STUDIES ON THE HUMAN C4b-BINDING PROTEIN GENE



Susan J. Lintin

Brasenose College

Oxford

A thesis submitted for the

Degree of Doctor of Philosophy

October 1987

MRC Immunochemistry Unit

Department of Biochemistry

University of Oxford

TO MY MUM AND DAD

Acknowledgements

I would like to thank all the members of the MRC Immunochemistry Unit for their valuable help and discussion over the last three years and particularly Carole Sargeant, Alex Law and Tony Day for proof reading sections of this thesis and help with the computer. I would like to thank my supervisor Dr. Ken Reid for his support and advice and also Dr. D. Bentley for all his encouragement and faith in me. I want to remember Professor R. R. Porter and regret that he is not able to see the completion of this thesis.

I am very grateful to Frank Caddick and Ken Johnson for their help in the preparation of the figures for this thesis and especially to Carolyn Brooks for the typing.

Finally I would like to thank Phil for his dedication in proof reading and for all his love and understanding over the last 2 years but especially in the last three months.

Nomenclature and Abbreviations

The nomenclature of the complement components are as described by the World Health Organisation (1968, 1981). A bar indicates the activated form of the component. The nomenclature for restriction enzymes is standard. One and three letter codes are used for the amino acids. All abbreviations are defined when they first appear. Some frequently used abbreviations are listed.

A600	Absorbance at 600 nm
APC	Antigen presenting cell
cDNA	Complementary DNA
DEAE	Diethyl aminoethyl
dNTP	Deoxy nucleoside triphosphate
DNA	Deoxy ribonucleic acid
DTT	Dithiothreitol
EBV	Epstein Barr Virus
EDTA	Ethylene diamine tetra acetic acid
EGF	Epidermal Growth Factor
HLA	Human Leukocyte Antigen
HSV	Herpes Simplex Virus
LDL	Low density lipoprotein
LFA-1	Leukocyte function-associated antigen-1
LHR	Long homologous repeat

MAC	Membrane Attack Complex
MHC	Major Histocompatibility Complex
mRNA	Messenger RNA
OD	Optical density
PFGE	Pulsed Field Gel Electrophoresis
PMSF	Phenyl methyl sulphonyl fluoride
RCA	Regulators of Complement Activators
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
SCR	Short Consensus Repeat
SDS	Sodium dodecyl sulphate
tRNA	Transfer RNA

Abstract

Studies on the human C4b-binding protein gene

Susan J. Lintin — Brasenose College, Oxford

Submitted for the degree of D.Phil — Michaelmas Term 1987

A 5' cDNA clone coding for human C4b-binding protein (C4bp) was isolated, characterised and sequenced to complete the cDNA sequence coding for amino-acid residues 1-32 thus confirming the protein sequence data of Chung et al. (1985c). The sequence extended to allow derivation of the putative leader sequence which was 32 residues in length and showed a high degree of hydrophobicity typical of other documented leader sequences. This clone together with others isolated by Chung et al. (1985a) were used to prepare cDNA probes which were subsequently used to hybridise genomic clones. The data from the blots suggested that the C4bp gene was up to 30 kb in size and thus consisted of large proportions of intron sequence (mRNA ≅2.5 kb). A commonly occurring restriction fragment length polymorphism was detected using the BglII restriction enzyme. Analysis of the genomic clones has shown that, with one possible exception, each internal protein repeat is encoded by a discrete exon. Use of the cDNA probes to hybridise RNA transfer blots, on which the RNA originated from various tissues and cell-lines, suggested that the liver is the major source of C4bp mRNA. Cross hybridisation was detected between the human C4bp cDNA probes and genomic DNA isolated from various species on transfer blots suggesting that genomic sequence homologous to that coding for C4bp has been conserved during evolution. Results from pulse field gel electrophoresis blot analysis suggests that although C4bp and the alternative pathway regulatory protein, Factor H, are genetically linked the physical distance separating them may be as great as 3 Mbp.

C O N T E N T S

Page

Acknowledgements i

Nomencalture and Abbreviations ii

Abstract iv

Contents

CHAPTER I INTRODUCTION

1.1 THE COMPLEMENT SYSTEM 1

1.2 THE COMPLEMENT PROTEINS 3
1.2.1 The classical pathway of activation 3
1.2.2 The alternative pathway of activation 4
1.2.3 Membrane Attack Complex (MAC) 6

1.3 THE REGULATORY PROTEINS 6
1.3.1 C3 convertase regulatory proteins 8
1.3.1.1. Decay acceleration 8
1.3.1.2 Cofactor activity 9
1.3.2 Inhibition of activation and membrane attack 10
1.3.3 Regulatory protein deficiencies and disease 10

1.4. THE RECEPTOR PROTEINS 12
1.4.1 The C3 receptors 13
1.4.1.1 Membrane-bound regulatory proteins 13
1.4.1.2 Cell adhesion glycoproteins 14
1.4.1.3 CR4 14
1.4.2 Anaphylatoxin receptors 14
1.4.3 C1q receptor and Factor H receptor 15
1.4.4 Murine complement receptors 15

1.5 C4b-BINDING PROTEIN (C4bp) 16
1.5.1 The structure of C4b-binding protein 16
1.5.2 The functions of C4b-binding protein 21

1.6 THE SHORT CONSENSUS REPEAT (SCR) SUPERFAMILY 23

1.7 THE COMPLEMENT SYSTEM GENES 26
1.7.1 The C1q B-chain gene 27
1.7.2 The C4 genes 28
1.7.3 C2 and Factor B genes 28
1.7.4 Activation linkage group (C2, C4 and Factor B) 29
1.7.5 Regulation linkage group 30

CHAPTER II METHODS

2.1 MATERIALS 33

2.2 OVERNIGHT CULTURES 35

		Page
2.3	PREPARATION OF DNA	36
2.3.1	Small scale isolation of plasmid DNA	36
2.3.2	Rapid isolation of plasmid DNA	36
2.3.3	Large scale isolation of plasmid DNA (500-1000 ml) . .	37
2.3.4	Isolation of the replicative form of bacteriophage M13mp8	38
2.3.5	Small scale isolation of bacteriophage λ DNA	39
2.3.6	Isolation of chromosomal DNA	39
2.3.7.	Preparation of high molecular weight DNA for Pulsed Field Gel Electrophoresis	40
2.4	DIGESTION OF DNA BY RESTRICTION ENDONUCLEASE ENZYMES .	41
2.4.1	Digestion of cloned and genomic DNA	41
2.4.2	Digestion of DNA in agarose blocks for Pulsed Field Gel Electrophoresis (PFGE)	41
2.5	RADIOLABELLING OF DNA	42
2.5.1	'Fill-in' reaction	42
2.5.2	'Nick-translation'	42
2.5.3	Primer extension labelling	43
2.5.4	Kinase labelling	43
2.6	GEL ELECTROPHORESIS	43
2.6.1	Polyacrylamide gel electrophoresis	43
2.6.2	Autoradiography	44
2.6.3	Agarose gel electrophoresis	45
2.6.4	Agarose gels for Pulsed Field Gel Electrophoresis . .	45
2.7	RECOVERY OF DNA FROM GELS	46
2.7.1	Elution of DNA from acrylamide gels	46
2.7.2	Electroelution of DNA from agarose gels	46
2.7.3	Recovery of DNA from low gelling temperature agarose .	47
2.8	ISOLATION OF CLONED DNA	48
2.8.1	Screening of plasmid clones	48
2.8.2	Screening of cosmid clones	49
2.8.3	Screening of bacteriophage plaques	52
2.9	DNA TRANSFER BLOT ANALYSIS	54
2.10	NUCLEOTIDE SEQUENCING TECHNIQUES	56
2.10.1	Chemical Degradation	56
2.10.2	Dideoxy sequencing	57
2.10.3	Analysis of nucleotide sequence data	61
CHAPTER III	ISOLATION AND CHARACTERISATION OF 5' cDNA CLONES	
3.1	INTRODUCTION	62
3.2	ISOLATION AND INITIAL CHARACTERISATION OF POSITIVE CLONES	63
3.3.	CHARACTERISATION AND SEQUENCING OF pBP51	64

Page

3.4 NORTHERN BLOT ANALYSIS 65

3.5 DISCUSSION 66

CHAPTER IV ISOLATION OF GENOMIC CLONES FOR HUMAN C4bp

4.1 INTRODUCTION 69

4.2 COSMID LIBRARIES, PLATING AND SCREENING 70

4.3 λ-LIBRARY SCREENING AND ISOLATION OF GENOMIC CLONES . 73

4.4. INITIAL CHARACTERISATION OF NEW GENOMIC CLONES 75

4.5 DISCUSSION 77

CHAPTER V CHARACTERISATION OF C4bp GENOMIC CLONES

5.1 INTRODUCTION 81

5.2 CHARACTERISATION OF INITIAL C4bp LAMBDA-GENOMIC CLONES 82
5.2.1 Characterisation of BP/Pva 84
5.2.2 Characterisation of BP/Sta 86
5.2.3 Characterisation of BP/Stb 88
5.2.4 Characterisation of BP/Pvb 89
5.2.5 Confirmation of λBP10 Restriction Map 90

5.3 CHARACTERISATION OF λBP3'2 91

5.4 CHARACTERISATION OF λBP5' CLONES 93

5.5 DISCUSSION 94

CHAPTER VI SOUTHERN BLOT ANALYSIS OF THE HUMAN C4bp GENE AND ITS LINKAGE TO THE FACTOR H GENE

6.1 INTRODUCTION 97

6.2 GENOMIC BLOT ANALYSIS OF C4bp GENE 98

6.3 RESTRICTION FRAGMENT LENGTH POLYMORPHISM IN THE C4bp GENE 99

6.4 PHYSICAL LINKAGE OF THE C4bp AND Factor H GENES . . . 99

6.5 INTERSPECIES CROSS-HYBRIDISATION OF THE C4bp cDNA SEQUENCE 102

6.6. DISCUSSION 103

Page

CHAPTER VII GENERAL DISCUSSION

7.1 SUMMARY 106

7.2 EVOLUTION OF THE C4bp GENE AND OTHER SCR CONTAINING GENES 110

7.3 THE REGULATORS OF COMPLEMENT ACTIVATION LINKAGE GROUPS (RCA) 115

7.4 CONCLUSION AND FINAL COMMENTS 117

REFERENCES

PUBLICATIONS

C H A P T E R I

INTRODUCTION

1.1 The complement system

Complement was originally defined as a complex system of proteins which 'complemented' the action of antibody recognition by the killing of cells. Elucidation of a mechanism of sequential activation of the early components leading to the assembly of the late components (C5-9) into a large protein complex that mediated cytolysis supported this definition. This somewhat limited definition has required considerable expansion in recent years, to include the biological activities of small peptides released during activation and degradation of various components in the system (Roitt et al., 1985). These activities include -

(i) Mediation of the acute inflammatory response by the action of C3a, C4a, and C5a (the anaphylatoxins) on smooth muscle, vascular permeability, mast cell degranulation and in the case of C5a as a chemotactic agent for neutrophils (Hugli, 1981).

(ii) Opsonisation and phagocytosis of microorganisms by the interaction of C3b coated target particles with C3b receptors present on many polymorphonuclear granulocytes, monocytes and macrophages (Joiner et al., 1984).

(iii) Deposition of C3b on the surface of immune complexes regulates the size of these insoluble aggregates thus preventing their deposition in tissue (Czop and Nussenzweig, 1976). Interaction of deposited C3b with C3b receptors on erythrocytes mediates immune complex clearance from the blood as the cells traverse the hepatic circulation (Fearon, 1984).

(iv) Regulation of specific immune responses (Bottger et al., 1985; Melchers et al., 1985; Morgan, 1986; Schultz et al., 1987) including antibody isotype switching from IgM to IgG; T-cell proliferation and regulation of the growth of activated B-cells all mediated by the cleavage products of C3.

The proteins of the complement system form two inter-related enzyme cascades termed the classical and the alternative pathways which converge after the activation of C5 to form a lytic complex common to both pathways. Many different ways of initiating the two pathways have been demonstrated. In vivo the most common activator for the classical pathway is the formation of antigen-antibody aggregates in the blood (Porter and Reid, 1979) whilst the alternative pathway is activated by the presence of a variety of polysaccharide containing substances (Fearon, 1979a).

The functional relationships of the different components in the activation pathways are shown in Fig 1.1 and the individual steps and components are reviewed in great detail in Reid, 1986; Reid and Porter 1981; Pangburn and Müller-Eberhard, 1984.

Each step in the activation pathways involves the conversion of an inactive precursor into its active form by limited proteolysis. During this activation a small fragment is cleaved from the zymogen and a nascent binding site or catalytic site exposed in the major fragment. This newly activated fragment is able to bind to the activating surface allowing the assembly of highly specialised protease complexes, consisting of two or three proteins, on the activator surface. The C3 and C5 convertases of the two pathways have equivalent specificities but are formed from different component proteins ($C\bar{4}\bar{b}\bar{2}\bar{a}$; $C\bar{3}\bar{b}B\bar{b}$). These components are however structurally and functionally related.

Each enzyme complex can activate many subsequent lytic complexes as a result the first step in the cascade is amplified giving rise to a very efficient overall response to a single initiating event. Similar amplification mechanisms have also been described in the blood clotting and fibrinolysis cascades.

1.2 The complement proteins

As already stated many of the proteins involved in the convertase complex of the two activation pathways are structurally and functionally related. As a result they can be grouped together e.g. C1r and C1s; C2 and Factor B; C3, C4 and C5; C6 and C7. Comparison of amino acid sequences demonstrates homologies between members of the groups which suggests that they also have common evolutionary links (Gagnon, 1984; Müller-Eberhard, 1986; Sottrup-Jensen et al., 1985; Tosi et al., 1987).

1.2.1 The classical pathway of activation

The C1-complex is a unique feature of the classical pathway and can best be envisaged as an interlocking enzyme system, the structure and activation of which is described in great detail in Reid, 1983; Porter and Reid, 1979; Colomb et al., 1984. In non-activated circumstances the constituents of the complex, C1q, C1r and C1s are associated by weak interactions. Binding of C1q to Fc regions of IgG and IgM during activation triggers movement within the rigid collagen-like structure of C1q. This movement strengthens interactions within the complex and brings about autoactivation of the serine protease, C1r, this in turn enzymatically activates C1s. (The action of C1q on C1r is not enzymatic.) $C\bar{1}s$ is a serine protease which has C4 and C2 as substrates.

$C\bar{1}s$ cleaves C4 at one site giving rise to the small C4a fragment, and the large active fragment $C\bar{4}b$. Activation exposes a previously well

protected intrachain thiolester bond in $C\bar{4}\bar{b}$ which is capable of reaction with nucleophiles on the complement activator surface or with water. This gives rise to two populations of $C\bar{4}\bar{b}$, one in the fluid phase and one which is surface bound to the complement activator (Campbell et al., 1980).

Utilization of a thiolester bond to bring about binding by the formation of an ester or amide bond to hydroxyl or amino groups by an acyl transfer reaction has also been described for C3 and α_2 macroglobulin as well as for C4 (Reviewed by Tack, 1985). C4 is unique in this group of analogous proteins in that it has two isotypes C4A and C4B in which the thiolester bond shows differences in chemical reactivities. C4A shows preferential binding to amino groups and C4B binds to both amino and hydroxyl groups but showing a preference for hydroxyl groups (Law et al., 1984; Dodds et al., 1986).

The interaction of $C\bar{4}\bar{b}$ with an activator surface acts as a nucleus for the assembly of the classical pathway convertase complexes. The zymogen form of C2 binds to $C\bar{4}\bar{b}$ and is activated by $C\bar{1}s$ via a single proteolytic cleavage which leaves the catalytic region $C\bar{2}a$ associated with the activator surface (Gagnon, 1984).

$C\bar{2}a$ is itself a serine protease which cleaves one site in C3 thus bringing about exposure of a thiolester bond in $C\bar{3}b$ which then reacts in the same ways as described for $C\bar{4}b$. Binding of $C\bar{3}b$ in the vicinity of $C\bar{4}\bar{b}\bar{2}a$ changes the specificity of the protease complex from a C3 convertase into a C5 convertase. The catalytic site of C5 convertase is still situated in the $C\bar{2}a$ component of the complex (Vogt et al., 1978).

1.2.2 The alternative pathway of activation

The activation of complement via the alternative pathway occurs independently of antibody interaction but the exact mechanism by which

the initial C3 convertase complex is assembled remains difficult to prove conclusively. A 'tick-over' mechanism was proposed by Nicholl and Lachmann, 1973, and was confirmed by Pangburn et al., 1981, and reviewed by Pangburn and Müller-Eberhard, 1984, which involves $C3(H_2O)$, a C3b-like molecule which is formed when the thiolester bond in native C3 is hydrolysed by water.

$C3(H_2O)$ is capable of binding Factor B and in conditions which favour activation the serine protease Factor D cleaves Factor B (Fearon and Austen, 1980). The catalytic site in $B\bar{b}$ is thus exposed and together $C3(H_2O)B\bar{b}$ forms the alternative pathway C3 convertase. This mechanism explains the paradoxical situation of a C3 convertase containing apparently activated C3.

Factor B is the alternative pathway analogue of C2 and it cleaves C3 in the same way to reveal the thiolester bond. Binding of $C3\bar{b}$ in the vicinity of $C3(H_2O)B\bar{b}$ changes the proteolytic activity to that of a C5 convertase, with the catalytic site still residing in $B\bar{b}$.

Both C3 convertases $C3(H_2O)B\bar{b}$ and $C\bar{4}\bar{b}\bar{2}a$ are inherently unstable and spontaneously decay by the release of $B\bar{b}$ and $C\bar{2}\bar{a}$ into the fluid phase. The stability of the alternative pathway convertase is increased by the binding of properdin (Fearon and Austen, 1975).

Once activated the thiolester bond in $C\bar{3}\bar{b}$ (and $C\bar{4}\bar{b}$) can react with water, as well as nucleophiles on the activator surface, to give rise to a fluid phase population of $C\bar{3}\bar{b}$ which can interact with Factor I (Section 1.3.1.2) or with cellular C3b receptors (Section 1.4). Surface bound $C\bar{3}\bar{b}$ can change the specificity of the convertase activity (Vogt et al., 1978) but it can also interact with cellular C3b receptors to mediate particle adherence seen during phagocytosis (Section 1.4). $C\bar{3}\bar{b}$ bound to the surface of immune aggregates acts to regulate their size and clearance from the blood (Czop and Nussenzweig, 1976; Fearon, 1984).

The fragments C3a, C4a and C5a released during the activation pathways can interact with specific cellular receptors and mediate inflammatory (Hugli, 1981) and immune responses (Morgan, 1986), and are collectively known as the anaphylatoxins.

Activation of C5 by either pathway involves cleavage of a single peptide bond in C5 which exposes a metastable C6 binding site in $C\bar{5}\bar{b}$. This then acts as a nucleus for the assembly of the membrane attack complex (Lachmann and Thompson, 1970).

1.2.3 Membrane Attack Complex (MAC)

Both pathways converge and share a common pathway leading to cell lysis by the membrane attack complex (C6-C9) (Reviewed by Müller-Eberhard, 1986). The complex once assembled, disrupts the membrane which leads to colloid osmotic lysis of the activator cell. A protein analogous to C9 called perforin has been discovered in the granules of cytotoxic T cells leading to speculation that a cascade activation system similar to complement will be present in T cells (Podack et al., 1986; Henkart et al., 1986; Jenne et al., 1987).

1.3 The regulatory proteins

Activation of the complement system leads not only to cell lysis but also to the release of potent mediators of the inflammatory and immune responses and hence it is extremely important to prevent inappropriate triggering of this system. Continuous low level production of C3b-like molecules (Section 1.2.2) leads to the 'tick-over' phenomenon seen in the alternative pathway which has the potential to fully activate the alternative pathway if $C\bar{3}\bar{b}$ levels are not regulated (Pangburn et al., 1981).

During activation in response to infection there are inherent mechanisms involved in the assembly of C3-convertases which prevent host cell damage.

(i) The location of classical pathway C3 convertase assembly is restricted to the activator surface by the initial requirement for the C1q interaction with antibody-antigen complexes (Porter and Reid, 1979). Similarly location of alternative pathway C3 convertase activation is dependent upon the environment in which $C3(H_2O)B$ is deposited. If deposition occurs on an activator surface Factor B is cleaved by Factor D leading to activation of the rest of the pathway. If deposition of $C3(H_2O)B$ occurs on host cells the regulatory role of Factor H is supported and no subsequent activation takes place (Fearon & Austen, 1980; Fearon, 1978).

(ii) If activation is triggered successfully assembly of the complex proteases are localised in close proximity to the activator by the transient nature of the exposed thiolester bond in C3 and C4.

(iii) The C3 convertases are inherently unstable (Section 1.2.2) and therefore the activities are controlled partly by dissociation of $C\overline{2a}$ and $B\overline{b}$ from the complexes.

These mechanisms are not sufficient alone to maintain the required homeostasis in C3 levels and complete regulation is brought about by specific control proteins. The sites of activity of the control proteins in the activation cascades are shown in Fig 1.2. These proteins can be viewed as belonging to one of two groups -

(a) Those targeted at the amplification steps regulating C3b and C4b levels and convertase formation (Reviewed by Holers et al., 1985).

(b) Those targeted at initiation of the cascades and assembly of the membrane attack complex.

1.3.1 C3 convertase regulatory proteins

This group consists of two types of protein (i) plasma proteins - C4-binding protein (C4bp) and Factor H; (ii) membrane bound proteins - complement receptor type I (CR1), decay accelerating factor (DAF) and membrane cofactor protein (MCP - formerly GP45-70). Recently a surface bound form of Factor H was reported (Schulz et al., 1984; Malhotra and Sim, 1985). A soluble form of CR1 has also been reported (Yoon and Fearon, 1985). All five proteins share one or both of the following activities.

1.3.1.1. Decay acceleration

As already described in Section 1.2.2 the C3 convertases are inherently unstable and decay by the release of $C\overline{2}a$ and $B\overline{b}$. The rate at which this decay occurs is accelerated by the presence of control proteins which have the ability to bind $C\overline{4}\overline{b}$ or $C\overline{3}\overline{b}$, thus displacing the other component of the complex. This activity is known as decay acceleration and is demonstrated by Factor H (Weiler et al., 1976; Whaley and Ruddy, 1976), C4bp (Gigli et al., 1979), DAF (Nicholson-Weller et al., 1982) and CR1 (Iida and Nussenweig, 1981; Fearon, 1979b).

The binding specificity of these proteins, for C3b/C4b, is not restricted to the fragments involved in convertase complexes with binding to the ligands also occurring in solution. This activity prevents the initial formation of the complexes and has been implicated as the primary role of DAF (Medof et al., 1984) (this is now disputed (Fujita et al., 1987)). The two membrane bound members of this group, CR1 and DAF, show different efficiencies in accelerating the decay of convertases on the same (intrinsic) or different (extrinsic) membranes. CR1 shows a more efficient extrinsic functional activity (Pangburn et al.,

1983) and DAF a more efficient intrinsic functional activity (Medof et al., 1984).

Cells infected with HSV-I express on their surfaces a glycoprotein gc-1 which has been shown to mediate binding of C3b/C4b coated particles (Friedman et al., 1984). The glycoprotein is encoded within the virus genome and becomes expressed after infection of cells. By binding $C\bar{3}\bar{b}/C\bar{4}\bar{b}$ coated particles the protein accelerates decay of the C3 convertases and is thus thought to protect infected cells from complement mediated neutralisation. Recently it has been reported that HSV-II expresses a genetically related glycoprotein on infected cells which also has $C\bar{3}\bar{b}/C\bar{4}\bar{b}$ binding specificities (McNearney et al., 1987).

1.3.1.2 Cofactor activity

Factor I is a plasma serine protease which proteolytically cleaves C3b/C4b causing a rapid loss of their ability to function in the convertase complexes (Ruddy & Austen, 1971; Cooper, 1975) and as such acts as a regulatory enzyme in the complement cascades. However it has no activity unless the C4b/C3b is associated with a cofactor protein, a role which is assumed by C4bp (Fujita et al., 1978), Factor H (Whaley and Ruddy, 1976), CR1 (Iida and Nussenzweig, 1981; Fearon, 1979b) and MCP (Seya et al., 1986).

The degradation reactions of $C\bar{3}\bar{b}/C\bar{4}\bar{b}$ follow a well characterised series of proteolytic cleavages which produce fragments capable of interaction with various cell surface proteins or receptors (Reviewed by Sim and Walport, 1987; Sim et al., 1987). Each step requires one of the cofactors and this, together with the pattern of degradation, is shown in Fig 1.3. The details of the cleavage reactions are reviewed extensively in Sim et al., 1986.

By a combination of these two activities homeostasis of C3 activation in both pathways is maintained by this group of regulatory proteins. These proteins do show preferential activity for one or the other convertase depending on the binding specificity for $C3\bar{b}$ or $C4\bar{b}$ and this is summarised together with other functions in Table 1.1. From this table it can be seen that *in vitro* at least CR1 is the most functionally diverse of this group and of the other two groups, membrane bound or plasma proteins, one member has preferential activity in the alternative pathway and one in the classical pathway. Analogous proteins in the two pathways have been described in the activation reactions - C2 and Factor B and similar pairing of regulatory proteins can be proposed for Factor H and C4bp.

C4bp is described in greater detail in Section 1.5 and more specific details of the other individual regulatory proteins can be found in Reid, 1986; Holers *et al.*, 1985; Sim *et al.*, 1986.

1.3.2 Inhibition of activation and membrane attack

C1-inhibitor protein (C1-inh) controls the activation of the C1 complex by binding reversibly to $C1r_2$-$C1s_2$ to prevent spontaneous activation of the classical pathway. It also binds irreversibly to activated $C\bar{1}\bar{r}_2$-$C\bar{1}\bar{s}_2$ to dissociate the C1 complex and to mediate catabolism of $C\bar{1}r$ and $C\bar{1}s$ (Ziccardi, 1982).

Control of the membrane attack complex is reviewed by Podack and Tschopp, 1985. S-Protein is the specific control protein involved in the lytic pathway where it prevents insertion of the $C5\bar{b}$-7 complex into the membrane and also prevents polymerisation of C9.

1.3.3 Regulatory protein deficiencies and disease

The *in vivo* relevance of regulating complement activation and maintaining homeostasis of C3 levels is highlighted by a consideration of the disease states and their symptoms attributed to regulatory protein deficiences. Factor I and H deficiences result in low C3 levels, accompanied by urticaria, angioedema and recurrent infections all due in the first instance to excessive activation of C3 followed by depletion of C3 levels in plasma (Alper *et al.*, 1970; Thompson and Lachmann, 1977; Barrett and Boyle, 1984).

A deficiency in decay accelerating factor is linked to paroxysmal noctural haemoglobinurea, a haemolytic process associated with erythrocyte bound complement (Pangburn *et al.*, 1983; Nicholson-Weller *et al.*, 1983).

An inherited deficiency of C1-inhibitor gives rise to the clinical disease of hereditary angioedema (Frank *et al.*, 1976). This is an autosomal dominant condition in which levels of circulating C1-inhibitor are much reduced, which permits chronic activation of the classical pathway. The oedema, which is commonly seen around the face and gut, is attributed to the action of a C2 derived peptide. Smith and Kerr, 1985, have, however, suggested that C2 is not the source of the kinin-like activity generated in the course of the disease. C3 levels remain unaffected in this condition as the activation of C4 or C2 is limited to the fluid phase.

A deficiency of C4bp has not been reported so far but the importance of regulation in the classical pathway is reflected in one group of patients with autoantibodies to the classical pathway C3 convertase (Halbwacks *et al.*,, 1980; Daha & Van Es, 1980; Gigli *et al.*, 1981). These patients are usually suffering from acute post-infectious glomerulonephritis or systemic lupus erythematosus when the antibody is detected. The antibody known as C4-nephritic factor has the ability to (i) stabilize the $\overline{C42}a$ complex from its intrinsic decay (ii) prevent the extrinsic decay of the complex by C4bp and (iii) protect C4b from

the degradative actions of C4bp and Factor I. As a result of these effects the patients have increased levels of circulating C3 convertase and correspondingly lowered levels of C3 and C5.

1.4. The receptor proteins

In the introductory paragraphs the diverse roles of complement activation, other than cytolysis, were listed all of which could be described as co-ordinating host response to infection. This co-ordination relies on recognition of substrates by various populations of cells involved in both humoral and cellular responses. The recognition is mediated by the interaction of ligands, complement fragments, with their membrane bound receptors. The ligands fall into one of two categories -

(i) Those bound to microorganisms or immune complexes during complement activation which promote the removal of the substrate by phagocytosis mediated by complement receptors on macrophages, monocytes and polymorphs. Ligands included in this group include $C3\bar{b}/C4\bar{b}$ and their degradation products and C1q (Sim et al., 1986).

(ii) Those freely diffusible in the fluid phase which act as local hormones and are responsible for secretion, altered metabolism and chemotaxis mediated through complement receptors on mast cells, macrophages and polymorphs. These interactions indirectly augment the accumulation of large numbers of phagocytes into the environment of the pathogen and thus produce a synergistic enhancement of the effect of membrane bound complement. C3a, C5a, (C4a) (Hugli, 1981) and possibly fluid phase degradation products of $C3\bar{b}/C4\bar{b}$ fall into this category.

Tissue distribution and the type of ligand involved in the interactions are well documented for most of the complement receptors. Many of the receptors have also been isolated and recently sequenced, usually

using cDNA cloning techniques. Some cellular responses induced by ligand binding have been well documented including roles in phagocytosis and acute inflammatory response, whilst the roles of lymphocyte complement receptors in immune regulation remain difficult to elucidate.

A number of different complement components are involved in receptor interactions, however C3 and its degradation products play a pivotal role, thus the major group of receptors are for C3 fragments.

1.4.1 The C3 receptors

The degradation products of C3 produced by the action of Factor I and its cofactors are ligands which interact with these receptors (Section 1.3.1.2 and Fig 1.3). The tissue distribution, ligands and functions of these receptors are shown in Table 1.2 and reviewed extensively by Sim and Walport, 1987; Sim et al., 1987; Fearon and Wong, 1983; Ross and Medof, 1985. Despite the diversity of the proteins they can be divided into two families, membrane-bound regulatory proteins and cell adhesion glycoproteins, together with a recently defined receptor activity attributed to CR4.

1.4.1.1 Membrane-bound regulatory proteins

This group includes CR1, CR2, MCP, DAF, Factor H and HSV gc-1 and gc-2, the role of which in the control of complement activation are discussed in Section 1.3. All share functional activities in decay acceleration and/or cofactor activity and they also have structural similarities discussed in Section 1.6 (Reviewed by Holers et al., 1985; Sim et al., 1986). CR2 is included in this group on the basis of the structural similarity even though no control functions have yet been attributed to it (Weis et al., 1986 and 1987a).

Although surface bound Factor H and MCP do not demonstrate adherence of $C3\bar{b}/C4\bar{b}$ coated particles they do recognise the soluble monomeric fluid phase C3 fragments. This recognition may be important in mediation of the inflammatory response (Sim et al., 1986).

1.4.1.2 Cell adhesion glycoproteins

CR3 and p150,95 are structurally similar to one another and also to LFA-1. All three have a two subunit polypeptide structure, the β subunit of which is common to all (Sanchez Madrid et al., 1983) whilst the α subunits are antigenically distinct although preliminary protein sequence data suggests they may be homologous (Springer et al., 1985).

CR3 is known to mediate adherence of iC3b coated particles and is the receptor of greatest importance in phagocytic uptake by polymorphs and monocytes, the expression on tissue macrophages being low or absent (Wright et al., 1983). Expression of p150,95 shows a reciprocal tissue distribution with high levels on tissue macrophages (Hogg et al., 1986). Even though particle adherence by p150,95 has not been demonstrated, it may serve as an equally important receptor in phagocytosis.

1.4.1.3 CR4

CR4 has been defined in terms of a receptor activity on monocytes and polymorphs (Ross and Medof, 1985) and on platelets (Vik and Fearon, 1987) which is distinct from CR1 and CR3, however it has not been identified at the molecular level.

1.4.2 Anaphylatoxin receptors

The activation fragments C3a, C4a and C5a (Section 1.2) have effects on a variety of cells, polymorphs, mast cells and monocytes. At this time it is not clear if there are distinct receptors for the three

ligands. It can however be predicted from the reported differences in function of C3a and C5a (Hugli, 1981) that there will be at least two different types of receptor. A specific receptor for C5a and C5a Des Arg on polymorphs has been demonstrated (Chenoweth and Hugli, 1978). Attempts to show the presence of C3a receptors on any cell population have proved difficult due to high affinity non-specific binding (Hugli, 1986).

The situation of cell activation by the anaphylatoxins is further complicated by the release of secondary mediators in response to anaphylatoxins which together contribute to a complex cascade effect (Stimler-Gerard, 1986).

1.4.3 C1q receptor and Factor H receptor

The C1q receptor is found on T and B lymphocytes, platelets, polymorphs and monocytes (Tenner and Cooper, 1980 and 1981) and interacts with the collagenous part of C1q after inactivation of the C1-complex by C1-inhibitor (Section 1.3.2.). Interaction of C1q with its receptor enhances cell-mediated cytotoxicity (Ghebrehiwet and Müller-Eberhard, 1978) and stimulates metabolism in polymorphs. Little is yet known about the structure of this receptor.

Factor H is reported to interact with B-lymphocytes, polymorphs and monocytes (Lambris and Ross, 1982; Schmitt *et al.*, 1981; Erdei and Sim, 1987) and mediates mitogenesis in mouse lymphocytes and respiratory burst in monocytes (Reviewed by Fearon and Wong, 1983). Interaction of Factor H with its receptor appears to stimulate secretion of IL-1 by monocytes (Iferroudjene *et al.*, 1987). The isolated receptor appears similar in size to Factor H but is antigenically distinct.

1.4.4 Murine complement receptors

Using cross hybridisation with human CR1+CR2, specific cDNA probes genomic and cDNA clones coding for mouse CR1 and CR2 have been isolated (Aegerter-Shaw et al., 1987). A second subfamily of clones was isolated in the same experiment, which were designated X/Y and were found to be distinct from CR1/CR2 by tissue distribution and size of mRNA's. It is speculated that the X/Y genomic sequences may code for the p65 protein, a C3b-binding protein (Wong and Fearon, 1985), thought to be the mouse homologue of MCP or DAF.

1.5 C4b-binding protein (C4bp)

C4b-binding protein (C4bp) was discovered independently by two groups. Initially the necessity for a cofactor for Factor I (C3b/C4b inactivator) was demonstrated by Shirashi and Stroud, 1975, a role which was attributed to a protein called C3b/C4b inactivator cofactor (Nagaswawa and Stroud, 1977). This was followed by the description of a mouse serum protein with a high affinity for actived C4 (Ferreira et al., 1977) and the human equivalent (Scharfstein et al., 1978). This binding protein was subsequently shown to have Factor I cofactor activity (Fujita et al., 1978) and was subsequently demonstrated to be one of the C3b/C4b inactivator cofactors, based on common antigenic determinants, size and function (Fujita and Nussenzweig, 1979). The role of this protein in regulation of the classical pathway C3 convertase was demonstrated by Gigli et al., 1979.

1.5.1 The structure of C4b-binding protein

C4bp is the largest complement component with a M_r of approximately 550 kDa, accurate measurement being complicated by the appearance of C4bp as a doublet on non-reducing SDS polyacrylamide gels (Scharfstein et al., 1978; Nagasawa and Stroud, 1980). The reason for

the appearance as a doublet is still unknown. After reduction C4bp runs as a single band of ~ 70 kDa. N-terminal sequence data (Reid and Gagnon, 1982; Dahlbäck and Stenflo, 1981) gives only one sequence suggesting that C4bp is composed of approximately eight 70 kDa polypeptide chains. An estimate of mol. wgt. obtained by calculation of sedimentation equilibrium gave a value of 570 kDa (Dahlbäck, 1983).

Purification procedures of C4bp (Scharfstein et al., 1978; Nagasawa and Stroud, 1980; Villiers et al., 1981; Reid and Gagnon, 1982) led to co-purification with the serum glycoprotein Protein S. Attempts to obtain completely pure C4bp required more stringent conditions and resulted in lower yields (Nagasawa et al., 1982; Dahlbäck, 1983).

Protein S is a single chain Vitamin-K dependent protein (DiScipio et al., 1977; Dahlbäck and Stenflo, 1981) the biological role of which is involved in the regulation of the natural anticoagulant pathways (Reviewed by Esmon, 1987). Its co-purification with C4bp led to the suggestion that the two proteins form a stable complex in plasma (Dahlbäck and Stenflo, 1981; Dahlbäck, 1983). The functional significance of this complex remains unproven, although it is known that C4bp and Protein S are in reversible equilibrium (Mammen, 1984) and that only the free form of Protein S functions in the anticoagulant pathway (Dahlbäck, 1986). Thus during infection and other acute phase situations, where circulating levels of C4bp are raised, corresponding levels of free Protein S fall giving a decrease in anticoagulant activity. The presence of Protein S has no effect on the interaction of C4bp and free or bound C4b (Dahlbäck and Hildebrand, 1983).

Limited proteolysis of C4bp with trypsin (Reid and Gagnon, 1982) gave a molecule which behaved like native C4bp on non-reducing SDS polyacrylamide gels but gave two major fragments when run in reducing conditions. Similar treatment with chymotrypsin (Nagasawa et al., 1982)

gave two fragments on non-reducing gels, a 48 kDa and 160 kDa fragment. The former when sequenced was found to originate from the N-terminal region of the C4bp polypeptide chain (Nagasawa et al., 1982; Dahlbäck and Müller-Eberhard, 1984) and the latter when run in reducing conditions gave a single 25 kDa fragment and was thus described as the 'core' consisting of the C-terminal regions disulphide linked together. Fig 1.4a summarises the proteolytic fragments of the C4bp molecule.

Electron microscopy studies (Dahlbäck et al., 1983) show a spider-like structure with flexible 33 nm long and 3 nm wide tentacles joined at a small central body. The number of tentacles observed varied between 4 and 7, therefore C4bp was interpreted as having seven chains in each molecule. This structure was supported by the gel analysis of the intermediate cleavage products obtained during a time course digestion of C4bp by chymotrypsin (Dahlbäck and Müller-Eberhard, 1984). This experiment followed the progressive decrease of mol. wgt. from 550 kDa to 160 kDa of the major fragment in non reducing condition which reflected subsequent removal of 48 kDa fragments of each tentacle.

Electron microscopic investigation of the chymotrypsin cleavage products confirmed that the 160 kDa fragment was the disulphide linked core which was visualized as a ring-like structure attached to short remnants of tentacle observed in the intact molecule (Dahlbäck and Müller-Eberhard, 1984).

Further electron microscopy studies suggested that C4b binds to the tips of the tentacles and Protein S to the central core via a short spacer-like structure (Dahlbäck et al., 1983). This finding has been confirmed by the use of C4bp monoclonal antibodies and binding studies with chymotryptic fragments (Hillarp and Dahlbäck, 1987).

Sedimentation coefficients and frictional ratios indicated that C4bp monomers had a highly elongated structure (Scharfstein et al.,

1978; Villiers et al., 1981, Dahlbäck, 1983) which supports the electron microscopy results. The ultrastructure of C4bp in solution was investigated using synchroton X-ray scattering and hydrodynamic analysis (Perkins et al., 1986), results from which agreed with the structure proposed by electron microscropy. The structure in solution was more compact than that expected, the difference being attributed to distortions during sample preparations for electron microscopy.

A different model for C4bp has been proposed (Villiers et al., 1985) which suggested that the C4bp molecule is a decamer with a second set of interchain disulphide bonds associating two 70 kDa chains into one subunit giving an overall pentameric structure. This study again used electron microscopy to visualise C4bp molecules (using the rotary shadowing technique not used by Dahlbäck et al., 1983). This produced convincing results to agree with the pentameric structure proposed. However, the additional results from the analysis, by SDS-PAGE of partially reduced C4bp agreed with the seven subunit model more convincingly than the pentameric structure.

C4bp has also been isolated from mouse (Ferreira et al., 1977; Kaidoh et al., 1981) and from guinea pig (Burge et al., 1981). The plasma levels of C4bp vary between species: human - 276 µg/ml; mouse - 150 µg/ml; guinea pig - 60 µg/ml. Both mouse and guinea pig C4bp are composed of multiple subunits all approximately 70-80 kDa in size. In mouse C4bp these chains are held together by non-covalent forces, rather than disulphide linkages (Kaidoh et al., 1981).

Isoelectric focusing of neuraminidase treated plasma samples (Rodriguez de Cordoba et al., 1983; Rodriguez de Cordoba and Rubinstein, 1987) show the presence of three forms of human C4bp-A, B and C which are coded for by a single autosomal locus with three codominant alleles, $C4bp^{*}1$, $C4bp^{*}2$, $C4bp^{*}3$. The gene frequencies of the alleles are 0.986,

0.010 and 0.004 respectively. Phenotype C4bpA corresponds to the genotype homozygous C4bp*1, C4bpB corresponds to heterozygous C4bp*1, 2; and C4bpC corresponds to heterozygous C4bp*1, 3.

Using these and other structural polymorphisms of CR1 and Factor H in family studies has lead to the proposal of a linkage group for the regulatory protein genes known as the RCA linkage group (Rodriguez de Cordoba et al., 1985). This and the analogous linkage group of activation proteins (Factor B, C2 and C4) will be discussed in Section 1.7.

The primary structure of human C4bp has been determined by a combination of protein and cloned cDNA sequence analysis (Chung et al., 1985a,c). It was found to consist of 549 amino acids, the N-terminal 491 of which can be divided into eight internally homologous, contiguous repeating units each approximately 60 amino acids in length and known as short consensus repeats (SCR's). The C-terminal 58 amino acids show no homology to the SCR's and contain two cysteine residues which probably participate in the interchain disulphide bridging seen in the overall structure.

During sequencing three possible polymorphic sites were identified (Chung et al., 1985a) residue 44 was either glutamine or threonine and residue 309 was either tyrosine or histidine whilst the change in the codon for residue 177 was a 'silent' change which did not alter the amino acid. The former two changes give rise to charge differences which may account for the structural polymorphisms reported by Rodriguez de Cordoba et al., 1983, however the distribution of gene frequencies of the three alleles (Rodriguez de Cordoba & Rubinstein, 1987) make this unlikely. Structural predictions for the N-terminal 491 amino acids suggest that residues will form predominantly β-sheet whilst the C-terminal 58 amino acids, involved in the formation of the 'core', are

likely to form a stable α helical structure. The overall structure of C4bp is summarised in Fig 1.4b.

Each short consensus repeat has four conserved cysteine residues which are thought to be involved in intrachain disulphide bonds. Preliminary results of Janatova et al., 1987 suggest that the disulphide bonds are restricted within each homology unit with the first and third and the second and fourth half cysteines in each being linked.

Mouse C4bp has also been sequenced and is known to consist of 413 amino acids (Kristensen et al., 1987a) the N-terminus of which can be divided into six short consensus repeats homologous to those found in human C4bp. There is an overall homology of 61% between mouse and human C4bp sequences, the major difference between the two sequences being the absence of the two cysteines in the C-terminal region of the mouse sequence. As these two residues are implicated in the interchain disulphide bridging their absence in the mouse sequence probably explains why the chains associate by non-covalent attractions in the mouse C4bp molecule.

1.5.2 The functions of C4b-binding protein - (summarised in Fig 1.5)

The cofactor activity, described in Section 1.3.1.2, is specific for C4b degradation due to the binding specificity of C4bp which does not bind to or act as a cofactor in the degradation of native C4 (Fujita et al., 1978). C4bp does recognize C4i (von Zabern et al., 1982) which is a C4b-like molecule, formed by treatment of native C4 with chaotropes or repeated freezing and thawing. The thiolester bond in C4i is broken but the C4a remains uncleaved, thus the breakdown pattern of C4i was the same as found for C4b except that the C4a containing fragments were correspondingly larger (von Zabern and Gigli, 1982).

Cleavage of C4b by Factor I using C4bp as the cofactor occurs in a well defined order to produce the C4 fragments shown in Fig 1.3 (Nagasawa et al., 1980). It was reported that C4bp can act as a cofactor in the breakdown of C3b by Factor I (Nagasawa and Stroud, 1977) although this was only seen with very high concentrations of C4bp. Factor H is far more efficient, the reaction requires twenty times less Factor H to degrade soluble C3b and 300 times less to degrade surface bound C3b. This suggests that C4bp and Factor H have analogous roles in the classical and alternative pathways (Fujita and Nussenzweig, 1979).

Initial binding studies suggested that one C4bp molecule could bind four or five C4b molecules (Scharfstein et al., 1978). Subsequent studies using ultracentrifugation techniques (Ziccardi et al., 1984) suggested the presence of four strong binding sites for C4b which have association constants of 1.2×10^7 M and two weaker binding sites. It could be expected that there would be seven C4b binding sites in C4bp (because there are seven identical chains), however, Perkins et al., 1986, suggest that the 'tentacles' are sufficiently close together in solution to offer steric hindrance to the full binding of C4b. The strong association between C4bp and C4b is central to the second function of C4bp in its role of regulation of the classical pathway C3 convertase, decay acceleration (Gigli et al., 1979).

Structural and functional studies (Chung et al., 1985b) examined the binding and cofactor activities of C4bp before and after limited proteolysis with pepsin, trypsin and chymotrypsin. The fragments generated were then characterised by amino acid sequencing which together allowed tentative assignment of cofactor activity to between residues 177 - 332 and binding activity to between residues 332 - 395, numbering according to Chung et al., 1985a. Characterisation of C4bp using monoclonal antibodies has confirmed that the C4b binding site is

not close to the amino-terminal of the subunits (Hillarp and Dahlbäck, 1987) - this would suggest that the appearance of C4b on the ends of the 'tentacles' in the electron microscopy (Dahlbäck et al., 1983) is a result of distortion during sample preparation.

1.6 The Short Consensus Repeat (SCR) superfamily

On completion of the amino acid sequence for C4bp (Chung et al, 1985a) the data were analysed using the DIAGON computer programme (Staden, 1982b). This demonstrated the presence of eight internally homologous regions, each approximately 60 amino acids in length, which could be aligned by the positions of conserved cysteine, tryptophan and proline residues Fig 1.6. The existence of similar internal homologous repeats had been demonstrated in Factor B (Morley and Campbell, 1984) and β_2-glycoprotein I (Lozier et al., 1984). Further DIAGON analyses with C4bp revealed that these too conformed to the same consensus sequence and could be aligned with the repeats in C4bp. It was suggested by Chung et al., 1985a, that each protein may have evolved from a primordial segment by a series of tandem duplication events.

Subsequently other proteins have been found to show significant homology with C4bp, Factor B and β_2-glycoprotein I, all with the homologies residing in regions of repeating units, each approximately 60 amino acids long. This has led to the proposal of a superfamily of structurally related proteins, comprising of both complement and non complement proteins which has been extensively reviewed by Reid et al., 1986, Reid et al., 1987 and Kristensen et al., 1987b and are shown in Fig 1.7.

Murine complement control proteins and receptors have been shown to contain multiple copies of SCR's, C4bp (Kristensen et al., 1987a),

Factor H (Kristensen and Tack, 1986) and X/Y gene products and CR1 and CR2 genes (Aegerter-Shaw et al., 1987, Parks et al., 1987).

The 60 amino acid homology unit common to the proteins in the superfamily, conforms to a consensus sequence having a framework of highly conserved residues consisting of one tryptophan, one proline and four cysteine residues. At two other positions conserved glycines and bulky hydrophobic amino acids are frequently found to be conserved. The term short-consensus repeat (SCR) is used to described these 60 amino acid homology units (Klickstein et al., 1987).

The initial discovery of SCR's in complement proteins seemed to correlate with C3b/C4b binding activity, however subsequent demonstrations of SCR's in non-complement proteins would suggest that the repeat units represent structural building blocks. The sequences related to function are then superimposed on the basic consensus sequence (in this characteristic the SCR is similar in concept to the immunoglobulin fold). Until C3b/C4b binding activity has been completely discounted in the non-complement proteins the functional/structural role of the SCR remains speculative.

Complete disulphide mapping data available for β_2-glycoprotein I shows that within each homology unit the four conserved cysteine residues form two internal disulphide links bridging Cys 2 to Cys 4 and Cys 1 to Cys 3 (Lozier et al., 1984). Preliminary results for similar studies with Factor H (Day et al., 1987b) and C4bp (Janatova et al., 1987) indicate a similar result. This pattern of disulphide linkage in the SCR's is likely to result in the formation of a compact domain giving rise to an overall appearance of "beads on a string" for proteins consisting almost entirely of SCR's. Physical measurements made on some members of the superfamily suggest that they have unusually elongated structures (β_2-glycoprotein I, C4bp and Factor H - Sim et al., 1986).

This would suggest that conservation of the cysteines in the repeat units is maintaining the unusually elongated structure.

Complement receptor type I (CR1) is the most recently sequenced member of the superfamily (Klickstein et al., 1987). It is known to have four allotypic variants differing by increments of approximately 50 kDa, these increments being accounted for by differences in the polypeptide chains rather than changes in glycosylation (Dykman et al., 1983; Wong et al., 1983). The most common allotypes are F(Mr $\tilde{\sim}$ 250 kDa) and S(Mr $\tilde{\sim}$ 290 kDa) and sequencing of the cDNA coding for an F allotype (Klickstein et al., 1987) showed the presence of four tandem direct repeats each 1.35 kb in length and coding for a peptide approximately 50 kDa in size. This repeat is termed a Long Homologous Repeat (LHR) and is composed of seven SCR's. Thus CR1 is unique in having organised the basic structural unit of SCR's into the higher order structural unit of LHR.

Analysis of multiple restriction fragment length polymorphisms of the gene for CR1 suggest that duplication of genomic sequences has occurred in association with the high molecular weight allotype (Wong et al., 1986 and 1987). This has subsequently been attributed to the gain of the coding sequence for one LHR.

High estimates of the degree of homology (Klickstein et al., 1987) between the LHR's provide evidence for duplication events during the evolution of this protein. The third LHR was shown to be a hybrid of the second and fourth LHR, a situation which could not have arisen by a single recombinatorial event. It is thought that this hybrid LHR may have arisen by a "gene conversion event" involving double crossing over or branch migration. Similar estimates of homology between SCR's in different proteins is being attempted to give an idea of the evolutionary links between superfamily members. Initial results are very complex

and cannot be explained as readily as the situation with the LHR's in CR1 (Tony Day, personnal communication).

The mechanisms of duplication, deletion and gene conversion put forward for the evolution of CR1 are consistent with the finding in analysis of the Factor B and IL-2 receptor genes (Campbell and Porter, 1983; Leonard et al., 1985) that the SCR's are precisely encoded by discrete exons. In the Factor B gene the exons coding for the SCR's are spliced to exons encoding the catalytic regions of typical serine protease active sites. A similar situation is expected to prevail in C2, C1r and C1s where the protein domains involved have been identified but the gene structure is not available.

It is apparent that mechanisms of tandem duplication of individual exons, groups of exons and complete genes together with exon shuffling and "gene conversion" have played important roles in the evolution of this superfamily of functionally diverse proteins.

1.7 The complement system genes (Reviewed by Campbell et al., 1987)

As the complement components have been sequenced many similarities between groups of proteins have been noted suggesting that they form multigene families which have resulted from the duplication of whole genes. Such groups include C3, C4 and C5 together with the non-complement protein α_2 macroglobulin (Sottrup-Jensen et al., 1985) and the components of the MAC, C6 and C7 (Müller-Eberhard, 1986). The genes coding for these proteins are not thought to be closely linked physically. Other proteins which appear to have evolved by gene duplication events include C2 and Factor B (Campbell and Bentley, 1985); C1r and C1s (Tosi et al., 1987). These pairs of genes have been closely linked physically, C2 and Factor B by sequencing (Wu et al., 1987) and C1r and C1s by pulsed field gel electrophoresis (Tosi et al., 1987).

Other components involved in the control of complement activation, which are all composed almost exclusively of short consensus repeats, are thought to have evolved by a mechanism of contiguous duplication. This type of evolution has been demonstrated in fibrinogen, albumin and transferrin (Doolittle, 1984). In these examples as in the SCR family the duplication does not always involve faithful copying but usually leads to high levels of homology.

Comparison of amino acid sequence from complement components with other published sequences has demonstrated the existence of functional subdivisions readily identifiable in other non-related proteins i.e. many of the complement proteins can be described as 'mosaic proteins' as defined by Doolittle, 1985. Sequencing of the genes for many of these proteins has not been completed but for Factor B (Campbell and Porter, 1983) the junctions between the recruited units fall at the intron junctions. Preliminary results in the analysis of the genes for C9 and C1-Inh (Carter and Fothergill, 1987; Marazziti et al., 1987) do not show this type of conserved exon organisation.

1.7.1 The C1q B-chain gene

C1q has a highly repetitive collagen-like region which has a Gly-X-Y repeat structure. In collagen genes groups of these repeats are coded for by discrete exons which differ in size by multiples of 9 bp which represents one repeat unit (Wozney et al., 1981). The breaking of the triplet structure automatically breaks the tertiary structure of collagen.

The Gly-X-Y subunits in the C1q- B chain gene are present in two exons with the intron falling at the position in the sequence corresponding to the 'hinge' region in the molecule. This 'hinge', however, occurs without an interruption to the repetitive Gly-X-Y sequence (Reid,

1985). The collagen-like subdivision are spliced to an as yet unique globular domain responsible for the antibody binding function.

1.7.2 The C4 genes

In man there are two highly polymorphic loci coding for the two isotypic forms of C4, C4A and C4B. Derived sequence for the two proteins have shown that there are only fourteen differences between the two, twelve of which are located in the C4d region and have been shown to give four isotypic amino acid variations thought to be responsible for the different reactivities of C4A and B (Section 1.2.1) (Belt *et al.*, 1984). Many polymorphic differences in the two loci have been reported suggesting more than forty alleles exist. There are differences in the numbers of genes expressed and the number present due to deletions and duplications. There are two sizes of gene found at the C4B locus, a situation which is very unusual (Reviewed by Campbell *et al.*, 1987).

