THE ISOLATION OF cDNA CLONES FOR HUMAN FIBRINOGEN

BY

AQEEL MOHAMMED ALI IMAM

a thesis submitted for the degree of

Doctor of Philosophy

in the

University of London

January 1984

Department of Biochemistry
St. Mary's Hospital Medical School
London W2 1PG.

ABSTRACT

Fibrinogen (factor I) is an acute phase plasma protein which participates in the final steps of the coagulation pathway. Human fibrinogen, like all vertebrate fibrinogens studied to date, is composed of dimers of Aα-, Bβ- and γ-chains which are synthesised in the liver as abundant polypeptides.

cDNA clones coding for the Aα-, Bβ- and γ-chains of human fibrinogen were isolated from an adult liver cDNA library. Clones were identified by hybridisation with mixtures of septadecamer oligonucleotides synthesised to regions of the polypeptide sequences showing low ambiguity on codon assignment. Sequencing of cDNA inserts subcloned into M13mp9 and comparision of the cDNA derived protein sequence with published amino acid sequences for fibrinogen polypeptides formed the basis of clone characterisation. cDNA insert sizes are 1,950 bp for Aα-fibrinogen, 1,450 bp for Bβ-fibrinogen and 950 bp for γ-fibrinogen. The recombinants do not share homologous sequences as revealed by the absence of cross-hybridisation.

Southern blot hybridisations of human genomic DNA digested with various restriction enzymes gave simple band patterns with each clone suggesting that fibrinogen genes represent a set of unique sequences in the human genome. Expression in adult human liver, examined by Northern blots, showed the clones for Bβ- and γ-fibrinogen hybridised to a single RNA sequence. The estimated mRNA sizes are 1.85 kb for Bβ-fibrinogen and 1.8 kb for γ-fibrinogen. Two mRNA sequences were detected with the clone for Aα-fibrinogen, a major band at 2.6 kb and a minor band at about 1.8 kb.
Chromosomal regional assignment performed by restriction fragment mapping of Human X CHO somatic cell hybrid DNAs tentatively localises the Aα-, Bβ- and γ-fibrinogen genes to human chromosome 4q26-qter.

To my Parents.

Acknowledgments

I would like to thank all members of the Biochemistry department (past and present) for their help during the course of my thesis. Thanks to Bob Williamson for his encouragement and advice. To Steve Humphries and Simon Wallis for reading the manuscript and making helpful suggestions. To Steve Brown and Julian Crampton for advice on M13 sequencing.

Thanks to Mike Eaton *et al*. (Celltech) for synthesising the oligomers.

Special thanks to fellow (flip) scientists Jack, Dave, Tim, Graham, John Joe, Sue, Conrad, Ola and Sissel who often put forward a more optimistic view of life. Thanks to Joy Dexter for typing the first draft of the manuscript and to Paul North for his expert help with the seemingly endless computer searches.

To my family and friends for their unfailing support.

..... it takes all the running
you can do to keep in the same place.

The Red Queen

(Through The Looking Glass
by L. C.)

ABBREVIATIONS

BSA	Bovine serum albumin
DNA	Deoxyribonucleic acid
cDNA	Complementary DNA
RNA	Ribonucleic acid
mRNA	Messenger RNA
rRNA	Ribosomal RNA
cpm	Counts per minute
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
ddA	Dideoxyadenosine triphosphate
ddC	Dideoxycytosine triphosphate
ddG	Dideoxyguanosine triphosphate
ddT	Dideoxythymidine triphosphate
DDT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
E.coli	Escherichia coli, a gram negative bacterium
g	Acceleration due to gravity
bp	Base pairs
kb	Kilobases
Oligomer	Oligonucleotide (deoxyoligonucleotide)
"n"mer	Oligonucleotide which is "n" nucleotides long
septadecamer	a 17mer
dodecamer	a 12mer
^{o}C	Degrees centigrade
mm	Millimeters
mM, M	Millimolar, molar
ul, ml, l	Microlitre, millilitre, litre
M13mp9	A derivative of the single stranded DNA bacteriophage M13
mg, ug, ng	Milligram, microgram, nanogram
V	Volts
mA	Milliamperes
SDS	Sodium dodecyl sulphate
SSC	Standard saline citrate
TCA	Trichloroacetic acid
UV	Ultraviolet light
W/V, V/V	Weight per volume, volume pr volume
IPTG	Isopropyl-β-D-thiogalacto-pyranoside
X-Gal	5-Bromo-4-chloro-3-indoyl-β-D-galactopyranoside

LIST OF FIGURES

Discussion

LIST OF TABLES

CONTENTS

1. INTRODUCTION

1.1. Molecular Biology and Human Genetics

The term 'human molecular genetics' is of comparatively recent origin. It has acquired a new meaning with recent developments in molecular biological techniques, such as recombinant DNA methodologies, rapid DNA sequencing and advances in gene mapping. The availability of these techniques has revolutionised gene isolation and it is now possible to study eukaryotic gene organisation and perform gene mapping of certain regions of the eukaryotic genome at a level of detail far beyond that possible even for procaryotic genomes two decades ago. Despite the complexity of the human genome - with a total haploid nucleotide base pair number of 3×10^9 - it should in the near future be possible, by restriction fragment mapping, to provide a genetic linkage map of the entire human genome (Solomon and Bodmer, 1979; Botstein et al., 1980). It can be said that as a result of these developments human molecular genetics has now truly come of age.

The human globin gene family illustrates well the advance in fine-structure gene mapping for eukaryotes. This has not only led to the elucidation of gene organisation in the α- and β- globin loci (Maniatis et al., 1980) but also to an understanding of the molecular basis of thalassaemias (Weatherall and Clegg, 1982), which remain the commonest group of monogenetic disorders in the world population. These studies have also revealed a

remarkable versatility in the regulation of mammalian gene expression (Mount and Stietz, 1983).

A number of individual human genes and gene families are being studied for which congenital disorders and genetically defined variants have been reported. Cloned probes specific for these genes will be powerful tools in defining the molecular nature of some of these genetic diseases. They may also be used for fetal diagnosis, without recourse to identifying the defect at the level of the gene. For instance, the availability of cloned cDNA for phenylalanine hydroxylase (Robson et al., 1982) can be used for fetal diagnosis of phenylketonuria an autosomal recessive disorder of phenylalanine metabolism known to affect 1 in 10-15,000 newborns of Caucasian decent. Such fetal diagnosis procedures have been developed for thalassaemias and sickle-cell anaemia (Kan and Dozy, 1978; Little et al., 1980; Maniatis et al., 1980).

A few examples, other than the haemoglobinopathies, where the molecular nature of a genetic defect has been determined using cloned gene probes include a form of osteogenesis imperfecta with type I α1 collagen chain dysfunction (Chu et al., 1983); isolated growth hormone deficiency (IGHD) type A (Phillips et al., 1981), an autosomal recessive trait with severe deficiency of the hormone; and haemophilia B (Giannelli et al., 1983), an X-linked trait with a frequency of 1 in 30,000 males and

characterised by deficiency in the intrinsic clotting factor IX. In all these cases, gene mapping studies revealed deletions ranging from 0.5 kb (type I α1 collagen) to 18 kb (factor IX) at the affected locus. More subtle genetic lesions, such as point mutations, can be defined by sequencing of the mutated gene, as for certain β^+-thalassaemias (Westaway and Williamson, 1981; Spritz et al., 1981). With the increasing number of cloned sequences of known function, the number and types of defined genetic dysfunctions will rise.

1.2. Hereditary Disorders of Blood Coagulation.

Blood clotting in vertebrates is a multistep process (figure 1). Consequently dysfunctions at any single stage can result in profound alterations of the finely balanced coagulation pathway, leading to clinical symptoms of varying severity. To date a number of disorders affecting this pathway have been identified. These include inherited factor VIII deficiency in haemophilia A and factor IX deficiency in Christmas disease (haemophilia B), congenital disorders of fibrinogen (see section 1.3.5.), prothrombin, factor VII and other clotting factors (Ratnoff and Bennett, 1973; Bloom, 1981). A common feature in all has been the heterogeneity of dysfunction. The molecular basis of many of these genetically determined heterogeneous states has not been established. The cloning of cDNAs and genomic sequences of coagulation factors forms the basis of investigating the molecular genetics of blood clotting disorders as exemplified by the cloning of factor IX cDNA (Choo et al., 1982) and its use in analysing familial factor IX deficiency (Gianelli et al., 1983).

Hereditary disorders of fibrinogen -the fibrinogenaemias- represent one of the most heterogeneous and least understood of the known disorders of blood coagulation. Much of our present understanding of the fibrinogenaemias relies on the knowledge of the structure and function of fibrinogen.

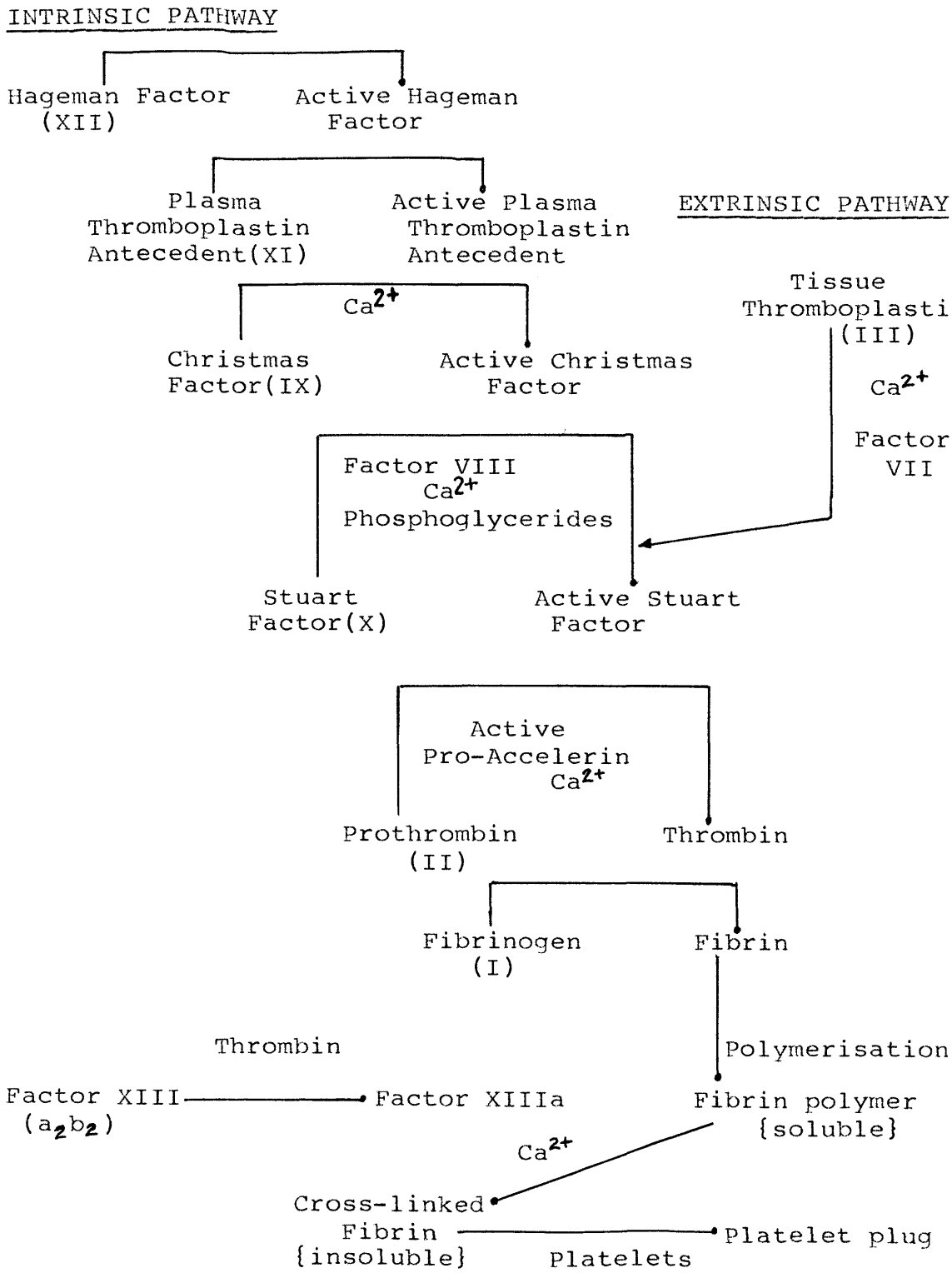

Figure 1. A schematic representation of the blood coagulation cascade showing INTRINSIC and EXTRINSIC pathways leading to clot formation.

1.3. Fibrinogen:Structure, Function, Biosynthesis, Genetics and Evolution

Fibrinogen (factor I) is an abundant (0.3mg/ml) acute phase plasma protein which participates in the final steps of the coagulation pathway. Native fibrinogen is a multimeric molecule (340,000) composed of dimers of non-identical subunits, designated, Aα-(64,000), Bβ- (56,000) and γ-(47,000) chains held together by disulphide bonds. (McKee et al., 1966). On attack by thrombin, fibrinogen is converted to fibrin which polymerises to form the insoluble matrix of a blood clot (see Doolitte, 1981; for a comprehensive review). Below I give an overview of the structure of fibrinogen and its function. This is followed by an account of fibrinogen biosynthesis , genetics and finally its evolution.

1.3.1. Fibrinogen Structure

(A) Tertiary Structure

The mechanism of clot formation by conversion of fibrinogen to fibrin can be understood only by an appreciation of the structure of fibrinogen. Much effort has gone into elucidating both the tertiary structure of the molcule as well as the primary sequence of the participating polypeptides. The tertiary structure of native fibrinogen described below is admittedly biased towards one model of the molecule. This is because there is a plethora of indirect evidence supporting this structure and one feels that unless overwhelming evidence to the

contrary is presented, this structure best explains the mechanism of blood clotting and clot dissolution.

The structure of fibrinogen is rather controversial (Mosesson et al., 1979; Hurdy-Clergeon et al., 1975; Doolittle, 1981). The molecule has properties both of a globular as well as a fibrous protein that confer on it unusual physicochemical behaviour and a variety of forms observed by electron microscopy. The unusual physicochemical properties have also made crystallisation of the protein a difficult task and only recently a modified form of bovine fibrinogen has been crystallised (Weisel et al., 1978). Earlier, various degrees of crystallisation were achieved only after limited proteolysis(Tooney and Cohen, 1977). As yet no successful diffraction studies of native fibrinogen have been reported.

Electron microscopy provides an alternative means of studying macromolcules and this method was used by Hall and Slayter (1959) to look at the structure of native fibrinogen. Using metal shadowing techniques they proposed a tri-nodular structure which has been supported by later electron microscopy studies (Fowler and Erickson, 1979), and a variety of other methods. These include fibre X-ray diffraction analysis (Weisel et al., 1978), enzymic fragmentation studies of native fibrinogen (Marder, 1970; Mihalyi et al., 1976), differential scanning calorimetric studies (Donovan and Mihayli, 1974) and the nature of amino acid sequence (Doolittle et al., 1978). All these studies are consistent with an elongated polydominal structure shown in figure 2.

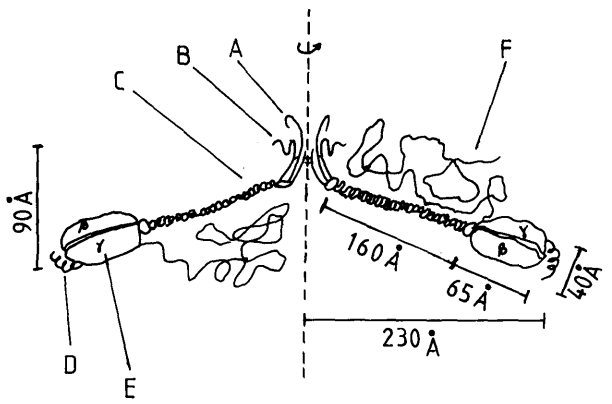

Figure 2. A detailed model of human fibrinogen molecule. A, fibrinopeptide A; B, fibrinopeptide B; C, coiled coil connection between domains; D, γ-chain cross-linking site; E, terminal domain; F, polar carboxy-terminal half of Aα-chain. The central axis depicts pseudo symmetry of the halves contributed by Bβ- and γ-chains in the terminal domains. (Adapted from Doolittle, 1981).

The molecule consists of a central globular domain, which comprises the N-terminal region of Aα-chain (A), Bβ-chain (B) and γ-chain, and two terminal globular domains on either side connected to the middle domain by a supercoiled α-helical region (C) - often called 'coiled coils'. The terminal domains (E) comprise mainly C-terminal Bβ- and γ-chain sequences. The extreme C-terminal sequence of γ-chain forms a protrusion (D) which participates in γ-γ chain cross-linking. The Aα-chain C-terminal sequence floats as a free appendage (F) and participates in α-α chain cross-linking. The molecule has a pseudosymmetry of rotation regarding the terminal domains (E).

An alternative, dodecahedron model of fibrinogen molecule was proposed by Koppel (1966) and supported by Pouit et al., (1972). Both used negative staining electron microscopy. However, none of the studies cited above, such as fibre X-ray diffraction, enzymic fragmentation analysis etc. support this model.

(B) Sequence Studies of Human Fibrinogen Polypeptides

Human fibrinogen, like all vertebrate fibrinogens studied to date, is composed of three pairs of polypeptides held together by a total of 29 disulphide bridges, including two sets of disulphide rings. Amino acid sequence studies of human fibrinogen were initiated by Blomback and colleagues. Using CNBr cleavage they isolated a fragment amounting to 15-20% of the starting mass but 40% of total cysteine residues. This came to be known as the N-terminal

disulphide knot (N-DSK), as it contained the N-terminal region of all three pairs of fibrinogen polypeptides, including fibrinopeptides A and B. Almost the entire region was sequenced and shown to have 51 amino acids of Aα-chain, 118 amino acids of Bβ-chain and 78 residues of γ-chain (Blomblack, 1976). Most of the sequencing was however done by two other groups. Lottspeich and Henschen (1977) reported a virtually complete sequence of γ-chain. The sequence of Bβ-chain and Aα-chain was determined by Doolittle and colleagues (Watt et al., 1978; Watt et al., 1979a; Watt et al., 1979b and Doolittle et al., 1979). Serum Aα-fibrinogen was thus shown to be 610 amino acids in length, Bβ-chain 461 amino acids long and γ-chain 411 amino acids in length. Protein sequence comparisons showed all three chains to be homologous, suggesting a common ancestory. Amino acid sequences also revealed the so called 'coiled coils' region in each polypeptide. Glycine and proline, which are incompatible with α-helix formation were notably few and the occurrence of non-polar amino acids had a distribution in synchrony over large stretches of the three chains. Computer assisted model studies revealed this region to be α-helical (Doolittle et al., 1978). The occurrence of a polar repetitive sequence in Aα-chain was also revealed by sequence analysis. It is known to exist as a highly exposed protruberence in the molecule and is readily cleaved off by virtually all known proteases. This is one of the steps of fibrinolysis during clot dissolution.

1.3.2. Mechanisms of Fibrin Clot Formation and Fibrinolysis

The function of fibrinogen is to provide a readily available soluble precursor to the insoluble matrix of a blood clot that forms after tissue injury. Fibrin clots are gelled and semi-solid masses having an all-pervading network of fibres which trap platelets, erythrocytes and leucocytes. The 'tissue plug' so formed prevents blood loss from a normally closed circulatory system. The triggering event of this transformation of a soluble precursor (fibrinogen) to an insoluble product (fibrin polymers), is thrombin catalysed cleavage of fibrinopeptides A (16 amino acids) and B (14 amino acids). Removal of these relatively small, polar peptides cause the parent molecules to polymerise spontaneously into a fibrin network. The mechanism of this polymerisation is based on the model of the protein shown in figure 2.

Fibre diffraction studies by Bailey and co-workers (1943) revealed no differences between fibrinogen and fibrin. This has been corroborated by later studies (Weisel *et al.*, 1978), suggesting that no conformatinal change occurs during this transformation. However, it was shown that although the molecular weights of fibrinogen and fibrin were virtually indistinguishable, the two could be distinguished electrophoretically with fibrinogen significantly more electron-negative (Mihalyi, 1950). It was concluded that thrombin removed the negatively charged fibrinopeptides. The amino acid sequence of human

fibrinogen revealed the fibrinopeptides A and B to indeed bear a net negative charge. This has also been shown for other species pointing to an evolutionarily conserved clotting mechanism. The spontaneous polymerisation that ensues fibrinopeptide release has been attributed to the elimination of mutually repulsive coulombic forces. At physiological pH (7.3), native fibrinogen has a net negative charge which is spread disproportionately over the central and terminal domains. On cleavage of fibrinopeptide A there is a considerable reduction in the excessive negative charge on the central domain. This, on cleavage of fibrinopeptide B, is reduced so much so that the central domain aquires a net positive charge. The negatively charged terminal domains interact with the positively charged central domain leading to the formation of a polymerisation unit.

Electron micrographs of fibrin reveal a characteristic banded pattern with a repeat distance of 220-240 Å, which is about half the proposed molecular length of fibrinogen (figure 2). This was hypothesised by Ferry (1952) to be a result of polymerisation occurring by halfway staggered overlap of fibrin monomers. Low angle X-ray studies on gel packed fibrin later confirmed the earlier observation of a repeat unit (Stryer et al., 1963), and the staggered overlap model is now accepted. This structure can result only from the accessibility of polymerisation sites on the parent molecule, following removal of fibrinopeptides shielding them. These sites have been designated donor

and acceptor sites often described as knobs and holes, respectively. A set of knobs on the central domain interact with a corresponding set of holes on the terminal domains of neighbouring molecules. The starting molecule being dimeric, causes the initial oligomer to grow in either direction. Release of fibrinopeptides A and B has been shown to be at different rates (Schainoff and Page, 1960). In mammals, cleavage of fibrinopeptide A occurrs first which allows end to end polymerisation so increasing fibre length. Following fibrinopeptide A release there is a conformational change which exposes fibrinopeptide B for thrombin attack (Blomback et al., 1978). Cleavage of fibrinopeptide B causes lateral polymerisation to increase fibre thickness. In this manner a meshwork of fibres begins to form. The clot formed is described as 'soft'- being dependent entirely on coulombic forces holding it together. It is turned into a hard clot by factor XIIIa mediated cross-linking of fibrinogen polypeptides. This factor is activated in much the same way as 'activation' of fibrinogen - thrombin catalysed cleavage the of N-terminal chain of 'a' subunit makes it into a calcium dependent transglutaminase. The peptide bonds formed are between α-α side chains and γ-γ side chains involving lysines (donors) and glutamines (acceptors) resulting in ϵ-(γ-glutamyl)lysine cross-links (Chen and Doolittle, 1970; Folk and Finlayson, 1977). The complete structure of γ-γ cross-links has been determined (Chen and Doolittle, 1971). Lysine donors and glutamine acceptors are both on the C-terminal end of γ-chain. The

joined side chains are antiparallel and bonded by two reciprocal cross-links, eight residues apart, between Gln 308 and Lys 316.

Aα-chain cross-links, unlike γ-chain cross-links, are in multimeric arrays. There are at least two glutamine acceptors and two lysine donors on each chain which participate in cross-link formation, though as yet their position has not been determined. Cross-links between γ-side chains stabilise end-to-end polymerisation and those between α-side chains stabilise lateral polymerisation (Doolittle, 1973), forming a hard clot held together by covalent bonds. This event is followed by platelet aggregation and a multistep healing process ensues.

Wound healing involves fibroblasts moving through the fibrin network and laying down a layer of collagen fibres. The superflous fibrin network is proteolytically dismantled. This is achieved by the activation of plasminogen to plasmin by urokinase. The open nature of fibrin clots and the extended polydominal structure of fibrin facilitates easy liquification of clots due to plasmin cleavage sites being easily accessable and only a few bonds are cleaved for clot dissolution.

Plasmin, a serine protease with trypsin like specificity, hydrolyses the lysyl and arginyl bonds in the fibres. Of an estimated total of 362 potential target bonds only about 50-60 are cleaved, reducing fibrin to its core fragment (Mihalyi et al., 1976). The first plasmin cleavage site has been determined to be 27 residues

from the α-chain C-terminus. The second proteolytic event occurs at the base of the α-chain protruberance (F, figure 2), releasing cross-linked α-fibrinogen side chains (Takagi and Doolittle, 1975a). At the same time a 42-residue peptide is cleaved from the N-terminal of β-chain (Takagi and Doolittle, 1975b). Finally the assault forming the basis of fibrinolysis is launched - cleavage of bonds in the interdominal 'coiled coils', reducing fibrin to its core fragements and releasing γ-side chain cross-linked terminal domains.

1.3.3. The Arrangment and Structure of Fibrinogen Genes

The availability of cloned cDNAs for rat and human fibrinogen (Crabtree and Kant, 1981; Kant *et al.*, 1983; Rixon *et al.*, 1983; and Chung *et al.*, 1983a and 1983b), has lead to the isolation of genomic sequences for fibrinogen polypeptides from these two mammalian species. Bovine fibrinogen cDNAs have been cloned as well (Chung *et al.*, 1981 and 1982), but the isolation of their corresponding genomic sequences has not been reported.

Rat fibrinogen genes have been isolated by Kant and Crabtree (1983) from EcoRI and Hae III genomic libraries. Detailed restriction mapping showed that all genomic isolates hybridizing to a single cDNA could be overlapped, and the Aα- and γ-chain genes appeared to be directly adjacent *in vivo*. There is a 12.8 kb region with identical restriction sites in the 5' region of Aα-chain gene isolates and 3' region of γ-chain gene sequences.

Furthermore, heteroduplex mapping studies showed complete homology spanning the 12.8 kb distance beween 5' end of Aα-chain gene and 3' end of γ-chain gene. This demonstrated that the γ- and Aα-chain genes are directly linked in a 5'-3' direction in vivo, separated by a distance of 12.8 kb (figure 3). From the known orientations of the cDNA probes, it was shown that both genes are transcribed from the same strand. The Aα- and γ-chain gene linkage region has been extended to about 57 kb by 'chromosomal walking'. At the 5' and the 3' ends of this region, highly repetitive DNA sequences were encountered. Similarly, repeats were observed at the 5' end of the Bβ-chain locus and extension of this to encompass 35 kb did not reveal any linkage to the Aα-γ gene cluster. However, the isolation of human fibrinogen genes has revealed linkage of Aα- and Bβ-chains genes (J.Kant, personal communication).

Genomic structure of rat fibrinogen genes has been studied by DNA sequencing and R-loop analysis (Crabtree and Kant, 1983). The γ-chain gene has 8 exons spannning 7 kb (figure 3). The Aα-chain gene spans 6.6 kb and contains at least 5 exons and 4 intervening sequences. There may be small exons in the gene not detected by R-looping. The structure of Bβ-chain has also been studied using R-looping and shows the presence of 6 exons.

Human Bβ-chain gene has been isolated by Chung et al., (1983a). R-loop mapping studies showed the gene to have 7 intervening sequences spanning about 10 kb. One intervening

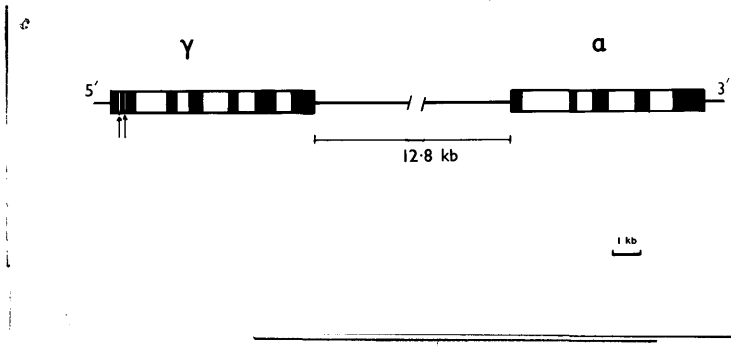

Figure 3. The arrangement and structure of rat Aα- and γ-fibrinogen genes. Coding regions are in black. Introns (intervening sequences) are in white. Arrows indicate the position of small introns.

sequence was 1.3 kb long and localised to the 5' end of the gene by sequencing. Sequence data also revealed that the the nucleotides coding for the signal sequence and first 8 amino acids of fibrinopeptide B constitute the first exon, preceeding the 1.3 kb intron. Altogether, 7 intervening sequences in the gene were shown to be present by selective sequencing of gene subclones, thus confirming earlier R-loop studies.

Chung et al., (1983a) suggest that although not all the structural and funtional domains of Bβ-chain have been defined, the arrangement of intervening sequences in the gene seems to be consistent with the hypothesis that introns serve to segregate a gene into structural domains (Gilbert, 1982). For example, the first exon comprises the signal sequence and first 8 amino acids of fibrinopeptide B. The second exon has thrombin cleavage site and fibrin polymerisation region. Similarly, the connecting region between interdomainal structure; including the disulphide ring forming cysteins, and the super helical folding sequence are respectively encoded by the third and fourth exons.

1.3.4. Biosynthesis and Assembly of Fibrinogen

Fibrinogen polypetides are synthesised in liver parenchymal cells (Forman and Banhart, 1964). Bβ-chain is glycosylated at residue 352 and γ-chain at residue 52 (Topfer-Peterson, 1976), while Aα-chain is free of any carbohydrate moiety (Pizzo et al., 1972). The process

by which hepatocytes synthesise, assemble and secrete fibrinogen is still not fully elucidated and is currently receiving intense scrutiny. It was suggested several years ago that a half molecule of fibrinogen might be synthesised from a single chain precursor undergoing proteolytic cleavage after folding of the non-identical chain regions, in the way insulin is formed from pro-insulin (Doolittle, 1973). This suggestion was prompted mainly by the apparently overwhelming problem of positioning and synchronising 29 disulphide bonds.

All the evidence gathered since that suggestion has however opposed such a scheme. Instead, the chains are specified by three discreet mRNAs coded by separate genes. Earlier studies had shown that the sizes of the mRNA identified immunologically were not long enough to code for a half-molecule (Bouma et al., 1975; Chung et al., 1980). In vivo pulse label studies in dogs were consistent with independent and simultaneous translation of three polypeptides (Kurdyk et al., 1979,1982). Experiments employing in vitro translations also indicated the existence of three discreet products, each slightly larger than the normal native chain, which indicated the existence of leader sequences that aid in transmembranous movement preceeding packaging and secretion (Nickerson and Fuller, 1979; Chung et al., 1980). Sequencing of cDNAs for human fibrinogen has confirmed the existence of these signal peptides (Kant et al., 1983; Rixon et al., 1983; Chung et al., 1983a and 1983b).

Additionally, for Aα-chain a 15 amino acid peptide at the C-terminus absent from serum Aα-chain, was observed (Rixon et al., 1983; Kant et al., 1983). It is not clear whether this propeptide is cleaved as the protein is secreted into the hepatocyte endoplasmic reticulum or cleavage occurs in the plasma by specific proteases after secretion.

The manner of assembly of the polypeptides to form a native, functionally active fibrinogen molecule has still to be elucidated. One suggestion is that the nascent chains are assembled while still bound to polysomes (Bouma et al., 1975). Presumably the driving force for such a mode of assembly would be the strong tendency of the chains to form coiled coils (Doolittle et al., 1978) stabilised by the formation of disulphide rings at their periphery; the rest of the molcule folding by itself.

1.3.5. Variant Human Fibrinogens

A large number of human fibrinogen variants have been reported in the last two decades (see Bloom, 1981, for a complete list). The vast majority have been detected and categorised on the basis of long thrombin times or altered reptilase times when compared to normal fibrinogens. Congenital disorders of fibrinogen synthesis can be divided mainly into two types:

(1) Deficient synthesis - Afibrinogenaemia
- Hypofibrinogaemia

(2) Defective synthesis - Dysfibrinogenaemia

Afibrinogenaemia is characterised by a complete absence (or virtually complete absence) of circulating fibrinogen, whereas in hypofibrinogenaemia there are reduced levels of apparently normal fibrinogen. It is not known whether the basic lesion in patients exhibiting deficient synthesis is related to gene abnormalities, such as deletions or promoter mutations. Examples of deficient synthesis confined to one chain have not been reported, suggesting a regulatory disorder affecting coordinate synthesis of fibrinogen polypeptides.

In dysfibrinogenaemia circulating fibrinogen exhibits normal or reduced amounts (hypo-dysfibrinogenaemia). These qualitative disorders of fibrinogen are seemingly due either to point mutations causing amino acid changes or to altered post-translational events. Alternatively a processing defect might operate either at the RNA or at the protein level.

1.3.5.1. Afibrinogenaemia and Hypofibrinogenaemia

(A) AFIBRINOGENAEMIA. The frequency of this disorder is very low in the human population. Absence of fibrinogen in humans was first described in 1920. Since then a little over a hundred patients have been recorded (Mammen, 1974; Flute, 1977). All coagulation assays in these individuals are normal, excepting fibrinogen, when measured by conventional clotting, chemical or immunological methods. Only by the use of very sensitive immunochemical techniques can trace amounts of fibrinogen be detected.

The disease is not restricted to humans and has been

observed in other mammals (Kammerman et al., 1971; Brevkink et al., 1972). Cosanguinity is often observed in parents and the disorder is autosomally inherited. Bleeding starts in first years of life and may be accompanied by other clinical syndromes, such as haemarthroses and menorrhagia. Cerebral haemorrhage has been noted as the cause of death.

(B) HYPOFIBRINOGENAEMIA. Parents of hypofibrinogenaemics show reduced levels of apparently normal fibrinogen, consistent with the heterozygous state. It is likely that hypofibrinogenaemia is a heterozygous form of afibrinogenaemia (Bloom, 1981). The condition is 'recessive' by clinical expression but'dominant' by laboratory criteria (Flute, 1977). Some patients thought to be suffering from hypofibrinogenaemia were later shown to be of a dysfibrinogenaemic sub-type (Jackson and Beck, 1970) and so labelled 'hypo-dysfibrinogenaemics'.

1.3.5.2. Inherited Dysfibrinogenaemias

Dysfibirnogenaemias are a heterogeneous group of disorders which appear to be due to a qualitative defect in fibrinogen molecule. The condition is inherited as an autosomal trait.

The first documented report of abnormal fibrinogen as an inherited disorder - a dysfibrinogenaemia - was by Menache (1964), later called fibrinogen Paris I. The disorder was characterised by an absence of functional fibrinogen. Presence of the protein was however

confirmed by immunological methods. Since then over 100 kindereds with dysfibrinogenaemias have been identified. However, as comparative investigations are still incomplete, the exact number of distinct variants is not known.

Heterogeneity of disfibrinogenaemias is reflected in the occurrence of variants such as fibrinogen Bethesda II (Gralnick et al., 1973), Bethesda III (Gralnick et al., 1979) and Philadelphia (Martinez et al., 1974); where reduced levels of circulating fibrinogen have been attributed to increased catabolism of the defective molecule. For fibrinogen Leuven (Verhaeghe et al., 1974), New York (Al-Mondhiry et al., 1975) and Valencia (Aznar, 1974), low levels of defective circulating fibrinogen have been ascribed to descreased synthesis (hypo-dysfibrinogenaemia). Further complications have sometimes been observed. In a family with fibrinogen St. Louis, factor VIII deficiency was also seen to be segregating (Sherman et al., 1972). In another family dysfibrinogenaemia was associated with von Willebrand's syndrome (Owen et al., 1979). Despite these anomalies an underlying pattern has been observed in studying all the dysfibrinogenaemias reported so far and the disorder has been clasified into three main types:

(a) Delayed fibrinopeptide release eg. fibrinogens Detroit, Bethesda II, New York, Valencia and Leuven.

(b) Delayed fibrin polymerisation eg. fibrinogens Paris I, Bethesda III and Philadelphia.

(c) Abnormalities of cross-linking or Plasmin mediated

digestion eg. fibrinogens Oslo and Tokyo.

Clinical features of dysfibrinogenaemias are not synonymous with excessive bleeding in all cases. In some families, presumed heterozygotes are asymptomatic though biochemical coagulation abnormalities do exist. Generally, bleeding tendency is more commonly associated with defective fibrinopeptide release rather than delayed fibrin polymerisation. For example, fibrinogen Detroit homozygotes show high tendency to bleed but Leuven homozygotes have no clinically apparant features (Bloom, 1981). Fibrinogen Oslo (Egeberg, 1967) on the other hand is hypercoaguable with reduced clotting time and a high incidence of venous thromboembolism.

