THE ISOLATION AND CHARACTERISATION OF HUMAN GLOBIN AND INTERFERON GENES.

by

Ian James Jackson.

a thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in the
University of London.

November, 1980.

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

ABSTRACT

The development of recombinant DNA technology has led to a new approach to the study of human genetic disease. The first stage of this "new genetics" is the isolation of genes. I have examined several methods for the isolation of human genes as recombinant molecules. Several methods for gene enrichment prior to *in vitro* recombination were developed, and as an alternative, partial digestion of DNA with restriction enzymes was characterised, to enable a library of human genes to be made.

A β-globin gene from a patient with homozygous β^0-thalassaemia was isolated in the λ-phage vector NM788. The DNA was enriched for the β-globin gene by size fractionation using agarose gel electrophoresis before cloning. The gene coding for fibroblast interferon was identified and isolated from a library of human genes, cloned in Charon 4A.

The β-globin gene was presumed to be defective, as the patient from whom it was isolated produces no β-globin protein. The gene was recloned into the plasmid pAT153 and examined for deletions or insertions with respect to the normal gene, by mapping with various restriction enzymes. No differences from the normal map were detected. The gene was also transcribed *in vitro* using an extract of human cells, and was shown to initiate transcription with equal efficiency when compared to a normal β-globin gene. The significance of this unexpected finding is discussed.

The isolated interferon gene-containing clone was mapped with restriction enzymes. The gene was found to be very close to one end of the isolated segment of human DNA. No evidence was found for the presence of intervening sequences within the gene.

CONTENTS.

Acknowledgments. i

1. INTRODUCTION.

1.1 The New Genetics.

- a) Gene Isolation. 1
- b) DNA Sequencing 5
- c) Methods for the Expression of Isolated Genes. 5
- d) *In Vitro* Mutagenesis. 7

1.2 Gene Structure and Function.

- a) Intervening Sequences and RNA Splicing. 8
- b) Gene Linkage. 10
- c) Gene Sequences. 11
- d) Functional Analysis of Isolated Genes. 13

1.3 Globin Gene Structure and Function.

- a) Globin Gene Linkage and Duplication. 16
- b) Globin Gene Sequences and Intervening Sequences. 20
- c) Globin Pseudogenes. 24
- d) Repetitive DNA in the Globin Loci. 27
- e) The Old Genetics and the New Genetics. 30

1.4 The Thalassaemia Syndromes.

- a) Alpha Thalassaemias. 31
- b) Beta Thalassaemias. 33
- c) Deletions in the Beta-Globin Locus: Long Range Effects.
- d) Perspectives. 39

1.5 Interferon. 43
- a) Interferon Production. 44
- b) Synthesis of Interferon by *E.coli*. 46
- c) Interferon Gene Families. 47

2. MATERIALS AND METHODS.

2.1 Materials. 50

2.2 Bacterial Strains. 52

2.3 DNA Manipulation Methods.

- a) Preparation of Human DNA. 53
- b) Restriction Enzyme Digestion. 54

c) Gel Electrophoresis. 54
d) Southern Transfer and Hybridisation. 55
e) Labelling of DNA by Nick Translation. 56
f) Ligation of DNA. 57
g) Isolation of DNA from Agarose Gels. 57
h) RPC-5 Chromatography. 58

2.4 Bacteriological Methods.

a) Plasmid Preparation 59
b) Transformation of HB101 with pAT153. 60
c) Colony Electrophoresis. 61
d) Preparation of Plating Cells. 62
e) Growth of Phage on Plates. 62
f) Large-Scale Growth of Phage. 62
g) Preparation of Phage DNA 64
h) Small-Scale Preparation of DNA From Phage. 65
i) Preparation of Extracts for In Vitro Packaging. 65
j) In Vitro Packaging of Phage. 66
k) Plaque Hybrisisation. 67

3. STRATAGIES AND TACTICS FOR THE ISOLATION OF HUMAN GENES.

3.1 Introduction.

a) Host/Vector Systems. 68
b) Gene Enrichment Methods. 70
c) Synthesis of Gene Libraries. 71

3.2 Developing the Technology.

a) The Gene Machine. 74
b) Small-Scale Preparative Gel Electrophoresis. 76
c) RPC-5 Chromatography. 81
d) Partial Digestion with EcoR1. 81

3.3 Manipulating the Vector.

a) DNA Preparation. 89
b) Isolation of the "Arms". 89
c) In Vitro Recombination of Human DNA with NM788.