1.7.3 C2 and Factor B genes

The complete amino acid sequence for Factor B was obtained from protein and cDNA sequencing (Christie and Gagnon, 1983; Morley and Campbell, 1984) and was shown to be homologous to the other serine proteases except in the N-terminus region which was composed of three short consensus repeats (Section 1.6). The gene coding for this protein showed each SCR to be coded in a discrete exon and the exon structure for the catalytic and binding sites in the serine protease fragment was typical of other serine protease genes (Campbell and Porter, 1983). There was an extra exon in this region which is speculated to code for a sequence involved in the unique specificity and/or mechanism of activation seen in Factor B (Campbell and Bentley, 1985).

The protein sequence of C2 is highly homologous to Factor B and shows the same arrangement of functional subdivisions. The genes for the two proteins have been shown by sequencing of cloned genomic DNA to be very closely linked, the 3' end of the C2 gene being separated from the 5' end of the Factor B gene by 421 bp (Wu et al, 1987). Results of genomic mapping and sequencing show a similar exon structure (D. Bentley, unpublished) suggesting that the genes arose by a mechanism of gene duplication. The mRNAs of C2 and Factor B are very similar (2.9 kb and 2.6 kb respectively) but the genes are very different in size, C2 being approximately 18 kb and Factor B being 6 kb. No suitable explanation for the insertion or deletion of intron sequence during the duplication event has yet been proposed (Campbell & Bentley, 1985).

As already mentioned some of the complement components arising by gene duplication are very closely linked in pairs. There are however more complex linkage groups of complement genes, one comprising of activation products and the other of regulatory products.

1.7.4 Activation linkage group (C2, C4 and Factor B)

The structural genes for C2, Factor B and C4 were mapped between the Class I and II genes of the HLA on chromosome 6 using the linkage of protein variants in classical population genetics (Weitkamp and Lamm, 1982). This linkage group thus became known as the Class III region and with the availability of cDNA probes, specific for each member, overlapping cosmid clones coding for the genes were isolated (Carroll et al., 1984).

These cosmids demonstrated that C2 and Factor B are encoded at single loci whilst C4 is coded for by two linked loci and that all four Class III genes lie within a 98 kb stretch of genomic DNA. The order of the genes was shown to be C2, Factor B, C4A and C4B and the physical

distances between loci were estimated; C2 and Factor B were very closely linked and separated by approximately 30 kb from C4A which is ∿10 kb from C4B. It has also been shown that the two genes coding for the cytochrome p450, 21-hydroxylase lie approximately 3 kb from the 3' ends of the C4 genes (Carroll et al., 1985). Recently pulsed field gel electrophoresis has been used to determine the orientation of the Class III genes with respect to Class I and II (Dunham et al., 1987). These studies have shown that C2 is linked to Class I and C4 to Class II with C2 lying telomeric to C4.

1.7.5 Regulation linkage group

Structural variants of C4bp (Rodriguez de Cordoba et al., 1983) and CR1 (Dykman et al., 1983; Wong et al., 1983) were used to identify a linkage between the two genes and to suggest a linkage disequilibrium between the C4bp*2 and C3bR*3 alleles (Rodriguez de Cordoba et al., 1984). Further studies, including the structural variants of Factor H (Rodriguez de Cordoba & Rubinstein, 1984), showed that this gene was linked to the loci for C4bp and CR1, leading to the proposal of the regulator of complement activation (RCA) linkage group (Rodriguez de Cordoba et al., 1985). This study also showed that this linkage group segregated independently of the Class III genes and C3 locus.

The discovery of new polymorphic variants of C4bp and Factor H (Rodriguez de Cordoba and Rubinstein, 1987) provided further linkage evidence for the RCA gene cluster. It also allowed further characterisation suggesting that C4bp and CR1 genes are very closely linked but that 5-10 cM separates them from the Factor H gene. The high crossover frequency between C4bp/CR1 and Factor H loci could be due to a large physical distance or the presence of a recombination 'hot spot' between them.

In addition to the size polymorphism in CR1 (Section 1.6) coded for at the structural locus, C3bR, variations in the numbers of CR1 molecules expressed on the surface of erythrocytes has also been reported to be under genetic control (Miyakawa et al., 1981; Wilson et al., 1982). This expression locus was designated the C3bRQ locus (Wilson et al., 1982) and was correlated with a genomic restriction fragment length polymorphism (Wilson et al., 1986). Linkage studies involving the C3bRQ locus shows that it is closely linked to the CR1 structural locus and is thus a member of the RCA (Rodriguez de Cordoba and Rubinstein, 1986).

Use of cDNA probes specific for C4bp and Factor H in studies using mouse/human somatic cell hybrids (E. Soloman, unpublished results) and CR1 (Weis et al., 1987b) has mapped the RCA linkage group to the long arm of chromosome 1 in the q32 band. Similarly, in mouse the genes coding for C4bp (Barnum et al., 1987), Factor H (D'Eustachio et al., 1986), CR1/CR2 and Y gene (Aegerter-Shaw et al., 1987) are found on the long arm of chromosome 1.

C4bp is an important control protein involved in the regulation of the classical pathway of complement activation. The availability of the complete amino acid sequence (Chung et al, 1985,a) permitted tentative assignments of functional regions in this unusual molecule and also allowed formulation of molecular explanations for its overall structure. Definition of the internally homologous regions within C4bp and the subsequent demonstration of similar repeats in other related and non-related proteins has placed C4bp within a large superfamily of proteins containing the short consensus repeats. The other regulatory proteins involved in both activation pathways are members of the superfamily but have also been shown to be functionally related and genetically linked to C4bp.

Availability of cDNA probes specific for C4bp has allowed the isolation of genomic clones with the intention of defining the exon structure coding for the unusual primary structure of the C4bp molecule. The cDNA probes have also been used in an attempt to define the physical distance involved in the linkage between Factor H and C4bp.

	Binding Specificity	Cofactor Activity	Decay Acceleration	Membrane Bound	Soluble
Factor H	C3b	+	+	+	+
C4bp	C4b	+	+	+	+
CR1	C3b/C4b	+	+	+	+
MCP	C3b/C4b	+	-	+	-
DAF	C4b/C3b	-	+	+	+ (1)
HSVgc-1	C3b	+	-	+	-

(1) Medof *et al.* (1985)

Table 1.1 Binding specificities, functional activites and status of the control proteins in complement activation.

Receptor	Ligands	Positive cells/ Tissues	Probable Roles
CR1	C3b/C4b (iC3b)	Erythrocytes Lymphocytes Monocytes Granulocytes Mast cells Kidney podocytes Follicular dendritic cells	Mediates binding to immune complexes Triggering of B cells Stimulates IL-1 production (1) Phagocytosis Regulation of C3b/C4 breakdown Alterations in APC (3) Retention of Ag during generation of memory cells (4)
CR2	C3b, iC3b C3dg, C3d E.B.V.	B lymphocytes Follicular dendritic cells	Triggering of activation and differentiation As for CR1
CR3	iC3b	Neutrophils, NK cells Monocytes Follicular dendritic cells	Phagocytosis As for CR1
CR4	C3d, C3dg	Monocytes Polymorphs Platelets (5)	Role as yet unknown
Surface bound Factor H	C3b	Tonsil B cells and various cell lines	Regulation of C3b breakdown ??
M.C.P.	C3b, C4b	Lymphocytes] (6) Monocytes, N.K. cells] Polymorphs]	Regulation of C3b/C4b breakdown.
P150,95	iC3b	Macrophages Monocytes, NK cells	Phagocytosis Ag presentation??
DAF	$C\bar{4}b\bar{2}a$ $C\bar{3}b\bar{B}b$	Erythrocytes Leukocytes	Acceleration of decay of convertases
Hsv gc-1	$C\bar{4}b\bar{2}a$ $C\bar{3}b\bar{B}b$	Virally infected cells	Acceleration of decay of convertases with a possible role in protection from complement inactivation

1. Bacle et al., 1987
2. O'Neil and Torres, 1987
3. Arvieux et al., 1987
4. Reynes et al, 1985
5. Vik and Fearon, 1987
6. Seya et al., 1987

Table 1.2: Principal ligands, roles and cellular distribution of C3 receptors [Positive cell lines which express the receptors are not given, see Sim et al., 1987. References if not given see Sim et al., 1986, 1987].

Fig 1.1 - The major activation steps in the classical and alternative pathways of the human complement system are shown. Diagram from Reid, 1986.

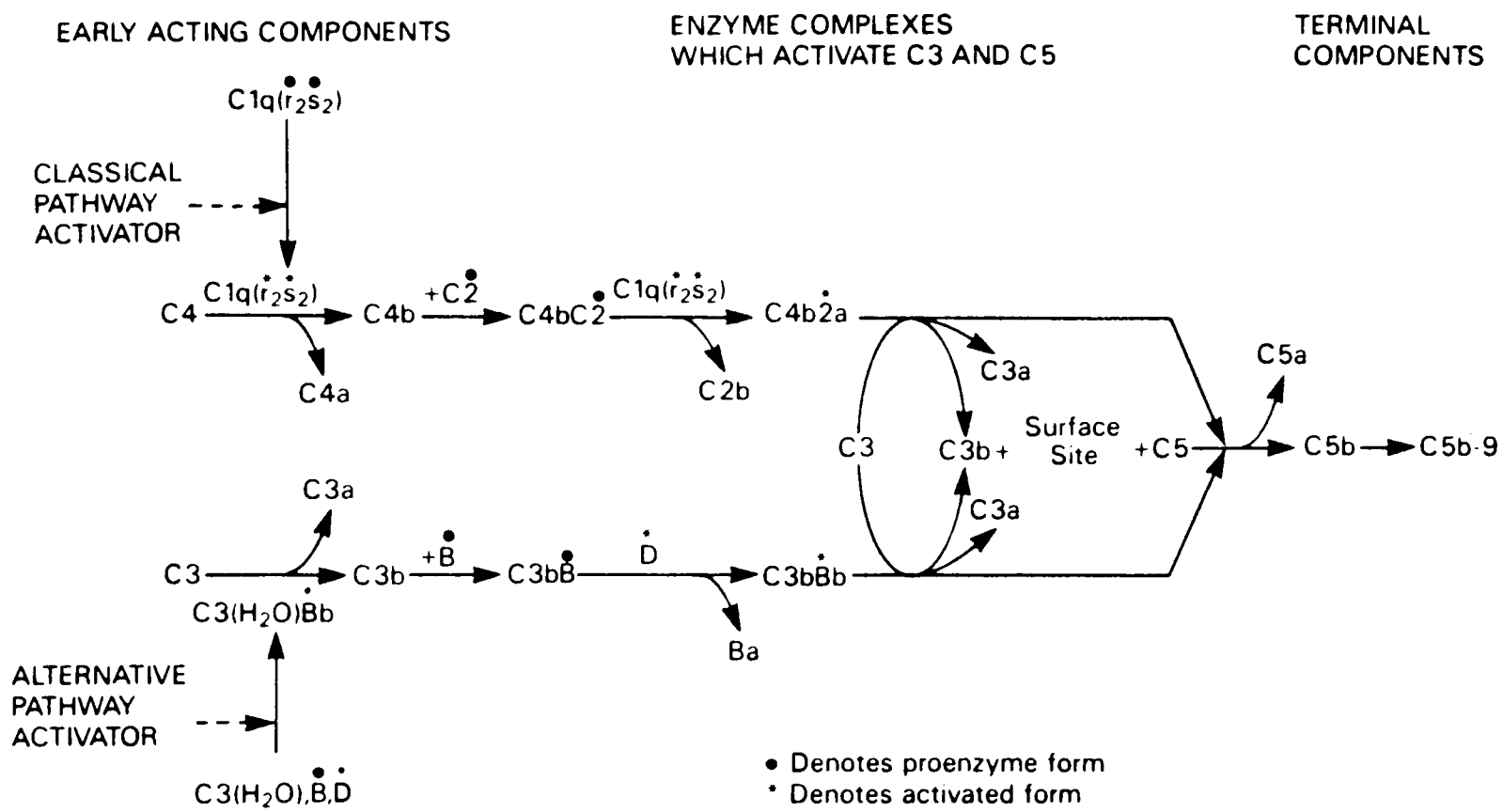

EARLY ACTING COMPONENTS
ENZYME COMPLEXES
WHICH ACTIVATE C3 AND C5
TERMINAL
COMPONENTS
C1q(r2s2)
CLASSICAL
PATHWAY
ACTIVATOR
C4
C1q(r2s2)
C4b
+C2
C4bC2
C1q(r2s2)
C4b2a
C4a
C2b
C3a
C5a
Surface
Site
C3
C3b +
+C5
C5b
C5b-9
C3a
C3
+B
C3b
C3bB
D
C3bBb
C3a
C3(H2O)Bb
Ba
ALTERNATIVE
PATHWAY
ACTIVATOR
C3(H2O),B,D
• Denotes proenzyme form
• Denotes activated form

Fig 1.2 - A simplified scheme of the human complement system is shown. The control proteins are boxed and the arrows with dashed lines indicate the components regulated by the control proteins. C3a, C4a and C5a, marked by +, are anaphylatoxins controlled by the anaphylatoxin inactivator. Diagram from L.P. Chung, D.Phil. thesis, Oxford.

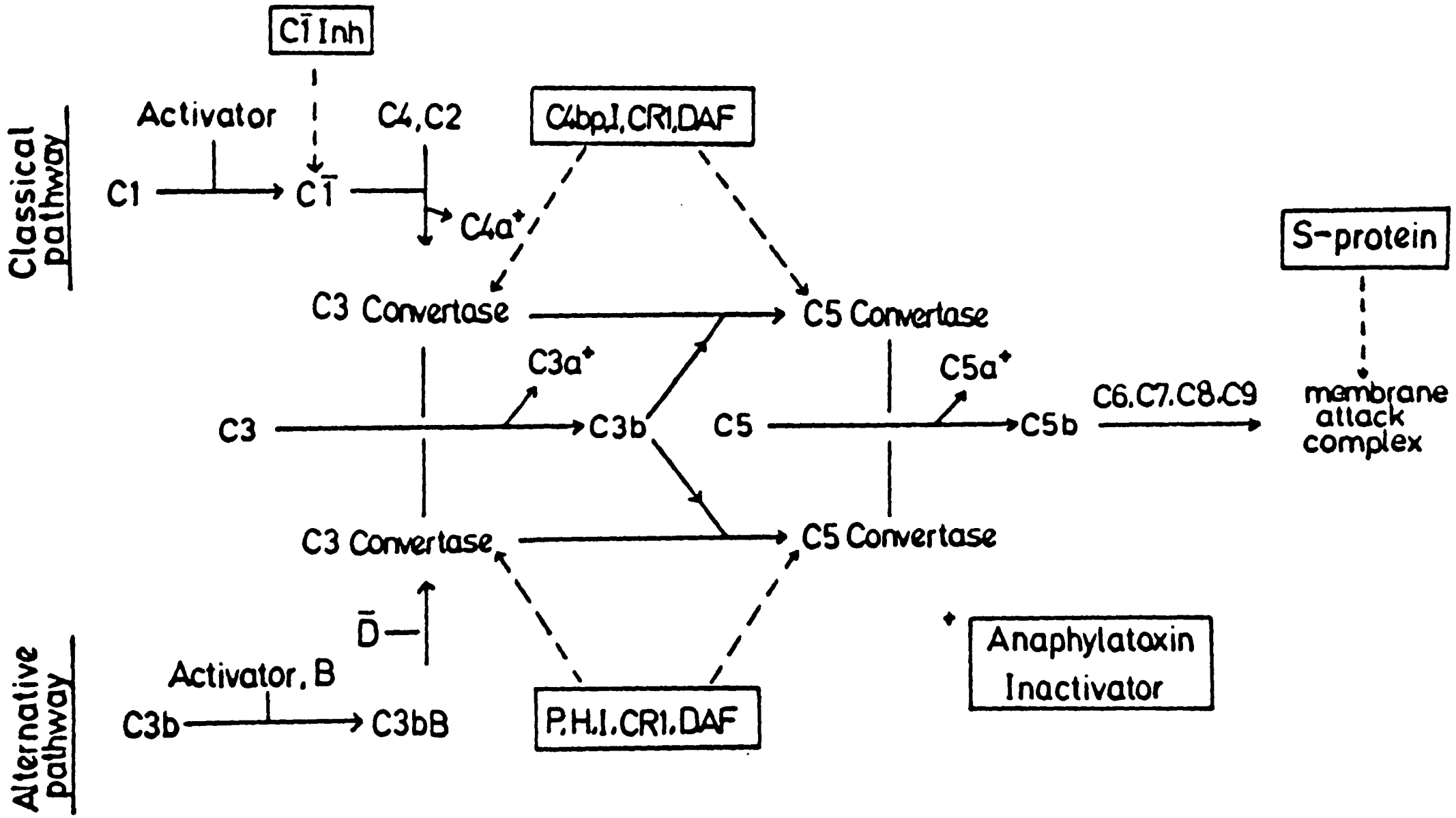
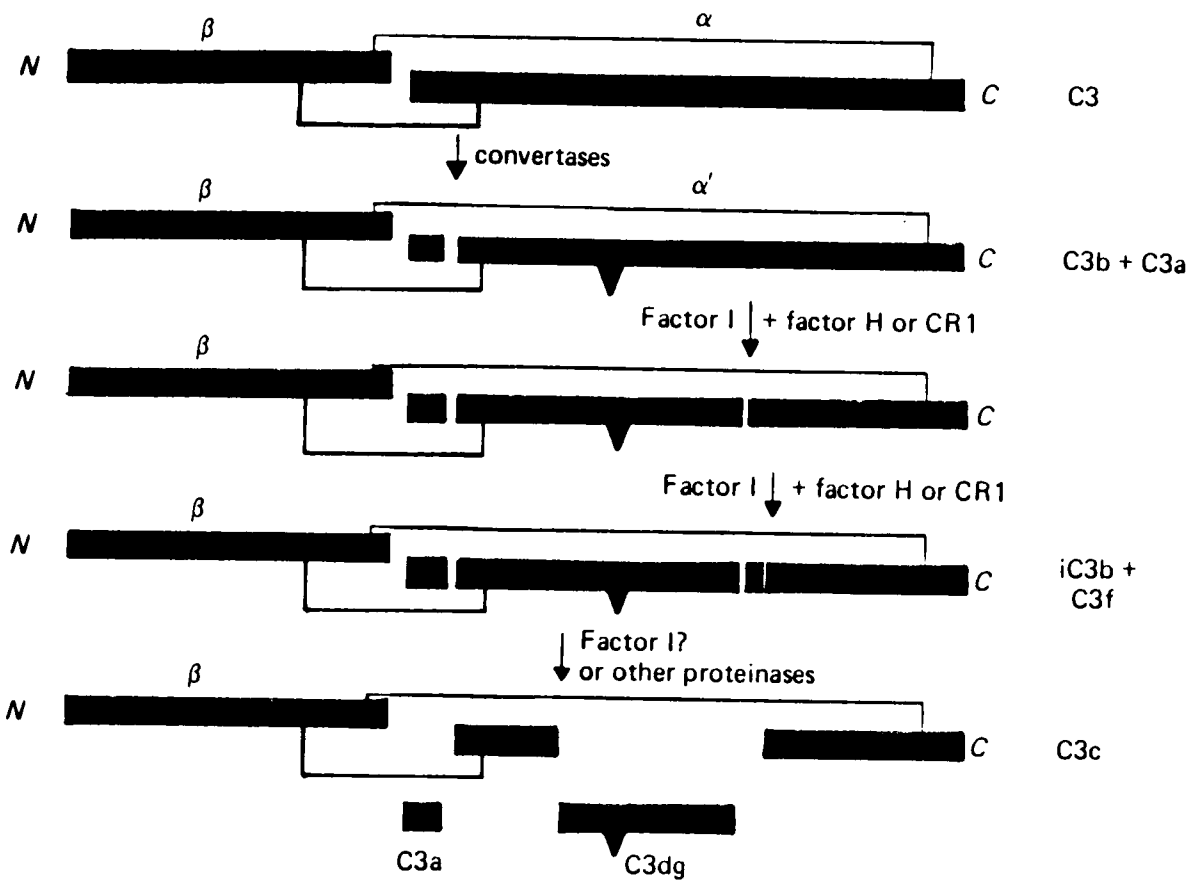

C1 Inh
Classical pathway
Activator
C4, C2
C4bp, I, CR1, DAF
C1
C1
C4a*
S-protein
C3 Convertase
C5 Convertase
C3a*
C5a*
C3
C3b
C5
C5b
C6.C7.C8.C9
membrane attack complex
C3 Convertase
C5 Convertase
D
Activator, B
C3b
C3bB
P.H.I.CR1.DAF
Alternative pathway
Anaphylatoxin Inactivator

Fig 1.3 - Summarises the cleavage reactions involved in the activation and subsequent degradation of C4 and C3. The thiolester site in C3b/C4b is marked with a spike. The diagram shows the Factor I cofactor involved in each step and the cleavage fragments produced by each reaction. Diagram from Sim et al., 1986.


β
α
N
C
C3
convertases
β
α′
N
C
C3b + C3a
Factor I + factor H or CR1
β
N
C
Factor I + factor H or CR1
β
N
C
iC3b +
C3f
Factor I?
or other proteinases
β
N
C
C3c
C3a
C3dg

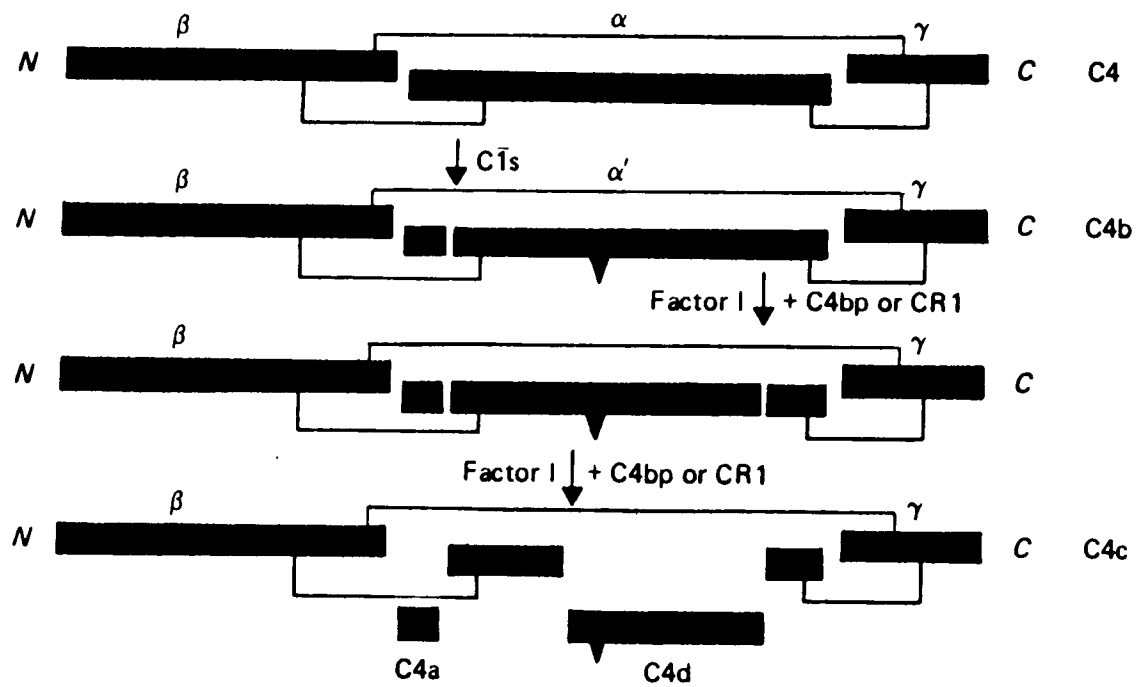

β
α
γ
N
C
C4
C1̄s
β
α′
γ
N
C
C4b
Factor I + C4bp or CR1
β
γ
N
C
Factor I + C4bp or CR1
β
γ
N
C
C4c
C4a
C4d

Fig 1.4a - Summarises the major sites of limited proteolysis by trypsin and chymotrypsin on C4bp together with the proposed positions of disulphide bonds within the molecule.

Diagram from Chung *et al.*, 1985b.

Fig 1.4b - Schematic model for the C4bp molecule showing it complexed to C4b. The dimensions shown are based on the measurements made by DahlBäck *et al.*, 1983 from electron micrographs. The possible arrangement of eight internal homology regions, found in the N-terminal 491 residues of each of the seven chains which form the 'tentacles' of the molecule are shown by O. Diagram from Reid *et al.*, 1986.

Fig 1.4a

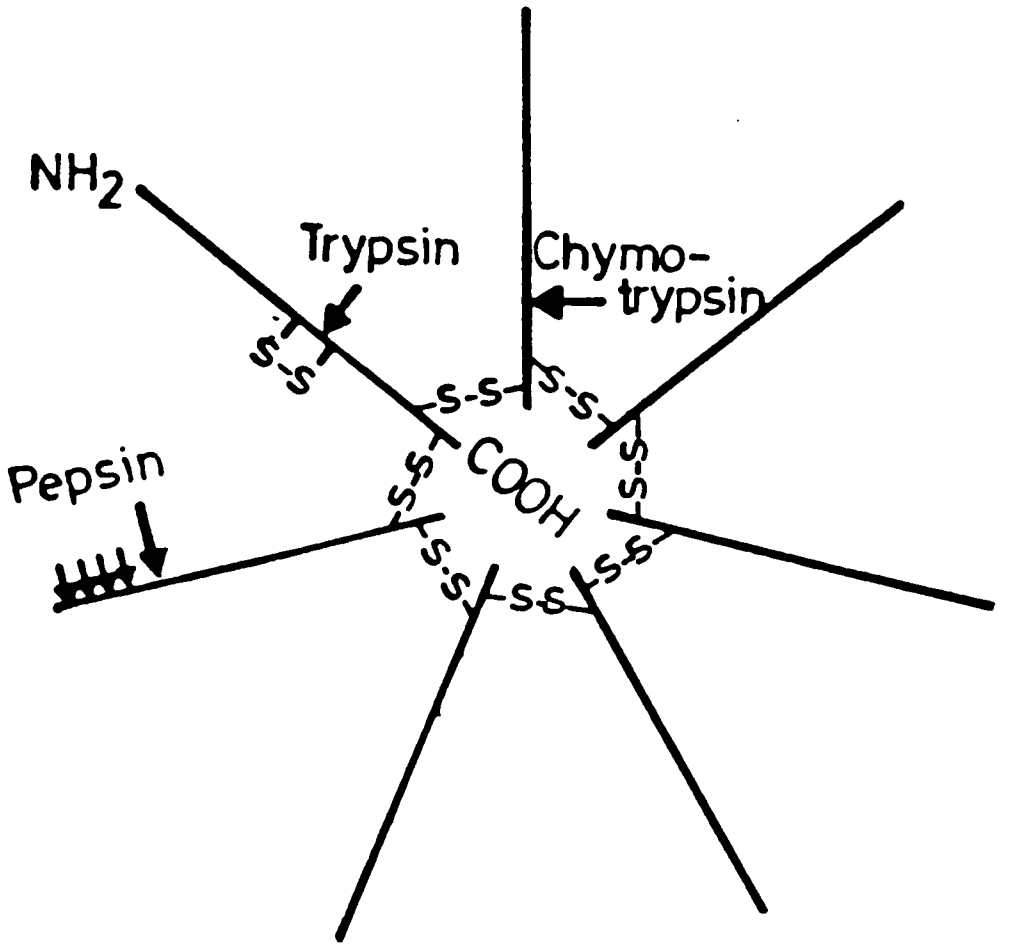
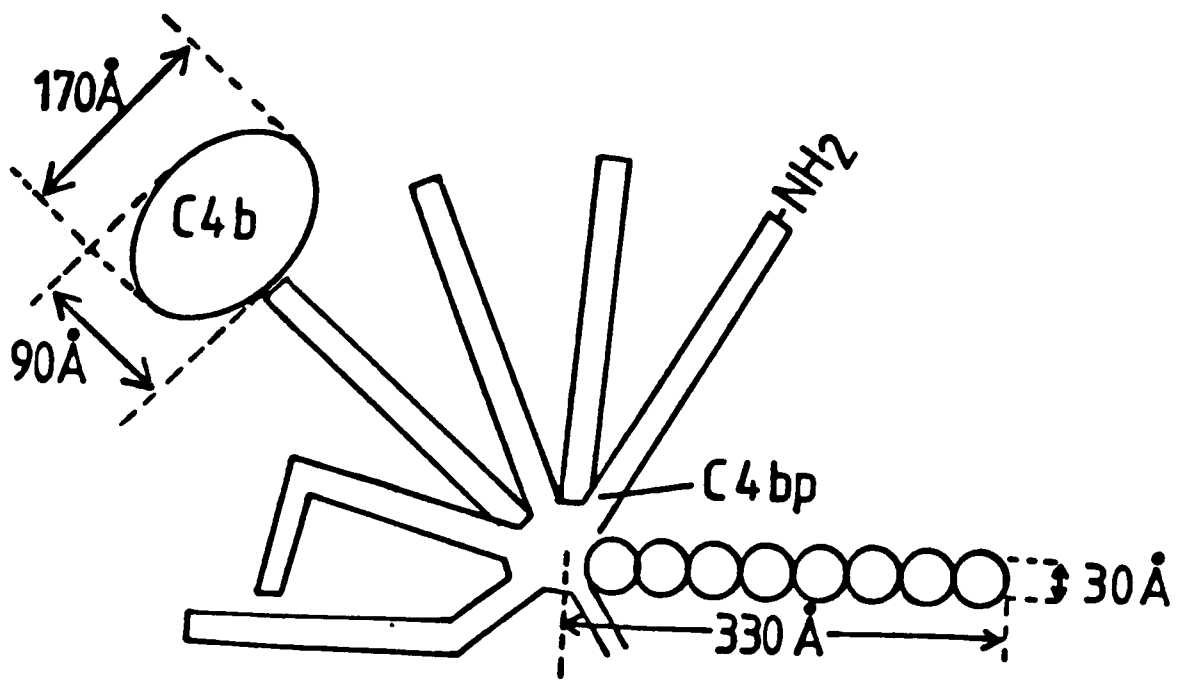


Fig 1.4b



○ = one repeating unit taken to be 42·5Å long

Fig 1.5 - Summarises the two functions of C4bp in its role controlling the classical pathway C3 convertase. Diagram from L.P. Chung, D.Phil. thesis, Oxford University.

Function of C4b binding protein (C4bp)

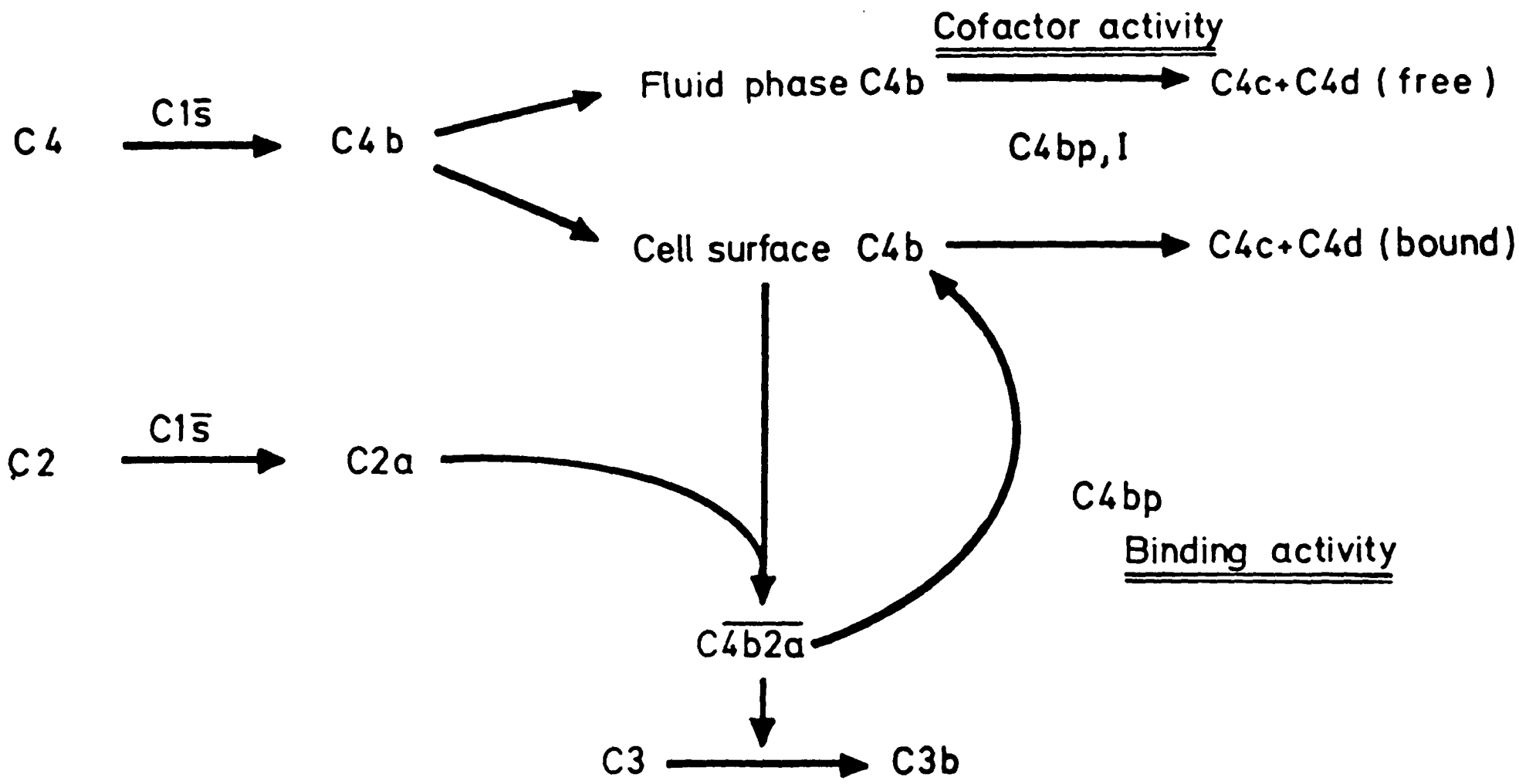


Fig 1.6 - Internal amino acid sequence homology in C4bp. Identical and functionally conserved amino acids that are present in four or more of the eight homologous regions are boxed. The invariant amino acids (four cysteines, proline and tryptophan) used to align the internally homologous regions and form the basis for the consensus sequence of the SCR family are indicated with *. Gaps, denoted by -, are introduced to maximise the homology and to allow space for the regions with short inserted sequences. Diagram from Chung et al., 1985a.

```
                 *       *                                                
```

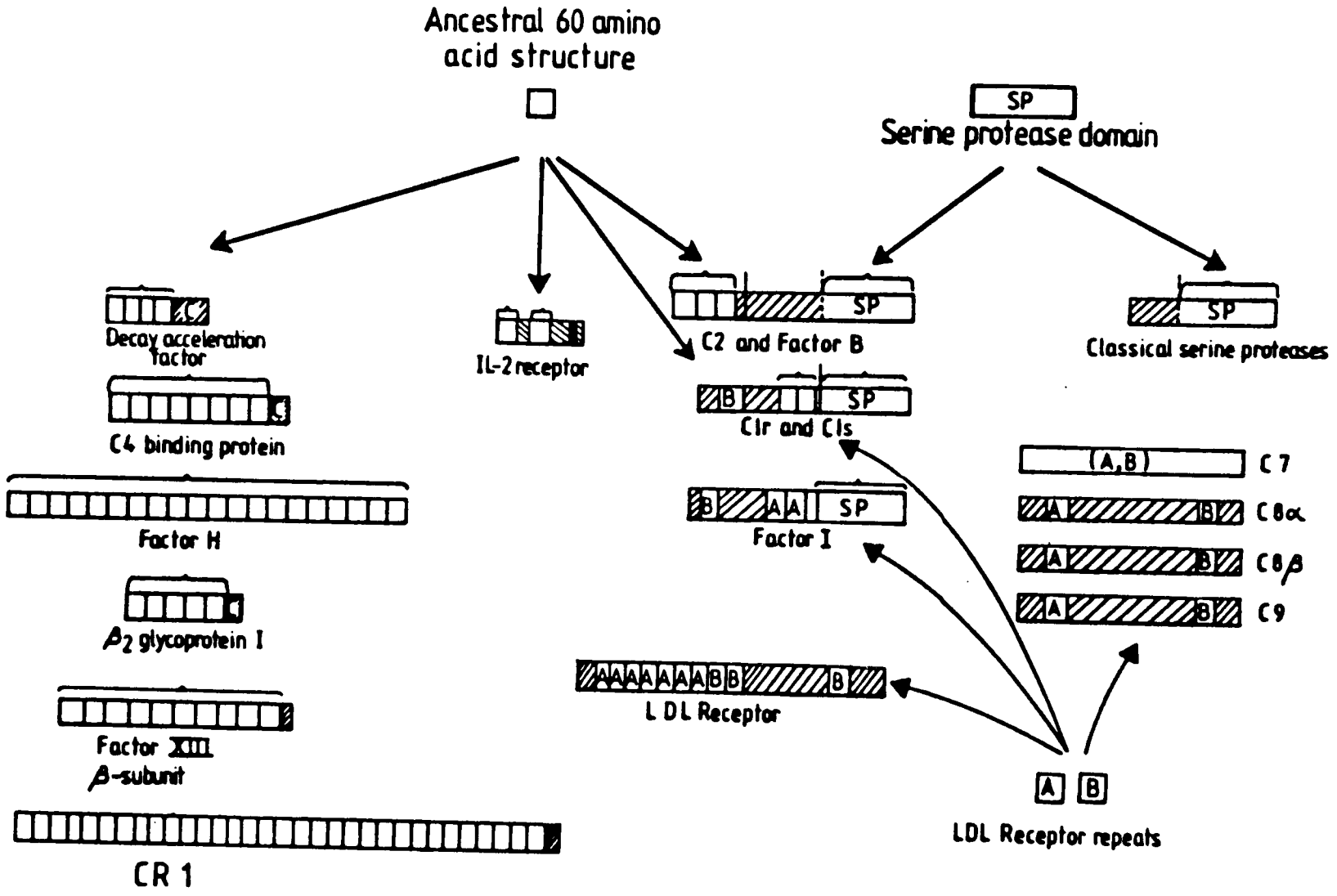
Fig 1.7 - Occurrence of the short consensus repeat (SCR) within the amino acid sequences of complement and non-complement proteins.

The SCR (depicted by □) described in the primary structure of C4bp (Section 1.5.1) occurs in Factor H, DAF, CR1, Factor B, C2, C1r and C1s, together with CR2 not shown in this diagram and also in non-complement proteins IL-2 receptor, β_2 glycoprotein I, Factor XIII β-subunit.

Serine protease units as seen in trypsin occur in C2 and Factor B, C1r and C1s, Factor I and Factor D. LDL receptor repeat structures type A occur in Factor I, C7, 8 and 9 and perforin. The LDL receptor class B repeat (the EGF domain) is in Factor I, C7, 8 and 9, C1r and C1s.

Diagonally-shaded sections represent sequences unrelated to domains discussed above.

The diagram comes from Campbell et al., 1987.


Ancestral 60 amino acid structure
SP
Serine protease domain
Decay acceleration factor
IL-2 receptor
C2 and Factor B
Classical serine proteases
C4 binding protein
C1r and C1s
(A,B)
C7
Factor H
Factor I
C8α
C8β
β2 glycoprotein I
C9
Factor XIII
β-subunit
LDL Receptor
A B
LDL Receptor repeats
CR 1

C H A P T E R I I

METHODS

2.1 Materials

(a) Enzymes

All enzymes were from New England Biolabs, BRL or Amersham except for:-

Lysozyme, RNase and DNase	-	Sigma
Calf intestine alkaline phosphatase	-	Boehringer Mannheim
DNA Pol I (Klenow fragment) for sequencing	-	Pharmacia
Proteinase K (Fungal)	-	BDH

(b) Chemical Reagents

Chemical Reagents were obtained from BDH, Sigma, Fisons and Boehringer Mannheim except for:-

Yeast extract	-	Oxoid
Agarose (unless otherwise stated)	-	Miles
Bactotryptone, agar	-	Difco
Dextran sulphate, Ficoll	-	Pharmacia

(c) Bacteria, Plasmids, Cosmids and Phages:-

E. coli strains were all derived from the K12 strain

MC1061	-	(Casadaban and Cohen, 1980)
JM101	-	(Messing et al., 1981)
TG1		
Q358	-	(Karn et al., 1980)
Q359	-	(Karn et al., 1980)

Plasmids	-	pAT153/PvuII/8	(Anson et al., 1984)
		pUC8	(Vieira and Messing, 1982)
Phage	-	M13 mp 8/9	(Messing and Vieira, 1982)
		EMBL3	(Frischauf et al., 1983)
Cosmids	-	PTCF	(Grosveld et al., 1982)
		PDV-Cos	(Gift from J. Rees).

(d) Media:-

2 x TY liquid media : 15 g bactotryptone, 10 g yeast extract, 5 g NaCl to 1 litre with water.

L-broth liquid media : 10 g bactotryptone, 5 g yeast extract, 10 g NaCl to 1 litre with water.

T-broth liquid media : 10 g bactotryptone, 5 g NaCl to 1 litre with water.

To make solid media 14 g agar/agarose was added to 1 litre of the appropriate liquid media.

Ampicillin to final concentration of 50 µg/ml (liq.) + 100 µg/ml (solid) was also added when necessary.

(e) Buffers:-

(Unless otherwise stated Tris = Tris-HCl)

Hogness freezing buffer - 4% glycerol/3.6 mM dipotassium hydrogen phosphate/1.3 mM potassium di-hydrogen phosphate/2 mM sodium citrate/1 mM $MgSO_4$. (Hogness and Simmons, 1964)

Lysis buffer (for plasmid prep.) - 50 mM glucose/10 mM EDTA/25 mM Tris, pH 8.0.

SM buffer - 100 mM NaCl/8mM $MgSO_4$. $7H_2O$/50 mM Tris, pH 7.5/2% w/v

gelatin.

NDS buffer - 500 mM EDTA/10 mM Tris/1% w/v lauroyl sarcosine, pH 9.5

x 10 Nick Translation buffer - 500 mM Tris, pH 7.5/70 mM $MgCl_2$/10 mM β mercaptoethanol

10 x TBE - 900 mM Tris, pH 8.0/900 mM boric acid/25 mM EDTA.

40% bis-acrylamide - 38 g acrylamide/2 g N,N'-Methylene-bis-acrylamide to 100 ml with water and deionised by stirring with 5 g 'Amberlite' MB-3 monobed resin followed by filtration and storage in a brown stoppered bottle.

1 x TAE - 40 mM Tris-acetate/ 1 mM EDTA

Hybridisation buffer - 1 M NaCl/50 mM Tris, pH 7.4/10% w/v dextran sulphate/0.2% w/v Ficoll/0.2% w/v BSA/0.2% w/v polyvinyl pyrolidone/0.1% w/v sodium pyrophosphate/0.1% SDS.† 50% formamide

Oligo hybridisation buffer - 0.2% w/v Ficoll/0.2% w/v polyvinyl pyrolidone/0.2% BSA/6 x SSC

20 x SSC - 3 M NaCl/300 mM Tri sodium citrate

x 10 kinase buffer - 500 mM Tris, pH 7.4/100 mM $MgCl_2$/100 mM DTT

TEMED - N,N,N',N' - Tetra methylethylene diamine

Chloroform - Chloroform; Isopropanol (24:1)

Phenol - Equilibrated and stored under 20 mM Tris, pH 7.5

2.2 Overnight cultures

Bacteria from a frozen glycerol stock or a single colony from a plate were used to inoculate 1-50 ml liquid media and incubated at 37°C overnight.

E. coli JM101 or TG1 were grown in 2 x Ty broth

E. coli MC1061 were grown in L-broth

E. coli Q358/Q359 were grown in T-broth 0.2% maltose/10 mM $MgSO_4$

2.3 Preparation of DNA

2.3.1 Small scale isolation of plasmid DNA (10-100 ml) (Birnboim and Doly, 1979)

Cultures of *E. coli* were grown in L-broth containing 50 µg/ml ampicillin to an A_{600} = 0.8 OD units, 1 ml was removed and 10% v/v Hogness freezing buffer (Section 2.1e) added, mixed gently and stored at -70°C as a stock. To the remainder of the culture chloramphenicol was added to 200 µg/ml (Clewell and Helsinki, 1972) and the culture incubated at 37°C overnight.

The cells were pelleted at 2,000 rpm for 20 min at 4°C and then resuspended in lysis buffer (200 µl for 10 ml culture) (Section 2.1e) containing 5 mg/ml lysozyme and left on ice for 30 min. Two volumes of 0.2 M NaOH/1% SDS was added and incubated on ice for 5 min, this was followed by one and a half volumes of 3 M NaOAc, pH 4.8 which was incubated for a further 30 min on ice. This solution was then centrifuged 11,000 rpm, the clear supernatant transferred to a clean tube and the plasmid DNA precipitated using $^{1}/3$ of the total volume, of isopropanol. The DNA was pelleted after precipitation at 11,000 rpm for 10 min and resuspended in 0.3 M NaOAc (pH 7.4) (500 µl for 10 ml starting culture). This was extracted using an equal volume of phenol/chloroform (1:1) (Section 2.1e) and the DNA was again precipitated after the extraction by the addition of 2.5 volumes of absolute ethanol. The pellet was washed twice with 70% ethanol, dried and resuspended 10 mM Tris/0.1 mM EDTA, pH 7.4 (20 µl for 10 ml starting culture). The plasmid DNA was incubated at 37°C for 30 min with heat inactivated pancreatic RNase (20 µg/ml) then stored at -20°C.

2.3.2 Rapid isolation of plasmid DNA (Ish-Horowicz and Burke, 1981)

For analysis of a large number of clones a rapid plasmid isolation

method was used. A single colony was used to inoculate 2 ml L-broth containing ampicillin (50 µg/ml) and incubated overnight at 37°C. Cells from 1.4 ml of culture were pelleted for 1 min in the microfuge and resuspended in 100 µl lysis buffer (1 vol) (Section 2.1e) and lysed with 2 vols 0.2 m NaOH/1% SDS on ice for 2 min; followed by the addition of 1.5 vols 3 M K^+/5 M OAc, pH 4.8, on ice for 5 min. The solution was centrifuged for 10 min in the microfuge and the supernatant transferred to a clean Eppendorf and $^2/3$ vols of isopropanol added to precipitate the plasmid DNA on ice for 30 min. The plasmid DNA was recovered by centrifugation in the microfuge for 5 min and the pellet washed with 70% ethanol and resuspended in 50 µl 10 mM Tris/0.1 mM EDTA, pH 7.4 and reprecipitated with $^1/10^{th}$ vol. 3 M NaOAc pH 7.4 and 2.5 vols absolute alcohol. The pelleting and washing was repeated after 30 min on ice and the final pellet resuspended in 50 µl, 10 mM Tris/0.1 mM EDTA, pH 7.4 and stored at -20°C.

2.3.3 Large scale isolation of plasmid DNA (500-1000 ml) (Radloff et al., 1967; Maniatis et al., 1982).

Cultures of E. coli were grown in L-broth containing ampicillin (50 µg/ml) and treated with chloramphenicol (Section 2.3.1). After overnight incubation the cells were harvested and lysed as described previously (Section 2.3.1). The phenol/chloroform extraction step was omitted and the DNA precipitated with isopropanol. The DNA pellet was resuspended in 15 ml of 150 mM NaCl/ 50 mM Tris/5 mM EDTA, pH 8.0 and ethidium bromide added to a final concentration of 0.02%, this solution was weighed and an equal weight of cesium chloride added. The final density of the solution was calculated and adjusted to 1.61 g/ml. The solution was then transferred to a 38 ml Beckman ultracentrifuge tube,

topped up with liquid paraffin, capped, sealed and centrifuged for 17 hr at 35,000 rpm at 18°C in a Ti60 rotor. The centrifuge was stopped without using the brake and the solution transferred away from the large RNA pellet to a clean centrifuge tube. This was capped and respun for 60 hr at 35,000 rpm at 18°C. Two bands were visible in the gradient, after spinning, the lower band of closed circular DNA was collected by puncturing the tube with a needle just below the band and collecting the band as it dripped out via the sterile needle. Ethidium bromide was removed by repeated extractions with equal volumes of 1-butanol. The DNA was recovered by addition of 3 vols water and 2 vols ethanol. The DNA was reprecipitated once, washed, dried and resuspended in 10 mM Tris/0.1 mM EDTA, pH 7.4. The DNA concentration was estimated assuming that A_{260} = 1.0 when DNA concentration is 40 µg/ml.

[More recent large scale DNA preparations use the programmable Beckman L8 ultracentrifuge to relax the gradient spin after 16 hr - Section 2.3.4].

2.3.4 Isolation of the replicative form of bacteriophage M13mp8

Ten ml of an overnight culture of *E. coli* TG1 was used to inoculate 1 litre of 2 x Ty liquid broth and incubated at 37°C until A_{600}=0.60D units. Ten ml of an overnight culture of *E coli* TG1, infected with a single viral plaque of bacteriophage M13 mp8, were added to the culture and incubated at 37°C for a further 1½-2 hr ($A_{600} \cong$ 1.0 OD unit). The cells were harvested and lysed as described previously (Section 2.3.1). The bacteriophage DNA was purified by centrifugation to equilibrium in CsCl gradients prepared as described (Section 2.3.3) but spun in Beckman L8 ultracentrifuge at 55,000 rpm for 16 hr and relaxed to 45,000 rpm for 1 hr. After butanol extractions and ethanol precipitations the DNA was resuspended in 10 mM Tris/0.1 mM EDTA, pH 7.4 and stored at 4°C.

2.3.5 Small scale isolation of bacteriophage λ DNA (Maniatis et al., 1982)

Approximately 10^5 plaque forming units (pfu) of bacteriophage were mixed with 2 ml of E. coli Q358 and incubated at 37°C for 30 min. This was then mixed with 9 ml molten T-top agarose, poured onto 14 cm T-broth agarose plates and incubated at 37°C to give confluent lysis. 12 ml S M buffer (Section 2.1e) was added to the plates and incubated at 37°C for 2 hr with gentle shaking. Bacterial debris was removed by centrifugation in the Sorval RC-5B at 8,000 rpm for 10 min at 4°C using the SS-34 rotor. Pancreatic DNase I and RNase A were added to the supernatant each at a final concentration of 10 µg/ml and incubated at 37°C for 30 min. An equal volume of 20% (w/v) polyethylene glycol/2M NaCl in SM buffer was added and incubated for 1 hr at 0°C. The bacteriophage particles were recovered by centrifugation at 10,000 rpm for 20 min at 4°C using the SS-34 rotor. The pellet was resuspended in 0.5 ml SM buffer and debris removed by centrifugation in a microfuge for 2 min. The supernatant was incubated in 0.1% SDS/5 mM EDTA, pH 8.0 at 68°C for 15 min. The aqueous phase was then extracted once with phenol and then subsequently with phenol/chloroform (1:1) and chloroform (Section 2.1e). The phage DNA was then precipitated with an equal volume of isopropanol and pelleted in the microfuge. The pellet was extensively washed with 70% ethanol, dried and resuspended in 50 µl 10 mM Tris/0.1 mM, pH 7.4 and stored at 4°C.

2.3.6 Isolation of chromosomal DNA (Bell et al., 1981)

Chromosomal DNA was isolated from white cells ('buffy coat) of blood, the 'buffy coat' being the layer of cells that forms between the red cells and plasma when whole blood is centrifuged at 2,500 rpm for 30

min. Blood was collected in EDTA, pH 7.5 at a final concentration of 10 mM, to prevent clotting.

Nuclei were prepared by adding lysis buffer, 0.32 M sucrose/10 mM Tris, pH 7.4/5 mM $MgCl_2$/1% Triton X-100 (40 ml/10 ml whole blood), to cells and centrifuging at 2,000 rpm for 15 min at 4°C. The supernatant was carefully decanted, leaving behind the nuclei and cell membranes in the pellet. Lysis was repeated twice more and the nuclei resuspended in 10 ml 150 mM NaCl/5 mM EDTA/50 mM Tris, pH 8.0. Proteinase K was added to 200 µg/ml; SDS to 0.5% and the suspension incubated at 37°C until all the protein and membranes were solubilised. The solution was then extracted once with an equal volume of phenol and then subsequently with equal volumes of phenol/chloroform (1:1) and chloroform (Section 2.1e). After each extraction the organic/aqueous mix was separated into phases by centrifugation at 2,000 rpm for 15 min at 4°C and the organic phase removed using a syringe leaving behind the aqueous phase and interface for subsequent extractions. After the final chloroform extraction the aqueous phase was removed to a clean tube and the chromosomal DNA precipitated with 2.5 vol of cold absolute ethanol. The DNA precipitated immediately and was removed with a sealed Pasteur pipette. It was briefly washed in 70% ethanol and resuspended overnight at 4°C in 10 mM Tris/0.1mM EDTA pH 7.4. The DNA was dialysed against 10 mM Tris/0.1 mM EDTA, pH 7.4 and then stored at 4°C at a concentration of approximately 1 mg/ml.

2.3.7. Preparation of high molecular weight DNA for Pulsed Field Gel Electrophoresis (Dunham et al - in press)

High molecular weight DNA from cells (from an HLA homozygous lymphoblastoid cell line - Ice 5) embedded in low-gelling temperature agarose treated with DE-52 cellulose ion-exchange resin (Section 2.7.3)

was a gift from Ian Dunham and was prepared as described by Dunham et al., 1987. Each block of agarose contained approximately $3x10^6$ cells equivalent to 18 µg DNA and provided enough DNA for three tracks on a gel. The blocks were stored in NDS buffer (Section 2.1e).