Most variant fibrinogens found so far occur in heterozygotes and the problem of separating normal and variant molecules is formidable. Consequently, the molecular defect in most variants has not been defined. Fibrinogen Detroit remains the only abnormal human fibrinogen in which the exact nature of the defect has been elucidated - a substitution of a serine for an arginine at residue 19 of the Aα-chain (Blomback, 1968). Unexpectedly, this mutation does not affect fibrinopeptide A release but causes steric hinderance making Bβ-chain inaccessable to thrombin attack (Kurdyk et al., 1976). An important practical point drawn from this result was that abnormality of fibrinopeptide release does not of necessity localise the molecular defect to that chain.

Lately another variant has been receiving intense

scrutiny. Fibrinogen Paris I (Menache, 1964) has an extended C-terminal γ-chain sequence (Budzynski et al., 1974) which does not participate in γ-γ chain cross-linking. The γ-chains from individuals with this disorder express features of both the nomal γ-chain and the variant. Differences between the two might arise due to differential splicing of a transcript from the same gene, giving rise to mRNAs specific for the normal γ-chain and its Paris I counterpart. There is indeed a precedence for such an event. Crabtree and Kant (1982), have shown that the two non allelic forms of γ-chains in rats, called γ_A and γ_B, arise by transcription of two mRNAs (1.7 kb and 2.2 kb) produced from a single gene by alternative splicing. The γ_B mRNA is identical with the γ_A sequence except for a 513 bp insert located 202 bp from the poly(A) tail. This insert is identical to the seventh and last intron of γ-fibrinogen gene (figure 4) suggesting that this intron is not spliced out in γ_B mRNA. It has been suggested that the γ' variant of human fibrinogen is also produced in a similar way (Wolfenstine-Todel and Masesson, 1980 and 1981).

Although a number of genes have been described that produce more than one mRNA by alternative splicing, γ-fibrinogen gene is different in that the two mRNAs are produced concurrently in the liver. This in contrast to α-amylase and immunoglobulin mRNAs which arise by differential splicing being associated with a developmental stage.

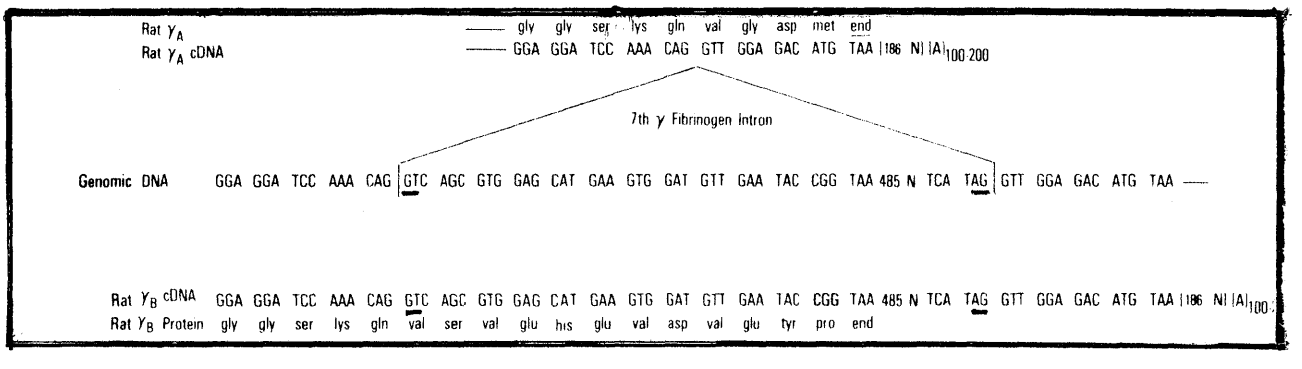

Figure 4. Alternative RNA splicing producing the γA and γB chains of rat fibrinogen. The consensus splice sites are underlined. (Adapted from Crabtree and Kant, 1982).

The position and sequence of the splice junction around the seventh intron, which remains unspliced in rat γB mRNA, are in agreement with the known consensus splice junctions (Breathnach et al., 1978; Lewin, 1980; Rogers and Wall, 1980). This suggests that there may be other sequences that are required for efficient splicing. The alternative splicing may reflect some intrinsic property of the RNA transcript so that up to 10% of mRNA transcripts fail to splice out the intron sequence. The relative abundance of γA and γB mRNAs may be determined by stochastic considerations (Crabtree and Kant, 1982).

1.3.6. Vertebrate Fibrinogen Evolution

Vertebrate fibrinogen has an ancient history. Even the most primitive of the vertebrates, the lamprey, has a fibrinogen molecule akin to the mammalian counterpart. There are two sets of three non-identical chains Aα , Bβ and γ that form this molecule and thrombin cleavage releases fibrinopeptides A and B (Doolittle, 1965). The junctions attacked by thrombin are the same as those attacked in mammalian fibrinogens. The similarity extends further to fibrin stabilisation by factor XIII giving γ-γ dimers (Doolittle and Wooding 1974). This is much in contrast to the haemoglobin molecule, which in lampreys exists as a single chain; aggregating only under reducing conditions. However, unlike vertebrate haemoglobins which have antecedents amongst invertebrates showing homology in primary sequence

and tertiary structure to the vertebrate form (Dayhoff, 1972), no such antecedents to fibrinogen have been found in these organisms. Lobster 'fibrinogen' has a very different architecture (Fuller and Doolittle, 1971) as does horseshoe crab coagulen (Solum, 1973). This leaves unanswered the question of where and when vertebrate fibrinogen evolved, though the presence of a fully differentiated form of the molecule in lampreys suggests a very ancient heritage and indicates that the three chains diverged perhaps as long ago as a billion years.

Amino acid sequence comparisions of fibrinogen polypeptides show the chains to be homologous and to have arisen by an ancestral sequence undergoing duplication and divergence (Doolittle, 1976: Doolittle et al., 1979). Homology between Bβ- and γ-chains is greater than that of either to the Aα-chain (Henschen and Lottspeich, 1977; Watt et al., 1978) indicating they have diverged more recently than the divergence of their ancestor from the Aα-chain. It has been estimated that the divergence of Bβ- and γ-chains occurred about 500 million years ago, while that of Aα-chain occurred 1500 million years ago (Henschen et al., 1980; Doolittle, 1980).

Selection pressure determines the rates of change of different parts of a protein and this is well illustrated by fibrinogen. The γ-γ cross-linking sites, which are critical for fibrin polymer stabilisation are highly conserved, there being only two replacements in 20 residues between bovine and human (Chen and Doolitte, 1971).

In contrast to this the fibrinopeptides, which play a rather non-specific masking role, are among the least conserved sequences studied to date (Doolittle and Blomback, 1964; Doolittle 1981). Not only do they show lack of sequence conservation, there is also no selection for size as demonstrated by the unusually long and varying size of fibrinopeptide B of non-mammalian vertebrates such as Xenopus (Wangh et al., 1983) chicken (Pindyck et al., 1977) and salmon (Murtaugh and Gladner, 1981).

The Aα-chain is evolving far more rapidly than either the Bβ- or γ-chains (Kant et al., 1983). This is reflected in the relatively low identity of 57% at the cDNA level between human and bovine Aα-chains (Rixon et al., 1983). This overall low identity is due to a hypervariable region in the chain spanning amino acids 472-552 (Henschen et al., 1980). Human and bovine Aα-chain cDNAs share no homology in this region. There is much speculation as to the origin of the hypervariable region, and it has been suggested that it may be due to differences in intervening sequence processing (Rixon et al., 1983).

The amino acid sequence of Aα-chain shows the presence of 10 tandem repeats consisting of 13 residues from amino acids 264 to 391 (Doolittle et al., 1979). The cDNA sequence of Aα-chain has shown this to originate from an 8 fold duplication event of a 39 bp sequence beginning with Pro 270 and ending with Glu 370 (Kant et al., 1983). The repeat is derived from a smaller repetitive unit; the dominant one being 5'CTGGAA3', occurring 13 times. Other

repeats occur less frequently. Occasionally these repeats are flanked by a 6 bp element of sequence 5'GGAACC3'. Smaller repetitive units have been observed as well.

The 13 amino acid repeat in the protein has clearly arisen from the duplication of the 39 bp genetic element and not by convergent evolution of dissimilar sequences. Analysis of the intron-exon structure of this region in the Aα gene might shed light on the evolution of this repeat. Duplicated genetic elements of smaller sizes have been observed in collagen (Yamada et al., 1980; Wozney et al., 1981) and α-fetoprotein (Eiferman et al., 1981). Duplications of large regions followed by divergence have led to genetically related proteins such as albumin and α-fetoprotein. The fibrinogen gene family exhibits duplications at both levels.

1.4. Aims and Objectives

The aim of this project was to isolate human Aα-, Bβ- and γ-fibrinogen cDNA clones for the analysis of fibrinogen genes in the human genome, and to study their inheritance in the human population.

The isolation and fine structure mapping of fibrinogen genomic sequences was to be the basis of identifying differences between normal and certain variant human fibrinogens, thus defining the molecular nature of genetic changes that either lead to the synthesis of a dysfunctional molecule or prevent normal expression of

fibrinogen genes.

1.4.1. Approaches for the Isolation and Chromosomal Assignment of Human Fibrinogen Sequences.

Eukaryotic DNA sequences have been isolated using three main approaches. Firstly, eukaryotic DNA recombinants encoding selectable genetic markers may be isolated by "rescue" procedures following transformation of eukaryotic cells under appropriate selection conditions (Lowy et al., 1980; Perucho et al 1980; Pellicer et al., 1980). Recent improvements, utilising cosmid vectors, have increased the efficiency of gene rescue (Lund et al., 1982; Lindermain et al., 1982). Secondly, it is possible to enrich mRNA species prior to cloning according to their physical and biological properties; employing size selection (Maniatis et al., 1976; Odink et al., 1981), polysomal immunoprecipitation (McGillivray et al., 1980; Robson et al., 1982) and biological inducibility (Anderson and Schimke, 1976; Axel et al., 1976; Crabtree and Kant, 1981). Finally, large representative cDNA libraries have become available which can be screened with synthetic DNA probes or immunological procedures. Recently a chicken tropomyosin cDNA was isolated using a specific antibody from a cDNA library constructed in an expression vector (Helfman et al., 1983). However, this technique is in its initial stages and other examples of successful applications are only in model systems (Broome and Gilbert, 1978; Villa-Komaroff et al., 1978; Clarke et al., 1979).

The use of synthetic oligonucleotides (oligomers) as probes for the direct screening of cDNA libraries has been particularly successful. Its technical simplicity and reliability gives it an advantage over the other approaches mentioned above. Synthetic DNA probes were therefore used to isolate fibrinogen sequences from an adult human liver cDNA library. The clones were used to study the organisation of fibrinogen genes in the human genome and their expression in human liver. Chromosomal regional assignment of human Aα-, Bβ- and γ-fibrinogen genes was performed by a somatic cell hyrid panel. The next section deals with the development, current uses and trends in synthetic DNA technology, and a brief review of gene mapping by somatic cell genetics.

1.5. Synthetic Oligonuclotides

1.5.1. An Overview

Synthetic oligonucleotides have played and continue to play a major role in the conceptual and technical advance of molecular biology. Some of the early work on deciphering the genetic code rested on the use of polyribonucleotides (Nirenberg and Mathaei, 1961) synthesised by polynucleotide phosphorylase - an enzyme which under controlled conditions adds predominantly a single nucleotide to a growing oligonucleotide strand. This approach has lately been used by Smith and his colleagues to synthesise oligodeoxynucleotides and larger stretches of DNA (Gillam et al., 1978 and 1979). Chemical

synthesis of DNA, developed by Khorana and co-workers (reviewed by Khorana, 1979) and by Itakura et al., (1975) has now superseded enzymatic DNA synthesis. The first successful application to biology of this technique was the synthesis of a complete yeast alanine tRNA gene by the diester method (Agarwal et al., 1970; Khorana et al., 1972). This has now been replaced by the more efficient triester method (Itakura et al., 1975). It differs from the diester method in having an extra protective group which allows the construction of a library of fully protected triester trimers. These may then be joined to give oligomers of increasing length. Some recent improvements in the method involve, block coupling of large oligomers, the use of HPLC for purification of intermediates and products, and the introduction of solid-phase synthesis. The latter renders the procedure amenable to automation. At present much effort is being focused towards construction of fully automated DNA synthesisers.

Ever since the development of recombinant DNA, chemical synthesis of DNA has ceased to be regarded merely as an intellectual exercise. Mutual interaction of the two methodologies resulted in the expression in E.coli of functional somatostatin hormone from a cloned chemically synthesised gene (Itakura et al., 1977). Earlier results from the cloning of a synthetic lac operator (Heyneker et al., 1976; Bahl et al., 1976) and a tyrosine suppressor gene (Ryan et al., 1977) had demonstrated that chemical

synthesis was capable of producing biologically functional DNA. However, only after the production of somatostatin as a biologically active product was the true potential of this approach fully realised. This prompted the chemical synthesis and expression in E.coli of other hormone genes, such as insulin (Crea et al., 1978; Goeddel et al., 1979a), thymosin (Wetzel et al., 1980) and human growth hormone (Goeddel et al., 1979b).

Total gene synthesis of other clinically important products and their expression in E.coli, has more recently been acheived for human β-urogastrone (Smith et al., 1982) and human leukocyte interferons (Edge et al., 1981 and 1983). For the latter, a 514 bp fragment, specifying the initiation and termination signals, and the 166 amino acid residues of human interferon -α1, was synthesised from a total of 66 oligonucleotides ranging in size from 14 to 21 bases. Such a feat would not have been possible without the developments of new solid-phase triester methods for oligonucleotide synthesis (Markham et al., 1980; Gait et al., 1980).

The flexibility of chemical gene synthesis allows the incorporation of desired features in an expressed product. Smith et al., (1982) took advantage of the known stability of β-urogastrone to trypsin protease and introduced a Lys-Lys sequence immediately preceeding the first amino acid of the hormone. The mature polypeptide could then be purified from the fused product expressed in E.coli, by trypsin cleavage of this sequence. Another feature of

chemical gene synthesis is the incorporation of preferred E.coli codons where possible, and the introduction of restriction sites both terminally and internally which respectively allow increased efficiency of expression in the host and further manipulation of the gene.

1.5.2. Oligonucleotide Mediated Priming of RNA Sequences

The use of oligonucleotides as primers for reverse transcriptase has assisted the cloning of specific gene sequences. The priming of RNA with oligo(dT), for cDNA synthesis using reverse transcriptase, repesents a non-specific event designed to reverse transcribe all oligo(dT) hybridising poly(A)$^+$ mRNA squences. Specific priming of RNAs for the purpose of enriching for a particular sequence requires some knowledge of either the protein or its RNA sequence. Villa-Komaroff et al.,(1978) were the first to demonstrate the feasability of this approach. The first two nucleotides preceeding the poly(A) tail in rat pro-insulin RNA were known, so a 'specific' primer incorporating these nucleotides with a shortened oligo(dT) was used for priming resulting in a selective enrichment of pro-insulin cDNA. For most sequences of interest however, such features are not known so an alternative scheme was used, employing oligomers synthesised to small regions of the protein sequence showing low ambiguity of codon assignment. Ambiguity of the oligomer sequence was further reduced from the knowledge of codon utilisation frequencies (Grantham et

al., 1980). Gastrin cDNA cloning by specific priming of hog antral mucosa poly $(A)^+$ RNA with a dodecamer of consensus sequence derived from codon utilisation frequencies, was reported by Noyes et al., (1979). Several groups have cloned specific sequences using this procedure (Chan et al., 1979; Goeddel et al., 1980; Sood et al., 1981.) The main disadvantage of this approach is its dependence on an exact match between a primer and its corresponding mRNA sequence allowing synthesis of cDNA (Houghton et al., 1980). The use of oligonucleotide mixtures has partially offset this problem (Shoulders et al., 1982) though there may be a need to characterise primed cDNAs even before cloning.

A more disturbing aspect of the procedure is the occurence of mis-match priming giving unexpected cDNAs. This has been observed in this department (J.T. Rogers, personal communication) as well as by others (Shoulders et al., 1982; Panabieres et al.,(1982).

1.5.3. Recent Developments and Current Trends in Oligonucleotide Technology

(a) DIRECT OLIGONUCLEOTIDE SCREENING.

Improvements in screening techniques have resulted in the establishment of favourable conditions for direct oligonucleotide screening of recombinant cDNA clone banks. The general approach, pionered by Wallace et al., (1981) is to chemically synthesise a mixture of oligonucleotides representing all possible codon

combinations for a given region of the protein; usually 5-6 amino acids long. The mixture has only one sequence that is perfectly complementary to the DNA encoding this region. Screening of a cDNA library with this mixture under stringent hybridisation conditions results in only the perfectly matched sequence to form a stable duplex with its complementary DNA. This method allows the isolation from an appropriate library, of cloned DNA sequences of any protein for which amino acid sequence data is available.

Hybridisation specificity of an oligomer is dependent on parameters such as length, base composition and the possibility of a mismatch. Oligonucleotide specificity increases with stringent conditions of decreasing ionic strength and increasing temperature, while signal strength decreases under these conditions. Factors affecting oligonucleotide hybridisation specificity have been determined and discussed by Wallace et al., (1979 and 1981) and Szostak et al. (1980).

The first successful application of this technique was reported by Suggs et al., (1981). Using a mixture of synthetic oligonucleotides they isolated a β_2-microglobulin clone from a cDNA library under previously determined hybridisation conditions (Wallace et al., 1981). A number of cDNA clones have now been isolated using this approach, which bear testimony to its general applicability.

(b) GENE DETECTION BY OLIGONUCLEOTIDES

Hybridisation of synthetic DNA probes to eukaryotic DNA was initially used for the detection of yeast CYC1 gene (Montgomery et al., 1978; Szostak et al., 1980). Technical improvements have recently led to gene detection in mammalian genomes using synthetic oligomers. Increased sensitivity of the technique allows alleles differing by single point mutations to be distinguished. Conner et al., (1983) used nanodecamer oligonucleotides to distinguish the normal human β-globin (β^A) allele and the sickle-cell β-globin (β^S) allele in the human genome, under controlled conditions. Their results showed that not only is it possible to detect single copy genes in the human genome using synthetic probes, but the method is sensitive enough to distinguish the β-globin genotype of individuals. A similar approach has been used for the detection of α1-antitrypsin deficiency by direct analysis of the mutated gene (Kidd et al., 1983). Direct detection by oligomers of the common Mediterranean β-thalassaemia gene has also been reported (Orkin et al., 1983). The technique has the potential to be applied to the diagnosis of any genetic disease involving a specific change in DNA sequence, particularly base substitutions.

(c) OLIGONUCLEOTIDE-DIRECTED MUTAGENESIS

Site-directed mutagenesis can be achieved by a number of currently available methods (Lathe et al., 1983). One of the most reliable method of obtaining a desired

mutation is provided by oligonucleotide-directed mutagenesis. With increasing availability of synthetic oligomers, it is likely to become the method of choice for the introduction of precise mutations in any DNA fragment. As this procedure has been recently reviewed in depth by Smith and Gillam (1981), it will not be dealt with here in detail.

The precision of this method is reflected by its ability to generate deletions of a single base (Gillam et al., 1980), three bases (Miyada et al., 1982) or more. Using a 21 base oligomer, Wallace et al., (1980) were successful in achieving a precise 14 base deletion of a yeast tRNA gene intervening sequence. It may also be used to give point mutations, such as those used to convert mammalian tRNA genes into functional amber suppressors by site-specific mutagenesis of the anticodon sequence (Laski et al., 1982; Temple et al., 1982). Similar site-specific mutagenesis altering the enzyme active site (Dalbadi-McFarland et al., 1982; Winter et al., 1982) makes it a powerful and general method for protein function studies.

One major drawback of the procedure has been low yields of mutant progeny due to excision repair either in vitro by Klenow polymerase or by DNA repair in vivo. Certain modifications of the steps involved have now increased mutant progeny frequency (Kramer et al., 1982; Zoller et al., 1982; Norris et al., 1983) which will increase the efficacy of this technique.

1.6. Somatic Cell Genetics and Mammalian Gene Mapping

Until recently it was not possible to map the mammalian genome with a resolution comparable with that of procaryotes. This was due primarily to the complexity of the mammalian genome, the lack of genetically defined mutants and the difficulty in obtaining recombinants. Two approaches that have reversed this trend are recombinant DNA techniques and somatic cell genetics.

Recombinant DNA methodologies have facilitated the isolation and sequencing of several genes together with their flanking regions thus mapping functional regions, such as transcriptional promoters and identifying point mutations and deletions. This high resolution mapping is contrasted by somatic cell genetics used to map genes at the level of chromosomes. More than 200 human genes have been assigned to chromosomal regions in the human genome using this method (Ruddle, 1981). Several more genes have been mapped either by linkage or by other mapping procedures such as chromosome cloning (Ricciuti and Ruddle, 1973) and *in situ* hybridisation (Gall and Pardue, 1969; Harper and Sanders, 1981).

Gene mapping by somatic cell genetics depends on parasexual events which facilitate the transfer of DNA from a donor cell, serving as the source of such material, to a recipient cell into which the donor DNA is delivered. The resulting heterogenote or hybrid cell segregates donor and recipient chromosomes allowing mapping of the donor genes. Altogether there are four ways of gene delivery, namely;

gene transfer by microcells (Fournier and Ruddle et al., 1977), chromosome-mediated gene transfer (McBride and Ozer et al., 1973; Klobutcher and Ruddle, 1981), DNA-mediated gene transfer (Wigler et al., 1977) and somatic cell hybridisation, with which I shall be primarily concerned.

1.6.1. Gene Mapping in Somatic Cell Hybrids

Somatic cell hybridisation is the fusion of two genetically distinct parental cells (for example, rodent X human) yielding hyrid cells which initially contain the full complement of parental genomes. In such interspecific hybrids however, there is almost invariably a partial loss of chromosomes of what has been termed, the donor cell. This provides a unique opportunity to map the genes of the donor by correlating the loss or retention of a particular chromosome with the disappearance or maintenance of a given gene.

The human genome has proved particularly amenable to this form of gene mapping due to a preferential loss of human chromosomes from rodent X human cell hybrids (Weiss and Green, 1967). The choice of cell types for the formation of such heterospecific hybrids is important though, as in rodent X human hybrids of cetain cell types there is a preferential loss of rodent chromosomes (Colten and Parkman, 1972; Croce, 1976). It seems that on the one hand selection acts to favour growth of hybrid cells with reduced genome complements, and on the other a given donor chromosome might be retained precisely because

it provides a selective advantage. This argument has always stood in the way of establishing whether chromosome loss is random. Croce et al., (1973) showed that hybrids of SV40 transformed human cells with untransformed mouse cells tended to retain human chromosome 7, possibly because this chromosome carries an integrated SV40 genome which confers a selective growth rate advantage on the cell hybrid.

Cultured somatic cells and somatic cell hybridisation have jointly overcome the difficulties in obtaining recombinants and genetically defined mutants for mammalian gene mapping. Cell hybridisation offers an approach that can be described as a form of recombination between parental chromosomes. The fusion of two cells generates a tetraploid hybrid and loss of chromosomes from this results in different combinations of parental chromosomes.

Somatic cells in culture have provided mutants that have been used for gene mapping by cell hybridisation and growth of hybrids on selective media. In this way a number of human genes have been mapped after fusion of normal human (donor) cells with cultured mouse (recipient) cell lines mutant in a given gene. Human thymidine kinase (tk) gene was the first gene to be mapped in this manner. After fusion of tk^+ human cells with tk^- mouse cells and growth of hybrids on HAT medium, the gene for cytosolic thymidine kinase was localised to human chromosome 17 (Weiss and Green, 1967). Similarly HGPRT was mapped to

the human X-chromosome (Nabholz et al., 1969), and APRT to chromosome 16, by alanosine selection (Kusano et al., 1971; Tischfeild and Ruddle, 1974).

The introduction of the hybrid clone panel (Creagan and Ruddle, 1975) has provided an alternative approach. Each clone in this panel has a small number of human chromosomes, comprising a subset of the human complement. As few as 8-10 hybrid clones carrying suitable subsets of human chromosomes is deemed sufficient to define all 24 chromosomes. Gene assignment using a hybrid clone panel can be by correlating a particular gene product (such as a surface antigen, a structural protein or an enzyme) with a donor chromosome or chromosomal fragment. For example, liver-type-6-phosphofructokinase isozyme (PFKL) has been assigned to human chromosome 21 by detection of the enzyme in somatic cell hybrids using a monoclonal antibody (Vora and Francke, 1981). The method being dependent on the expression of the donor gene precludes mapping of genes expressed only in certain differentiated states, or of sequences having no coding function. The use of nucleic acid hybridisation techniques to detect donor sequences directly has overcome this drawback. Its feasability was first demonstrated when human α- and β-globin genes were assigned to chromosomes 16 and 11 respectively, by solution hybridisation of cDNAs to DNA prepared from a mouse X human somatic cell hybrid panel (Deisseroth et al., 1977 and 1978). Success of this method depends on the absence of appreciable cross-hybridisation with recipient cell DNA.

This represents a serious limitation and has been eliminated by restriction fragment mapping which has the advantage of being independent of interspecies cross-hybridisation of nucleotide sequences. It depends only on restriction fragment length differences observed between species and using this procedure Swan et al., (1979) mapped the mouse immunoglobulin κ-chain gene to chromosome 6 in a mouse X CHO hybrid clone panel.

The applicability of restriction fragment mapping is evident from the ever increasing number of mammalian genes being mapped using this approach. It is also being used to assign sequences of unknown function derived from chromosome specific libraries. In this department, for instance, the mapping of chromosomes 4, 19 and X is being carried out using isolates from libraries specific to these chromosomes (Murray et al., 1982).

1.6.2. Resolution Power of Mapping by Somatic Cell Genetics

Somatic cell genetic procedures are low resolution mapping methods as they can be expected to assign genes to chromosomes and then within a chromosome to a genetic distance of only 5-10 centimorgans (Ruddle, 1981). In this respect, translocation mapping has been particularly useful as a regional mapping technique. Exceptions to the joint inheritance of syntenic genes occur whenever there is a spontaneous chromosome break or translocation either in the parental cell or in the hybrid cell line. Miller et al., (1971) obtained hybrid clones possessing human HGPRT but

lacking G6PDH due to post-fusion translocation, thus aiding regional mapping of the latter. Similarly, a translocation removing about 25% of chromosome 1 short arm caused segregation of the genes for phosphoglucomutase-1 and 6-phosphogluconate DH from the Pep C locus (which remained on the truncated autosome). This placed the genes for both enzymes in the deleted region (Douglas et al., 1973)

An alternative approach involving deletion mapping of chromosomes has been used by Gusella et al. (1980). From a human X CHO somatic cell hybrid containing the single human chromosome 11, they constructed a series of hybrids in which varying portions of the chromosome had been deleted. Using these hybrids they were able to map the isoenzyme markers lactate dehydrogenase A (LDHA) and acid phosphatase-2 (ACP-2) to specific regions of chromosome 11 short arm.

A further development of this approach has seen the utilisation of a repetitive sequence as a probe (Gusella et al., 1982). This sequence occurs about 2000x in the human genome and is present on chromosome 11. DNAs prepared from the original hybrid containing the single human chromosome 11, and from hybrids of its deletion mutants are probed with the repetitive human DNA sequence to detect fragments containing copies of the repeat. Correlation of the presence or absence of a fragment containing the repeat with that of a biochemical or immunological marker is used to order deletion mutants. Loss of fewest markers represents a single terminal

deletion from one arm, used as the starting point in ordering markers along the chromosome.

This method can also be applied to other human chromosomes. Once a particular copy of the repeat is localised near a region of interest, the repetitive sequence may be cloned and flanking single copy sequences used as probes to identify mutations, chromosomal rearrangements and to study defects involved in human genetic diseases

2. Materials and Methods

2.1. Materials

2.1.1. Chemicals

Except for the following, all chemicals were Analar grade supplied by British Drug Houses (BDH), Poole, England. Trizma (Tris-base), MOPS, sodium chloride, ethidium bromide, Orange G, bromophenol blue, xylene cyanol, dextran sulphate and XGal (5-Bromo-4-chloro-3-indoyl-β-D-galactopyranoside) were from Sigma Chemical Co., St. Louis, Missouri, USA. IPTG (Isopropyl-β-D-thiogalactopyranoside), and urea were purchased from Bethesda Research Laboratories (BRL), Cambridge, England.

2.1.2. Isotopes

All radio isotopes were obtained from the Radiochemical Centre, Amersham, England.

2.1.3. Chromatography, Electrophoresis etc.

Sephadex and Ficoll were supplied by Pharmacia, Upsalla, Sweden; and polyvinyl pyrolidone and Bovine Serum Albumin (BSA Fraction V) were from Sigma Chemical Co. Agarose was from BRL, Cambridge, England. Hypaque sodium was from Winthrop Laboratories, Suburbiton-upon-Thames, Surrey. Acrylamide and bis acrylamide were obtained from BDH, Poole, England.

2.1.4. Photography and Autoradiography

FP4 film, Phenisol and PQ Universal developer and Hypam Fixer were from Ilford, Basildon, England. Medical X-ray films were either RX film from Fuji Photo Co. Ltd., Tokyo, Japan or XAR5 film from Eastman Kodak Co., Rochester, New York, USA. Calcium-tungstate-phosphor intensifying screens were either from X-ograph Ltd., Malmesbury, Wiltshire, England or from Okamoto, Japan. Short wave ultra violet light source was a transilluminator from UV Products Inc., Gabriel, California, USA.

2.1.5. Scintillation Counting

2,5 diphenyloxazole (PPO) and 1,4-di-2-(5 phenyloxazoyl) benzene (POPOP), were obtained from Fisons, Loughborough, England. Toluene was from BDH, Poole, England.

2.1.6. General Enzymes

RNase A and lysozyme were from Sigma Chemical Co., E.coli DNA polymerase I (Klenow fragment), proteinase K and T4 DNA ligase were from Boehringer Mannheim, West Germany. Bacterial alkaline phosphatase was from BRL, Cambridge, England. T4 polynucleotide kinase was from PL-Biochemicals, Milwaukee, Wisconsin, USA.

2.1.7. Restriction Enzymes

All restriction enzymes were from BRL, Cambridge, England, unless otherwise stated.

2.1.8. Nucleic Acids

Phage λC1857 DNA and ØX174 DNA were from BRL, Cambridge, England. Herring sperm DNA and polyadenylic acid were from Sigma Chemical Co. Oligo$(dT)_{12-18}$, 2'deoxyribonucleotides and 2',3'dideoxyribonucleotides were obtained from Boehringer Mannheim, West Germany.

2.1.9. Filters

Hardened ashless No.540 filters were purchased from Whatman Ltd., Maidstone, Kent. Nitrocellulose filters were from Schleicher and Schull, Dassel, West Germany.

2.1.10. Bacteriological Materials

Agar was from Difco Laboratories, Detroit, Michigan, USA. Tryptone and yeast extract were either from Difco Laboratories or Oxoid Ltd., Basingstoke, England. Thiamine HCl, tetracycline hydrochloride, ampicillin and chloramphenicol were from Sigma Chemical Co. 90mm petri-dishes and 96 well microtitre plates were from Sterilin Ltd., Teddington, England.

2.2.1. Bacteriological and Phage Strains

The host for all fibrinogen cDNA recombinants was E.coli K12 MC1061 of genotype: AraD 139, Δ(ara, leu) 7697, lacX74, gal U^-, gal K^-, hsr^-, hsm^+, StrA.
E.coli K12 JM103 of genotype: Δ(lac pro), thi, StrA, SupE, end A sbc B, hsd R^-, F′ traD36, proAB, $LacI^q$, Z M15) was the host for the M13 phage derivative M13mp9.

2.2.2. Human x CHO Hybrid Cell Lines

Human lymphoblastoid x CHO A23 somatic cell hybrids Gal 7 and Dis 20 were generously provided by Dr. Ben Carritt (Galton Laboratories, London). Hybrids Paw 4 and Paw 13 were constructed by Dr. C. Ingles of this department, from a human lypmhoblastoid cell line with a chromosome 4-12 translocation. All hybrid cell lines were maintained in this department by Dr. C. Ingles.

2.3. General Methods

2.3.1. Restriction Enzyme Digestion

DNA to be restricted was dissolved in distilled deionised water at a concentration not greater than 200 ug/ml. One tenth volume of 500 mM Tris-HCl (pH 8.0), 200 mM $MgCl_2$, 500 mM NaCl, 10 mM dithiothreitol and 1 mg/ml BSA was added to the DNA solution. For genomic DNA digests, spermidine was added to a final concentration of 4 mM. Enzymes were added to give a 2.5 fold excess for the allocated digestion time ie. 1 ug DNA digested with 2.5 units enzyme for one hour. All digestions were performed in eppendorf tubes at 37°C. After incubation, digests were heat shocked a 65°C for 5 min. and placed on ice. Small aliquots were electrophoresed on 1 or 2% agarose gels to determine the extent of digestion. Completed digests were either run directly or after deproteinisation with phenol, extraction of the aqueous phase with chloroform/isoamyl alcohol (24:1) and ethanol precipitation followed by resuspension in distilled water to the initial concentration.

2.3.2. Ethanol precipitation of Nucleic Acids

DNA was routinely precipitated by the addition of 1/10 volume 3M sodium acetate (pH 4.8) and 2.5 volumes 100% ethanol. Samples were chilled overnight at -20°C or 1 hr. at -70°C. Centrifugation was for 10 min either in a Sorvall HB4 rotor at 16000g or in an eppendorf microfuge (Eppendorf 5412). The pellet was washed twice in 70% ethanol and dried under vacuum. DNA was dissolved in an appropriate volume of sterile deionised water or sterile TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). For the further removal of impurities, DNA was reprecipitated by the addition of 1/2 volume 7.5M ammonium acetate and 2 volumes 100% ethanol. Samples were chilled, centrifuged and pellets resuspended as described above.

2.3.3. Agarose Gel Electrophoresis

Horizontal slab gel electrophoresis was performed to analyse DNA and RNA samples. For DNA gels, the appropriate amount of agarose was dissolved by boiling in E buffer (40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.5 with glacial acetic acid). For RNA gels, agarose was dissolved in MOPS buffer (20 mM sodium-MOPS, 5 mM sodium acetate, 1 mM EDTA) and formaldehyde added to 2.2 M after cooling to 65°C. Gels were poured at 60°C and allowed to set before removing the slot-former. Samples were loaded in 2% (w/v) Ficoll and 0.2% (w/v) Orange G dye marker. Electrophoresis was carried out at 30 volts until samples entered the gel matrix, and then at 70-100 volts until the Orange G marker reached the bottom of the gel.

2.3.4. Acrylamide Gel Electrophoresis

Preparative and analytical acrylamide: bis acrylamide (30:1) gels were formed in 160 mm x 195 mm x 1.5 mm mould and run in TBE buffer (90 mM Tris, 90 mM boric acid, 2.4 mM EDTA). For restriction fragment fractionation, 5% or 6% gels (with respect to acrylamide concentration) were used routinely.

2.3.5. Visualisation of Nucleic Acids

Following electrophoresis, DNA agarose gels were stained in E buffer containing 0.5 ug/ml ethidium bromide for 15 min. For RNA gels run under denaturing conditions, formaldehyde was removed by soaking the gel in two changes of 0.1M ammonium acetate for 1 hr. and staining with 0.5 ug/ml ethidium bromide in 0.1M ammonium acetate and 0.1M B-mercaptoethanol. Nucleic acids were visualised using short wave UV light and photographed on Ilford FP4 film or on polaroid type 52 Polapan film.

2.3.6. Gel Filtration

Gel filtration was used to remove unincorporated radiolabelled nucleotides from nick-translations, kinase labellings etc. 1 ml plastic syringes were plugged with polypropylene wool and topped with G-50 sephadex pre-equilibrated with 3xSSC. The syringe was placed in a disposable plastic centrifuge tube and spun in a Sorval HB4 rotor at 1500g for 3 min. The 'dry' sephadex column was layered with labelled DNA containing BEG dye (0.2% bromophenol blue, 0.1 M EDTA, 50% glycerol) as marker, and spun as above.

The void volume so collected in the tube was counted.

2.3.7. TCA Precipitation of DNA

Acid precipitation of kinase labelled DNA was performed as follows. An aliquot of the reaction mix, typically 1/10 reaction volume, was spotted on Whatman 540 filter and DNA immobilised by chilling the filter in an ice cold solution of 5% w/v TCA, 1% w/v tetra-sodium pyrophosphate. Unincorporated label was washed by whirlimixing followed by a second wash with 5% w/v TCA performed as above. Filters were finally washed in 100% ethanol and dried prior to counting.