3.4 Isolation of a β-Globin Gene From a Patient with Homozygous β^{o}-Thalassaemia.

a) The Patient. 95
b) Gross Structure of the β-Globin Gene. 96
c) Construction of a β^{o}-Thalassaemic Gene-Containing Phage. 100
d) Identification and Isolation of

a β^{o}-Thalassaemic Gene-Containg Recombinant 103

3.5 Isolation of the Gene for Fibroblast Interferon From a Human Gene Library.
- a) Screening the Library. 109
- b) Problems of Amplification 112

4. STRUCTURE AND FUNCTION OF A CLONED β^{O}-THALASSAEMIC GENE.

4.1 Structural Studies.
- a) Structure of Pad2. 116
- b) Subcloning of the β^{o}-Thalassaemic Gene into pAT153. 119
- c) Restriction Enzyme Mapping of β /7.5. 122

4.2 *In Vitro* Transcription of Normal and Thalassaemic Globin Genes. 127

4.3 Discussion. 130

5. STRUCTURAL ANALYSIS OF THE FIBROBLAST INTERFERON GENE.

5.1 Construction of the Nearest-Site Map
- a) Deduction of the Map. 132
- b) Orientation of the Map. 136
- c) Positioning of the Map. 136

5.2 Arrangemnt of the Cloned EcoR1 Fragments. 138

5.3 Structure of the Fibroblast Interferon Gene. 144

5.4 Discussion 147

6. CONCLUSIONS.

6.1 Gene Structure and IVS Function. 151

6.2 *In Vitro* Analysis of Gene Expression. 153

6.3 Future Problems. 157

Epilogue. 160

REFERENCES 161

List of Figures.

1.1 Map of deletions in the $\gamma/\delta/\beta$-globin gene locus. 38

3.1 Schematic diagram of the gene-machine. 75

3.2 Gel electrophoresis of gene machine fractions. 77

3.3 Southern blot of gene-macine fractions. 78

3.4 Small-scale DNA isolation from agarose gel 80

3.5 DNA fraction from agarose gel preparation. 82

3.6 Gel electrophoresis of RPC-5 fractions. 83

3.7 Southern blot of RPC-5 fractions. 84

3.8 Gel analysis of partial EcoR1 digestions. 87

3.9 Southern blot of partial digestions. 88

3.10 Sucrose gradient fractionation of phage arms. 91

3.11 DNA preparations of human DNA recombinants. 93

3.12 Southern blot of β^{o}-thalassaemic DNA. 98

3.13 Map of the δ/β-globin locus. 99

3.14 Preperations of human size fractionated DNA recombinants. 102

3.15 Plaque hybridisation to identify thalassaemic globin gene. 105

3.16 Plaque purification of thalassaemic gene. 106

3.17 Electrophoresis of thalassaemic gene. 107

3.18 Plaque hybridisation to identify fibroblast interferon recombinants. 111

3.19 Electrophoresis of isolated interferon genes. 113

3.20 Southern blot of isolated interferon genes. 114

4.1 Electrophoresis and Southern blot of Pad2. 117

4.2 Rudimentary map of Pad2. 120

4.3 Colony electrophoresis of Pad2 subclones. 123

4.4 Electrophoresis to analyse 5' and 3' ends of Pad2. 125

4.5 Map of β-globin gene. 126

4.6 *In vitro* transcription of normal and thalassaemic globin genes. 129

5.1 Southern blot of λ.H.IFNβ/Pad3. 133

5.2 Nearest site map of λ.H.IFNβ/Pad3. 137

5.3 Electrophoresis of digests of λ.H.IFNβ/Pad3. 139

5.4 Electrophoresis of isolated KpnI fragments. 143.

5.5 EcoR1 and SstI digests of λ.H.IFNβ/Pad3. 145

5.6 Map of λ.H.IFNβ/Pad3. 146

List of Tables

4.1 Sizes of fragments hybridising to the β-globin probe in Figure 4.1. 118

5.1 Sizes of fragments visualised in Figure 5.1. 134

5.2 Sizes of human DNA-containing fragments in Figure 5.3. 140

Acknowledgements.

