* 1653–1661 (1987).

Rodriguez de Cordoba, S.; Dykman, T.R.; Ginsberg-Fellner, F.; Ercilla, G.; Aqua, M.; Atkinson, J.P.; Rubinstein, P.: Evidence for linkage between the loci coding for the binding protein for the fourth component of human complement (C4BP) and for the C3b/C4b receptor. Proc. natn. Acad. Sci. USA *81:* 7890–7892 (1984).

Rodriguez de Cordoba, S.; Ferreira, A.; Nussenzweig, V.; Rubinstein, P.: Genetic polymorphism of human C4-binding protein. J. Immun. *131:* 1565–1569 (1983).

Rodriguez de Cordoba, S.; Lublin, D.M.; Rubinstein, P.; Atkinson, J.P.: Human genes for three complement components that regulate the activation of C3 are tightly linked. J. exp. Med. *161:* 1189–1195 (1985).

Rodriguez de Cordoba, S.; Rubinstein, P.: Genetic polymorphism of human factor H (βIH). J. Immun. *132:* 1906–1908 (1984).

Rodriguez de Cordoba, S.; Rubinstein, P.: Quantitative variations of the C3b/C4b receptor (CR1) in human erythrocytes are controlled by genes within the regulator of complement activation (RCA) gene cluster. J. exp. Med. *164:* 1274–1283 (1986).

Rodriguez de Cordoba, S.; Rubinstein, P.: New alleles of C4-binding protein and factor H. Immunogenetics *25:* 267–268 (1987).

Schendel, D.J.; O'Neill, G.; Wank, R.: MHC-linked class III genes. Analyses of C4 gene frequencies, complotypes and associations with distinct HLA haplotypes in German Caucasians. Immunogenetics *20:* 23–31 (1984).

Schneider, P.M.; Carroll, M.C.; Alper, C.A.; Rittner, C.; Whitehead, A.S.; Yunis, E.J.; Colten, H.R.: Polymorphism of human complement C4 and steroid 21-hydroxylase gene. Restriction fragment length polymorphisms revealing structural deletions, homoduplications, and size variants. J. clin. Invest. *78:* 650–657 (1986).

Schumaker, V.N.; Zavodsky, P.; Poon, P.H.: Activation of the first component of complement. Annu. Rev. Immunol. *5:* 21–42 (1987).

Schwartz, D.C.; Cantor, C.R.: Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell *37:* 67–75 (1984).

Sim, R.B.; Malhotra, V.; Ripoche, J.; Day, A.J.; Micklem, K.J.; Sim, E.: Complement receptors and related complement control proteins. Biochem. Soc. Symp. *51:* 83–96 (1986).

Singer, M.F.; Skowronski, J.: Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. Trends biochem. Sci. *10:* 119–122 (1985).

Spies, T.; Morton, C.C.; Nedospasov, S.A:; Fiers, W.; Pious, D.; Strominger, J.L.: Genes for the tumor necrosis factor α and β are linked to the human major histocompatibility complex. Proc. natn. Acad. Sci. USA *83:* 8699–8702 (1986).

Stafford, H.A.; Tykocinski, M.L.; Holers, V.M.; Lublin, D.M.; Atkinson, J.P.; Medof, M.E.: Polymorphism of the DAF gene. Complement *4.* 227 (1987).

Strachan, T.: Molecular genetics and polymorphism of class I HLA antigens. Br. med. Bull. *43:* 1–14 (1987).

Teisberg, P.; Akesson, I.; Olaisen, B.; Gedde-Dahl, T., Jr.; Thorsby, E.: Genetic polymorphism of C4 in man and localization of a structural C4 locus to the HLA gene complex of chromosome 6. Nature, Lond. *264:* 253–254 (1976).

Teisberg, P.; Olaisen, B.; Jonassen, R.; Gedde-Dahl, T., Jr.; Thorsby, E.: The genetic polymorphism of the fourth component of human complement: Methodological aspects and a presentation of linkage and association data relevant to its localization in the HLA region. J. exp. Med. *146:* 1380–1389 (1977).

Trowsdale, J.: Genetics and polymorphism: class II antigens. Br. med. Bull. *43:* 15–36 (1987).

Uring-Lambert, B.; Goety, J.; Tongio, M.M.; Mayer, S.; Hauptmann, G.: C4 haplotypes with duplications at the C4A or C4B loci: frequency and associations with Bf, C2 and HLA-A, B, C, DR alleles. Tissue Antigens *24:* 70–72 (1984).

Weiss, J.J.; Fearon, D.T.; Klickstein, L.B.; Wong, W.W:; Richards, S.A.; deBruyn Kops, A.; Smith, J.A.; Weiss, J.H.: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor of human B lymphoytes. Proc. natn. Acad. Sci. USA *83:* 5639–5642 (1986).

Weiss, J.H.; Morton, C.C.; Bruns, G.A.P.; Weis, J.J.; Klickstein, L.B.; Wong, W.W.; Fearon, D.T.: A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor map to 1q32. J. Immun. *138:* 312–315 (1987).

White, P.C.; Grossberger, D.; Onufer, B.J.; Chaplin, D.D.; New M.I.; Dupont, B.; Strominger, J.L.: Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man. Proc. natn. Acad. Sci. USA *82:* 1089–1093 (1985).

White, P.C.; New, M.I.; Dupont, B.: HLA-linked congenital adrenal hyperplasia results from a defective gene encoding a cytochrome P-450. Proc. natn. Acad. Sci. USA *81:* 7505–7509 (1984).

White, P.C.; New, M.I.; Dupont, B.: Structure of human steroid 21-hydroxylase genes. Proc. natn. Acad. Sci. USA *83:* 5111–5115 (1986).

Whitehead, A.S.; Colten, H.R.; Chang, C.C.; Demars, R.: Localisation of the human MHC-linked complement genes between HLA-B and HLA-DR by using HLA mutant cell lines. J. Immun. *134:* 641–643 (1985).

Wong, W.; Cahill, J.; Kennedy, C.; Bonaccio, E.; Wilson, J.; Klickstein, L.; Ralson, L.; Fearon, D.: Analysis of genomic polymorphism in the human CR1 gene. Complement *4:* 240 (1987).

Wong, W.W.; Wilson, J.G.; Fearon, D.T.: Genetic regulation of a structural polymorphism of human C3b receptor. J. clin. Invest. *72:* 685–693 (1983).

Wu, L.C.; Morley, B.J.; Campbell, R.D.: Expression of the human complement protein factor B gene: evidence for the role of two distinct 5′ flanking elements. Cell *48:* 331–342 (1987).

Yu, C.Y.; Belt, K.T.; Giles, C.M.; Campbell, R.D.; Porter, R.R.: Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity. Eur. molec. Biol. Org. J. *5:* 2873–2881 (1986).

Yu, C.Y.; Campbell, R.D.: Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants. Application to the study of C4 null alleles. Immunogenetics *25:* 383–390 (1987).

Yunis, E.J.; Awdeh, Z.; Johnson, A.; Suciu-Foca, N.; Robinson, M.A.; Hartzman, R.; Raum, D.; Fleischnick, E.; Alper, C.A.: Complotype genetic loci segregate more frequently with HLA-DR than with HLA-B. Immunogenetics *21:* 25–31 (1985).

R.D. Campbell
MRC Immunochemistry Unit
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU (UK)