2.3.8. Determination of Radioactivity

Measurement of P^{32} radioactivity was by Cerenkov counting and counting efficiency was approximately 30%. TCA precipitated P^{32} labelled DNA was immobilised on Whatman 540 filters, dried and counted in toluene based scintillant (TBS: 0.5% w/v PPO, 0.03% w/v POPOP in toluene).

2.3.9. Autoradiography

Agarose and acrylamide gels were dried under vacuum on 3mm Whatman paper, on a Bio-Rad slab gel dryer. Nitrocellulose filters from hybridisations were air dried. Exposure to X-ray film was at $-70^{o}C$ using an intensifying screen. Films were sensitised if required by pre-flashing to an OD_{650} of 0.1-0.3.

2.4. Specific Methods

2.4.1. Plasmid Vector and Bacterial Strain

Plasmid vector pKT218 is a derivative of PBR322 (Talmadge and Gilbert, 1980). Bacterial host for the above vector was E.coli K12 MC1061 (Casadavan and Cohen, 1980).

2.4.2a. Preparation of Plasmid DNA

Plasmid DNA was prepared by a modification of the cleared lysate method (Birnboim and Doly, 1979). Liquid cultures were grown in autoclaved L-broth (5 g/l yeast extract,19 g/l tryptone, 5 g/l NaCl, 1 g/l D+glucose), supplemented with tetracycline (10 ug/ml).

20 ml cultures were grown overnight at 37°C and diluted 1 in 100 into fresh L-broth (typically 500 mls). The inocula were further grown to OD_{650}=0.8 and chloramphenicol added to a final concentration of 0.17 ug/ml. Plasmids were amplified overnight at 37°C. Bacteria were harvested by centrifugation in a Sorvall GSA rotor for 10 min.at 8000g and 4°C. The pellets were drained well of any remaining L-broth and each pellet resuspended in a small volume of buffer 1 (50 mM glucose, 25mM Tris-HCl, pH 8.0; 10 mM EDTA and 5 mg/ml lysozyme). This suspension was incubated at room temperature for 10 min. followed by the addition of two volumes of buffer 2 (0.2 M NaOH, 1% SDS) and gentle shaking on ice for 5-10 min. Potassium acetate (pH 4.8) was added to 1M, the mixture vortexed thoroughly and kept on ice for 30 min. Cell debris and chromosomal DNA were removed by centrifugation in a Sorvall GSA rotor for 10

min. at 8000g. The supernatant containing supercoiled plasmid DNA was precipitated by the addition of 60% volume of propan-2-ol and incubation at room temperature for 5 min. Precipitated DNA was centrifuged at 16000g in Sorvall HB-4 rotor for 10 min. and the pellet dissolved in TE buffer. To this were added 1/10 volume 10 mg/ml ethidium bromide and solid caesium chloride to 1.56 g/ml. The solution was cleared by centrifugation in a Sorvall HB-4 rotor, as described above, and plasmid DNA centrifuged to equilibrium at 140,000g in the Sorvall vertical rotor TV-865B; for 16 hrs. at 15^{o}C. Supercoiled plasmid DNA band was visualised under UV illumination and collected by piercing the tube with a syringe needle. Ethidium bromide was removed by extraction with caesium chloride saturated isoamylalcohol and the DNA solution dialysed extensively against TE buffer. DNA was ethanol precipitated, resuspended in TE buffer and plasmid DNA concentration estimated by reading at OD_{260} (50 ug/ml DNA has OD_{260}=1).

2.4.2b. Plasmid Miniprep

Miniprep of plasmid DNA was essentially a scaled down version of plasmid DNA preparation described above in 2.4.2a. A 20 ml. overnight culture grown at 37^{o}C was centrifuged at 8000g and 4^{o}C for 10 min. Harvested cells were resuspended in a small volume (400 ul) of buffer 1 and incubated at room temperature for 5 min. Buffer 2 (800 ul) was added and the mixture shaken gently on ice for 10 min. Potassium acetate (pH 4.8) was added to 1M, and the mixture

vortexed thoroughly on ice. Cell debris and chromosomal DNA were removed by centrifugation at 8000g for 10 min. The supernatant was carefully removed and DNA precipitated by the addition of 2.5 volumes of 100% ethanol followed by incubation at room temperature for 5 min. DNA was pelleted by centrifugation at 16000g for 10 min. and resuspended in a small (150-200 ul) volume of sterile TE buffer. Excess RNA was digested by the addition of heat-shocked RNase A to 10 ug/ml and incubation on ice for 30 min. DNA was phenol extracted, ethanol precipitated and resuspended in 150 ul of sterile TE buffer. Plasmid DNA concentration was determined as described above.

2.4.3. Nick-translation of DNA

High specific activity labelling of double stranded DNA was performed by incorporating α-^{32}P-phosphate labelled deoxyribonucleotides into double stranded DNA by nick-translation (Rigby et al., 1977), using the Nick-Translation Kit (code N.500) supplied by Amersham. 100 ng of DNA was labelled for 2hrs at 15^{o}C in a reaction mixture containing Tris-HCl (pH 7.8), $MgCl_2$, 2-β-mercaptoethanol, 20 pg DNase I, 2.5 units DNA polymerase I, 20 uM each of dATP, dGTP and dTTP, and 25 uCi of α-^{32}P dCTP (>400 Ci/mmol). The reaction was terminated by the addition of EDTA to 0.2M. Unincorporated nucleotides were removed by gel filtration through G-50 sephadex pre-equilibrated with 3xSSC, as described above (2.3.8.).

2.4.4. T4 Polynucleotide Kinase Labelling of Oligonucleotides

Oligonucleotides were end-labelled at the 5' position with γ-^{32}P ATP (5000 Ci/mM) using T4 polynucleotide kinase (Maxam and Gilbert, 1980). 0.5-1.0 ug of oligonucleotides were labelled in a mixture containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, 25 uCi of γ-^{32}P ATP (5000 Ci/mM) and 2.5 units of T4 polynucleotide kinase. The mixture was incubated at 37^{o}C for 15 min. and the reaction stopped by the addition of 10 mM EDTA.

2.4.5. Southern Transfers

Fractionation of DNA for Southern transfers was performed on 0.8-1.% agarose gels with DNA size markers. Following electrophoresis the gel was denatured for 2 hrs in 0.5 M NaOH, 1.5M NaCl to render the DNA single stranded. The gel was neutralised in 0.5 M Tris-HCl (pH 5.5), 1.5M NaCl. (Renaturation of DNA strands does not occur during neutralisation). DNA was blotted in 20xSSC (SSC is 0.15 M NaCl, 0.015 M NaCitrate) onto nitrocellulose filter as described by Southern (1975). The filter was washed for 5 min in 3xSSC and DNA immobilised by baking at 80^{o}C for 2 hrs. Non-specific binding sites on the filter were saturated by pre-treatment prior to hybridisations. All treatments were at 65^{o}C for transfers probed with nick-translated DNA. Filters were washed in 3xSSC, 0.1% SDS for 30 min followed by prehybridisation in 3xSSC, 50 ug/ml

sonicated herring sperm DNA, 10xDenhardts (0.2% w/v ficoll, polyvinyl pyrolidone, bovine serum albumin - Denhardts, 1966), 50 mM sodium phosphate, pH 6.8, 10% dextran sulphate and 0.1% tetra-sodium pyrophosphate, at 65°C for 16 hrs. Solution of same composition was used for hybridisation. Nick-translated probes were denatured by immersing in boiling water for 5 min added directly to the hybridisation solution. For Southern transfers of human genomic DNA, hybridisation was for 12-16 hrs at 65°C. This duration was reduced if the filter bound DNA was in excess in the hybridisation reaction, eg. when blots of fibrinogen recombinant DNAs were hybridised. After hybridisation, filters were washed repeatedly (>5 times) in 3xSSC, 0.1% SDS. If appropriate a low salt (high stringency) wash was added to the protocol (2x30 min).

For transfers probed with end-labelled oligonucleotides, filters were washed in 3xSSC, 0.1% SDS; followed by prehybridisation in 6xSSC, 0.5% SDS 10xDenhardts, 100 ug/ml sonicated herring sperm DNA and 0.5% w/v tetra-sodium pyrophosphate, at 37°C for 60 min. Hybridisation was for 30 min at 37°C in 6xSSC, 10xDenhardt 50 ug/ml sonicated herring sperm DNA and 0.5% w/v tetra-sodium pyrophosphate. Filters were washed 3 times in 6xSSC and 0.05% w/v tetra-sodium pyrophosphate for 20 min at 37°C. If required higher stringency washes were performed at 42°C, 46°C and 50°C.

2.4.6. Preparation of Genomic DNA

(a) From blood (buffy coat)

Heparinised or EDTA treated blood was stored as 10 ml fractions at -70°C. Thawed fractions were kept at 4°C prior to DNA preparation. To each 10 ml blood fraction were added 90 ml lysis solution (0.32 M sucrose, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100) pre-cooled to 4°C. Nuclei of lysed cells were collected by centrifugation at 1600xg for 10 min at 4°C. The pellet was resuspended, with a pipette, in 4.5 mls of 0.075 M NaCl, 0.024 M EDTA (pH 8.0). To this resuspension were added 1/20 volume 10% SDS, 10 mg Proteinase K and distilled water bringing the volume to 5 mls. This mixture was incubated for 12-16 hrs at 37°C. Equal volume of phenol (saturated with 20 mM Tris-HCl, pH 8.0) was added and mixed gently, followed by centrifugation at 1600g for 5 min. The aqueous phase was removed, taking care to avoid phenol, and extracted with an equal volume of chloroform : isoamylalcohol (v/v 24:1), by gentle mixing and centrifugation as above. DNA was precipitated at room temperature from the aqueous phase by the addition of 1/10 volume 3M sodium acetate (pH 5.0), and 2.5 volumes of 100% ethanol. Precipitated DNA was dissolved in 1 ml TE buffer (10 mM Tris-HCl; pH 7.6, 1 mM EDTA) at 4°. DNA concentration was determined measuring $O.D._{260}$. Typically, 200-500 ug DNA were obtained from 10 mls of blood.

(b) From tissue culture cells

Culture medium was removed by aspiration and cells washed twice in PBSA (140 mM NaCl, 2.5 mM KCl, 4.2 mM Na_2HPO_4.

$12H_2O$, 1.5 mM KH_2PO_4). A final wash was given in versene (140 mM NaCl, 2.5 mM KCl, 3.2 mM Na_2HPO_4, 1 mM Na_2EDTA 0.2% w/v Phenol Red). Cells were suspended in STE buffer (150 mM NaCl, 10 mM EDTA, 10 mM Tris-HCl, pH 8.0) supplemented with 0.5% SDS and Proteinase K to a final concentration of 0.25 mg/ml. The suspension was incubated at 37°C overnight (12-16 hrs), phenol extracted twice, chloroform/isoamyl-alcohol extracted once, and DNA precipitated by the addition of 2 volumes 100% ethanol to the aqueous phase. Precipitated DNA was given a 70% ethanol wash to remove excess salt and resuspended in TE buffer. A second precipitation was performed by the addition of 1/20 volume 4 M NaCl and 2.5 volume 100% ethanol at -20°C . The DNA pellet was redissolved in TE and stored at 4°C. Concentration determination was by measuring $O.D._{260}$.

2.4.7. RNA preparation

Glassware used for RNA preparation was washed thoroughly, rinsed with distilled water containing 0.1% diethylpyrocarbonate (DEPC) and autoclaved. Adult human liver RNA was a gift of Dr. S. Wallis. A postmortem specimen of adult human liver was taken 30 minutes after death, frozen immediately in liquid nitrogen and stored at -70°C. Total RNA was prepared by the LiCl-Urea method (Auffray and Rougeon, 1980). Frozen liver tissue was homogenized in 8 volumes of 8 M Urea, 4 M LiCl and left for 24-48 hrs at 4°C. RNA was pelleted by centrifugation for 20 minutes a 1600g, resuspended in 8 M urea and extracted with urea saturated phenol followed by

extraction with phenol saturated in TE buffer. The aqueous phase was chloroform/isoamylalcohol extracted and ethanol precipitated. RNA was resuspended in sterile distilled water and stored at -70°C.

2.4.8. Northern Blots

RNA for Northern transfers to nitrocellulose was run on a 1% agarose denaturing gel in MOPS buffer (20 mM Na-MOPS, 5 mM sodium acetate, 1 mM EDTA). RNA samples were denatured by boiling in 2.2 M formaldehyde and 50% deionised formamide. 5-10 ugs of RNA were run per track. To visualise the RNA, the gel was washed for 2x30 min in 0.1 M ammonium acetate, stained with 0.5 ug/ml ethidium bromide in 0.1 M ammonium acetate for 60 min, destained in 0.1 M ammonium acetate for 45 mins and irradiated on a short wave UV transilluminator.

Transfer of RNA to nitrocellulose, RNA immobilisation on the filter, prehybridisation and hybridisation with nick-translated fibrinogen cDNA clones, were as described for Southern transfers (2.4.5.). Post-hybridisation washes were in 2xSSC, 0.1% SDS for 2x30 min at 65°C.

2.4.9. Synthesis of Oligonucleotides

Septadecamer oligonucleotides specific for the Aα-, Bβ- and γ- chains of human fibrinogen were synthesised by Dr. M.A.W. Eaton (Celltech Ltd., Slough), essentially as described by Smith et al., (1982).

2.4.10. Screening of Adult Human Liver cDNA Library

An adult human liver cDNA library was generously provided by Dr. D.E. Woods and is described elsewhere (Woods et al., 1982). The library was plated on 82 mm diameter nitrocellulose filters, and grown overnight at 37°C on L-agar plates (15 g/l agar in L broth) supplemented with 10 ug/ml tetracycline, until colonies were 0.5 mm in diameter. Filters with appropriate colony density (not greater than 1000 colonies/filter) were replica plated onto 82 mm nitrocellulose disc filters. Two sterile Whatman filters were laid on a sterilised glass plate and the master filter laid face up, on top of these. 3 or 4 orientation marks were made on the master filter with a marker pen in order to identify any bacterial colonies required for further study. The replica filter was prewetted on a fresh L-agar plate and placed dry side down on top of the master filter. Two more Whatman filters were laid on top of this sandwich and pressed down firmly with a sterile glass plate for even distribution of pressure. The replica was marked to align with the master. The two filters were peeled apart and the replica grown, colonies up, as usual. A second replica was taken from the master, which was finally regrown for 60 min before more replicas were made.

Replica filters were incubated at 37°C for about 4 hrs until colonies were 0.5 mm in diameter. Filters were then transferred to L-agar plates containing 170 ug/ml chloramphenicol and plasmids amplified overnight at 37°C.

Filters were screened using a modification of the procedure of Hanahan and Meselson (1980). Colonies were lysed by placing each nitrocellulose filter onto large 3 mm Whatman filter sheets saturated with 0.5 M NaOH, for 15 min. The alkali was neutralised by transferring the nitrocellulose disc to a Whatman sheet saturated with 1M Tris (pH 7.6), for 5 min. DNA was immobilised by 5 min on 1.5M NaCl, 0.5 M Tris-HCl (pH 7.6), followed by 5 min on 0.3 M NaCl and bakingfor 2 hrs. at 80^{o}C. Non-specific binding sites were saturated by pre-treatment of the filters prior to hybridisation with end- labelled oligonucleotides. Filters were pre-washed overnight in 3xSSC, 0.1% SDS at 65^{o}C. Any remaining colony proteins were removed by rubbing the filter surface with a gloved finger and washing filters for a further 30 min at 65^{o}C in the above solution. Filters were prehybridised with 6xSSC, 5xDenhardts, 0.1% w/v SDS, 50 ug/ml sonicated herring sperm DNA and 0.01% w/v tetra-sodium pyrophosphate, for 2 hrs at 37^{o}C. Hybridisation with 0.5 ug end-labelled probe was for 20 min at 37^{o}C, in a mixture of same composition. Post-hybridisation washes were in 6xSSC, 0.1% tetra-sodium pyrophosphate for 2x30 min at 37^{o}C or higher temperature.

2.5. DNA Sequence Analysis

DNA sequence analysis was performed by subcloning PstI excised cDNA inserts of putative fibrinogen recombinants into the PstI site of M13mp9 polylinker, and sequencing the subclones by the dideoxy chain termination method

(Sanger et al., 1977). The principles of M13 subcloning and sequencing are shown in figure 5a.

(a) Subcloning into M13mp9

Double stranded (replicative form) of phage M13mp9 DNA (5 ug) was digested with PstI. An aliquot was checked on a 1% agarose gel for completion of digestion (figure 5b) and the remainder diluted in sterile deionised water to a concentration of 10 ng/ul. This preparation was then used subsequently for all subclonings.

Putative fibrinogen cDNA recombinants (5 ug) were similarly digested with PstI, checked on a 1% agarose gel (figure 12a) and diluted to 30 ng/ul in distilled deionised water. Ligations were performed in a small volume (10 ul) for 4 hrs. at room temperature by incubating 30 ng of PstI cleaved cDNA recombinants with 10 ng of M13mp9 DNA digested with PstI, in a mixture containing 10 mM dATP, 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 50 mM NaCl and 10 units of T4 DNA ligase. The reaction was stopped by heat shocking to 65^{o}C for 10 min. Ligated DNA was chilled on ice and used to transfect competent E.coli K12 JM103 cells (Messing et al.,1981).

(b) Preparation and Transfection of Competent Cells

A single colony from a streak of E.coli K12 JM103 on a minimal agar plate was used to inoculate 15 mls of 2xYT (YT medium is 8g/l tryptone, 5g/l yeast extract, 5g/l NaCl), and grown overnight at 37^{o}C. 10 mls of fresh 2xYT medium were inoculated with 0.1 ml of the overnight culture and grown

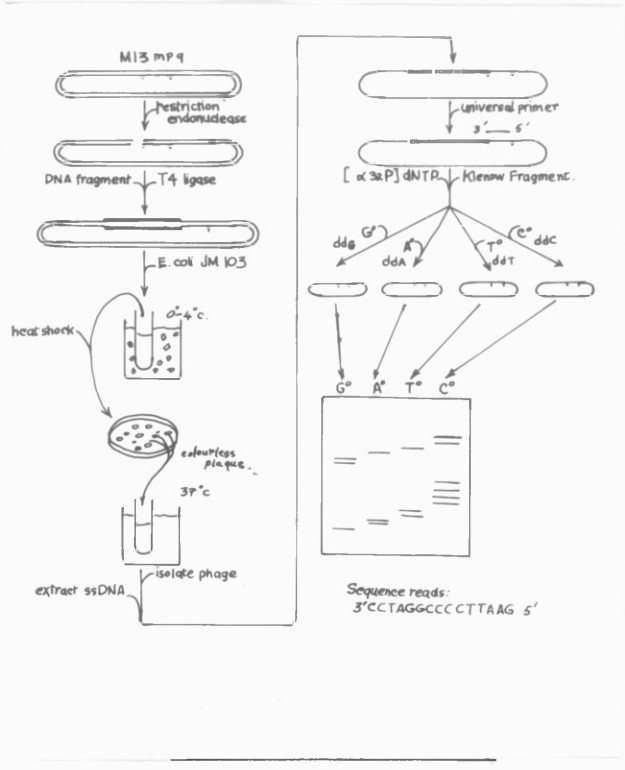

Figure 5a. An outline of the principles of M13 subcloning and sequencing.

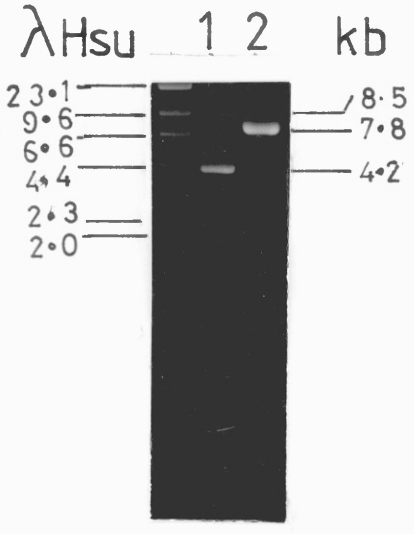

Figure 5b. A 1% agarose gel of 0.5 ug M13mp9 supercoiled DNA (lane 1) and DNA digested with PstI (lane 2).

to an OD_{660}=0.3-0.4. Cells were harvested by centrifugation at 7000g for 5 min and resuspended in 5 mls of ice cold 50 mM $CaCl_2$. The suspension was kept on ice for 20 min and centrifuged as before. Competent cells were resuspended gently in 2.5 mls of 0.5 M $CaCl_2$ on ice, aliquoted into 200 ul fractions and either used immediately for transfections or stored overnight at 4°C without detectable loss of transfection efficiency. Ligated DNA was added to 200 ul of competent cells and kept on ice for 30 min. The mixture was heat shocked for 2 min in a water bath pre-heated to 42°C. Transfected cells were plated on YT agar (15 g/l agar in YT medium).

(c) Plating of Transfected Cells

Transfected E.coli K12 JM103 cells were added to 3 mls of soft agar (6 g/l agar in YT medium), at 45°C containing 2.6 ug IPTG, 16.6 ug X-Gal and 20 ul of freshly grown E.coli K12 JM103 (OD_{600}=0.4-0.6). The mixture was swirled gently, poured on to prewarmed YT agar plates and left to set on a smooth flat surface. Plates were incubated overnight at 37°C to allow phage growth.

(d) Preparation of Single Stranded Phage DNA

White plaques (recombinants) were transferred with sterile truncated 'yellow tips' (Sarstedt No. 70.760) into 1 ml YT medium in a 20 ml 'Universal Container' (Sterilin), and grown for 8 hrs at 37°C. To increase DNA yield, 10 ul of a fresh exponentially growing culture of E.coli K12 JM103 was added to each inoculum prior to the above incubation. 1ml cultures were transferred to 1.5 ml

eppendorf tubes and microfuged for 10 min. The supernatant was carefully removed and phage precipitated by the addition of 1/5 volume 20% PEG-6000, 2.5 M NaCl solution and incubated for 15 min at room temperature. Phage was collected by microfuging for 5 min. The supernatant was decanted and the inside walls of the tube wiped with Kleenex tissue, carefully avoiding the pellet. 100 ul of TES buffer (20 mM Tris-HCl; pH 7.4, 10 mM NaCl, 0.1 mM Na_2 EDTA) was added and the pellet resuspended by vortexing for 2 seconds. Resuspended virus particles were extracted with phenol by vortexing for 2 seconds, leaving to stand for 5 min and vortexing again as before. The aqueous phase was extracted with chloroform/isoamylalcohol (24:1) and DNA precipitated by the addition of 1/30 volume of 3M sodium acetate (pH 4.8), 2.5 volumes of 100% ethanol and chilling at -70°. DNA was pelleted by microfuging for 10 min, resuspended in 50 ul of TES buffer and stored at -20°C. Aliquots of M13 recombinant DNAs were run on a 1% agarose gel.

(e) Single-Lane Screening

Recombinants were further characterised by single lane screening, using single stranded M13mp9 vector DNA as control. 250-500 ug viral DNA was annealed to 1.25 ug of M13 universal primer in Hin buffer (7 mM Tris-HCl; pH 7.5, 7 mM $MgCl_2$ and 7 mM NaCl), by heating at 85-90°C for 2 min. in a water bath, and slow cooling to room temperature. This was achieved by heating the DNA and M13 primer in eppendorf tubes immersed in test tubes filled with at least 5 mls of water at 85-90°C. Following the incubation the test

tubes were placed on the benchtop for about 45 min. allowing slow equilibriation to room temperature.

To each eppendorf tube, containing annealed DNA, 1.0 uCi α^{32}P dCTP, 1.5 uM dTTP, 30 uM dATP, 30 uM dGTP, 0.35 mM ddT and 0.05 units DNA pol I (Klenow fragment) were added. The mixture was incubated at room temperature for 15 minutes, following which unlabelled dCTP was added to 70 mM, and the chase performed at room temperature for 15 min. The reaction was stopped by the addition of an equal volume of formamide-dye mix (95% v/v deionised formamide, 10 mM Na_2EDTA, 0.1% w/v xylene cyanol, 0.1% w/v bromophenol blue). Samples were boiled for 5 minutes and 1/4 volume (3 ul) run on a 6% acrylamide 8M urea denaturing gel at 30 mA, until bromophenol blue reached the bottom of the gel. Gels were dried under vacuum and exposed overnight at room temperature to Fuji Rx X-ray film.

(f) Sequencing of Recombinants

Recombinant single stranded M13 phage DNAs were used as templates for sequencing reactions primed with the M13 universal primer.

2.5 ug of M13 universal primer was annealed to 0.5 ug of template DNA, by heating for 5 min. at 85-90^{o}C, and slow cooling to room temperature, as described above. 1.0 ug of annealed DNA was transferred to reaction tubes containing 1 uCi α-^{32}P dCTP, 4 mM DTT, 0.2 unit DNA polymerase I (Klenow fragment) and the appropriate mixture of deoxy and dideoxy nucleotides; i.e. reactions specific for adenine had 200 uM ddA, 30 uM dGTP, 30 uMdTTP and 1.6 uM

dATP; and those specific for guanine had 140 uM ddG, 30uM dATP, 30 uM dTTP and 1.6 uM dGTP. Similarly, reactions specific for thymine had 400 uM ddT, 30 uM dATP, 30 uM dGTP and 1.6 uM dTTp; while those specific for cytosine had 70 uM ddC and 25 uM each of dATP, dGTP and dTTP.

Reactions were performed at room temperature for 15 min following which 1/5 volume 0.5 mM dCTP was added and the chase carried out as before. The chase served to extend chains prematurely terminated due to the initial limiting dCTP concentration. The reaction was stopped by the addition of equal volume of formamide-dye mix (see 2.4.9.e above). Samples were denatured in a boiling water bath and 1/5 volume run on a 6% acrylamide 8M urea denaturing gel at 30 mA, until either bromophenol blue or xylene cyanol reached the bottom. Gels were dried and exposed as described earlier.

(g) Computer Search of Protein Sequences

Identification of cDNA insert sequences was performed by searching an amino acid sequence protein data file, using a programme written by Mr. Paul North of this department.

(h) Dot Matrix Analysis

Nucleotide sequence comparisions were performed by a dot matrix analysis programme also written by Mr. P. North.

3. Results

3.1. The Isolation of cDNA Clones for Human Aα-, Bβ- and γ-fibrinogen

The strategy for isolating cDNA clones specific for the three chains of human fibrinogen was based on the known amino acid sequence for each chain (Blomback et al., 1976, Lottspeich and Henschen, 1977, Doolittle et al., 1979 and Watt et al., 1979). This enabled the synthesis of oligonucleotides to regions of the polypeptides giving complementary DNA sequences of low ambiguity. Labelled oligonucleotides were used for the direct screening of a representative cDNA clone bank of human liver tissue to isolate recombinants showing specific hybridisation to these probes.

3.1.1. Synthetic Oligonucleotides Specific for the Aα-, Bβ- and γ-chains of Human Fibrinogen

(a) SEQUENCES OF OLIGONUCLEOTIDES.

The amino acid sequence of each chain of human fibrinogen was searched for stretches of six residues that would give septadecamer oligonucleotides of low complexity and redundancy. Figure 6 shows regions of the protein sequences chosen and their corresponding cDNAs. Oligonucleotides for the Bβ- and γ-chains were each mixtures of 4 sequences. Two oligonucleotide mixtures were synthesised for the Aα-chain since the sequence of this

protein did not give low mix oligomers. Aα-oligomer$_1$ (a.a. 582-587) was a mixture of 32 and Aα-oligomer$_2$ (a.a. 472-477) a mixture of 64 different sequences.

(b) SIZE AND PURITY OF OLIGONUCLETIDES.

All four oligomers were end labelled with γ-^{32}P ATP (5,000 Ci/mM) to a specific activity of 1x10^7 cpm/ug DNA, using T4 polynucleotide kinase. Oligo(dT)$_{12-18}$ was similarly labelled to a specific activity of 6x10^7 cpm/ug. Labelled oligomers were run on a 20% acrylamide/8M urea sequencing gel (figure 7) and sized relative to the oligo(dT)$_{12-18}$ ladder (lane C). Aα-oligomer$_1$ (lane A) migrates primarily as a 13mer with minor bands going up to the 18 base position. On the other hand Aα-oligomer$_2$ (lane B) has a size range of 13 bases to 19 bases in length. Size ranges of Bβ-oligomer (lane D) and γ-oligomer (lane E), are from 13 bases in length to 17 bases.

The presence of more than a single band 17 bases long in each mixture was due to a broad 'size cut' taken during purification (M.A.W. Eaton, personal communication). None of the bands in the oligomers co-migrated exactly with the oligo(dT) bands of corresponding length. This variation in migration of the oligomer bands can be attributed to DNA sequence differences (A. Markham, personal communication). The apparently short size of Aα-oligomer$_1$ however, might be the result of prematurely terminated synthesis.

3.1.2. Screening of an Adult Human Liver cDNA Library

(a) PRIMARY SCREENING

An adult human liver cDNA library (Woods et al., 1982) was plated on a nitrocellulose filter 82 mm in diameter, to a density of approximately 1000 colonies/filter. Replicas of this filter were screened with oligomers specific for each chain of human fibrinogen using a modification of the procedure of Hanahan and Meselson, (1980). Positive control for specific hybridisation was provided by a unique oligomer for human α1- antitrypsin. This septadecamer oligomer was synthesised to the 3' untranslated region of the mRNA using the published cDNA sequence of human α1-antitrypsin (Kurachi et al., 1981). The sequence of this probe is shown below:

α1- antitrypsin cDNA: 5' CCTGGATGACATTAAAG 3'

All five oligonucleotides were labelled with γ-^{32}P ATP to a specific activity of 10^7cpm/ug and hybridised separately to the filter replicas in duplicate. At low stringency post-hybridisation washes (6xSSC, 37^{o}C), the filters showed background of varying intensity (figure 8a). Background was particularly high on filters A_1 and A_2, which were hybridised with Aα-oligomer$_1$ and Aα-oligomer$_2$ respectively, so that no clear signal was detectable over the background. Higher stringency washes at 46^{o}C for 20 min and 50^{o}C for 2 min, were performed in 6xSSC. Figure 8(b) shows the decrease in background on filters A_1 and A_2; the latter showing a single positive. The absence

of a signal on filter A_1 is possibly due to the short size of Aα-oligomer$_1$. Any duplex formed between this oligomer and its complementary DNA sequence would be unstable and melted during high stringency washes.

(b) SECONDARY SCREENING

All duplicating positives showing strong hybridisation were picked into microtitre wells and grown alongside clones that did not show specific hybridisation. As an additional negative control, plasmid vector pKT218 was grown alongside each panel of positively and negatively hybridising clones. Panels were plated in duplicate ordered arrays on nitrocellulose filters and hybridised as above with the appropriate end-labelled oligonucleotide.

Figure 9 shows the filter hybridisation result of ordered arrays after high stringency washes at 46^{o}C for 20 min and 50^{o}C for 2 min. Panel (A) was hybridised with Aα-oligomer$_2$ and the single positive which hybridised to this probe in the primary screening also gave a strong signal here (lane 1). The probe also showed strong hybridisation to two recombinants (lane 3) which were negative during primary screening. However, it also hybridised to the plasmid vector pKT218, which suggested that the observed result in lane 3 was an artefact.

Hybridisation of Bβ-oligomer is shown in panel (B). There are four very strongly hybridising clones (lane 2) and two weakly hybridising clones (lane 1 and bottom of lane 2). There is no detectable hybridisation to plasmid vector pKT218. Panel (C) was hybridised with γ-oligomer.

Of the three positives picked from the primary screening, only one showed strong hybridisation to the probe (lane 1). There was no detectable hybridisation to pKT218 (lane 2).

3.1.3. Plasmid Preparation and Oligonucleotide Hybridisation to Southern Blots of PstI Digested Recombinants

Positives from the secondary screen were picked and grown in 25ml cultures under tetracycline selection. Plasmid DNA minipreps (section 2.4.2.) were performed using a modification of the cleared lysate procedure (Birnboim and Doly, 1979).

Of the six putative Bβ-chain recombinants, only the strongly hybridising clones designated pBF1, pBF2, pBF4 and pBF5 were grown and plasmid DNA prepared. Similarly, the putative Aα-chain recombinant (pAF1) and the putative γ-chain recombinant (pGF1) were used for plasmid DNA preparation. All clones on digestion with PstI released an insert (figure 10 panels A and B). This demonstrated that the PstI site had been reconstructed from the annealing of polyguanosine tails of PstI cleaved vector pKT218 and the polycytosine tails of double stranded cDNA, by DNA repair activity of the host (Otsuka,1981). Southern blots of PstI digested recombinants (figure 10) were hybridised with the appropriate end-labelled oligomer and showed each insert to hybridise with its specific oligomer (figure 11). Panel (A) shows γ-oligomer hybridisation to PstI excised insert of pGF1 (lane c). Duplicate tracks of PstI digested pAF1 (lane

a and lane b), were respectively hybridised with Aα-oligomer$_2$ and Aα-oligomer$_1$. The former shows hybridisation to the excised insert while no signal is given by the latter, thus confirming the earlier negative result obtained during primary screening of the liver cDNA library (figure 8b, filter A_1).

Inserts of PstI digested recombinants pBF1, pBF2, pBF4 and pBF5, each hybridised with the Bβ-oligomer (figure 11, panel B). Recombinants pBF1 (lane a) and pBF5 (lane d), representing the two size classes were used for further analysis.

Large scale preparation of supercoiled plasmid DNA from recombinants pAF1, pBF1, pBF5 and pGF1 was as described in in section 2.4.2. Purity of supercoiled DNAs was achieved by centrifugaion to density equilibrium on a 1.56 g/ml CsCl gradient. Typically, 600-800 ug of supercoiled DNA was obtained from a 500 ml L-broth culture. Similar yeilds were observed for the parental plasmid pKT218. Supercoiled plasmid DNAs were digested with PstI and run on a 1% agarose gel (figure 12a). Recombinants pAF1 (lane 1) and pGF1 (lane 2) released an additional small PstI insert fragment (figure 12b) demonstrating the presence of an internal PstI site. Insert sizes are shown in Table 1.

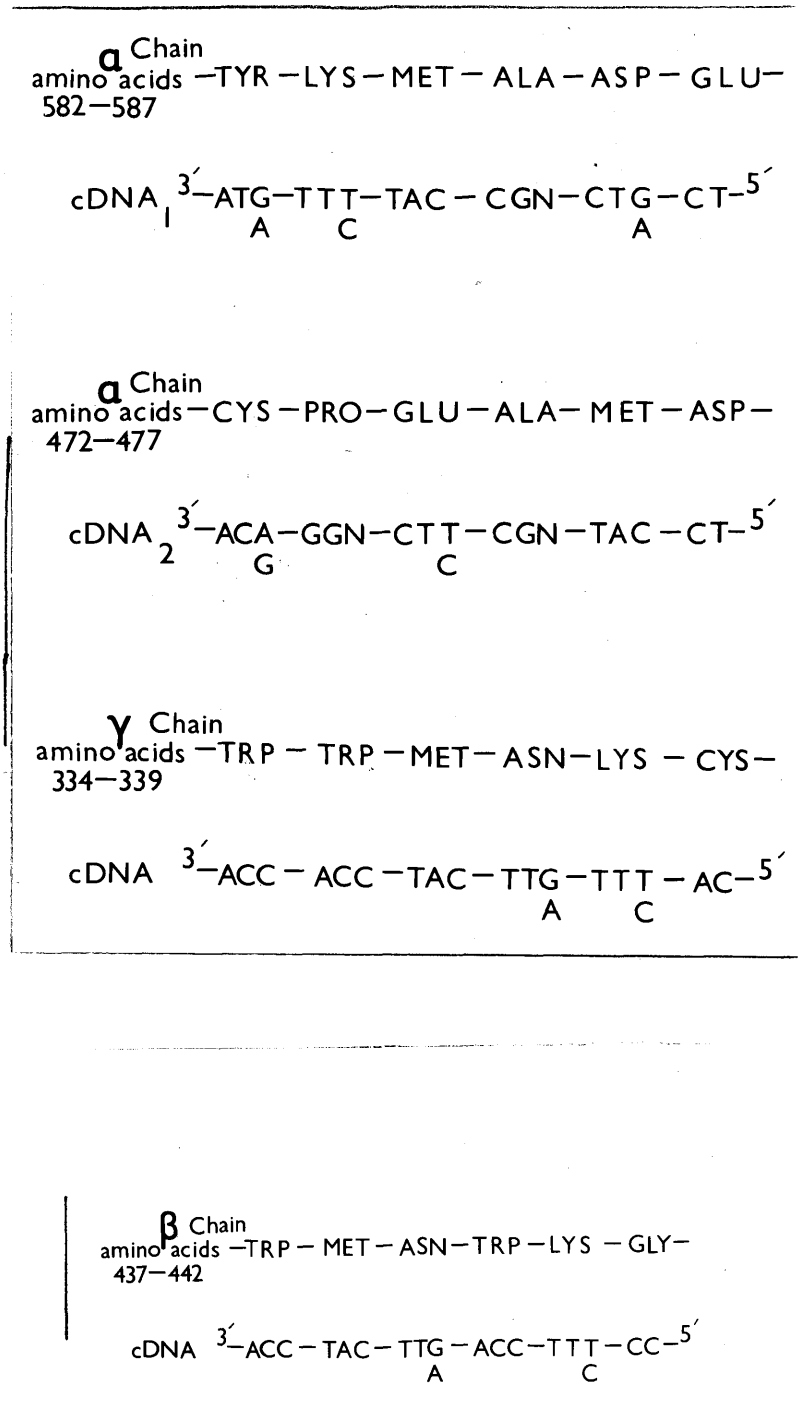

Figure 6. Sequences of synthetic oligonucleotides specific for the Aα-, Bβ- and γ-chains of human fibrinogen used to screen an adult human liver cDNA library.