I would like to thank my supervisor, Bob Williamson, for his help, advice and encouragement and for providing an exciting and fun group among which to work.

Special thanks are also due to the following;

Peter Little, for endless hours of discussion, argument, encouragement and beer.

Dave Westaway, similarly, for much of the same, plus a great deal of practical co-operation on the globin project.

Mike Houghton and Norman Carey at Searle for the opportunity to work on the interferon project.

Sergio Ottolenghi for the spleen from the thalassaemic patient.

Rhyaiza de Freund for doing the experiment in Chapter 4.

Emma Whitelaw for her handy hints and maternal instincts.

Ray, Rob, Mike and Gill for making up the rest of the globin lab.

The rest of the Biochemistry Department at St. Mary's for their help, especially the Five o'clock Crowd; Tim, Janet, Sue, Jack, Lesley *et al*.

Kass, for keeping me calm, cool and collected (relatively speaking), offering her sane (albeit often unwelcome) opinions, and teaching me that the true way to success is to mellow out.

1. INTRODUCTION.

1.1 The New Genetics.

"New Genetics" is a term sometimes applied to the ways of analysing gene function directly, by studying the DNA of the gene itself and its interactions, rather than studying the phenotypic effects of genes. The principle of the new genetics was made possible by four technical advances. These are, gene isolation, DNA sequence analysis, functional analysis of isolated genes and *in vitro* mutagenesis.

In the first part of this introduction I will review some of these methods and I will attempt an overview of the current knowledge of gene structure and function.

The approach taken by the new genetics is to isolate a gene, as part of a recombinant plasmid or bacteriophage, to determine its structure, ultimately by DNA sequencing, to add the gene to a system which mimics its *in vivo* function, and finally to create specific mutations in the gene and to examine their effect on its expression.

a) Gene Isolation.

The gene used as the starting material for this approach must contain the putative control elements of the gene. It must therefore be isolated from the DNA contained in the chromosomes, rather than from a DNA copy of the mRNA (cDNA).

However, in order to isolate a chromosomal gene, a cDNA-recombinant is usually isolated first. This provides a hybridisation probe for the chromosomal gene, with which the subsequent cloning is fairly straightforward. A few genes have been isolated without recourse to cloned cDNA probes. They are genes which code for "superabundant" mRNAs which are readily purified, and radioactive probes made by reverse transcription or by iodination. Genes isolated in this way include human γ-globin, mouse α-globin, mouse immunoglobulin λ-chain and silk-moth fibroin genes (Blattner et al, 1978; Brack and Tonegawa, 1977; Ohshima and Suzuki, 1977).

In the main, however, it is the isolation of the cDNA probe which is likely to pose the most problems in the future. Cloning of the cDNA from very abundant mRNA such as globin (Maniatis et al, 1976; Wilson et al, 1978; Little et al, 1978), chick ovalbumin (McReynolds et al, 1977), and immunoglobulin (Wall et al, 1978) is a relatively straightforward process. Other cDNA molecules made from relatively less abundant mRNAs such as β-endorphin (Roberts et al, 1979) or amylase (McDonald et al, 1980), relied on a certain amount of pre-enrichment of the mRNA by size fractionation and immunoprecipitation but are even so complementary to a large proportion of the message population.

The isolation of cDNA from specific, rare mRNAs is much more difficult. The isolations of the cDNAs from human fibroblast and leukocyte interferons by several laboratories probably embody the current "state of the art" of isolating rare cDNAs.