2.4 Digestion of DNA by restriction endonuclease enzymes

2.4.1 Digestion of cloned and genomic DNA

Restriction enzyme digests were carried out using the conditions recommended by the manufacturer. Analytical digests used 10-50 ng DNA in a final volume of 10-20 µl if end-labelled or 0.5-1.0 µg DNA in a similar volume for agarose mini gels. Preparative digests and digests for large agarose gels, used 0.5-10 µg DNA in 50-100 µl reactions. Routinely the amount of enzyme used was in excess of the amount recommended by the manufacturer (1 unit of enzyme is the amount stated by the manufacturer to fully digest 1 µg of λ DNA in 1 hr). Genomic DNA was digested for 16-24 hr, all other DNA samples were fully digested after 2 hr.

2.4.2 Digestion of DNA in agarose blocks for Pulsed Field Gel Electrophoresis (PFGE) (Anand, 1986)

Enzymes used in PFGE analysis are all infrequently cutting enzymes and fall into one of the following classes (i) those with 8 bp recognition sequences or (ii) those containing one or more CpG dinucleotides within their recognition sequences.

Each block was cut into three equal parts and one third used per digest. Before digestion the blocks were washed three times in cold buffer, 10 mM Tris, pH 7.4/0.1 mM EDTA/0.1 mM PMSF (phenyl methyl sulphonyl fluoride), on ice for 30 min. Each section of block was transferred to an individual Eppendorf tube and 500 µl of the

appropriate restriction enzyme buffer (as recommended by the manufacturer) but lacking gelatin, DTT and spermidine, added. This was incubated on ice for 30 min and then replaced with 100 µl of the complete restriction enzyme buffer and 20 units of enzyme and incubated at the appropriate temperature for 3 hr. The reaction was stopped by the removal of the buffer and enzyme, followed by the addition of 200 µl NDS (Section 2.1e) and Proteinase-K (1mg/ml) and incubation at 50°C for 2 hr. The plugs were then loaded directly into the gel (Section 2.6.4).

2.5 Radiolabelling of DNA

2.5.1 'Fill-in' reaction (Wu and Taylor, 1971; Maxam and Gilbert, 1980)

The recessed 3' ends, produced by digestion of double stranded DNA with suitable restriction endonuclease enzymes were labelled directly by incubating with 100 µM cold dNTPs/5 units DNA polymerase I (Klenow fragment)/medium salt restriction enzyme buffer/1-50 µCi α^{32}P[dNTP], for 20 min at room temperature. To ensure all fragments were 'filled-in' efficiently the reaction was chased with cold dNTPs at a final concentration of 250 µM, for a further 10 min at room temperature. The reaction was stopped by heating at 70°C for 20 min.

2.5.2 'Nick-translation' (Rigby et al., 1977)

Ten-50 ng DNA in 10 mM Tris/0.1 mM EDTA pH 7.4 was mixed in nick-translation buffer (Section 2.1e) with 20 µCi α^{32}P[dATP], 40 µM dCTP, dGTP, dTTP and 5 µl Amersham 'nick-translation enzyme (2.5 U DNA Polymerase I and 50 pg DNase) in a total volume of 100 µl. The reaction was incubated at 16°C for 2½ hr then terminated with 10 mM EDTA. The unincorporated label was removed by filtration on a 5 ml Sephadex G50-SF column.

2.5.3 Primer extension labelling (Feinberg and Vogelstein, 1983)

Ten-30 ng DNA in 10 mM Tris/0.1 mM EDTA, pH 7.4 was labelled using the "standard multiprime DNA labelling system" from Amersham. This is based on the use of random sequence hexanucleotides to prime DNA synthesis by the 'Klenow' fragment of DNA Polymerase I as described by Feinberg and Vogelstein, 1983.

2.5.4 Kinase labelling (Van de Sande et al., 1973; Maxam and Gilbert, 1977)

Oligonucleotides were synthesised using an Applied Biosystems automated synthesiser, after synthesis the oligo was eluted from the support resin and deprotected by incubation at 55°C in 2 ml concentrated ammonia solution. 360 µl of the ammonia solution was further purified by alcohol precipitation for kinase labelling and the remainder stored at 4°C.

Forty-50 ng of purified oligo was labelled using T4 kinase by incubation in kinase buffer (Section 2.1 e) with 60 µCi γ^{32}P[dATP] and 5.0 U T4 kinase, in a total volume of 20 µl, for 30 min at 37°C. The reaction was stopped with 10 mM EDTA and 5 µl formamide dyemix (Section 2.6.1b) added. The labelled oligo was purified from the unincorporated label by running on a 20% denaturing acrylamide gel (Section 2.6.1) and the probe eluted by excision of the appropriate band from the gel (identified by autoradiography) and incubation of the gel piece in 1 ml water for 2 hr at 37°C.

2.6 Gel Electrophoresis

2.6.1 Polyacrylamide gel electrophoresis (Sanger and Coulson, 1978; Maxam and Gilbert, 1980)

Glassplates (40 cm and 20 cm) were spaced with 0.3 - 0.6 mm strips

of plastic card and taped together with waterproof electrical tape, prior to taping the top plate was treated with a 2% solution of dimethyl dichlorosilane in 1:1:1 trichlorethane. Sample wells 2 mm - 2 cm in width were formed at the top of the gel by insertion of a shaped plastic card, 'comb' between the glass plates after pouring the gel.

(a) Native gels - 4 -10% gels were prepared by mixing stock acrylamide solution (Section 2.1e) with 5 ml 10 x TBE (Section 2.1e) in a total volume of 50 ml. The polymerisation was catalysed with 0.4 ml ammonium persulphate 10% and 40 µl TEMED. A glycerol dye mix, 30% glycerol in 5 mM EDTA, 0.1% xylene cyanol and bromophenol blue was used for running samples.

(b) Denaturing gels - gels were again prepared from the acrylamide stock solution (Section 2.1e) with 5 ml 10 x TBE (Section 2.1e) and catalyzed in the same way. 21 g of urea was dissolved in the acryl-amide/TBE and then made up to volume with water prior to polymerisation. A formamide dye mix was used, 60% formamide in 35 mM EDTA, 0.1% xylene cyanol and bromophenol blue and all samples containing formamide dyes were boiled for 5 min then snap cooled on ice prior to loading.

Gels were electrophoresed in 1 x TBE buffer for 1½ - 4 hr at a constant current of 15 mA for 20% gels and 25-30 mA for lower percentage gels.

2.6.2 Autoradiography (Laskey and Mills, 1977)

Gels were covered with cling-film and autoradiographed using X-OMAT 'S' X-ray film (Kodak) at -70°C with an intensifying screen (Cronex Lightening Plus, Dupont). The films were sensitised by pre-flashing and developed in an ME-1 X-OMAT automatic processor.

2.6.3 Agarose gel electrophoresis (Southern, 1979)

Agarose horizontal slab gels of varying agarose percentage (0.7-1.2% in 1xTBE) were run using the BRL apparatus. 20 x 24 x 0.8 cm gels were poured using a 20 x 100 µl well former and were run in 1 x TBE running buffer. Gels were stained by the addition of 20 µg ethidium bromide/100 ml gel before pouring. (Genomic blots were stained after running by submerging the gel in 1 x TBE containing ethidium bromide, 2 mg/litre, for 30 min). Samples were mixed with glycerol dyes and electrophoresed overnight at 50 - 70 mA.

11 x 11 x 0.5 cm minigel apparatus (made by Biochemistry Department Workshop) were used for rapid restriction mapping and for less accurate resolution the 'Uniscience' minigel apparatus (Johnson and Crossman, 1977) was used. 50 and 30 ml of agarose in 1 x TBE was used for gels in the respective apparatus and the running buffer was 1 x TBE. Gels were stained with 3 µl of 1% ethidium bromide solution in the running buffer. Samples were mixed with glycerol dyes and electrophoresed at 30 - 100 mA for 1-3 hr.

The ethidium bromide stained DNA was visualised under short wave ultra-violet light (Sharp et al., 1973) using a transilluminator and photographed using a Polaroid MP-4 land camera fitted with a red filter. Exposures were for 1 sec at f4.5.

2.6.4 Agarose gels for Pulsed Field Gel Electrophoresis

A 20 x 20 x 0.8 cm 1.5% agarose (Sigma Type I) gel in 0.5 x TAE (Section 2.1e) was used and run in a chamber based on the design of Carle and Olsen, 1984, with 4 litres of 0.5 x TAE as the running buffer. Markers were λ DNA oligomers (Anand, 1986). Gels were run initially at 50 V, 9°C for 30 min with a pulsed time of 50s, this was to run in the

markers and these parameters were then adjusted as described in the results, Chapter VI.

Gels were stained, after electrophoresis, using 200 µl 0.5% ethidium bromide in 500 ml 0.5 x TAE for 30 min and destained for between 30 min - 1 hr in deionised water. The stained DNA was visualised and photographed as described for agarose gels (Section 2.6.3).

2.7 Recovery of DNA from gels

2.7.1 Elution of DNA from acrylamide gels (Maxam and Gilbert, 1980)

The section of gel containing DNA of interest (identified by the superimposition of the autoradiograph on the gel) was excised and eluted in 2 M-ammonium acetate at 37°C, overnight with constant shaking. After elution the supernatant was removed and the gel slice washed with 0.5 vol 2M ammonium acetate. After pooling the supernatants were centrifuged in a microfuge for 5 mins to remove residual acrylamide. The DNA was then precipitated by the addition of 2.5 vols absolute alcohol and incubation at -70°C for 30 min followed by pelleting in the microfuge. The DNA was resuspended in 300 µl 0.3 M sodium acetate, pH 7.4 and ethanol precipitated twice more before washing in 70% ethanol and allowing to dry open on the bench. The DNA was resuspended in 25-50 µl 10 mM Tris/0.1 mM EDTA, pH 7.4.

2.7.2 Electroelution of DNA from agarose gels (Dretzen et al., 1981)

The DNA, appropriately digested, was run on an agarose minigel (Section 2.6.3) of appropriate agarose percentage for good separation of the band of interest. When this band was well separated an incision through the gel was made, to the right of the band, this was done by visualising the stained DNA with a UV transilluminator. A strip of DEAE-paper, pretreated for 3-4 hr in 2.5 M NaCl and stored in 1 mM EDTA

(pH 8.0) at 4°C, was inserted into the slit. The sides of the gel were trimmed and the gel turned through 90°C in the gel box, enough running buffer to just cover the electrodes was added and the gel connected so as to run the DNA onto the DEAE paper. When all the DNA from the band had collected on the paper, the paper was removed, rinsed in distilled water, cut into small pieces and placed in a 0.5 ml Eppendorf tube. The DNA was eluted in 250 µl of buffer, 20 mM Tris (pH 7.5)/1 mM EDTA/1.5 M NaCl, at 37°C overnight. The buffer was spun out into a larger Eppendorf, through a hole pierced in the bottom, using a microfuge. The paper was then washed twice with 50 µl of the same buffer at room temperature and once at 65°C. The washings were then pooled and the DNA precipitated with ^{1}/10th vol 3 M sodium acetate, pH 7.4 and 2.5 vol absolute alcohol, pelleted and washed with 70% ethanol, the pellet was resuspended in 500 µl 10 mM Tris/0.1 mM EDTA, pH 7.4. This was then extracted with an equal volume of phenol/chloroform (1:1) (Section 2.1e) after which the DNA was reprecipitated as outlined above and finally resuspended in 50 µl 10 mM Tris/0.1 mM EDTA, pH 7.4.

2.7.3 Recovery of DNA from low gelling temperature agarose

A 2% solution of Sigma type VII low gelling temperature agarose in water, produced by heating to above 50°C, was mixed with a 5 ml packed volume of DE-52 cellulose (Whatman) (pre-equilibriated and stored at 4°C in 5 x PBS, Phosphate buffered saline) and held at 50°C for 30 min. The resin was removed by centrifugation at 2,000 rpm for 5 min and the agarose transferred to a clean tube. This purification was repeated and followed by clarifying spins to remove traces of resin, keeping the agarose at 50°C for 10 min between spins to keep it molten. The molten agarose was then split into 25 ml aliquots and ^{1}/10th vol of 10 x TBE (Section 2.1e) added. One melted aliquot of purified agarose was

sufficient for one gel and was diluted with 1 x TBE if a lower percentage agarose was needed. The gels were poured and run at 4°C using the 'Uniscience' minigel apparatus (Section 2.6.3) with a 4 x 2 cm well former and loading a maximum of 2 µg total DNA per well. The gel was run at 4 mA until the required band was well separated, this band was then cut out, using a uv transilluminator to visualise the DNA. The excised band was placed in a pre-weighed Eppendorf tube and its weight calculated. A volume of water (in µl) equivalent to three times the weight of excised agarose (in mg) was added, heated at 70°C for 10 min and cooled to room temperature.

The solution was then extracted with phenol, phenol/chloroform (1:1) and finally chloroform (Section 2.1e) and the DNA precipitated by addition of ^{1}/10th vol 3 M sodium acetate (pH 7.5) and 2.5 vol absolute alcohol. The pellet was washed extensively with 70% ethanol, air dried and resuspended in a appropriate volume of 10 mM Tris/0.1 mM EDTA, pH 7.4.

2.8 Isolation of cloned DNA

2.8.1. Screening of plasmid clones (Denhardt, 1966; Gergen et al.,1979)

Colonies were plated onto agar/ampicillin plates (Section 2.1d) and grown overnight at 37°C. Bacterial colonies were transferred to Whatman 541 filters by leaving the filter in contact with the colonies for about 10 min at room temperature. The filters and plates were marked in identical fashion so they could be matched up later. The plates were sealed and stored at 4°C while the filters were processed. Colonies were lysed *in situ* by floating the filters colony side up on the following solutions;

0.5 M NaOH	-	5 min
0.5 M Tris, pH 7.4	-	5 min

0.5 M Tris, pH 7.4 - 5 min

2 x SSC (Section 2.1e) - 5 min

- followed by a short wash in absolute alcohol and allowing the filters to dry at room temperature.

The filters were moistened in 6 x SSC (Section 2.1e) and then prehybridised at 42°C for at least 3 hr in hybridisation buffer (Section 2.1e) containing 50% formamide, 100 µg/ml sonicated salmon sperm DNA and 10 µg/ml E. coli/pAT153/PvuII-8/DNA. This mixture was boiled for 10 min and snap cooled on ice before use. The filters were hybridised overnight in the above buffer containing 5×10^5 cpm/ml of [^{32}P]-labelled probe, also boiled and snap cooled before use.

After hybridisation the filters were washed as follows:

3 times in 2 x SSC (Section 2.1e)/0.1% SDS for 20 min at room temperature

3 times in 0.2 x SSC for 20 min at 68°C.

The filters were air dried, wrapped in Saran wrap and autoradiographed (Section 2.6.2). The autoradiograph was used to match with original plates via the markings on the plates.

To rescreen the clones the colony was scraped from the surface of the original plate using a sterile wooden toothpick, diluted in L-broth, replated and screened as described above. The rescreening procedure was repeated until a colony, not contaminated with neighbouring bacteria could be picked from the plate and the plasmid DNA isolated for analysis by restriction mapping and nucleotide sequencing.

2.8.2 Screening of cosmid clones (Steinmetz et al., 1985)

(a) Replica plating

Colonies were plated onto 20 x 20 cm nitrocellulose filters (Schleicher and Schull) on 22 x 22 cm agar/ampicillin plates (Section

2.1e) to a density of 1 x 10^5 colonies per plate and grown overnight at 37°C. The colonies were pin-head size by next morning and replicas taken (Hanahan and Meselson, 1983). The master filter was transferred onto a glass plate and a new identically sized filter, prewetted on an L-ampicillin plate, placed onto it with the agar side downwards. A second glass plate was then placed on top and pressed evenly all over. After removing the top plate an asymmetric pattern of markings were keyed through both filters using a sterile needle and india ink. The replica was carefully peeled off and placed colony side upwards on a fresh L-ampicillin plate. The master was returned to its plate. The process was then repeated with a second and third replica filter but this time with the master filter on top between the glass plates so as to allow transfer of identical markings to subsequent replicas. The initial replica was incubated at 37°C until colonies were pinhead size on L-ampicillin plates and stored at 4°C as the short term masters. The last two replicas were incubated until the colonies were 1 mm in diameter and then lysed and used for screening (Section 2.8.2 b & c). The original masters were recovered by incubation at 37°C on on L-ampicillin plates containing 1 x Hogness freezing buffer (Section 2.1e) until the colonies were raised and healthy but still only pinhead size. These plates were then sealed and frozen at -70°C as the long term masters.

b) Lysis and binding of DNA -

4 large, shallow, plastic trays lined with Whatman 3 MM paper were used to lyse the bacterial colonies on the nitrocellulose. The Whatman was soaked in:-

Tray 1;	10% SDS
Tray 2 & 3;	0.5 M NaOH
Tray 4;	10 M Tris pH 8.0

(Grosveld et al., 1981; Maniatis et al., 1982)

Excess liquid was poured off each tray just before placing the nitrocellulose, colony side upwards onto the Whatman, the filters were left on the 10% SDS for 4 min, then transferred to the 0.5 M NaOH for 2 min each and then neutralised on the 1 M Tris, pH 8.0 for 4 min (the Whatman paper was replaced between the two sets of replicas). The replicas were collected in a plastic box containing 1.5 M NaCl/0.5 M Tris, pH 8.0 and washed for 15-20 min at room temperature.

After this initial wash each filter was transferred individually, to a solution of 2 x SSC/0.1% and wiped very gently with a tissue to remove bacterial debris. The filters were then collected into a solution of 2 x SSC and washed again at room temperature, for 15 min. The filters were then air dried and baked for 2-3 hr at 80°C. The filters were ready to be screened.

(c) Hybridisation of the filters (Bernards and Flavell, 1980)

After baking the filters were wet thoroughly in 6 x SSC for 15 min at room temperature and then transferred to a prewarmed buffer, 1 M NaCl/1 mM EDTA, pH 8.0/50 mM Tris, pH 8/0.1% SDS and washed at 42°C for 1 hr. The filters were then sealed into polythene bags using a heat sealer and 10 ml of the prehybridisation buffer (Section 2.1e) containing competitor DNA's and boiled etc. as described in Section 2.1e, was added. Air bubbles were removed and the bag sealed, these were then incubated at 42°C for a minimum of 3 hr before adding the labelled probe. The probes used had a specific activity of between 10^7-10^9 cpm/µg and between 10^5-10^6 cpm/ml was added to each bag and incubated at 42°C overnight. The filters were then washed and autoradiographed as described Section 2.8.1.

(d) Picking positive clones and rescreening.

After developing, the asymmetric markings on the filters were transferred to the autoradiograph and then the positive signal lined up

on matching replicas. The master plates were warmed to room temperature (or thawed - Steinmetz et al., 1985), the autoradiograph was placed on the light box, covered with clingfilm and the master filter then transferred, carefully lining up the markings, onto the top of the autoradiograph. Colonies corresponding to the positive singals were then scraped of using a sterile toothpick and resuspended in 500 µl L-ampicillin broth. Appropriate dilutions were then made e.g. 10^{-2}-10^{-5}, and 100 µl of each plated onto 7 cm diameter nitrocellulose filters (Amersham) on L-ampicillin plates. These were then incubated overnight at 37°C and replicas taken from a dense and sparse plate for each clone, as described above. Lysis, hybridisation and washing of the filters is as described for the original library filters. This rescreening was continued until a single positive colony was isolated.

2.8.3 Screening of bacteriophage plaques (Maniatis et al, 1982)

E. coli Q359 cells in T-broth/0.2% Maltose/10 mM $MgSO_4$ were grown at 37°C for 20 hr with constant shaking. An equivalent of 10^5 plaque forming units (pfu) was added to 1.5 ml of Q359 culture and incubated at 37°C for 30 min. The mixture was added to 7.5 ml T-top agarose, mixed and poured onto 14 cm T-agar plates and incubated at 37°C overnight. The plates were left at 4°C for at least 1 hr to allow the top agarose to harden. Dry nitrocellulose filters (Amersham) were placed in direct contact with the plaques and keyed with ink asymmetrically, marking both filter and solid medium. After 60 sec the filter was removed and immersed, DNA side up, in the following solutions

1.5 M NaCl/0.5 M NaOH (Denaturing solution) - 60 sec

1.5 M NaCl/0.5 M Tris pH 7.4 (Neutralising solution) - 5 min

2 x SSC - 5 min

- the filters were then air dried on Whatman 3 MM paper. Multiple replicas could be taken from each plate, using the original marks on the plate for each subsequent copy.

The DNA was baked onto the nitrocellulose at 80°C for 2 hr before hybridising, washing and autoradiography (Section 2.8.2).

Rescreening of plaques - agar plugs were taken from the original plates using the ink marks on the plates to line up the autoradiographs and put into 500 µl SM buffer (Section 2.1e) 1% chloroform, was added these were stored at 4°C. Various dilutions of phage solution were plated out as above using 500 µl of overnight Q358 culture. The infected cells were plated using 3.5 ml of T-top agarose on 8 cm T-agar plates and incubated at 37°C. The plaques were screened, hybridised, washed and autoradiographed using the method described above.

Single, isolated plaques were chosen from the re-screen plates and stocks of infectious particles prepared as above. These were then used to give plate lysates from which DNA was isolated (Section 2.3.5) for restriction mapping and nucleotide sequencing.

When oligonucleotides were used in the re-screening the filters were prepared in the same way as described above, the hybridising conditions were different. After baking the filters were soaked thoroughly in 6 x SSC and then transferred to oligo-hybridisation buffer (Section 2.1e) containing 100 µg/ml sonicated salmon sperm DNA and incubated at 42°C for 2-3 hr. Oligonucleotides labelled by kinase labelling methods (Section 2.5.4) to a specific activity of $\sim 1 \times 10^9$ cpm/µg + filters hybridised using 2×10^5 cpm/ml hybridisation buffer and incubated overnight at 42°C. Filters were washed at room temperature in 6 x SSC/0.1% SDS prewarmed to 42°C, changing the washing buffer every 30 min until the background on the negative control filter fell to 2-3 cpm. The filters were then air dried and autoradiographed.

2.9 DNA transfer blot analysis (Southern, 1979; Wahl et al., 1979)

(a) Using nitrocellulose (Amersham)

Samples in $^{1}/_{10}$ vol. glycerol dye mix were electrophoresed in agarose gels/1 x TBE (Section 2.6.3) at 50 mA for approximately 16 hr. The gel was stained with ethidium bromide, photographed and then washed in the following solutions -

(1) Twice in 0.25 M HCl for 15 min at room temperature

(2) Twice in 0.5 M NaOH/1M NaCl at room temperature

(3) Twice in 3 M NaCl/0.5 M Tris, pH 7.4 at room temperature

The gel was rinsed with water between washings. Blotting of agarose minigels used shorter washing times and depending on the size of fragments the acid wash was totally omitted.

The gel was placed on two sponges and two pieces of Whatman 3 MM, all soaked in 20 x SSC (Section 2.1.e). The surrounds of the gel were covered with cling-film and glass rods placed at the ends of the gel. A piece of nitrocellulose (Amersham) briefly floated on water was placed on top of the gel and this covered with two pieces of Whatman 3 MM also soaked in water. Approximately two inches of paper towels, glass plate and book were placed on the top to weight the blot. The buffer then passes through the gel by capillary action carrying the DNA onto the nitrocellulose. Miniblots were weighted only with paper towels and a tissue culture microtitre plate.

With cloned DNA multiple copies of the blot were made, the first blot was taken for 30 min and subsequent copies for 1 hr and 2 hr. Miniblots were taken for 15 min, 30 min and 1 hr, respectively, and single copies of genomic blots were taken for 24 hr. The sponges were kept moist with additional 20 x SSC and the towels replaced when necessary to maintain the capillary action. The DNA was baked onto the nitrocellulose and hybridised as described in Section 2.8.2.

(b) Using Gene-Screen Plus membrane (Dupont)

This membrane was used for the pulsed field gel electrophoresis blots. After running the gel was treated as follows -

(1) Once in 0.25 M HCl for 15 min at room temperature

(2) Once in 0.4 M NaOH/0.6 M NaCl for 30 min at room temperature

(3) Once in 1.5 M NaCl/0.5 M Tris, pH 7.5 for 30 min at room temperature

The blotting arrangements were as described above using Gene-screen plus cut to the size of the gel and marked on the concave side whilst still dry. The membrane was first floated on deionised water to wet thoroughly and then equilibrated in 10 x SSC (Section 2.1e) for 15 min. During blotting the marked, concave side of the membrane was in contact with the gel. 10 x SSC was used for the transfer buffer. Blotting was carried out for 24 hr after which the membrane was treated in 0.4 M NaOH for 30-60 sec followed by 0.2 M Tris pH 7.5/2 x SSC for 5 min. Filters were not baked and wrapped in Saran wrap to prevent complete drying.

Hybridisation was carried out as in Section 2.8.2, however, the hybridisation buffer contained 1% SDS. After 18-20 hr at 42°C with radiolabelled probe, specific activity 3-5 x 10^8 cpm/µg and 1 x 10^6 cpm/ml hybridisation buffer the filters were washed as follows -

Three times in 2 x SSC at room temperature for 5 min

Twice in 0.2 x SSC/1% SDS at 65°C for 30 min

Once in 0.2 x SSC at room temperature for 30 min.

The filters were not dried completely before wrapping in Saran wrap and autoradiographing for 48 hr.

Between the C4bp and Factor H probes the blot was stripped by washing twice in 2 mM Tris, pH 7.4/1 mM EDTA/1% SDS at 80°C for 1

hr. Filters were autoradiographed overnight to check removal of probe before rehybridising with the second probe.

2.10 Nucleotide Sequencing Techniques

2.10.1 Chemical Degradation (Maxam and Gilbert, 1980)

(a) Modification Reaction

DNA was digested (Section 2.4) and radiolabelled (Section 2.5.1) and the reaction terminated by heating at 70°C for 20 min. A second digest was then used to give fragments of DNA labelled at one end only, these were then electrophoresed on an acrylamide gel and the appropriate band eluted from the gel (Section 2.7.1). The final pellet was dissolved in 25 µl water and 6 µl 0.75 mg/ml sonicated DNA added. This was then aliquoted between four eppendorfs as follows:-

T + C: 10 µl, A + G: 10 µl, C: 5 µl, G: 5 µl - these were then scintillation counted to check an even distribution of DNA.

The 'C + T stop' solution was prepared, 0.3 M NaOAc, pH 7.0/0.1 mM EDTA/10 ng tRNA and also the 'G stop', 1.5 M NaOAc, pH 7.0/1.0 M β-mercaptoethanol/10 ng tRNA. The partial cleavage reactions were carried out as follows, all were vortexed and incubated at room temperature.

A + G; 25 µl 90% formic acid - incubate 10 min

T + C; 15 µl water + 30 µl anhydrous hydrazine - incubate 10 min

C; 20 µl saturated salt solution + 30 µl hydrazine - incubate 5 min

Each of the above reactions is stopped with 200 µl 'C and T stop'

G; 200 µl 0.05 sodium cacodylate pH 7.9/10 mM $MgCl_2$/1 mM EDTA and 1 µl dimethyl sulphate - incubate 5 min.

This reaction is stopped with 50 µl 'G stop'.

All reactions were precipitated with 1 ml absolute alcohol and pelleted by spinning for 10 min in a microfuge. The pellets were carefully washed with 70% ethanol.

(b) Cleavage reaction

The drained pellets were resuspended in 50 µl 1M piperidine solution (freshly diluted), vortexed, and heated to 90°C for 20 min in a Tecam 'Dribloc' unit. The samples were cooled on ice and spun for 4 min, the supernatants were transferred to clean tubes containing 50 µl 0.6 M NaOAc, pH 7.0. The samples were precipitated with 3 volumes of absolute alcohol. The pellets were washed with 70% ethanol and resuspended in 10 µl formamide dye mix (Section 2.6.1).

8%, 10% and 20% denaturing gels were then run (Section 2.6.1b) and autoradiographed (Section 2.6.2).

2.10.2 Dideoxy sequencing (Messing, 1983)

(a) Preparation of the vector

4 µg of the replicative form of bacteriophage M13 mp8 was cut with 20 units of Hinc II enzyme at 37°C for 2 hr, then with 2 µg calf intestinal phosphatase for 5 min. The DNA was extracted extensively with phenol/chloroform mixture (Section 2.1e) and recovered by ethanol precipitation.

(b) Subcloning

DNA restriction fragments were isolated from acrylamide or agarose gels as described (Section 2.7).

10 ng cut and phosphatased vector was mixed with 1-10 ng DNA in 1 mM rATP/50 mM Tris, pH 7.4/10 mM $MgCl_2$/20 mM DTT/50 µg/ml BSA in a final volume of 10µl. This was incubated for 12-16 hr at 16°C with 0.5 µl (approximately 2 units) DNA ligase (Winter et al., 1981).

(c) Transformation of E. coli TGI cells

The cells were made competent by the following treatment. 100 µl of an overnight culture of TGI was diluted into 10 ml 2 x Ty broth and grown to an A_{600} of 0.4 units. The cells were pelleted by

centrifugation at 1,300 rpm at 4°C for 7 min in MSE 6L. The cells were resuspended, gently in 5 ml 50 mM $CaCl_2$/10 mM Tris, pH 7.4 and incubated for 20 min on ice. The cells were pelleted again, resuspended in 1.5 ml of the same buffer and stored on ice. The cells are viable for up to 24 hr.

100 µl of the competent cells were added to the ligation mix, incubated at 0°C for 45 min and then at 42°C for 5 min. The transformed cells were added to 3 ml molten L top agar containing 25 µl of 2.5% solution of isopropylthiogalactoside (IPTG in 10 mM Tris, pH 7.5/0.1 mM EDTA and 25 µl of 2% bromochloroindolylgalactoside (BCIG) in dimethylformamide, kept at 45°C. Finally 0.5 ml of overnight TGI cells was added gently mixed and poured onto 2 x Ty plates.

Incubation of the plates overnight at 37°C resulted in the production of blue and white plaques; areas of lower density cells caused by M13, a non-lytic phage, reducing the growth rate of infected cells. Blue plaques represent bacteria containing wild type M13 phage which has β-galactosidase activity due to complementation between the M13 and TGI. IPTG induces enzyme synthesis which hydrolyses BCIG, a substrate analogue, releasing a blue indolyl derivative. The cloning site in M13 is within the β-galactosidase gene so M13 containing insert DNA results in inactivation of the β-galactosidase gene. The M13 can no longer complement the 'M15' protein of the host (Messing et al., 1977), no dye is released and white plaques result.

(d) Preparation of phage DNA (Sanger et al., 1980)

Agar plugs of white plaques were grown in 1.5 ml of one hundreth dilution of an overnight culture of TGI in 2 x Ty broth at 37°C for 5 hr, with constant shaking. The medium was transferred to an Eppendorf tube and spun for 10 min in a microfuge at room temperature. The supernatant was transferred to a fresh tube and respun for 10 min.

Again the supernatant was transferred to a clean tube and the residual kept at 4°C as a seed for future preparation of phage DNA. To the supernatant was added 200 µl 20% polyethylene glycol/2.5 M NaCl, mixed and incubated for 20 min on ice. This was centrifuged for 10 min, the supernatant discarded and the pellets resuspended in 100 µl 10 mM Tris, pH 8.0/0.1 mM EDTA. These were then extracted with 50 µl phenol/chloroform (Section 2.1e) and precipitated with 10 µl 3 M NaOAc pH 7.0 and 100 µl isopropanol for 1 hr at -20°C. The DNA was pelleted, washed in 70% ethanol, dried and resuspended in 50 µl 10 mM Tris, pH 8.0/0.1 mM EDTA. The single stranded phage DNA was stored at -20°C.

(e) Relative orientations of integrated fragments (Herrman et al., 1980)

The orientation of the cloned restriction fragments was determined as follows. 2 µl of each clone was mixed with 2 µl of a reference clone in a 0.5 ml Eppendorf tube in the presence of 4 µl 600 mM NaCl/70 mM $MgCl_2$/60 mM Tris, pH 7.4 and 1 µl 50% glycerol/1% SDS/bromophenol blue. 50 µl of liquid paraffin was added to the top and the tubes incubated at 67°C for 1 hr, snap cooled on ice and electrophoresed on a 0.8% agarose minigel. Cloned DNA containing fragments of opposite orientation hybridised with each other and migrated slower due to an increase in size.

(f) Sequencing reaction (Sanger et al., 1977; Biggin et al., 1983)

A universal oligonucleotide primer (Duckworth et al., 1981) was used to prime synthesis using the single standed DNA as a template. The sequence of the oligomer was:

5' d(CACAATTCCACACAAC) 3'

The composition of the mixes used was

'A' mix:- 0.125 mM dCTP; 0.125 mM dGTP; 0.125 mM dTTP; 15 µM ddATP

'C' mix:- 6.25 mM dCTP; 0.125 mM dGTP; 0.125 mM dTTP; 40 µM ddCTP

'G' mix:- 0.125 mM dCTP; 6.25 µM dGTP; 0.125 mM dTTP; 80 µM ddGTP

'T' mix:- 0.125 mM dCTP; 0.125 mM dGTP; 6.25 µM dTTP; 250 µM ddTTP

Approximately 1.25 ng primer was hybridised with 8 µl single stranded DNA in 13.6 mM Tris, pH 8.5/6.8 mM $MgCl_2$, in a total volume of 11 µl, for 5' at 65°C and then kept at room temperature until ready to use. 2 µl of primed clone were added to four tubes; T, C, G and A. To this was added 2 µl of the correct mix and 2 µl of enzyme premix; 2U DNA Polymerase I, 'Klenow' fragment (Pharmacia) and 2.5 µCi α^{32}P[dATP] in 11 mM DTT/0.15 µM dATP/10 mM Tris, pH 8.0. The reaction was started by mixing the reactant in a capless eppendorf tube in a microfuge. The reaction was incubated at room temperature for 15 min and then chased for another 15 min with 2 µl 0.5 mM dNTP's. The reaction mixture was stopped by the addition of 4 µl formamide dye mix (Section 2.6.1b) and the samples boiled for 3 min.

(g) Electrophoresis of sequenced samples

The samples were run on either (a) 6% denaturing gels (Section 2.6.1) at 30 mA for 5 hr or (b) buffer gradient gels at 30 mA for 1 hr and 40 min (Biggin et al., 1983). Prior to preparing both types of gels the top plate was treated with a 2% solution of dimethyl dichlorosilane in 1,1,1 trichloroethane (Section 2.6.1).

After electrophoresis the gels adhering to the back plate were fixed in 10% acetic acid for 15 min and then transferred to Whatman 3 MM paper before drying down on a gel drying machine for 1 hr. The gels were then autoradiographed for 1-3 days at room temperature, without a screen and the film was not pre-flashed.

Buffer gradient gels

0.5 x TBE gel mix:- 6% acrylamide/0.5 x TBE/7 M Urea

2.5 x TBE gel mix:- 6% acrylamide/2.5 x TBE/7 M Urea/
5% sucrose/0.05% bromophenol blue

7 ml 2.5 x TBE mix were mixed with 35 µl 10% ammonium persulphate and 14 µl TEMED. 40 ml 0.5 x TBE mix were mixed with 200 µl 10% ammonium persulphate and 80 µl TEMED. 6 ml of 0.5 x TBE mix were taken up into a 10 ml pipette, followed by 6 ml 2.5 x TBE mix, and a gradient introduced by introducing a couple of air bubbles through the interface. This mix was poured between the gel plates and the remaining space filled with 0.5 x TBE mix. 0.5 TBE was used as the running buffer in the top reservoir and 1 x TBE in the bottom.

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

Sequence data was analysed using a VaxII/computer, sequences were read from gels and entered by hand into the computer. Programmed used in data analysis include CUTSIT, ANALYSEQ and Search.

C H A P T E R I I I

ISOLATION AND CHARACTERISATION OF 5' cDNA CLONES

3.1 Introduction

Using synthetic oligonucleotides as probes Chung et al., 1985a isolated plasmid clones from a liver cDNA library which coded for human C4bp. These initial clones together with overlapping clones, subsequently isolated using cDNA probes, allowed the prediction of the amino acid sequence for C4bp from residue 32-549. Clone pBP2 contained the 3' two-thirds of the coding sequence (starting at the codon for threonine, 175) and the entire 3' non-translated region whilst pBP6 extended the 5' sequence to include the codon for threonine 32. Protein data for the N-terminal residues 1-40 was available (Chung et al., 1985c) which overlapped with the sequence derived from the 5' cDNA sequence and completed the amino-acid sequence residues 1-31.

From Northern blot analysis (Chung et al., 1985a) using human liver RNA, the C4bp mRNA was estimated to be 2.5 kb in size. The coding region for residues 1-549 and the 3' non-translated region account for 1.9 kb of the message. This leaves a total of 600 bp of cDNA to code for the 5' untranslated, leader peptide and poly A^+ tail.

Using a cDNA probe originating from the most 5' region in clone pBP2 an overlapping clone was isolated, again from a human liver cDNA library. This new clone covered the amino acid residues 1-31 and extended into the putative leader sequence. A cDNA probe was prepared from the 5' region of this clone to be used in the isolation and characterisation of genomic clones coding for the human C4bp gene.

Numbering of cDNA nucleotides in this and subsequent chapters is that used by Chung et al., 1985a, where nucleotide 1 corresponds to the 1st base pair of the codon for threonine, 32.

3.2 Isolation and initial characterisation of positive clones

The cDNA library used in this study was a gift from Dr. D. Bentley and was prepared from human liver poly A^+ RNA. The inserts in the library were prepared using S1 nuclease treated and DNA polymerase I (Klenow fragment) repaired double stranded cDNA (Bentley and Porter, 1984). The use of the self-priming mechanism in the synthesis of the second cDNA strand introduces 'loop-backs' into the double stranded DNA and subsequent treatment with S1 nuclease means that the loop-back region was cloned. This results in the 5' region of inserts being inverted onto the opposite strand, however comparison with known cDNA and/or amino acid sequence should allow pinpointing of the loopback. The vector used in the making of the library was pAT153/PvuII/8 (Anson et al., 1984). The library contained more than $1x10^5$ transformants before amplification and more than 50% of the inserts were greater than 1 kb in size.

Approximately 30,000 colonies were plated directly onto 23x23 cm nitrocellulose filters, replicas taken, lysed and baked as described for cosmid screening (Section 2.8.2). The filters were hybridised using the A400 probe (Fig 3.1) for 24 hr, washed and autoradiographed. Nine positives were identified and rescreened (Section 2.8.2) again using the A400 probe and plasmids prepared from single positive colonies. (The initial screening was carried out by L. P. Chung).

The size of the inserts in the positive clones was determined by agarose gel electrophoresis of BamH I/HindIII digested plasmids (BamH I and HindIII flank the insert close to the cloning site in the vector).

The HinfI and AvaII digestion patterns were ascertained by polyacrylamide gel electrophoresis of fill-in labelled plasmid digests.

The clones fell into three size categories, the largest of which pBP62 and pBP51 were approximately 1.7 kb and showed HinfI/AvaII digestion profiles identical to those seen in the 3' region of pBP2. To confirm the exact positions of the inserts with respect to the complete cDNA sequence the ends of the inserts were fill-in labelled at the BamH I/HindIII sites then cut with HinfI or AvaII to produce fragments labelled at one end only. The fragments were then gel purified (Section 2.7.1) and used for Maxam and Gilbert sequencing (Section 2.10.1).

The preliminary sequencing results are shown in Fig 3.2. Both clones start at approximately the same point in the 3' region of sequence but not in the poly A^+ tail. This is a feature seen in other C4bp cDNA clones and is attributed to premature termination during synthesis of the second strand with DNA Polymerase I (Klenow fragment). pBP62 has an internal BamH I site at the 5' end followed by unidentified sequence. pBP51 has 5' sequence coding for residue 91 (serine) through to residue 48 (cysteine), the sequence is inverted and on the opposite strand suggesting a loopback has occurred during cDNA synthesis. pBP51 was selected for further mapping and sequencing.

3.3. Characterisation and sequencing of pBP51

Further characterisation of pBP51 involved mapping with combinations of the following enzymes PvuII, KpnI (Asp 718), BglI, DdeI, HinfI, BamH I and ClaI. Construction of a restriction enzyme map and comparison with the known cDNA map are given in Fig 3.3. The loopback was positioned approximately 500 bp from the BamH I vector site and the mapping confirmed that the sequence should extend into the putative leader sequence.

A BglI/HinfI fragment was isolated using polyacrylamide gel electrophoresis. The sticky ends were previously filled-in using cold nucleotides ($\frac{1}{10}^{th}$ of the total digest was labelled with α^{32}P[dATP] to allow identification on the gel). The fragment was eluted overnight, and then extensively purified by ethanol precipitation (Section 2.7.1). It was then cloned into the Hinc site of M13 mp8 and single stranded DNA isolated from twenty white plaques. The Hinf/Bgl band co-purified with a Hinf fragment from the 3' region of the insert, thus four populations of clones were identified using T-tracking analysis. One representative from each population was sequenced using the dideoxy termination method (Section 2.10.2). The sequence for both orientations of the HinfI/BglI fragment was obtained and the amino acid sequence determined for all six reading frames using computer analysis. The sequence for the fragment is shown in Fig 3.4.

Clone pBP51 covers approximately four-fifths of the total coding sequence for C4bp, the putative leader sequence and some of the 5' non-translated region. The 5' region of 500 bp is inverted and on the opposite strand due to the 'loopback' introduced during cDNA synthesis. This did not cause any ambiguity because of the availability of a detailed restriction map for the known cDNA sequence and peptide data for the N-terminus region of the protein.

3.4 Northern blot analysis

Chung et al., 1985a, used the A500 probe to hybridise an RNA transfer blot, the RNA for which was isolated from human liver. This identified a 2.4-2.5 kb mRNA species coding for human C4bp, the majority of which was present in the 18S fraction. A smaller, very much fainter, band was noted at approximately 1.8 kb, the source of which was unknown.

Poly A^+ RNA isolated from U937, HepG2 and Raji cell-lines together with RNA from both γ interferon stimulated and unstimulated cultured endothelial cells was used to prepare transfer blots by D. Goundis and J. Ripoche (MRC Immunochemistry Unit). These blots were made available for hybridisation with C4bp cDNA probes. (Each track on the gels contained approx. 10 µg RNA.)

The nick-translated S1 probe (see Fig 4.1) was used to hybridise the filters as described for genomic DNA transfer blots (Section 2.9). The blots were autoradiographed for 4 days, no signals were seen in tracks originating from U937, endothelial cells, HepG2 or Raji cells but the probe did hybridise strongly to a 2.5 kb band in the track of liver RNA suggesting that the hybridisation and transfer procedures were not at fault. Again a small, fainter band at approximately 1.8 kb was detected in the liver track (Fig 3.5).

3.5 Discussion

The coding sequence for residue 1 (asparagine) through to residue 33 (cysteine) were sequenced from clone pBP51 and are confirmed by the N-terminus peptide sequence (Chung et al., 1985c). These are the first 33 amino acids in the first internal repeat of C4bp and include two cysteines and a proline residue which form part of the consensus sequence on which the short consensus repeat superfamily is based.

The sequence immediately 5' to the first codon in the coding region is the putative leader sequence. This region of the derived protein sequence shows a high proportion of hydrophobic residues with two universal translation initiation codons (ATG), one at position $^{-}13$ and one at $^{-}32$, the latter being preceded by four polar residues. Leader peptides show no well defined consensus sequence of amino acids but the presence of polar residues followed by a region of relatively high

hybrophobicity are well documented features of other leader sequences (Watson, 1984). The most 5' (ATG) codon (i.e. the one at $^{-}$32, closest to the CAP site) is considered to be the point of translation initiation. Thus the putative leader sequence is 32 residues in length for human C4bp. The extent of the 5' non-translated region can only be defined using S1 nuclease mapping and cannot be derived accurately in cDNA clones.

The hydrophobic nature of the leader sequence is vital for its role in anchoring the peptide in the membrane during translation and thus mediating secretion of the mature protein from the cell into the fluid phase (Mechanisms of secretion from cells are reviewed by Blobel, 1980). After completion of translation the leader peptide is cleaved from the secreted protein and is thus never detected or sequenced in the mature protein. As a result of this cleavage the leader peptide sequence is available only by derivation from the cDNA sequence.

The cDNA sequence for mouse C4bp has been completed (Kristensen et al., 1987a) and includes the putative leader sequence. Fig 3.6 shows a comparison of the leader sequences in human and mouse, they show 63% identity which is comparable with an overall identity of 61% throughout the sequences. The mouse sequence also has two translation initiation codons, one at -13 (as in human) and the other at -56 (as opposed to -32 in human) which would suggest that the mouse leader is 56 residues in length.

Sequence from clone pBP51 was used to prepare a cDNA probe specific for the 5' region for use in subsequent genomic studies.

The absence of hybridisation signals using the C4bp cDNA probes with RNA transfer blots, on which the RNA originates from various cell types, suggests that the C4bp gene is either not expressed or is expressed in very low levels (cf. the level of expression in liver) in these

cells. The results of these hybrisation experiments are summarised in Table 3.1 together with the results obtained by J. Ripoche using cDNA probes specific for human Factor H. In the absence of results using well characterised monoclonal antibodies to detect C4bp at the protein level with these cell lines the hybridisation results remain indicative although in no way conclusive.

The origin of the smaller 1.8 kb band seen on C4bp hybridised liver RNA transfer blots remains unknown. Truncated mRNA's of DAF (Caras et al., 1987), CR1 (Hourcade et al., 1987) and Factor H (Ripoche et al., 1987; Schwaeble et al., 1987) have been demonstrated by hybridisation of cDNA probes to RNA transfer blots and by cDNA sequencing. A correspondingly truncated form of the protein has been demonstrated only for Factor H, using Western blot analysis. The signal given by the 1.8 kb species hybridising to the C4bp cDNA probe is very much fainter than that of the major 2.5 kb species and no evidence of a short form of C4bp was reported by Chung et al., 1985a, during cDNA sequencing.

Much of the cDNA mapping and final sequencing was completed with the help of Andrew Lewin, a Biochemistry Part II student, and was presented at the 12th Complement Workshop, Chamonix, September 18th-21st, 1987.

Table 3.1

	Liver	U937	Raji	HepG2	Endothelial cells
Factor H	+	-	?	-	+
C4bp	+	-	-	-	-

A very weak signal was obtained when up to 100 µg Poly A^+ RNA originating from Raji cells was used to prepare the transfer blot (M. Fontaine per comm). No signal was obtained by J. Ripoche using 10 µg RNA on the blot in Fig 3.5. Factor H mRNA has also been identified in RNA originating from macrophages and monocytes, the expression of C4bp has not been investigated in these cells in this study.

Fig 3.1 - The cDNA probes isolated from plasmid clone pBP2. A400 covers coding for residues 175-306 (excluding residues 206-221, not present due to the loopback present at approximately residue 241 and marked in the diagram with $^{\int}$).

A500 covers coding for residues 307-470. Both probes were prepared by elution from acrylamide gels and were labelled by nick-translation to a specific activity of between $1\text{-}5\times10^{8}$ cpm/µg. The probes were initially a gift from L. P. Chung.

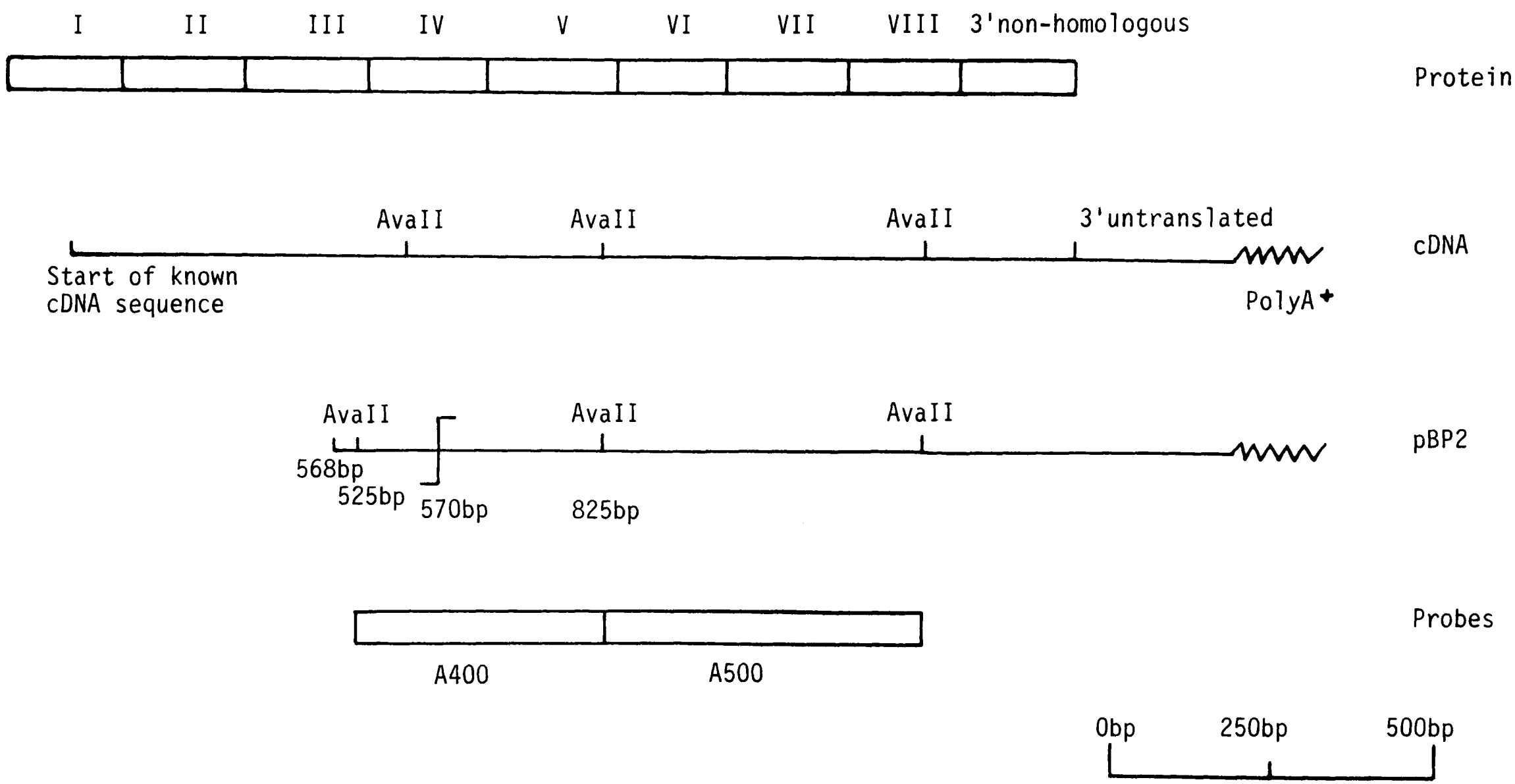
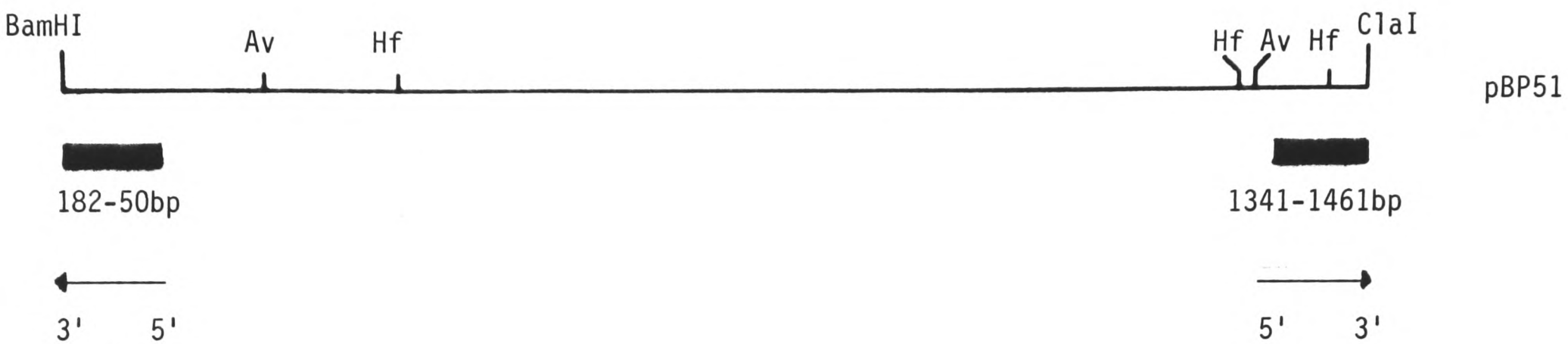

I
II
III
IV
V
VI
VII
VIII
3'non-homologous
Protein
AvaII
AvaII
AvaII
3'untranslated
cDNA
Start of known
cDNA sequence
PolyA+
AvaII
AvaII
AvaII
pBP2
568bp
525bp
570bp
825bp
Probes
A400
A500
0bp
250bp
500bp

Fig 3.2 - The preliminary mapping and sequencing results of the cDNA clones pBP51 and pBP62. The restriction sites in the vector BamH I and ClaI flank the insert DNA. The solid bars represent the fragments sequenced using the Maxam and Gilbert chemical degradation method, with the nucleotides covered numbered below. The arrows show the direction of the sense strands in the clones. The restriction sites in the inserts are marked - Av; AvaII, Hf; HinfI, M; BamH I.