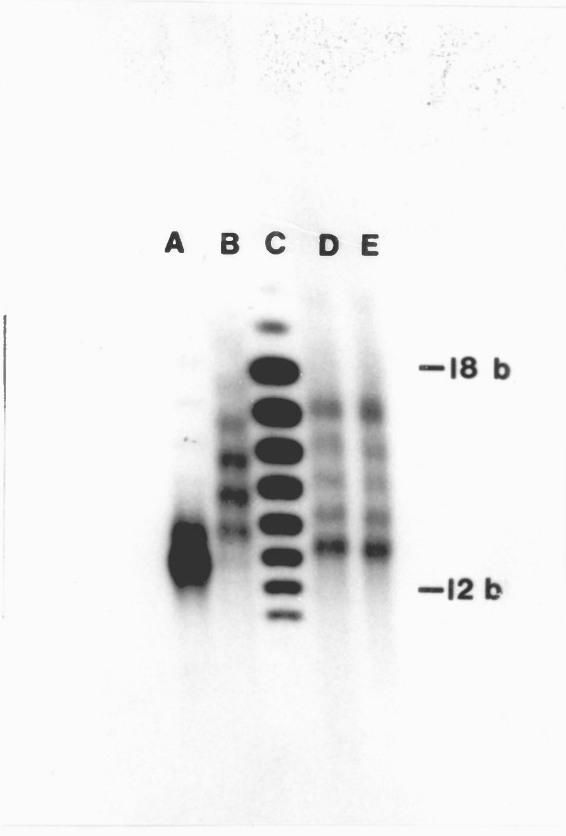

Figure 7. Sizing of synthetic oligomers. End-labelled oligomers (0.1ug) were run on a 20% acrylamide sequencing gel. Lane A is Aα-oligomer 1; lane B is Aα-oligomer 2; lane C is oligo(dT)$_{12-18}$; lane D is Bβ-oligomer and lane E is γ-oligomer. Exposure to X-ray film was for 2 mins. at room temperature.

(a)

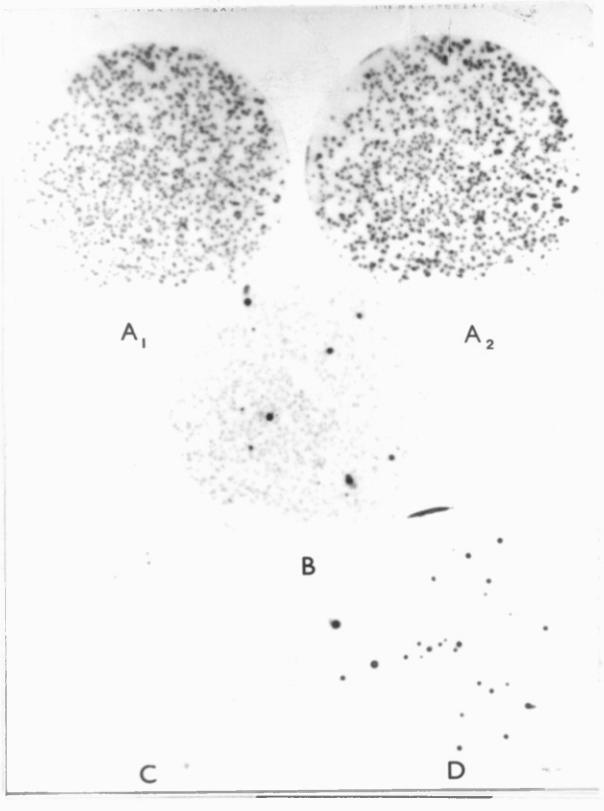

Figure 8. Primary screening of 1000 clones from an adult human liver cDNA library. All filters are replicas of a single master. Filters A_1 and A_2 were respectively hybridised with Aα-oligomer 1 and oligomer 2; filter B was hybridised with Bβ-oligomer; filter C hybridised with γ-oligomer and filter D with α1-antitrypsin oligomer.

(a) Low stringency washes were at 37^{o}C for 20 min.

(b)

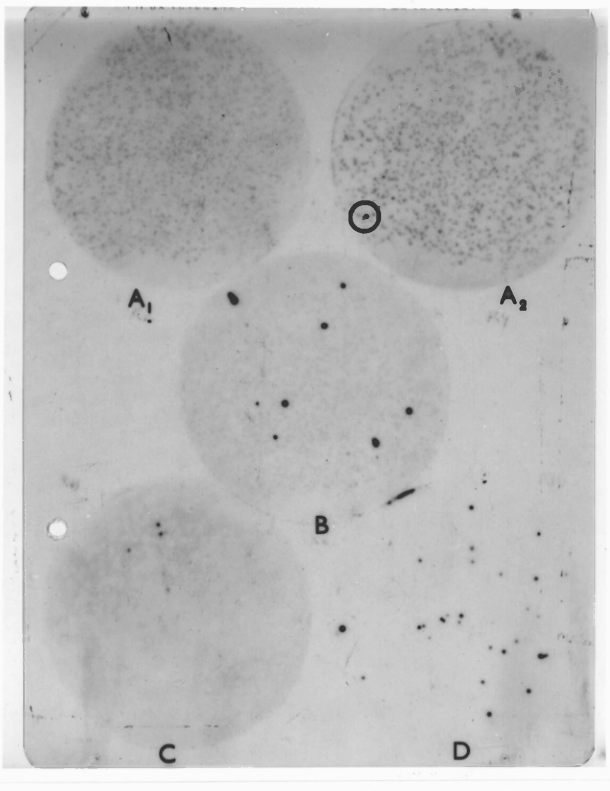

Figure 8.

(b) High stringency washes were at 46°C for 20 min. and 50°C for 2 min.
The single positive on filter A_2 is circled.

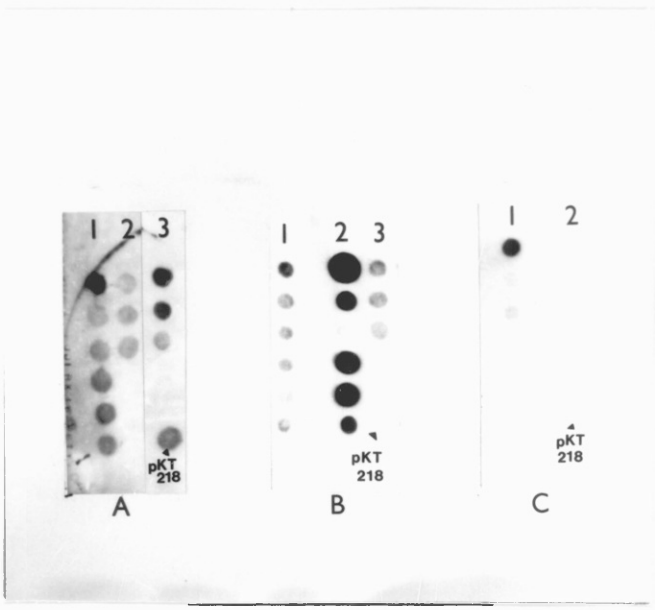

Figure 9. Secondary screening of putative fibrinogen cDNA recombinants shown in figure 8. Hybridisation of panel (A) was with Aα-oligomer 2, panel (B) with Bβ-oligomer and panel (C) with γ-oligomer. High stringency post-hybridisation washes were at 50°C for 2 min.

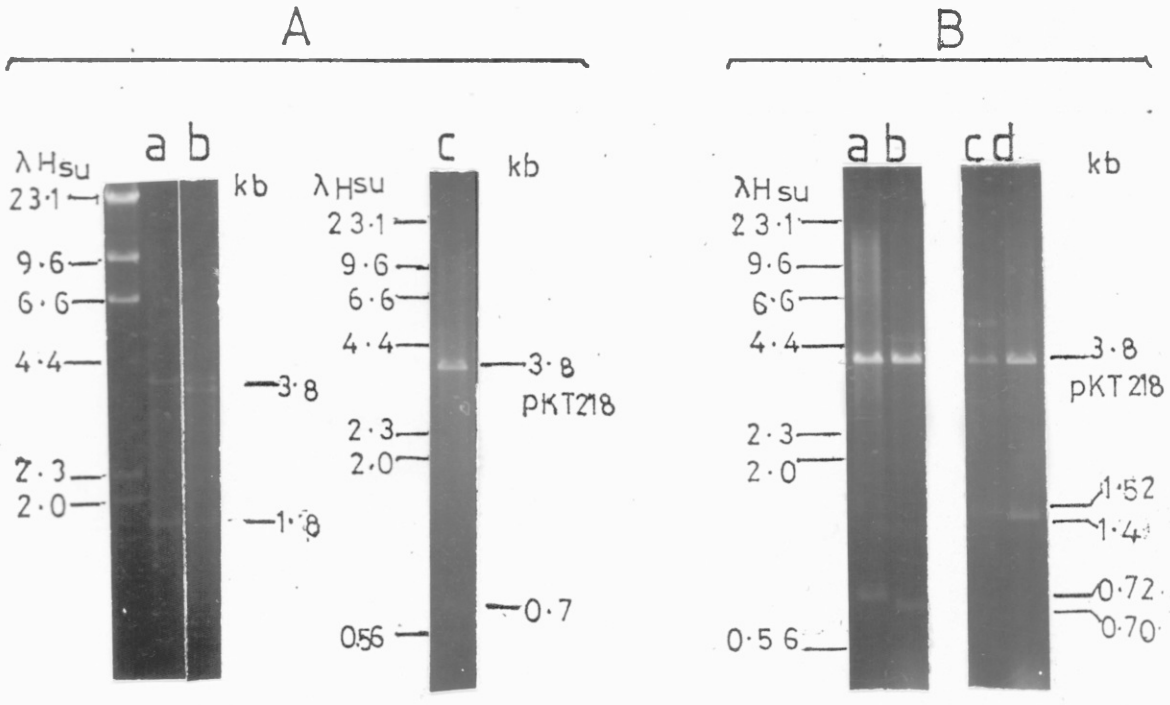

Figure 10. 1% agarose gels with PstI digests of minipreps.
Panel A. Lanes (a) and (b) have PstI digested pAF1. Lane (c) has PstI digested pGF1.
Panel B. Lane (a) has pBF1, lane (b) has pBF2, lane (c) has pBF4 and lane (d) has pBF5.

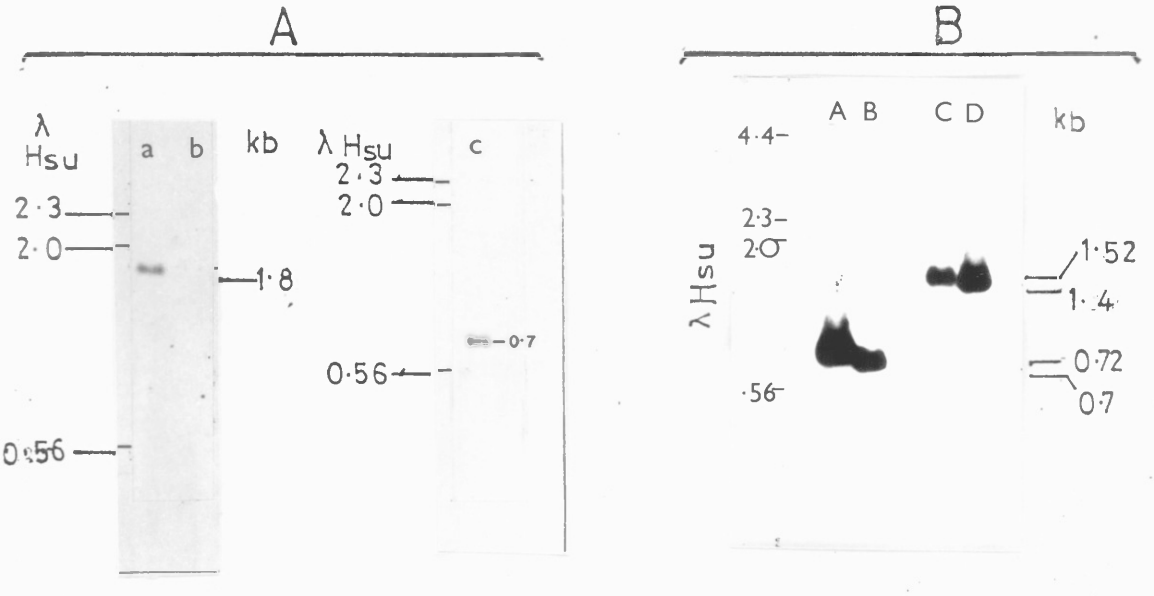

Figure 11. Southern blot hybridisation of PstI restricted recombinant minipreps shown in Figure 10.
Panel A. Lane (a) was hybridised with A α-oligomer 2 and lane (b) with A α-oligomer 1. Lane (c) was hybridised with γ-oligomer.
Panel B. Lanes (a)-(d) were hybridised with B β-oligomer.

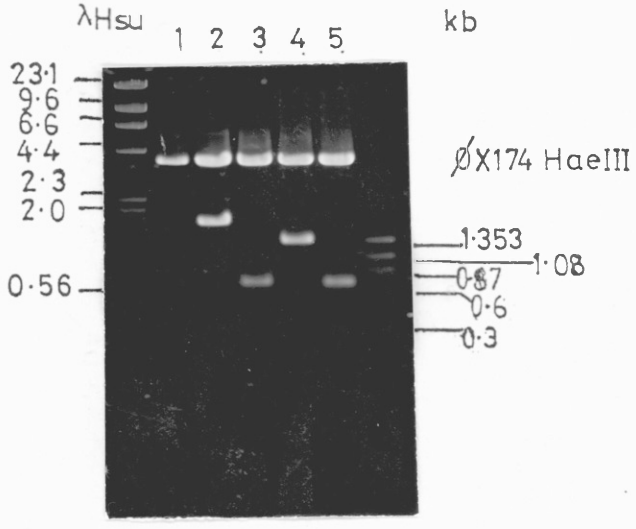

Figure 12a. Fractionation of PstI digested plasmid preps. 1 ug plasmid DNAs pKT218 (lane 1), pAF1 (lane 2), pBF1 (lane 3), pBF5 (lane 4) and pGF1 (lane 5) were run on a 1% agarose gel. Insert fragment sizes were determined relative to λHsu and ØX174 HaeIII markers.

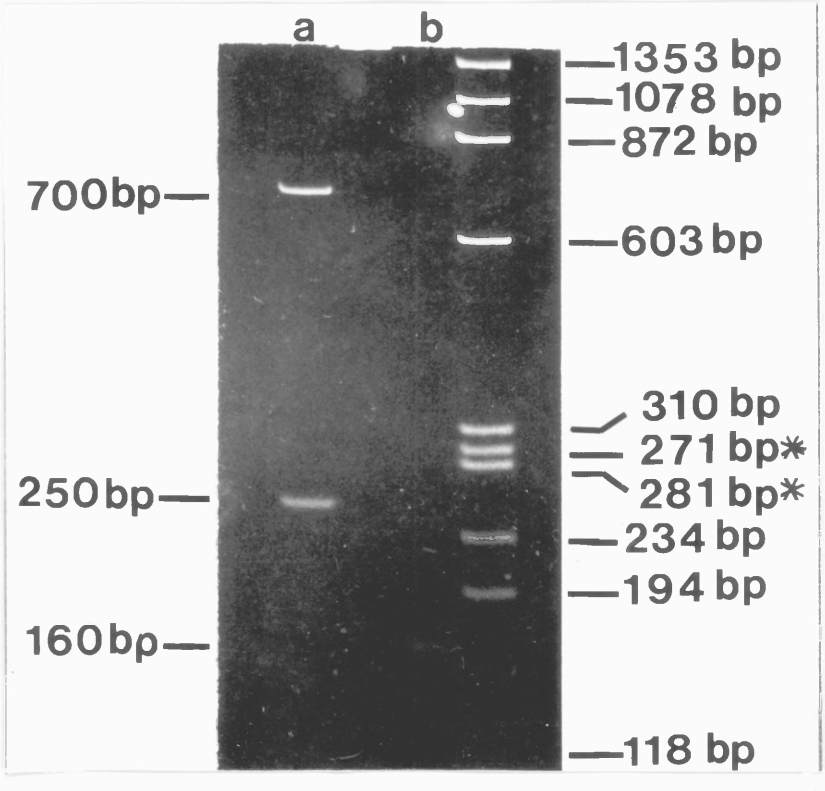

Figure 12b. A 5% acrylamide gel of PstI digested clones (5 ug/lane) showing small PstI excisable fragments of pAF1 insert (lane a) and pGF1 insert (lane b). ØX174 HaeIII DNA was used as markers.
* Anomalous mobility.

Table 1

Recombinant	PstI excisable insert fragments (kb)	Total Insert Length (kb)
pAF1	1.79 0.16	1.95
pBF1	0.72	0.72
pBF5	1.4	1.4
pGF1	0.70 0.25	0.95

Table 1.

The total insert lengths of the recombinants pAF1, pBF1, pBF5 and pGF1 as determined from their PstI excisable insert fragments sizes.

3.2. Characterisation of the Putative cDNA Clones for Human Fibrinogen

The cDNA clones were characterised by sequencing and comparing the corresponding amino acid sequence with the known protein sequence for each chain of human fibrinogen.

cDNA recombinants were subcloned into phage M13mp9 replicative form (double stranded) DNA and the resulting recombinant single stranded phage DNAs sequenced by the dideoxy chain termination method (Sanger et al., 1979). Prior to subcloning and sequencing, cross-hybridising clones were identified and restriction maps constructed for unique recombinants.

3.2.1. Cross-Hybridisation of cDNA Recombinants

Clones pBF1 and pBF5 were isolated by hybridisation under stringent conditions with the same probe (Bβ-oligomer) which suggested that they may homologous. To confirm this, clone pBF1 was nick-translated to a specific activity of 10^7cpm/ug and 10^6 counts hybridised to a Southern blot of PstI digested recombinants pAF1, pBF1, pBF5 and pGF1 (figure 13). Filters were washed extensively at high stringency (0.1xSSC, 65^{o}C). The probe cross-hybridised only with clone pBF5 (lane D), demonstrating that this is the only recombinant with which it has extensive homology.

A further cross-hybridisation study was performed. Southern blots of PstI digested recombinants pAF1, pBF5 and pGF1 were run in triplicate (figure 14a) and hybridised separately with nick-translated pAF1, pBF5 and pGF1

labelled to a specific activity of $6x10^7$cpm/ug. Filters were washed at low stringency (3xSSC, 65°C) to optimise the detection of homologous sequences. No cross-hybridising sequences were detected for the three clones (figure 14b).

3.2.2. Restriction Mapping and the Detection of Poly(A)

A preliminary restriction map was determined for recombinants pAF1, pBF5 and pGF1, using the enzymes PstI, EcoRI, BamHI and HindIII. The plasmid vector has a unique site for each of these enzymes thus facilitating mapping of the insert.

(A) MAPPING OF pAF1

Digestion of pAF1 with EcoRI excised a 775 bp fragment (figure 15, lane 3), demonstrating the presence of an internal EcoRI site. Since the distance between the EcoRI and PstI sites of pKT218 is 235 bp, the internal EcoRI site was placed 540 bp from the rightward PstI cloning site (figure 17a). This was confirmed by secondary digestion of EcoRI digested pAF1 with PstI (lane 1). The 775 bp band was cleaved into a 540 bp insert fragment and a 235 bp band representing the EcoRI—PstI fragment of pKT218, as deduced from the EcoRI+PstI double digest of this DNA (lane 2).

Secondary digestion with PstI also released a 160 bp fragment observed earlier (figure 12b, lane b), and a 1250 bp fragment. The former represented the distance between the internal PstI site and the leftward PstI cloning site, while the latter was the distance between the internal PstI and EcoRI sites (figure 17a).

(b) MAPPING OF pGF1

Digestion of pGF1 with HindIII released a 425 bp fragment (figure 15b, lane 4), of which 265 bp was the distance between the HindIII and PstI sites of pKT218. The internal HindIII site was therefore placed 160 bp from the rightward PstI cloning site (figure 17c). This 160 bp fragment was seen on restriction of HindIII digested pGF1 with PstI (lane 1). Besides this three other bands were also seen. A 250 bp band observed earlier (figure 12b, lane a), placed the internal PstI site 250 bp from the leftward PstI cloning site (figure 17c). Of the other two bands, one 265 bp in length was also present in the PstI+HindIII double digest of pKT218 (lane 2), and the other 540 bp long was assigned as the distance between the internal PstI and Hind III sites (figure 17c).

(C) MAPPING OF pBF5 AND pBF1.

Clone pBF5 was mapped with BamHI, Bg1II, PstI and EcoRI. BamHI digestion of pBF5 excised a 1430 bp fragment (figure 15c, lane 2). Since the PstI—BamHI fragment of pKT218 is 610 bp, the internal BamHI site was placed 820 bp from the rightward PstI cloning site(figure 17b). Secondary digestion with PstI released the expected 820 bp fragment (lane 6). A 630 bp fragment was also seen (lane 6) and assigned as the distance between the internal BamHI site and the leftward PstI cloning site (figure 17b). However, the PstI and BamHI sites of pKT218 are also about the same distance and a 610 bp fragment is released when this DNA is restricted with these enzymes (lane 5). The two fragments

were distinguished by a diagnostic restriction with EcoRI as the insert has no site for this enzyme, while an EcoRI site exists between the BamHI and PstI sites of pKT218. Restriction of PstI+BamHI double digested pBF5 DNA with EcoRI resulted in the appearance of 235 bp and 375 bp fragments (lane 4). These represented EcoRI cleavage products of the 610 bp pKT218 PstI—BamHI fragment seen in lane 3. The 630 bp was not cleaved (lane 4), thus confirming it to be an insert fragment.

Earlier restriction analysis revealed the presence of a BglII site in pBF5 insert. This site was mapped readily as pKT218 does not have a site for this enzyme (Talmadge and Gilbert, 1980). Secondary digestion with BglII of BamHI digested pBF5 gave cleavage products of 1200 bp and 230 bp (lane 7) derived from the 1430 bp BamHI fragment pBF5 (lane 2). The BglII site was therefore placed 1200 bp from the plasmid BamHI site and 230 bp to the right of the internal BamHI site (figure 17b).

Mapping the internal BamHI and BglII sites of pBF5 enabled the smaller cross-hybridising clone (pBF1) to be assigned to the rightward region of pBF5. It appeared that the 720 bp insert of pBF1 was not long enough to have both the BamHI and the BglII sites mapped in pBF5 insert. Figure 15d (lane 2) shows the presence of a BglII site in clone pBF1. No internal BamHI site was detected in this recombinant. This was confirmed during sequence identification of the clone.

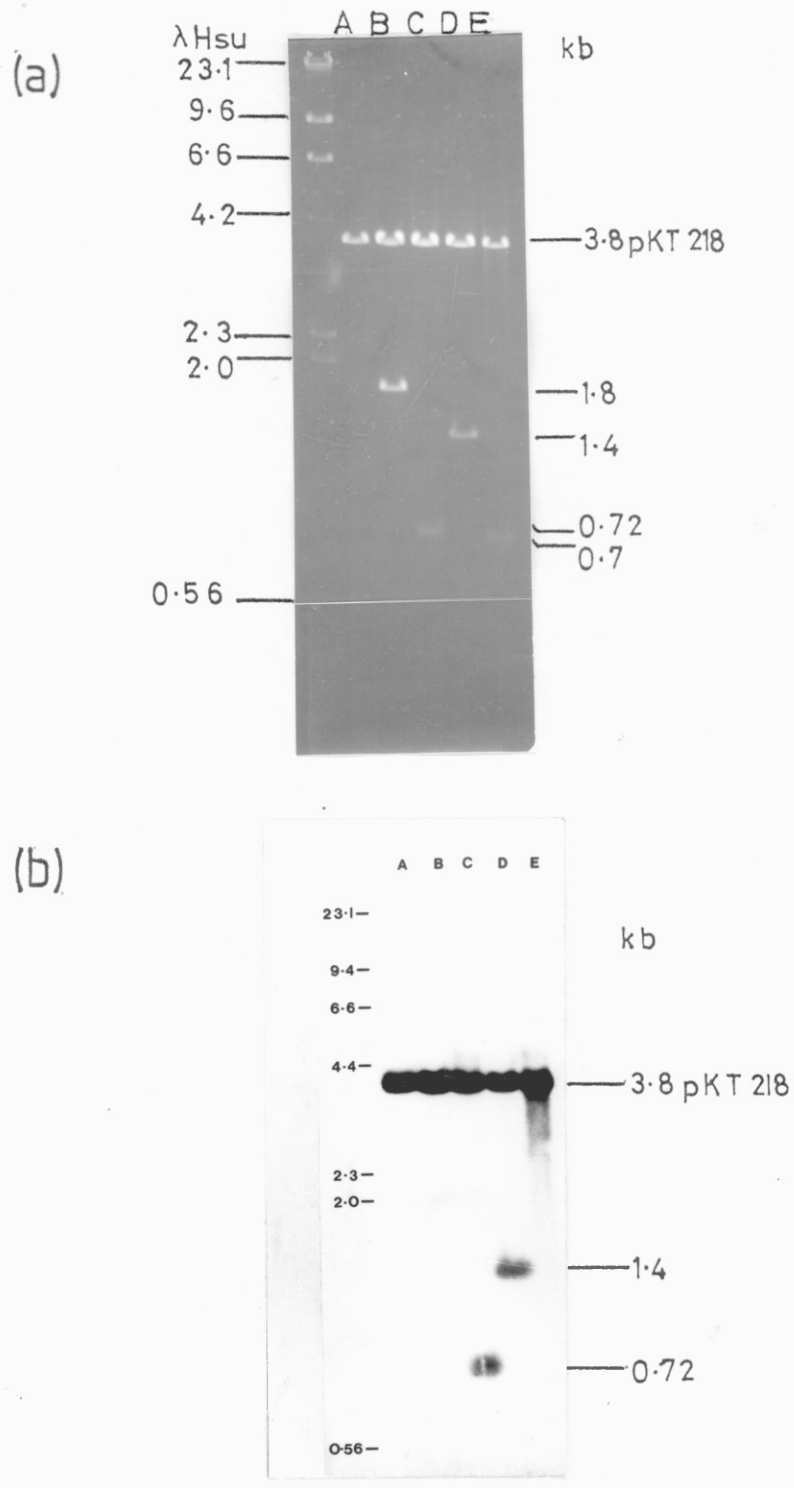

Figure 13. Cross-hybridisation of pBF1 and pBF5.
(a) 0.5 ug of PstI digested pKT218 (lane A), pAF1 (lane B), pBF1 (lane C), pBF5(lane D) and pGF1 (lane E) fractionated on a 1% agarose gel.
(b) Southern blot of the gel in figure 13(a) hybridised with nick-translated pBF1. Washes were at high stringency (0.1xSSc, 65°C, 5x30 min.).

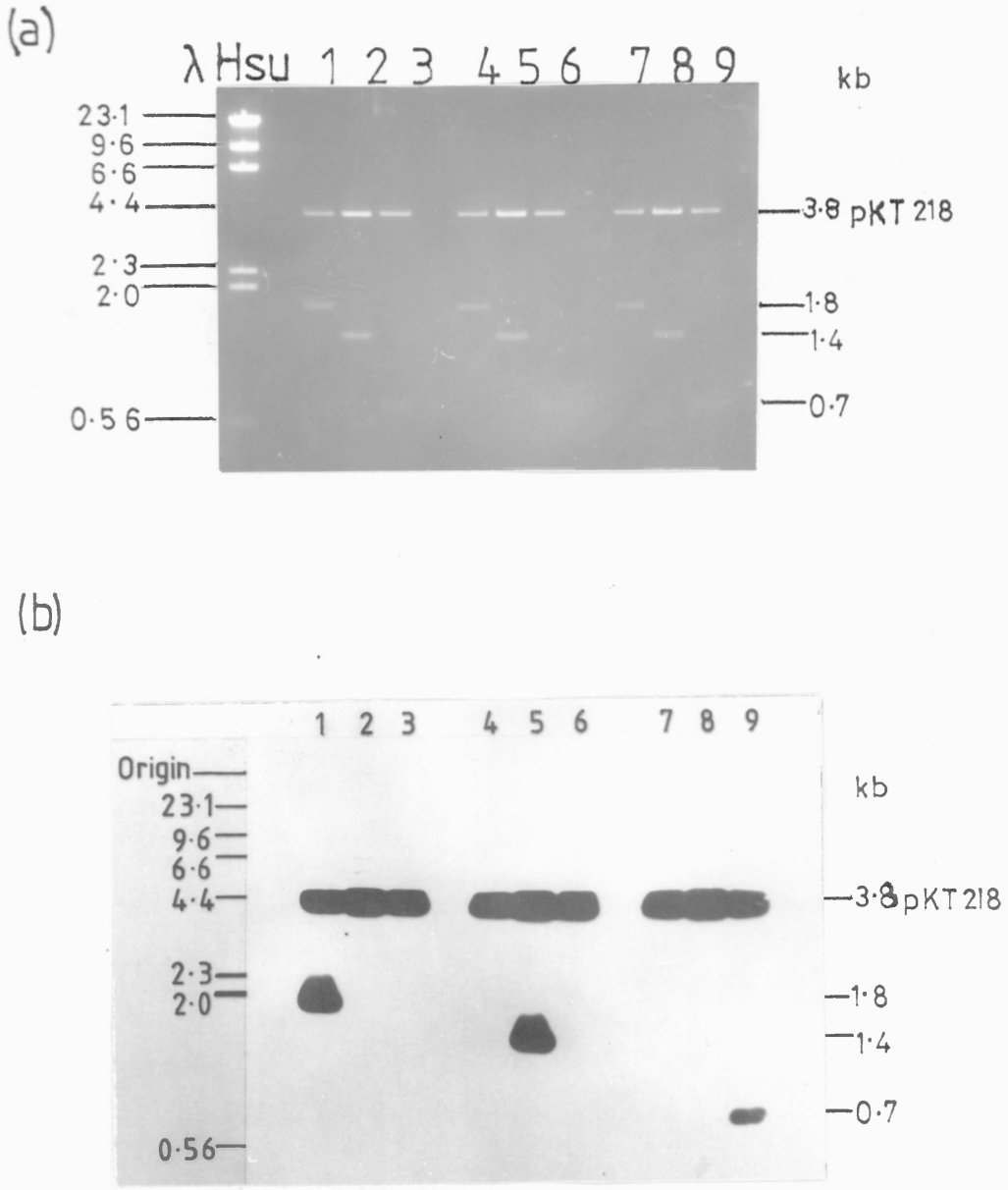

Figure 14. Cross-hybridisation analysis of clones pAF1, pBF5 and pGF1.
(a) 1% agarose gel of 0.1 ug each of PstI digested pAF1 (lanes 1, 4 and 7), pBF5 (lanes 2, 5 and 8) and pGF1 (lanes 3, 6 and 9).
(b) Southern blot of the gel in figure 14(a), hybridised with nick-translated clones pAF1, pBF5 and pGF1.
Lanes 1-3 were hybridised with pAF1, lanes 4-6 hybridised with pBF5 and lanes 7-9 with pGF1. Low stringency washes were at (3xSSC, 65°C, 2x30 min.).

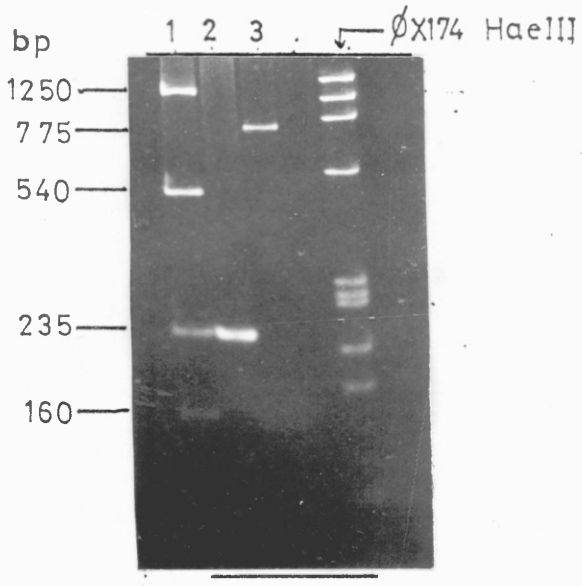

Figure 15a. Restriction mapping of pAF1. Plasmid DNAs (5ug/lane) were digested with restriction endonucleases and fractionated on a 5% acrylamide gel. EcoRI+PstI digested pAF1 (lane 1), EcoRI+PstI digested pKT218 (lane 2) and EcoRI digested pAF1 (lane 3).

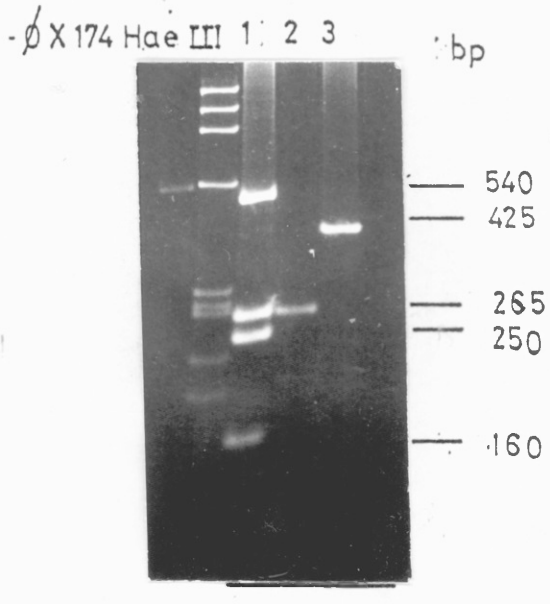

Figure 15b. Restriction mapping of pGF1. Plasmid DNAs (5 ug/lane) were digested with restriction endonucleases and fractionated on a 5% acrylamide gel. PstI+HindIII digested pGF1 (lane 1), PstI+HindIII digested pKT218 (lane 2) and HindIII digested pGF1 (lane 3).