As the protein product of the gene can be identified by a biological assay, then the approach of Nagata et al (1980a) and Derynck et al (1980a) is possible. They were able to identify the interferon mRNA by injecting it into Xenopus oocytes and assaying the translation products. By binding cDNA-containing plasmids to filters, they were able to hybrid-select interferon mRNA from the mRNA population, and by repeated fractionation of the plasmid library they were able to focus in on the plasmid which contained the interferon cDNA.

The induction of interferon in cells presumably causes an increase in the interferon mRNA levels. Tanaguchi et al (1979) selected by hybridisation those sequences which were present in induced but absent from uninduced cells. They used these sequences as a probe to isolate a small number of clones, among which the fibroblast interferon cDNA was easy to find.

As the partial protein sequences of the interferons are known, small oligonucleotides can be synthesised which have complementarity to the predicted mRNA sequence. Goedell et al (1980a,b) used these to prime specific cDNA synthesis, which was used to probe a library of cDNA clones. Houghton et al (personal communication) used similarly primed specific fibroblat interferon cDNA as the material which was initially cloned.

Other workers, rather than isolating genes coding for known proteins, have focussed onto particular classes of expressed gene. By differential hybridisation, Williams and Lloyd (1979), isolated several cDNAs from mRNAs expressed at different times during Dictyostelium development. J. Crampton (personal communication) has identified cDNA recombinants from messages present in high abundance in lymphocytes, but at low abundance in fibroblasts.

Once a hybridisation probe is available, the methods for the isolation of the corresponding chromosomal gene are well characterised (see section 3.1). Furthermore, the gene from one species can often be identified by cross-hybridisation with the gene from another. The yeast and Drosophila actin genes were isolated by probing libraries of genes with a Dictyostelium actin probe (Gallwitz and Seidal 1980; Tobin et al, 1980) and the human ε-globin chromosomal gene was identified using a rabbit β-globin probe (Proudfoot and Baralle, 1979).

There are ways by which genes can be isolated without using hybridisation probes, but knowledge of the mRNA structure will always be needed in order to distinguish between coding sequences from intervening and flanking sequences.

Some genes can be isolated on the basis of their function, rather than their sequence. This entails using eukaryotic cells as hosts for the genes. Vectors capable of replicating in both yeast cells and in E.coli have been used to identify some yeast genes by their expression in yeast cells (Hsiao and Carbon, 1979; Nasmyth and Tatchell, 1980). In the so-called "plasmid-rescue" technique (Perucho _et al_, 1980) a recombinant plasmid is identified by the expression, in mammalian cells, of the gene carried and the plasmid is subsequently rescued by transformation of E.coli. This method will have some limited application and has been used to isolate the chicken thymidine kinase gene.

b) DNA Sequencing.

After the gene has been isolated, its structure must be determined.

Methods for the rapid determination of DNA sequences have improved considerably, even in the short time since the first such method was described (Sanger and Coulson, 1975). The method of Maxam and Gilbert (1977) is now most often used. It depends on end-labelling of the DNA fragment at a defined site and uses base-specific chemical cleavage to generate a series of fragments of increasing length, each terminating where the specific base occurs. The position of these known bases relative to the end-label is found by gel electrophoresis. The series of fragments generated for each base allows the sequence to be simply read off an autoradiograph.

The method of Sanger, Nicklen and Coulson (1977) uses a primer hybridised to the DNA or RNA to define one end of the molecule and generates a series of fragments terminating with a particular base by extending from the primer, including a small amount of a base-specific chain terminator. This method uses a single stranded substrate, and so is best suited to sequencing mRNA, or DNA cloned in the single-stranded vector M13 (Schreier and Cortese, 1979).

c) Methods for the Expression of Isolated Genes.

The expression of isolated eukaryotic genes was the last of the four components of the new genetics to be developed, and the methodology is still in a confused state. The approaches can broadly be divided into *in vitro* or cellular mehtods.