BamHI
Av
Hf
Hf Av Hf ClaI
pBP51
182-50bp
1341-1461bp
3' 5'
5' 3'

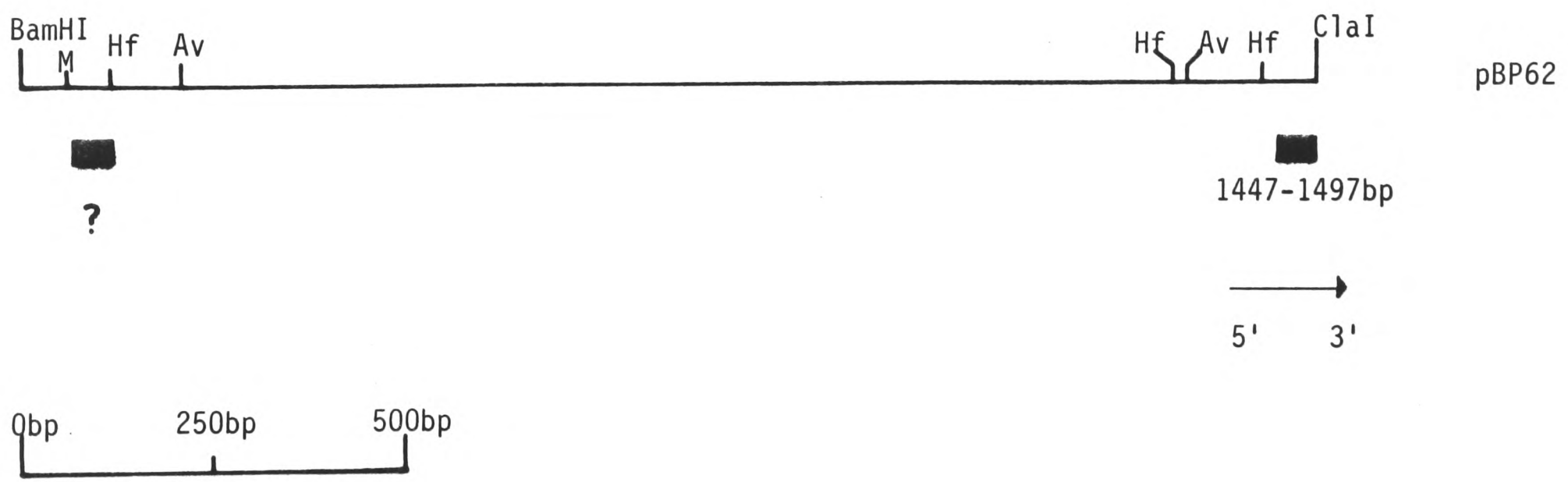
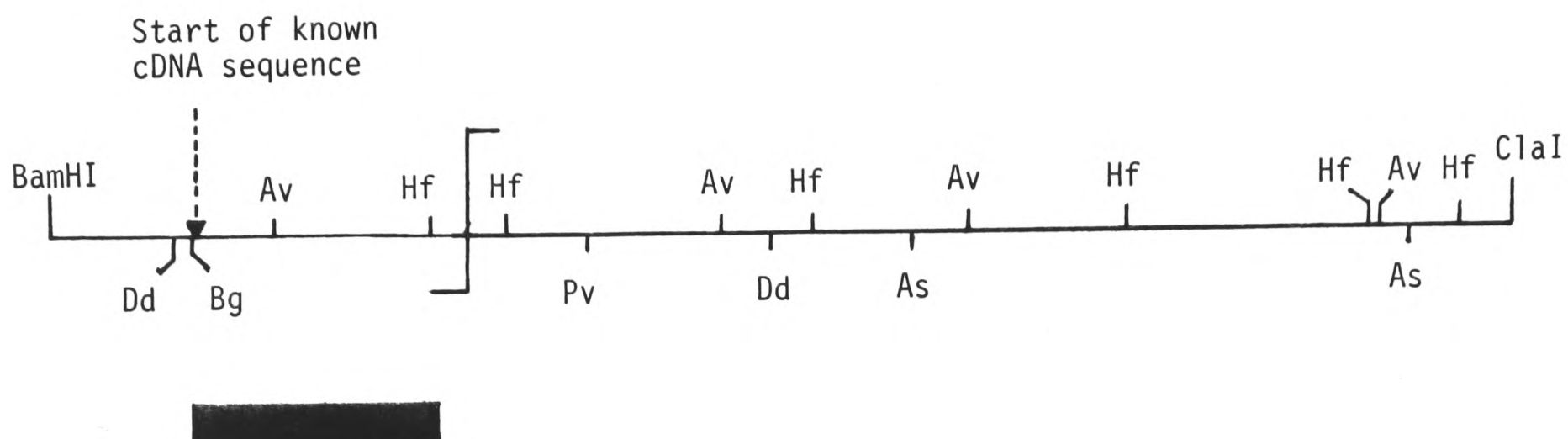
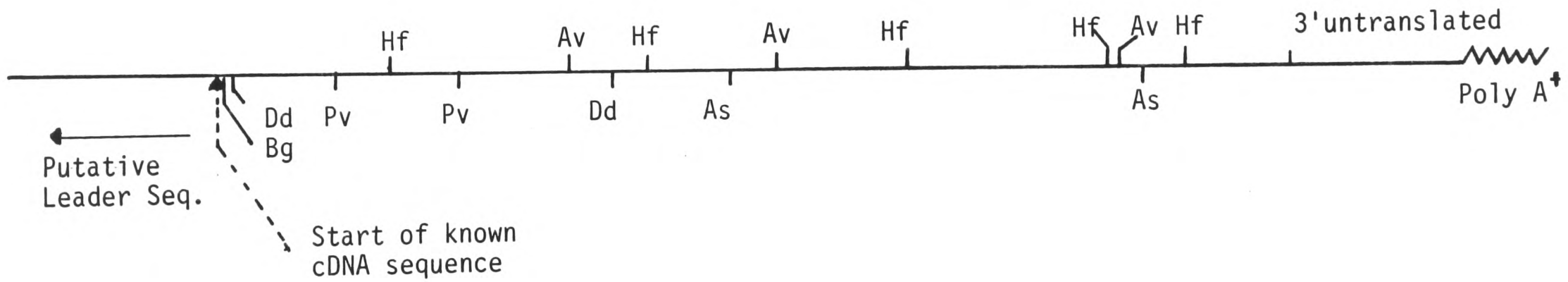

BamHI
M
Hf
Av
Hf Av Hf ClaI
pBP62
?
1447-1497bp
5' 3'
0bp
250bp
500bp

Fig 3.3 - Restriction map constructed for pBP51 compared with the known cDNA restriction map, the markings are as in Fig 3.2 As, Asp718; Bg, BglI; Dd, DdeI; Pv, PvuII. The loopback is marked ⌠ and occurs approximately 500 bp from the BamH I site. The solid bar shows the HinfI/BglI fragment purified for sequencing using the dideoxy termination method.


Start of known
cDNA sequence
BamHI
Av
Hf
Hf
Av
Hf
Av
Hf
Hf
Av
Hf
ClaI
Dd
Bg
Pv
Dd
As
As


Hf
Av
Hf
Av
Hf
Hf
Av
Hf
3'untranslated
Dd
Bg
Pv
Pv
Dd
As
As
Poly A+
Putative
Leader Seq.
Start of known
cDNA sequence

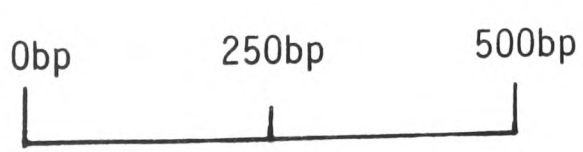

0bp
250bp
500bp

Fig 3.4 - cDNA sequence and derived amino acid sequence for the HinfI/BglI fragment. The sequence is inverted and on the opposite strand due to the presence of a loopback in clone pBP51. The underlined amino acid sequence corresponds to the N-terminus peptide sequence (Chung et al., 1985c) and covers residue 1 (asparagine) through to residue 33 (cysteine). The two universal translation initiation codons (ATG) are boxed and lie at positions -13 and -32. Sequence 5' to residue -40 is not shown in the diagram, the cDNA sequence was only obtained on one strand and appeared to contain cloning artefacts. This region is in the 5' non-translated region and is very close to the loopback artefact in the clone. The hydrophobic residues in the putative signal sequence are marked *.

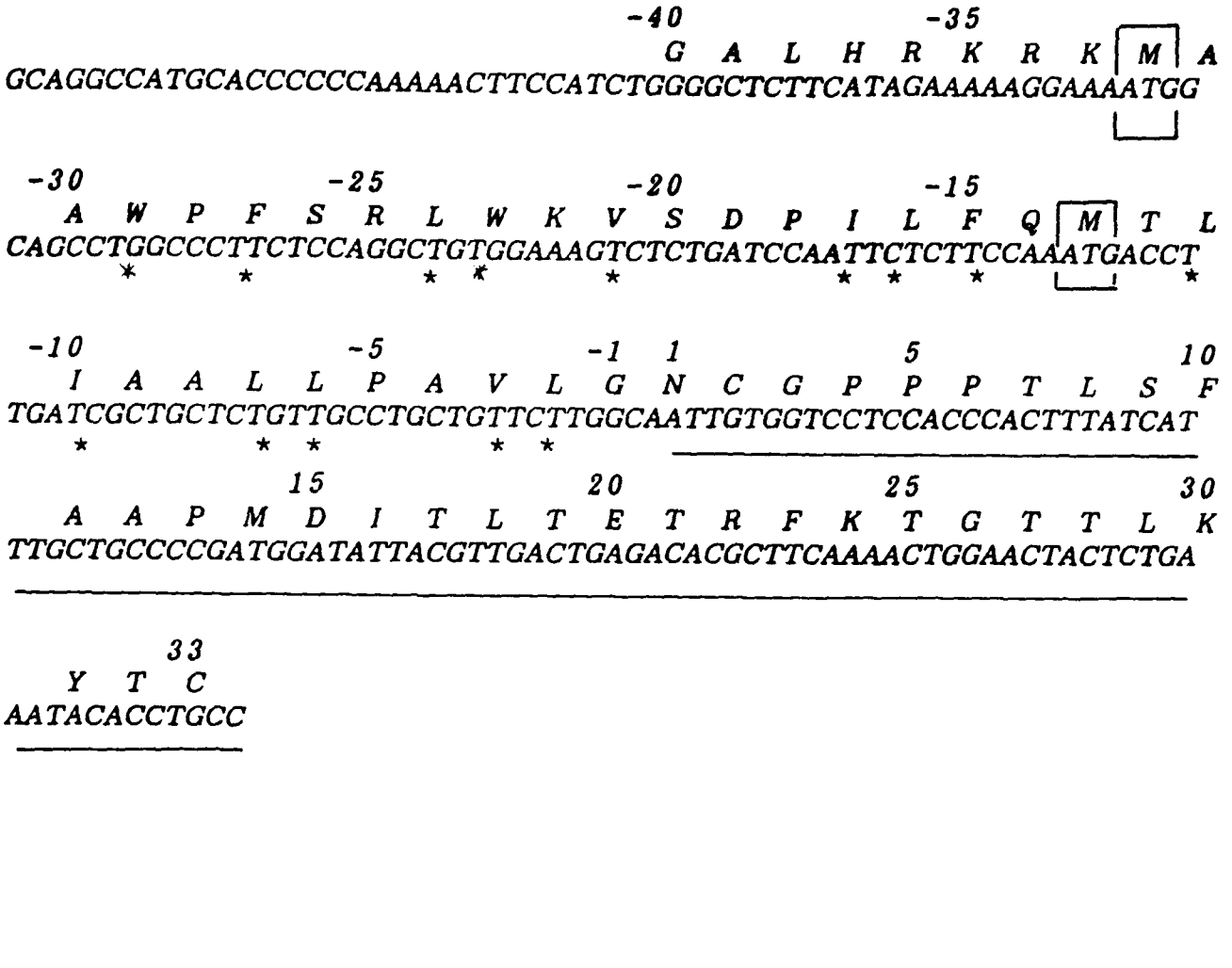

-40 -35
G A L H R K R K M A
GCAGGCCATGCACCCCCCAAAAACTTCCATCTGGGGCTCTTCATAGAAAAAGGAAAATGG
-30 -25 -20 -15
A W P F S R L W K V S D P I L F Q M T L
CAGCCTGGCCCTTCTCCAGGCTGTGGAAAGTCTCTGATCCAATTCTCTTCCAAATGACCT
-10 -5 -1 1 5 10
I A A L L P A V L G N C G P P P T L S F
TGATCGCTGCTCTGTTGCCTGCTGTTCTTGGCAATTGTGGTCCTCCACCCACTTTATCAT
15 20 25 30
A A P M D I T L T E T R F K T G T T L K
TTGCTGCCCCGATGGATATTACGTTGACTGAGACACGCTTCAAAACTGGAACTACTCTGA
33
Y T C
AATACACCTGCC

Fig 3.5 - RNA transfer blot hybridised with the S1 cDNA probe (Fig 4.1). The major mRNA species, 2.5 kb codes for the 70 kDa C4bp monomer and corresponds to the strongest signal in the track consisting of mRNA isolated from liver. The source of the fainter, 1.8-1.9 kb band in this track remains unknown. The other tracks on this blot correspond to RNA isolated from HepG2 cells, U937 and Raji cells, no signals could be detected even when the autoradiograph was left for 10 nights at -70°C.

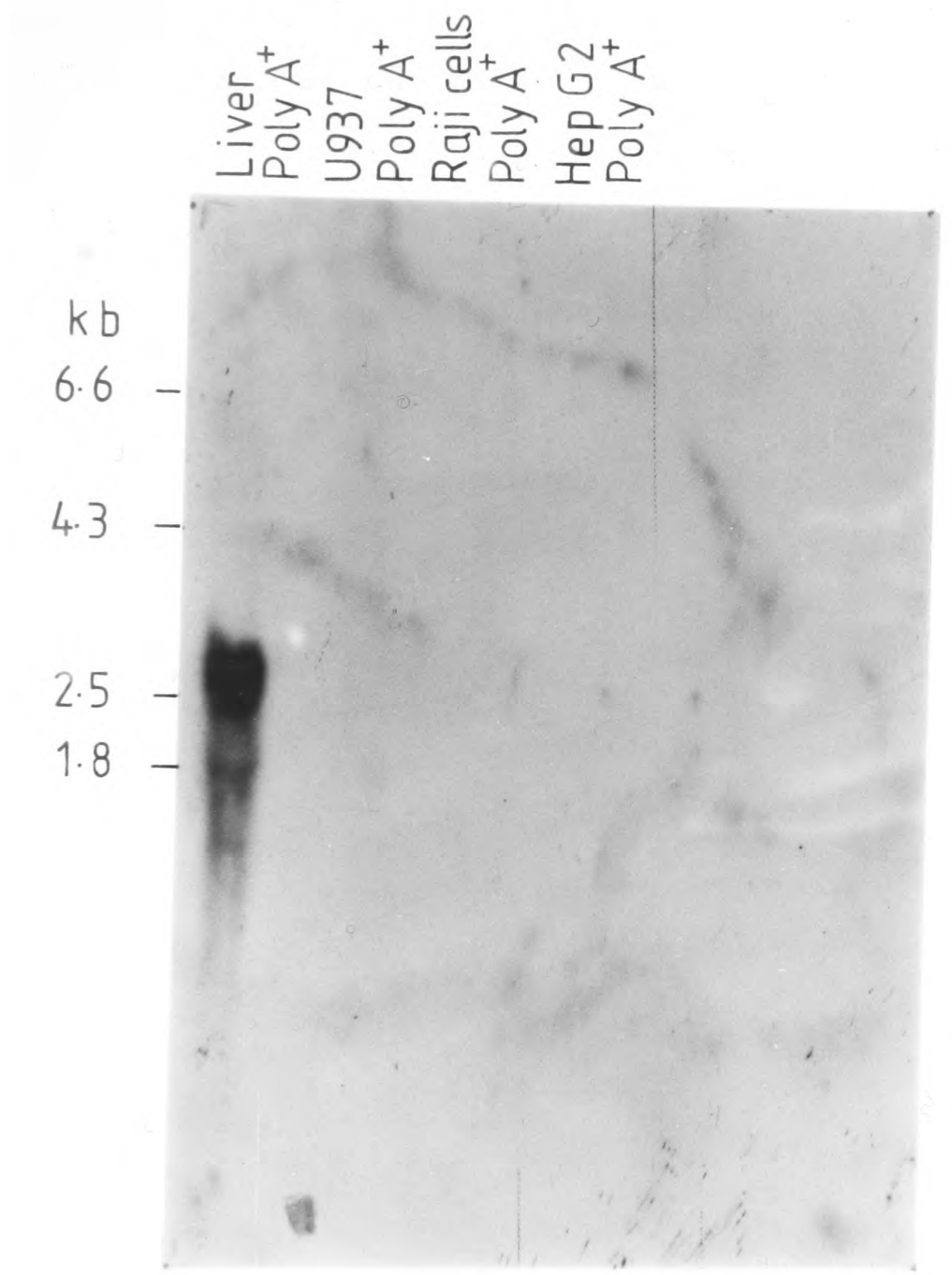
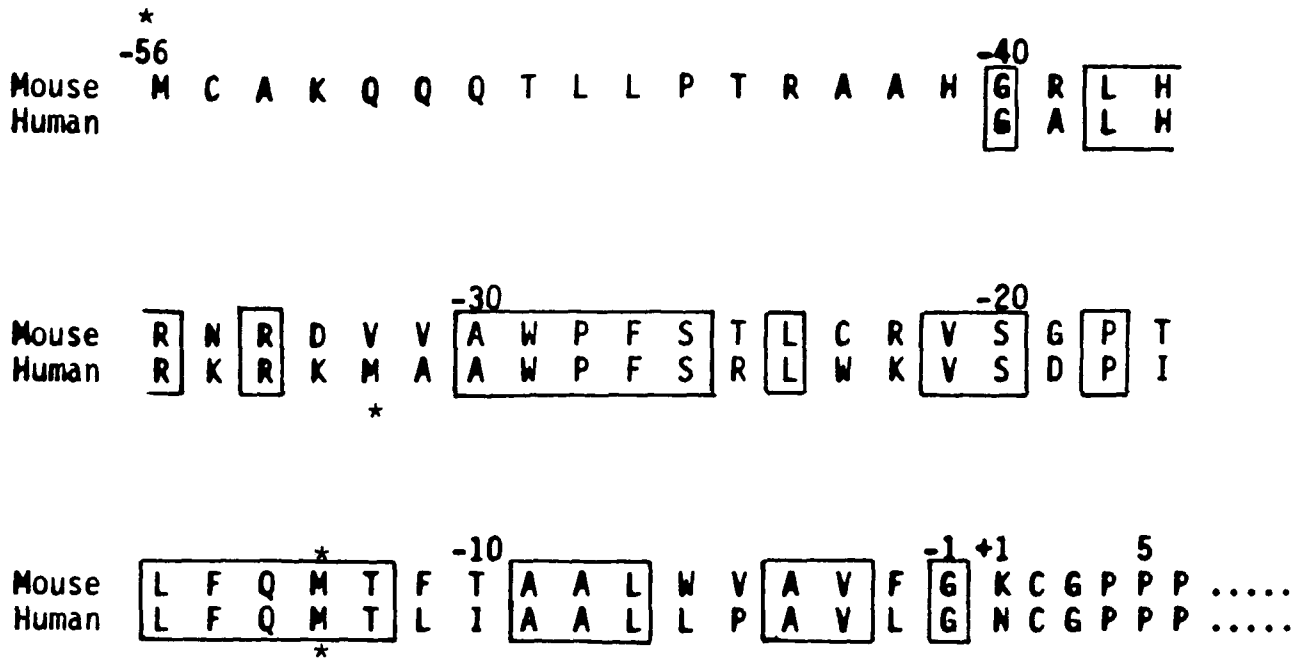

Liver
Poly A+
U937
Poly A+
Raji cells
Poly A+
Hep G2
Poly A+
kb
6·6
4·3
2·5
1·8

Fig 3.6 - Comparison of the mouse and human putative leader sequences. The residues which show identity are boxed and the initiation codons marked with *. The sequences show 63% identity and both have two initiation codons.

Homology between human and mouse C4bp leader peptides



C H A P T E R I V

ISOLATION OF GENOMIC CLONES FOR HUMAN C4bp

4.1 Introduction

The major aim of this project was to clone the human C4bp gene to examine any correlation between the internal homology units, described in the primary amino acid sequence, with the exon organisation. The discovery of homology units, conforming to the same consensus sequence, in other complement and non-complement proteins lead Chung et al., 1985a, to propose a multigene family of proteins containing the repeat unit. These proteins are thought to have evolved by tandem duplication of a primordial gene segment. This theory was supported by the demonstration, in the Factor B gene, of the intron junctions coinciding with repeat boundaries i.e. one repeat unit was encoded by one exon.

Genetic linkage studies by Rodriguez de Cordoba et al., 1984, showed that the C4bp and CR1 loci were closely linked and preliminary results suggested that a third control protein gene, Factor H, was also linked to the C4bp/CR1 loci. This had allowed the proposal of a linkage group of control proteins analogous to the Class III linkage group on chromosome 6 (Rodriguez de Cordoba et al., 1985). Cloning of the Class III genes in an overlapping cluster of cosmid clones covering 98 kb had allowed definition of the order of genes and the physical distances separating them (Carroll et al., 1984).

At this time work was in progress to isolate a cDNA clone coding for human Factor H and it was felt that a cDNA probe would be available in the near future. It was thus decided to attempt to isolate cosmid

clones containing the human C4bp gene with a view to linking the C4bp and Factor H genes using overlapping cosmid clones.

4.2 Cosmid libraries, plating and screening

Cosmid vectors are used in the cloning of large fragments of eukaryotic genomic DNA and are derived directly from plasmids. They contain the cohesive ends ('cos sites') of λ phage, and as recognition of this sequence is the only pre-requisite for the activity of λ packaging enzymes, cosmids can be packaged into phage protein heads *in vitro*. Once packaged the protein heads are used to infect *E. coli* host cells following which the cosmid DNA is released into the cytoplasm where it circularises and exists as super-coiled extrachromosomal DNA (Steinmetz *et al.*, 1985).

Cosmids also contain an origin of replication and selective marker gene hence recombinants can be selected for by growing on appropriate antibiotic-containing media and multiple copies of the cosmid plus insert DNA can be purified from one host cell. There is usually a cloning site or multiple cloning sites into which the insert can be ligated. Cosmid vectors are replicated by the host cell enzymes, do not cause lysis of the host and are not packaged into protein coats after the initial infection and therefore do not need to include genes necessary for these activities. As a result the vector size is kept to a minimum, most cosmids being approximately 5 kb in size. *In vitro* packaging is efficient with linear DNA up to 50 kb in size and hence inserts of up to 45 kb can be cloned using these vectors.

Donor DNA is randomly sheared or digested prior to cloning. This was achieved by partial digestion of genomic DNA with restriction enzymes during construction of the libraries used in this study. After digestion the DNA was size fractionated and only DNA between 35-45 kb

used for ligation. This prevents small fragments ligating into the same vector and hence stops the linking of non-contiguous segments of the genome.

Once the ligation, packaging and infection of the host cells is complete the library can be plated out as described in Section 2.8.2. At this stage the library may or may not be amplified, this involves allowing recombinants to grow on plates to form small colonies which are then scraped off into Hogness freezing broth and stored at -70°C. The frozen stocks thus contain more than one copy of each original recombinant and aliquots can be plated out at a later date (Steinmetz et al., 1985).

The first cosmid library screened was a gift from Dr. F. Grosveld and was the HPFH cosmid library described by Grosveld et al., 1982. Approximately 400,000 colonies from the library were plated directly onto nitrocellulose filters (23x23 cm) at a density of approx. 100,000/plate. Replicas were taken and hybridised etc as described in Section 2.8.2. using the A400 probe (Fig 4.1). Following autoradiography the probe appeared to have hybridised to every colony on the filter giving an unworkable background, any positive signals were completely obscured. After stripping, the filters were re-hybridised with the A400 probe (freshly nick-translated) using hybridisation buffer containing higher concentration of competitor DNA's. This did not improve the background.

A further set of replicas were taken from the master filters and treated as described in Section 2.8.2, paying particular attention to the removal of bacterial debris after lysis. These filters were hybridised with A500 probe, the background problem was again encountered.

This same library had been used by Carroll et al., 1984, to isolate the overlapping cosmid clones used in the linkage of the Class III genes and no such problems had been encountered.

Other libraries screened at this time included two constructed by Dr. R. D. Campbell and one by Dr. A. Palsdottir. All three used the PTCF cosmid vector, the genomic DNA was prepared by partial digestion with MboI and the libraries were all amplified. The high background was a problem throughout although some possible positive signals from one of Dr. Campbell's libraries were rescreened. The signal on the rescreens were indistinguishable from the background.

The A400/A500 probes were prepared from AvaII digests of pBP2 (Chung et al., 1985a). The two fragments originated from the middle of the insert and thus contained no vector sequence. The fragments were purified by elution from a 4% polyacrylamide gel on which the probe fragments were well separated from pAT/AvaII fragments. No smearing of bands could be detected on the autoradiograph suggesting that no degradation had taken place during the digest. It was however decided to prepare a new probe - S1 (Fig 4.1).

This probe consisted of the entire insert from pBP6 (Chung et al., 1985a) excised from the vector using the BamH I/Hind III sites in the vector (this leaves 3bp and 8 bp of vector sequence at each end of the probe which was not thought to be enough to hybridise at the stringencies used in these experiments). The insert (1.54 kb) and the vector (3.3 kb) were well separated on an agarose gel before eluting the insert DNA onto DEAE paper, and by ethidium bromide staining no smearing between the bands was detected.

The S1 probe was used to screen one of Dr. Campbell's cosmid libraries again, together with a set of replica filters from Carole Sargeant which were from an unamplified library made from MboI partially

digested genomic DNA in PDV-cos vector. The background on both sets of filters showed some improvement but was still higher than that seen with probes generated from the Class III region. Again possible positive colonies from the PDV-cos library were picked from the master plates but gave no strong signals in the rescreen.

Recently the Nco and 5' probes were each used to hybridise one cosmid library filter. The two filters used were a pair of replicas from the same master and hybridised with approximately the same number of counts/ml hybridisation buffer. Both probes were labelled to approximately the same specific activity. As can be seen in Fig 4.2, the 5' probe gave lower background than the Nco probe. This could prove useful in future attempts to isolate cosmids.

4.3 λ-library screening and isolation of genomic clones

λ replacement vectors are used in the construction of genomic libraries and like cosmid vectors the foreign DNA is ligated between 'cos-sites'. This again utilizes the *in vitro* packaging enzymes to produce a viable infective particle which transfers the cloned DNA into host cells. Unlike the cosmid vectors the amplification of the recombinant phage relies on the ability of the phage to replicate and package new infective particles prior to host cell lysis and plaque formation. This necessitates the inclusion of 50% of the phage genome (30 kb) in the vector leaving room for 15-20 kb of insert DNA. Thus the major drawback in the use of these vectors lies in the reduced length of the cloned inserts which leads to an increase in the number of overlapping clones which must be isolated. It was hoped that any background problems encountered during the cosmid library screening with C4bp cDNA probes would not recur when using λ vectors.

The λ genomic library used to isolate all the genomic clones covering the human C4bp gene was constructed by Dr. A. Bentley using the λ-replacement vector EMBL-3 (Frischauf et al., 1983). EMBL-3 is derived from λ1059 (Karn et al, 1980) into which a multiple cloning site has been inserted between the λ arms (which contain the essential genetic elements for maintenance of the phage life cycle). The foreign DNA is ligated into the cloning site and as a result a single restriction enzyme digest releases the insert from the vector. Genomic DNA was prepared by partial digestion with MboI and the fragments ligated into the BamH I cloning site in the vector. One library complexity was represented by approximately 2.2×10^6 plaques.

In the original screening experiment approximately 1×10^6 plaques were screened. The phage were plated at a density of 1×10^5 plaques per 14 cm plate using Q359 as the host strain. Q359 is able to select for recombinant constructions and is thus used in the initial screening. (The plaques are very small with this host strain and it is replaced by Q358 a non-selective host, for the rescreens.) The filters were prepared and hybridised (as described in Section 2.8.3) with the A400 probe, as shown in Fig 4.1. Sixteen possible positive clones were isolated from the master plate and rescreened.

Nine clones rescreened which could be divided into two groups - strong signals from λBP 3, 4, 5, 10, 11 - weak signals from λBP, 1. 2, 6 and 7. These were all further plaque purified until every plaque on an 8 cm plate corresponded to a positive signal. This final set of replica filters were washed with 0.5 M NaOH/1.5 M NaCl and neutralised with 0.5 M Tris, pH 7.4 to remove the A400 probe and rehybridised with the A500 probe. Signals with the A500 probe were seen with clones λBP 3, 4, 5, 10, 11.

A single plaque from all nine plates was picked and stored at 4°C as a stock from which lysates were prepared for the isolation and characterisation of DNA (Chapter V).

Subsequently DNA from λBP10 was fully characterised and shown to contain coding sequence for only approximately one-third of the C4bp mRNA. New cDNA probes were prepared which were specific for the regions of coding sequence not represented in the fully characterised genomic clone. These probes, Nco and 5' probe (Fig.4.1) were used to screen the same genomic library again.

Approximately $2x10^6$ plaques were plated onto twenty 14 cm plates and nitrocellulose replica filters made, ten of which were hybridised with the Nco probe and the remainder with the 5' probe.

The Nco probe gave nine strong signals and ten weaker signals of which eighteen rescreened → λBP3'.

The 5' probe gave one strong signal and nine weaker signals of which five rescreened → λBP5'.

The rescreens were carried out using the Nco and 5' probes and the clones plaque purified until all plaques corresponded to positive signals. Multiple nitrocellulose replica filters were taken at this stage and hybridised with other cDNA probes to attempt to define the regions of coding in each clone.

All the 3' clones hybridised with the A500, Nco and 3' probes and none hybridised to the A400 probe. None of the λBP5' clones hybridised to any other cDNA probes. A single plaque from each plate was picked and stored at 4°C.

4.4. Initial characterisation of new genomic clones

The distribution of enzyme sites in the 3' region of the cDNA made the preparation of informative cDNA probes relatively straightforward

(Fig 4.1). It was possible to predict from the hybridisation of λBP3' clones to the cDNA probes that the majority of the missing 3' coding sequence should be represented in the clones. No such convenient sites were available in the 5' region so it was decided to try using unique 17-long oligonucleotides from specific repeat units and the signal peptide to define the extent of coding sequence present in λBP5' clones. The sequences used in this study are shown in Fig 4.3.

The oligos were kindly synthesised by Tony Day using an Applied Biosystems automated synthesiser. This uses chemically modified nucleotides during the multi-step synthesis which are known as phosphoramidites and hence this synthesiser is said to use the phosphoramidite method of synthesis.

After deprotection and cleavage from the solid silica support on which the oligo is synthesised, the oligo is in a solution of concentrated ammonium hydroxide in which form it can be stored at 4°C. Before use an aliquot of the crude oligo solution is purified by ethanol precipitation and labelled using T4 kinase with γ^{32}P[dATP]. The conditions for the labelling reaction were optimised by titrating the amount of oligo needed to saturate a standard kinase labelling reaction (Section 2.5.4). The results are shown in Fig 4.4 - subsequent labelling reactions used 1 µl of purified oligo solution.

The oligos were used to screen replicas from the final rescreen of λBP5' clones - one clone hybridised to C4bp-signal (Fig 4.5a) but none hybridised to C4bp-II. This clone λBP5' 12 thus contained coding sequence for the signal peptide but none of the λBP5' contained coding sequence for the second internal repeat.

C4bp-II was then used to rescreen the original λBP clones isolated with the A400 probe - λBP3 hybridised to C4bp-II, which suggests it contains coding for the second repeat (Fig 4.5b).

Use of the oligo's therefore defined two 5' clones, λBP 5'12 which includes sequence coding for the signal peptide and λBP3 which includes sequence coding for the second internal repeat, hopefully one or both include sequence coding for the first repeat.

4.5 Discussion

The combined use of cDNA probes and oligonucleotides has thus permitted the isolation of four λ-genomic clones which appear to include the coding sequence for the human complement control protein, C4bp. The clones and probes to which they hybridise are summarised in Fig 4.6.

Attempts to isolate cosmid clones encoding the C4bp genomic sequence have, so far, been unsuccessful due in part to the background problem encountered with the cDNA probes. The use of different techniques during cDNA probe preparation has attempted to rule out the possibility of vector contamination. The use of probes on filters previously hybridised with Class III cDNA and genomic probes with no problems also excludes any errors during replica filter preparation. The C4bp cDNA probes have been used successfully to hybridise genomic blots and have given clean well defined bands with no background smearing in the tracks. This suggests the absence of repetitive sequence, such as Alu repeats in the coding sequence, a situation which has been reported in the cDNA sequence of one of the complement control proteins, DAF (Caras et al., 1987).

The cDNA probes do give some background when used to screen plasmid vector libraries, however in this case the positive signals are readily distinguishable. This would perhaps implicate the selective marker gene or the origin of replication, (common to all plasmid and cosmid vectors), as a possible source of homology with the probe sequence. Similar background problems have been found during the screening of

cosmid libraries with cDNA probes specific for human Factor H (Bob Sim, personal communication) suggesting that this is a problem common to the region of the genome containing the complement control proteins. Subsequent use of the same factor H probes on λ-library filters gave no background and has allowed the successful isolation of genomic clones for human Factor H.

When using the S1 probe to screen cosmid libraries the background showed some improvement and possible positives were identified which did not rescreen. These clones were isolated from the unamplified PDV-cos library suggesting that possibly part of the inserts including the C4bp gene sequence inhibit growth of the host and cause loss of the clone during amplification. This would suggest that screening unamplified cosmid libraries in the future may prove more fruitful.

Genomic DNA fragments, used in ligations during the construction of all the libraries screened in the isolation of human C4bp genomic clones, were prepared by partial digestion with the restriction enzyme MboI. It may be that there is a high density of MboI recognition sequences within the C4bp gene which are very susceptible to cleavage and even in partial digests give rise to small fragments not included in the fractions used for cosmid ligations. As a result the C4bp gene fragments are represented in lambda libraries but not cosmid libraries constructed in this way. It may be more productive to screen cosmid libraries made from genomic DNA prepared by digestion with BamH I or Sau3A, or by physical shearing.

Anecdotal reports of other sequences that cannot be found in libraries of eukaryotic genomes have been made, sequences that include the hypervariable region 3' to the α globin locus (Wyman et al., 1985) and the junction fragment from a translocation between chromosome 6 and 12 in a murine plasma cytoma (Perlmutter et al., 1984).

Wyman et al., 1985, report the absence of some random human DNA fragments from λ-genomic libraries grown on rec$^+$ E. coli host strains commonly used in the propagation of recombinant DNA libraries. They suggest that this may be due to the recognition of secondary structures in the insert DNA by products of the host recB, recC and sbcB genes. The recognition then directly or indirectly affects the representation of the insert sequences in libraries using these host strains. In support of this theory was the identification of clones from host strains carrying recB, recC and sbcB mutations which could not subsequently be propagated in rec$^+$ host strains. The inserts of these clones were then analysed for the presence of secondary structures using electron microscopy. This showed the presence of extensive simple snap-back and stem-and-loop structures. This is in agreement with the observation of Leach and Stahl, 1983, that phage which have been engineered to contain perfect palindromic inverted repeats, which are capable of forming cruciform secondary structure, were not viable in rec$^+$ host strains. The same engineered phage constructions could be propagated in recB, recC, sbcB mutant hosts. This report favours the direct nucleolytic action of exonuclease V and exonuclease I (the gene products of the recB,C and sbcB loci respectively) on cruciform structures as an explanation for the inviability of these phage in rec$^+$ hosts.

Although these reports are confined to sequences missing from λ libraries it could be used to explain the absence of sequences from cosmid libraries. If the sequences causing the secondary structures were separated by the increased extent of digestion during the preparation of smaller inserts for lambda cloning the same sequences could be propagated in lambda but not cosmid vectors.

E. coli K-12 derived *in vitro* packaging systems were used in the preparation of both the cosmid and lambda libraries used in this study. It is therefore unlikely that the problems of biasing against insert sequences containing the heptameric EcoK restriction site reported by Rosenberg, 1985, apply in this study.

Most of the problems encountered in attempts to isolate cosmid clones with C4bp cDNA probes are attributable to the background problems and possibly use of the 5' probe will alleviate this. Use of unamplified libraries constructed from genomic DNA partially digested with BamH I or Sau3A may prove more fruitful in the future.

Fig 4.1 - cDNA probes used in the screening of cosmid libraries, isolation and characterisation of λ clones.

A400 - Covers coding for residues 175-307 (excluding 206-221 not present because of a 'loop-back' in the clone) which includes the protein repeats III, IV and V. A500 covers coding for residues 307-470 which includes protein repeats V, VI, VII and VIII. These probes were prepared by AvaII digestion of pBP2 and isolated from polyacrylamide gels.

S1 - insert from pBP6 and covers coding for residues 32-530 which includes protein repeats I-VIII and 3' non-homologous region was prepared by BamH I/HindIII digestion and was purified from agarose gels by elution onto DEAE paper.

Nco - covers coding for residues 385-530 (rpts VII, VIII and 3' non-homologous region), was prepared by NcoI/HindIII digestion of clone pBP6 and was isolated from DE52 purified low gelling temperature agarose (Sigma type IV).

3' - covers coding for residues 483-530 (rpts VIII and 3' non-homologous region) was prepared by Asp718/HindIII digestion of pBP6 and was purified from polyacrylamide gels.

5' - covers coding for residues 1-72 (rpts I and II), the signal peptide and 5' non-translated, it was prepared by BamH I/HinfI digestion of clone pBP51 and was purified from polyacrylamide gels.

All probes were labelled using nick translation or primer extension to a specific activity of $2\text{-}8\times10^{8}$ cpm/μg (Section 2.5).

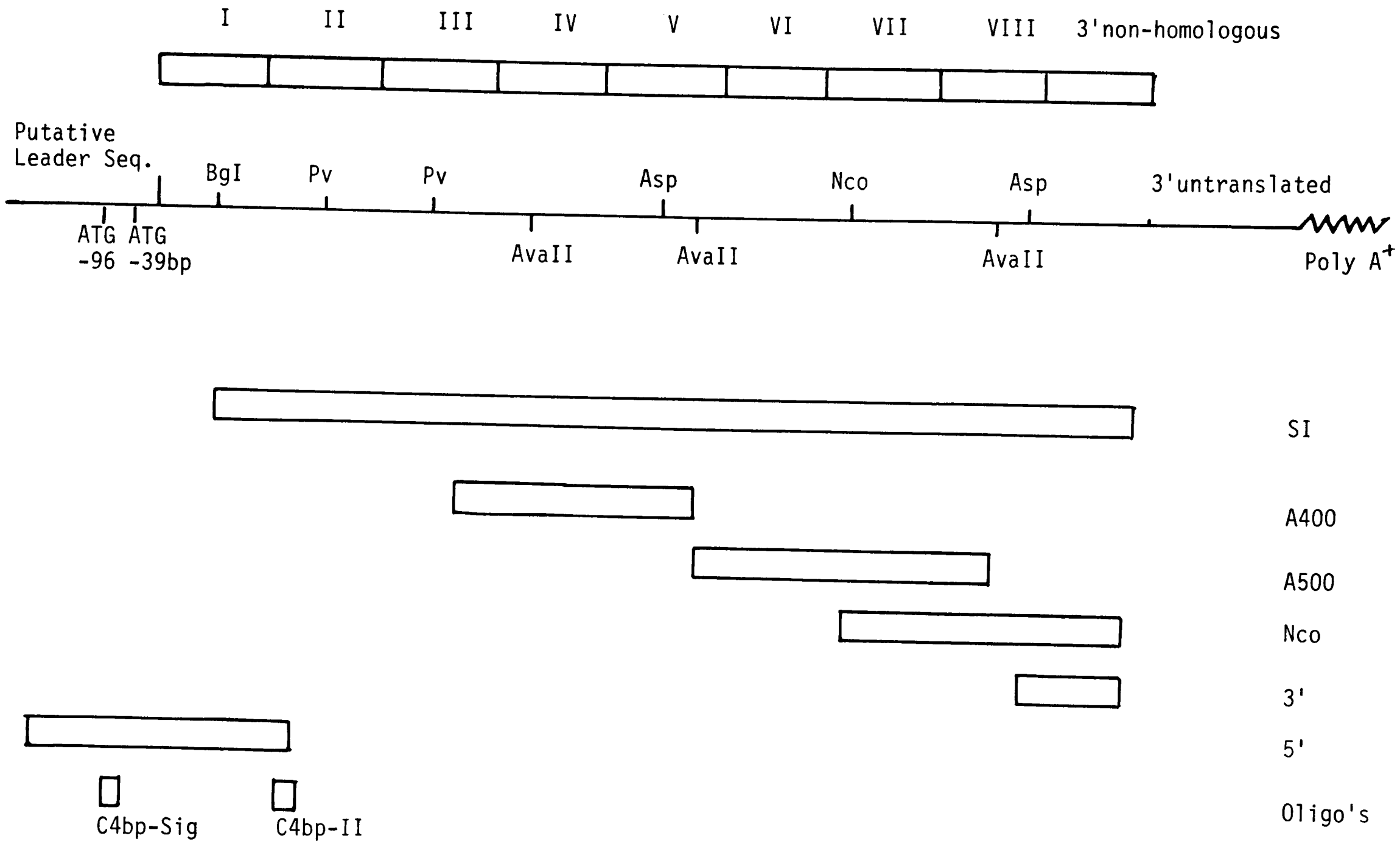

I
II
III
IV
V
VI
VII
VIII
3'non-homologous
Putative
Leader Seq.
BgI
Pv
Pv
Asp
Nco
Asp
3'untranslated
ATG
-96
ATG
-39bp
AvaII
AvaII
AvaII
Poly A+
SI
A400
A500
Nco
3'
5'
Oligo's
C4bp-Sig
C4bp-II
0bp
250bp
500bp

Fig 4.2 - Replica cosmid filters from the same master filter hybridised with the two most recently prepared cDNA probes.

(a) Nco probe
(b) 5' probe

Both probes were labelled to a specific activity of approx. 2×10^8 cpm/µg

Both filters were hybridised in 10 ml prehybridisation buffer with 4×10^5 cpm/ml overnight at 42°C and washed -

3x20' 2xSSC and 0.1% SDS at room temperature

3x20' 0.2xSSC at 68°C

and autoradiographed overnight.

Fig 4.2a

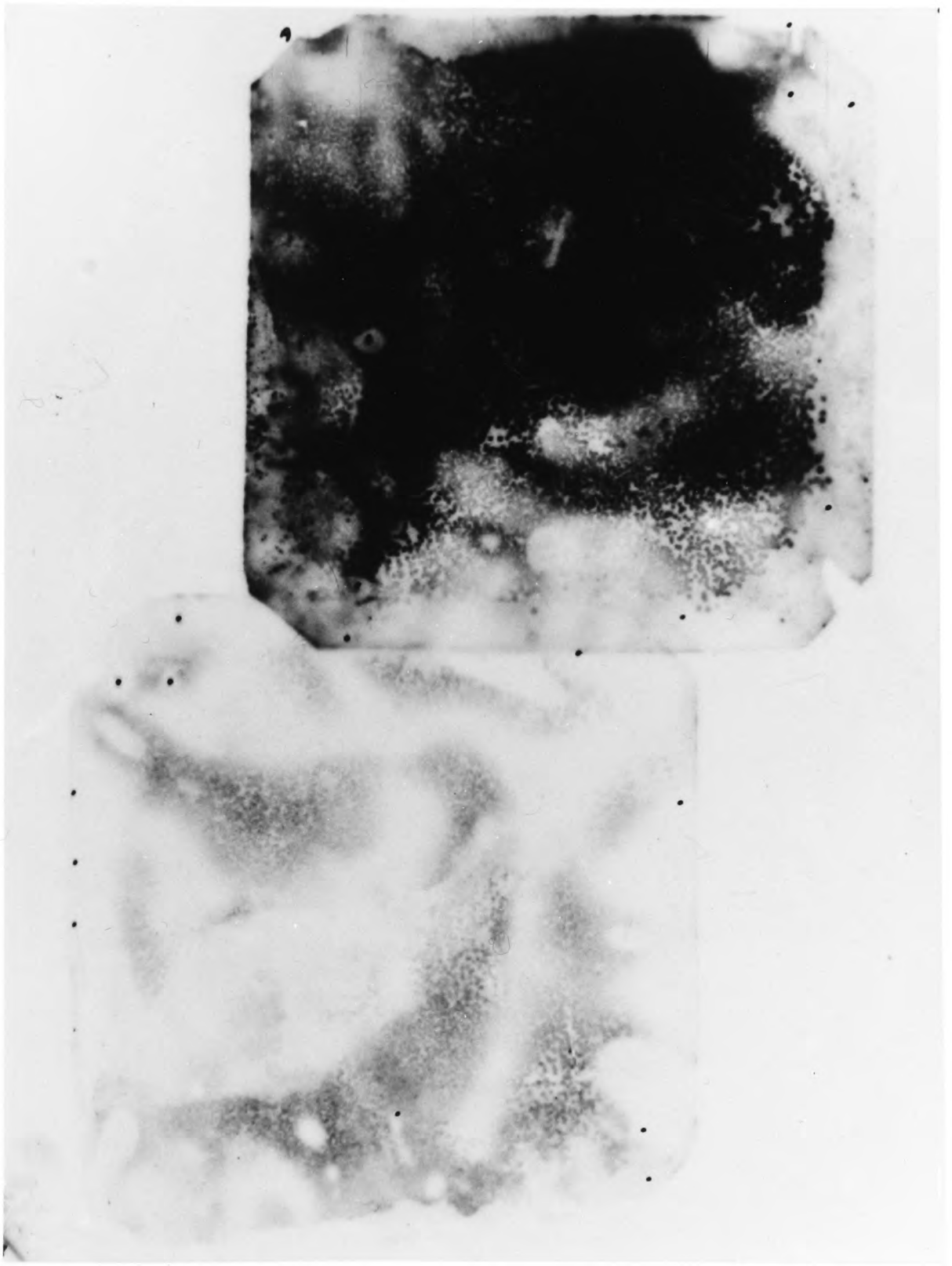


Fig 4.2b

Fig 4.3 - The sequences of the two 17-long oligonucleotides synthesised.

C4bp-Sig covers residues -12 to -16 in the signal peptide (non-sense strand).

C4bp-II covers residues 66 to 70 in the second repeat (sense strand).

C4bp-Sig

A	GGT	CAT	TTG	GAA	GAG	A
Leu	Thr	Met	Gln	Phe	Leu	Iso
-11	-12	-13	-14	-15	-16	-17

C4bp-II

C	AGA	CAC	CCA	GGA	GAG	T
Cys	Arg	His	Pro	Gly	Glu	Leu
65	66	67	68	69	70	71

Fig 4.4 - 20% acrylamide gel showing the result of the titration of the labelling reactions for the two oligonucleotides. The different amounts of oligo were labelled with 2 µl γ^{32}P[dATP]: 1 µl T4 kinase enzyme: 1 µl x 10 kinase buffer in a 10 µl reaction incubated at 37°C for 30 mins. Subsequent labelling reactions used 1 µl of undiluted purified oligo in a 20 µl labelling reaction with 6 µl γ^{32}P[dATP] (Section 2.5.4).

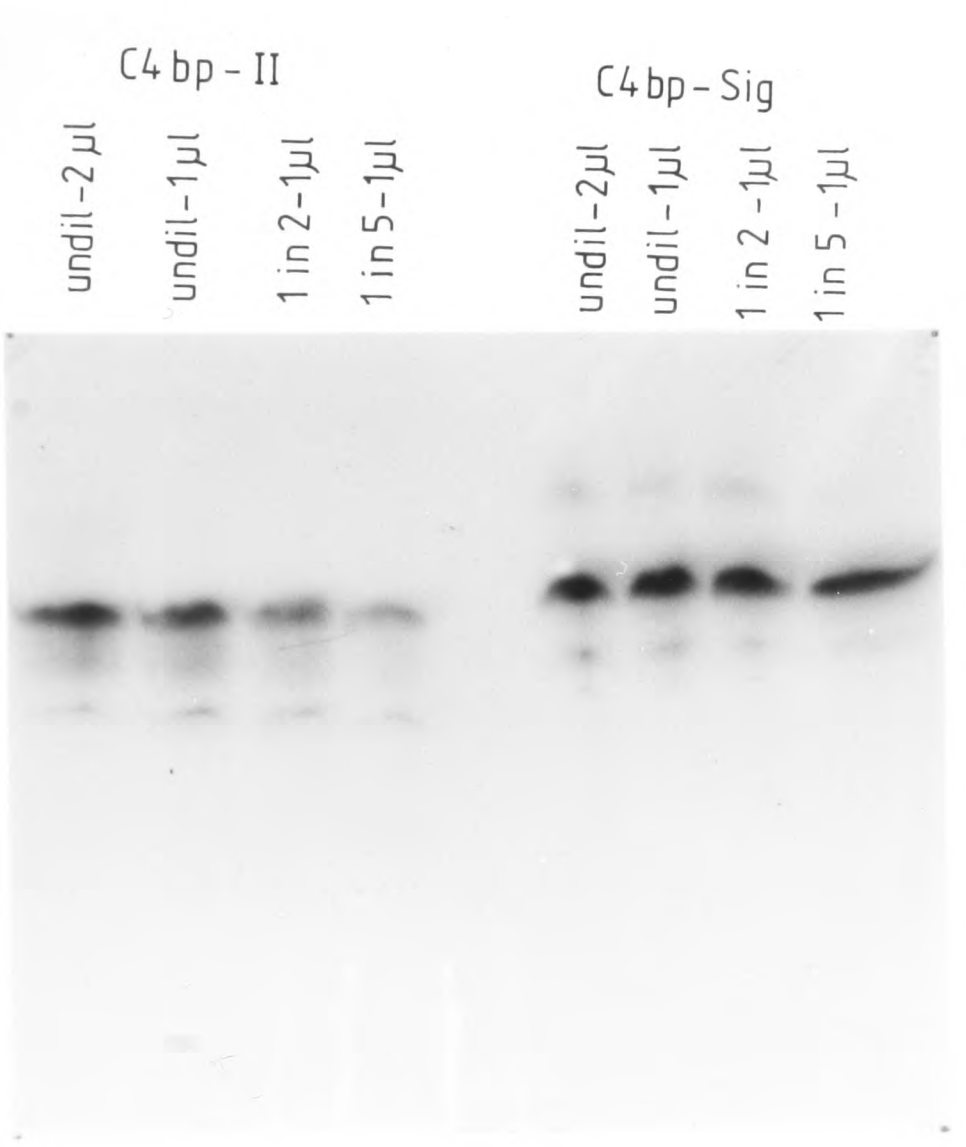
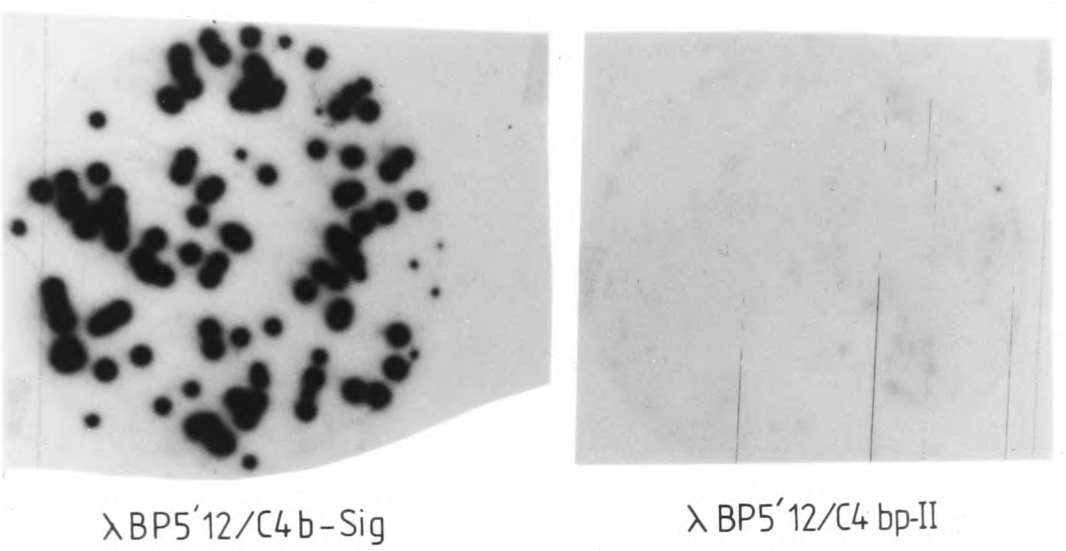

C4 bp - II
undil - 2 µl
undil - 1 µl
1 in 2 - 1 µl
1 in 5 - 1 µl
C4 bp - Sig
undil - 2 µl
undil - 1 µl
1 in 2 - 1 µl
1 in 5 - 1 µl

Fig 4.5 - (a) Results of the hybridisation of replica filters from λBP 5'12 rescreen with the two oligo's.

(b) Results of the hybridisation of replica filters from λBP3 with A400 cDNA probe and C4bp-II oligo.