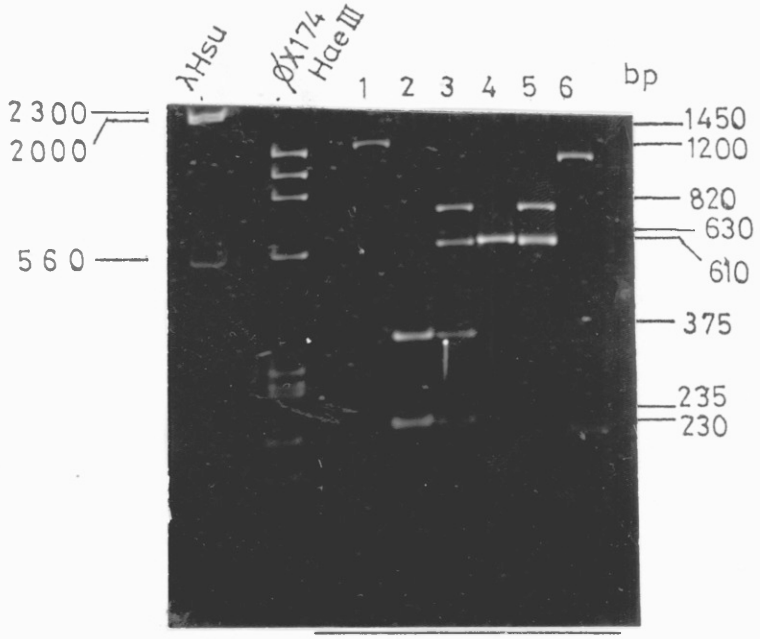

Figure 15c. Restriction mapping of pBF5. Plasmid DNAs (5 ug/lane) were digested with restriction endonucleases and fractionated on a 5% acrylamide gel. BamHI digested pBF5 (lane 1), BamHI+PstI+EcoRI digested pKT218 (lane 2), BamHI+PstI+EcoRI digested pBF5 (lane 3), BamHI+PstI digested pKT218 (lane 4), BamHI+PstI digested pBF5 (lane 5) and BamHI+BglII digested pBF5 (lane 6).

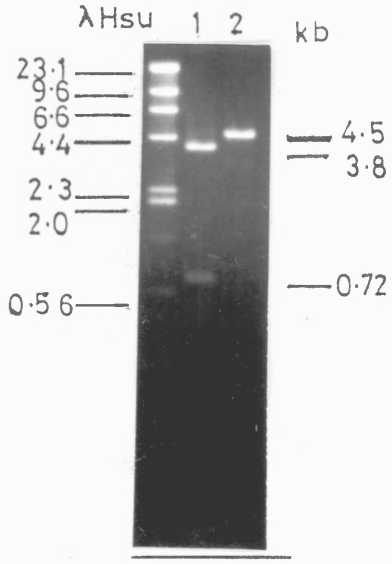

Figure 15(d). Digestion of pBF1(0.5 ug/lane) with PstI (lane 1) and BglII (lane 3).

(d) LOCALISATION OF POLY(A) IN pAF1, pBF5 AND pGF1 INSERTS.

Cloning of cDNAs at the PstI site using G-C tailing offers two advantages. Firstly, this method reconstructs the PstI site on either side of the insert (Williams, 1981; Otsuka, 1981) thus enabling the cDNA insert to be excised readily and secondly, any positive hybridisation with end-labelled oligo(dT) would of necessity imply the presence of poly(A) sequences in the cDNAs.

Apart from demonstrating the presence of a 3' untranslated region in cDNAs, oligo(dT) hybridisation can also be used to orientate the cDNA insert as relative to the vector. In this study the cDNA inserts of clones pAF1, pBF5 and pGF1 were orientated as relative to the β-lactamase promoter - a fact useful for expression studies of these recombinants in E.coli.

Figure 16(a) shows a ' mapping gel' that was Southern blotted and the filter hybridised with γ-^{32}P ATP labelled oligo(dT)$_{12-18}$ (sp. activitity of 5×10^7 cpm/ug.) Hybridisation conditions and post hybridisation washes were as for oligomer hybridisations described earlier. Figure 16(b) shows oligo(dT) hybridising to the 0.76 kb EcoRI fragment and the 1.8 kb PstI fragment of pAF1 (lanes 2 and 3). Secondary digestion of the latter with EcoRI yielded a 0.5 kb oligo(dT) hyridising fragment (lane 4). Similarly BamHI digested pBF5 showed oligo(dT) hybridising to the 1.4 kb fragment (lane 7) and to the 1.45 bp PstI excised insert (lane 8). The poly(A) bearing internal fragment was localised by BamHI secondary digestion of

the PstI digested clone. This showed that the oligo(dT) probe hybridised to the 0.8 kb insert BamHI--PstI fragment (lane 9).

The same method showed that the oligo(dT) hybridised to the larger PstI excised insert fragment of pGF1 (lane 13), suggesting that the insert fragment bearing the poly(A) sequence is the HindIII—PstI fragment (figure 17c). Digestion of pGF1 with HindIII released a 0.46 kb oligo(dT) hybridising fragment (lane 14) thus confirming the earlier suggestion.

The positive control for oligo(dT) hybridisation is shown in lane 12. This is an EcoRI digest of clone pB4 - a cDNA clone for human apolipoprotein E, isolated in the department, and shown by sequencing to contain a poly(A) 35 residues long (S. Wallis, personal communication). The negative control was provided by various restriction digests of pKT218. No detectable hybridisation was observed with this DNA. Table 2 summarises the restriction digests performed and the result of the oligo(dT) hybridisation.

It is apparent that each of the the recombinants pAF1, pBF5 and pGF1 have a poly(A) and therefore are likely to have the 3' untranslated region. Restriction maps locating the position of the poly(A) sequence (figures 17a, 17b and 17c) for the three recombinants show that the cDNA insert in each is in an incorrect orientation for their expression as fibrinogen peptides from the β-lactamase promoter of pKT218, which transcribes from right to left.

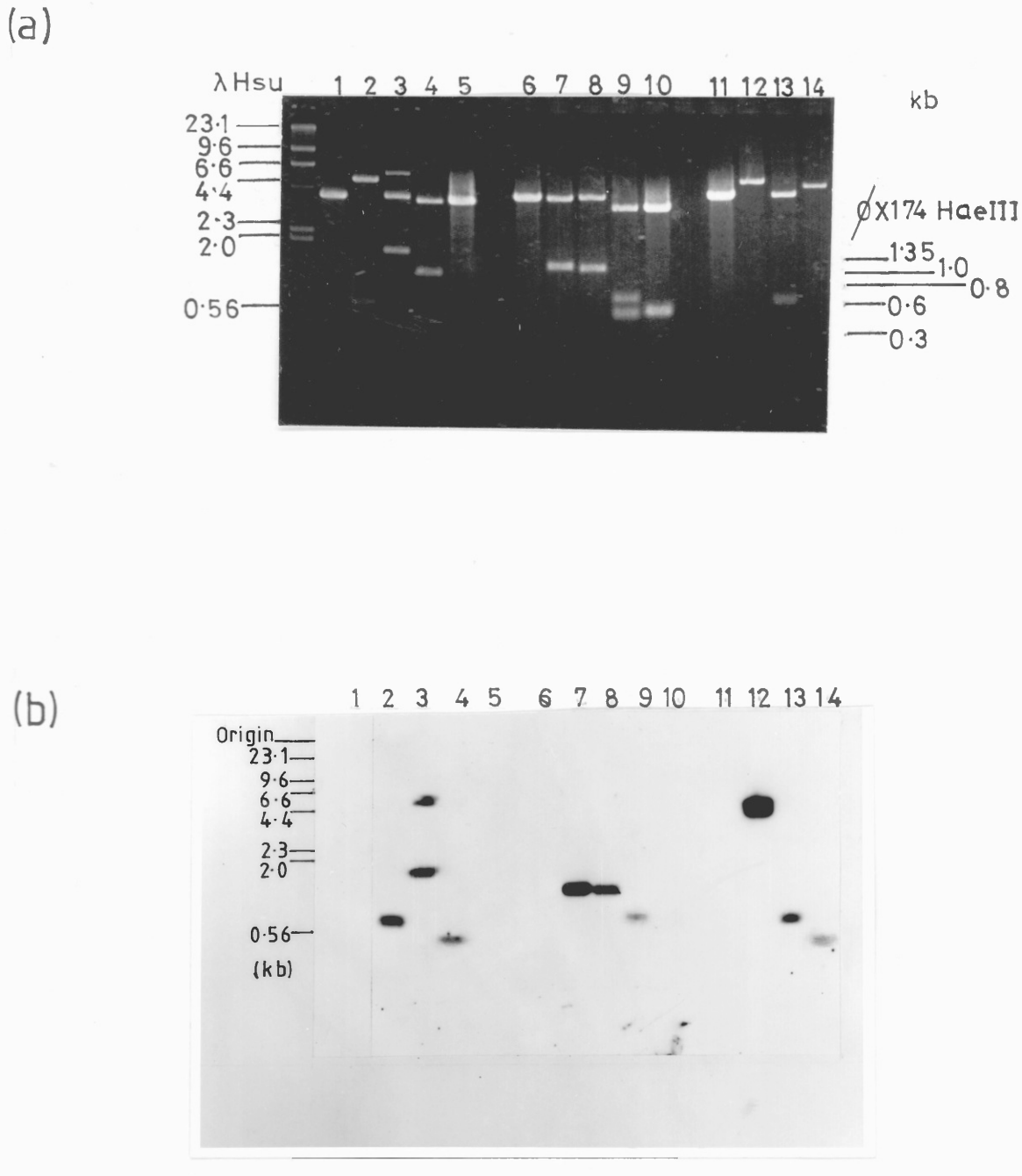

Figure 16. Localisation of poly(A) in cDNA inserts of pAF1, pBF5 and pGF1.
(a) A mapping gel of pAF1, pBF5 and pGF1. Plasmid DNAs (1 ug/lane) were digested with the appropriate restriction endonuclease and fractionated on a 1% agarose gel.
(b) Southern blot of the gel in figure 16(a) hybridised with end-labelled oligo(dT)$_{12-18}$.
Table 2 summarises the restriction enzyme digests performed and the result of the oligo(dT) hybridisation.

Table 2

Lane	DNA	RESTRICTION FRAGMENTS (kb)					
		EcoRI	PstI	EcoRI + PstI	BamHI	BamHI + PstI	Hind III
1	pKT218	3.8					
2	pAF1	4.9 0.76					
3	pAF1		5.45* 3.8 1.8				
4	pAF1			3.34 1.25 0.50 0.23			
5	pKT218			3.6 0.2			
6	pKT218				3.84		
7	pBF5				3.7 1.4		
8	pBF5		3.8 1.4				
9	pBF5					2.8 0.8 0.6	
10	pKT218					2.8 0.6	
11	pKT218	3.8					
12	pB4	4.2					
13	pGF1		3.8 0.7				
14	pGF1						3.8 0.46

* Partial.

Table 2.

A summary of the restriction enzyme fragments shown in figure 16a and the result of the oligo(dT)$_{12-18}$ hybridisation in figure 16b. All oligo(dT) hydridising fragments are underlined. The 5.45 kb oligo(dT) hybridising fragment in lane 3 is a partial.

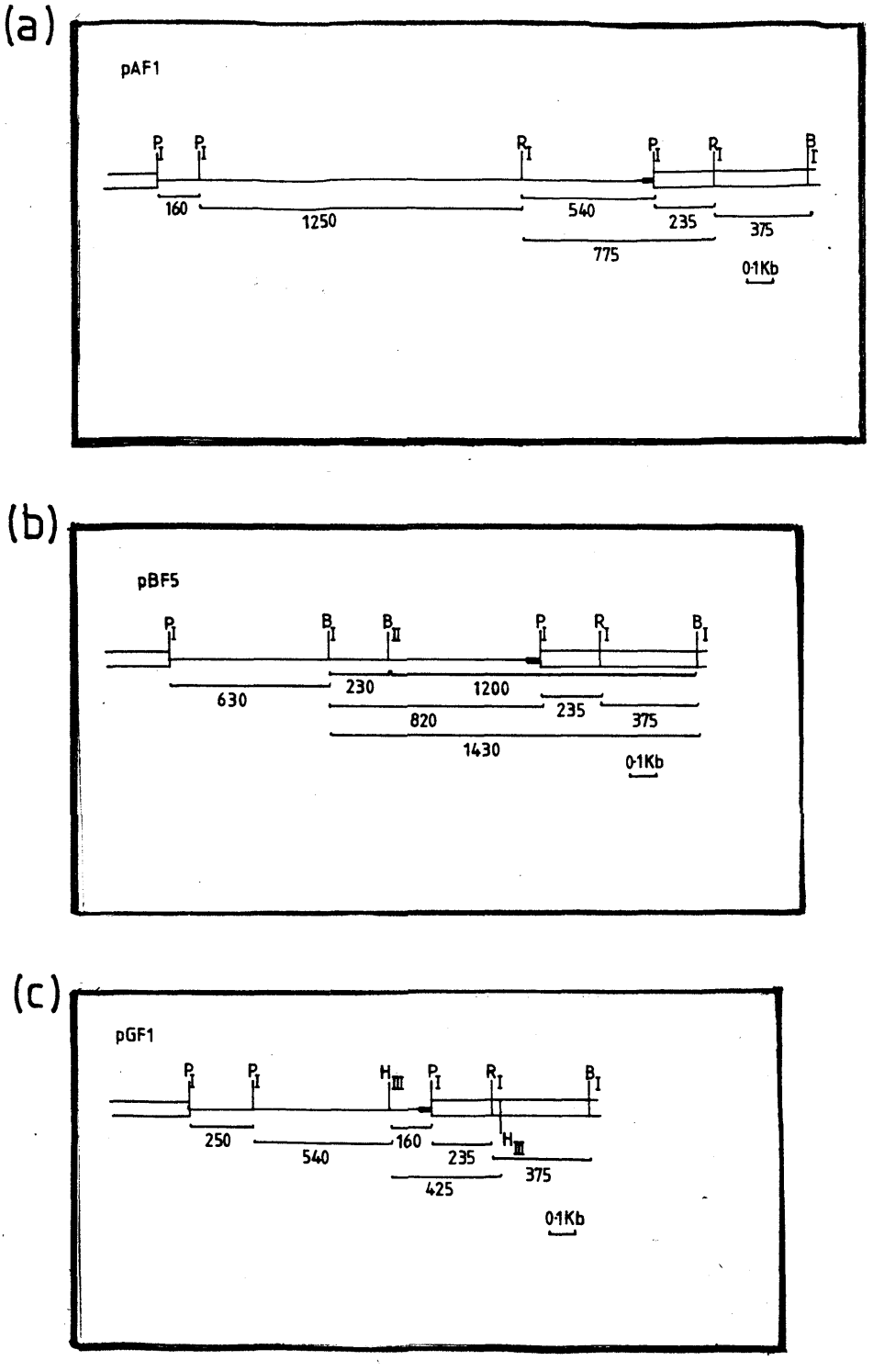

Figure 17. Restriction maps of pAF1 (a), pBF5 (b) and pGF1 (c). Boxed area represents plasmid pKT218 DNA. Line joining boxed area is cDNA insert with shaded area representing poly (A) (not to scale). Restriction sites indicated are: B_I = BamHI, B_{II} = BglII, H_{III} = HindIII, P_I = PstI and R_I = EcoRI.

3.2.3. Sequencing Strategy

Sequence indentification of the cDNAs was by the dideoxy chain termination method (Sanger et al ., 1977). The cDNA recombinants were subcloned into the polylinker of M13mp9 replicative form DNA, and the recombinant single stranded phage DNA sequenced by priming from the universal priming site in the presence of dideoxynucleotides, using DNA polymerase I (Klenow fragment). The products of the reaction were fractionated on a 6% acrylamide/8M urea sequencing gel and the sequence read after overnight autoradiography at room temperature.

Two points were taken into consideration when deciding sequencing strategy. Firstly, only an identifying DNA sequence was required which meant that a small region (150 bp - 250 bp) of the cDNA need be sequenced. Secondly, this region had to be in the coding block which meant eliminating all those enzyme digests that gave small fragments bearing poly(A). For example, PstI and HindIII digest of pGF1. Such fragments could consist entirely of 3'non-coding region, thus giving no identification of the clone.

It has been reported that very small DNA fragments (<250 bp) clone at higher efficiency due to a lower frequency of recircularisation on steric grounds (Shore et al., 1981). Digestion of pAF1 with PstI generates a 3.8 kb vector fragment, a 1.79 kb large insert fragment and a small 160 bp insert fragment. The latter due to its small size was expected to be cloned in M13 at higher efficiency

and also to give more stable recombinants than the former two fragments. The size limit of about 1 kb clonable fragment in M13mp9, was also used to advantage in generating subclones of pGF1. PstI digestion of this recombinant gives a 3.8 kb pKT218 fragment which was not expected to be cloned efficiently, the larger 700 bp insert fragment and the small 250 bp insert fragment. Of the two insert fragments, the latter was expected to be cloned at higher efficiency. However, the possibility of the 700 bp fragment being cloned, albiet, at low efficiency was not eliminated.

Recombinant pBF1 was used instead of pBF5 for subcloning into M13. PstI digestion of pBF1 releases a 720 bp insert fragment which would be cloned at higher efficiency than the 1.4 kb insert fragment of pBF5. Size selection in M13 cloning would in this case also result in the 3.8 kb vector fragment being cloned at a lower efficiency.

3.2.4. Subcloning of cDNA Recombinants into M13mp9 and the Preparation of Recombinant Single Stranded M13 DNA.

Recombinants pAF1, pBF1 and pGF1 were subcloned as described in section 2.5.a. Cloning into the polylinker inactivates α-complementation, distinguishing recombinants (colourless plaques) from parentals (blue plaques). Plasmid DNAs digested with PstI were ligated to PstI cleaved replicative form of M13mp9 DNA using T4 DNA ligase. In this reaction the cohesive termini of PstI linearised M13 vector

and PstI cleaved insert DNA anneal and the staggered nicks joined by T4 DNA ligase with concomitant hydrolysis of ATP (Weiss and Richardson, 1968) so reconstructing the PstI site. Ligated DNAs were transfected into competent E.coli K12 JM103 cells by heat shock at 42°C for 2 min. Cells were plated on YT agar plates in soft agar containing IPTG (a gratuitous inducer of β-galactosidase) and X-Gal (a histochemical substrate of the enzyme). Overnight incubation at 37°C allowed phage growth and the colour reaction to develop.

Recombinants (colourless plaques) from each of the three subcloning experiments were picked and single stranded phage DNA prepared as described in Materials and Methods (2.5.d). pAF1 subclones were grouped as MA series, pBF1 subclones as MB series, and pGF1 subclones as MG series. The mobilities of recombinant phage DNAs was compared with that of single stranded M13mp9 DNA on a 1% agarose gel (figure 18). The varying mobilities of the recombinant phage DNAs was due to the different sizes of the expected inserts and also the variation that exists in the degree of secondary structure of single stranded DNAs.

3.2.5. Single Lane Screening of Subclones

Subclones were screened for the purpose of grouping like recombinants so that a representative from each group could be sequenced. This screening was performed by carrying out a single lane sequencing reaction, in this case dTTP/ddT ('T tracking'), for a number of recombinants and

comparing their tracks on a sequencing gel. Figure 19 shows the 'T tracking' of a number of M13 subclones. All recombinants show a unique 'T tracking' pattern. This pattern is also different from that given by M13mp9 DNA (lanes 1 and 9) which was taken as evidence of the recombinants being bona fide, rather than deleted M13mp9 giving colourless plaques.

Similar single lane screens were performed on other subclones and like recombinants were grouped together. A representative from each was then subsequently sequenced.

3.2.6. Sequencing of cDNA Recombinants Subcloned in M13

Sequencing of M13 subclones revealed pAF1 to be an Aα-chain cDNA clone, pBF1 to be a Bβ-chain clone and pGF1 was identified as a γ-chain clone. MA1 DNA sequence is shown in figure 20(a). The single nucleotide difference from the Aα- fibrinogen cDNA sequence of Kant et al.,(1983) is indicated. Using a computer search programme, written by Mr. Paul North of this department, the sequence was used to search a protein data file updated to 1982. This identified MA1 insert as the anti-coding strand of Aα-fibrinogen cDNA. The complete sequence of MA1 insert, which is the 160 bp PstI insert fragment of pAF1, and its corresponding protein sequence is shown in figure 21(a).

Subclone MB14 was similarly sequenced and shown to have an insert corresponding to the coding strand of Bβ-fibrinogen cDNA. The partial sequence is shown in figure 20(b). The sequence starts at the PstI cloning site,

proceeds through the poly(G) tract; interrupted by a single C at nucleotide 22, and into the insert commencing at the codon for amino acid 277 of Bβ-fibrinogen. A total of 140 bases were read from two different runs and are shown in figure 21(b) with their corresponding amino acid sequence.

Figure 20(c) shows the sequencing gel of MG4. Computer assisted search of the protein data file with this DNA sequence identified MG4 insert to represent the coding strand of γ-fibrinogen cDNA. The complete sequence of the 250 bp PstI insert fragment and its corresponding protein sequence is shown in figure 21(c). The starting codon of the sequence is amino acid 183 of γ-fibrinogen.

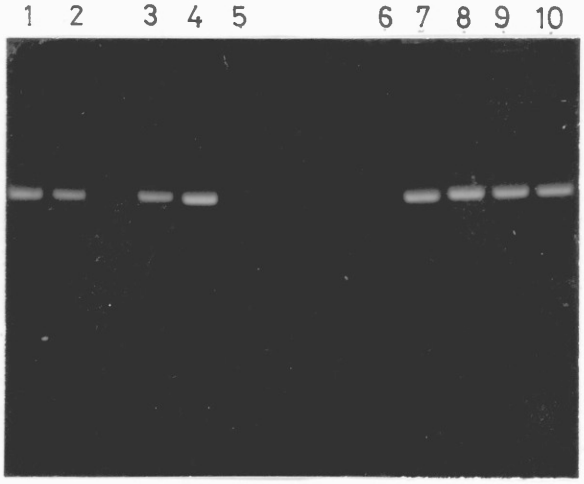

Figure 18. A 1% agarose gel of M13 recombinant phage minipreps. Mobilities of M13 subclones MA1 (lane 1), MA2 (lane 2), MB3 (lane 3), MB14 (lane 5), MG1 (lane 6), MG2 (lane 8), MG4 (lane 9) and MG5 (lane 10) as relative to M13mp9 DNA (lanes 4 and 7).

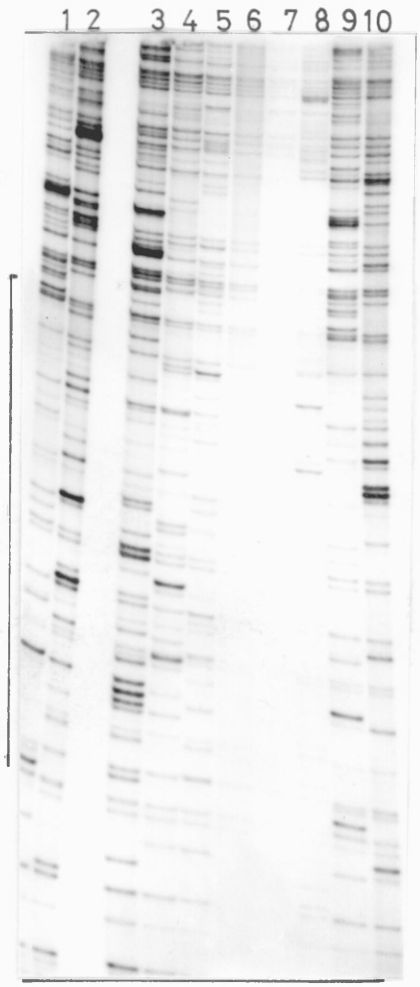

Figure 19. Single lane screening on a 6% acrylamide sequencing gel. 'T tracking' of M13 subclones MA1 (lane 2), MA2 (lane 3), MB3 (lane 4), MB14 (lane 5), MG1 (lane 6), MG2 (lane 7), MG4 (lane 8) and MG5 (lane 10). Lanes 1 and 9 have M13mp9.

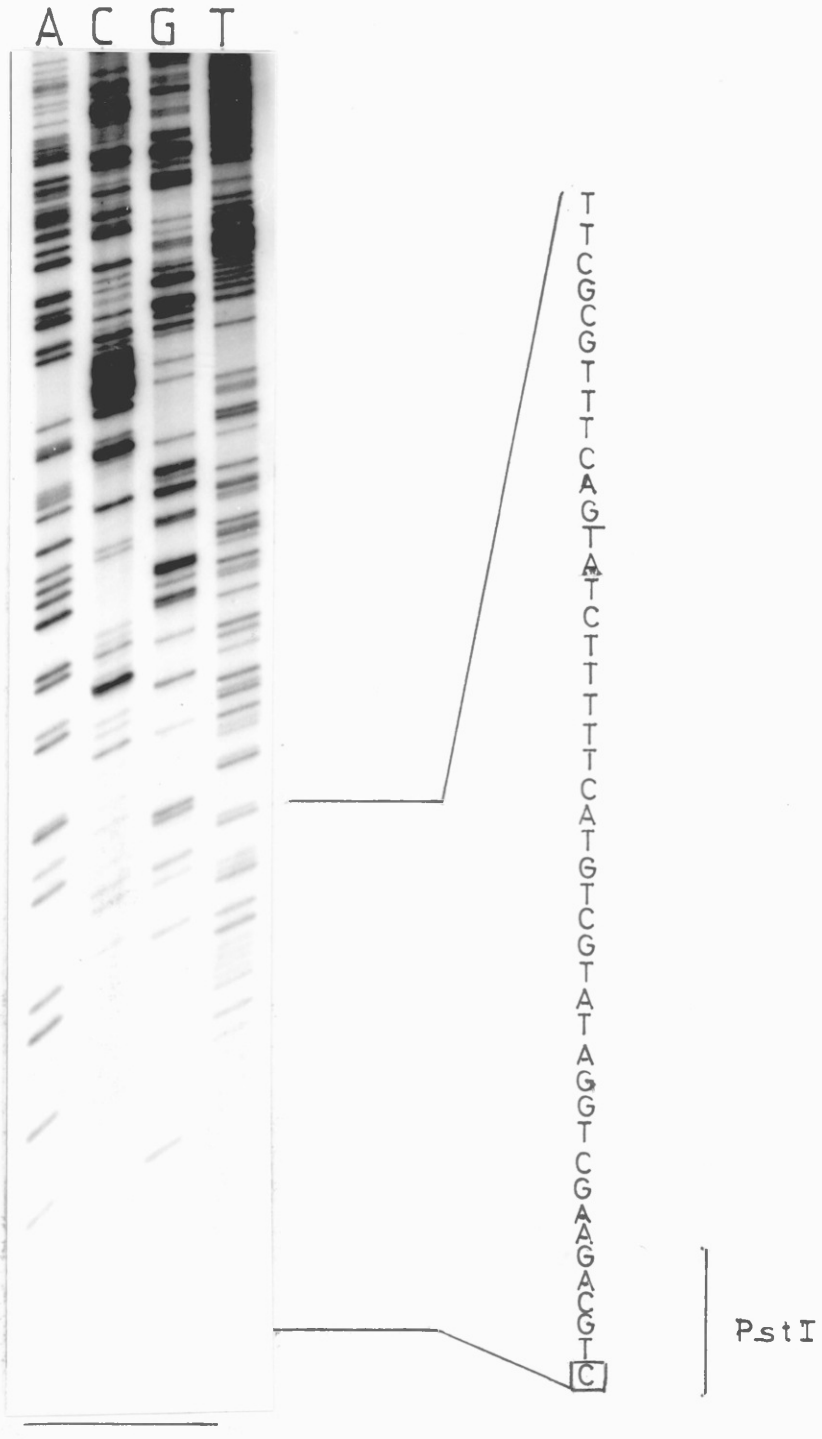

Figure 20a. A 6% acrylamide sequencing gel of subclone MA1 showing the sequence of the anti-coding strand. Boxed nucleotide represents the single base difference from the sequence of Kant et al., (1983).

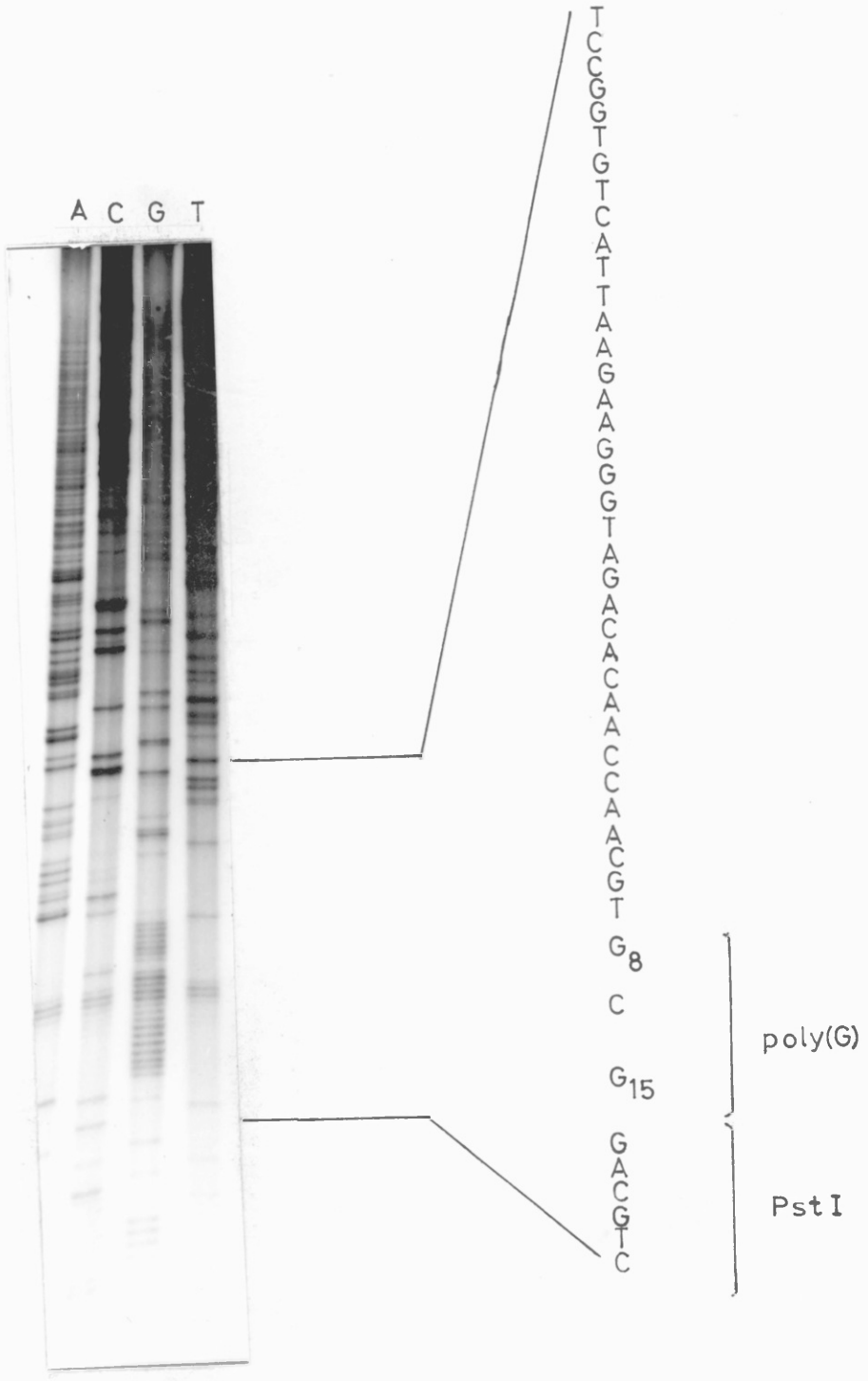

Figure 20b. A 6% acrylamide sequencing gel of subclone MB14 showing the sequence of the coding strand, the poly(G) tract and the PstI site.

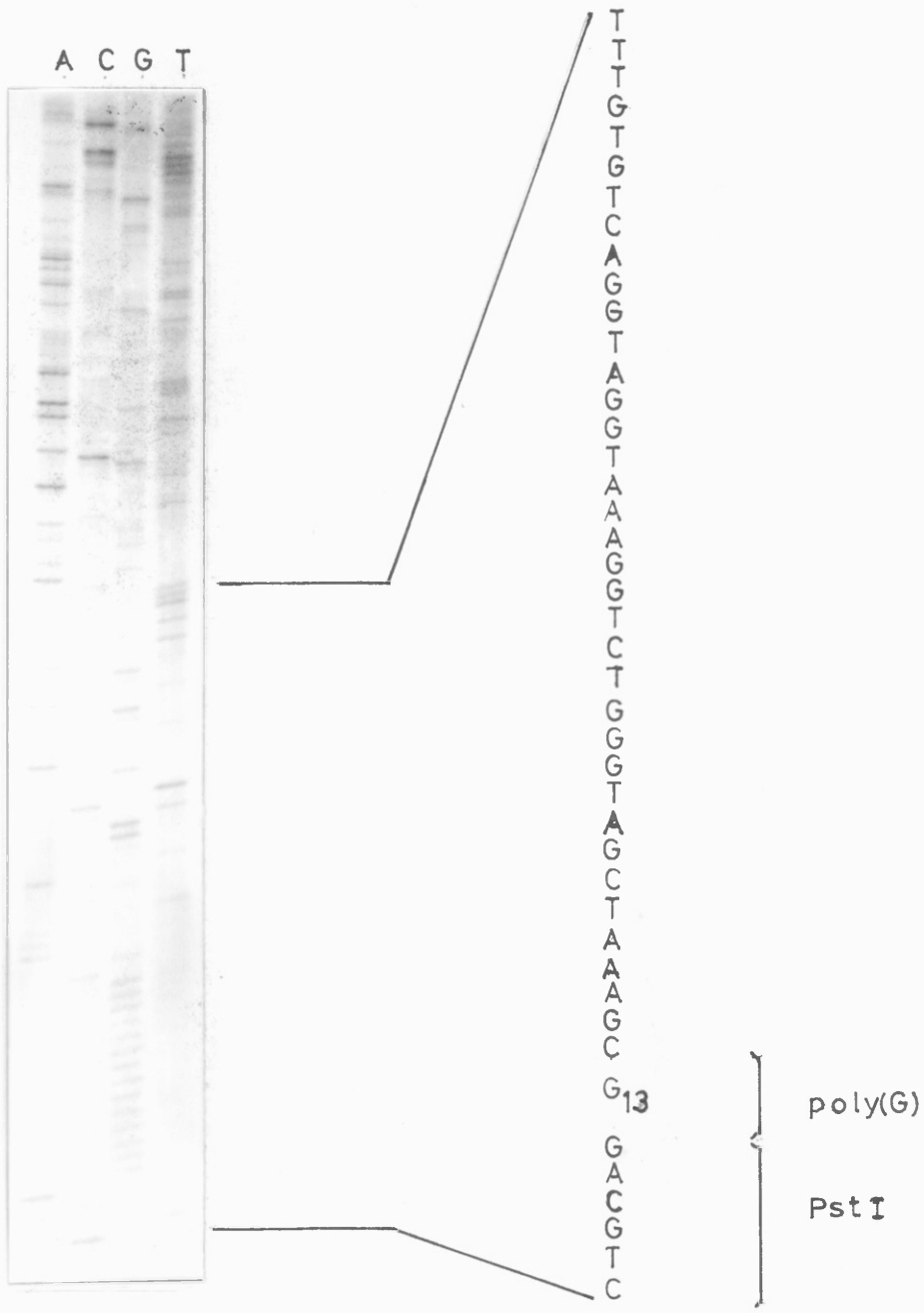

Figure 20c. A 6% acrylamide sequencing gel of subclone MG4 showing the sequence of the coding strand, the poly (G) tract and the PstI site.