In vitro methods have been developed over the past three years. The most satisfactory results have been obtained using as a template genes transcribed by RNA polymerase III. These genes direct the synthesis of small RNA products such as tRNA and ribosomal 5S RNA. Extracts of *Xenopus* oocyte nuclei (Birkenmeir *et al*, 1978), of whole oocytes (Ng *et al*, 1979), or of cultured cells (Wu, 1978) have been prepared, and these have been shown to initiate and to terminate transcription of these genes correctly.

Several systems which will allow *in vitro* transcription of protein encoding (RNA polymerase II) genes have been described. Weil *et al* (1979) demonstrate accurate initiation of the adenovirus late genes using a cell-free extract plus exogenous DNA and RNA polymerase II. Manley *et al* (1980) show that the same genes can be transcribed from added DNA using endogenous RNA polymerase in an extract of HeLa cells. The transcripts are not terminated using either of these systems. No method has been found which will allow *in vitro* transcription of rRNA genes by RNA polymerase I.

The "cellular" methods of expression include microinjection of *Xenopus* oocyte nuclei (Mertz and Gurdon, 1977), the use of the mammalian virus SV40 as a vector to carry the gene into the cell (Goff and Berg, 1979), and the transformation of tissue culture cells by a selectable marker (Pellicer *et al*, 1978). The last technique has been extended so that additional genes, both physically linked (Mantei *et al* 1979) or unlinked (Wigler *et al*, 1979) genes have been introduced into mouse cells by cotransformation.

Although the strictly *in vitro* methods are preferable in that they are easier to work with and can be manipulated to modify their protein composition as well as the template, they require a vast excess of DNA to be added over the "natural" amounts, and the compartmentation and long range interactions imposed by the chromosomal structure are absent.

The genes introduced into the cells by transformation, although chromosomally integrated, still may not be expressed normally as they are not integrated into their "natural" location on the chromosome. This has been overcome in yeast cells; a ura3 gene has been replaced by exogenous DNA (Scherer and Davis, 1979).

d) In Vitro Mutagenesis.

In order to study the sequence requirements of gene expression, the function of both normal and *in vitro* mutated genes has been studied in the various systems described above. At its simplest this mutagenesis would involve gross deletions of DNA sequences by the elimination of restriction enzyme fragments. Deletions of rather less DNA can be produced by treating the end of a DNA molecule with exonuclease III and S1 nuclease. Several methods are available for introducing single base changes into a piece of DNA, some are dependent on a restriction enzyme site being near to the site of mutagenesis (Shortle and Nathans, 1978) whilst others are capable of changing any one base into any other (Hutchinson *et al*, 1978).

Although it may be tedious to carry out in practice, it is theoretically possible to change any base in a cloned DNA segment, and to test its function in some artificial expression system.

1.2 Gene Structure and Function.

The new technologies have enabled the structure of a number of genes from a number of eukaryotic species to be analysed at various levels; from a gross map provided by restriction enzyme mapping, to the DNA sequence itself.

a) Intervening Sequences and RNA Splicing.

Intervening sequences (IVSs) are segments of DNA present in the region of a gene which codes for RNA, but which are themselves not present in the mature RNA product. In other words, the mature RNA is not colinear with the DNA which codes for it. Intervening sequences were first observed in the 28S ribosomal RNA genes of Drosophila (White and Hogness, 1977; Glover and Hogness, 1977), but at the time they were beleived to be inactive genes.

The first evidence that intervening sequences are present in active genes came from Jeffreys and Flavell (1977), who used the Southern transfer technique (Southern, 1975) to construct a map around the rabbit β-globin genes in total cellular DNA. They noted that the distance between two intragenic restriction enzyme sites was 600bp greater in the cellular DNA than in the cDNA (or mRNA). This was true of all cell types examined. It was these data obtained from total cellular DNA which permitted the discovery of cloned genes containing IVSs to be accepted as relating to functional genes, and not inactive or spuriously rearranged genes.

Almost all protein-coding genes of eukaryotes or eukaryotic viruses examined have intervening sequences (Crick, 1979). The only definite exceptions to date are histone genes (Shaffner et al, 1978), the adenovirus-2 polypeptide IX gene (Alestrom et al, 1980) and a human leukocyte interferon gene (Nagata et al, 1980b).