Fig 4.5a



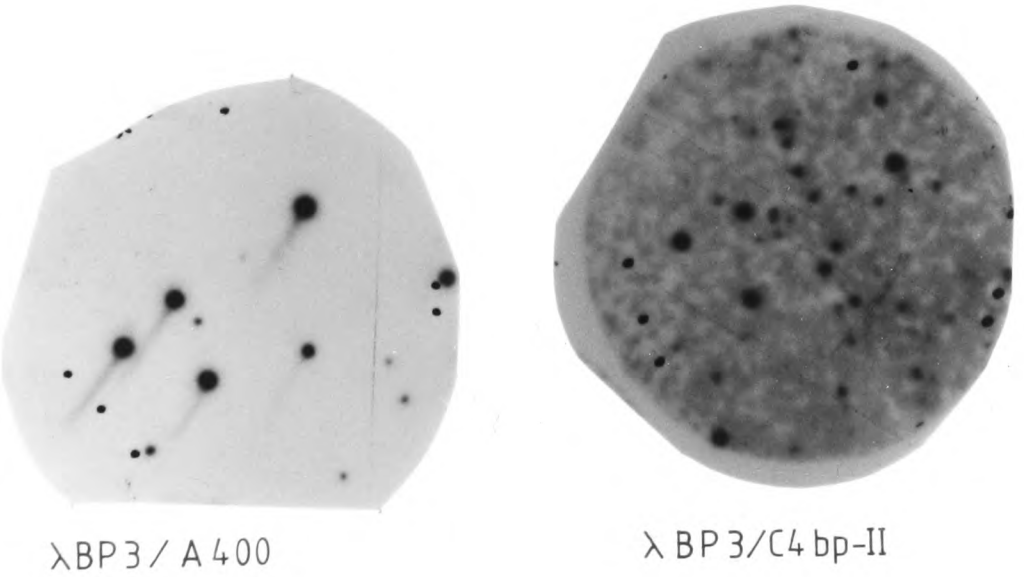


Fig 4.5b

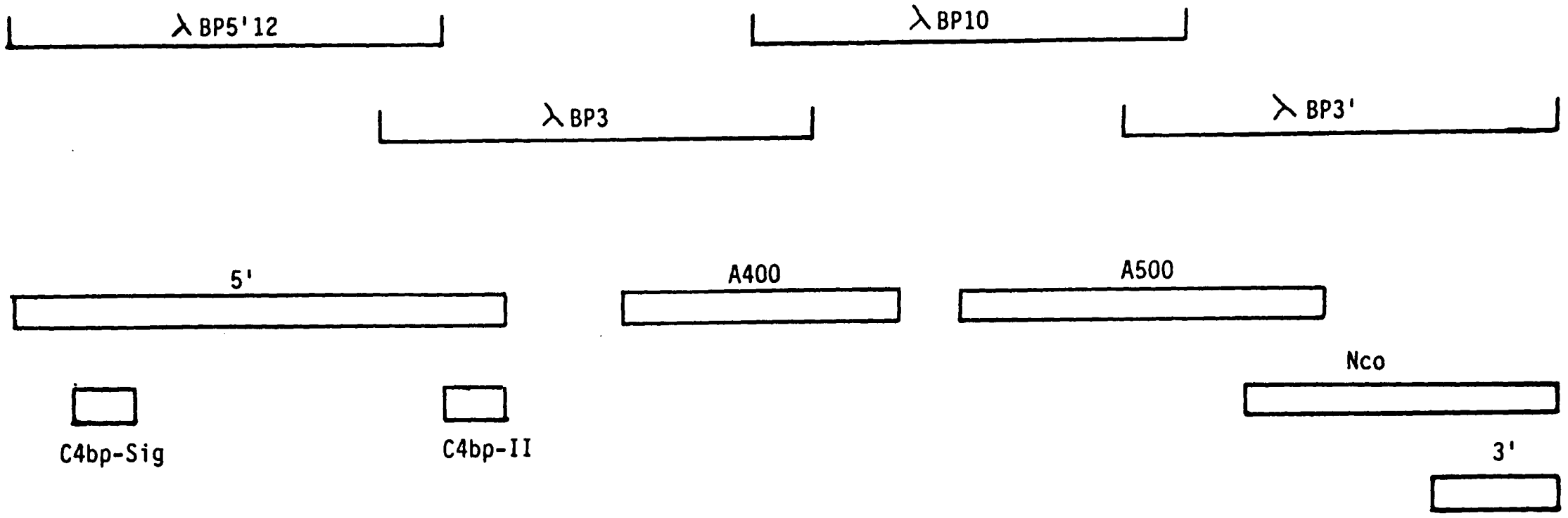
Fig 4.6 - Summary of the 4 overlapping λ genomic clones isolated using the cDNA and oligo probe shown. This diagram is not drawn to scale.

λ BP5'12 - C4bp-sig, 5' probe

λ BP3 - C4bp-II, A400

λ BP10 - A400, A500, Nco

λ BP3' - A500, Nco, 3' probe.


λ BP5'12
λ BP10
λ BP3
λ BP3'
5'
A400
A500
Nco
3'
C4bp-Sig
C4bp-II

C H A P T E R V

CHARACTERISATION OF C4bp GENOMIC CLONES

5.1 Introduction

Initial screening of the lambda genomic library isolated two distinct groups of positive clones, one group hybridised to the A400 probe (Fig 5.1) only and the other group, hybridised to the A400 and A500 probes. DNA was isolated from the latter group assuming that this would include the majority of the sequence coding for human C4bp.

Sequencing of the cDNA coding for C4bp had demonstrated that the primary structure could be divided into eight short consensus repeats (SCR's) and a 58 amino acid long 3' non-homologous region (Chung et al., 1985a). Factor B had been shown to contain three SCR's and sequencing of the gene demonstrated that each SCR was coded in a separate exon and that the coding for all three SCR's was represented in approximately 2 kb of genomic sequence (Campbell and Porter, 1983).

During analysis of the original clones preliminary results from blot analysis of genomic DNA using the S1 probe, which covers most of the cDNA sequence, suggested that the gene coding for C4bp was greater than 30 kb in size. Estimates of the size of the mRNA were 2.5 kb which suggested that much of the genomic sequence consisted of intron sequence.

Sequencing strategies discussed for the C4bp gene included a shotgun approach involving random shearing and cloning directly into M13 vectors for complete sequencing. In the light of the predicted size and proportion of intron sequence it was decided to use a more cautious mapping, subcloning and limited sequencing approach with the initial

clones.

Characterisation of one of the original clones using this approach showed it to include one-third of the total coding sequence. Subsequent isolation of new genomic clones for C4bp has utilised different cDNA probes and oligonucleotide probes to allow preliminary characterisation of the clones during the early rescreening and before DNA isolation and mapping.

5.2 Characterisation of initial C4bp lambda-genomic clones

Plate lysates were obtained using phage stocks prepared from plaque purified single positive clones as described in Chapter IV and DNA was then purified using the method outline in Section 2.3.5. Initial restriction enzyme mapping analysis of the different clones used enzymes with six base pair recognition sequences (BamH I, BglII, HindIII, EcoRI, StuI, PvuII) in digests on their own and in combination with SalI. The multiple cloning sites in EMBL-3 have SalI sites included which flank the human genomic DNA insert and can be used to excise the insert or, in double digests, for the identification of the ends of the insert.

Preparation of comprehensive restriction maps for the clones proved difficult due to the absence of accurate restriction maps of the vector arms. Difficulties in digesting the prepared DNA with some enzymes, due to DNase activity originating from early steps in the DNA preparation, further complicated the picture.

Eventually one of the clones was selected and used in a major subcloning experiment to create a 'mini-library' representing fragments of the lambda clone in a plasmid vector. λBP10 was used because PvuII and StuI digests of this clone gave fragments ranging from approximately 12 kb to 3 kb which were blunt ended and suitable for direct cloning into the PvuII cloning site of pAT153/PvuII/8 (Anson et al., 1984).

After ligation the plasmids were used to transform E. coli, MC1061 cells which were plated onto 14 cm L-broth ampicillin plates and the colonies screened using Whatman 541 filters (Section 2.8.1). By using two enzyme digests in separate subcloning experiments and by optimising the ligation transformation and plating conditions it was hoped to obtain overlapping plasmid subclones representative of the entire lambda-clone insert.

The numbers of recombinants screened from each digest were -

PvuII; 1226 colonies

StuI: 680 colonies

The filters were screened with A400 probe initially then washed with 0.5 M NaOH, neutralised and re-hybridised with A500 probe. (The cDNA probes are shown in Fig 5.1). This gave rise to four distinct populations of clones, one group which hybridised to the A400 probe only and one group which hybridised to both A400 and A500 probes, each with representatives originating from both enzyme digests. 15 colonies from each class were picked and used in the preparation of plasmid DNA using the rapid isolation protocol described in Section 2.3.2.

BP/Pv 1-15 - A400

BP/Pv 16-30 - A400/A500

BP/St 1-15 - A400

BP/St 16-30 - A400/A500

The inserts were characterised by agarose gel electrophoresis of BamH I/ClaI digested plasmids, the band patterns suggested that the majority of clones in each population were the same. Differences seen in the band patterns of some clones could be attributed to the presence of more than one insert suggesting that single colonies were not picked from the master plates. One representative from each population was selected and large volume cultures prepared and used to isolate plasmid

DNA.

BP/Pva	-	A400
BP/Pvb	-	A400/A500
BP/Sta	-	A400
BP/Stb	-	A400/A500

5.2.1 Characterisation of BP/Pva (Hybridised to A400)

Initial characterisation involved combinations of single and double digests using enzymes with six base pair recognition sequences Asp718; BamH I; BglII; NcoI; StuI etc. Digests were analysed using agarose gel electrophoresis followed by blotting analysis, hybridising with A400 and S1 probes.

Further digests with four base pair recognition sequence enzymes, HinfI; AvaII; DdeI; RsaI etc gave very complex patterns when analysed with agarose gel electrophoresis. Blot analysis of the digests identified which fragments contained coding sequence for C4bp. Further detailed mapping of the HinfI digest in combination with the six base pair recognition sequence enzymes, using polyacrylamide gel electrophoresis to analyse the fill-in labelled fragments, permitted the construction of a detailed restriction map for subclone BP/Pva and identified the fragments containing coding sequence (Fig 5.2).

A preparative digest of BP/Pva using HinfI was carried out and the majority filled-in with cold nucleotides using DNA Polymerase I (Klenow fragment) to produce blunt-ends suitable for cloning into Hinc cut M13mp8 ($\frac{1}{10}^{th}$ of the digest was fill-in labelled with α^{32}P[dATP] to allow identification of the fragments by autoradiography). The fragments were separated on a 4% polyacrylamide gel and the insert fragments excised, eluted overnight and ethanol precipitated extensively. The following purified bands were then used in sequencing analysis -

BP/Pva-1 - 1400 bp

BP/Pva-2 - 520 bp (mixture of insert and vector fragments)

BP/Pva-3 - 160 bp

BP/Pva-3 was cloned directly into M13, single stranded DNA was analysed using T-tracking to identify both orientations of the insert and then one representative of each group was sequenced completely. The clone contained 10 bp of intron sequence followed by coding for residue 188 (lysine) through to residue 240 (asparagine) which corresponds to the HinfI site at 624 bp in the cDNA sequence (Fig 5.3). The intron junction thus coincided with the N-terminus of the predicted fourth repeat boundary (Fig 1.6 and Chung et al., 1985a).

BP/Pva-2 consisted of a mixture of the vector and insert 520 bp HinfI fragments which were cloned as a mixture into M13 and the replicative forms (RF) analysed for the presence or absence of a StuI site. The 520 bp HinfI insert originating in the genomic DNA of C4bp was known to contain a StuI site from mapping (Fig 5.2) and hence the RF is linearised, the vector fragment contains no StuI site and the RF is not linearised. The BP/Pva-2 clones for which the RF's were linearised by StuI were used in the preparation of single stranded DNA and again both orientations were identified and sequenced.

This clone included coding sequence for the C-terminus of the fourth repeat from the HinfI site at 624 bp in the cDNA through to the codon for residue 248 (proline). The intron junction again coincided with the boundary of the protein repeat. The definition of the 3' end of the exon included one more residue (Pro248) in the protein repeat than that predicted for repeat IV from homology considerations (Chung et al., 1985a, Fig 1.6)). The positioning of the intron junction was based on the presence of GT dinucleotide which maintained the "AG-GT" rule proposed for the 5' and 3' splice junctions of exons by Breatnach and Chambon, 1981.

BP/Pva-1 consisted of a 1400 bp fragment which included 1 kb of pAT vector sequence (2322-23 bp) and 400 bp of C4bp genomic sequence (see inset, Fig 5.2). The original subclone (Bp/Pva) was generated from a PvuII digest cloned into the PvuII site in pAT vector thus regenerating the PvuII recognition site. Based on the sequencing of BP/Pva-2,3 it was predicted that the coding sequence in the 400 bp fragment was for the protein repeat III and that the PvuII site thus corresponded to the PvuII site at position 366 bp in the cDNA (Fig 5.3). The gel purified fragment BP/Pva-1 was thus further digested with PvuII and the mixture of fragments cloned directly into M13mp8. T-tracking analysis of the M13 clones generated identified the one orientation containing the coding sequence closest to the sequencing primer site and subsequent annealing reactions identified the opposite orientation.

This fragment included coding for residue 154 (serine) corresponding to the PvuII site, through to residue 187 (glutamic acid). The intron junction coincided with the repeat boundary at the C-terminus of protein repeat III.

BP/Pva-1 included, at the 3' end of the exon, the first nucleotide of the codon for residue 188 (lysine) whilst BP/Pva-3 included, at the 5' end of the exon, the remaining two nucleotides of the codon. Similarly BP/Pva-2 included the first nucleotide of the codon for residue 249 (asparagine) at the 3' end of exon IV. The introns appear to fall in the same phase of reading frame defined as Phase I by Rogers, 1986, i.e. after the first base pair of the codon.

5.2.2 Characterisation of BP/Sta (Hybridised to A400)

Initial characterisation again involved single and double digests of the subclone using six-base cutting enzymes. Early digests using SalI suggested that approximately half of the insert consisted of

sequence from the 20 kb arm of the EMBL-3 vector. Digests were again analysed by agarose gel electrophoresis and blots analysed by hybridisation with A400 and S1 cDNA probes. No fragments hybridised to the S1 probe alone suggesting that λBP10 did not extend to include coding for the second protein repeat. (Subsequent hybridisation of rescreen filters of λBP10 with the oligonucleotide C4bp-II confirmed this result.) The restriction map constructed for BP/Sta is shown in Fig 5.4.

The map showed that BP/Sta overlapped with BP/Pva and extends into the 20 kb vector arm thus identifying the 5' limit of λBP10. BP/Sta included coding for protein repeats III and IV and thus extends to include the 5' intron junction of exon III. This was shown by the sequencing of an 800 bp HindIII/PvuII fragment BP/Sta-1. This fragment was purified by elution from a polyacrylamide gel and cloned directly into M13. The two orientations were identified by T-tracking and one of each sequenced. This fragment was shown to contain coding for residue 154 (serine), corresponding to the PvuII site at position 366 bp in the cDNA sequence, through to residue 124 (isoleucine). The intron coincided with the protein repeat boundary at the N-terminus of protein repeat III and included two nucleotides of the codon for residue 124 (isoleucine).

Subclone BP/Sta contained 5.1 kb of C4bp genomic DNA together with approximately 8.0 kb of the 20 kb vector arm. 2.7 kb of the insert overlapped with subclone BP/Pva and extended the 5' data to define the 5' limit of the λBP10 genomic clone. Sequencing of an 800 bp HindIII/PvuII fragment, BP/Sta-1, identified the 5' end of the exon coding for protein repeat III. Blotting analysis showed that λBP10 did not extend to include sequence for protein repeat II.

5.2.3 Characterisation of BP/Stb (Hybridised to A400 and A500)

Initial characterisation again involved analysis of single and double enzyme digests using agarose gel electrophoresis and blot analysis, hybridising with A400 and A500 cDNA probes. These probes hybridised to the subclone during rescreening of the subclones. Double digests with SalI suggested that BP/Stb included sequence originating in the 9.0 kb arm of the vector. One fragment HindIII/SalI, BP/Stb-1, hybridised to the A500 probe alone, suggesting that the subclone extended to include sequence coding for the sixth protein repeat. Also a PvuII/BamH I fragment, BP/Stb-2, hybridised to both probes, A400 and A500, the sequencing of which is described in Section 5.2.4. The restriction map for BP/Stb is given in Fig 5.5.

The 1.1 kb BP/Stb-1 fragment contained no other useful restriction sites to give fragments suitable for direct cloning into M13 vectors for direct sequencing. The 1.1 kb fragment was filled-in to give blunt ends and purified from a polyacrylamide gel and subcloned again into the PvuII site of pAT (subclone BP/Stb-1). Mapping of the HinfI sites in this subclone was attempted to pinpoint the HinfI site corresponding to the HinfI site at position 1008 bp in the cDNA sequence (Fig 5.3) to allow sequencing from the middle of the coding region. Multiple small HinfI fragments were identified and proved impossible to map, thus a HinfI digest of subclone BP/Stb-1 was used in a shotgun cloning experiment into M13mp8. The clones containing coding were identified by hybridisation of replica filters from the M13 plate with the A500 probe.

White plaques corresponding to positive signals were isolated and used to prepare single stranded DNA which was analysed using T-tracking to identify the four orientations of insert containing coding (Fig 5.6). Three major populations of insert were identified and one representative from each group was sequenced. One insert included sequence coding for

residue 368 (serine) corresponding to the HinfI site at position 1008 bp in the cDNA sequence, through to residue 376 (aspartic acid). The intron junction coincided with the C-terminal boundary of protein repeat VI. Again the definition of the 3' end of the exon, based on the position of the GT dinucleotide, included extra residues (glycine, 375 and aspartic acid, 376) in the protein repeat compared with the repeat based on homology predictions (Fig 1.6 and Chung et al., 1985a).

The other two representative clones, when sequenced contained sequence originating in the 298 bp pAT Hinf fragment. This suggests that the A500 probe is homologous to this region of the vector. The remainder of the clones gave different T-tracking patterns, each with different regions in common suggesting that small HinfI fragments had ligated together. Confirmation of the 5' end of the sixth exon has not yet been completed.

BP/Stb thus contained 3.7 kb of C4bp genomic DNA together with approximately 4.0 kb of the short vector arm. Hybridisation analysis with the A400 and A500 probes suggests the clone includes coding for protein repeats V and VI. Sequencing of the HinfI fragment originating in the BP/Stb-1 fragment confirmed the presence of sequence covering the C-terminal amino acid residues of protein repeat VI.

5.2.4 Characterisation of BP/Pvb (Hybridised to A400 and A500)

Mapping and blotting analysis of this subclone demonstrated that it overlapped with BP/Stb in a 700 bp StuI/ClaI fragment which hybridised to the A400 and A500 cDNA probes. This suggested that the overlap included coding for protein repeat V. The restriction map of BP/Pvb is shown in Fig 5.5. The Asp718 site in this region corresponds to the Asp718 site positioned at 749 bp in the cDNA sequence (Fig 5.3).

The digests generating fragments BP/Pvb-1 (500 bp Asp718/ClaI) and

BP/Pvb-2 (200 bp Asp718/StuI) were filled-in to give blunt ends, gel-purified and the appropriate fragments cloned into M13mp8. The four orientations were identified from T-tracking and one representative from each sequenced. BP/Pvb-1 included sequence coding for residue 282 (tyrosine) which corresponds to the Asp718 site at 749 bp in the cDNA, (Fig 5.3), through to residue 313 (glutamic acid). BP/Pvb-2 included sequence coding for residue 282 (tyrosine) through to residue 249 (asparagine). The sequencing gel corresponding to the intron junction in this fragment is shown in Fig 5.7.

Thus subclone BP/Pvb contained sequence coding for the fifth protein repeat, the intron junctions again coincided with the repeat boundaries and the base pair organisation of the first codon of the exon suggested that the intron was still in phase I of the reading frame.

5.2.5 Confirmation of λBP10 restriction map

Alignment of all four subclones gives an overall map for the insert of λBP10 (Fig 5.8). The intron sequence between the exons coding for protein repeats IV and V is 4.5 kb in length and is contained almost entirely within one StuI fragment. The pAT subclone containing this fragment will not hybridise to cDNA probes as a result of which no overlapping subclone from this region was available. To confirm the alignment of the subclones the 13.5 kb insert from λBP10 was subcloned intact into pUC8 plasmid vector using direct cloning of a SalI digest into the SalI site in the pUC8 polylinker. The recombinants containing the C4bp insert were identified by hybridisation with the S1 probe and mapped using the same enzymes used to generate the maps for BP/Pva etc. This confirmed that the map shown in Fig 5.8 was complete and that no small fragments in the large intron had been missed.

The 13.5 kb of human genomic DNA included in the clone, λBP10

included coding for residues 124-376 of human C4bp. At least two more λ-genomic clones were needed to cover the remaining C4bp gene sequence. The same λ-genomic library, made by Dr. A. Bentley, was screened for a second time using the Nco and 5' probes (Fig 5.1). These probes are specific for the cDNA sequence not represented in the exons found in λBP10. The probes used and the strategy employed in the isolation of the new λ genomic clones for C4bp was outlined in Chapter IV.

5.3 Characterisation of λBP3'2

All the genomic clones isolated using the Nco probe also hybridised to the A500 and 3' probes (Fig 5.1) suggesting that the clones extend to include the sequence for protein repeat eight. One purified plaque was chosen at random and used to prepare DNA with the plate lysate protocol (Section 2.3.5). In this preparation more care was taken to remove all traces of DNase from the phage solution before releasing the λ-DNA from the protein coats. This allowed the isolation of DNA from which the insert could be mapped intact.

Single and double enzyme digests were analysed using agarose gel electrophoresis. During the mapping of the λBP3'2 clone the small 11x11 cm minigel apparatus was used for which the acid/alkali treatment times of the gel were reduced and three copies of each blot made as described in Section 2.9. The blots were hybridised with all three cDNA probes and one was then stripped and re-hybridised with commercially available λ-DNA labelled by nick-translation. This identified the restriction fragments originating in the vector.

The absence of an accurate restriction map for the EMBL-3 vector arms made the initial steps in the construction of the map for this clone difficult. However, comparison of the details of the deletions

made during construction of λ1059 (Karn et al., 1980) and subsequently EMBL-3 (Frischauf et al., 1983) from wild type lambda, with the sections of vector arm mapped in subclones BP/Sta,b allowed construction of a fairly reliable map of EMBL-3 (see Fig legend 5.13).

Using results of enzyme digests and the multiple probe hybridisation analysis as shown in Fig 5.9, the complete map of the vector was confirmed and a map of the insert constructed (Fig 5.10). The overlapping cDNA probes used in this experiment covered protein repeats VI, VII and VIII and the 3' non-homologous region. They were generated from cDNA clones using six-base recognition sequence enzyme sites NcoI and Asp718 which are also used in the mapping of genomic clones. Thus the hybridisation pattern of fragments from NcoI and Asp718 digests of λBP3'2 pinpointed the position of exons coding for protein repeats VII and VIII. Hybridisation of the PvuII/ClaI 2.0 kb fragment to only the A500 probe showed that coding for protein repeat VI was present and hence λBP3'2 overlaps with λBP10 in this region.

Further use of this combination of probes and sites was used to generate plasmid subclones ideal for rapid sequencing. NcoI and Asp718 digests were filled-in and ligated into pAT153/PvuII/8 and transformation of MC1061 cells carried out. The transformed cells were plated directly onto nitrocellulose filters, which formed the master plates from which two replicas were taken and lysed as described for cosmid screening (Section 2.8.2). One set of filters was hybridised with the A500 probe, the other with the Nco probe and one set was then rehybridised after stripping, with the 3' probe. Overlaying of the autoradiographs revealed distinct populations of colonies hybridising to one or two of the probes. The coding in each subclone should be at one end of the insert and hence fragments suitable for sequencing should be easily identified e.g. the 3' end of the coding for protein repeat VII

will be at one end of the insert of an Nco derived subclone hybridising to Nco and A500 probes.

The final sequencing remains to be completed so it is not known if the 1.9 kb PvuII fragment hybridising to the 3' probe extends to include the sequence for the 3' non-homologous region. Use of a unique 17-long oligonucleotide originating in this region would indicate whether extensive sequencing was necessary in this region of the clone. Comparison with the genomic blot data, Chapter VI, suggests that the coding for the 3' non-homologous region is included within the 1.9 kb PvuII fragment.

5.4 Characterisation of λBP5' clones

The use of the 5' probe in the isolation of new λBP genomic clones identified one strong positive signal which subsequently hybridised to the C4bp-sig oligonucleotide probe. This suggests that λBP5' contains coding sequence for the signal peptide. The other oligo - C4bp-II did not hybridise to any of the λBP5' clones but it did hybridise to λBP3 which was originally isolated using A400 cDNA probe. Thus two new λ genomic clones are needed to complete the 5' region of the C4bp gene. The λBP3 clone overlaps with the fully characterised λBP10, based on the hybridisation to the A400 probe - no mapping data is yet available for this clone. Preliminary mapping results for λBP5'12 have not given conclusive evidence of the presence or absence of coding for protein repeat I in this clone. The use of oligonucleotides both in rescreening and blot analysis has proved very successful in the preliminary characterisation of this region of the C4bp gene and will speed up completion of the mapping and sequencing of the 5' genomic clones.

In the light of these most recent results the size of the complete C4bp gene has been amended and is now thought to be closer to 40 kb than the original estimates of 30 kb. The early estimates were based on results from genomic DNA blot analysis (Chapter VI). As already demonstrated in the λBP10 restriction map, large restriction fragments can consist entirely of intron sequences and are thus not identified in an analysis using cDNA probes.

5.5 Discussion

Four lambda genomic clones - λBP5'12, λBP3, λBP10, and λBP3'2 were isolated from an EMBL-3 genomic library, (using a combination of cDNA probes and oligonucleotide probes), which cover most of the coding sequences for human C4bp.

λBP10 and λBP3'2 have been mapped completely using restriction enzyme digests and shown to overlap to give a complete map covering 29 kb of genomic DNA. This includes two-thirds of the coding sequence for human C4bp gene. The map is shown in Fig 5.11. Together the two clones contain coding for residues 124-491 and, based on genomic blot data, extend to include amino acids 492-549 i.e. the protein repeats III-VIII and probably the 3' non-homologous region.

The coding so far characterised is represented in six discrete exons, the intron junctions of which correspond to the repeat boundaries defined in the primary protein structure. Each exon is approximately 200 bp in size whilst the introns vary from between 3.0-6.0 kb in size.

Overlapping plasmid subclones constructed to cover the entire λBP10 insert sequence were used in nucleotide sequencing to define the exon sequence together with splice junction and some intron sequence. The intron-junction sequences are summarised in Fig 5.12. Analysis of these

sequences shows that the "AG-GT" rule defined by Breathnach and Chambon, 1981, is obeyed in the introns sequenced. The consensus at the donor splicing site AG-GT $\begin{matrix}A\\G\end{matrix}$ AGT is conserved faithfully whilst the acceptor splice site (Py)n $N\begin{matrix}C\\T\end{matrix}$AG-GT is less well conserved. The (T)n region can be seen in the sequence shown in Fig 5.7.

The intron junctions sequenced so far fall in homologous positions within the first codon of the exon where the intron falls after the first nucleotide of the codon. As already mentioned each discrete exon encodes one of the short consensus repeats described in the protein structure.

The repeat boundaries are positioned within the amino acid sequence on the basis of the positions of the first and last cysteine residues of the repeat consensus sequence. In most of the intron junctions so far sequenced the repeat boundary coincides exactly with the predicted position. However the intron between the exons coding for protein repeats IV/V and VI/VII forms junctions which are shifted to a position 5' to that predicted. This slight deviation could result from restrictions on codon use or splicing rules preventing precise joining of the repetitive units. It is also possible that intron positions may have shifted by imprecise splicing or mutations that activated cryptic splicing sites. This type of event, leading to addition or loss of exon sequences, can only be tolerated if it occurs in the regions coding for the non-conserved residues of the protein sequence. In the case of the SCR's this flexibility is found at the extremities of the exon sequences outside the first and last cysteine residues of the consensus sequence. If the addition/loss of coding extended to affect the first and last cysteine residues the repeat would be unable to fold independently and possibly lose its function.

Although the introns were by no means sequenced in full, limited intron sequence was obtained during the sequencing of both orientations of junction-containing fragments. This sequence was used in computer database searches to see if any homologies to repetitive DNA sequences existed, no good matches were identified. The same sequences were used with programmes to predict secondary structure and this showed the presence of sequence able to form hairpin loops together with stem and loop structures.

Mapping of the lambda vector arms in the subclones BP/Sta,b and hybridisation data of labelled λ-DNA to restriction fragments of λBP3'2, when compared with deletion details of the EMBL-3 construction permitted the completion of restriction maps for the long and short arms of EMBL-3 - shown in Fig 5.13.

Preliminary mapping of λBP5'12 and hybridisation patterns of the oligonucleotide probes to λBP3 and λBP5'12 suggest the total size for the C4bp gene could be in excess of 40 kb. This is bigger than original estimates based on genomic DNA blotting data, the discrepancy is due to the identification, in cloned DNA, of fragments containing intron sequence only and hence not identified by cDNA probes. Use of oligonucleotides specific to regions coding for protein repeat II and the signal peptide together with the 5' cDNA probe should enable the rapid completion of the mapping and sequencing in the 5' region of the gene.

Some of the preliminary data on sequence and mapping was presented at the 17th FEBS meeting in Berlin (Aug 1986) and is published in Lintin and Reid, 1986. The remaining results were presented at the 12th Complement Workshop, Chamonix 18-21st September, 1987.

Fig 5.1 - cDNA probes used in the mapping of λBP genomic clones.

A400/A500 cover coding for residues 175-307 (excluding 206-221 not present because of the loopback in clone) which includes protein repeats III, IV and V and residues 307-470, protein repeats V, VI, VII and VIII respectively.

S1 - insert from pBP6 and covers coding for residues 32-530 which includes protein repeats I-VIII and 3' non-homologous region.

Nco - 450 bp from an NcoI/HindIII digest of pBP6 covers coding for residues 385-530 and includes protein repeats VII, VIII and 3' non-homologous region.

3' - 150 bp probe from an Asp718/HindIII digest at pBP6 covers coding for residues 483-530 and includes protein repeats VIII and 3' non-homologous region.

5' - 450 bp probe from a BamH I/HinfI digest of pBP51 covers residues 1-72 (repeats I and II), the putative leader sequence and 5' non-translated.

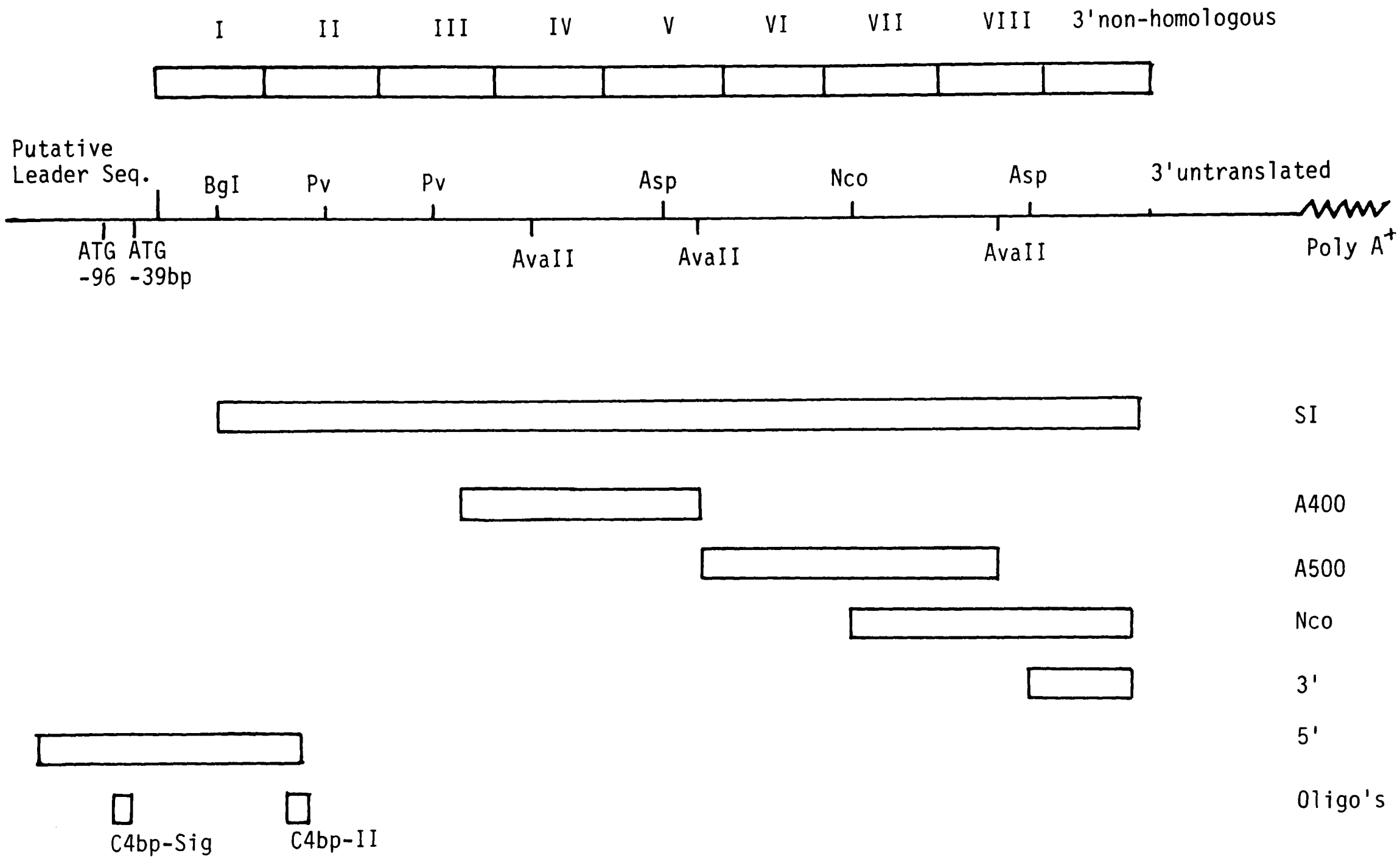

I
II
III
IV
V
VI
VII
VIII
3'non-homologous
Putative
Leader Seq.
BgI
Pv
Pv
Asp
Nco
Asp
3'untranslated
ATG
-96
ATG
-39bp
AvaII
AvaII
AvaII
Poly A+
SI
A400
A500
Nco
3'
5'
Oligo's
C4bp-Sig
C4bp-II
0bp
250bp
500bp

Fig 5.2 - Restriction map of BP/Pva. The restriction sites BamH I and ClaI represent the vector sites flanking the insert and correspond to PvuII sites in the original λBP10 insert. Restriction sites within the insert are marked; As, Asp718; M, BamH I; BgII, BglII; Hf, HinfI; Nc, NcoI and St, StuI. The solid bars represent the fragments which hybridised to the cDNA probes and which were gel purified and used in sequence analysis. The hatched boxes represent the exon sequences and the numbers refer to the amino acids encoded by each exon. Roman numerals correspond to the number of the protein repeat coded for by the exon. The inset demonstrates the sequencing strategy for the fragment BP/Pva-1.

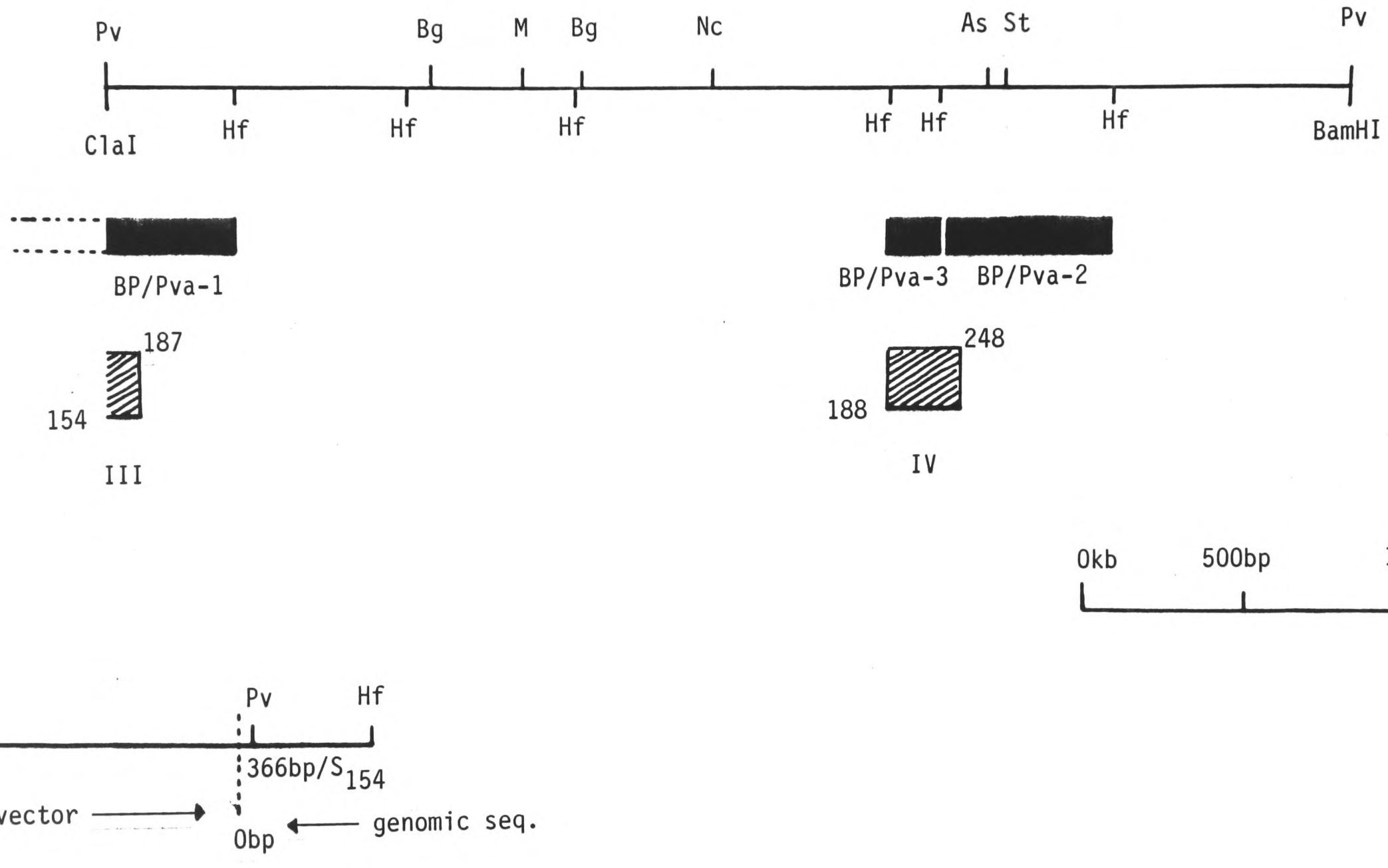
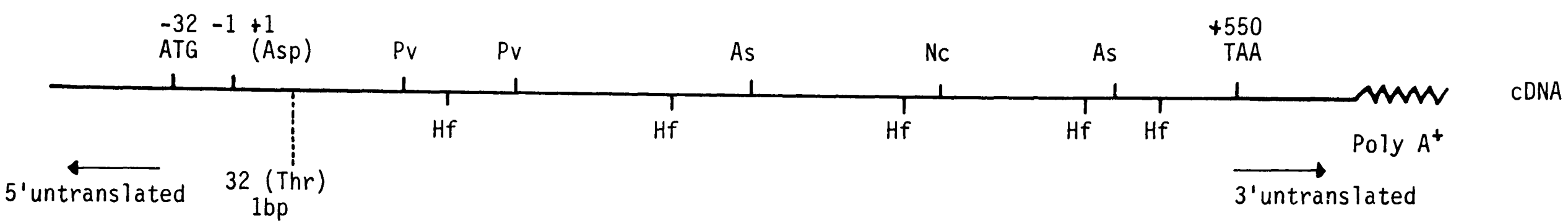

Pv
Bg
M
Bg
Nc
As St
Pv
ClaI
Hf
Hf
Hf
Hf Hf
Hf
BamHI
BP/Pva-1
BP/Pva-3
BP/Pva-2
187
154
III
248
188
IV
0kb
500bp
1kb
Hf
Pv
Hf
2322bp
366bp/S154
pAT vector
0bp
genomic seq.

Fig 5.3 - A limited restriction enzyme map of the cDNA sequence coding for human C4bp and the primary protein structure. The protein structure is marked with the eight internally homologous repeating units (SCR's), the numbers refer to the residues at the predicted boundaries (Chung et al., 1985a). The restriction sites marked on the cDNA are As, Asp718; Pv, PvuII; Nc, NcoI and Hf, HinfI and are the sites utilized during subcloning and sequencing of genomic clones. The residue 32 (threonine) is marked as this represents the first bp of the cDNA numbering system used by Chung et al., 1985a.


-32 -1 +1
ATG (Asp)
Pv
Pv
As
Nc
As
+550
TAA
cDNA
Hf
Hf
Hf
Hf
Hf
Poly A+
5'untranslated
32 (Thr)
1bp
3'untranslated

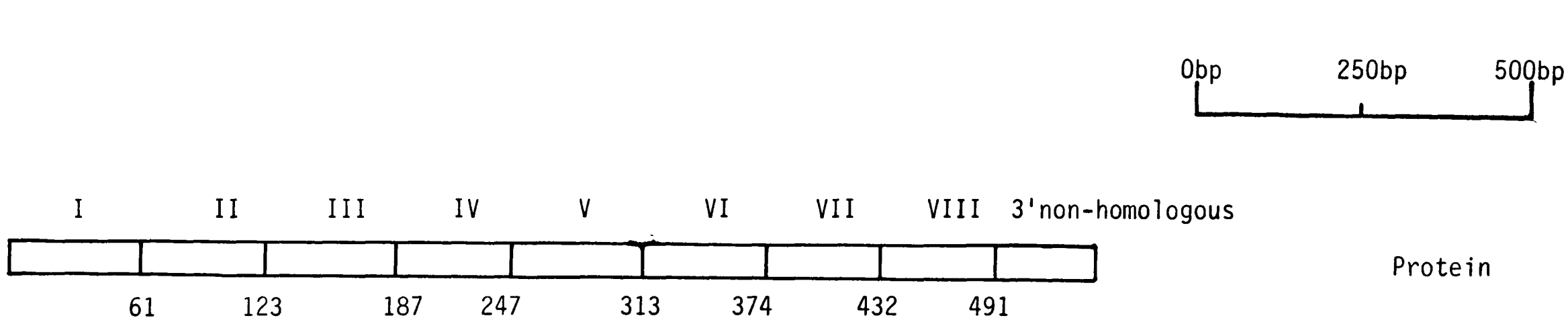

0bp
250bp
500bp
I
II
III
IV
V
VI
VII
VIII
3'non-homologous
Protein
61
123
187
247
313
374
432
491

Fig 5.4 - Restriction map of BP/Sta. The notation in this diagram is the same as in Fig 5.2, BamH I and ClaI sites correspond to StuI sites in the genomic DNA. Other restriction enzymes used include Cl, ClaI; SaI, SalI. The SalI site represent the 5' limit of the λBP10 insert, the remainder of the subclone is sequence originating in the 20 kb vector arm.

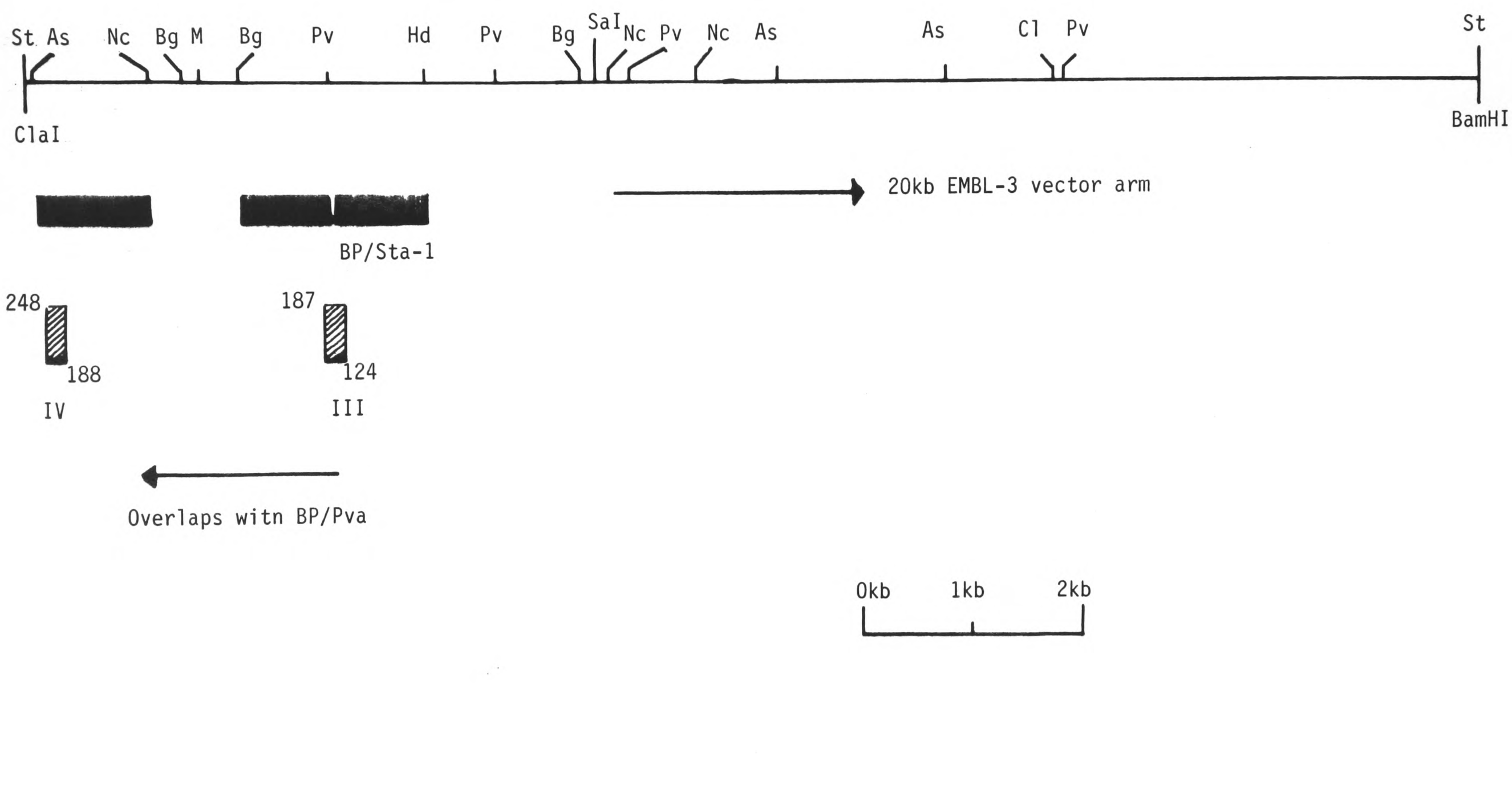
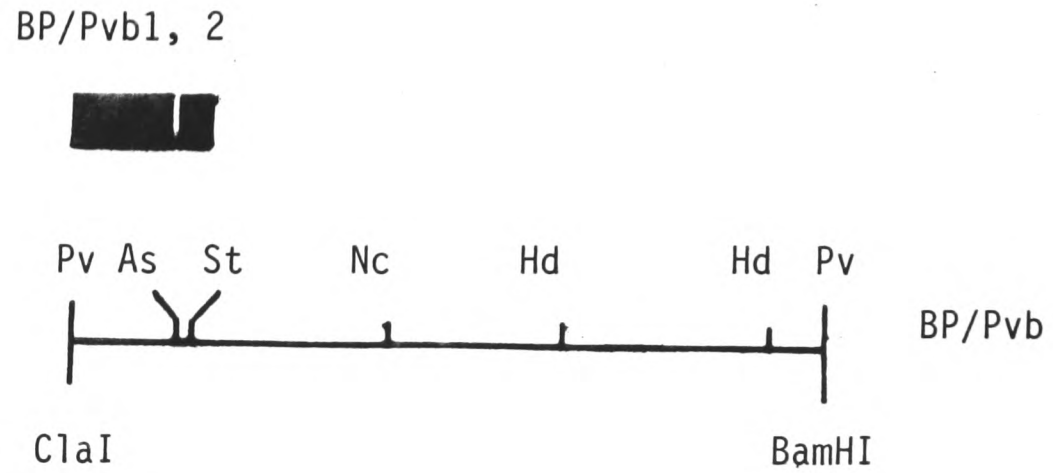

St
As
Nc
Bg
M
Bg
Pv
Hd
Pv
Bg
SaI
Nc
Pv
Nc
As
As
Cl
Pv
St
ClaI
BamHI
20kb EMBL-3 vector arm
BP/Sta-1
248
188
IV
187
124
III
Overlaps witn BP/Pva
0kb
1kb
2kb

Fig 5.5 - Restriction maps of the overlapping subclones BP/Stb and BP/Pvb. The notation in this diagram is as for Fig 5.2. BP/Stb includes approximately 4.0 kb of the sequence originating in the 9.0 kb vector arm. The SalI site thus defines the 3' limit of λBP10. The overlap of the two subclones occurs in the region coding for protein repeat V which was sequenced from the BP/Pvb fragments 1 and 2.


BP/Pvb1, 2
Pv As St Nc Hd Hd Pv
BP/Pvb
ClaI
BamHI

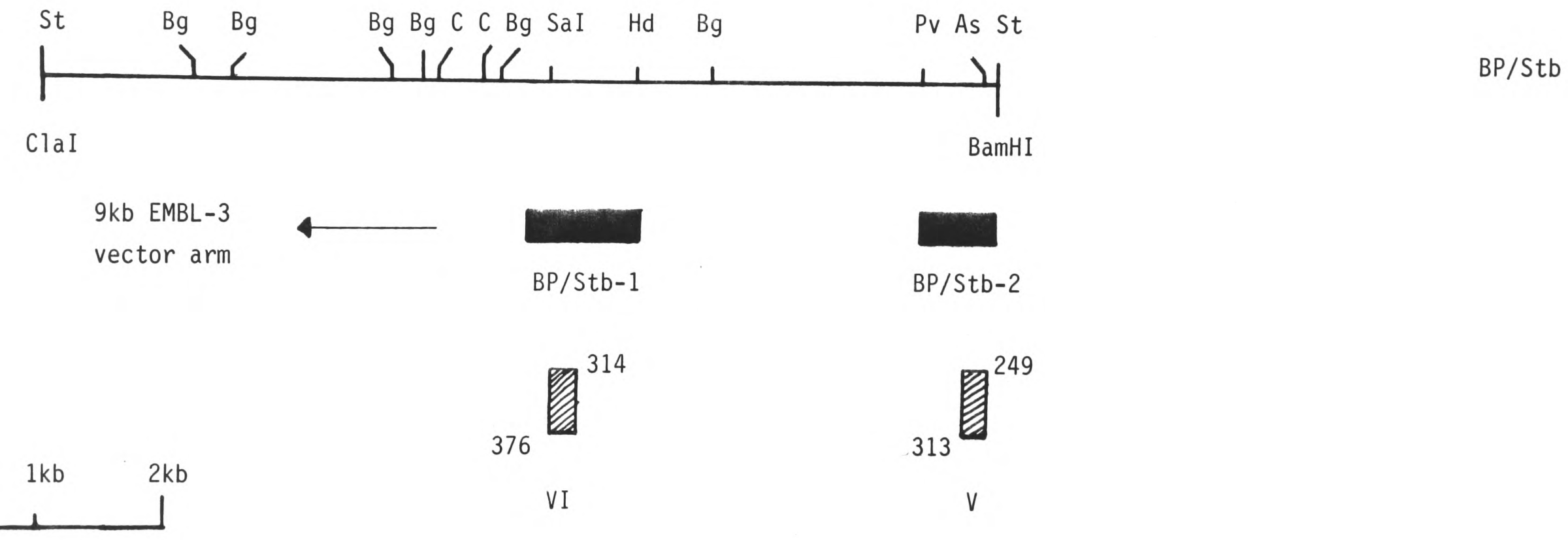

St Bg Bg Bg Bg C C Bg SaI Hd Bg Pv As St
BP/Stb
ClaI
BamHI
9kb EMBL-3
vector arm
BP/Stb-1
BP/Stb-2
314
376
249
313
VI
V
0kb 1kb 2kb

Fig 5.6 - T-tracking analysis of fifteen M13 clones originating from a HinfI digest of BP/Stb-1 and hybridising to A500 probe. Subsequently a further twenty clones were analysed in the same way. The three major populations are marked with 2, 3,4, one representative of each was sequenced.