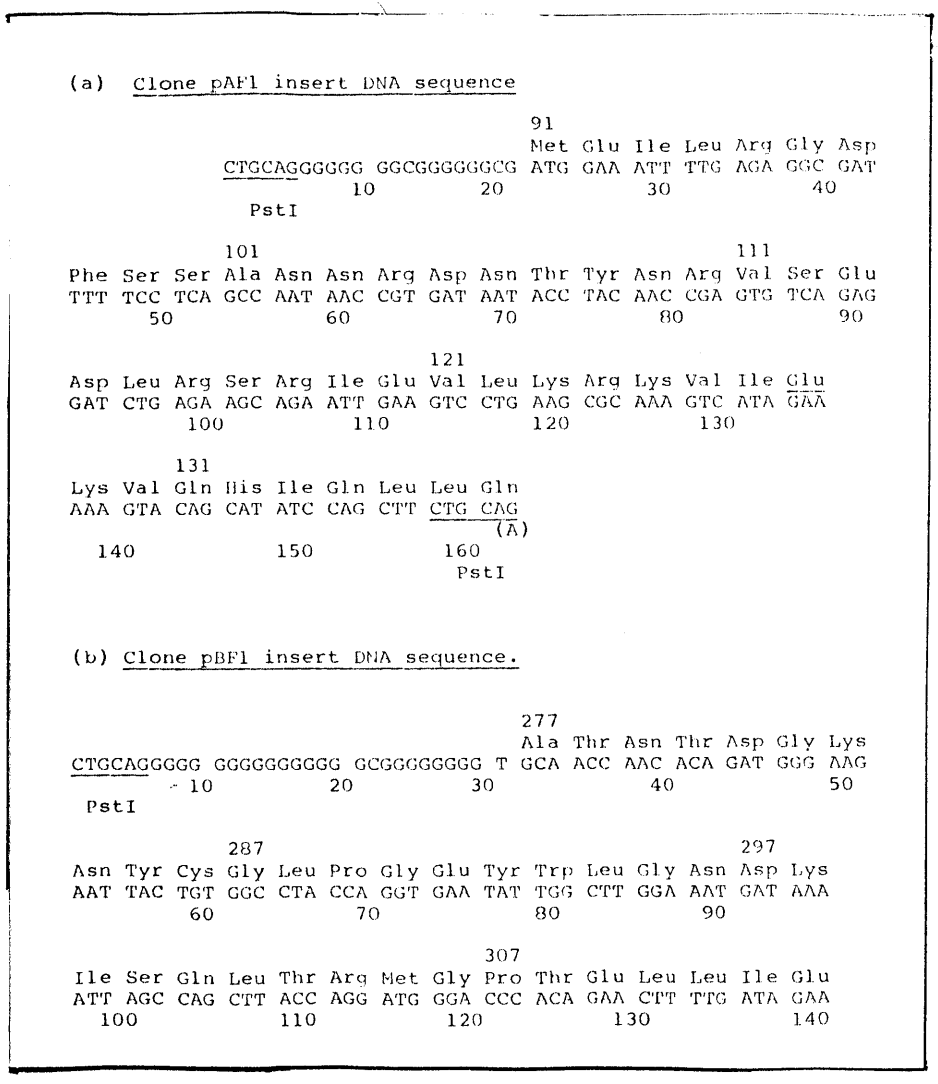

Figure 21. Partial nucleotide sequences of clones pAF1 (a) and pBF1 (b) inserts and their corresponding amino acid sequences. The numbering of amino acids is as reported in the literature. The nucleotide in parantheses is that reported by Kant et al., (1983).

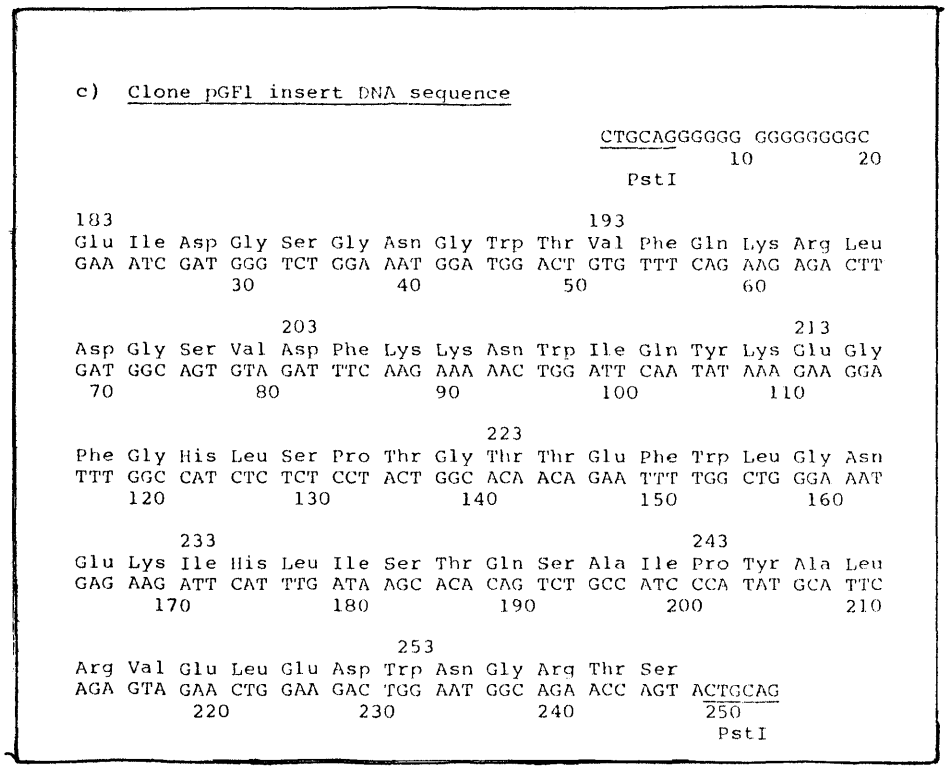

Figure 21 (contd.). Partial nucleotide sequence of pGF1 insert and its corresponding amino acid sequence. The numbering of amino acids is as reported in the literature.

3.3. Genomic Analysis, Expression in Adult Human Liver and Chromosomal Assignment of Fibrinogen Genes

Having established that pAF1 is an Aα-fibrinogen cDNA clone, pBF1 and hence pBF5 are Bβ-fibrinogen cDNA clones and pGF1 is a γ-fibrinogen cDNA recombinant, these clones were used as probes for the analysis of genomic organisation, expression in adult human liver and chromosomal localisation of human fibrinogen genes.

Genomic analysis was performed by Southern blot hybridisation of human DNA using nick-translated cDNA recombinants as probes. Genomic DNA was digested with restriction enzymes recognising a six base sequence to minimise cleavage within the gene. In this way an estimate of the number of genomic sequences hybridising to each probe was made which determined whether the gene for each chain of human fibrinogen is a unique sequence or constitutes a multigene family. This also facilitated the chromosomal assignment of the genes for human fibrinogen by restriction fragment mapping in somatic cell hybrids.

The expression of fibrinogen genes in adult human liver was examined by Northern blot hybridisations of total RNA prepared from this tissue. This determined the number of mRNA species hybridising to each clone and gave an estimate of their size.

3.3.1. Genomic Analysis of Aα-, Bβ-and γ-fibrinogen Genes

Ten micrograms of human DNA were digested to completion with each of the enzymes, EcoRI, BamHI, HindIII, BglII and PstI and fractionated on a 1% agarose gel. DNA was blotted onto nitrocellulose (Southern, 1975) and the filters hybridised separately with 10^6 counts of nick-translated pAF1, pBF5 and pGF1 (figure 22a, 22b, 22c and 22d). All probes were labelled to a specific activity of $1x10^8$cpm/ug. Dextran sulphate was added to 10% so increasing reaction kinetics (Wahl et al., 1979) and hybridisations were performed at 65°C for 12-16 hrs. Following post-hybridisation washes (0.1xSSC, 65°C) and overnight exposure to Kodak X-AR5 film, genomic fragments hybridising to the clones could be visualised. The simple band patterns observed suggested that none of the fibrinogen chain genes constitutes a multigene family and that each gene represents a unique sequence in the human genome.

The insert of recombinant pAF1 has an EcoRI site which divides it into 1.4 kb and 0.54 kb fragments (figure 17a). It was therefore surprising that Southern blots of EcoRI digested human DNA hybridised with pAF1 consistently showed a single band 2.5kb long (Figure 22, panel D). It is possible that the second genomic band is either too large for efficient transfer or is cleaved into fragments of sizes not retained on nitrocellulose filters. A third possibility is that both genomic fragments are of almost the same size and therefore not resolvable on an agarose gel.

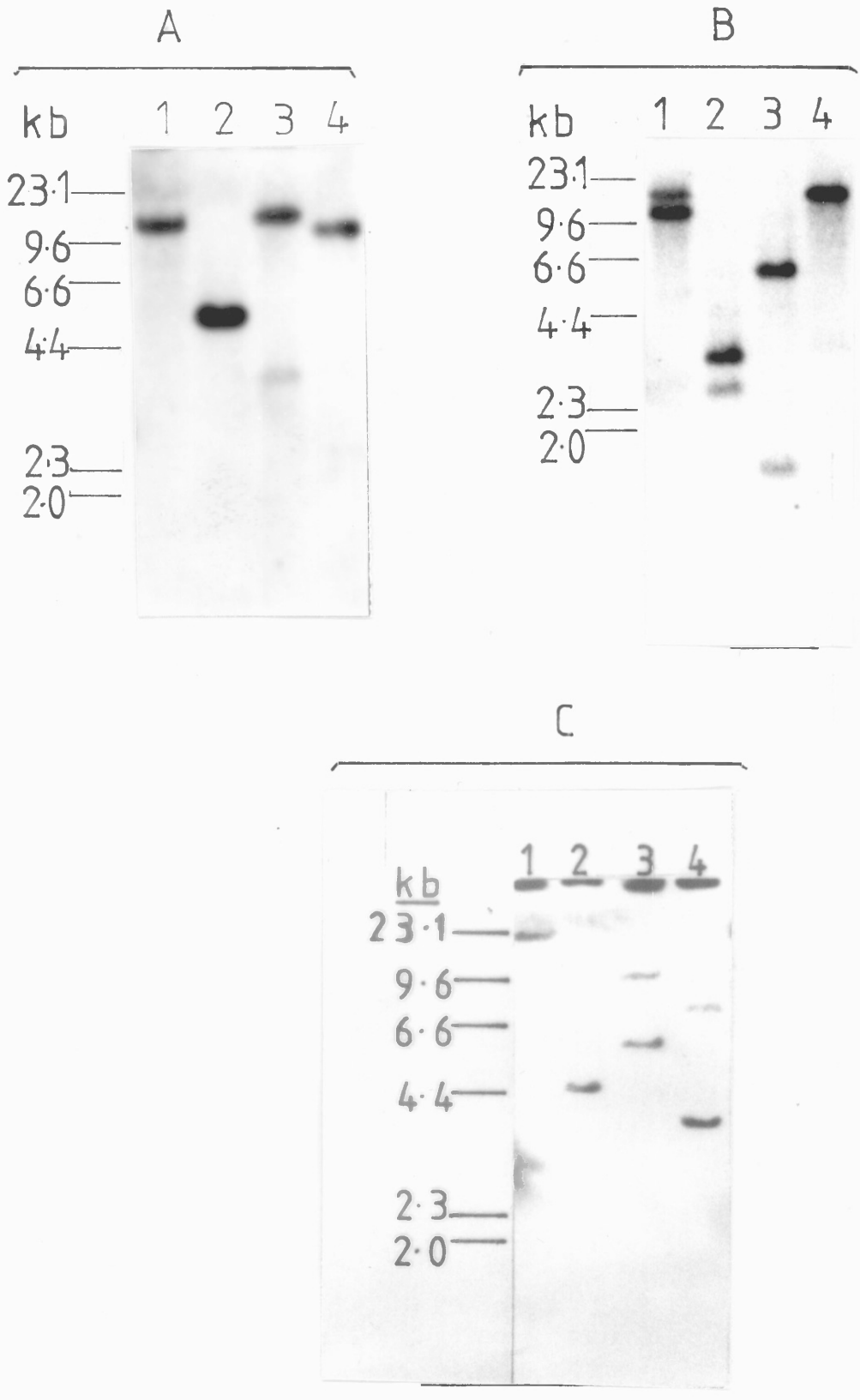

Figure 22. Southern blot hybridisation of human genomic DNA (10 ug/lane) digested with BamHI (lane 1), HindIII (lane 2), BglII (lane 3) and PstI (lane 4) run on a 1% agarose gel. Panel A was hybridised with nick-translated pAF1, panel B with pBF5 and panel C with pGF1.

[TABLE 3 GIVES BANDS SIZES]

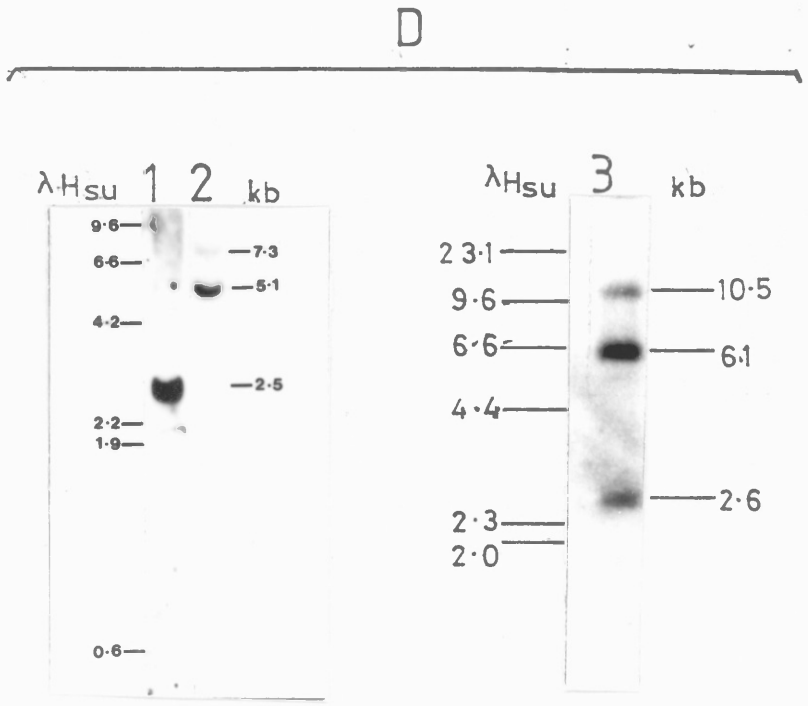

Figure 22. Panel D is a Southern blot hybridisation of EcoRI digested human genomic DNA (10 ug/lane) run on 1% agarose gels and hybridised with nick-translated pAF1 (lane 1), pGF1 (lane 2) and pBF5 (lane 3).

Table 3

	BamHI (lane 1)	HindIII (lane 2)	BglII (lane 3)	PstI (lane 4)
Panel A	12.46	5.1	13.2	11.2
Panel B	15.3 11.1	3.2 2.7	5.8 1.8	15.1
Panel C	23.1	4.1	10.5 5.5	7.4

Table 3.

Fragments sizes (in kb) from Southern blot hybridisations of human genomic DNA shown in figure 22 (panels A, B and C), Panel A was hybridised with pAF1, panel B was hybridised with pBF5 and panel C was hybridised with pGF1.

3.3.2. The Expression of Fibrinogen Genes in Adult Human Liver

A Northern blot of total RNA prepared from a sample of post mortem adult human liver was performed by fractionating the RNA (5 ug/lane) on a 1% agarose/2.2 M formaldehyde denaturing gel and blotting onto nitrocellulose filters. The filters were hybridised in the presence of 10% dextran sulphate with 10^6 counts of nick-translated pAF1, pBF5 and pGF1 (specific activity of $1X10^8$cpm/ug). Figure 23 shows the result of the Northern blot hybridisation after washes at 1XSSC, 65°C. Lane A was hybridised with pAF1, lane B was hybridised with pBF5 and lane C with pGF1. Clones pBF5 and pGF1 each hybridised to a single mRNA of almost equal size. This is to be expected from the sizes of the two polypeptides (460 amino acids of Bβ-chain and 410 amino acids of γ-chain), assuming that the untranslated regions of the mRNAs are not significantly different in length and also the rather poor resolution on agarose gels.

The clone for Aα-fibrinogen (pAF1) hybridises to two RNAs. There is major band at 2.6 kb and a minor band at about 1.8 kb. The latter is unlikely to be a cross-hybridising Bβ- or γ-fibrinogen mRNA as the cDNAs do not share homologous sequences (figue 14b). It is possible that the minor band is a degradation product of the 2.6 kb mRNA.

The intensity of hybridisation of pAF1 to the major Aα-fibrinogen RNA is stronger than the intensities of pBF5 and pGF1 hybridisation to the Bβ- and γ-fibrinogen RNAs. However,as the insert of pAF1 is larger than the inserts of

pBF5 and pGF1, it is to expected that a greater proportion of the former probe hybridises to its RNA sequence giving a stronger signal. The mRNAs for the three chains may therefore have similar abundance. The technique however is not sensitive enough to detect small differences that might exist in the relative abundance of the RNAs.

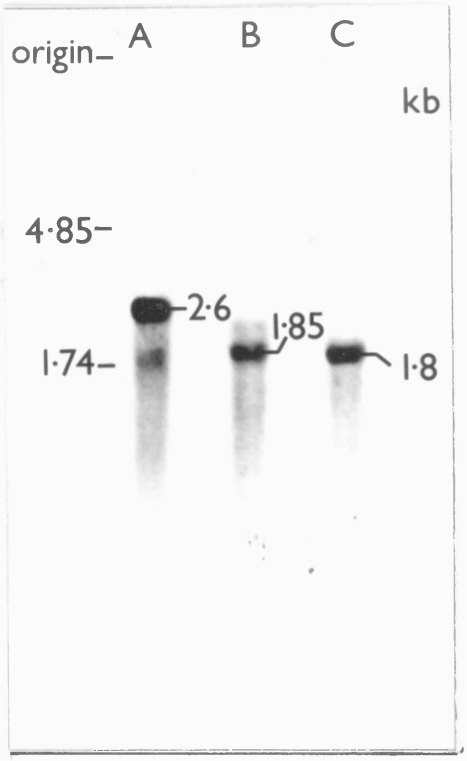

Figure 23. Detection of mRNAs for Aα-, Bβ- and γ-fibrinogen. Total RNA from adult human liver was run on a 1% agarose denaturing gel. Lanes (a), (b) and (c) were respectively hybridised with nick-translated pAF1, pBF5 and pGF1. Human 18S (1.74kb) and 28S (4.8kb) rRNAs were used as markers.

3.3.3. Chromosomal Regional Assignment of The Genes for Aα-, Bβ- and γ-fibrinogen.

Human X CHO somatic cell hybrids were used for assigning fibrinogen genes to human chromosomes. During the course of this investigation all three genes of fibrinogen were assigned to human chromosome 4 (J.C. Kaplan, personal communication). Regional assignment of these genes on human chromosome 4 was performed by restriction fragment mapping in two somatic cell hybrids Paw 4 and Paw 13. Both have a translocated human chromosome 4 with a break point at 4q26 resulting in the 4q26-4qter region being present in each. Hybrids Gal 7 and Dis 20 were also included as they provided additional evidence of the presence of fibrinogen genes on chromosome 4. The human chromosome 'profile' of these hybrids is shown in figure 24.

A Southern blot of EcoRI digested DNAs from the above hybrids and EcoRI digested genomic DNAs of human and chinese hamster was hybridised with 10^6 counts of nick-translated pAF1 and pGF1 (specific activities of $1x10^8$ cpm/ug) as described above in section 3.3.1. Figure 25(a) shows the regional assignment result of human Aα- and γ-fibrinogen genes. The 3.1 kb and 4.5 kb bands are cross-hybridising chinese hamster genomic fragments (lane 1) and are present in all hybrid DNAs (lanes 4-7). The 2.6 kb human Aα-fibrinogen band is present in Paw 4 (lane 4), Paw 13 (lane 5) and Gal 7 (lane 7), but absent in Dis 20 (lane 6). Similarly, the 5.2 kb and 7.1 kb human γ-fibrinogen bands are present in the Paw hybridise (lane 4 and 5) and Gal 7

(lane 7), but absent in Dis 20 (lane 6). These results are consistent with the two genes being on chromosome 4q26-qter.

The regional assignment result of human Bβ-fibrinogen is shown in Figure 25(b). The band at 12.1 kb is a cross-hybridising chinese hamster DNA sequence (lane 3) and is present in all hyrid DNAs (lanes 4-7). The 6.1 kb band is human Bβ-fibrinogen genomic fragment. It is present in hybrids Gal 7 (lane 4), Paw 4 (lane 6) and Paw 13 (lane 7). Hybrid Dis 20 (lane 5) shows the presence of a very minor 6.1 kb band, together with two other minor bands. The presence of the 6.1 kb band represents a chromosome 4 contamination in Dis 20. The result of this blot is however consistent with the Bβ-fibrinogen gene being on chromosome 4q26-qter.

Hybridisation to human genomic DNA (lane 2) shows only the 6.1 kb band. Genomic fragments of 10.1 kb and 2.6 kb observed earlier (Figure 22, panel D) were only faintly visible on the original autoradiogram.

Human Chromosomes	Hybrids			
	Gal 7	Dis 20	Paw 4	Paw 13
1	+	-		
2	+	-		
3	-	+		
4	+	-	4q26-qter	4q26-qter
5	+	+		
6				
7	+	-		
8	+	-		
9	+	-		
10	-	+		
11	+	+		
12	+	+	+	+
13		-		
14	+	-		
15	+	-		
16				
17	+	+		
18		+		
19	+	-		
20	-	+		
21				
22	+	+		
X	+	-		

Figure 24. Human chromosomes and chromosomal regions present in human X CHO somatic cell hybrids as determined by enzyme marker analysis.

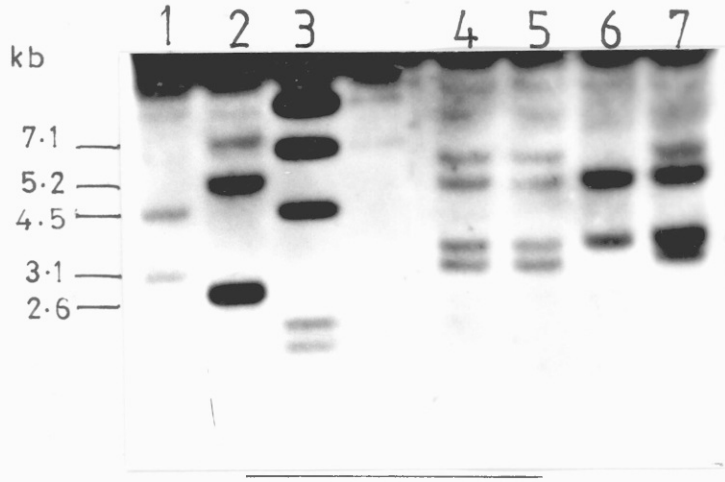

Figure 25a. Regional assignment of human Aα- and γ-fibrinogen genes. Southern blot of EcoRI digested DNAs fractionated on a 1% agarose gel and hybridised with nick-translated pAF1, pGF1 and λDNA.

Lane 1 is chinese hamster DNA, lane 2 is human DNA, lane 3 is λHsu, lane 4 is Paw 4 hybrid DNA, lane 5 is Paw 13 hyrid DNA, lane 6 is Dis 20 hybrid DNA and lane 7 is Gal 7 hybrid DNA.

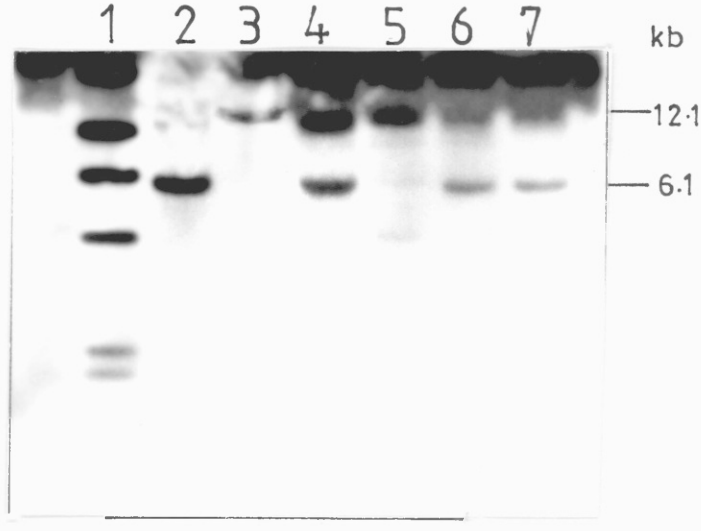

Figure 25b. Regional assignment of human Bβ-fibrinogen gene. Southern blot of EcoRI digested DNAs fractionated on a 1% agarose gel and hybridised with nick-translated pBF5 and λDNA.

Lane 1 is λHsu, lane 2 is human DNA, lane 3 is chinese hamster DNA, lane 4 is Gal 7 DNA, lane 5 is Dis 20, lane 6 is Paw 4 hybrid DNA and lane 7 is Paw 13 hybrid DNA.

4. DISCUSSION

4.1. Direct Oligonucleotide Screening of cDNA Clone Banks

4.1.1. Specificity of Direct Oligonucleotide Screening

The use of oligonucleotide mixtures for the direct screening of recombinant cDNA libraries has become an established technique and this approach was adopted as the means of isolating cDNA clones for human fibrinogen. The feasibility of this method was enhanced by the presence of regions in the amino acid sequences of fibrinogen polypeptides that would give oligonucleotide mixtures of very low complexitiy. For both Bβ- and γ-fibrinogen, amino acid sequences were chosen to give oligonucleotides which were mixtures of only 4 different sequences. This maximises the chances of correct cDNA clone isolation. Even for Aα-fibrinogen, where oligonucleotide mixtures of 32 and 64 different sequences were used, probe complexity was relatively low so that under stringent conditions non-specific hybridisation was unlikely.

17 base long oligonucleotides specific for each chain of human fibrinogen were synthesised. DNAs of this length are capable of encoding 4^{17} different sequences. This figure also represents the complexity of a RNA pool which might contain sequences showing chance complementarity to a given oligonucleotide. Since the estimated complexity of human liver RNA is expected to be considerably less than this (2×10^7 is the complexity of mouse liver RNA as determined by Hastie and Bishop, 1976) oligomers of this

length are sufficiently specific for screening its cDNA clone bank. On the same basis oligonucleotides of even shorter length (14 bases) have been sucessfully used for the direct screening of human liver cDNA libraries (Breslow et al., 1982a and 1982b; Carroll and Porter, 1983).

4.1.2. Direct Screening for Human Fibrinogen cDNAs

The estimated mRNA level for fibrinogen in human liver is about 6%, as determined by cell-free translation of total liver mRNA and immunoprecepitation of translated products (D. Chung, personal communication). Assuming equal abundance of the three sequences, the mRNA level for each chain would be about 2 % and screening 1000 clones of an adult human liver cDNA library should therefore give about 20 positives with each septadecamer oligonucleotide. This assumes that the occurrence of cDNAs in a pool is a reflection of the relative abundance of corresponding mRNAs. Hybridisations under stringent conditions gave a single positive for Aα-chain, 9 positives for Bβ-chain and 3 positives for γ-chain (figure 8b). The abundance of cDNA sequences hybridising specifically to these probes was lower than expected. Chung et al., (1983b) using a γ-fibrinogen oligomer virtually identical to the one described in figure 6, were able to detect 18 positives out of 1000 clones of an adult human liver cDNA library. This figure represents 6x higher abundance for γ-fibrinogen than that observed in this study. Similar direct comparisons between the observed and

published frequencies of Aα- and Bβ-chain cDNAs cannot be made meaningfully because Chung et al., (1983b) used bovine cDNA clones as probes for the isolation of corresponding human cDNA sequences. The oligonucleotides used in this investigation would detect only those cDNA clones representing 3' coding regions of the mRNA sequence.

The abundance of α1-antitrypsin cDNAs in this library was found to be close to the expected value of 2% on screening the same set of filters with an oligonucleotide specific for this cDNA (figure 8b). This implies that there has not been any significant alteration in the relative abundance of this mRNA sequence during cloning and subsequent partial amplification of the cDNA library. The reduced frequencies of fibrinogen cDNA clones may represent artefacts of the cloning procedure or perhaps are a reflection of fibrinogen level variations in the individuals from which the libraries were made. The relative abundance of cDNA clones for the three chains however is the same as that reported recently (Chung et al., 1983b), with Bβ-chain cDNAs being more frequent than Aα- and γ-chain cDNAs ; the latter two occuring at similar frequencies. The higher abundance of Bβ-chain cDNA clones may be a reflection of greater cloning efficiency of this sequence or else higher levels of mRNA for this polypeptide in liver. Unequal amounts of mRNAs for other coordinately regulated genes such as α- and β- globin genes have been observed (Lodish, 1971).

A notable feature of the initial oligonucleotide screening was the high background on filters hybridised

with Aα-oligonucleotides 1 and 2 which are of higher complexity (figure 8a). Under the same conditions other oligonucleotides, which were either unique (α1-antitrypsin) or low mixtures (Bβ- and γ-fibrinogen) gave hybridisation signals detectable over the background (figure 8a). This variation in background intensities was probably due to the presence of sequences in Aα-oligonucleotides 1 and 2 which formed stable duplexes with partially homologous sequences under non-stringent conditions. Increase in stringency of post-hybridisation washes eliminated this problem (figure 8b).

4.2. The Fibrinogen cDNA Clones

4.2.1. Recombinant cDNA Clones for Human Fibrinogen

Partial DNA sequences of pAF1, pBF1 and pGF1 (figures 21a, 21b and 21c) unequivocally confirm their identity as cDNA clones for the Aα-, Bβ- and γ-chains of human fibrinogen. The sequence of pAF1 insert agrees with that of the corresponding region of Aα-fibrinogen cDNA reported by Kant et al., (1983) except for a single nucleotide. At the third base position of the codon for glutamine 137 a change from A to G results in the creation of a PstI site. The presence of this site has been confirmed by PstI digests (figure 12b). The Aα-chain cDNA sequence of Rixon et al., (1983) also reports the presence of this site. It is possible that such a change could represent misreading by reverse transcriptase and DNA polymerase I during cDNA cloning. Systematic errors during enzymatic processes used

in molecular cloning have been shown to result in only about 1 base change in 3700 bases (Feilds and Winter, 1981). It therefore seems likely that the A-G change represents a genuine base transition giving rise to a silent site substitution. The frequency of this putative DNA polymorphism is not known. Putative amino acid polymorphisms for Aα-fibrinogen have also been revealed from the cDNA sequences of Kant et al., (1983) and Chung et al., (1983a).

The sequence of clone pAF1 (figure 21a) also showed a discrepancy between the published protein sequence of the Aα-chain (Doolittle et al., 1979) and that derived from its cDNA. Residue 128 has been assigned glutamic acid rather than glutamine. This is in agreement with the Aα-chain cDNA sequences of Kant et al., (1983) and Rixon et al., (1983). Similar sequence comparisons did not reveal any discrepancies between the published cDNA sequences of Bβ- and γ-fibrinogen, and those determined from sequence analysis of clones pBF1 and pGF1 (figures 21b and 21c).

All 4 cDNA clones lack 5' coding regions. The 3'-bias can be ascribed both to the cloning procedure, which results in a preponderance of cDNA clones containing the 3' region of mRNAs (Craig and Hall, 1982), and the use of oligonucleotides specific to the carboxy-terminus of the polypeptides. The insert of clone pAF1 commences at residue 91 of Aα-fibrinogen, contains the rest of the downstream coding region and the 3' noncoding region including the poly(A) sequence. Similarly, the insert of

pGF1 starts at amino acid 183 of γ-fibrinogen and extends up to and including the poly (A).

Of the two Bβ-chain recombinants analysed, only pBF1 was sequenced and shown to commence at residue 277 of the polypeptide. The extent of downstream sequence represented in this clone is not known. From the insert length (0.72 kb) however, it is likely that this clone includes the entire 3' coding region and also part of the untranslated region. Recombinant pBF5, like pAF1 and pGF1, has however been shown to have a 3' untranslated region as revealed by the presence of a poly(A) sequence (figure 16b). Due however to the 3' heterogeneity of Bβ-fibrinogen mRNA (Chung et al., 1983a), pBF1 and pBF5 may not extend to the same position in the downstream noncoding region.

The starting residue of pBF5 insert can be determined since the internal BamHI site is 630 bp from the 5' terminus (figure 17b). Taking into account 15-20 bases of G-C tails, this represents a coding capacity for 203-205 amino acids. The BamHI site has been shown to span amino acids 266-268 of Bβ-fibrinogen (Kant et al., 1983; Chung et al., 1983a) thus placing the starting codon of pBF5 between residues 61-65 of the polypeptide.

Recombinants pAF1, pBF5 and pGF1 do not have cross-hybridising sequences (figure 14). From the known limited homology of fibrinogen chains at the amino acid level (Doolittle et al., 1979) it appeared unlikely that the corresponding mRNA sequences would have sufficient homology for any cross-hybridisation to occur even under conditions

of low stringency (3xSSC). The result also showed the absence of homologous or partially homologous sequences in the 3' noncoding region.

4.2.2. Sequence Comparison of Human and Rat γ-fibrinogen cDNAs

Amino acid sequence comparisons of homologous proteins was until recently the only means of determining evolutionary divergence between them. With the availability of cDNA and genomic sequences, similar comparative studies can now be performed at the DNA level. These studies have revealed silent nucleotide substitutions which point to mutational changes being more extensive than previously postulated and that genetic drift plays a significant role in evolution.

The 250 bp sequence of pGF1 (figure 21c), representing human γ-fibrinogen, was compared with the corresponding region of rat γ-fibrinogen cDNA sequence (Crabtree and Kant, 1982) by dot matrix analysis (figure 26). A 4 bp match was used to minimise background scatter and the high degree of homology in this region of rat and human γ-fibrinogen mRNAs is evident. Base-by-base comparison revealed a total of 41 base changes, of which 24 (59%) are silent site substitutions and the remaining 17 (41%) cause 12 amino acid changes. This strong homology at the nucleotide level comparable to that at the amino acid level has also been reported for the carboxyl ends of the two cDNAs (Kant *et al.*,1983). Only 58 out of 486 bp (12%)

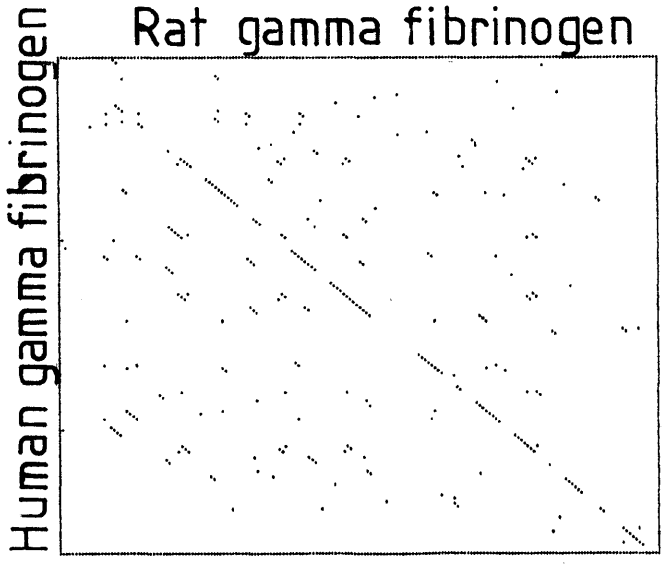

Figure 26. Dot matrix analysis comparing the human γ-fibrinogen cDNA sequence in figure 21c with the corresponding region of rat γ-fibrinogen cDNA (Crabtree and Kant, 1982). Each dot represents a perfect match of 4 bp.

differ in this region, of which 36 (62%) are silent site substitutions and the other 22 (38%) produce 16 amino acid changes. Certain coding regions, especially amino acids 323-338, show an even higher level of homology with only 1 in 20 bp alterations and a single amino acid difference. This is indicative of functional constraints on this portion of the γ-chain which is known to be involved in a number of functions including binding to calcium and platelets (Marguerie et al., 1977 and 1982; Kloczewiak et al., 1982).

The high level of homology extends into the 3'-untranslated region where only 33 out of 187 bp (17.6 %) are different when best alignment is considered. The homology appears preferentially conserved around the hexanucleotide 5' AATAAA 3' which is thought to be the signal sequence for polyadenylation (Proudfoot and Brownlee, 1976) or RNA processing (Montell et al., 1983). Homology in the 3' noncoding region has also been reported for trout protamine mRNAs (Gedamu et al., 1981) and between members of the human β-tubulin gene family (Lee et al., 1983) implying some form of evolutionary constraint on these regions.

In the regions of human and rat γ-fibrinogen cDNAs so far examined there is a greater frequency of silent site substitutions (59-62%) than replacement site substitutions. A similar phenomenon has been observed in a comparison of human and rabbit β-globin mRNA sequences (Kafatos et al., 1977) where of the 48 base substitutions

in the coding region, 32 (67%) were silent. Theoretically, there are 526 possible single base substitutions in the 61 sense codons, of which only 134 (25%) are silent (Jukes and King, 1979). The observed deficit of replacement site substitutions therefore points to selection acting against such substitutions in coding regions. The preferential accumulation of silent nucleotide changes in protein coding sequences has been confirmed by a survey of a wide range of eukaryotic, prokaryotic and viral genes (Jukes, 1980) . These changes are fixed by genetic drift, thus contributing towards evolutionary change.