The number and size of IVS shows considerable variation. The Bombyx fibroin gene has only one (Tsujimoto and Suzuki, 1979) as does one of the alleles of the rat insulin gene (Lomedico et al, 1979). All globin genes examined have two intervening sequences (see section 1.4), and the ovalbumin gene has seven (Dugiczyk et al, 1978; O'Hare et al, 1979). The gene with the largest number of intervening sequences currently known is the Xenopus vitellogenin gene, which has at least 33 (Wahli et al, 1980), and the gene which is known at the moment to have the most DNA in its IVSs is the mouse dihydrofolate reductase gene, which has only 1kb of coding potential, but at least 41kb of intervening sequence (Nunberg et al, 1980).

In addition to the genes transcribed by RNA polymerase II, some genes transcribed by RNA polymerase I and III have intervening sequences. Although it is thought that the Drosophila 28S gene with the IVS is not transcribed (Dawid, quoted in Abelson, 1979), the active yeast mitochondrial 21S rRNA gene also has an IVS (Bos et al, 1978). Some tRNA genes, transcribed by RNA polymerase III, have small intervening sequences (Goodman et al, 1977; Valenzuela et al, 1978).

The IVSs in the protein-encoding genes are highly variable in size, ranging from tens to thousands of base pairs in length. In rRNA genes the IVSs found have been of the order of thousands of bp., whereas the IVSs of tRNA genes are very small, between 14 and 19bp.

Of the several models proposed by which a contiguous RNA molecule could be made from a split gene, the experimental data support only one; that the intervening sequences are transcribed and subsequently removed from the transcript. Tilghman et al (1978b) showed by R-looping that a mouse β-globin-containing 15S RNA species contained transcripts of the IVSs. Pulse-chase experiments show that the 15S RNA and the globin mRNA have a precursor/product relationship (Ross and Knecht, 1978).

Gel electrophoresis has demonstrated that a series of mRNA precursors can be detected from those genes with larger numbers of intervening sequences. Thus, ovalbumin has seven IVSs and seven pre-mRNA bands. This indicates that the IVSs are spliced out in an ordered, rather than random manner (Roop et al, 1978)

It is likely that initiation of transcription occurs at the "cap-site"; the base which becomes the first base of the mature mRNA, after the cap structure. Tsai et al (1980) have used solution hybridisation to search in chick oviduct nuclei for transcription from DNA upstream of the 5' end of the mature ovalbumin mRNA. Within the limits of the method none was detected; there are therefore less than 0.1 transcripts per nucleus from this region.

Precursor tRNA molecules have been isolated which have been shown by fingerprinting to contain a transcript of the intervening sequence (Knapp et al, 1978). The pre-tRNA can be processed to mature tRNA in vitro (see section 1.3)

b) Gene Linkage.

Linkage between related genes has been shown in several cases. The sea-urchin histone genes were shown to be linked when they were isolated as recombinants (Kedes, 1975; Clarkson, 1976). The first demonstration of physical linkage in a higher eukaryote, however, used Southern blotting to show that the human β- and δ-globin genes were separated by about 5kb (Flavell _et al_, 1978b; Mears _et al_, 1978). Similar techniques similarly indicated that the two γ-globin genes, the $^{A}\gamma$- and δ-globin and the two α-globin genes are closely linked (Little _et al_, 1979; Bernards _et al_, 1979a; Orkin, 1979).

Cloning has since demonstrated linkage between these and other genes, including Drosophila heat shock genes (Craig _et al_, 1979), chicken ovalbumin and two other oestrogen-induced genes (Royal _et al_, 1979), silkmoth chorion protein genes (Jones _et al_) and human leukocyte interferon genes (Nagata _et al_, 1980b). No physical linkage has been reported between two unrelated genes.

c)Gene Sequences.

The major interest of the new genetics is in the DNA sequences involved in the expression of the gene.