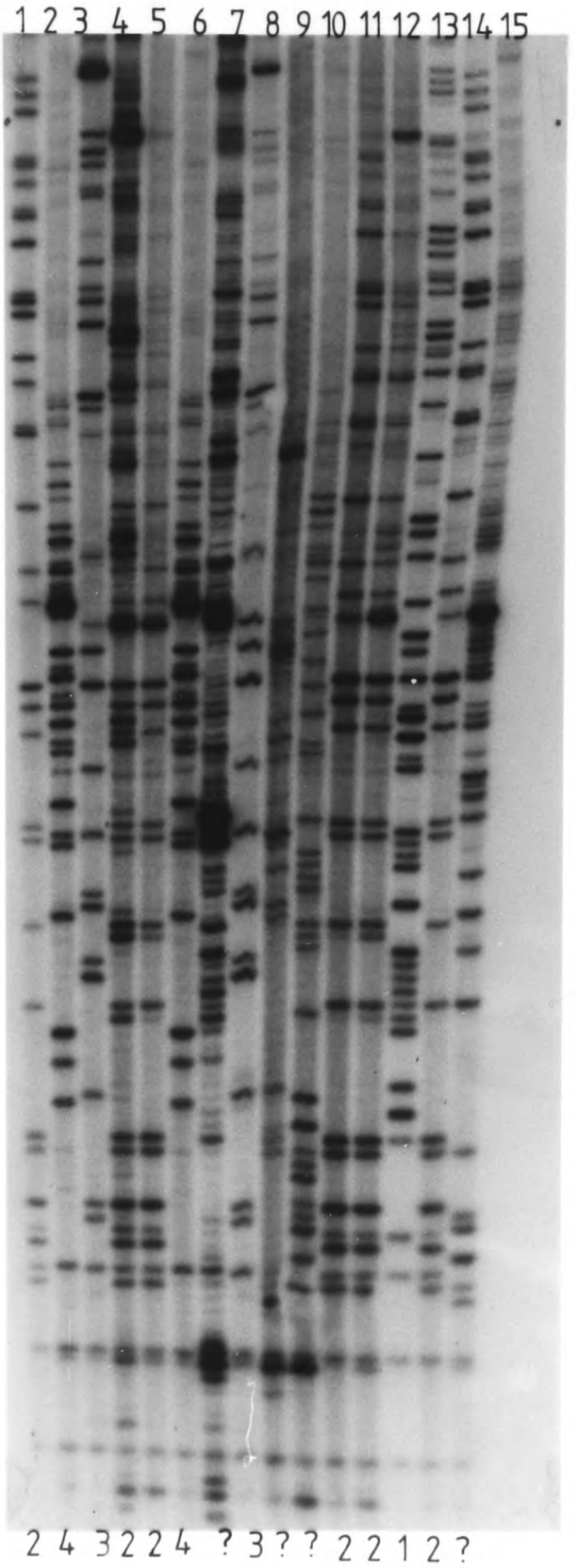
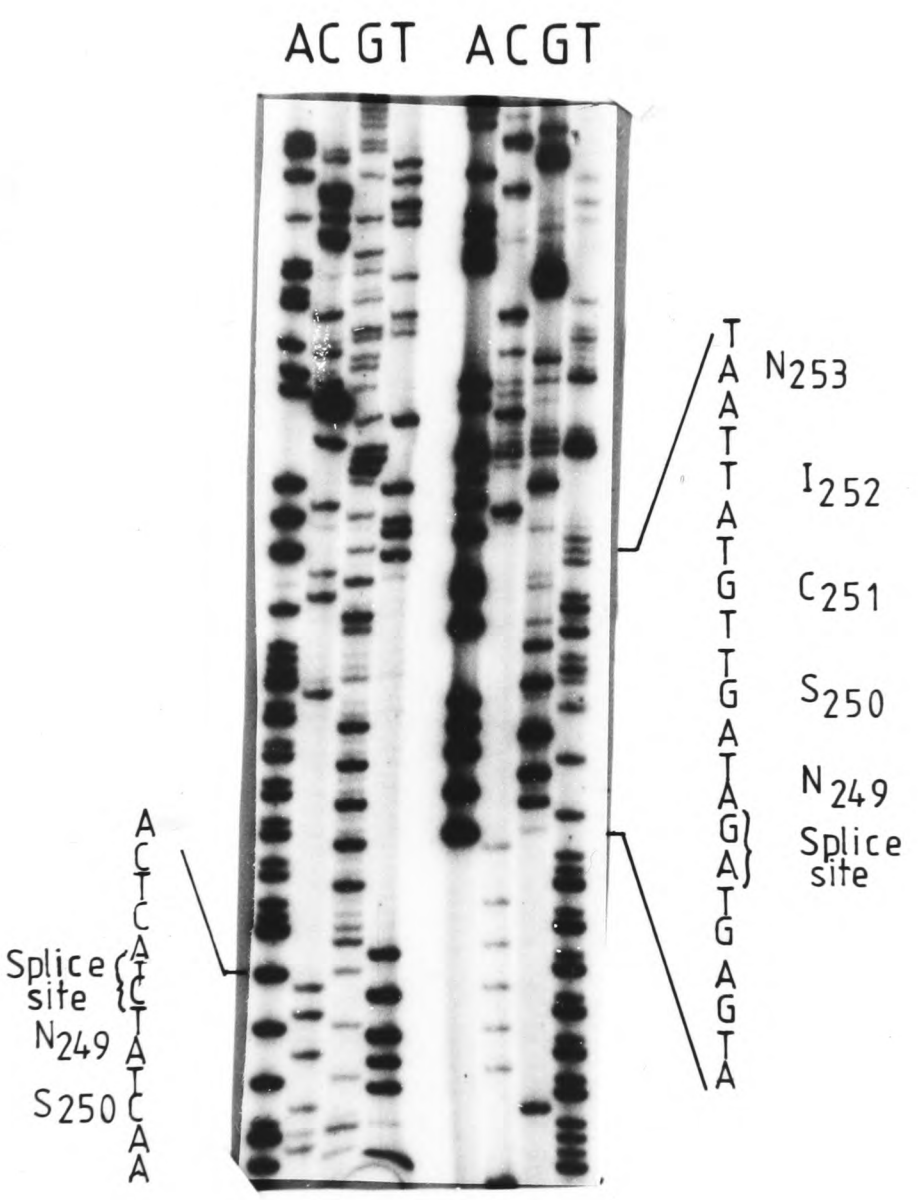

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
2 4 3 2 2 4 ? 3 ? ? 2 2 1 2 ?

Fig 5.7 - Autoradiograph of polyacrylamide gradient gel showing the dideoxy sequencing analysis, in both orientations of BP/Pvb-2 (200 bp Asp 718/StuI) fragment originating in subclone Bp/Pvb. The sequence shows the intron-junction at the 5' end of the exon coding for protein repeat V. The acceptor splice site "AG" and the (T)n region are included. The first nucleotide of the codon for residue 249 (asparagine) is situated at the 3' end of exon IV and is not seen in this sequence.



Sequence through the intron/exon boundary at the 5′ end of sequence for Rpt V

Fig 5.8 - Summarises the overlapping subclones which cover the λBP10 insert. The entire clone covers 13.5 kb and includes coding for amino acid residues 124-376 which is divided into four exons, each of which codes for one short consensus repeat (III, IV, V and VI). The diagram also shows which regions are covered by the cDNA probes used in the isolation and mapping of the subclones.

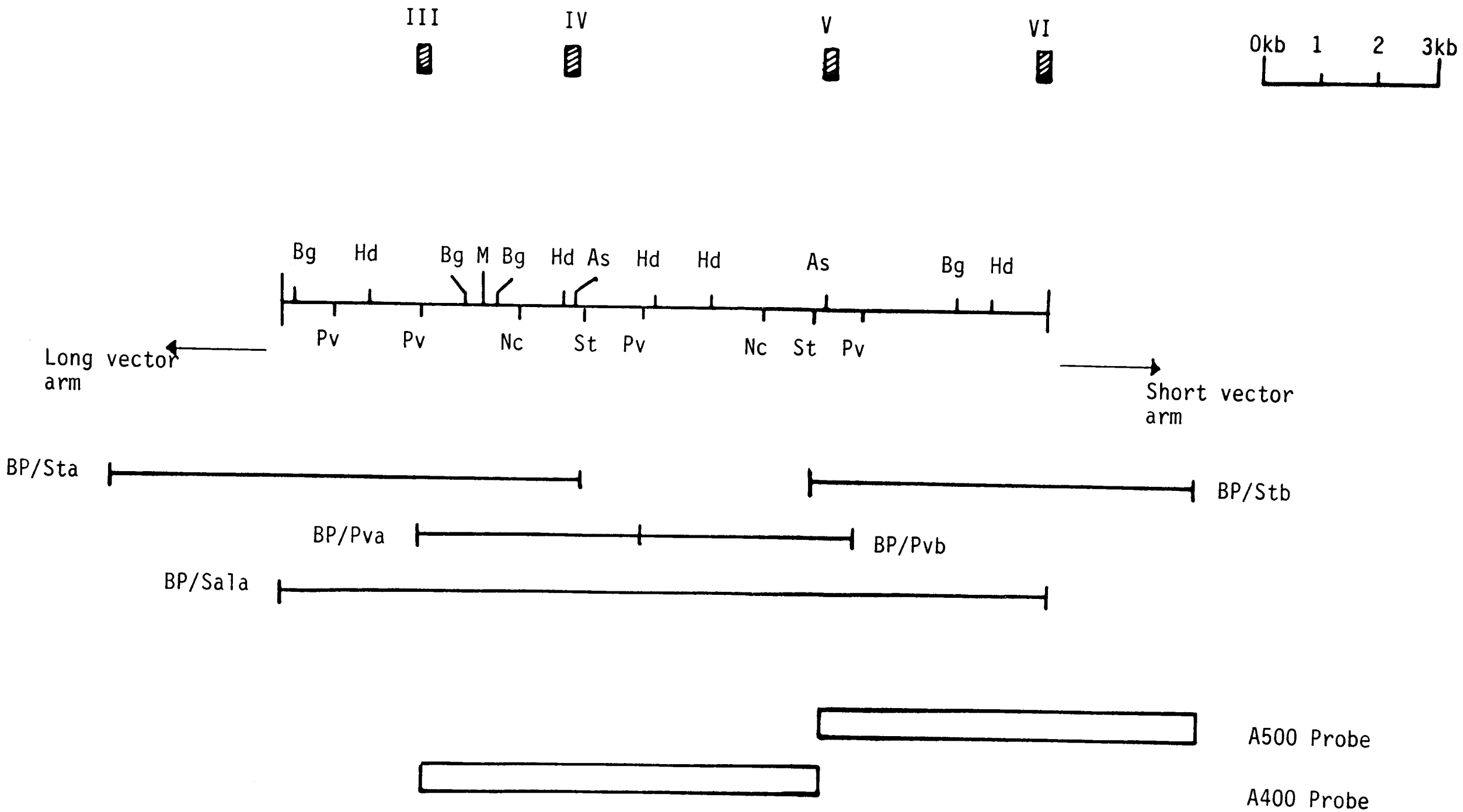

III
IV
V
VI
0kb 1 2 3kb
Bg Hd Bg M Bg Hd As Hd Hd As Bg Hd
Pv Pv Nc St Pv Nc St Pv
Long vector arm
Short vector arm
BP/Sta
BP/Stb
BP/Pva
BP/Pvb
BP/Sala
A500 Probe
A400 Probe

Fig 5.9 - (a) Photograph of the ethidium bromide stained agarose gel showing the restriction fragments generated from double digests with Asp718 and PvuII, used in the construction of the restriction map for λBP3'2.

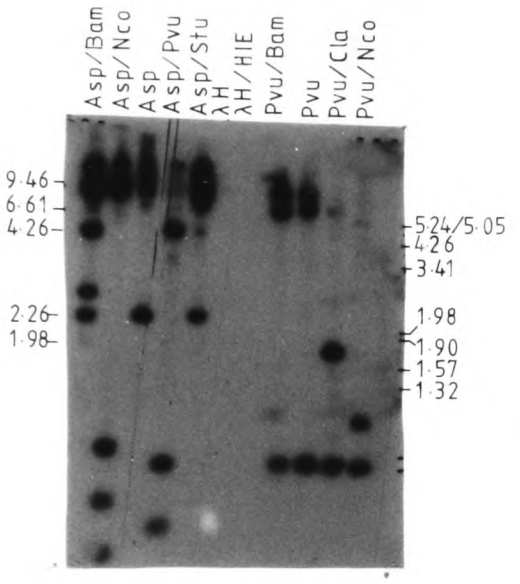
(b) Results of hybridisation of a transfer blot from the gel with CDNA probe A500. The band at 1.9 kb in the PvuII/ClaI digests demonstrates the region involved in the overlap with λBP10.

(c) Hybridisation results with the Nco probe. The band at 1.90 kb in the PvuII digest suggests that this fragment includes coding for the 3' non-homologous region - see Chapter VI.

(d) Hybridisation results with the 3'probe.

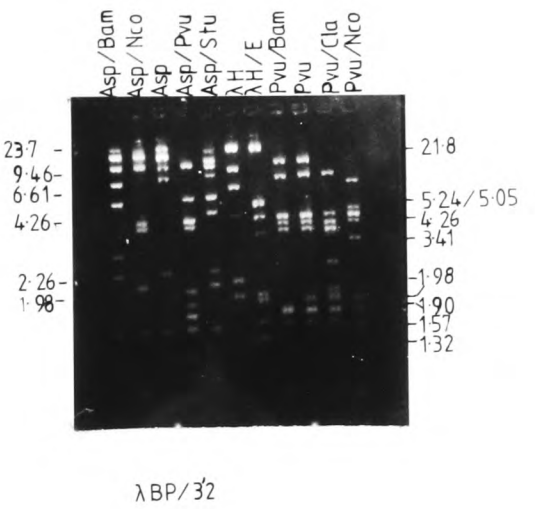
Results from this and other gel/blot combinations allowed construction of a restriction map for the λBP3'2 insert. Use of nick-translated wild-type λ as a probe on transfer blots confirmed the derived maps for the EMBL-3 vector arms.

Fig 5.9b

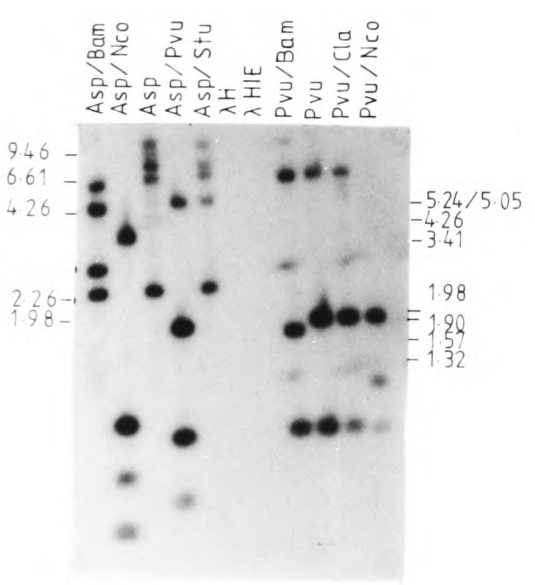


λBP/3'2 - ASOO Probe

Fig 5.9a

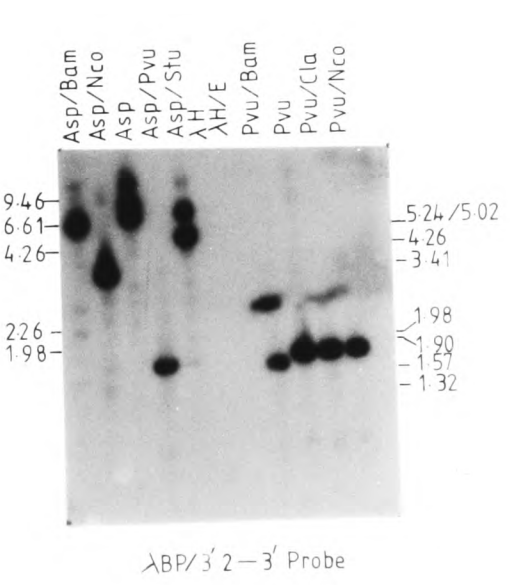


λBP/3'2



λBP/3'2 Nco Probe

Fig 5.9c



λBP/3'2—3' Probe

Fig 5.9d

Fig 5.10 - The complete restriction map constructed for the λBp3'2 insert. The restriction sites are marked as in other figures. This clone covers 16kb of genomic DNA and includes coding for residues 314-491 (which includes protein repeats VI, VII and VIII) and possibly the 58 amino acids in the 3' non-homologous region. Sequence confirmation of the exon sequences is not complete. The positioning of the exons is based on the hybridisation of the cDNA probes shown together with the enzyme sites in the λBP3'2 clone corresponding to cDNA sites (Fig 5.3).

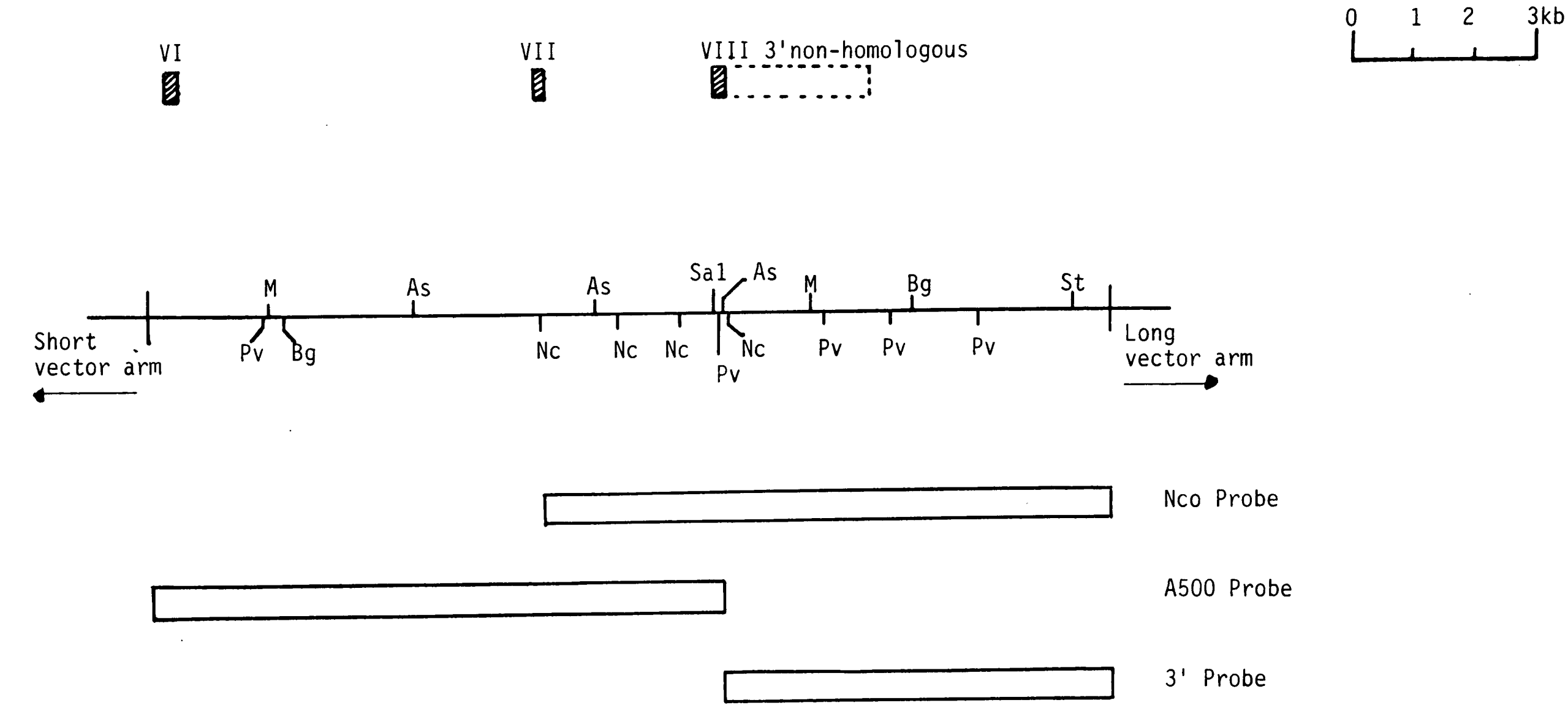

0 1 2 3kb
VI
VII
VIII 3'non-homologous
Short vector arm
M
Pv
Bg
As
Nc
As
Nc
Nc
Sal
Pv
As
Nc
M
Pv
Bg
Pv
St
Pv
Long vector arm
Nco Probe
A500 Probe
3' Probe

Fig 5.11 - The overall map covered by the overlapping λ clones λBP10 and λBP3'2. The position of the exons corresponding to the cDNA sequence and alignment of the protein internal repeats is shown. The approximate sizes of the introns are marked.

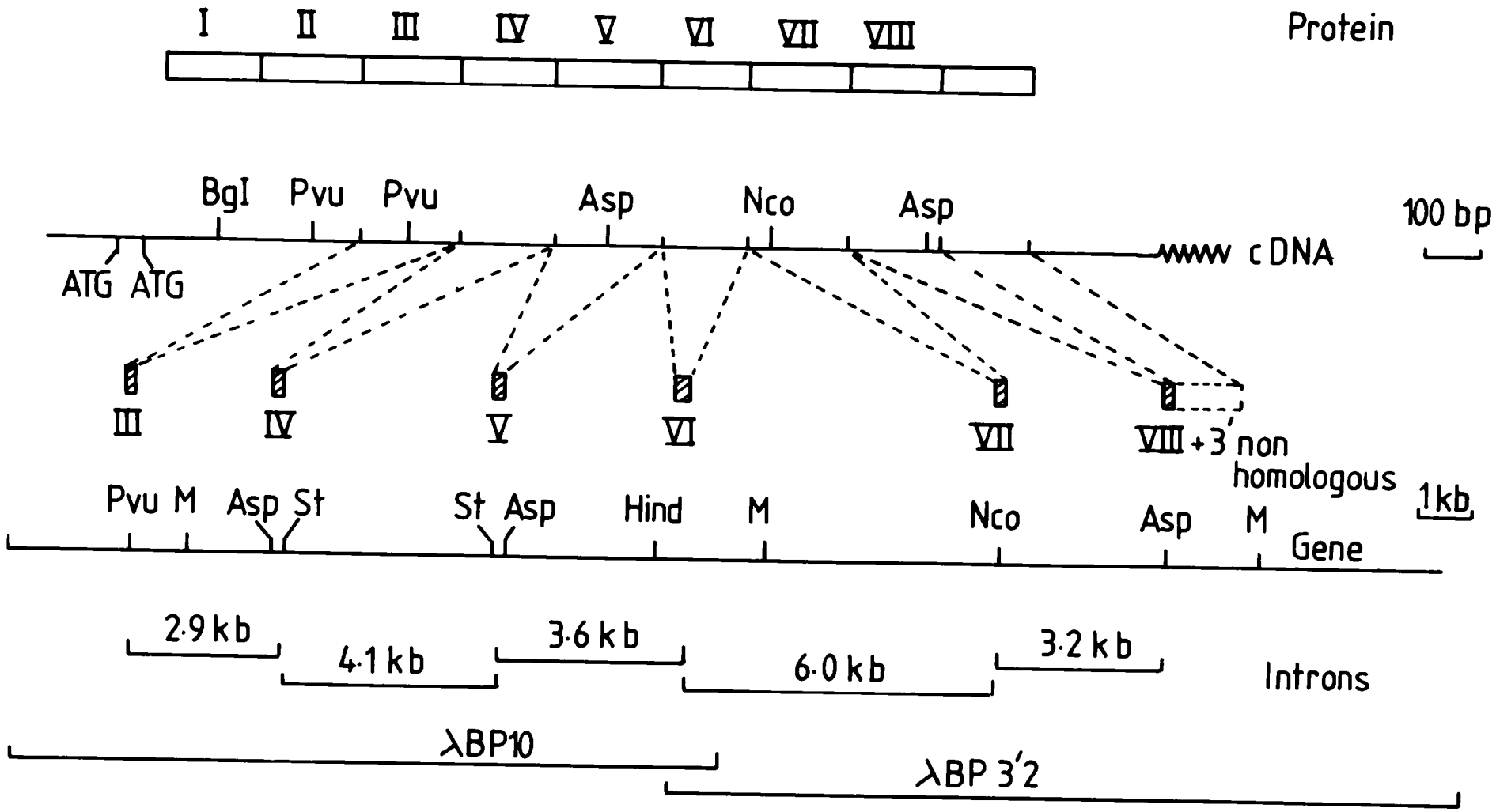

I
II
III
IV
V
VI
VII
VIII
Protein
BgI
Pvu
Pvu
Asp
Nco
Asp
100 bp
ATG
ATG
cDNA
III
IV
V
VI
VII
VIII + 3 non homologous
Pvu
M
Asp
St
St
Asp
Hind
M
Nco
Asp
M
1kb
Gene
2·9 kb
4·1 kb
3·6 kb
6·0 kb
3·2 kb
Introns
λBP10
λBP 3'2

Fig 5.12 - The splice site sequences for exon III - VI.

INTRON/EXON BOUNDARY SEQUENCES FOR THE C4bp GENE

EXON III

| | | |

Fig 5.13 - EMBL-3 is a lambda replacement vector derived from λ1059 by the insertion of a multiple cloning site between the lambda arms and non-essential stuffer region. Comparison of the restriction sites predicted by computer analysis of wild type lambda sequence with sites approximately positioned from the deletion details (Karn et al., 1980) enabled construction of the maps. Most of the deletions involved regions in close proximity to the cloning sites which have been mapped fully in the subclones Bp/Sta and b. This gave exact physical measurements in these regions which when compared with the positions in wild type λ pinpointed the deletions in EMBL-3.

The BamH I sites (and the multiple cloning site insertion point) at the ends of the vector arms correspond to the sites at positions 27972 and 34499.

The long arm of the vector is the same as the wild type over residues 1-19716 apart from the sBam 1° mutation at position 5505. The b189 deletion removed approximately 8 kb between the PvuII at 19716 and the BamH I site at 27972.

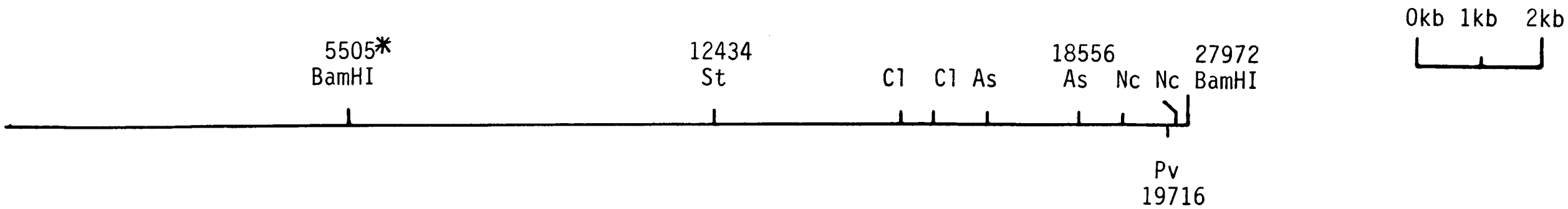
The short arm of the vector contains two deletions, KH54 which removes approximately 2 kb from between the two BglII sites at 35711 and 38103 and nin5 which removes approximately 3.0 kb from between the StuI site at 39992 and the Cla site at 43824.

The numbering refers to positions in wild type lambda and the box regions show the regions included in BP/Sta,b.

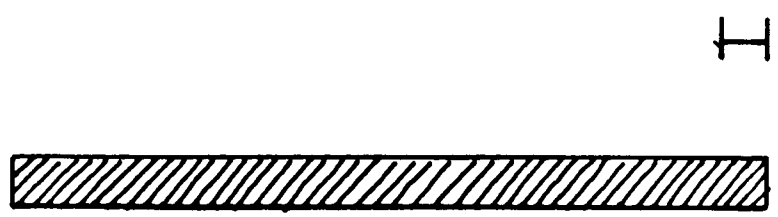


EMBL-3 short vector arm (9kb)





EMBL-3 long vector arm (20kb)



C H A P T E R VI

SOUTHERN BLOT ANALYSIS OF THE HUMAN C4bp GENE AND ITS LINKAGE TO THE FACTOR H GENE

6.1 Introduction

Analysis of the human C4bp gene was undertaken using hybridisation of cDNA probes to transfer blots of human genomic DNA (Southern blots). This gave preliminary information on the overall size of the gene together with initial alignments of restriction fragments and analysis of restriction fragment length polymorphisms (RFLP's). Information using this approach has developed alongside the cloning data and was not available prior to the initiation of the cloning experiments described in Chapters IV and V.

The hybridisation of cDNA probes, specific for C2 and Factor B, to common fragments on genomic blots was used to confirm the close physical linkage of the corresponding genes on chromosome 6 (Carroll et al., 1984). When a Factor H specific cDNA probe became available (Ripoche et al., 1986 and Day et al., 1987a) a similar type of approach was used to physically link the structural genes for C4bp and Factor H. Recently this approach has been extended to cover larger physical distances using the technique of pulsed field gel electrophoresis.

Studies on the structure and function of C4bp in the sera of mammalian species other than human are limited to the mouse and guinea pig. Difficulties in the performance of phylogenic studies of the complement system as a whole arise from the choice of unsuitable detection systems i.e. target cells, antibody sources and developing reagents (Linscott, 1986). Marche and Kindt, 1986, used genomic probes

originating in the T-cell receptor gene of a rabbit T-cell line (RL-5) to investigate interspecies conservation of sequences by cross hybridisation to genomic DNA from other mammalian species. cDNA probes originating from human C4bp clones were used to analyse Southern blots of genomic DNA from other mammalian species, aves and chondrichthyes (sharks), to show conservation of sequence between species.

6.2 Genomic blot analysis of the C4bp gene

Genomic DNA was prepared from three unrelated individuals LW; SL; and KA, using the method described in Section 2.3.6. This was then used in digests with various six base pair recognition enzymes (BamH I, BglII, HindIII, PvuII etc). The digests were then analysed using Southern blotting techniques described in Section 2.9. The blots were hybridised initially using A400 and A500 cDNA probes, (Fig 6.1) and subsequently with the S1 and Nco probes. The preliminary results from this approach indicated that the size of the gene coding for human C4bp was approximately 30 kb, based on the addition of the sizes of fragments in the more frequently cutting digests. When this was compared with the estimated size of mRNA of 2.5 kb it was apparent that a high proportion of the genomic sequence coding for C4bp would be composed of intron sequence.

Overlaying of the autoradiographs corresponding to the use of the different probes permitted alignment of the genomic restriction fragments. Fig 6.2 shows an autoradiograph of one genomic blot hybridised with the S1 probe showing the corresponding signals from the other fragment cDNA probes summarised on the overlaid acetate.

The results from this type of analysis (together with a limited number of double enzyme digests at the genomic DNA level) have been used

to confirm the size and alignment of fragments in the genomic clones and restriction maps obtained (Chapter V).

6.3 Restriction fragment length polymorphism in the C4bp gene

A restriction fragment length polymorphism was detected in the BglII digest with the DNA from individual LW. This can be defined by presence or absence of a single BglII recognition site. When the site is absent a single large fragment, approximately 11.0 kb in size, is detected which hybridises to both A400 and A500 probes. When the site is present two smaller fragments are detected, one hybridising with the A500 probe only at 2.6 kb and the other hybridises to both A400 and A500 probes at approximately 8.5 kb (Fig 6.8).

A further fifteen genomic DNA samples (a gift from A. Palsdottir) together with the LW sample were analysed using BglII digests (Fig 6.3). This demonstrated three distinct band patterns corresponding to the homozygote or heterozygote status of the individuals

LL - homozygous for the large fragment (11.0 kb)

SS - homozygous for the small fragments (2.6 kb and 8.5 kb)

LS - heterozygous (11.0 kb; 8.5 kb; 2.6 kb)

Of the sixteen examples used in this analysis three originate from one family in which the father is homozygous for the small fragments and the mother and daughter are heterozygous. This suggests that the polymorphism is inherited and not acquired.

6.4 Physical linkage of the C4bp and Factor H genes

Following the cloning of the cDNA coding for human Factor H reported by Ripoche *et al.*, 1986 and Day *et al.*, 1987a, a cDNA probe was made

available for use in genomic blot analysis in conjunction with C4bp cDNA probes. The factor H probe covered coding sequence for the C-terminus of the protein and included approximately 2.0 kb of 3' sequence. When this probe was used to hybridise liver Poly A^+ RNA transfer blots it demonstrated two species of mRNA approximately 4.7 and 5.7 kb in size (Ripoche et al., 1986). The 4.7 kb species is thought to code for the familiar 155 kDa Factor H protein and the 5.7 kb species is now thought to be an artefact. The truncated forms of Factor H, described in Chapter III, originate from the N-terminus of the protein and do not therefore hybridise to the probe used in this study (Ripoche et al., 1987; Day et al, 1987a). This probe thus covers less than 50% of the total sequence coding for the 155 kDa protein.

Initial results from the hybridisation of genomic blots with this probe suggested that a minimum estimate for the size of the Factor H gene (again based on addition of restriction fragment sizes originating from the more frequently cutting enzymes) was in the range of 80 kb.

When the autoradiographs of blots hybridised with both C4bp and Factor H cDNA were overlaid, a PvuII fragment of approximately 7.0 kb could be aligned. This hybridised to the A500 C4bp probe i.e. in the 3' region of the sequence. Fig 6.4 shows the same blot as Fig 6.2 but hybridised with the Factor H probe. (On the gel the PvuII digest was loaded 8 hr earlier than the remainder in an attempt to increase resolution in the 7.0 kb region.) The acetate overlay in Fig 6.4 shows the bands which could be superimposed between the two autoradiographs. These bands hybridised to the A500 and/or the S1 probes and the 3' Factor H probe suggesting that if the two genes were closely linked it would have to be in the more unusual orientation 5'→3' 3'←5' to allow for the estimated large size of the Factor H gene.

Use of a cDNA probe isolated from the extreme 3' end of the C4bp cDNA (Nco probe, Fig 6.1) to hybridise the same blots suggested that the bands in common originated in the middle of the C4bp gene. The hybridisation results using the Nco probe, (Fig 6.2), demonstrated a 1.9 kb PvuII fragment which hybridised to the Nco probe only whereas the 7.0 kb PvuII fragment hybridised to both the A500 and Nco probes. This indicated that the 7.0 kb fragment lies 5' to the 1.9 kb fragment (this has subsequently been confirmed from the cloned DNA genomic restriction map).

It can thus be concluded that the signals on the blots which appear to be common to both C4bp and Factor H originate from similar sized fragments from different regions of the chromosome. These resolve at the same place under the gel conditions used.

Subsequent work by Rodriguez de Cordoba and Rubinstein, 1987, suggested that the genetic distance between the loci for C4bp and Factor H is between 5 and 10 cM. Although genetic distances (cM) cannot be translated into physical distances (kb) this did suggest that the size of fragments involved in any linkage would be large. This, together with the large sizes of the individual genes involved, suggested that the resolving power of ordinary agarose gels was not adequate to define this linkage. Thus in collaboration with Ian Dunham (MRC Immunochemistry Unit) a long range mapping approach using pulsed field gel electrophoresis was undertaken.

Fig 6.5 and 6.6 show autoradiographs of Southern blots from pulsed field gel electrophoresis, hybridised with the C4bp S1 probe. Details of the enzymes and pulse times are given in the figure legends, otherwise the gels and blots were prepared as described in Sections 2.9., 2.6, 2.4. Subsequent hybridisation of the blots using the factor H probe revealed only one band which could be aligned and this is shown in

Fig 6.5. The size of this NruI fragment was too large to be estimated accurately from this gel. More recently hybridisation of blots from gels with double enzyme digests showed the C4bp probe hybridised to a NruI/NotI fragment approximately 1 M bp in size.

These preliminary results suggested that the range of the restriction map for the RCA locus (including the genes for C4bp and Factor H) was going to extend into the megabase range. Any further mapping was postponed until cDNA probes from other members of the linkage group (CR1, CR2 etc) could be included in the study.

6.5 Interspecies cross-hybridisation of the C4bp cDNA sequence

Genomic DNA was prepared from the white cells of blood from various mammalian and non-mammalian species as described in Section 2.3.6. BglII was then used to digest the various DNA samples together with one human sample and the digests analysed using Southern blotting. The blots were hybridised with the C4bp S1 probe in hybridisation buffer (Section 2.1e) containing 50% formamide and competitor DNA's, at 42°C for 16 hr. The washing conditions used were less stringent than described earlier (Section 2.8.2).

3 times in 2xSSC (Section 2.1e) and 0.1% SDS for 20 mins at room temperature

2 times in 5xSSC for 15 min at 65°C

2 times in 3xSSC for 15 min at 65°C

Similar amounts of DNA were digested for each track and all the digests appeared to be complete on ethidium bromide stained gels. Fig 6.7 demonstrates the hybridisation patterns seen and suggests the presence of sequences related to human C4bp cDNA in most species with the exception of shark. Many of the patterns, with the exception of

that seen with rat DNA, are complex. The human DNA track shows more bands than seen under more stringent conditions suggesting possible cross hybridisation with other related sequences.

These results show that genomic sequence homologous to the human C4bp sequence has been conserved during evolution. It cannot however be concluded that these sequences are active C4bp genes.

6.6. Discussion

The sizes of the genomic DNA fragments and the probes to which they hybridise are compared with the restriction map derived from the overlapping lambda clones in Fig 6.8. This type of comparison allows further confirmation of the restriction map for λBP10 generated initially from overlapping plasmid subclones. Hybridisation of the Nco probe to one BglII fragment (10.5 kb) and PvuII fragment (1.9 kb) which are included in the λBP3'2 clone suggests that this clone extends to include coding for the 3' non-homologous sequence.

Characterisation of the BglII restriction fragment length polymorphism (RFLP) and the hybridisation pattern of the probes to the small fragments (8.5 kb - A500/A400: 2.6 kb - A500) pinpointed the position of the polymorphic site within the cloned genomic DNA. This also confirmed the extent of the overlap between the two lambda clones in the region of exon coding for protein repeat VI.

The distribution of the BglII RFLP seen within the family included in the group of sixteen individuals suggests that this is an inherited rather than acquired trait. The number of DNA samples analysed for the BglII RFLP was relatively small, however representatives of both homozygous patterns (LL, SS) and the heterozygous pattern (LS) were demonstrated. This suggests that within the population as a whole this RFLP

is relatively common. As a result of this observation the RFLP described cannot be correlated directly with the alleles coding for the three protein phenotypes which have been suggested to show gene frequencies of

$C4bp^{*}1$ - 0.986

$C4bp^{*}2$ - 0.010

$C4bp^{*}3$ - 0.004

(Rodriguez de Cordoba and Rubinstein, 1987)

Hybridisation of the Nco probe to the 1.9 kb PvuII fragment and to the very large BamH I fragment suggests that the overlapping signals from the C4bp and Factor H cDNA probes were due to similar sized fragments with separate origins. Subsequent definition of the sizes of the individual genes involved (≅40 kb and ≅90 kb) suggests that it is very likely for the frequent cutting restriction enzymes (as opposed to enzymes used in PFGE) to give fragments of the same size. Recent mapping used pulsed field gel electrophoresis of large genomic fragments generated using rare cutting enzymes and this has suggested that the distance separating the C4bp and Factor H loci is in the mega-base range. Attempts to link the two genes, even using pulsed field gel electrophoresis, will be extremely difficult and inaccurate without access to cDNA probes coding for other members of the linkage group.

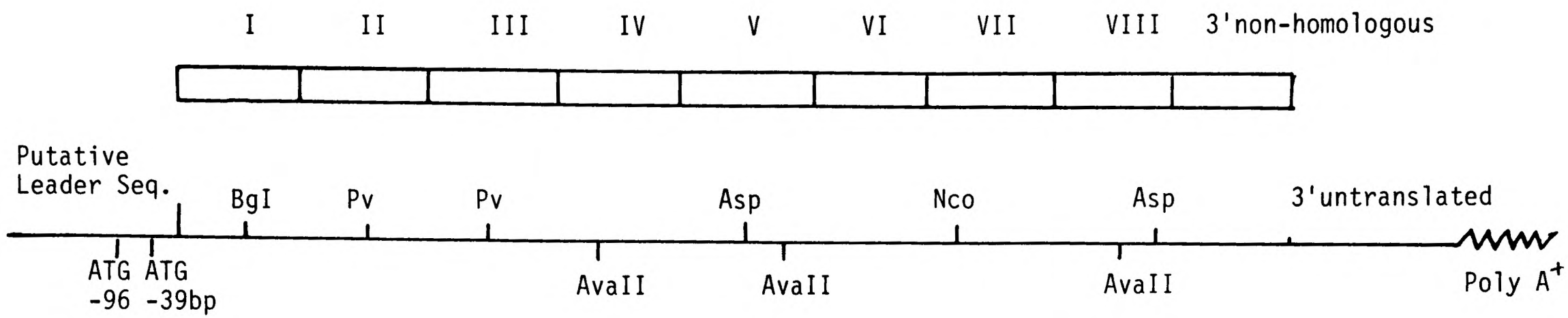
The results of interspecies cross-hybridisation of a human C4bp cDNA probe with sequences in mammalian and non-mammalian genomic DNA suggests that this sequence has been conserved during evolution. The results cannot be used to predict the presence of an active C4bp gene in these species. The only samples which did not hybridise to the probe originated from the shark (chondrichthyes), the most primitive species analysed in this study.

In a recent study by Kaidoh and Gigli, 1987, human C4b and C3b were used as substrates to evaluate the presence of molecules serving the function of C4bp, Factor H and Factor I in the sera of a number of species. The results showed that the serum of mammals, birds, amphibia, reptiles and most fish were able to cleave C4b and C3b to some extent. Considerable quantitative difference in the cleavage efficiencies were noted. The cleavage fragments of human C4b and C3b generated by the sera of the various species are comparable with fragments obtained using the human regulatory proteins. This suggests that the overall structure of C4 and C3 has been conserved through evolution. Serum samples of the more primitive species e.g. sharks and hagfish failed to participate in the cleavage of fluid phase C4b (information on the activity with surface bound C4b and C3b is not given).

These results correlate with the cross-hybridisation patterns (Fig 6.7) seen with the human C4bp probe, an extension of this study to include the use of a human Factor H cDNA probe may help to define the evolution of the two pathways of complement activation.

As mentioned in Section 1.5.1, C4bp has been isolated from two other species only, mouse and guinea pig, and of these only the mouse protein sequence has been determined. The mouse protein sequence was derived from cDNA sequence (Kristensen et al., 1987a) and it shows each monomer to contain six short consensus repeats (cf. eight in the human C4bp monomer). Estimates of the size of the mouse gene based on the results in Fig 6.7 suggest that the gene will be between 20 and 30 kb in size, a feature consistent with each SCR being coded for by a single, discrete exon. Mechanisms for the evolution of the SCR containing proteins based on tandem duplication of a single primordial gene segment can be explained by the addition or deletion of genetic material including one or more exons coding for the SCR's.

Fig 6.1 - cDNA probes used in the blotting analysis - A400 covers coding for protein repeats III, IV and V. A500 covers coding for protein repeats V, VI, VII and VIII. Nco covers coding for protein repeats VII, VIII and 3' non-homologous region S1 covers coding for protein repeats I-VIII and 3' non-homologous region.


I
II
III
IV
V
VI
VII
VIII
3'non-homologous
Putative
Leader Seq.
BgI
Pv
Pv
Asp
Nco
Asp
3'untranslated
ATG ATG
-96 -39bp
AvaII
AvaII
AvaII
Poly A+

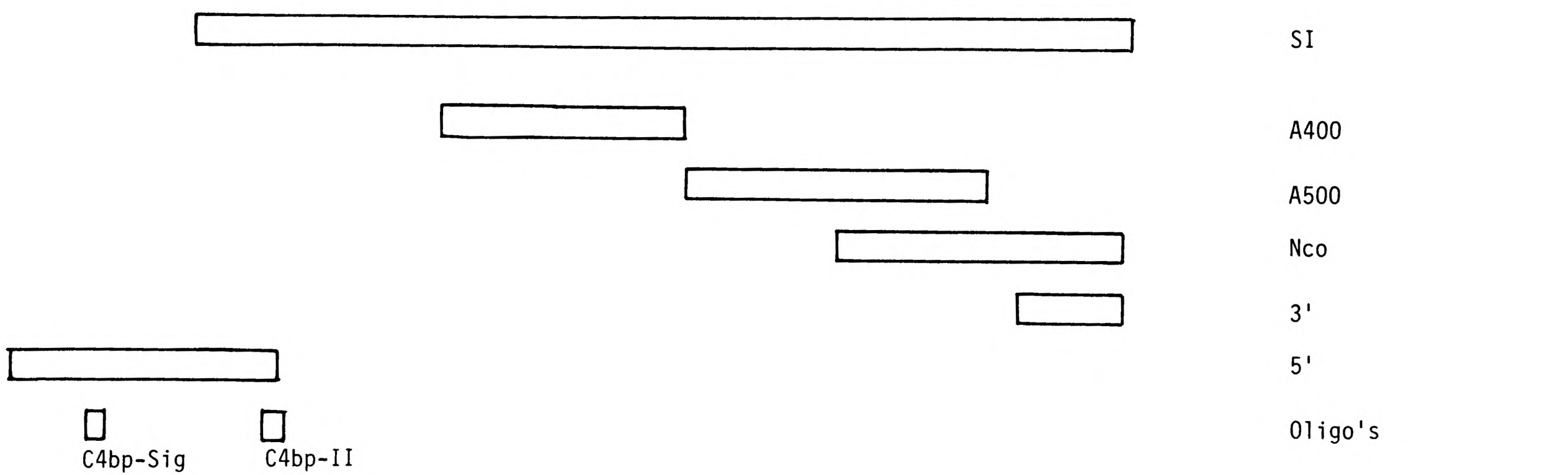

SI
A400
A500
Nco
3'
5'
Oligo's
C4bp-Sig
C4bp-II
0bp
250bp
500bp

Fig 6.2 - Autoradiograph showing the results of Southern blot analysis of genomic DNA isolated from three unrelated individuals (KA; SL; LW). This blot was hybridised with the S1 probe. The acetate overlay summarises the results of hybridisation with the shorter cDNA probes. The PvuII digests and the markers on the right hand side were loaded approximately 8 hr earlier than the remainder to increase the resolution of larger fragments.

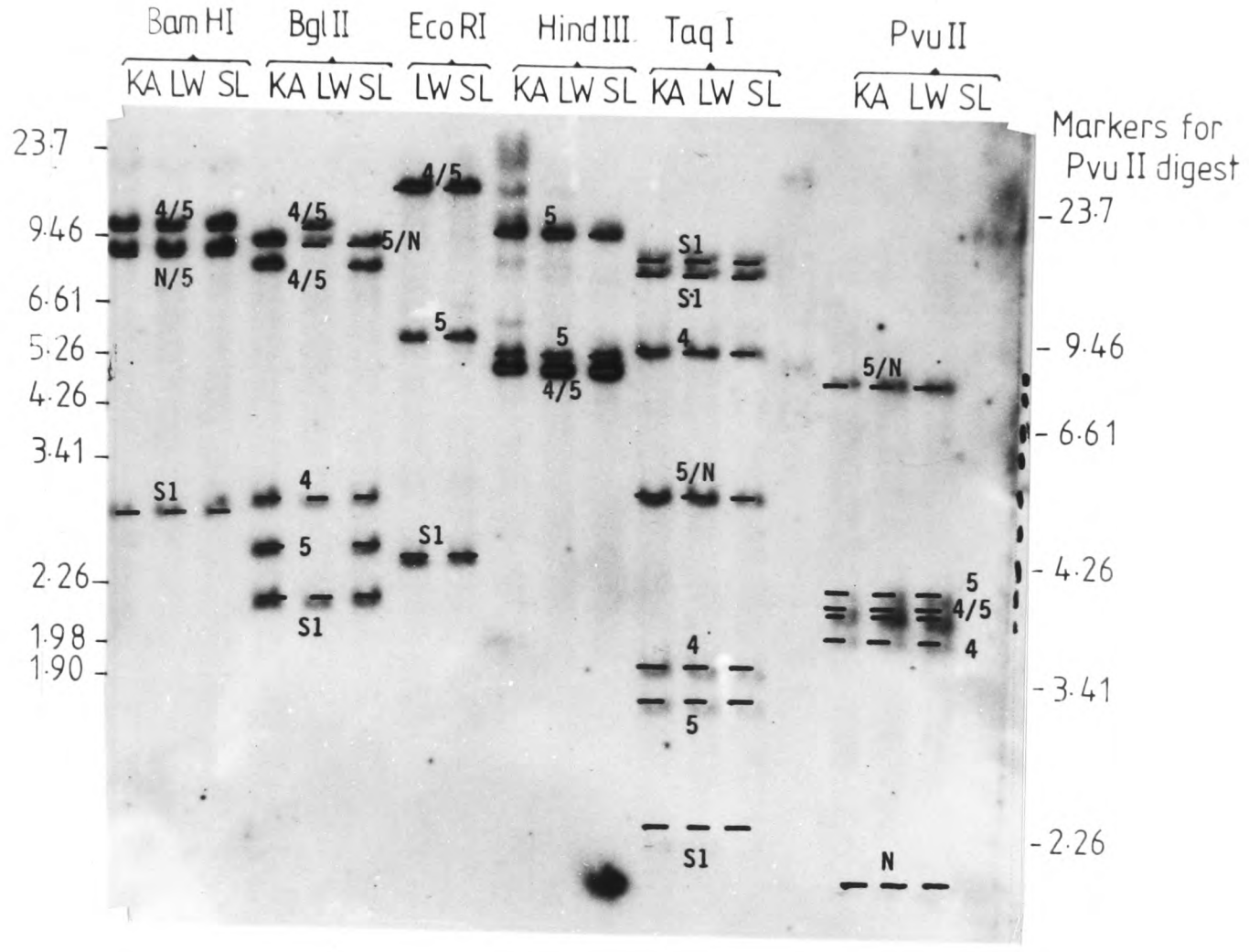

Bam HI
Bgl II
Eco RI
Hind III
Taq I
Pvu II
KA LW SL
KA LW SL
LW SL
KA LW SL
KA LW SL
KA LW SL
23·7
9·46
6·61
5·26
4·26
3·41
2·26
1·98
1·90
4/5
N/5
S1
4
5
5/N
Markers for
Pvu II digest
23·7
9·46
6·61
4·26
3·41
2·26
N

Fig 6.3 - Southern blot analysis of sixteen genomic DNA samples digested with BglII and hybridised with the A500 probe. Three distinct patterns are demonstrated which reflect the homozygous/heterozygous status of the individuals -

LL - homozygous large (11.0 kb fragment)

SS - homozygous small (8.5 kb and 2.6 kb fragments)

LS - heterozygous (11.0 kb; 8.5 kb and 2.6 kb fragments)

The related individuals are marked with M, Mother; F, Father and D, Daughter.

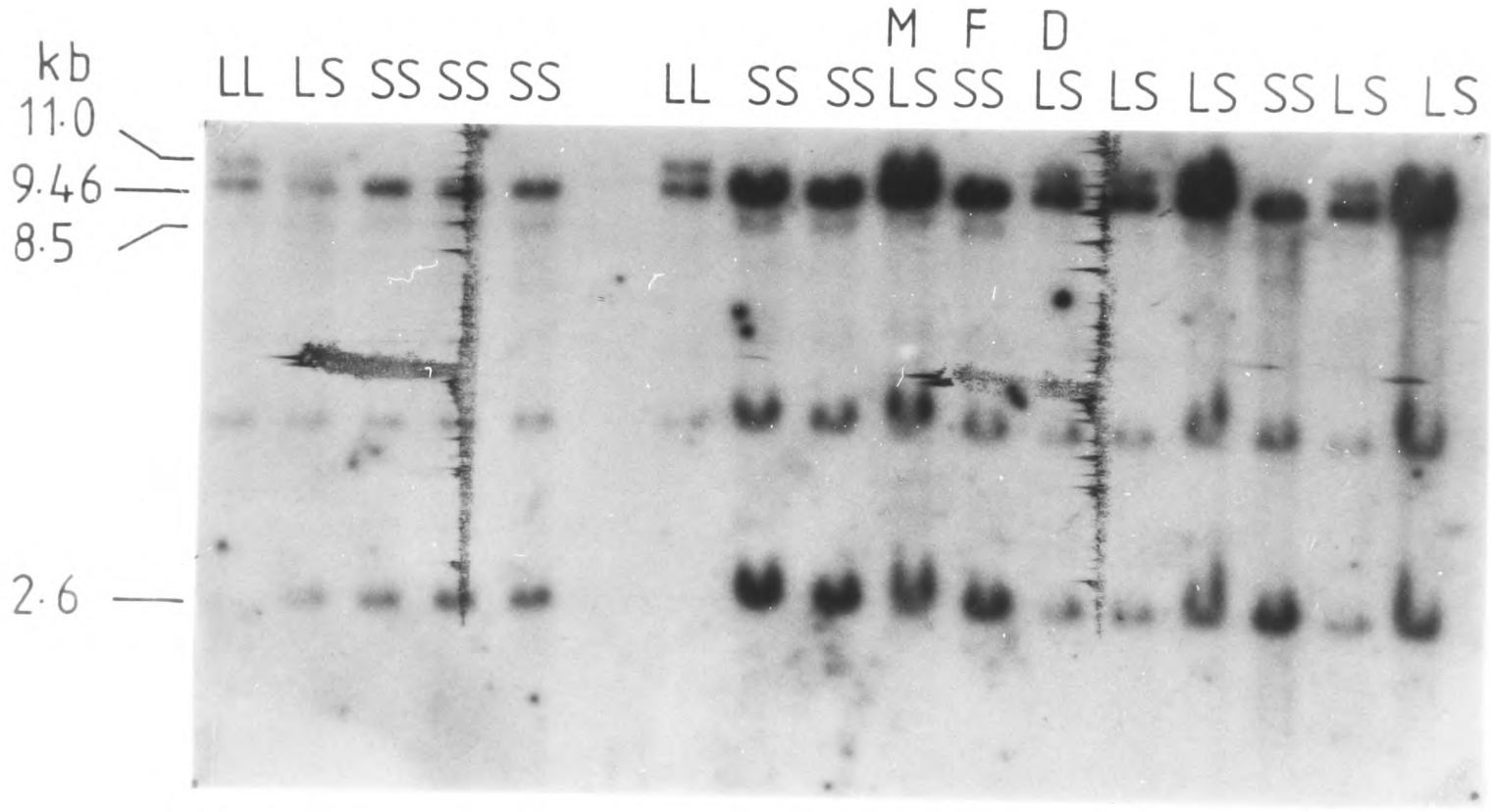

kb
11·0
9·46
8·5
2·6
LL LS SS SS SS
M F D
LL SS SS LS SS LS LS LS SS LS LS

Fig 6.4 - Autoradiograph showing results of the hybridisation of a Southern blot using the Factor H cDNA probe (which covers approximately 50% of the coding sequence). This blot uses the same filter as was prepared for the C4bp cDNA probe in Fig 6.2. (The nitrocellulose filter was not stripped but instead the signal was allowed to decay between probes to prevent distortion of the filter). The acetate overlay shows the signals which appear in common between the C4bp and Factor H probes.