A closer examination of the nature of base substitutions in the fibrinogen sequences showed nucleotide transitions involving A-G or T-C to be more frequent (76%) than transversions A or G - T or C. This bias in base substitutions has also been noted by Jukes (1980), and will tend to reduce the theoretical maximum rate of sequence divergence. For example, a T residue common between two sequences will preferentially change to C in one sequence at the first substitution. A second substitution at this site will preferentially revert C to T, or else alter the second sequence to C.

The availability of bovine and rat Aα- and Bβ-fibrinogen mRNA sequences would allow similar comparative studies with corresponding human fibrinogen sequences. It seems unlikely that they would show as high a degree of overall evolutionary conservation as that shown by γ-fibrinogen, due primarily to the presence of

fibrinopeptides and also the existence of a hypervariable region in the Aα-chain.

4.3. The Organisation and Expression of Human Fibrinogen Genes

4.3.1. Genomic Analysis and Hepatic Expression

A characteristic feature of the eukaryotic genome is the occurence of multigene families of sequence-related members. Such families have been found both for genes encoding products of terminally differentiated cells (e.g. globins, ovalbumin, immunoglobulins) as well as for genes that are expressed universally (histones, arginosuccinate synthetase, tubulins, actins). To determine whether the fibrinogen gene family also falls into this category, Southern blots of human DNA digested with various restriction enzymes were hybridised with the cDNA clones for fibrinogen polypeptides (figure 22). Each clone hybridised to give simple band patterns even at low stringency (3xSSC), implying that the corresponding genomic sequences do not constitute multigene families consisting either of related functional genes or pseudogenes. These observations have been confirmed by the work of two other groups. Chung et al., (1983a) have reported the isolation of the gene for human Bβ-fibrinogen which occurs once per haploid genome. Similarly, unique genomic sequences for rat Aα- and γ-fibrinogen have been reported (Kant and Crabtree, 1983).

The absence of multiple copies of fibrinogen genes for each of the polypeptides has a bearing on the earlier

suggestions that separate genes might code for adult and fetal Aα-fibrinogen (Witt and Tesch, 1979) and also for platelet derived and hepatic γ-fibrinogen (Jandrot-Perrus et al., 1979). These possibilities have now to be discounted. It could however be argued that these tissue-specific differences are the result of rearranged genes or that developmental or tissue specific differential splicing or transcription of unrearranged genomic sequences may be involved. Young et al., (1981) have demonstrated that the α-amylase mRNAs which accumulate in two different tissues of the mouse, namely the salivary gland and the liver, are transcribed from the same gene (Amy 1^{A}). Gross rearrangement of the gene does not account for the tissue-specific pattern of expression. Instead, the two mRNAs are differentially transcribed or processed from the same genomic sequence.

The hepatic expression of human fibrinogen genes was examined by Northern blot hybridisation analysis of RNA prepared from human liver (figure 23). The Bβ- and γ-chains each appear to be coded by a single mRNA sequence. Two mRNA sequences were observed with the Aα-chain probe. There was a major band at 2.6 kb and a minor band at about 1.8 kb (figure 23, lane A). The latter is not long enough to code for the normal Aα-fibrinogen and as yet no heterogeneity for serum Aα-chain has been reported. It is also unlikely that the minor sequence is either a Bβ-chain or a γ-chain cross-hybridising sequence as the three recombinants do not have sequences that cross-hybridise

even at low stringency (figure 14b). Hetrogeneity has been shown to exist at the 3'-untranslated region of Aα-fibrinogen mRNAs. This was reflected in the isolation by Rixon et al.(1983), of Aα-fibrinogen cDNAs with variable 3'-end termination sites. One clone had a poly(A) sequence 17 bp downstream from the other. Clearly, this rather small difference between the lengths of the two mRNAs cannot account for the existence of the minor band observed in this study.

Variation at the 3'-non coding region of human Bβ-fibrinogen is more pronounced. Almost 2/3 of all Bβ-chain cDNAs isolated by Chung et al.(1983), had short 3'-noncoding regions which varied in length from 98 bases to 167 bases, while the remainder had a 431 base long 3'-noncoding region. Sequences of the shorter regions were colinear with the latter and also colinear wih the 3'-end of the gene, suggesting variable transcription termination as the cause of this heterogeneity. The Northern hybridisation data of Chung et al., (1983a) was consistent with their findings at the cDNA level. Two distinct mRNAs were observed, a major band 1.6 kb long and a minor band at 1.85 kb position representing the 431 bp 3'-noncoding sequence was observed. In contrast to this observation, only a single mRNA sequence 1.85 kb long was detected during this investigation (figure 23, lane B). However, the Northern blot data presented here is for liver RNA prepared from a single individual and the absence of a second band at 1.6 kb position might reflect variation of

Bβ-chain mRNA 3'-heterogeneity between individuals.

Size heterogeneity in the mRNA 3' noncoding region of mouse dihydrofolate reductase (Setzer et al., 1980 and 1982) mouse α-amylase (Tosi et al., 1981) and human β-tubulin (Lee et al., 1983) has been observed and is thought to be due to read through of the 5' AATAAA 3' sequence. This seems to be a common feature of many eukaryotic genes (Heilig et al., 1980; Lewis et al., 1981; Unterman, 1981), and as yet no functional differences have been discerned between RNAs possessing this type of heterogeneity.

Human γ-fibrinogen heterogeneity exists within the coding region giving rise to the γ'-variant with an extended C-terminus (Wolfenstein-Todel, 1980). The amino acid sequence of this region has been determined and is thought to arise by differential RNA splicing (Wolfensten-Todel, 1981). Another γ-chain variant with a possible caboxy-terminal heterogeneity has been reported recently (Francis et al., 1983). The estimated abundance of this variant is 3%, while that of the γ'-variant is 10% of the total γ-fibrinogen synthesised in human liver. Under the conditions utilised for Northern blot hybridisations, mRNAs corresponding to these low abundance variants cannot be detected and only a single, 1.8 kb long, mRNA was obsreved using the γ-chain probe.

4.3.2. The Fibrinogen Chromosomal Locus: Gene Synteny and Evolutionary Conservation of Linkage Groups

There has been much interest in the genomic organisation of sequences with related function. In a number of procaryotes, genes coding for enzymes involved in sequential steps of a metabolic pathyway are often closely linked to form a gene cluster - an operon- which undergoes coordinate gene induction or repression. The lac, gal,ara, arg, his and trp operons of E.coli, each comprise several sets of genes which individually code for enzymes sequentially catalyzing a specific stage of their respective metabolic pathways. Moving up the evolutionary ladder this linkage relationship seems not to exist. Yeast does not show linkage of genes with related function. On the other hand such a linkage relationship is observed in higher eukayotes where gene clusters have arisen by tandem duplication.

The term 'synteny' (Renwick, 1971) applies to the occurrence of two or more genes on a single chromosome in eukaryotes. Spandidos and Siminovitch (1977) provided the first indication that some genes encoding enzymes of a common metabolic pathway in eukaryotes are syntenic. The evidence demonstrated the existence of genes for folate metabolising enzymes on a single chromosome in CHO cells. Since then, a number of mammalian gene families have been shown to exist as large complexes on a single chromosome. These include the major histocompatability antigen genes in mouse and man (Klein, 1975; Festenstein and Demant, 1978),

human growth hormone and chorionic somatotropin genes (Barsh et al., 1973), murine urinary protein genes (Bennett et al., 1982) and the mouse kallikrien gene family (Mason et al., 1983). However, a priori, no assumption can be made of synteny of genes that are either coordinately expressed or encode a heteromeric protein. Mammalian α- and β-globin genes exist on small clusters present on separate chromosomes (Fritsch et al., 1980). Similarly, other functionally related genes are dispersed throughout the genome. Drosophila actin genes are located at multiple loci on different chromosomes (Fyrberg et al., 1980) as are chicken tubulin genes (Cleveland et al.,1981). It was therefore of interest to see if all three genes for human fibrinogen were syntenic or indeed very close to each other. Regional assignment of these genes, performed by Southern blot hybridisation of DNAs from somatic cell hybrids constructed from a human lymphoblastoid cell line with a chromosome 4-12 translocation, showed that the Aα-, Bβ- and γ-fibrinogen genes are present on human chromosome 4q26-qter (figures 25a and 25b). Olaisen et al., (1982) assigned the γ-fibrinogen gene to human chromosome 4 long arm by linkage to the MNS locus (4q29-q31) and in situ hybridisations performed in this department have shown that the gene for Aα-fibrinogen is also near this locus (T. Robbins, personal communication). As human Aα- and Bβ-fibrinogen genes are physically linked (J. Kant, personal communication), it appears that all three genes of human fibrinogen are within a small region of the chromosome.

The existence of linkage groups and their evolutionary conservation are important features of genome organisation and evolution in higher eukaryotes. Comparative studies of the nature and organisation of genetic material in different mammalian species throws light on questions concerning the importance of gene duplication, the evidence for tetraploidisation and the stability of particular chromosomal segments over an evolutionary timescale. These studies also reveal the nature of chromosomal changes occuring most frequently and the close linkage of functionally related loci during mammalian evolution. Lundin (1979) has discussed two separate types of homology based on comparative studies of mammalian genetics. That which exists within a species when a common ancestral gene undergoes duplication giving rise to "paralogous genes", and that which exists between different species when they are said to have "orthologous genes" with similar properties and showing discent from the same ancestor. The former may either come about as a result of genomic doubling involving the whole genome (tetraploidisation) or just small segments or individual loci. The level of amino acid sequecne homology between the three polypeptides of human fibrinogen led to the suggestion that the fibrinogen locus has arisen by duplication and divergence of a common ancestral sequence (Doolittle et al., 1979). The extent of these duplications may be greater than that perceived so far. It has been tentatively suggested recently that κ-caesins have evolved from mammalian γ-fibrinogen (Jolles

and Henschen, 1982). If this is indeed the case, it would be interesting to see whether the gene for κ-caesin occurs in tandem to γ-fibrinogen gene or at least shows synteny with the fibrinogen locus, forming an extended linkage group.

The conservation of chromosome segments and linkage groups during mammalian evolution has been revealed by genetic comparisons of man and mouse. Ohno (1967) proposed the conservation of the mammalian X-chromosome. Although genetic rearrangements affecting this chromosome during mammalian evolution have been reported (Franke and Taggart, 1979), gene synteny seems to have been maintained. Similar conservation of linkage groups and chromosomal segments has now been shown to exist for the autosomes. Of particular interest are the findings that five loci on the short arm of human chromosome 1 are also syntenic in the mouse (Lalley et al., 1978) and there are altogether ten instances involving eight chromosomes in which two or more loci syntenic in the mouse are also syntenic in man (Searle, 1981). Such a conservation of gene synteny might also exist for the fibrinogen locus. The three human fibrinogen genes occur in a very small region of chromosome 4 with human Aα- and Bβ-fibrinogen genes being physically linked. In a similar manner rat Aα- and γ-fibrinogen genes have been shown to be linked and rat Bβ-fibrinogen may also be on the same chromosome as the Aα-γ gene cluster (J. Kant, personal communication). It thus appears that although gene linkage relationships at this locus have been

altered between rat and man by some form of DNA rearrangement, synteny of all three genes has possibly been maintained. Ohno (1973) suggested that gene linkages have been retained essentially by chance during evolution because of the very slow rate at which rearrangements, such as inversions or translocations are fixed during evolution. It is also possible that gene linkage homologies are an indication of some selective advantage (Bodmer, 1975) considering the extent of genomic rearrangemnts observed during mammalian evolution. There may indeed be a selective advantage for all the genes of an acute phase protein like fibrinogen to be on the same DNA segment and under the same localised conditions of gene expression. Conservation of gene synteny between functionally unrelated markers, such as myosin heavy chain genes being syntenic with thymidine kinase and galactokinase in human and rodent genomes (Leinwand et al., 1983) though implies that this selection may not necessarily be at the phenotypic level. Further speculation as to how evolutionary conservation of chromosomal segments might operate would require more information on the role of intergenic DNA, the rates of DNA rearrangements and their fixation.

5. Perspectives

The availability of cloned human Aα-, Bβ- and γ-fibrinogen cDNA sequences will allow further investigation into the genetics and molecular biology of the fibrinogen chromosmal locus. It should be possible to isolate cosmid

clones of human fibrinogen genes and establish physical linkage of all three fibrinogen genes by genome walking. Human Aα- and Bβ-fibrinogen genes have already been shown to be within a few kilobases of each other by physical mapping (J. Kant, personal communication). Physical linkage of γ-fibrinogen gene to the Aα-Bβ gene cluster has not been demonstrated due primarily to the occurence of highly repetitive sequences in the intergenic DNA. Overlapping genomic clones from a cosmid library will improve the chances of establishing linkage between the two regions. Physical linkage together with restriction mapping of the genomic clones would determine whether all three genes are in the same orientation forming a closely linked set of transcriptional units, as shown for rat Aα- and γ-fibrinogen genes (Kant and Crabtree, 1983) and for the coordinately expressed B2 high-sulphur keratin genes of sheep (Powell et al., 1983).

Restriction mapping and sequencing of fibrinogen genomic clones will also identify differences between normal and certain variant human fibrinogens. Of particular interest would be the sequence of the γ-fibrinogen gene as it would shed light on the origin of the expression of the normal γ^{1}-chain variant. This sequence information may also establish the molecular basis of γ-fibrinogen Paris I variant, which like the γ^{1}-chain variant has an additonal 20 amino acids at the carboxy-terminal end, but differs from the latter in not functioning normally in clotting and cross-linking

reactions.

Fibrinogen genes are coordintely expressed and synthesis of the protein is influenced by a number of factors including stress, injury and adrenal hormones (Majumdar et al., 1967; Bouma et al., 1975; Beuving and Vonder, 1978; Greininger et al., 1978). There is much speculation as to how coordinate regulation of genes is achieved. The proposition that a set of genes could be coordinately induced by means of cis interactions with a common regulatory molecule, as put forward by Britten and Davidson (1969) and later elaborated by Davidson and Britten (1979), assumes that some form of homologous sequences exist in the vicinity of functionally related genes which are induced in concert. Sequencing of 5' upstream region of Aα-, Bβ- and γ-fibrinogen genes might reveal the presence of such a sequence. Repeat sequences common to a set of coordinately expressed genes have been observed in a number of gene families. These include yeast his genes (Donahue et al., 1982; Struhl et al., 1982; Donahue et al., 1983), Drosophila heat shock genes (Holmgren et al., 1981; Ingolia and Craig, 1981; Pelham, 1982) and silk moth chorion genes (Jones and Kafatos, 1980). Genes regulated by glucocorticoids, such as human proopiomelanocortin (Cochet et al., 1982), mouse mammary tumour virus (Donehower et al., 1981), and rat growth hormone (Barta et al., 1981) have been shown to share a 24 bp consensus repeat sequence element at their 5' end (Davidson et al., 1983). It is conceivable that there

might be a similar sequence 5' upstream to fibrinogen genes as expression of these genes is regulated by glucocorticoids.

A longer term investigation could be based on "surrogate" or reverse genetics involving site-directed mutagenesis. This could either focus on the 5' region of the genes to look for regions involved in gene regulation and expression or else on exon sequences, revealing functionally important domains of the protein.

Finally, population studies have revealed that high levels of fibrinogen are correlated with increased risk of ischaemic heart disease (Mead and North, 1977; Mead et al., 1980). As yet there is no evidence that elevated levels of fibrinogen are genetically determined. Assuming that there is a genetic basis to increased plasma levels of fibrinogen, DNA polymorphisms detected with any of the three probes could be used to look for a possible linkage with the disease thus providing a genetic marker. Additonally, these polymorphisms will serve as markers for the construction of a genetic linkage map of human chromosome 4.

6. REFERENCES

Agarwal, K.L., Bucki, H., Caruthers, M.H., Gupta, N., Khorana, H.G., Kleepe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U.L., van de Somde, J.H., Sgaramella, V., Weber, H. and Yamada, T. (1970) Nature 227, 27-37.

Al-Mondhiry, H.A.B., Bilezikian, S.B. and Nossel, H.L. (1975) Blood 45, 607-619.

Anderson, J.N. and Schimke, R.T. (1976) Cell 7, 331-338.

Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303-314.

Axel, R., Fiegelson, P., Schultz, G. (1976) Cell 7, 247-254.

Aznar, J., Fernandes-Pavon, A., Reganon, E., Vila, V. and Arellna, F. (1974) Thrombosis et Diathesi Haemorrhagica 32, 564-577.

Bahl, C.P., Marians, K.J., Wu, R., Stawinsky, J. and Narang, S.A. (1976) Gene 1, 81-92.

Bailey, K., Astbury, W.T. and Rudall, K.M. (1943) Nature 151, 716-717.

Barsh, G.S., Seeburg, P.H. and Gelinas, R.E. (1983) Nucleic Acids Res. 11, 3939-3958.

Barta, A., Richards, T.I., Baxter, J.D. and Shine, J. (1981) Proc. Natl. Acad. Sci. USA. 78, 4867-4871.

Bennett, K.L., Lalley, P.A., Barth, R.K. and Hastie, N.D. (1982) Proc. Natl. Acad. Sci. USA. 79 1220-1224.

Beuving, G. and Vonder, G.M.A. (1978) Gen. Comp. Endocrinol. 35, 153-159

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

Blomback, M., Blomback, B., Mammen, E. and Prasad, A.S. (1968) Nature 218, 134-137.

Blomback, B., Hessel, B. and Hogg, D. (1976) Thromb. Res. 8, 639-658.

Blomback, B., Hessel, B., Hogg, D. and Thirkildsen, L. (1978) Nature 275 501-505.

Bloom, A.L. (1981) in Haemostasis and Thrombosis, Bloom, A.L.and Thomas, D.P. eds., pp 321-370, Churchill Livingstone, London.

Bodmer W.F. (1975) in Chromosome Variation in Human Evolution (Ed. A.J. Boyer) 53-61. Symposia of Society for Study of Human Biology. Vol. XIV. Taylor and Francis London.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Am. J. Hum. Genet. 32, 314-331.

Bouma, H., Kwan, S-W. and Fuller G.M. (1975) Biochemistry 14, 4787-4792.

Breathnack, R., Bensist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA. 75, 4853-5857.

Breslow, J.L., Ross, D., McPherson, J., Williams, H., Kurmit, D., Nussbaum, A.L., Karathanasis, S.K. and Zannis, V.I. (1982a) Proc. Natl. Acad. Sci. USA. 79, 6861-6865.

Breslow, J.L., McPherson, J., Nussbaum, A.L., Karathansis, S.K. and Zannis, V.I. (1982b) J. Biol. Chem. 257, 14639-14641.

Brevkink, H.J., Hart, M.C. and von Arkel, C. (1972) Zentralblatt fur Veterinarmedizin 19, 661-676.

Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357.

Broom, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA. 75, 2746-2749.

Budzynski, A.Z., Marder, V.J., Menache, D. and Guillin, M-C. (1974) Nature 252, 66-68.

Carroll, M.C.and Porter, R.R. (1983) Proc. Natl. Acad. Sci. USA. 80, 264-267.

Chan, S.J., Noyse, B.E., Agarwal, K.L. and Steiner, D.F. (1979) Proc. Natl. Acad. Sci. USA. 76, 5036-5040.

Chen, R. and Doolittle, R.F. (1970) Proc. Natl. Acad. Sci. USA. 66, 472-479.

Chen, R. and Doolittle, R.F. (1971) Biochemistry 10, 4486-4491.

Choo, K.H., Gould, K.G., Rees, D.J.G. and Brownlee, G.G. (1982) Nature 299, 178-180.

Chu, M-L., Williams, C.J., Pepe, G., Hirsch, J.L., Prockop, D.J. and Ramirez, F. (1983) Nature 304, 78-80.

Chung, D.W., MacGillivray, R.T.A. and Davie, E.W. (1980) Ann. N.Y. Acad. Sci. 343, 210-217.

Chung, D.W., Rixon, M.W., MacGillivray, R.T.A. and Davie, E.W.(1981) Proc. Natl. Acad. Sci. USA. 78, 1466-1470.

Chung, D.W., Rixon, M.W. and Davie, E.W. (1982) in Proteins in Biology and Medicine (Bradshaw, R.A., Hill, R.L., Tang, J., Liang, C.C., Tsao, T.C. and Tso, C.L. eds.) pp 309-328, Academic Press, New York.

Chung, D.W., Que, B.G., Rixon, M.W., Mace, Jr., M. and Davie, E.W. (1983a) Biochemistry 22, 3244-3250.

Chung, D.W., Chan, W-Y. and Davie, E.W. (1983b) Biochemistry 22, 3250-3256.

Clarke, L., Hitzeman, R. and Carbon, J. (1979) Methods in Enzymology 68, 436-442.

Cleveland, D.H., Hughes, S., Stubblefield, E., Kirschner, M. and Varmus, H. (1981) J. Biol. Chem. 256, 3130-3134

Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982) Nature 297, 335-339.

Colten, H.R. and Parkman, R. (1972) Science 176, 1029-1031.

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, B.R. (1983) Proc. Natl. Acad. Sci. USA. 80, 278-282.

Crabtree, G.R. and Kant, J.A. (1981) J. Biol. Chem. 256, 9718-9723.

Crabtree, G.R. and Kant, J.A. (1982) Cell 31, 159-166.

Craig, R.K. and Hall, L. (1982) in Genetic Engineering Vol. 4, Ed. R. Williamson (pp 58-125) Academic Press, London.

Crea, R., Kraszewsi, A., Hirose, T. and Itakura, K. (1978) Proc. Natl. Acad. Sci. USA. 75, 5765-5769.

Creagan, R.P. and Ruddle, F. (1975) The Clone Panel. In Rotterdam Conferance, 1974; 2nd Int. Workshop on Human Gene Mapping. Original Article Series, Bergsma, D.(ed.), pp. 109-111.

Croce, C., Girardi, A.J. and Koprowski, H. (1973) Proc. Natl. Acad. Sc. USA. 70, 3617-3620.

Croce, C. (1976) Proc. Natl. Acad. Sci. USA. 73, 3248-3252.

Dalbadie-McFarland, G., Cohen, L.W., Riggs, A.D., Morin, C., Itakura, K. and Richards, J.H. (1982) Proc. Natl. Acad. Sci. USA. 79, 6409-5413.

Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.

Davidson, E.H., Jacobs, H.T. and Britten, R.J. (1983) Nature 301, 468-470.

Dayhoff, M.O.(1972) Atlas of Protein Sequence and Structure. Vol. 5. National Biomedical Research Foundation, Silver Spring, Maryland, USA.

Deissoroth, A., Niehius, A., Turner, P., Velz, R., Anderson, W., Ruddle, F., Laurence, J., Creagan, R. and Kucherlapati, R. (1977) Cell 12, 205-208.

Deissoroth, A., Neihius, A., Giles D., Laurence, J., Turner, P., and Ruddle, F.H. (1978) Proc. Natl. Acad. Sci. USA. 75, 1456-1460.

Donehower, L.A., Huang, A.L. and Hager G.L. (1981) J. Virol. 37, 226-238.

Donahue, T.F., Farabough, P.J. and Fink, G.R. (1982) Gene 18, 47-59.

Donahue, T.F., Davis, R.S., Lucchini, G. and Fink, G.R. (1983) Cell 32, 89-98.

Donovan, J.W. and Mihalyi (1974) Proc. Natl. Acad. Sci. USA. 71, 4125-4128.

Doolittle, R.F. and Blomback, B. (1964) Nature 202, 147-152.

Doolittle, R.F. (1965) Biochem J. 94, 735-741.

Doolittle, R.F. (1973) Adv. Prot. Chem. 27, 1-109.

Doolittle, R.F. and Wooding, G.L. (1974) Biochem. Biophys. Act. 271, 277-282.

Doolittle, R.F. (1976) Fed. Proc. 35, 2145-2149.

Doolittle, R.F., Goldbaum, D.M. and Doolittle, L.R. (1978) J. Mol. Biol. 120, 311-325.

Doolittle, R.F., Watt, K.W.K., Cottrell, B.A., Strong, D.D. and Riley, M. (1979) Nature 280, 464-468.

Doolittle, R.F. (1980) in Protides of Biological Fluids, Proceedings of the 28th Colloquium (Peters, M. ed.), pp 41-46, Pergamon Press, New York.

Doolittle, R.F. (1981) in Haemostasis and Thrombosis; Bloom, A.L. and Thomas, D.P. eds, pp 163-191.

Douglas, G.R., McAlpine, P.J. and Hammerton, J.L. (1973) Proc. Natl. Acad. Sci. USA. 70, 2737-2740.

Edge, M.D., Greene, A.R., Heathcliffe, G.R., Meacock, P.A., Schuch, W., Scanlon, D.B., Atkinson, T.C., Newton, C.R. and Markham, A.F.(1981) Nature 292, 756-761.

Edge, M.D., Green, A.R., Heathcliffe, G.R., Morre, V.E., Faulkner, N.J., Camble, R., Potter, N.N., Trueman, P., Schuch, W., Henman, J., Atkinson, T.C., Newton, C.R. and Marham, A.F. (1983) Nucl. Acids Res. 11, 6419-6435.

Egeberg, O. (1967) Thrombos. Diathes, Haemorrh. (Stuttz) 17, 176-187.

Eiferman, F.A., Young, P.R., Scott, R.W. and Tilghman, S.M. (1981) Nature 294, 713-718.

Feilds, S. and Winter, G. (1981) Gene 15, 207-214.

Ferry, J.D. (1952) Proc. Natl. Acad. Sci. USA. 38, 566-569.

Festenstein, H. and Demant, P. (1978) HLA and H2:Basic Immunogenetics, Biology and Clinical Relevance (Edward Arnold, London).

Flute, P.T. (1977) Br. Med. Bull. 33, 253-259.

Folk, J.E. and Finlayson, J.S. (1977) Adv. Prot. Chem. 31, 1-133.

Forman, W.B. and Barnhart, M.I. (1964) J. Am. Med. Assoc. 187, 128-132.

Fournier, R.E.K. and Ruddle, F. (1977) Proc. Natl. Acad. Sci. USA 74, 319-323.

Fowler, W.E. and Erickson, H.P. (1979) J. Mol. Biol. 134, 241-249.

Francis, C.W., Kraus, D.H. and Marder, V.J. (1983) Biochim. Biophys. Acta 744, 155-164.

Fritsch, E., Lawn, R., Maniatis, T. (1980) Cell 19, 959-972.

Fuller, G.M. and Doolittle, R.F. (1971) Biochemistry 10, 1305-1311.

Fyrberg, E.A., Kindle, K., Davidson, N. and Sodja, A. (1980) Cell 19, 365-378.

Gait, M.J., Singh, M.and Sheppard, R.C. (1980) Nucl. Acids Res. 8, 1081-1096.

Gall, J. and Pardue, M.L. (1969) Proc. Natl. Acad. Sci. USA. 63, 378-383.

Gedamu, L., Wosnick, M.A., Connor, W., Watson, D.C. and Dixon, G.H. (1981) Nucl. Acids Res. 9, 1463-1482.

Giannelli, F., Choo, K.H., Rees, D.J.G., Boyd, Y., Rizza, C.R. and Brownlee, G.G. (1983) Nature 303, 181-182.

Gilbert, W. (1982) Nature 271, 501.

Gillam, S., Jahnke, P. and Smith, M. (1978) J. Biol. Chem. 253, 2532.

Gillam, S., Jahnke, P., Astell, C., Phillips, C., Hutchinson, A. and Smith, M. (1979) Nucl. Acids Res. 6, 2973.

Gillam, S., Astell, C.R. and Smith, M. (1980) Gene 12, 129-137.

Goeddel, D.V., Kleid, D., Bolivar, F., Heyneker, H.L., Yansuma, D.A., Crea, R., Hirose, T., Krazenski, A., Itakura, K. and Riggs, A. (1979a) Proc. Natl. Acad. Sci. USA. 76, 106-110.

Goeddel, D.V., Heyneker, H.L., Hozami, T., Arentzen, R., Itakura, K., Yansuna, D.G., Ross, M.J., Miozzain, G., Crea, R. (1979b) Nature 281, 544-548.

Goeddel, D.V., Shepard, H.M., Yelverton, E., Leung, D.W., Sloma, A., Pestka, S. and Crea, R. (1980) Nucl. Acids Res. 8, 4057-4074.

Gralnick, H.R., Givelber, H.M. and Finlayson, J.S. (1973) Thrombosis et Diathesis Haemorrhagica 29, 562-571.

Gralnick, H.R., Coller, B.S., Fratanoti, J.C. and Martinez, J. (1979) Blood 53, 28-46.

Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980) Nucl. Acids Res. 8, r47- r62.

Greininger, G., Hertzberg, K.M. and Pindyck, J. (1978) Proc. Natl. Acad. Sci. USA. 75, 5506-5510.

Gusella, J.F., Keys, C., Varsanyi-Breine, A., Kao, F.T., Jones, C., Puck, T.T. and Housman, D. (1980) Proc. Natl. Acad. Sci. USA. 77, 2829-2833.

Gusella, J.R., Jones, C., Kao, F.T., Housman, D. and Puck, T.T. (1982) Proc. Natl. Acad. Sci. USA. 79, 7804-7808.

Hall, C.F.and Slayter, H.S. (1959) J. Biophys. Biochem. Cytol. 5, 11-16.

Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.

Harper, M.E. and Sanders, G.F. (1981) Chromosoma 83, 431-439.

Hastie, N.D. and Bishop, J.O. (1976) Cell 9, 761-774.

Helfman, D.M., Feramisco, J.R., Fiddes, J.C., Paul, T.G. and Hughes, S. (1983) Proc. Natl. Acad. Sci. USA. 80, 31-35.

Henschen, A. and Lottspeich, F. (1977) Thromb. Res. 11, 869-880.

Henschen, A., Lottspeich, F., Topfer-Petersen, E., Kehl, M. and Timpl, R. (1980) in Protides of Biological Fluids (Petes, H. ed.) pp 47-50. Pergamon Press, New York.

Heilig, R., Perrin, F., Gannon, F., Mandel, J.L. and Chambon, P. (1980) Cell 20, 625-637.

Heyneker, H.L., Shine, J., Goodman, H.M., Boyer, H.W., Rosenberg, J., Dickerson, R., Narang, S.A., Itakura, K., Lin, S-Y and Riggs, A.D. (1976) Nature 263, 748-752.

Holmgren, R., Cources, V., Morimoto, R., Blackcman, R. and Meselson, M. (1981) Proc. Natl. Acad. Sci. USA. 78, 3775-3778.

Houghton, M., Stewart, A.G., Doel, S.M., Emtage, J.S., Eaton, M.A.W., Smith, J.C., Patel, T.P., Lewis, H.M., Porter, A.G., Birch, J.R., Cartwright, T. and Carey, N.H. (1980) Nucl. Acids Res. 8, 1913-1931.

Hurdy-Clergeon, G., Marguerie, G., Pouit, L. and Suscillon, M. (1975) Thrombosis Res. 6, 533-541.

Ingolia, T.D. and Craig, E.A. (1981) Nucl. Acids Res. 9, 1627-1642.

Itakura, K., Katagin, N., Narang, S.A., Bahl, C.P., Mairans, K.J. and Su, R. (1975) J. Biol. Chem. 250, 4592-4600.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. and Boyer, H.W. (1977) Science 198, 1056-1063.

Jackson, D.P. and Beck, E.A. (1970) Inherited Abnormal Fibrinogens, in Brinkhouse, K.M. (ed.) Haemophilia and new haemorrhagic states. Chapel Hill:University of North Carolina Press, pp225-232.

Jandrot-Perrus, M., Mosesson, M.W., Deninger, M.H. and Menache, D. (1979) Thrombosis and Haemostasis 42, 78 (abstract).

Jolles, P. and Henschen, A. (1982) Trends in Biochem. Sci. (TIBS) 81, 325-328.

Jones, C.W. and Kafatos, F.C. (1980) Cell 22, 855-867.

Jukes, T.H. and King, J.L. (1979) Nature 281, 605-606.

Jukes, T.H. (1980) Science 210, 973-978.

Kafatos, F.C., Efstratiadis, A., Forget, B.G. and Weissman, S.M. (1977) Proc. Natl. Acad. Sci. USA. 74, 5618-5622.

Kammerman, B., Gemur, J. and Stunzi, H. (1971) Zentralblatt fur Veterinarmedizin 18, 192-205.

Kan, Y.W. and Dozy, A.M. (1978) Proc. Natl. Acad. Sci. USA. 75, 5631-5635.

Kant, J.A., Lord, S.T. and Crabtree, G.R. (1983) Proc. Natl. Acad. Sci. USA. 80, 3953-3957.

Kant, J.A. and Crabtree, G.R. (1983) J. Biol. Chem. 258, 4666-4667.

Khorana, H.G., Agarwal, K.L., Bucki, K.L., Caruthers, M.H., Gupta, N.K., Kleepe, K., Kumar, A., Ohtsuka E., Raj Bhandry, V.l., Van de Saud, J.H., Sgaramella, V., Terao, T., Weber, H. and Yamada, T. (1972) J. Mol. Biol. 72, 209-221.

Khorana, H.G. (1979) Science 203, 614-625.

Kidd, V.J., Wallace, B.R., Itakura, K. and Woo, S.L.C. (1983) Nature 304, 230-234.

Klein, J. (1975) Biology of the Mouse Histocompatibility - 2 Complex (Spinger-Verlag, Berlin).

Klobutcher, L.A. and Ruddle, F.H. (1981) Ann. Rev. Biochem. 50, 533-554.

Kloczewiak, M., Timmons, S. and Hawinger, J. (1982) Biochem. Biophys. Res. Commun. 107, 181-187.

Koppel, G. (1966) Nature 212, 1608-1609.

Kramer, W., Schughart, K. and Fritz, H-J. (1982) Nucl. Acids Res. 10, 6475-6500.

Kurdyk, B., Blomback, B. and Blomback, M. (1976) Thromb. Res. 9, 25-36.

Kurdyk, B., Blomback, B., Yu, S., Redman, C. and Goldstein, J. (1979) J. Cell Biology 83, 430a (abstract).

Kurdyk, B., Okada, M., Redman, C.M. and Blomback, B.(1982) Eur. J. Biochem. 125, 673-682.

Kusano, J., Long, C. and Green, H. (1971) Proc. Natl. Acad. Sci. USA. 68, 82-86.

Lalley, P.A., Minna, J.D. and Fracnke, U. (1978) Nature 274, 160-163.

Laski, F.A., Belagage, R., Rajbhandary, U.L. and Sharp, P. (1982) Proc. Natl. Acad. Sci. USA. 79, 5813-5817.

Lathe, R.F., Lecocq, J.P. and Everett, R. (1983) in Genetic Engineering Vol. 4. ed. R. Williamson, Academic Press Ltd., London, pp 1-56.

Lee, M.G-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Cell 33, 477-487.

Leinwand, L.A., Fournier, R.E.K., Nadal-Finard, B. and Shows, T.B. (1983) Science 221, 766-768.

Lewin, B. (1980) Cell 22, 324-326.

Lewis, J.A., Kurtz, D.T. and Malera, P.W. (1981) Nucl. Acids Res. 9, 1311-1321.

Lindermain, W., Hauser, H., deWilke, I.G. and Schutz, G. (1982) Nucl. Acids Res. 10, 1243-1256.

Little P.F.R., Annison, G., Darling, S., Williamson, R., Camba, L. and Modell, B. (1980) Nature 285, 144-147.

Lodish, H.F. (1971) J. Biol. Chem. 246, 7131-7138.

Lottspeich, R. and Henschen, A. (1977) Hoppe-Seyle's Z. Physiol. Chem. 358, 935-938.

Lowy, I., Pellicer A., Jackson, J., Simm, G.K., Silverstein, S. and Axel, R. (1980) Cell 22, 817-823.

Lund, T., Grosveld, F.G. and Flavell, R. (1982) Proc. Natl. Acad. Sci. USA. 79, 520-524.

Lundin, L.G. (1979) Clin. Genet. 16, 72-81.

McBride, O.W. and Ozer, H.L. (1973) Proc. Natl. Acad. Sci. USA 70, 1258-1262.

McGillivray, R.T.A., Friezner-Degen, S.J., Chandra, T., Woo , S.L.C., Davie and Davie, E.W. (1980) Proc. Natl. Acad. Sci. USA. 77, 5153-5157

McKee, P.A., Rogers, L.A., Marler, E. and Hill, R.L. (1966) Arch. Biochem. Biophys. 116, 271-279.

Majumdar C., Tsukada, K. and Liebeman, I.(1967) J. Biol. Chem. 242, 700-704.

Maniatis, T., Kee, S.G., Efstradiatis, A. and Kafatos, F.C. (1976) Cell 8, 163-182.