A number of junctions between coding and intervening sequences have been sequenced. In most cases it is not possible to pinpoint exactly the junction as there is a small redundancy of sequence. If there is a unique splice point on the RNA it will only be defined when the excised IVS can be sequenced. The first attempt to identify conserved sequences at the splice point was unsatisfactory, as it relied on the false assumption that the 5' and 3' ends of an IVS would be operationally the same (Catterall _et al_, 1978). Better analysis by Breathnach et al (1979) allowed "Chambon's rule" to be formulated (Abelson, 1979). This rule indicates that all potential splice junctions can be alligned so that the sequence is:

coding region---$\downarrow$GT----IVS----AG$\downarrow$--- coding region.

The arrows indicate the putative positions of cleavage. There are further homolgies near the splice point but the divergence from a concensus sequence is rather great overall.

Although the assignment of function to sequence, on the basis of sequence data alone, must be tentative, sequencing of several genes has led to the proposition that a sequence about 22bp from the initiation site of RNA polymerase II genes has the same functional significance as the Pribnow box in prokaryotes (Pribnow, 1975; Rosenberg and Court, 1979). The Pribnow box occurs at about -10 with respect to the first base of the gene and has a generalised sequence of:

TATAATG.

Benoist et al (1980) suggest that the so-called "Hogness box" in eukaryotes has the sequence:

TATatAtt,

in which the upper case bases are the most conserved.

A further region of homology has been noted between diferent genes at around -70, named the "CCAAT box" after the five most coserved bases.

On the other hand it should be borne in mind that the conservation of sequence noted at the 5' ends of several RNA polymerase III transcribed genes (Korn and Brown, 1978), have subsequently proved not to be the major signals for trancriptional initiation (see section 1.3).

Termination of mRNA transcription often appears to end in a region of several A residues. As mRNA is polyadenylated, it is not possible to precisely know which base is the last transcribed base, and which is the first added in polyadenylation. Upstream of all termination sites at which a poly(A) tail is added, the sequence AATAAA is seen. This is absent from histone genes, which of course are polyA(-). A homologous sequence, related to the sequence TTTTCACTGC, is found near to many (but not all) RNA polymerase II termination sites (including histone H1) (Benoist et al, 1980).

d)Functional Analysis of Isolated Genes.

Once methods had been developed which gave faithful expression of isolated genes, then mutagenesis could begin to define the functionally important regions of the gene.

i)Splicing

Hamer and Leder (1979a,b) have investigated the function of intervening sequences by incorporating various fragments of the mouse β-globin gene into the late region of SV40. The transcripts of the virus indicate that the presence of an IVS (a viral IVS in this case) is necessary to stabilize in some way the primary transcript. They also found that splicing of the mouse β-globin major IVS could take place in the monkey cells, and still took place even if only 18bp of globin-coding RNA was to the 5' side of the splice point.

Wallace et al (1980) constructed yeast suppressor tRNA genes with the IVS precisely removed. After introducing this gene into yeast cells the suppressor function was still expressed, indicating that the IVS is not essential to the function of this gene.

Knapp _et al_ (1978; 1979) and Peebles _et al_ (1979) have studied in vitro the splicing enzymes involved in yeast tRNA expression. They find that the reaction can be dissected into two steps; a cleavage step followed by a ligation. If the cleaved half-molecules are purified and mixed together, they can be ligated by a purified enzyme. The work suggests that the half-molecules spontaneously will reassemble to allow correct joining of their ends. This may be a function of the large degree of secondary and tertiary structure of the tRNAs in general, or it may be due to the unusual specificity of the ligation enzyme. It is not possible to ascertain whether similar mechanisms will apply to rRNA or mRNA splicing reactions.

ii)_Transcription._

The _in vitro_ systems all give correct initiation. The polymerase III systems produce a 5' triphosphate moiety, which is indicative of transcriptional initiation. The 5' end of the products of the _in vitro_ transcription of adenovirus-2 late genes by polymerase II has been sequenced and shown to be identical to the _in vivo_ transcripts (Weil _et al_, 1979).