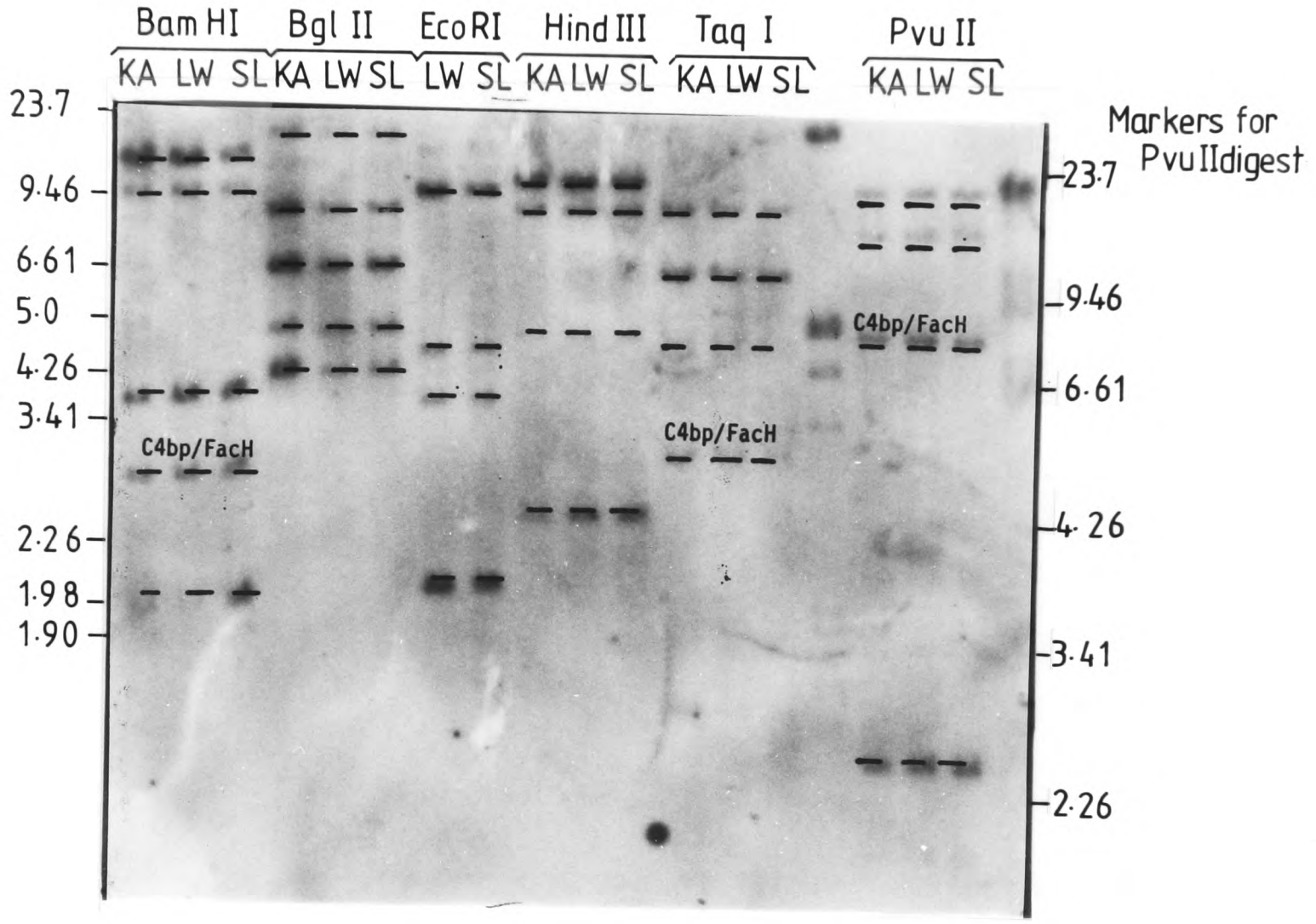

Bam HI
Bgl II
EcoRI
Hind III
Taq I
Pvu II
KA LW SL KA LW SL LW SL KA LW SL KA LW SL KA LW SL
23·7
9·46
6·61
5·0
4·26
3·41
2·26
1·98
1·90
C4bp/FacH
C4bp/FacH
C4bp/FacH
Markers for
PvuIIdigest
23·7
9·46
6·61
4·26
3·41
2·26

Fig 6.5 - Autoradiograph showing preliminary results from the transfer blots of initial pulsed field gels hybridised with S1 probes. The enzymes used all have either CpG dimers within the recognition sequences or have 8bp recognition sites, because of which they are described as rare cutting enzymes and generate larger fragments of genomic DNA. The pulse time used during this separation was 50 sec. The blots were made on nylon filters which were stripped (as described in Section 2.9) between the C4bp and Factor H probes. The large NruI fragment marked also hybridised to the Factor H probe.

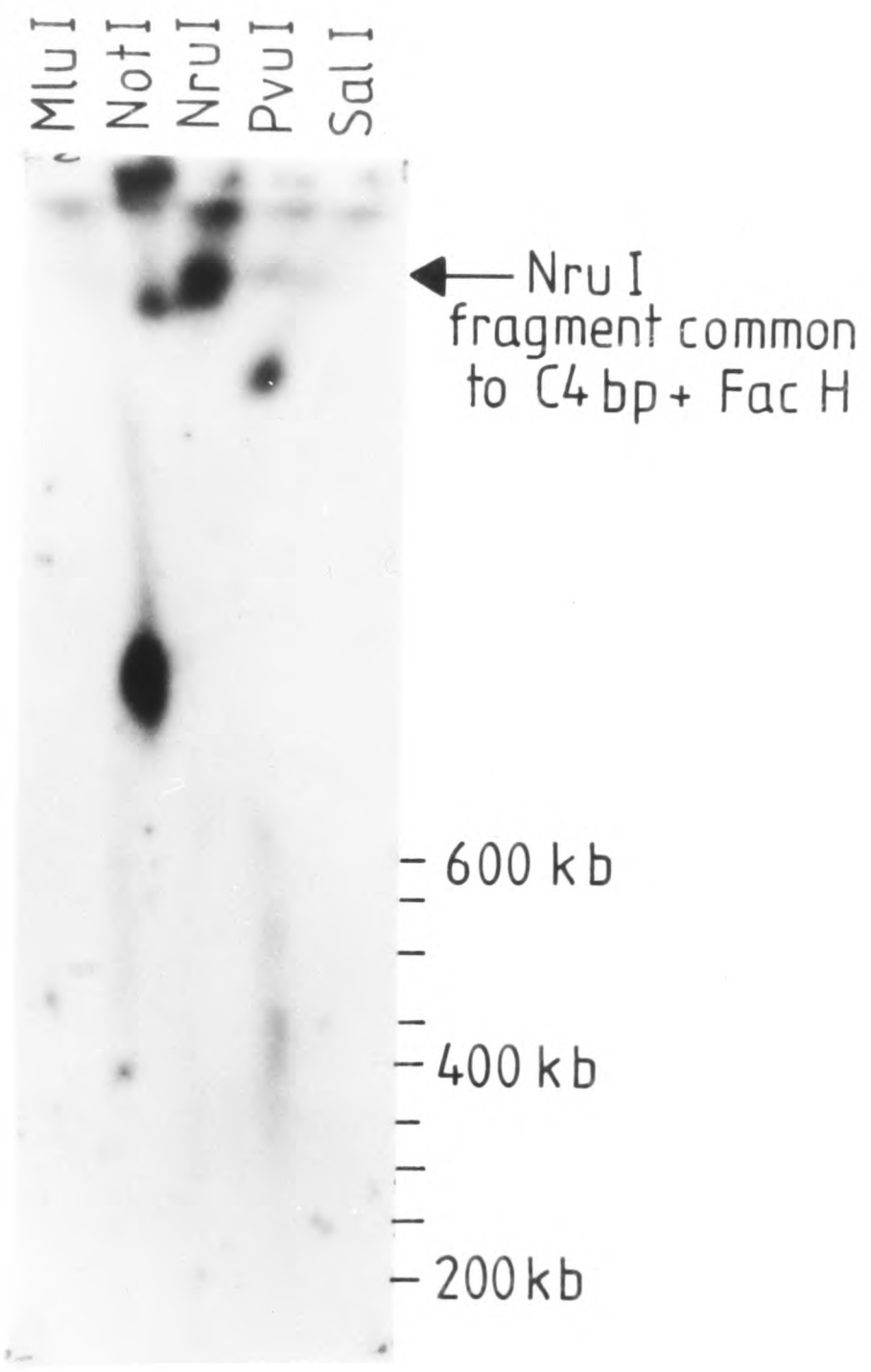

Mlu I
Not I
Nru I
Pvu I
Sal I
Nru I
fragment common
to C4 bp + Fac H
600 kb
400 kb
200 kb

Fig 6.6 - Further pulsed field gel blot again hybridised with the C4bp cDNA probe S1. No common bands were detected with the Factor H probe. The pulse time during this separation was 40 sec.

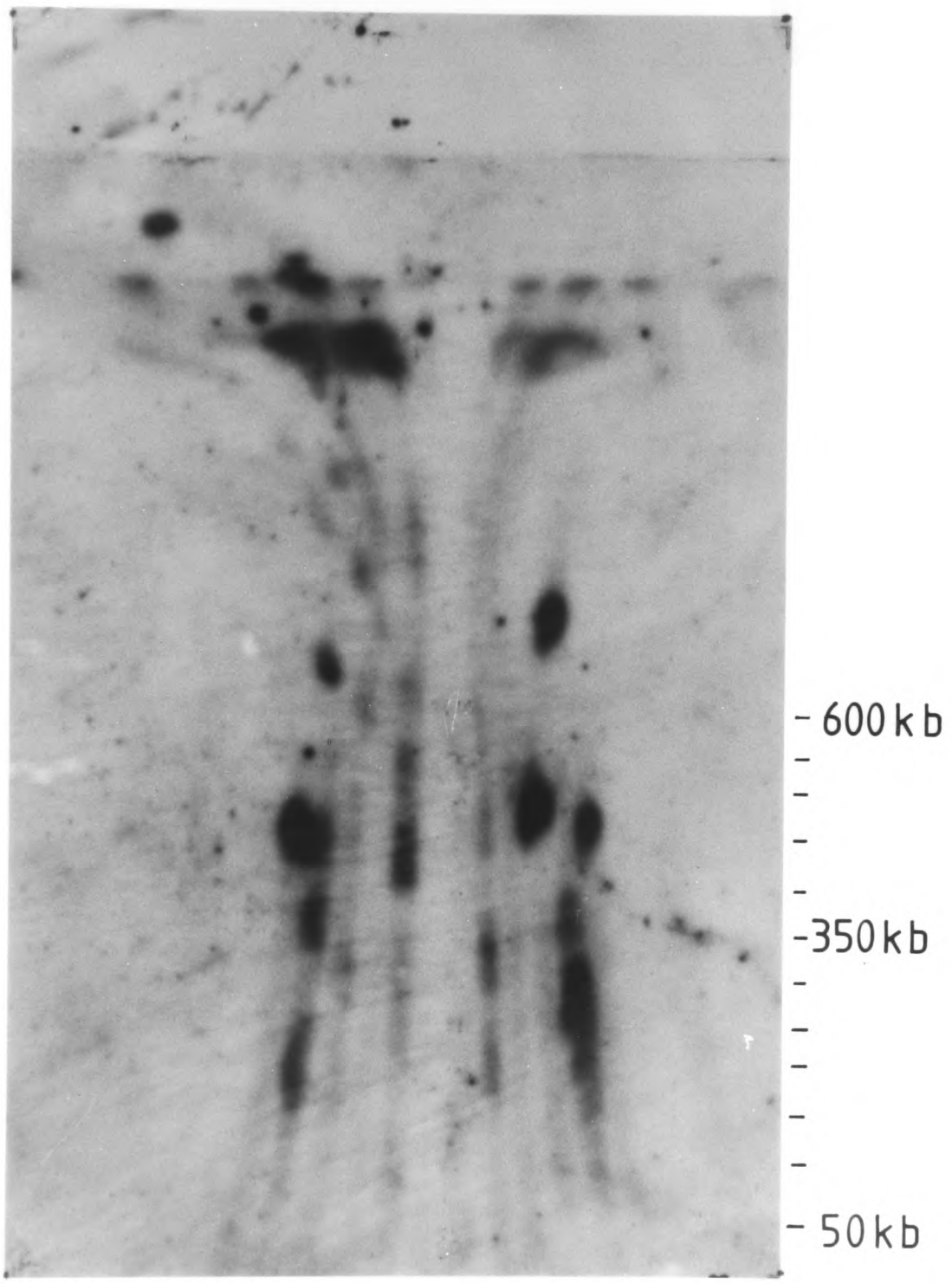

600 kb
350kb
50kb
Bss HII
Cla I
Nae I
Na I
Pae R71
Rst II
Sac II
Sma I
Sfi I
Xho I
Xma III

Fig 6.7 - Results of the interspecies cross-hybridisation using the human C4bp cDNA probe S1 to hybridise genomic DNA isolated from various mammalian and non-mammalian species. The hybridisation and washing conditions are described in the text.

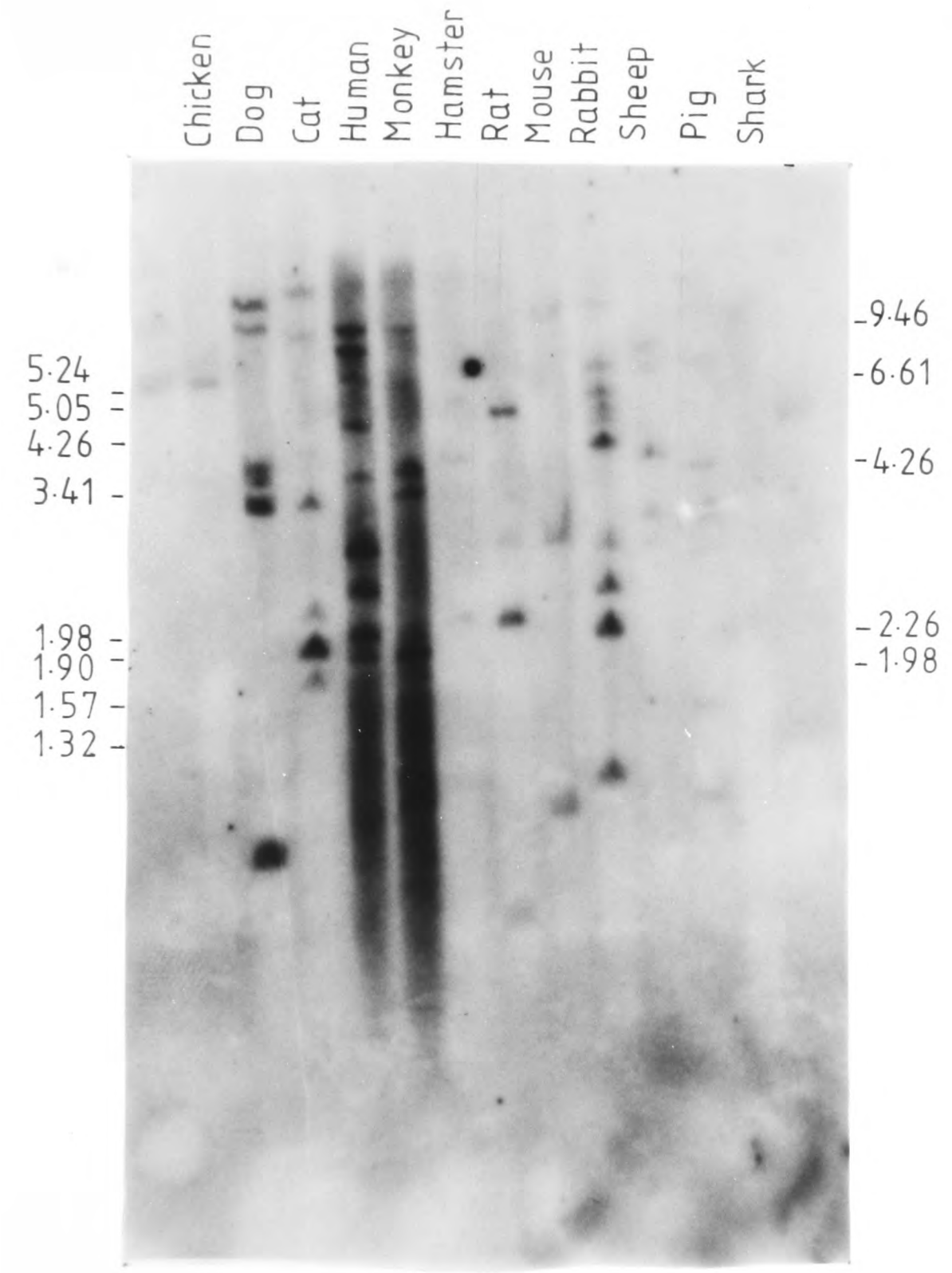
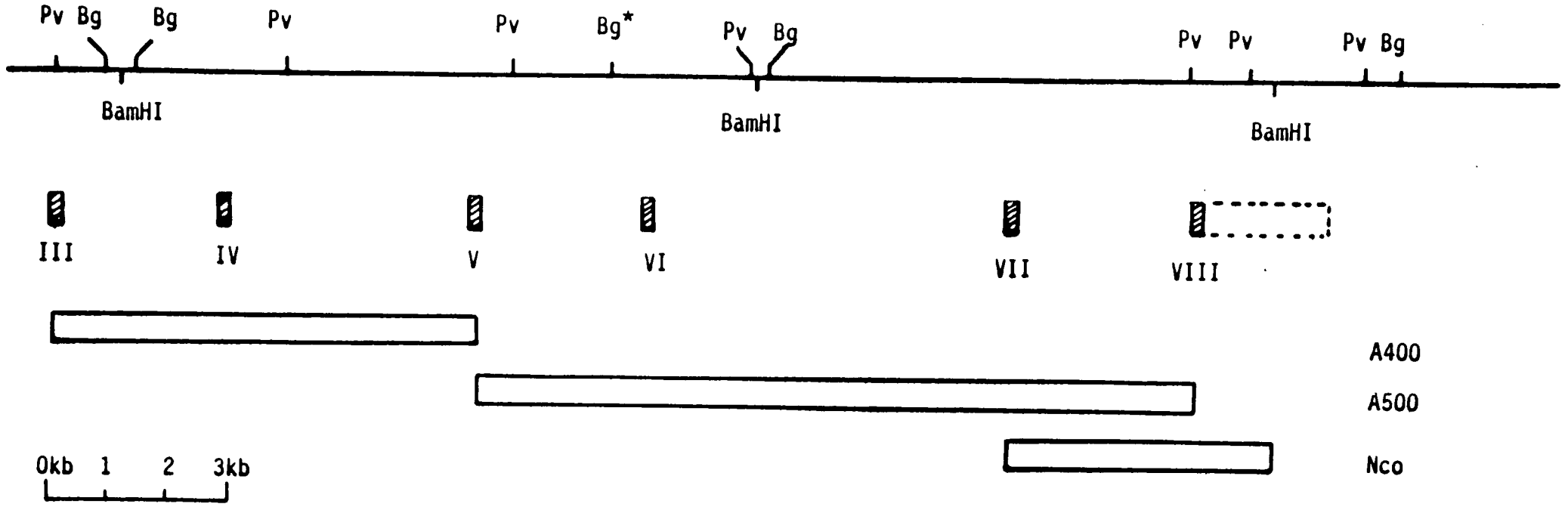

Chicken
Dog
Cat
Human
Monkey
Hamster
Rat
Mouse
Rabbit
Sheep
Pig
Shark
5·24
5·05
4·26
3·41
1·98
1·90
1·57
1·32
9·46
6·61
4·26
2·26
1·98

Fig 6.8 - Shows the restriction map constructed from the overlapping λ-genomic clones together with a summary of the genomic DNA fragments detected during blot analysis. The cDNA probes are shown and the fragments to which they hybridise on Southern blots are also shown. The positions of the exons are marked with hatched boxes and the corresponding protein repeat identified with Roman numerals. The polymorphic site, BglII is marked with *.



Genomic Fragments

PvuII	7.0kb	A500/Nco
	4.3kb	A500
	4.0kb	A400/A500
	3.8kb	A400
	2.0kb	Nco

BamHI	>15.0kb	Nco
	10.7kb	A400/A500
	9.0kb	A500/Nco

Bgl(SmaTl)		10.5kb	A500/Nco
		8.5kb	A400/A500
11.1kb	A400/	3.0kb	A400
	A500	2.6kb	A500

C H A P T E R VII

GENERAL DISCUSSION

7.1 Summary

A cDNA clone, pBP51, was isolated from a human liver cDNA library and found to overlap with clone pBP6, isolated by Chung et al., 1985a, and extended into the 5' untranslated region. This clone included sequence coding for the thirty-second residue, threonine, through to the codon for the first residue, asparagine, thus confirming the N-terminal amino acid sequence reported by Chung et al., 1985c. This sequence included the first three residues of the consensus sequence (cys-pro-cys) in the first internal protein repeat (SCR) of the C4bp monomer. cDNA sequence data 5' to the codon for the first amino acid residue allowed derivation of the putative leader sequence. The AUG codon of the methionine residue at position -32 was assigned as the translational start, yielding a thirty-two amino acid long leader sequence which agrees closely with previously recorded leader sequences (Watson, 1984). The methionine residue at - 32 is preceded by polar residues and followed by a high proportion of hydrophobic residues which is typical of other documented putative leader sequences and reflects the role of the leader sequence during protein synthesis and secretion (Blobel, 1980). The completed cDNA sequence is shown in Fig 7.1.

A cDNA probe was prepared from the 5' region of this clone and used in the subsequent isolation of human C4bp genomic clones. Sequence from this clone was also used to generate unique 17-long oligonucleotides

which are being used to characterise and map the genomic clones covering the 5' region of the C4bp gene.

Preliminary results from the hybridisation of RNA transfer blots, prepared using RNA originating from various cell-lines and tissues, with cDNA probes specific for human C4bp suggests that the C4bp gene is expressed mainly in the liver. The limits of sensitivity of this approach mean that expression of genes at very low levels cannot be ruled out. However, when compared with the levels of mRNA detected in liver tissue, the expression of the C4bp gene in other tissues is likely to be insignificant. The different tissue expressions of C4bp and its analogue in the alternative pathway, Factor H, provide preliminary indications suggesting differential activation of the two complement pathways in local tissues and in generalised infection.

Clone pBP51, together with other cDNA clones, isolated by Chung et al., 1985a, were used to prepare a series of overlapping cDNA probes which together covered the sequence coding for human C4bp. These probes were used to hybridise Southern blots of genomic DNA and to screen genomic libraries to isolate C4bp clones. Initially attempts were made to identify cosmid clones to take advantage of the large inserts cloned into these vectors. This approach has, so far, proved unsuccessful. This is due, in part, to the unworkable background obtained with many of the probes, but also to a possible under-representation of this region of chromosome 1 in the cosmid libraries screened so far.

The same probes have been used to successfully identify, isolate and characterise lambda genomic clones. A total of thirty-seven genomic clones were identified which could be categorised by the hybridisation of the various cDNA and oligonucleotide probes during rescreening. One member from each of four categories was selected to use in the preparation of DNA for further detailed mapping. Three clones overlap

λBP3, λBP10 and λBP3'2 and include approximately three quarters of the coding sequence (internal protein repeats II-VIII and 3' non-homologous 58 residues). It is not demonstrated conclusively that λBP5'12 overlaps with λBP3 but λBP5'12 does include DNA coding for the putative leader sequence based on the hybridisation of the C4bp-sig oligonucleotide.

The clones are summarised in Fig 7.2. The intron distances in the 5' region of the gene map are estimated from Southern blot data and hybridisation of the two oligonucleotides to the genomic clones. Detailed mapping of these clones is in progress and until it is completed these estimates remain inconclusive as no account can be made of large introns using the Southern blot approach. The use of oligonucleotide probes specific for the well defined internal protein repeats is proving very efficient and is demonstrating that the 5' region of the C4bp gene is more extensive than first envisaged.

Detailed mapping and sequencing completed to date has shown that each of the internal protein repeats, defined in the primary amino acid sequence of the C4bp monomer as short consensus repeats (SCR's), are coded for by discrete exons in the corresponding gene sequence. The positioning of the intron junctions within the gene sequence was based on preservation of the "AG-GT" rule (Breathnach and Chambon, 1981) together with a comparison of the cDNA sequence with genomic sequence. This definition of the intron junctions placed some of the repeat boundaries in slightly different positions when compared with those predicted by Chung et al., 1985a. The deletions and insertions at the ends of the repeats created by the changes in boundaries do not affect the residues involved in the SCR consensus sequence.

The organisation of the structural gene coding for mouse C4bp has recently been studied by Barnum et al., 1987, which demonstrates that at least four of the six SCR's defined in the primary amino acid sequence

(Kristensen et al., 1987a), are precisely encoded by separate exons. The second repeat unit appears to be encoded by two exons, based on the presence of sequence conforming to the splice acceptor consensus sequence 5' to the coding sequence for the second half of protein repeat II. The position of the 'split' within the exon sequence corresponds to the position of a similar 'split' in the exon coding for the protein repeat II in the genomic sequence of murine Factor H (Vik et al., 1987). Human haptoglobin precursor was shown to include a region of sequence homologous to human Factor H which corresponded to an SCR except that it lacks one of the highly conserved cysteine residues (Maeda et al., 1984). In the corresponding gene this repeat is split into two exons.

Mapping information on the 5' region of the structural gene coding for murine factor H demonstrates that the exon coding for the putative leader sequence is separated from the exon coding for protein repeat I by an intron approximately 13 kb in length (Vik et al., 1987). The preliminary mapping data of the corresponding region of the human C4bp gene has revealed an unexpectedly extended sequence suggesting that here too a large intron may separate the exons coding for the putative leader sequence and protein repeat I.

Hybridisation of genomic Southern blots with the series of cDNA probes has allowed confirmation of the cloned DNA restriction map and also investigation of a restriction fragment length polymorphism (RFLP) defined by the hybridisation pattern of BglII digests. The high frequency at which the RFLP was detected in a very small sample of individuals suggested that it was not linked with the protein polymorphism described by Rodriguez de Cordoba and Rubinstein, 1987.

Attempts to physically link the structural genes for the regulatory proteins C4bp and Factor H used the hybridisation of cDNA probes to identify common genomic DNA fragments. This showed that the genes were

not very close in physical distance unlike the structural genes coding for the activation proteins C2 and Factor B (Wu et al., 1987). Extension of this approach to use pulsed field gel electrophoresis has suggested that the physical distance involved in the linkage between the two regulatory protein genes is in fact in the Mbp range. Any detailed mapping of this region of chromosome 1 will thus require extensive use of cDNA probes specific for genes lying between these two loci.

Use of cDNA probes specific for human C4bp in the hybridisation of genomic Southern blots, on which the genomic DNA originated from various mammalian and non-mammalian species, suggested that genomic sequence homologous to the human C4bp sequence has been conserved in evolution. The most primitive species included in this study was shark and this showed no cross-hybridisation with the human cDNA probe suggesting that the sequence has diverged too far to support cross-hybridisation. These results correlate with a functional study of the regulatory proteins in various species reported by Kaidoh and Gigli, 1987. The results of the hybridisation study do not conclusively prove the presence of active C4bp genes in these varied species.

7.2 Evolution of the C4bp gene and other SCR containing genes

One of the most striking features revealed in the amino acid sequence of C4bp is the internal sequence homology which divides the N-terminal 491 amino acid residues into eight repeating units (now referred to as SCR's). A mechanism of contiguous duplication was proposed to describe the elongation of the protein sequence in the C4bp monomer (Chung et al., 1985a). Recently a study to determine the disulphide-bonding pattern in the C4bp monomer has indicated that intra-chain disulphide bonds are restricted within each SCR (Janatova et

al., 1987). This implies that each SCR can fold independently and hence each internal duplication thus represents a protein domain as defined by Doolittle, 1985.

As described in the preceding chapters of this thesis the domain structure of the protein is reflected in the intron distribution in the gene coding for the human C4bp monomers. A similar intron distribution has been demonstrated in other proteins which have highly repeated domain structures, for example the proteins constructed from varying numbers of copies of the immunoglobulin fold. The mechanism proposed for the evolution of these repeat structures involves unequal crossing over between homologous chromosomes during meiosis. This gives rise to non-reciprocal recombinant pairs of chromosomes, one half of which includes an insertion of one or more repeated sequences and the other includes a corresponding deletion of sequence (Fig 7.3).

This exchange of material relies initially on the misalignment of paired chromosomes during meiosis followed by recombination of homologous sequences on the individual chromosomes. This pairing of homologous sequence can only occur if two copies of the same sequence are presented in tandem on the chromosomes and the mechanism which generates this initial duplication is not well understood. Suggestions put forward are reviewed by Maeda and Smithies, 1986, and include RNA mediated transfer and recombination between repeated elements placed on the 5' and 3' sides of the primordial segment.

As the number of repeat units increases on the chromosomes so the frequency of homologous recombination increases to give rise to a rapidly expanding population of the repeated sequences aligned down the chromosome. This type of mechanism is used to explain the evolution of multigene families where the duplication of complete genes or groups of genes occurs whereas in the evolution of the C4bp gene the repeated unit

consists of one or more exon sequences.

The conservation of sequence in the exons compared with the divergence of intron sequence usually indicates that homologous recombination occurs between exon sequences. This suggests that the presence of interspersed intron sequence decreases the probability of this mechanism being supported. The large size of the introns found in the C4bp gene and the obvious success of the mechanisms to generate multiple copies of the exons coding for the SCR suggest that either the presence of interspersed intron sequence has little effect on the mechanism or recombination is occuring between intron sequence, possibly between repetitive sequences e.g. Alu.

The intron/exon organisation described for the C4bp gene supports the theory that it has evolved by a mechanism of duplication of tandemly repeated exon sequences. This same mechanism is used to described the evolution of clusters of genes such as the histone and tRNA genes. The affects of this tandem duplication of repeating units on the overall restriction map for the region of the genome is described by Lewin, 1985. This suggests that the characteristic map pattern will be circular if the repeating units are identical. It is unlikely that the repeat unit sequence in the intron regions has been conserved completely in the C4bp gene, however, repetition of restriction sites can be detected which, together with the conserved pattern of intron sizes shown in Fig 7.2, may reflect the evolutionary mechanisms of the development of this gene.

The evolution of other proteins consisting almost exclusively of repeated SCR domains can be explained by the same mechanism. This group includes the other C3 convertase regulatory proteins, Factor H (Day et al., 1987a), CR1 (Klickstein et al., 1987) and DAF (Caras et al., 1987) together with CR2 (Weis et al., 1987a) a complement receptor found on B

cells which has no known regulatory role. Non-complement proteins included in this group are β_2 glycoprotein I (Lozier et al., 1984) and Factor XIII β subunit (Ichinose et al., 1986). These proteins together with other SCR containing proteins are summarised in Fig 7.4. The C-terminal regions of all these proteins, with the exception of Factor H, show no homology to the SCR's, a situation already described for C4bp. The functions of these non-homologous regions are attributed to membrane-spanning and cytoplasmic portions in DAF, CR1 and CR2 but remain uncharacterised in β_2 glycoprotein I and Factor XIII β subunits.

Initial results suggest that the SCR's in human Factor H are encoded by discrete exons (M. McAleer, personal communication), and this has also been demonstrated in the murine Factor H gene (Vik et al., 1987). It would thus appear that the genomic organisation of the C4bp gene is characteristic of the genes coding for the family of proteins consisting almost exclusively of SCR's. The mechanisms of evolution proposed for the C4bp gene can thus be extended to describe the evolution of the family of genes as a whole. Divergence of sequence not affecting the SCR consensus sequence has permitted the divergence of functions during evolution of the group.

The regions coding for the cytoplasmic and membrane-spanning sequences in the CR1, CR2 and DAF proteins are likely to be carried in separate exons, if they follow the precedent set by other membrane-bound proteins including immunoglobulins. These proteins together with C4bp suggest that evolutionary mechanisms other than straight-forward duplication may have occurred to include, within the gene structure, sequence coding for other functional and structural domains.

This type of exon shuffling is taken one step further in the genes coding for C2 (D. Bentley, unpublished) and Factor B (Campbell and Porter, 1983). In the genes coding for these two activation proteins

the three exons coding for the SCR's are spliced to exons encoding the catalytic residues which form typical serine protease active sites. A similar situation is predicted for the genes encoding C1r and C1s. Both these pairs of genes appear to have arisen by a mechanism which results in exon shuffling followed by gene duplication. The evolution of these genes thus supports the hypothesis of gene evolution occurring by the joining of exons coding for discrete protein domains by combination of the intron sequences. Exon shuffling is thus able to generate genes coding for novel proteins with diverse functions from pre-existing exon sequences (Gilbert, 1985; Blake 1985 ; Doolittle, 1985).

Indications that the evolution of the CR1 gene may involve a more complex mechanism than straightforward duplication of exon sequences has been obtained by Klickstein et al., 1987, during cDNA sequencing studies. This revealed that in the primary sequence of CR1, groups of seven SCR's are organised into second order repeats called Long Homologous Repeats (LHR's) which are subsequently duplicated. Analysis of genomic clones and RFLP's in the CR1 gene (Wong et al., 1987) suggest that this overall organisation of SCR's into LHR's is reflected in the gene organisation. Thus unequal crossing over events have involved widely dispersed homologous sequences in this region of the chromosome in order to duplicate long segments of genomic DNA (up to 25 kb) intact.

Sequencing of the cDNA coding for the F-allele gene product identified the presence of four LHR's, A, B, C and D. Comparison of the homologies between the amino acid sequences of these repeats demonstrated that LHR-C is a hybrid of LHR-B and D. This suggests that a mechanism of gene conversion has occurred during the evolution of CR1. The term gene conversion in this context is used to described the local transfer of DNA sequences from one gene to a related gene, an event that resembles a double cross-over or branch migration event. Mechanisms to

account for this overall effect are reviewed by Maeda and Smithies, 1986.

The definition of the LHR in the primary structure of CR1 has allowed the definition of the size polymorphism of CR1 in molecular terms. The F allele being represented by four copies of the genomic sequence coding for the LHR and the S allele by five copies of the LHR sequence (Wong et al., 1986).

7.3 The Regulators of Complement Activation linkage group (RCA)

A linkage between the loci coding for the structural genes of human C4bp, CR1 and Factor H was established by Rodriguez de Cordoba et al., 1985, and this lead to the definition of the RCA linkage group. C4bp and CR1 are encoded by two loci within the group which appear to be very closely linked, based on a strong linkage disequilibrium between their alleles (Rodriguez de Cordoba et al., 1984). Subsequently the control locus for the expression of CR1 was linked to the RCA (Rodriguez de Cordoba & Rubinstein, 1986) and speculation that genes coding for other complement receptors and regulatory proteins were also linked to this group began.

The identification of a partial cDNA clone coding for CR2 (Weis et al., 1986) allowed preparation of a specific probe. When this probe was used in in situ hybridisation studies it was demonstrated that the structural gene coding for CR2 mapped to the q32 band on chromosome 1 (Weis et al., 1987b). This provided evidence that the CR2 locus is included in the RCA linkage group. Rey-Campos et al., 1987, reported the presence of an RFLP in the gene coding for DAF and used this to formally link the DAF locus to the RCA group. Within the cluster the DAF locus is closer to the CR1/C4bp loci than to the Factor H locus.

Analysis of the murine complement receptor gene family (Weis et al., 1987c; Aegerter-Shaw et al., 1987) has identified two related subfamilies of murine genes which hybridise with CR1 and CR2 specific cDNA probes. One subfamily consists of three distinct genes, one of which may be a pseudogene and the gene products of the other two are as yet unknown. The second subfamily is likely to contain the gene sequences coding for murine CR1 and CR2 based on mRNA size and tissue distribution. The genes coding for CR1 and CR2 have been mapped to the RCA group also on chromosome 1 in the mouse system. Thus the definition of a physical linkage between the two subfamilies using pulsed field gel electrophoresis suggests that both subfamilies are members of the RCA group.

Similarly, isolation of genomic clones coding for murine Factor H has identified three distinct clusters of clones, one of which codes for the full length Factor H protein. The other clusters code for H-related sequences thought to represent other genes in the superfamily (Vik et al., 1987). These sequences have also been mapped to the murine RCA group.

The RCA gene cluster has developed into a complex multigene family consisting of at least six loci with the prospect of more members being included in the future when gene products for the CR1 and Factor H-related genomic sequences are identified in both human and mouse systems. The initial results from the pulsed field gel electrophoresis study using cDNA probes specific for C4bp and Factor H, which suggested that the C4bp and Factor H loci can be mapped only in the Mbp range, are confirmed. Using cDNA probes specific for C4bp, CR1 and CR2 and DAF Carroll et al., 1987, have mapped the corresponding genes and demonstrated a physical linkage between the loci using pulsed field gel electrophoresis. It is reported that all four genes are clustered on a 950

kb Not/Nru fragment (cf. Section 6.4) and that the cluster is split by Mlu to give two closely linked pairs of genes; CR1 and CR2 on a 450 kb fragment and C4bp and DAF on a 500 kb fragment.

The RCA gene cluster thus consists of a minimum of six loci coding for the structurally and functionally related proteins. The cluster appears to cover between 2-3 Mbp of chromosome 1. Each individual gene appears to consist of varying numbers of copies of the tandemly repeated exons coding for the short consensus repeat protein domains. Preliminary results from genomic clones covering genes coding for some of the group members, Factor H and CR1 suggest that their intron/exon organisation will be similar to that described for the human C4bp gene. If this is found to hold for all the members of the group the intron/ exon distribution in this region of chromosome 1 will show an unusually regular distribution of similar sized exons within long sections of intron sequence.

The organisation of the genes coding for proteins with similar functions, involved in the production of C4b/C3b fragments and/or acting as receptors for C3b fragments appears to reflect their evolutionary origins. To date there is very little information to suggest that the expression of the genes is co-ordinated in any way. Extensive studies will be necessary to reveal any orchestration of expression in different tissues and under the influences of different cytokines. The availability of cDNA probes specific for each member of the group will make this type of study a more viable project.

7.4 Conclusion and final comments

The main aim of this project was to examine any correlation between the internal repeat units (SCR's) seen in the primary sequence of C4bp

with the exon organisation of the corresponding gene. The isolation and characterisation of genomic clones covering most of the coding sequence for C4bp has shown that, with one possible exception, each SCR is encoded by a discrete exon. This has allowed discussion of the molecular mechanisms involved in the evolution of the C4bp gene in terms of non-reciprocal recombination following unequal crossing of parental chromosomes at meiosis. If it is assumed that the genes encoding other SCR containing proteins maintain this exon organisation the evolutionary discussion can be extended to include mechanisms of "gene conversion" in the CR1 gene and exon shuffling in the C2, Factor B, C1r and C1s genes.

The definition of the SCR as an independently folding protein domain together with its representation at the genomic level in a discrete exon allows discussion of its participation in the major rearrangements described by Gilbert, 1985, to generate new genes encoding novel proteins from pre-existing diverse functional units.

The linkage of the genes encoding C4bp together with other complement receptors and regulatory proteins on chromosome 1 probably reflects the evolutionary origins of the genes. The genes of the RCA linkage group code for proteins which are involved in control of the activation pathways, formation of C3b fragments and as receptors for the fragments. It is thus attractive to speculate that mechanisms co-ordinating the expression of these genes exist, this is in no way proven. Detailed mapping of the region of chromosome 1 covering the complete RCA linkage group is a major project which if undertaken will encompass many of the most recent innovations in molecular biology techniques including pulsed field gel electrophoresis and efficient chromosome walking techniques.

Fig 7.1 - The complete cDNA and amino acid sequence for human C4bp. The universal translation initiation codons (ATG) at position -32 and -13 and the termination codon (TAA) are marked with *. The putative signal peptide is underlined and the first residue of the mature, secreted protein chain, asparagine is boxed.

```
                                                
```

Fig 7.2 - Summary of the λ genomic clones covering the gene sequence for human C4bp. Precise intron sizes in the 5' region of the gene remain to be confirmed in the clones λBP5'12 and λBP3. Preliminary results from clone λBP5'12 suggest that it contains coding for the signal peptide only. The restriction enzyme sites marked in the gene map correspond to those in the cDNA and were used to pinpoint exons and in the sequencing. The exons are represented by hatched boxes and the Roman numerals correspond to the internal protein repeats shown diagramatically above the cDNA restriction map.

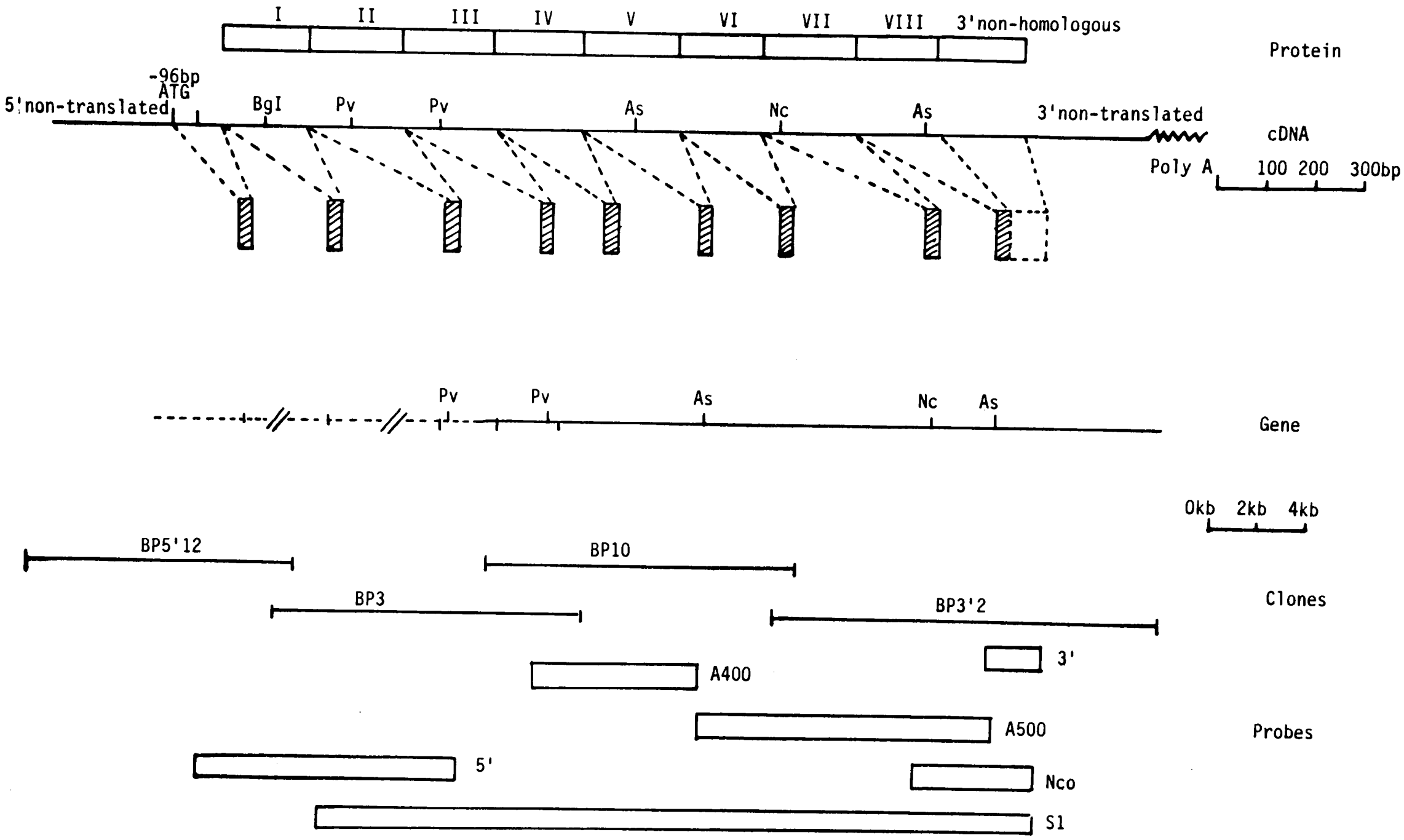
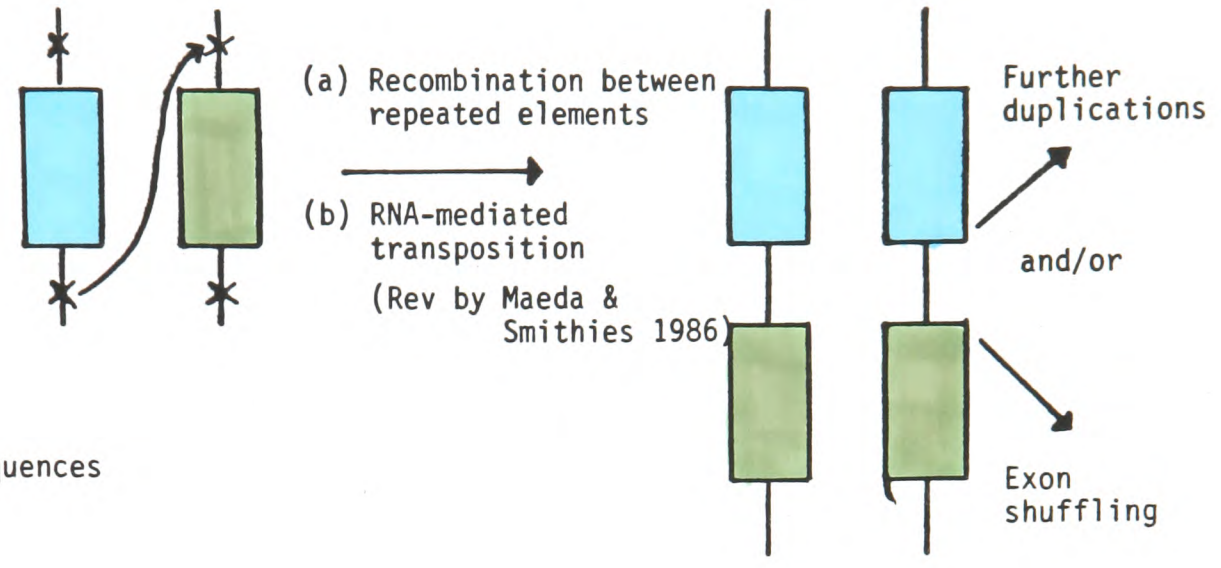

I
II
III
IV
V
VI
VII
VIII
3'non-homologous
Protein
-96bp
ATG
5'non-translated
BgI
Pv
Pv
As
Nc
As
3'non-translated
cDNA
Poly A
100 200 300bp
Pv
Pv
As
Nc
As
Gene
0kb 2kb 4kb
BP5'12
BP10
BP3
BP3'2
Clones
3'
A400
A500
Probes
5'
Nco
S1

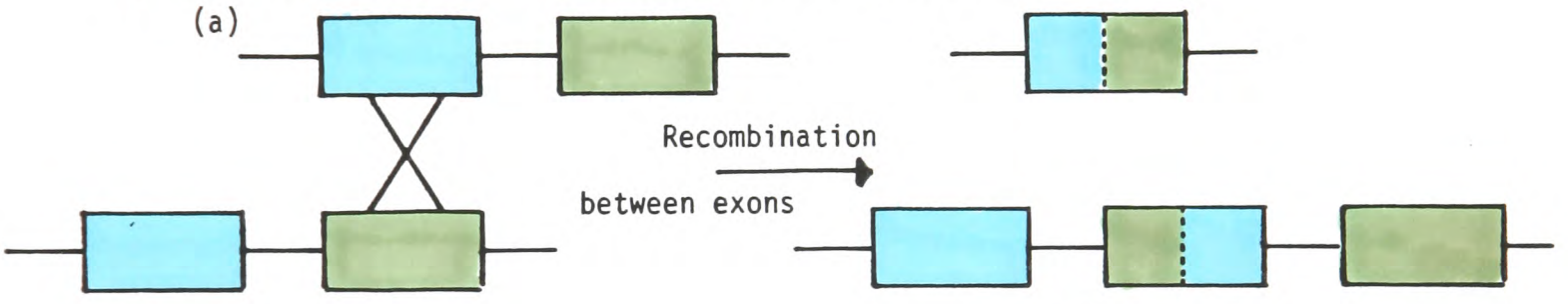
Fig 7.3 - Demonstrates diagramatically the mechanisms involved during the evolution of the C4bp gene and the multigene family of SCR containing genes. Each box represents one "repetitive unit" consisting of one or a group of exons encoding SCR's.

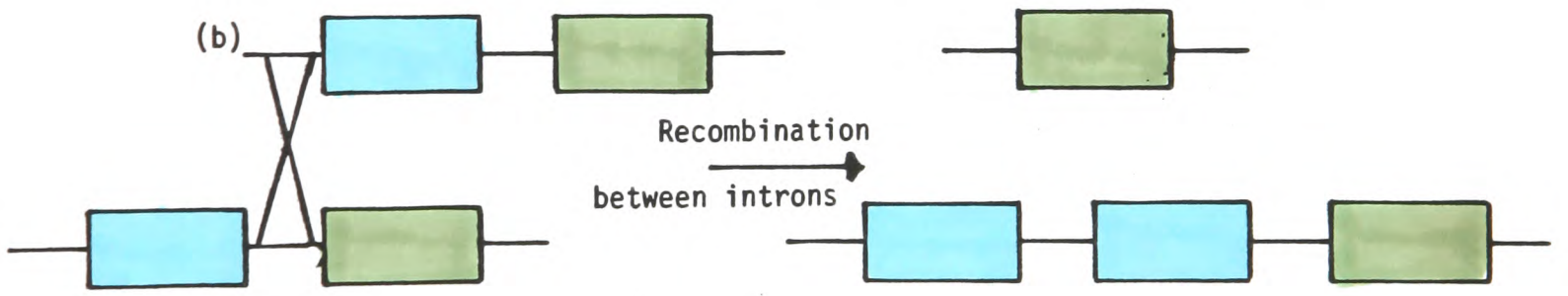
(i) Initial Duplication Event



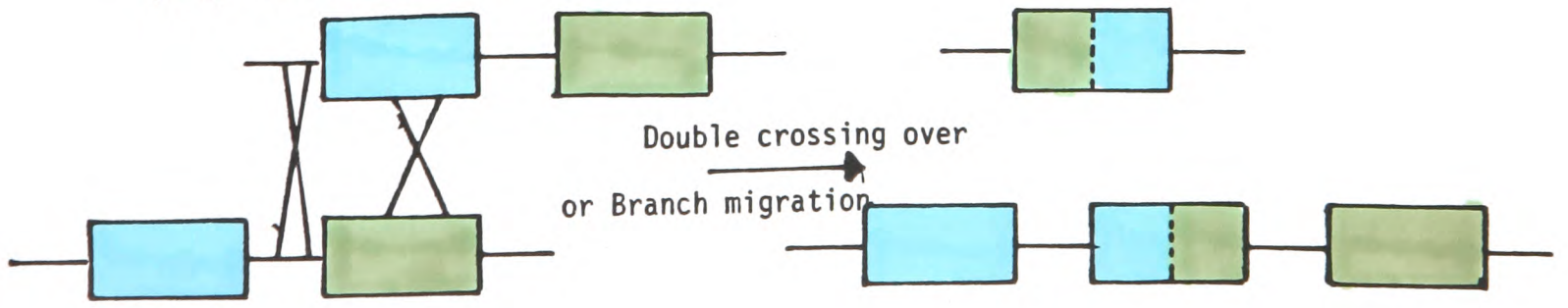
X - Repetitive Sequences
eg Alu

(ii) Unequal crossing over between mispaired parental chromosomes





(iii) "Gene conversion" in CR1



(compare with (ii a)

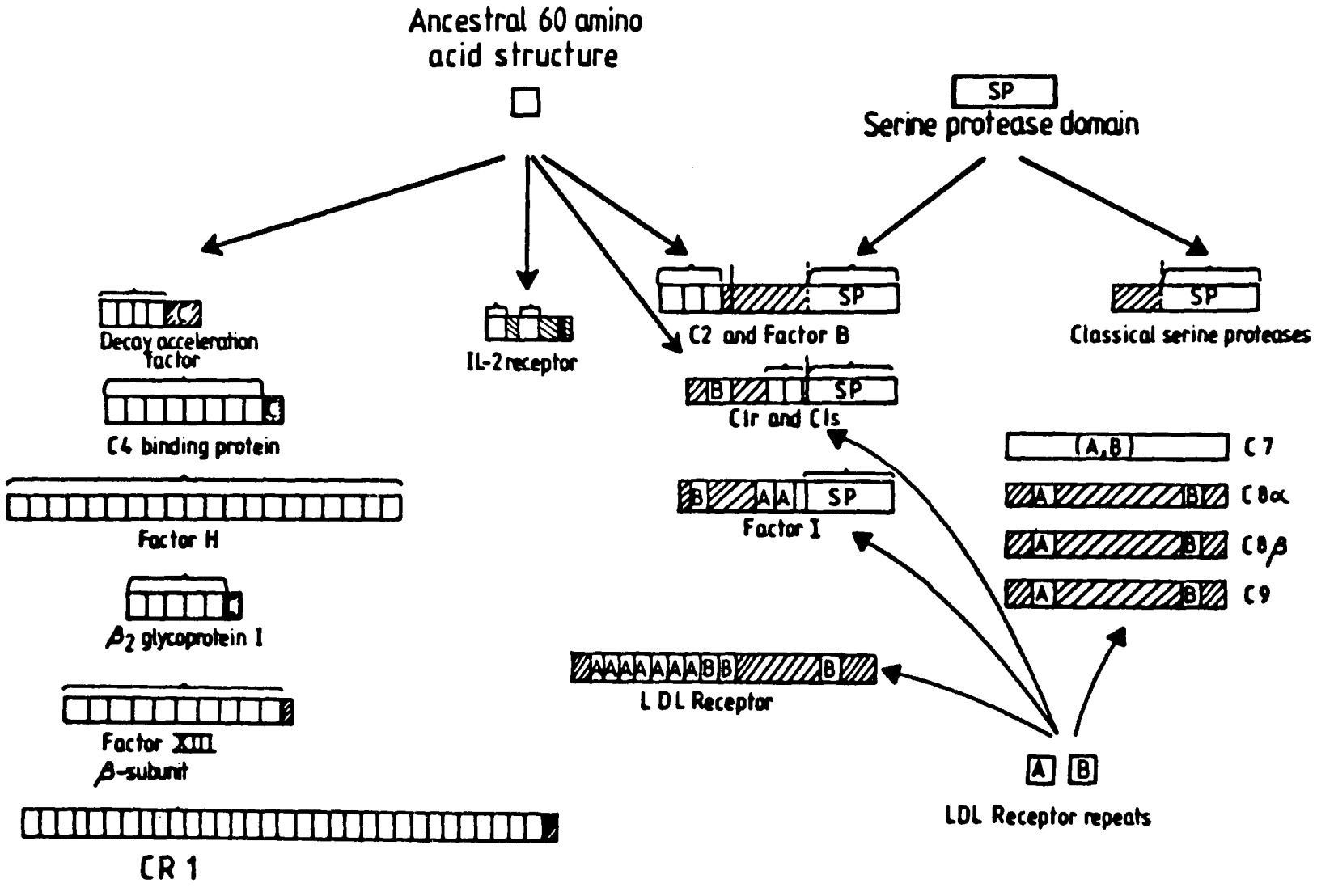
Fig 7.4 - Occurrence of the short consensus repeat (SCR) within the amino acid sequences of complement and non-complement proteins.