Maniatis, T., Fritsch, E.F. and Lauer, J. (1980) Ann. Rev. Genetics 14, 145-178.

Mammen, E.F. (1974) Seminars in Thrombosis and Haemostasis 1, 184-201.

Marder, V.J. (1970) Thrombosis Diathesis Haemorrhagica Supplement 39, 187-195.

Marguerie, G., Chagnel, G. an Suscllon, M. (1977) Biochem. Biophys. Act. 490, 94-103.

Marguerie, G.A., Ardaillou, N., Cherel, G. and Plow, E.F. (1982) J. Biol. Chem. 257, 11872-11875.

Markham, A.F., Edge, M.D., Atkinson, T.C., Greene, A.R., Heathcliffe, G.R., Newton, C.R. and Scanlon, D. (1980) Nucl. Acids Res. 8, 5193-5205.

Martinez, J., Holburn, R.R., Shapiro, S.S. and Erslev, A.J. (1974) Journal of Clinical Investigation 53, 600-611.

Mason, A.J., Evans, B.A., Cox, D.R., Shine, J. and Richards, R.I. (1983) Nature 303, 300-307.

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

Mead, T.W. and North, W.R.S. (1977) Brit. Med. Bull. 33, 283-288.

Mead, T.W., North, W.R.S., Chakrabarty, R., Stirling, Y., Haines, A.P. and Thompson, S.G. (1980) Lancet 1, 1050-1053.

Menache, D. (1964) Thromb. Diath. Haemorrh. 13, 173-185.

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

Mihalyi, E. (1950) Act. Chem. Scand. 4, 351-358.

Mihalyi, E., Weinberg, R.M., Towne, D.W. and Freidman, M.E. (1976) Biochemitry 15, 5372-5381.

Miller, O.J., Cook, P.R., Khan, P.M., Shin, S. and Siniscalo, M. (1971) Proc. Natl. Acad. Sci. USA 68, 116-120.

Miyada, C.G., Soberon, X., Itakura, K. and Wilcox, G. (1982) Gene 17, 167-177.

Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) Nature 305, 600-605

Montgomery, B.l., Hall, B.D., Gillam, S. and Smith, M. (1978) Cell 14, 673-680.

Mosesson, M.O., Escaig, J. and Feldmann, G. (1979) Brit. J. Haem. 43, 469-477.

Mount, S. and Steitz, J. (1983) Nature 303, 380-381.

Murray, J.M., Davies, K.E., Harper, P.S., Meredith, L., Muller, C.R., Goodfellow, P.N. and Williamson, R. (1982) Nature 300 69-71.

Murtaugh, P.A. and Gladner, J.A. (1981) Fed. Proc. 40, 3, part 1, p. 778 (Abstr. 3114).

Nabholz, M., Miggiano, V.and Bodmer, W. (1969) Nature 223, 358-363.

Nachman, R.L., Marcus, A.J. and Zucke-Franklin, D. (1967) J. Lab Clinical Med. 69, 651-658.

Nickerson, J.M. and Fuller, G.M. (1979) Proc. Natl. Acad. Sci. USA. 78, 303-307.

Niewiarowski, S. (1981) in Haemostasis and Thrombosis (ed Bloom, A.L. and Thomas, D.P.) pp 73-83, Churchill Livingstone.

Nirenberg, M.W. and Mathaei, J.H. (1961) Proc. Natl. Acad. Sci. USA. 47, 1588-1602.

Norris, K., Norris, F., Christiansen, L. and Fiil, N. (1983) Nucl. Acids Res. 15, 5103-5112.

Noyes, B.A., Mevarech, M., Stein, R. and Agarwal, K.L. (1979) Proc. Natl. Acad. Sci. USA. 76, 1770-1774.

Odink, K., Fey, G., Wiebauer, K. and Diggleman, H. (1981) J. Biol. Chem. 256, 1453-1458.

Ohno, S.(1967) Sex Chromosomes and Sex-Linked Genes (edit. Labhart, A., Mann, T. and Samuels, L.T.) Monographs on Endocrinology, 1 (Springer-Verlag).

Ohno, S. (1973) Nature 244, 259-262.

Olaisen, B., Teisberg, P. and Gedde-Dahl Jr., T. (1982) Hum. Genet. 61, 24-26.

Orkin, S.H., Markham, A. and Kazazian, Jr., H. (1983) J. Clin. Invest. 71, 775-779.

Owen, C.A., Bowie, E.J.W., Fass, D.N., Perez, R.A., Cole, T.L. and Stewart, M. (1979) Mayo Clinic Proceedings 54, 375-380.

Panabieres, F., Fort, Ph., Piechaczyk, M., Blanchard, J.M. and Jeanteur, Ph. (1982) Gene 19, 321-326.

Pelham, H.R.B. (1982) Cell 30, 517-528.

Pellicer, A., Robbins, D., Wold, B., Sweet, R., Jackson, J., Lowy, I., Roberts, J.M., Sim, G.K., Silverstein, S. and Axel, R. (1980) Science 209, 1414-1421.

Perucho, M., Hanahan, D., Lipsich, L. and Wigler, M. (1980) Nature 285, 207-210.

Phillips, J.A., Hijelle, B.L., Seeburg, P.H. and Zachman, M. (1981) Proc. Natl. Acad. Sci. USA. 78, 6372-6375.

Pindyck, J., Mosesson, M.W., Bannerjee, D. and Galanakis, D. (1977) Biochim. Biophys. Act. 492, 377-386.

Pizzo, S.B., Shwartz, M.L., Hill, R.L. and McKee, P.A. (1972) J. Biol. Chem. 247, 636-645.

Pouit, L., Marcille, G., Suscillon, M. and Hollard, D. (1972) Thromb. Diasthesis Haemorrhagica 27, 599.

Powell, B.C., Sleigh, M.J., Ward, K.A. and Rogers, G.E. (1983) Nucl. Acids Res. 11, 5327-5346.

Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.

Ratnoff, O.D. and Bennette, B. (1973) Science 179, 1291-1298.

Renwick, J.H. (1971) Ann. Rev. Genet. 5, 81-120.

Ricciuti, F. and Ruddle, F.H. (1973) Genetics 74, 661-678.

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

Rixon, M.W., Chan, W-Y, Davie, E.W. and Chung, D.W. (1983) Biochemistry 22, 3237-3244.

Robson, K.H., Chandra, T., McGillivray, R.T.A., Woo, S.L.C. (1982) Proc. Natl. Acad Sci. USA. 79, 4701-4705.

Rogers, J. and Wall, R. (1980) Proc. Natl. Acad. Sci. USA. 77, 1877-1879.

Ruddle, F.H. (1981) Nature 294, 115-120.

Ryan, M.J. (1977) Fed. Proc. Am. Soc. Exp. Biol. 36, 732 (abstr.)

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA. 74, 5463-5467.

Searle, A.G. (1981) Symp. Zool. Soc. Lond. 47, 63-84.

Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980) Cell 22, 361-370.

Shainoff, J.R. and Page, I.H. (1960) Circulation Res 8, 1013-1022.

Sherman, L.A., Gaston, L.W., Kaplan, M.E. and Spivak, A.R. (1972) J. Clin. Invest. 51, 590-597.

Shore, D., Langowski, J. and Baldwin, R.L. (1981) Proc. Natl. Acad. Sci. USA. 78, 4833-4837.

Shoulders, C.C. and Barelle, F.E. (1982) Nucl. Acids Res. 10, 4873-4882.

Smith, M. and Gillam, S. (1981) In "Genetic Engineering:Principles and Methods" (J.K. Setlow and A. Hollaender, eds.) Vol. 3, 1-32. Plenum Press, New York.

Smith, J., Cooke, E., Fotheringham, I., Pheby, S., Derbyshire, R., Eaton, M.A.W., Doel, M., Lilley, D.M.J., Pardon, J.F., Patel, T., Lewis, H. and Bell, L.D. (1982) Nucl. Acids Res. 10, 4467-4482.

Soloman, E. and Bodmer, W.F. (1979) Lancet 1, 923.

Solum, N.O. (1973) Throm. Res. 2, 55-70.

Sood, A.K., Pereira, D. and Weissman, S.M. (1981) Proc. Natl. Acad. Sci. USA. 78, 616-620.

Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.

Spandidos, S.A. and Simonovitch, L. (1977) Cell 12, 235-242.

Spritz, R.A., Jagadeeswaran, P., Choudary, P.V., Biro, P.A., Elder, J.T., de Riel, J.K., Manley, J.L., Gefter, M.L., Forget, B.G. and Weissman, S.M. (1981) Proc. Natl. Acad. Sci. USA. 78, 2455-2459.

Struhl, K. (1982) Nature 300, 284-287.

Stryer, L., Cohen, C. and Lanridge, R. (1963) Nature 197 793-794.

Suggs, S., Wallace, B., Hirose, T., Kawashima, E., Itakura, K., (1981) Proc. Natl. Acad. Sci. USA. 78, 6613-6617.

Swan, D., D'Eustachio, P., Leinwand, L., Seidman, J., Keithley, D. and Ruddle, F.H. (1979) Proc. Natl. Acad. Sci. USA. 76, 2735-2739.

Szostak, J.W., Stiles, J.I., Tye, B.-K., Chiu, P., Sherman, F. and Wu, R. (1980) Methods in Enzymology (R. Wu, ed.) vol. 68, pp419-428.

Takagi, T. and Doolittle, R.F. (1975a) Biochemistry 14, 5149-5156.

Takagi, T. and Doolittle, R.F. (1975b) Biochemistry 14, 940-946.

Talmadge, K. and Gilbert, W. (1980) Gene 12, 235-241.

Temple, G.F., Dozy, A.M., Roy, K.L. and Kan, Y.W. (1982) Nature 296, 537-540.

Tischfield, J.A. and Ruddle, F.H. (1974) Proc. Natl. Acad. Sci. USA. 71, 45-49.

Tooney, N.M. and Cohen, C. (1977) J. Mol. Biol. 110, 363-385.

Topfer-Peterson, E., Lottspeich, F. and Henschen, A. (1976) Hoppe-Seylers Z. Physiol. Chem. 357, 1509-1513.

Tosi, M., Young, R.A., Hagenbuckle, O. and Schibler, U. (1981) Nucl. Acids Res. 9, 2313-2323.

Unterman, R.D., Lynch, K.R., Nakhasis, H.K., Dolan, K.P., Hamilton, J.W., Cohn, D.V. and Fiegelson, P. (1981) Proc. Natl. Acad. Sci. USA. 78, 3478-3482.

Verhaeghe, R., Verstrete, M., Vermylen, J. and Veryler, C. (1974) Br. J. Haematol. 26, 421-433.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomdico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA. 75, 3727-3725.

Vora, S. and Franke, U. (1981) Proc. Natl. Acad. Sci. USA. 78, 3738-3742.

Wahl, G.M., Stem, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA. 76, 3683-3687.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979) Nucl. Acids Res. 6, 3543-3557.

Wallace, R.B., Johnson, P.F., Tanaka, S., Schold, M., Itakura, K. and Abelson, J. (1980) Science 209, 1396-1400.

Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kwashima, E.H. and Itakura, K. (1981) Nucl. Acids Res. 9, 879-894.

Wangh, L.J., Holland, L.J., Spolski, R.J., Aprison, B.S. and Weisel, J.W. (1983) J. Biol. Chem. 258, 4599-4605.

Watt, K.W.L., Takagi, T. and Doolittle, R. (1978) Proc. Natl. Acad. Sci. 75, 1731-1735.

Watt, K.W.K., Takagi, T. and Doolittle, R.F. (1979a) Biochemistry 18, 68-76.

Watt, K.W.K., Cottrell, B.A., Strong, D.D. and Doolittle, R.F. (1979b) Biochemistry 18, 5410-5416.

Weatherall, D.J. and Clegg, J.B. (1982) Cell 29 7-9.

Weisel, J.W., Warren, S.G. and Cohen, C. (1978) J. Mol. Biol. 126 159-183.

Weiss, M.C. and Green, H. (1967) Proc. Natl. Acad. Sci. USA. 58, 1104-1111.

Weiss, B. and Richardson, C.C. (1968) J. Biol. Chem. 243, 4556-4563.

Westaway, D. and Williamson, R. (1981) Nucl. Acids Res. 9, 1777-1788.

Wetzel, R., Heyneker, H.L., Goeddel, D.V., Thurani, P., Shapiro, J., Crea, R., Low, T.K.L., McClure, J.E., Thurman, G.B. and Goldstein, A.L. (1980) Biochemistry 19, 6096-6104.

Wigler, M., Silverstein, S., Lee, L-S., Pellicer, F., Cheng, Y-C. and Axel, R. (1977) Cell 11, 223-232.

Williams, J.G. (1981) in Genetic Engineering vol. 1. Ed. Williamson, R., Academic Press Ltd., London. pp 1-59.

Williams, C.J. and Prockop, D.J. (1983) J. Biol. Chem. 258, 5915-5921.

Winter G., Fersht, A.R., Wilkinson, A.J., Zoller, M. and Smith, M. (1982) Nature 229, 756-758.

Witt, I.and Tesch, R. (1979) Thromb. and Haemostasis 42, 79 (abstract).

Wolfenstein-Todel, C. and Mosesson, M.W. (1980) Proc. Natl. Acad. Sci. USA. 77, 5069-5073.

Wolfenstein-Todel, C. and Mosesson, M.W. (1981) Biochemistry 20, 6146-6147.

Woods, D.E., Markham, A.F., Ricker, A.T., Goldberger, G. and Colten, H.R. (1982) Proc. Natl. Acad. Sci. USA. 79, 5661-5665.

Wozney, J., Hanahan, D., Morimoto, R., Boedtker, H. and Doty, P. (1981) Proc. Natl. Acad. Sci. USA. 78, 712-716.

Yamada, Y., Avvedimento, V.E., Murdryj, M., Ohkuba, H., Vogeli, G., Irani, M., Pastan, I. and de Gommbrugghe, B. (1980) Cell 22, 887-892.

Young, R.A., Hagenbuchle, O. and Schibler, U. (1981) Cell 23, 451-458.

Zoller M.J. and Smith, M. (1982) Nucl. Acids Res. 10, 6487-6500

Volume 11 Number 21 1983 Nucleic Acids Research

Isolation and characterisation of cDNA clones for the Aα- and γ-chains of human fibrinogen

A.M.A.Imam, M.A.W.Eaton[1], R.Williamson and S.Humphries

Department of Biochemistry, St.Mary's Hospital Medical School, University of London, London W2 1PG, and [1]Department of Chemistry, Celltech Ltd., 250 Bath Road, Slough SL1 4DY, UK

Received 26 July 1983; Revised and Accepted 13 October 1983

ABSTRACT

cDNA clones coding for the Aα- and γ-chains of human fibrinogen have been isolated from an adult liver cDNA library. Clones were identified by hybridisation with mixtures of synthetic oligonucleotides 17 bases long, predicted using amino acid sequence data for each chain. The cDNA insert sizes are 1,950bp for Aα-fibrinogen and 950bp for γ-fibrinogen. The clones do not show any cross-hybridisation. Each cDNA hybridises to a unique sequence in the human genome. In adult human liver, Northern blots give an estimated messenger RNA size of 2.6kb for Aα-fibrinogen and 1.8kb for γ-fibrinogen.

INTRODUCTION

Fibrinogen (factor I) is an abundant protein present at 0.3 mg/ml in human plasma and synthesised in the liver. It is an acute phase protein which participates in the final steps of the coagulation pathway in higher vertebrates. A number of genetically determined variants for human fibrinogen have been reported (1). Elevated levels of fibrinogen and other coagulation factors, such as factor VII and factor VIII, have been correlated with a predisposition to ischaemic heart disease (2,3).

Native fibrinogen (MW 340,000) is composed of dimers of Aα- (MW 64,000), Bβ- (MW 56,000) and γ- (MW 47,000) chains linked by disulphide bridges. The Bβ- and γ-fibrinogen chains are glycosylated at residues 364 and 52 respectively. The N-termini of the Aα- and Bβ-fibrinogen chains respectively terminate in fibrinopeptides A and B, which are cleaved sequentially by thrombin to give fibrin monomers. These polymerise to form a network of fibres. A transamidation reaction catalysed by factor XIIIa covalently links the fibres, following which platelet aggregation completes clot formation.

The complete amino acid sequences of all three chains have been determined (4-7). The chains show sequence homology, suggesting that a common ancestral sequence underwent duplication and divergence. The chains are rich in Trp and Met, a fact used to our advantage in determining regions of low ambiguity for the synthesis of gene-specific oligonucleotide mixtures to be used in isolating the clones.

The isolation of clones coding for rat and human fibrinogen has recently

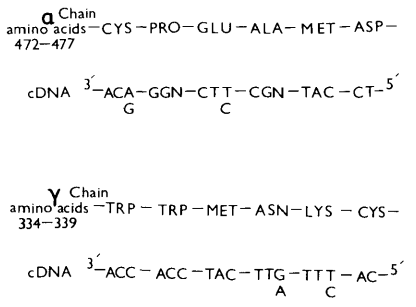

Fig.1 Sequence of synthetic oligonucleotide mixtures used for screening the adult human liver cDNA library. The A α-fibrinogen probe is a mixture of 64 different oligonucleotides. The γ-fibrinogen probe is a mixture of 4 oligonucleotides.

been reported (8-12). Here we report the isolation and characterisation of cDNAs corresponding to the human A α- and γ-fibrinogen chains. We use these clones in a preliminary analysis of the structure and expression of human fibrinogen genes.

MATERIALS AND METHODS

Screening of an adult human liver cDNA library

The 17-base long oligonucleotide probes were synthesised by the solid-phase phosphotriester method as described by Smith et al (13). The cDNA library used was generously provided by Dr. Derek Woods and has been previously described (14). 1,000 clones were plated on a 82mm nitrocellulose filter, which was replica-plated onto nitrocellulose filters of the same size, amplified using chloramphenicol and screened with the oligonucleotide mixtures using a modification of the method of Hanahan and Meselson (15). The oligonucleotides were 5'- labelled with γ-[32P]ATP (5,000 Ci/mM), using T4 polynucleotide kinase (PL Biochemicals) (16). Oligo(dT), 12-18 nucleotides in length (Boehringer), was similarly labelled to a specific activity of 1×10^7 cpm/ug. Hybridisation to the filter-bound DNA was performed in a solution containing 0.9M NaCl/0.09M sodium citrate/standard Denhardt's solution/0.05% sodium pyrophosphate/50 ug/ml sonicated herring sperm DNA/10 ug/ml poly(A) and labelled oligonucleotide probe (5×10^6 cpm/ug) for four hours at 37°. After hybridisation the filters were washed in 0.9M NaCl/0.09M sodium citrate/0.05% sodium pyrophosphate at 37° for 2x30 min and then sequentially at 46° for 15 min and 50° for 2 min. Washed filters were exposed to preflashed Fuji X-ray film at -70°, using an Okamoto tungstate intensifying screen.

Plasmid preparation and DNA analysis

Bacterial clones were grown in L-broth supplemented with tetracycline (25 ug/ml). Plasmid DNA was prepared by a modification of the cleared lysate method (17) and purified on CsCl gradients. Restriction enzymes were purchased from

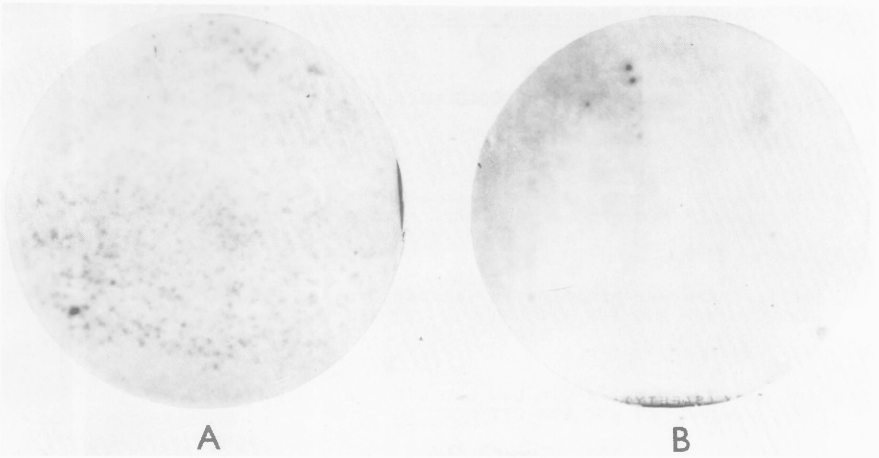

Fig.2 Screening of 1,000 recombinants from the cDNA library with A α-oligonucleotide (A) and γ-oligonucleotide (B). Filter hybridisation and post-hybridisation washes were as in Materials and Methods. Exposure to X-ray film was for 16 hrs.

Bethesda Research Labs; DNA digests were performed according to the manufacturer's recommendation, and the cleaved fragments were analysed on agarose or acrylamide gels. Restricted human DNA samples (10 ugs/slot) were run on 1% agarose gels and blotted onto nitrocellulose filters as described by Southern (18). Plasmid probes were labelled by nick translation (19), using α-[32P]dCTP, 400 Ci/mM (Amersham), and hybridised with the addition of dextran sulphate to 10% in the prehybridisation and hybridisation solutions (20). DNA sequencing was carried out by the enzymatic method of Sanger (21). The cDNA clones were digested with PstI and subcloned into the PstI site of the M13mp9 polylinker. The resulting recombinant single stranded M13 phage DNA was used as a template for sequencing reactions, primed with the M13 universal primer.

RNA preparation and Northern blotting

A post mortem specimen of male adult human liver was taken 30 min after death, frozen immediately in liquid nitrogen and stored at -70°. RNA was prepared as described by Auffray and Rougeon (22), denatured by boiling in 2.2M formaldehyde/50% formamide, run on a 1% agarose/2.2M formaldehyde denaturing gel in MOPS (20mM Na-MOPS/5mM sodium acetate/1mM EDTA) buffer and blotted onto nitrocellulose paper.

RESULTS

Identification of A - and -fibrinogen cDNA clones

1,000 recombinants of an adult liver cDNA library were screened with 17-base long mixtures of oligonucleotide specific for A α- and γ-fibrinogens. Fig 1 shows the amino acid sequences used to construct the oligonucleotides and the corresponding mixtures of DNA fragments. The A α-fibrinogen probe is a

Clone pAF1 insert DNA sequence

```
                                          91
                                          Met Glu Ile Leu Arg Gly Asp
                CTGCAGGGGGG GGCGGGGGGCG ATG GAA ATT TTG AGA GGC GAT
                          10        20          30            40
                  PstI

              101                                          111
Phe Ser Ser Ala Asn Asn Arg Asp Asn Thr Tyr Asn Arg Val Ser Glu
TTT TCC TCA GCC AAT AAC CGT GAT AAT ACC TAC AAC CGA GTG TCA GAG
      50            60          70          80            90

                                121
Asp Leu Arg Ser Arg Ile Glu Val Leu Lys Arg Lys Val Ile Glu
GAT CTG AGA AGC AGA ATT GAA GTC CTG AAG CGC AAA GTC ATA GAA
         100          110          120          130

        131
Lys Val Gln His Ile Gln Leu Leu Gln
AAA GTA CAG CAT ATC CAG CTT CTG CAG
                              (A)
  140           150          160
                              PstI
```

Clone pGF1 insert DNA sequence

```
                                                CTGCAGGGGGG GGGGGGGGC
                                                          10        20
                                                  PstI

183                                             193
Glu Ile Asp Gly Ser Gly Asn Gly Trp Thr Val Phe Gln Lys Arg Leu
GAA ATC GAT GGG TCT GGA AAT GGA TGG ACT GTG TTT CAG AAG AGA CTT
             30          40          50          60

                 203                                        213
Asp Gly Ser Val Asp Phe Lys Lys Asn Trp Ile Gln Tyr Lys Glu Gly
GAT GGC AGT GTA GAT TTC AAG AAA AAC TGG ATT CAA TAT AAA GAA GGA
 70          80          90          100         110

                                223
Phe Gly His Leu Ser Pro Thr Gly Thr Thr Glu Phe Trp Leu Gly Asn
TTT GGC CAT CTC TCT CCT ACT GGC ACA ACA GAA TTT TGG CTG GGA AAT
    120         130         140         150         160

        233                                        243
Glu Lys Ile His Leu Ile Ser Thr Gln Ser Ala Ile Pro Tyr
GAG AAG ATT CAT TTG ATA AGC ACA CAG TCT GCC ATC CCA TAT
      170         180         190         200
```

Fig.3 Partial nucleuotide sequence of clones pAF1 and pGF1 inserts and the amino acid sequences derived from them. The numbering of amino acids is as previously reported (4-7). The nucleotide in parentheses is that reported by Kant et al (9).

mixture of 64 different oligonucleotides, and the γ-fibrinogen probe a mixture of four oligonucleotides. For Aα- fibrinogen a single positive colony (Fig 2A) was detected, while three positives (Fig 2B) were found using the γ-fibrinogen probe. The positive clones were picked into ordered arrays and rescreened with each of the oligonucleotide mixtures. Plasmid DNA from each clone

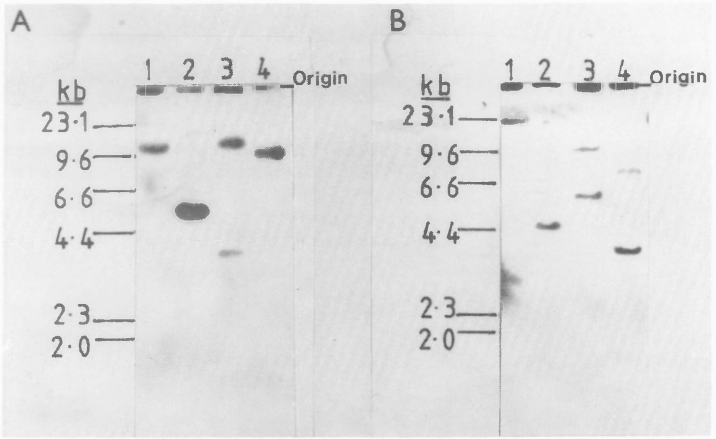

Fig.4 Southern blot hybridisation of human DNA digested with BamHI (lane 1), HindIII (lane 2), BglII (lane 3), and PstI (lane 4). Panels A and B were respectively hydridised with nick-translated pAF1 and pGF1. $2x10^7$ cpm of each probe ($2x10^8$ dpm/ug) was incubated with filters for 16 hrs in 10 mls. Filters were washed in 0.1xSSC, 0.1% SDS at 65° for 4x30 min and were exposed to X-ray film for 16 hrs at -70°. Band sizes for panel A are; 13.9kb (lane 1), 5.9kb (lane 2), 15.6kb and 3.9kb (lane 3), and 13.1kb (lane 4). Similarly, band sizes for panel B are; 21.3kb (lane 1), 4.3kb (lane 2), 11.3kb and 5.8kb (lane 3), and 7.9kb and 3.5kb (lane 4). λ DNA digsted with HindIII is shown on the left of both panels.

was digested with PstI. All recombinants contained an insert that was excised by this enzyme. Restriction of clone (pAF1), isolated using the probe for Aα-fibrinogen, with PstI gives two fragments: a 1,790bp fragment, and a small 160bp fragment. The largest of the γ- fibrinogen chain cDNA clones, (pGF1), on restriction with PstI gives a 700bp fragment, and a 250bp fragment. Therefore, the total insert length is 1,950bp for the Aα-fibrinogen clone (pAF1) and 950bp for the γ-fibrinogen clone (pGF1).

Each recombinant was sequenced by subcloning PstI fragments into the PstI site of the M13mp9 polylinker. The recombinant phage single strand DNA was used as a template for sequencing reactions, giving the sequence shown in Fig 3. The DNA sequence of pAF1 confirms that the clone codes for Aα- fibrinogen, and pGF1 for γ-fibrinogen.

For the Aα-chain, the sequence spans amino acids 91 to 137, and agrees with published amino acid data except at position 128 where we predict glutamic acid rather than glutamine (6). The sequence shown is from the small PstI fragment. The recombinant is long enough to code for all the remaining amino acids as well as the 3'-untranslated region and the poly(A) tail. The presence of poly(A) sequences in the clone was confirmed by probing with γ-[32P]-labelled oligo(dT) (data not shown).

200bp of clone pGF1 were sequenced. The 62 residue amino acid sequence

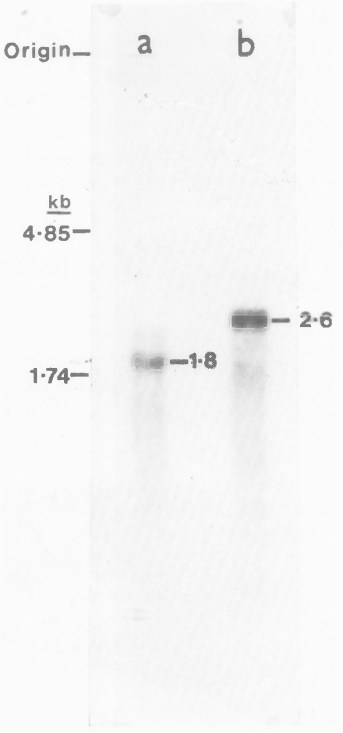

Fig.5 Detection of mRNAs for Aα- and γ-fibrinogen. 10 ugs of total RNA from human liver was run on a 1% agarose denaturing gel. Lanes (a) and (b) were hybridised with nick-translated clones pGF1 and pAF1 respectively. Human 18S (1.74kb) and 28S (4.85kb) rRNAs were used as markers. Filters were washed 1xSSC, 0.1% SDS at 65°; for 2x30 min, and exposed to X-ray film for 14 hrs at -70°.

agrees precisely with the published protein sequence of human γ-fibrinogen (5). The cDNA sequence spans amino acids 183-244 of the protein . This recombinant is also long enough to contain the remaining coding sequence and 3'- untranslated region. Once again, the presence of a poly(A) sequence in the 3'-portion of the cDNA was confirmed using labelled oligo(dT).

Analysis of the Aα- and γ-fibrinogen genes

In order to determine if the Aα- and γ-cDNA clones share any sequence homology, a cross-hybridisation study was performed. Plasmid DNA from pAF1 and pGF1 was digested with PstI and duplicate samples run on an agarose gel. The DNA was blotted onto nitrocellulose filters which were hybridised either with nick-translated pAF1 or pGF1. Even when the filters were washed at low stringency (3xSSC), there was no detectable hybridisation between the clones (not shown).

When the recombinants were used as probes of Southern blots of human genomic DNA digested with various restriction enzymes, only simple band patterns were observed (Fig 4).

The size of the mRNA for Aα- and γ-fibrinogen was determined using Northern blots. Total RNA from adult human liver was run on a denaturing agarose gel, the RNA transferred to nitrocellulose and the blot probed with nick-translated Aα- or γ-fibringen probe (Fig 5). By comparison with 18S and 28S ribosomal RNA markers, the Aα-fibrinogen mRNA is 2.6kb in length, and the γ-fibrinogen mRNA 1.8kb in length.

DISCUSSION

The liver synthesises a number of proteins involved in coagulation, including the abundant protein fibrinogen. Consequently, only 1,000 clones of an adult human liver cDNA library were screened with fibrinogen-specific oligonucleotides. Clones pAF1 and pGF1, corresponding to the Aα- and γ-fibrinogen chains respectively, were isolated by screening with 17-base long oligonucleotides complementary to regions of each protein chain showing low ambiguity in codon assignment.

The partial DNA sequence of the pAF1 insert agrees with the corresponding region of the cDNA sequence for Aα- fibrinogen reported by Kant et al (9), except at one place. At the codon for glutamine 137, a change from A to G creates a PstI site. This internal PstI site has been confirmed by our restriction digest data . The Aα fibrinogen cDNA sequence of Rixon et al (10) also shows the presence of this site. At present we are determining the frequency of occurence of this PstI site polymorphism in the human population. In agreement with the published Aα-fibrinogen cDNA sequences (9,10), we also assign residue 128 as glutamic acid.

The 200bp sequence of pGF1 confirms the identity of the clone, and the predicted amino acid sequence agrees precisely with the published sequence of human γ-fibrinogen. Recently, the entire sequence of the rat γ-fibrinogen cDNA has been reported (23). By comparison with the rat sequence, there are 8/62 amino acid changes (12.9%) and 30/206 base changes (14.5%) in the region we have sequenced. Of the 30 base changes, 19 are silent substitutions; the remaining 11 cause 8 amino acid changes. Therefore, this region of γ-fibrinogen is not as highly conserved in evolution as the carboxyterminal region (9).

Both clones give simple band patterns when used as probes of Southern blots of total human genomic DNA, demonstrating that they represent unique sequences and donot constitute a multigene family.

The sizes of the Aα- and γ-fibrinogen mRNAs we detect in human liver are similar to those reported in the rat (8) (Aα- 2.3kb , γ- 1.95kb). These observations in human liver confirm that each fibrinogen chain is coded by an independent mRNA and not by a common RNA transcript that either undergoes subsequent differential splicing, or is translated to give a polyprotein. In the Northern blot, the intensity of the hybridisation signal with the Aα-fibrinogen probe is greater than for γ- fibrinogen. However, the Aα-fibrinogen insert is roughly twice as large as the γ-fibrinogen insert, and therefore we estimate that the two mRNAs are present in almost equal abundance.

ACKNOWLEDGEMENTS

We thank Dr. S. Wallis for supplying adult human liver RNA, Paul North for help with the computer search of protein sequences, and Padi Kosol for help in preparing the manuscript. This work was supported by grants from the British Heart Foundation and the Medical Research Council.

REFERENCES

1. Bloom,A.L. (1981) in Haemostasis and Thrombosis, Bloom,A.L. and Thomas,D.P. eds., pp.321-370, Churchill Livingstone, London.
2. Mead,T.W. and North,W.R.S. (1977) Brit.Med.Bull. 33, 283-288.
3. Mead,T.W., North,W.R.S., Chakrabarty,R., Stirling,Y., Haines,A.P. and Thompsom,S.G. (1980) Lancet 17, 1050-1053.
4. Blomback,B., Hessel,B. and Hogg,D. (1976) Thromb.Res. 8, 639-658.
5. Lottspeich,R. and Henschen,A. (1977) Hoppe-Seyler's Z. Physiol.Chem. 358, 935-938.
6. Doolittle,R.F., Watt,K.K., Cottrell,B.A., Strong,D.D. and Riley,M. (1979) Nature 280, 464-468.
7. Watt,K.W.K., Tagaki,T. and Doolittle,R.F. (1979) Biochemistry 18, 68-76.
8. Crabtree,G.R. and Kant,J. (1981) J.Biol.Chem. 256, 9718-9723.
9. Kant,J.A., Lord,S.T. and Crabtree,G.R. (1983) Proc.Natl. Acad.Sci.USA 80, 3953-3957.
10. Rixon,M.W., Chan,W-Y., Davie,E.W. and Chung,D.W. (1983) Biochemistry 22, 3237-3244.
11. Chung,D.W., Que,B.G., Rixon,M.W., Mace,M. and Davie,E.W. (1983) Biochemistry 22, 3244-3250.
12. Chung,D.W., Chan,W-Y. and Davie,E.W. (1983) Biochemistry 22, 3250-3256.
13. Smith,J., Cook,E., Fotheringham,I., Pheby,S., Derbyshire,R., Eaton,M.A.W., Doel,M., Lilley,D.M.J., Pardon,J.F., Patel,T., Lewis,H. and Bell,L.D. (1982) Nucl. Acids Res. 10, 4467-4482.
14. Woods,D.E., Markham,A.F., Ricker,A.T., Goldberger,G. and Colten,H.R. (1982) Proc.Natl.Acad.Sci.USA 79, 5661-5665.
15. Hanahan,D. and Meselson,M. (1980) Gene 10, 63-67.
16. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
17. Birnboim,H.C. and Doly,J. (1979) Nucl.Acids Res. 7, 1513-1523
18. Southern,E. (1979) Methods Enzymol. 68, 152-176.
19. Rigby,P.W.J., Deickmann,M., Rhodes,C. and Berg,P. (1977) J.Mol.Biol. 113, 237-251.
20. Jeffreys,A.J. and Flavell,R.A. (1977) Cell 12, 429-439.
21. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc.Natl. Acad.Sci.USA 74, 5463-5467.
22. Auffray,C. and Rougeon,F. (1980) Eur.J.Biochem. 107, 303-314.
23. Crabtree,G.R. and Kant,J.A. (1982) Cell 31, 159-166.