The SCR (depicted by □) described in the primary structure of C4bp (Section 1.5.1) occurs in Factor H, DAF, CR1, Factor B, C2, C1r and C1s, together with CR2 not shown in this diagram and also in non-complement proteins IL-2 receptor, β_2 glycoprotein I, Factor XIII β-subunit.

Serine protease units as seen in trypsin occur in C2 and Factor B, C1r and C1s, Factor I and Factor D. LDL receptor repeat structures type A occur in Factor I, C7, 8 and 9 and perforin. The LDL receptor class B repeat (the EGF domain) is in Factor I, C7, 8 and 9, C1r and C1s.

Diagonally-shaded sections represent sequences unrelated to domains discussed above.

The diagram comes from Campbell et al., 1987.


Ancestral 60 amino acid structure
SP
Serine protease domain
Decay acceleration factor
C4 binding protein
Factor H
β2 glycoprotein I
Factor XIII β-subunit
CR 1
IL-2 receptor
C2 and Factor B
SP
Classical serine proteases
B
SP
C1r and C1s
B A A SP
Factor I
A B
LDL Receptor
A B
LDL Receptor repeats
(A,B) C7
A B C8α
A B C8β
A B C9

References

Aegerter-Shaw, M., Cole, J.L., Klickstein, L.B., Wong, W.W., Fearon, D.T., Lalley, P.A. and Weis, J.H. (1987) J.Immunol. 138, 3488-3494.

Alper, C.A., Abramson, N., Johnston, R.D., Jandl, J.H. and Rosen, F.S. (1970) New Engl.J.Med. 282, 349-353.

Anand, R. (1986) TIG 2, 278-283.

Anson, D.S., Choo, K.H., Rees, D.J.G., Gianelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1984) EMBO J. 3, 1053-1060.

Arvieux, J., Yssel, H. and Colomb, M.G. (1987) Complement 4, 124-224 (Abstr. 12).

Bacle, F., Haeffner-Cavaillon, W. and Kazatchkine, M.D. (1987) Complement 4, 124-244 (Abstr. 18)

Barnum, S., Kenney, J., Kristensen, T., Noack, D., Seldon, M., D'Eustachio, P., Chaplin, D. and Tack, B. (1987) Complement 4, 124-224 (Abstr. 20).

Barrett, D.J. and Boyle, M.D.P. (1984) J.Paedr. 104, 76-81.

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.

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

Bernards, R. and Flavell, R.A. (1980) Nucl. Acid Res. 8, 1521-1534.

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) Nucl. Acid Res. 7, 1513-1523.

Blake, C.C.F. (1985) Int.Rev. of Cytol. 93, 149-185.

Blobel, G. (1980) Proc.Natl.Acad.Sci.USA 77, 1496-1500.

Bottger, G.C., Hoffman, T., Metzger, S., Hadding, U. and Bitter-Suermann, D. (1985) Complement 2, 12 (Abstr. 29).

Breathnach, R. and Chambon, P. (1981) Ann.Rev.Biochem. 50 349-384.

Burge, J., Nicholson-Weller, A. and Austen, K.F. (1981) J.Immunol. 126, 232-235.

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

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

Campbell, R.D. and Bentley, D.R. (1985) Immunol.Rev. 87 19-37.

Campbell, R.D., Law, S.K., Reid, K.B.M. and Sim, R.B. (1987) Ann.Rev. Immunol., in press.

Caras, I.W., Davitz, M.B., Rhee, L., Weddell, G., Martin, D.N. Jr. and Nussenzweig, V. (1987) Nature 325, 545-549.

Carle, G.F. and Olsen, M.V. (1984) Nucl. Acid Res. 12, 5647-5664.
Carroll, M.C., Campbell, R.D., Bentley, D.R. and Porter, R.R. (1984) Nature 307, 237-241.
Carroll, M.C., Campbell, R.D., Porter, R.R. (1985) Proc.Natl.Acad.Sci.USA 82, 521-525.
Carroll, M.C., Alicot, E.A., Katzman, P., Klickstein, L.B. and Fearon, D.T. (1987) Complement 4, 124-244 (Abstr. 46).
Carter, P.E. and Fothergill, J.E. (1987) Complement 4, 124-244 (Abstr. 47).
Casadaban, M.J. and Cohen, S.N. (1980) J.Mol.Biol. 138, 179-207.
Chenoweth, D.E. and Hugli, T.E. (1978) Proc.Natl.Acad.Sci.USA 75, 3943-3947.
Christie, D.L. and Gagnon, J. (1983) Biochem.J. 209, 61-70.
Chung, L.P., Bentley, D.R. and Reid, K.B.M. (1985a) Biochem.J. 230, 133-141.
Chung, L.P. and Reid, K.B.M. (1985b) Bioscience Rep. 5, 855-865.
Chung, L.P., Gagnon, J. and Reid, K.B.M. (1985c) Mol.Immunol. 22, 427-435.
Clewell, D.B. and Helsinki, D.R. (1972) J.Bact. 110, 1135-1146.
Colomb, M.G., Arlaud, G.J. and Villiers, C.L. (1984) Phil.Trans.R.Soc. Lond.B 306, 283-292.
Cooper, N.R. (1975) J.Exp.Med. 141, 890-903.
Czop, J. and Nussenzweig, V. (1976) J.Exp.Med. 143, 615-630.
Daha, M.R. and Van Es, L.A. (1980) J.Immunol. 125, 2051-2054.
Dahlbäck, B. and Steflo, J. (1981) Proc.Natl.Acad.Sci.USA 78, 2512-2516.
Dahlbäck, B. (1983) Biochem.J. 209, 847-856.
Dahlbäck, B., and Hildebrand, B. (1983) Biochem.J. 209, 857-863.
Dahlbäck, B., Smith, C.A. and Müller-Eberhard, H.J. (1983) Proc.Natl.Acad. Sci.USA 80, 3461-3465.
Dahlbäck, B. and Müller-Eberhard, H.J. (1984) J.Biol.Chem. 259, 11631-11634.
Dahlbäck, B. (1986) J.Biol.Chem. 261, 12022-12027.
Day, A.J., Ripoche, J., Lyons, A., McIntosh, B., Harris, T.J.R. and Sim, R.B. (1987a) Bioscience Rep. 7, 201-207.
Day, A.J., Ripoche, J., Willis, A.C. and Sim, R.B. (1987b) Complement 4, 124-244 (Abstr. 64)
Dehardt, D.T. (1966) Biochem.Biophys.Res.Commun. 23, 641-646.
D'Eustachio, P., Kristensen, T., Wetsel, R.A., Riblet, R., Taylor, B.A. and Tack, B.F. (1986) J.Immunol. 137, 3990-3995.
DiScipio, R.G., Hermodson, M.A., Yates, S.G. and Davie, E.W. (1977) Biochemistry 16, 698-706.
Dodds, A.W., Law, S.K.A. and Porter, R.R. (1986) Immunogenetics 24, 279-285.
Doolittle, R.F. (1984) in "The plasma proteins" 2nd edn. 4, 317-360 (ed.

Putnam, S.W.).
Doolittle, R.F. (1985) Trends Biochem.Sci. 10, 233-237.
Dretzen, G., Bellard, M., Sassone-Corsi and Chambon, P. (1981) Anal.Biochem. 112, 295-298.
Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) Nucl. Acids Res. 9, 1691-1706.
Dunham, I., Sargeant, C.A., Trowsdale, J. and Campbell, R.D. (1987) Proc.Natl.Acad.Sci.USA, in press.
Dykman, T.R., Cole, J.L., Iida, K. and Atkinson, J.P. (1983) Proc.Natl.Acad. Sci.USA 80, 1698-1702.
Erdei, A. and Sim, R.B. (1987) Biochem.J. 246, 140-156.
Esmon, C.T. (1987) Science 235, 1348-1352.
Fearon, D.T. and Austen, K.F. (1975) J.Exp.Med. 142, 856-863.
Fearon, D.T. (1978) Proc.Natl.Acad.Sci.USA 75, 1971-1975.
Fearon, D.T. (1979a) CRC Crit.Rev. in Immunol. 1, 1-32.
Fearon, D.T. (1979b) Proc.Natl.Acad.Sci.USA 76, 5867-5871.
Fearon, D.T. and Austen, K.F. (1980) New.Engl.J.Med. 303, 259-263.
Fearon, D.T. and Wong, W.W. (1983) Ann.Rev.Immunol. 1, 243-271.
Fearon, D.T. (1984) Immunol. Today 5, 105-110.
Feinberg, A.P.and Vogelstein, D. (1983) Anal.Biochem. 132, 6-13.
Ferreira, A., Takahashi, M. and Nussenzweig, V. (1977) J.Exp.Med. 146, 1001-1018.
Frank, M.M., Gelfand, J.A. and Atkinson, J.P. (1976) Ann.Int.Med. 84, 580-593.
Friedman, H.M., Cohen, G.H., Gisenberg, R.J., Sidel, C.A. and Cines, D.D. (1984) Nature 309, 633-635.
Frischauf, A.M., Lehrach, H., Poustka, A. and Murray, N. (1983) J.Mol.Biol. 170, 827-842.
Fujita, T., Gigli, I. and Nussenzweig, V. (1978). J.Exp.Med. 148, 1044-1051.
Fujita, T., and Nussenzweig, V (1979) J.Exp.Med. 150, 267-276.
Fujita, T., Inoue, T., Ogawa, K., Iida, K. and Tamura, N. (1987) Complement 4, 124-224 (Abstr. 87).
Gagnon, J. (1984) Phil.Trans.R.Soc.Lond.B 306, 301-309.
Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979) Nucl. Acid Res. 7, 2115-2136.
Ghebrehiwet, B. and Müller-Eberhard, H.J. (1978) J.Immunol. 120, 27-30.
Gigli, I., Fujita, T. and Nussenzweig, V. (1979) Proc.Natl.Acad.Sic.USA 76, 6596-6600.

Gigli, I., Sorvillo, J., Mecarelli-Halbwachs, L. and Leibowitch, J. (1981) J.Exp.Med. 154, 1-12.

Gilbert, W. (1985) Science 227, 823-824.

Grosveld, F.G., Dahl, H.H.M., de Boer, E. and Flavell, R.A. (1981) Gene 13, 227-237.

Grosveld, F.G., Lund, T., Murray, E.J., Mellor, P.L., Dahl, H. and Flavell, R.A. (1982) Nucl.Acid.Res. 10, 6715-6731.

Halbwachs, L., Leveille, M., Lesarte, P.H., Wattel, S. and Leibowitch J. (1980) J.Clin.Invest. 65, 1249-1256.

Hanahan, D. and Meselson, M. (1983) Methods in Enzymol. 100, 333-342.

Henkart, P.A., Millard, P.J., Reynolds, C.W. and Henkart, M.P. (1986) J.Exp.Med. 160, 75-93.

Herrmann, R., Neugebauer, K., Pirkl, E., Zentgraf, H. and Schaller, H. (1980) Molec.Genet. 177, 231-242.

Hillarp, A. and Dahlbäck, B. (1987) Complement 4, 124-244 (Abstr. 118 and 119).

Hogg, N., Takacs, L., Palmer, D.G., Selvendran, Y. and Allen C. (1986) Eur.J.Immunol. 16, 240-248.

Hogness, D.S. and Simmons, J.R. (1964). J.Mol.Biol. 9, 411-438.

Holers, V.M., Cole, J.L., Lublin, D.M., Seya, T. and Atkinson, J.P. (1985) Immunol. Today 6, 188-191.

Hourcade, D., Holers, V.M., Miesner, D.R. and Atkinson, J.P. (1987) Complement 4, 124-244 (Abstr. 129).

Hugli, T.E. (1981) CRC Critical Rev. Immunol. 1, 321-366.

Hugli, T.E. (1986) Complement 3, 111-127.

Ichinose, A., McMullen, B., Fujikawa, K. and Davie, E.W. (1986) Biochem. 25 4633-4638.

Iferroudjene, D., Ripoche, J., Gilbert, D., Schouft, M.T. and Fontain, M. (1987) Complement 4, 124-244 (Abstr. 130).

Iida, K. and Nussenzweig, V. (1981) J.Exp.Med. 153, 1138-1150.

Ish-Horowicz, D. and Burke, J.F. (1981) Nucl. Acid Res. 9, 2989-2998.

Janatova, J., Reid, K.B.M. and Willis, A. (1987) Complement 4, 124-244 (Abstr. 133).

Jenne, D.E., Masson, D., Rey, C. and Tschopp, J. (1987) Complement 4, 124-244 (Abstr. 135).

Johnson, P.H. and Grossman, L.I. (1977) Biochemistry 16, 4217-4224.

Joiner, K.A., Brown, E.J. and Frank, M.M. (1984) Ann.Rev.Immunol. 82, 461-491.

Kaidoh, T., Natsume-Sakai, S. and Takahashi, M. (1981) J.Immunol. 126, 463-467.

Kaidoh, T. and Gigli, I. (1987) J.Immunol. 139, 194-201.

Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc.Natl.Acad. Sci.USA 77, 5172-5176.

Kristensen T. and Tack, B.F. (1986) Proc.Natl.Acad.Sci.USA 83, 3963-3967.

Klickstein, L.B., Wong, W.W., Smith, J.A., Weis, J.H., Wilson, J.G. and Fearon, D.T. (1987) J.Exp.Med. 165, 1095-1112.

Kristensen, T., Ogata, R.T., Chung, L.P., Reid, K.B.M. and Tack, B.F. (1987a) Biochemistry 26, 4668-4674.

Kristensen, T., D'Eustachio, P., Ogata, R.T., Chung, L.P., Reid, K.B.M. and Tack, B.F. (1987b) Fed.Proc. 46, 2463-2469.

Lachmann, P.J. and Thompson, R.A. (1970) J.Exp.Med. 131, 643-657.

Lambris, J.D. and Ross, G.D. (1982) J.Exp.Med. 155, 1400-1411.

Laskey, R.A. and Mills, A.D. (1977) FEBS Letts. 82, 246-258.

Law, S.K.A., Dodds, A.W. and Porter, R.R. (1984) EMBO J. 3, 1819-1823.

Leach, D.R.F. and Stahl, F.W. (1983) Nature 305, 448-451.

Leonard, W.J., Depper, J.M., Kanehisa, M., Krönke, M., Peffer, N.J., Svetlik, P.B., Sullivan, M. and Greene, W.C. (1985) Science 230, 633-639.

Lewin, B. (1985) Genes II (2nd Edn.) Ch.23. John Wiley & Sons.

Linscott, W.D. (1986) Prog.Vet.Microbiol.Immun. 2, 54-77.

Lintin, S.J. and Reid, K.B.M. (1986) FEBS Lett 204, 77-81.

Lozier, J., Takahashi, N. and Putnam, F.W. (1984) Proc.Natl.Acad.Sci.USA 81, 3640-3644.

Maeda, N. Yang, F., Barnett, D.R., Bowman, B.H. and Smithies, O. (1984) Nature 309, 131-135.

Maeda, N. and Smithies, O (1986) Ann.Rev.Genet. 20, 81-108.

Malhotra, V. and Sim, R.B. (1985) Eur.J.Immunol. 15, 935-941.

Mammen, E.F. (1984) ed. Protein C (Seminars in Thrombosis and Haemostais, Thieme-Stratton, New York 10.

Maniatis, T., Fritsch, E.P. and Sambrook, J. (1982) 'Molecular cloning, a Laboratory Manual', Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Marazziti, D., Eggertsen, G., Stanley, K.K. and Fey, G.H. (1987) Complement 4, 124-244 (Abstr. 174).

Marche, P.N. and Kindt, T.J. (1986) J.Immunol. 137 1729-1734.

Maxam, A.M. and Gilbert, W. (1977) Proc.Natl.Acad.Sci.USA 74, 560-564.

Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.

McNearney, T.A., Odell, C., Holers, V.M. Spear, P.G. and Atkinson, J.P. (1987) Complement 4, 124-244 (Abstr. 182).

Medof, M.E., Kinoshita, T. and Nussenzweig, V. (1984) Complement 1, 168-173.

Medof, M.E., Walter, E., Rutgers, J., Knowles, D. and Nussenzweig, V. (1985) Complement 2, 53 (Abstr. 151).

Melchers, F., Erdei, A., Schultz, T. and Dierich, M. (1985) Complement 2, 54 (Abstr.152).

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) Proc.Natl.Acad.Sci.USA 74, 3642-3646.

Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl. Acid Res. 9, 309-321.

Messing, J. and Vieira, J. (1982) Gene 19, 269-276.

Messing, J. (1983) Methods in Enzymol. 101, 20-78.

Miyakawa, Y., Yamada, B., Kosaka, K., Tsuda, F., Kosugi, E. and Mayumi, M. (1981) Lancet ii 493-497.

Morgan, E.L. (1986) Complement 3, 128-136.

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

Müller-Eberhard, H.J. (1986) Ann.Rev.Immunol. 4, 503-528.

Nagasawa, S., Mizuguchi, K., Ichihara, C. and Koyama, J. (1982) J.Biochem (Japan) 92, 1329-1332.

Nagasawa, S. and Stroud, R.M. (1977) Immunochem. 14, 749-756.

Nagasawa, S., Ichikara, C. and Stroud, R.M. (1980) J.Immunol. 125, 578-582.

Nagasawa, S. and Stroud, R.M. (1980) Mol.Immunol. 17, 1365-1372.

Nicholl, P.A.E. and Lachmann, P.J. (1973) Immunol. 24, 259-275.

Nicholson-Weller, A., Burge, J., Fearon, D.T., Weller, P.F. and Austen, K.F. (1982) J.Immunol. 129, 184-189.

Nicholson-Weller, A., March, J.P., Rosenfeld, S.I. and Austen, K.F. (1983) Proc.Natl.Acad.Sci.USA 80, 5066-5070.

O'Neill, K.M. and Torres, M.E. (1987) Complement 4, 124-244 (Abstr. 214).

Pangburn, M.K. and Müller-Eberhard, H.J. (1984) Springer Semin. Immunol. 7, 163-192.

Pangburn, M.K., Schreiber, R.D. and Müller-Eberhard, H.J. (1983) Proc.Natl. Acad.Sci.USA 80, 5430-5434.

Pangburn, M.K., Schreiber, R.D. and Müller-Eberhard, H.J. (1981) J.Exp.Med. 154, 856-862.

Parks, D.L., Gruner, B.A., Hourcade, D.E. and Holers, V.M. (1987) Complement 4, 124-244 (Abstr. 224).

Perkins, S.J., Chung, L.P. and Reid, K.B.M. (1986) Biochem.J. 233, 799-807.

Perlmutter, R.M., Ram, D. and Hood, L. (1986) UCLA Symp. on Molecular and Cellular Biology in Genes and Cancer (eds. Bishops, J.M., Rowley, J.D., Greaves, M.) 17, 489-499.

Podack, E.R. and Tschopp, J. (1984) Mol.Immunol. 21, 589-603.

Podack, E.R., Young, I.D. and Cohen, Z.A. (1985) Proc.Natl.Acad.Sci.USA 82, 8629-8633.

Porter, R.R. and Reid, K.B.M. (1979) Adv. Protein Chem. 33, 1-71.

Radloff, R., Bauer, W. and Vinograd, J. (1967) Proc.Natl.Acad.Sci.USA 57, 1514-1521.

Reid, K.B.M. and Porter, R.R. (1981) Ann.Rev.Biochem. 50, 433-464.

Reid, K.B.M. (1983) Biochem.Soc.Trans. 11, 1-12.

Reid, K.B.M. and Gagnon, J. (1982) FEBS Lett. 137, 75-79.

Reid, K.B.M. (1985) Biochem.J. 231, 729-735.

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. and Tack, B.F. (1986) Immunol. Today 7, 230-233.

Reid, K.B.M., Campbell, R.D., Day, A.J., Dodds, A.W., Law, S.K., Lintin, S.J., Ripoche, J., Sim, R.B., Wu, L.-C., Yu, C.-Y. In: "Progress in Immunology VI (B. Cinader and R.G. Miller, eds.) Academic Press Inc. (1986).

Rey-Campos, J., Rubinstein, P. and Rodriguez de Cordoba, S. (1987) J.Exp. Med. 166, 246-252.

Reynes, M., Aubert, J.P., Cohen, J.H.M., Audonin, J., Tricottet, V., Drebold, J. and Kazatchkine, M.D. (1985) J.Immunol. 135, 2687-2694.

Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J.Mol.Biol. 113, 237-251.

Ripoche, J., Day, A.J., Willis, A.C., Belt, K.T., Campbell, R.D. and Sim, R.M. (1986) Bioscience Rep. 6, 65-72.

Ripoche, J., Fontaine, M., Day, A.J., Davrinche, C. and Sim, R.B. (1987) Complement 4, 124-244 (Abstr. 246).

Rodrigues de Cordoba, S., Ferreira, A., Nussenweig, V. and Rubinstein, P. (1983) J.Immunol. 131, 1565-1569.

Rodrigues de Cordoba, S. and Rubinstein, P. (1984) J.Immunol. 132 1906-1908.

Rodrigues de Cordoba, S., Dykman, T.R., Ginsberg-Fellner, F., Ercilla, G., Aqua, M., Atkinson, J.P. and Rubinstein, P. (1984) Proc.Natl.Acad.Sci. USA 81, 7890-7892.

Rodrigues de Cordoba, S., Lublin, D.M., Rubinstein, P. and Atkinson, J.P. (1985) J.Exp.Med. 161, 1189-1195.

Rodrigues de Cordoba, S. and Rubinstein, P. (1986) J.Exp.Med. 164, 1274-1283.

Rodrigues de Cordoba, S. and Rubinstein, P. (1987) Immunogenetics. 25, 267-268.

Rogers, J. (1986) TIG 2, 223.

Roitt, I., Brostoff, J. and Male, D. (1985) in Immunology, Ch.7, Chuchill Livingstone.

Rosenberg, S.M. (1985) Gene 39, 313-315.

Ross, G.D. and Medof, M.E. (1985) Ad.Immunol. 37, 217-267.

Ruddy, S. and Austen, K.F. (1971) J.Immunol. 107, 742-750.

Sanchez-Madrid, F., Nagy, J.A., Robbins, E., Simon, P. and Springer, T.A. (1983) J.Exp.Med. 158, 1785-1803.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci.USA 74, 5463-5468.

Sanger, F. and Coulson, A.R. (1978) FEBS Letts. 87, 107-110.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B. (1980) J.Mol.Biol. 143, 161-178.

Scharfstein, J., Ferreira, A., Gigli, I. and Nussenzweig, V. (1978) J.Exp.Med. 148, 207-222.

Schmitt, M., Musel, H-H., Hamman, K.P., Scheiner, O., Dierich, MP. (1981) Eur.J.Immunol. 11, 739-745.

Schultz, T.F., Schreiner, D., Alsenz, J., Lambris, J.D. and Dierich, M.P. (1984) J.Immunol. 132, 392-398.

Schultz, T.F., Pernaegger, G., Petzer, A., Eingentler, A., Myones, B.L., and Dierich, M.P. (1987) Complement 4, 124-244 (Abstr. 261).

Schwaeble, W., Schultz, T.F., Zwirner, J., Dierich, M.P. and Weiss, E.H. (1987) Complement 4, 124-244 (Abstr. 264).

Seya, T., Turner, J.R., Atkinson, J.P. (1986) J.Exp.Med. 163, 837-855.

Seya, T., Ballard, L., Bora, N. McNearney, T., and Atkinson, J.P. (1987) Complement 4, 124-244 (Abstr. 268).

Sharp, P.A., Sugden, B. and Sambrook, J. (1973) Biochemistry 12, 3055-3063.

Shiraishi, S. and Stroud, R.M. (1975) Immunochemistry 12, 935-939.

Sim, R.B. and Walport, M.J. (1986) in "Complement in Health and Disease"

Sim, R.B., Malhotra, V., Ripoche, J., Day, A.J., Michelem, K. and Sim, E. (1986) Biochem.Soc.Symp. 51, 83-96.

Sim, R.B., Malhotra, V., Day, A.J. and Erdei, A. (1987) Immunol. Letts 14, 183-190.

Chapter 6, MTP Press, Lancaster, U.K. (K. Whaley, ed.).

Smith + Kerr, M.S. (1985), Immunol 56 561 - 570

Sottrup-Jensen, L. Stepanik, T.M., Kristensen, T., Lonblad, P.B., Jones, C.M., Wierbicki, D.M., Magnusson, S., Dimdey, H., Wetsel, R.A., Lundwall, A., Tach, B.F. and Fey, G.H. (1985) Proc.Natl.Acad.Sci.USA 82, 9-13.

Southern, E. (1979) Methods in Enzymol. 68, 152-176.

Springer, T.A., Teplow, D.B. and Dreyer, W.J. (1985) Nature 314, 540-542.

Staden, R. (1982a) Nucl. Acid Res. 10, 4731-4751.

Staden, R. (1982b) Nucl. Acid Res. 10, 2951-2961.

Staden, R. (1984) Nucl. Acid Res. 12, 521-538.

Steinmetz, M., Stephan, D., Dastoornikoo, G.R., Gibb, E. and Romaniuk, R. (1985) In: "Immunological Methods" Vol.III p.1-19 (eds. Lefkovits, I. and Pernis, B.) Academic Press, N.Y.

Stimler-Gerard, N.P. (1986) Complement 3, 137-151.

Tack, B.F. (1985) In: "Complement", Springer-Verlag Berlin (eds. Müller-Eberhard, H.J. and Meishcer, P.J.) 49-72.

Tenner, A.J. and Cooper, N.R. (1980) J.Immunol. 125, 1658-1664.

Tenner, A.J. and Cooper, N.R. (1981) J.Immunol. 126, 1174-1179.

Thompson, R.A. and Lachmann, P.J. (1977) Clin.Exp.Immunol. 27, 23-29.

Tosi, M., Duponche, C., Meo, T., Julier, C., (1987) Biochemistry, in press.

Van de Sande, J., Kleppe, H., and Khorana, H.G. (1973) Biochemistry 12 5050-5055.

Vieira, J. and Messing, J. (1982) Gene 19, 259-268.

Vik, D.P., Keeney, J.B., Bronson, S., Westlund, B., Kristensen, T., Chaplin, D.D., and Tack, B.F. (1987) Complement 4, 124-244 (Abstr. 296).

Vik, D.P. and Fearon, D.T. (1987) J.Immunol. 138, 254-258.

Villiers, M.B., Reboul, A., Thielens, N.M. and Colomb, M.G. (1981) FEBS Lett 132, 49-54.

Villiers, M.B., Thielens, N.M., Villiers, C.L. and Colomb, M.G. (1985) Eur.J.Immunol. 15, 941-945.

Vogt, W., Schmidt, G., Von Buttlar, B. and Dieminger, L. (1978) Immunol. 34, 29-40.

von Zabern, I. and Gigli, I. (1982) J.Immunol. 128, 1439-1442.

von Zabern, I., Bloom, E.L., Chu, V. and Gigli, I. (1982) J.Immunol. 128, 1433-1438.

Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc.Natl.Acad.Sci.USA 76, 3683-3687.

Watson, M.E.E. (1984) Nucl. Acid Res. 12, 5145-5164.

Weiler, J.M., Daha, M.R., Austen, K.F. and Fearon, D.T. (1976) Proc.Natl.

Acad.Sci.USA 73, 3268-3272.

Weis, J.J., Fearon, D.T., Klickstein, L.B., Wong, W.W., Richards, S., De Bruyn Kops, A., Smith, J.A. and Weis, J.H. (1986) Proc.Natl.Acad.Sci. USA 83, 5639-5643.

Weis, J.J., Toothaker, L.E., Burrow, S.R., Weis, J.H. and Fearon, D.T. (1987a) Complement 4, 124-244 (Abstr. 304).

Weis, J.H., Morton, C.C., Bruns, G.A.P., Weis, J.J., Klickstein, L.B., Wong, W.W. and Fearon, D.T. (1987b) J.Immunol. 138, 312-315.

Weis, J.H., Aegerter-Shaw, M., Cole, J.L., Tantravahi, R.V., Miller, M.P., Kurtz, C.B., Diehl, A.D. and Weis, J.J. (1987c) Complement 4, 124-244 (Abstr. 303).

Weitkamp, L.R. and Lamm, L.U. (1982) Cytogenet. Cell Genet. 32, 130-143.

Whaley, K. and Ruddy, S. (1976) J.Exp.Med. 144, 1147-1163.

Wilson, J., Wong, W., Schur, P. and Fearon, D.T. (1982) New Engl.J.Med. 307, 981-986.

Wilson, J.G., Murphy, E.E., Wong, W.W., Klickstein, L., Weis, J.H. and Fearon, D.T. (1986) J.Exp.Med. 164, 50-59.

Winter, G., Fields, S., Gait, M.J. and Brownlee, G.G. (1981) Nucl. Acid Res. 9, 237-245.

Wong, W., Wilson, J.G. and Fearon, D.T. (1983) J.Clin.Invest. 72, 685-693.

Wong, W.W. and Fearon, D.T. (1985) J.Immunol. 134, 4048-4056.

Wong, W.W., Kennedy, C.A., Bonaccio, E.T., Willson, J.G., Klickstein, L.B., Weis, J.H. and Fearon, D.T. (1986) J.Exp.Med. 164, 1531-1546.

Wong, W., Cahill, J., Kennedy, C., Bonaccio, E., Wilson, J., Klickstein, L., Rabson, L. and Fearon, D.T. (1987) Complement 4, 124-244 (Abstr. 309).

Wozney, J., Nanahan, D., Tate, V., Boedtker, H. and Doly, P. (1981) Nature (Lond.) 294, 129-135.

Wright, S.D., Rao, P.E., Van Voorhis, W.C., Craigmyle, L.S., Iida, K., Talle, M.A., Westberg, E.F., Goldstin, G. and Silverstein, S.C. (1983) Proc.Natl.Acad.Sci.USA 80, 5699-5703.

Wu, R. and Taylor, E. (1971) J.Mol.Biol. 57, 491-511.

Wu, L.C., Morley, B.J. and Campbell, R.D. (1987) Cell 48, 331-342.

Wyman, A.R., Wolfe, L.B. and Botstein, D. (1985) Proc.Natl.Acad.Sci.USA 82, 2880-2884.

Yoon, S.H. and Fearon, D.T. (1983) J.Immunol. 134, 3332-3338.

Ziccardi, R.J. (1982) J.Immunol. 128, 2505-2508.

Ziccardi, R.J., Dahlbäck, B. and Müller-Eberhard (1984) J.Biol.Chem. 259, 13674-13679.

Studies on the structure of the human C4b-binding protein gene

Susan J. Lintin and K.B.M. Reid+

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

Received 30 May 1986

Protein and cDNA sequence data have shown human C4b-binding protein to contain eight internally homologous repeat units, each approx. 61 amino acids in length. Repeat units conforming to the same consensus sequence have been found in other complement and non-complement proteins. Southern blot analysis together with isolation, characterisation and sequencing of genomic clones has allowed the study of intron/exon organisation in the human C4b-binding protein gene and the identification of a *Bgl*II restriction fragment length polymorphism.

Complement *C4b-binding protein* *Gene structure* *Polymorphism*

1. INTRODUCTION

C4b-binding protein (C4bp) is one of a group of regulatory proteins, which includes complement receptor type 1 (CR1), factor H and gp45–70 [1], within the complement cascade which maintain homeostasis of C3 levels. C4bp performs this regulation in the classical pathway of complement activation [2,3] by accelerating the decay of C3 convertase (C4bC2a), through binding to C4b thus displacing C2a and acting as a cofactor for factor I which proteolytically splits C4b [4,5]. These functions rely upon the ability of C4bp to bind C4b; similar dependence of C4b/C3b binding is noted in the regulatory function of the plasma protein factor H of the alternative pathway and the cell surface molecules CR1, DAF and gp45–70 [1].

C4bp is a serum glycoprotein of 550 kDa and is thought to be composed of 7 identical disulphide-bond-linked chains [6]. Each chain is composed of 8 contiguous repeating homology units, each approx. 61 amino acids long, making up the N-terminal 491 amino acids while the C-terminal 58 amino acids show no homology to the rest of the chain [7]. The repeating units conform to a consensus sequence having a framework of highly conserved residues consisting of one tryptophan, two proline and four half-cystine residues and a number of other partially conserved positions involving glycine or hydrophobic residues. These repeating units are widespread being found in a number of complement proteins (C2, factor B, C4bp, factor H and CR1) and non-complement proteins (β_2-glycoprotein, IL-2 receptor; blood clotting factor XIII, haptoglobin-2 precursor) (review [8]).

Two phenotypes of human C4bp have been identified [9] and have been closely linked to the loci for CR1 and factor H and thus form a recently described gene cluster of complement regulatory proteins [10] which has been mapped to chromosome 1 ([11] and Solomon, E., unpublished). Use of the cDNA probes and Southern blot analysis has allowed the mapping of the C4bp gene and identification of a *Bgl*II restriction fragment length polymorphism. The isolation of genomic clones has allowed the study of the intron/exon

+ To whom correspondence and reprint requests should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50 © 1986 Federation of European Biochemical Societies

organisation in the gene which will be informative with respect to understanding the evolution of this group of regulatory proteins.

2. MATERIALS AND METHODS

2.1. *Enzymes*

Restriction endonucleases were purchased from New England Biolabs, Boehringer Mannheim and Amersham International. The nick translation 'kit' and all radiolabelled nucleotides were also from Amersham International.

2.2. *Southern blot analysis*

Genomic DNA was prepared from whole blood [12]; for analysis 8–10 μg was digested with 20 units of enzyme for 20 h, electrophoresed in 0.7% agarose and transferred to nitrocellulose (Amersham) [13]. These were then hybridised with cDNA probes (fig.1a), nick-translated to a specific activity of $\simeq 10^8$ cpm/μg of DNA [14] in buffer containing 50% formamide and 6×10^6 cpm/ml for 30 h. The blots were then washed three times at room temperature in 2 × SSC + 0.1% SDS and three times at 68°C in 0.2 × SSC.

2.3. *Genomic clone isolation and characterisation*

Approx. 10^6 plaques from a λ phage library (constructed by A. Bentley) were screened by standard techniques [15] using the nick-translated cDNA probes. λDNA was prepared [15] and partially characterised by Southern blotting. Restriction fragments from a *Pvu*II digest of λDNA was subcloned into pAT153/*Pvu*II/8 [16] and detected

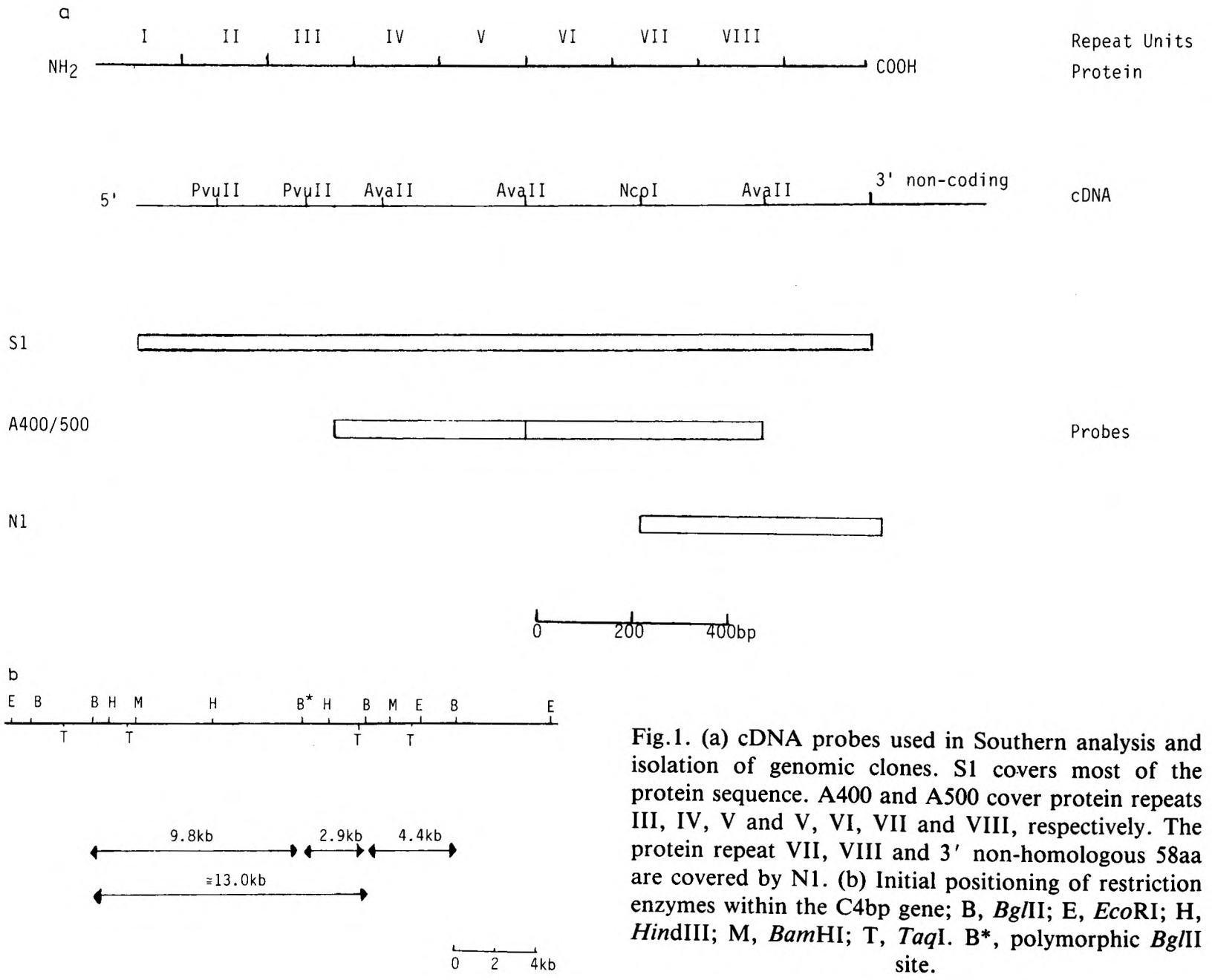


Fig.1. (a) cDNA probes used in Southern analysis and isolation of genomic clones. S1 covers most of the protein sequence. A400 and A500 cover protein repeats III, IV, V and V, VI, VII and VIII, respectively. The protein repeat VII, VIII and 3′ non-homologous 58aa are covered by N1. (b) Initial positioning of restriction enzymes within the C4bp gene; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; M, *Bam*HI; T, *Taq*I. B*, polymorphic *Bgl*II site.

by colony hybridisation with the A400 and S1 cDNA probes. Plasmid DNA was prepared by the alkaline-SDS method [17] and DNA sequencing carried out using the Maxam and Gilbert method.

3. RESULTS AND DISCUSSION

From initial genomic DNA blotting results it has been possible to construct a restriction map (fig.1b) using the four available cDNA probes which together cover most of the cDNA sequence. The overall size of the gene is approx. 30 kb as estimated by Southern blotting using 5′ and 3′ cDNA probes. From Northern blots the C4bp mRNA is estimated to be 2.5 kb long [7] and a large proportion of the gene must therefore be composed of intron sequences.

A restriction fragment length polymorphism was detected using the restriction endonuclease *Bgl*II and the S1 and A500 probes. A total of 18 genomic DNA samples (15 of which were a gift from A. Palsdottir) were then analysed with the same enzyme and probes (fig.2). Using the cDNA probes available the *Bgl*II polymorphism was mapped as shown in fig.1b and is due to the presence or absence of a *Bgl*II site 2.9 kb upstream of the 5′-end of the 4.4. kb *Bgl*II fragment.

Of the 18 samples 3 were members of a family, in which the father is homozygous for the smaller fragments and the mother and daughter are heterozygous. Thus the polymorphism is inherited and in a Mendelian fashion.

The polymorphism has been detected in 50% of the 18 samples tested and is therefore unlikely to show a simple correlation with the protein polymorphism which has gene frequencies of 0.981 and 0.018 [9]. During this study no other restriction fragment length polymorphisms using 10 other common restriction enzymes have been detected in the C4bp gene which suggests the C4bp gene, unlike the C4 and related MHC associated genes, may not be highly polymorphic although a

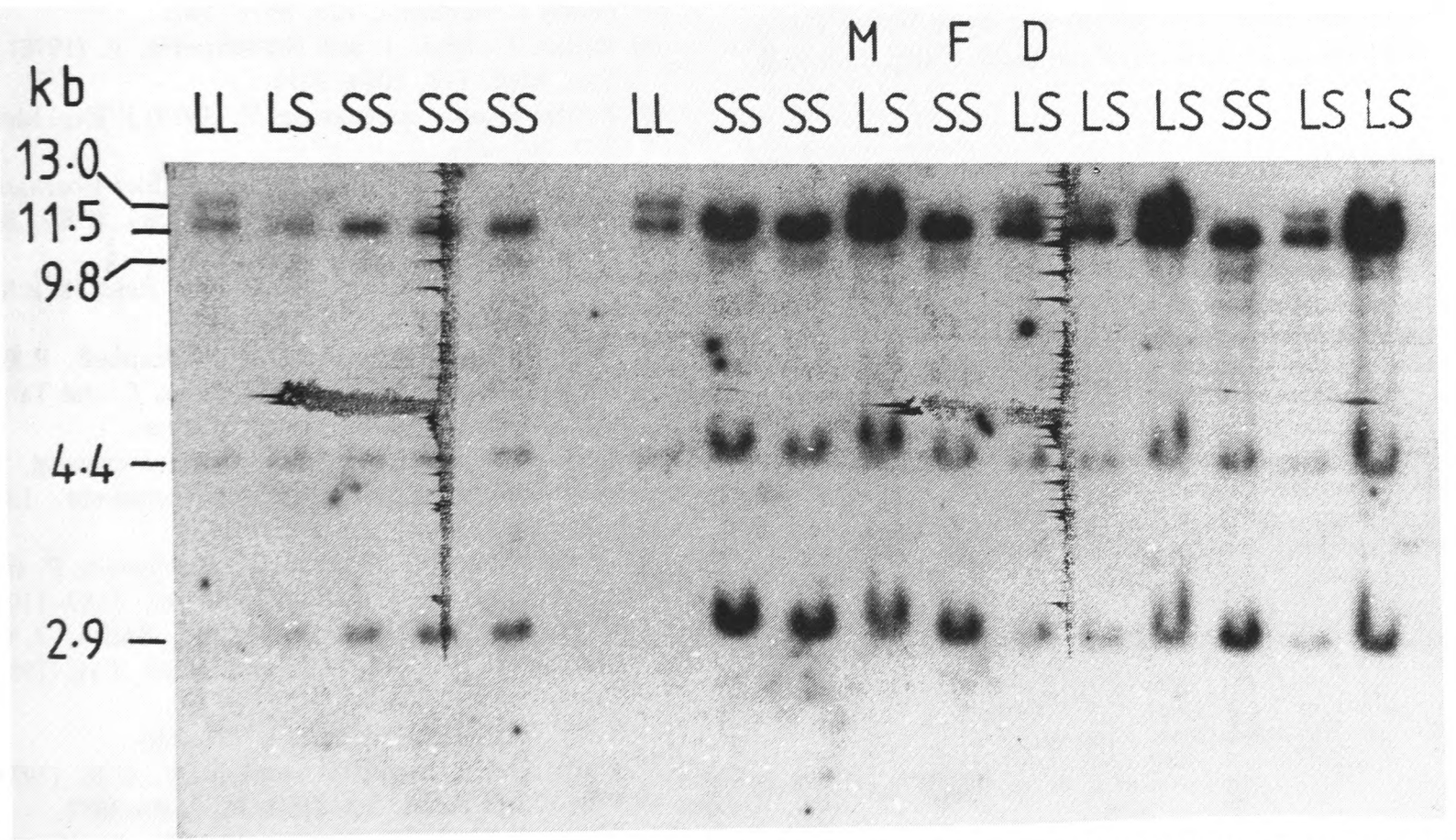


Fig.2. Southern analysis of the *Bgl*II restriction fragment length polymorphism using the A500 probe. Three different patterns can be seen, the 11.5 kb and 4.4 kb fragments being common in all. Samples marked SS have two extra bands at 9.8 kb and 2.9 kb (and are homozygous for the small fragments); LL, have one extra band at approx. 13.0 kb (and are homozygous for the large fragment). LS have all five bands (and are heterozygous). The samples marked M (mother), F (father), D (daughter) are members of one family.

number of differences between the derived protein sequence from the cDNA of an individual and the protein sequence of C4bp isolated from pooled plasma have been reported [7].

Preliminary sequencing data from one of the subclones of the 4 isolated genomic clones, clone BP/Pv.a, have determined the location of two intron/exon boundaries (fig.3) which from comparison with the published amino acid sequence of C4bp [7] corresponds to the C-terminal of the third internal repeat, and N-terminal of the fourth. From these early data it can be speculated that each of the 60 amino acid repeats will be precisely encoded within one exon, as has been found for the three internally homologous repeats in the complement protein factor B [18].

A number of complement proteins contain these internal repeating units [8], the N-terminal Ba and C2b portions of factor B and C2, both contain 3 and the chains of C4bp and factor H [19] contain 8 and 20 repeating units respectively; the precise number in CR1 is not known yet [20] but there could be as many as 36. Evidence for the encoding of the protein repeat within one exon in the three complement genes so far examined, C4bp, factor B and C2 [8], would suggest that a mechanism of tandem duplication of a single exon has occurred. In factor B and C2 the repeat structure has been combined with a catalytic chain containing a serine proteinase domain [18]. The evolution of these proteins can now be discussed in terms of exon shuffling as described for the LDL receptor [21].

The complement proteins containing the repeating units are all functionally, as well as structurally, related since they all interact with C3b or C4b [8]. However, the discovery of the repeat in non-complement proteins would imply that the presence of repeating units may be a general structural feature of a superfamily of related proteins.

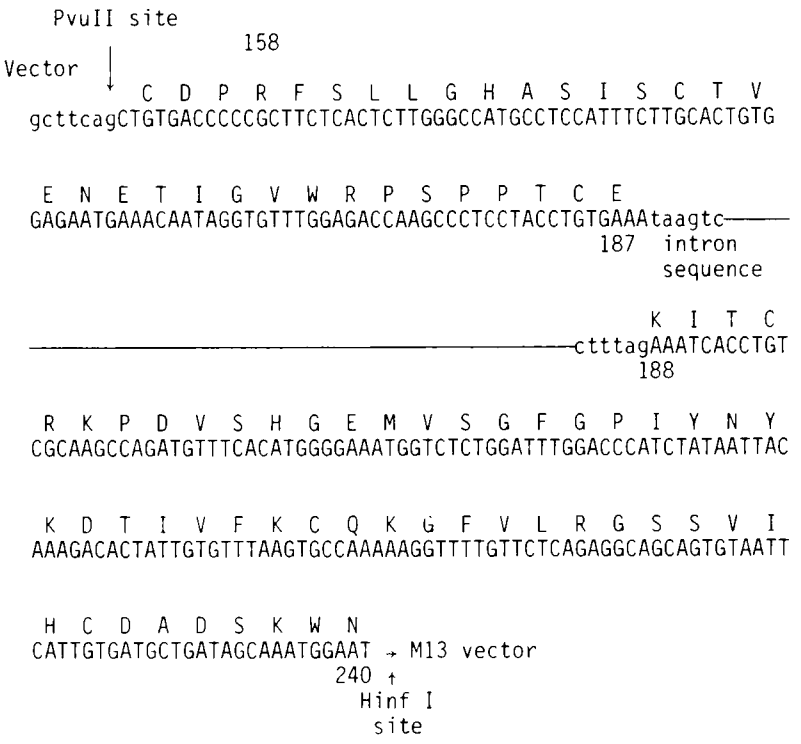


Fig.3. The nucleic acid sequence of the C4bp gene showing the coding region from Cys 155 to Glu 187 and some intron sequence and coding region from lysine 188 to asparagine 240. Amino acid 187 marks the end of the third protein repeat in C4bp.

REFERENCES

[1] Holers, V.M., Cole, J.L., Lublin, D., Seya, T. and Atkinson, J.P. (1985) Immunol. Today 6, 188–192.
[2] Gigli, I., Fujita, T. and Nussenzweig, V. (1979) Proc. Natl. Acad. Sci. USA 76, 6596–6660.
[3] Nagasawa, S., Ischihara, C. and Stroud, R.M. (1980) J. Immunol. 125, p578–582.
[4] Fujita, T., Gigli, I. and Nussenzweig, V. (1978) J. Exp. Med. 148, 1044–1051.
[5] Fujita, T. and Nussenzweig, V. (1979) J. Exp. Med. 150, 267–276.
[6] Dahlbäck, B., Smith, C.A. and Müller-Eberhard, H.J. (1983) Proc. Natl. Acad. Sci. USA 80, 3461–3465.
[7] Chung, L.P., Bentley, D.R. and Reid, K.B.M. (1985) Biochem. J. 230, 133–141.
[8] Reid, K.B.M., Bentley, D.R., Campbell, R.D., Chung, L.P., Sim, R.B., Kristensen, T. and Tack, B.F. (1986) Immunol. Today, in press.
[9] De Cordoba, S.R., Ferreira, D., Nussenzweig, V. and Rubinstein, P. (1983) J. Immunol. 131, 1565–1569.
[10] De Cordoba, S.R., Lublin, D., Rubinstein, P. and Atkinson, J. (1985) J. Exp. Med. 161, 1189–1195.
[11] Klickstein, L.D., Wong, W.W., Smith, J.A., Morton, C., Fearon, D.T. and Weiss, J.H. (1985) Complement 2, 44.
[12] Sykes, B.C. (1983) Lancet i, 787–788.
[13] Wahl, G.M., Stem, M. and Start, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683–3687.
[14] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Borg, P. (1977) J. Mol. Biol. 113, 237–251.
[15] Maniatis, T., Fritsch, G.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[16] Anson, D.S., Choo, K.H., Rees, D.J.G., Giannelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1986) EMBO J. 3, 1053.
[17] Birnboim, H.C. and Daly, J. (1979) Nucleic Acids Res. 7, 1513–1523.
[18] Campbell, R.D., Bentley, D.R. and Morley, B.J. (1986) Phil. Trans. R. Soc. Lond. 306, 367–378.
[19] Kristensen, T. and Tack, B.F. (1986) Proc. Natl. Acad. Sci. USA, in press.
[20] Wong, W.W., Klickstein, L.D., Smith, J.A., Weis, J.H. and Fearon, D.T. (1985) Proc. Natl. Acad. Sci. USA 82, 7711–7715.
[21] Sudhof, T.C., Goldstein, J.L., Brown, M.S. and Russell, D.W. (1985) Science 228, 815–822.

Studies on the structure of the human C4b-binding gene and its close linkage to the factor H gene.

S.J. Lintin, J. Ripoche, D.R. Bentley, K.B,M. Reid.
M.R.C. Immunochemistry Unit, Oxford, U.K.

C4b-binding protein (C4BP) is a regulatory protein for the classical pathway of complement activation and has been closely linked with other regulatory proteins, CR1 and factor H to form a multigene family. Each 70,000Mr chain in C4BP contains eight contiguous and internally homologous repeat units of 60 amino acids, similar to those found in factor H, CR1 and the N-terminal portions of factor B and C2.

From a genomic lambda library 4 clones were isolated, using a cDNA probe, which together cover approx. 16kb of the gene. Southern blotting techniques and limited DNA sequencing work have so far indicated that, like factor B, the internally homologous repeats will be precisely encoded by separate exons. Using Southern blotting an RFLP with BglII has been identified and also evidence, at the gene level, for the close linkage C4BP to factor H is being sought using the same technique.

Biological Chemistry (Hoppe-Seyler), (1987), 367, 255

STUDIES ON THE STRUCTURE OF THE HUMAN C4b-BINDING PROTEIN GENE AND CHARACTERISATION OF THE PROTEIN SIGNAL SEQUENCE

*S. J. Lintin, A. Lewin and K. B .M. Reid. MRC Immunochemistry Unit, Oxford University, Oxford, OX1 3QU, U.K.

Human C4b-binding protein (C4bp) is functionally, structurally and genetically linked to other complement control proteins and receptors. The gross features of the gene for this protein appear to be conserved in a variety of other species. Using a stretch of cDNA covering sequence for residues 173-307 as a probe, a 1.75 kb clone was isolated from a cDNA liver library. This clone was found to contain sequence encoding residue 518 and extending into the N-terminal sequence. The 5' data from this clone confirmed the peptide data for the first 31 amino acids and also gave the derived signal sequence. The majority of residues in the signal peptide were either hydrophobic or uncharged and contained two universal translation-initiation codons (ATG) at positions -14 and -32. The derived human signal peptide showed 63% identity when compared with same region of the mouse sequence. This clone provided a suitable probe to screen a lambda genomic library to isolate the 5' sequence of the human C4bp gene. Other clones isolated and characterised from the same library have shown that each exon encodes one of the 60 amino acid repeats described in the primary protein sequence of C4bp and other complement and non-complement proteins. Using Southern blotting the genomic DNA of species other than human was analysed for cross hybridisation with a human C4bp probe to suggest the presence of gene or pseudo-gene sequence of C4bp in these species.

Complement (1987) 4, 124-244 (Abstr. 166)

The redacted book chapter "Molecular Biology of the Human Complement Class III Products of the MHC: Factor I and Its Cofactors" is available here:
http://dx.doi.org/10.1016/B978-0-12-174685-8.50033-0