By deleting DNA from the 5' end of a _Xenopus_ 5S RNA gene, it was shown that not only could all the evolutionarily conserved elements at the 5' end be deleted without hampering gene function, but also the first 50bp of the gene could be removed before synthesis of a 5S-sized RNA ceased. (Sokanju _et al_, 1980). The 3' end of this centrally placed control element was identified at about 90bp from the 5' end of the gene (Bogenhagen et al, 1980). It seems, however, that this unexpected location of the "promoter" may be limited to polymerase III transcribed genes.

Similar deletions from the 5' end of the chicken conalbumin gene indicate that the only absolute sequence requirement is the Hogness box; all DNA 5' of this is dispensible, which includes, of course, the conserved sequences at -70bp (Corden et al, 1980). These same workers show that substitution of the T at the third position in the box (TATAAA) with a C (TACAAA) also eliminates its function.

On the other hand, Grosscheld and Birnstiel (1980), using an oocyte microinjection assay, suggest that the Hogness box of sea-urchin histone genes can be removed without completely abolishing transcription. They also find that there is a negative control element, upstream of the Hogness box, the removal of which increases transcription of the gene.

Several groups have looked for correct initiation of genes introduced into cells by the cotransformation method. Breatnach et al (1980) find no faithful initiation on a chicken ovalbumin gene introduced into a mouse cell, although the transcript is correctly spliced. Dierks et al (1980) find 60% of the transcripts of a rabbit gene transformed into a mouse cell-line are correctly initiated. If the DNA upstream of -70 is deleted, this proportion is reduced to 25%, and if the Hogness box is removed there are no correct initiations.

The controversy presented here will only be resolved if the relationship between *in vitro*, microinjected, chromosomally integrated, and real gene expression is resolved.

1.3 Globin Gene Structure and Evolution.

A large body of data accumulated over the years on the haemoglobin proteins of humans permitted some conclusions to be made regarding the genetics of haemoglobin before molecular genetic studies were possible. The new genetics, therefore, quickly took up the globin system as an easily studied model for gene structure and function.

The structure of the globin proteins from different species allows them to be assigned to two families, β-like and α-like globins. Each haemoglobin molecule has four subunits, two from each family. The families are clearly homologous and diverged by evolution from a stem molecule over a period of 500 million years. The myoglobin molecule is likewise related, and diverged by 1000 million years.

a) Globin Gene Linkage and Duplication.

It has been known for many years that the human globin genes are duplicated, and that the families form close linkage groups. The major adult globin proteins are α- and β-globin. In addition there is a minor β-like protein known as δ-globin. Foetal haemoglobin is made up of α- plus β-subunits, and there are embryonic subunits; ε- and ζ-globin.

Where protein studies have been possible, good evidence has been found for gene duplication. The close similarity between β- and δ-globin suggests that they are the products of recently duplicated genes. That the adult α-globin genes are also duplicated was known because individuals are found who have three different α-globin alleles; for example Hb Buda, Hb Pest and Hb A. Every individual has two different γ-globin genes, differing at amino acid 136. ${}^{G}\gamma$-globin has glycine at this position whilst ${}^{A}\gamma$-globin has alanine. The embryonic globins, ε and ζ, are expressed so early in embryonic life that very little was known about them until recently.

Evidence suggesting the linkage of members of the β-globin gene family came initially from two abnormal haemoglobins; Hb Lepore and Hb Kenya. Both of these proteins are fusion polypeptides with a COOH-terminus derived from β-globin and a NH2-terminus from δ- and γ-globin respectively. (Baglioni, 1962; Huisman et al, 1972). They appear to be products of fused genes formed by unequal recombination between chromosomes. Thus the γ- and δ-genes are to the 5' side of the b-gene.

i)Duplication of Globin Genes.

The postulated gene duplications were confirmed directly with the advent of the new genetics. Furthermore, cross hybridisation has allowed the study to be extended to the α- and β-globin genes of the primates. In these species it is found that the loci are all duplicated, as in humans, except in the lowest primates which have single genes at each locus (Zimmer et al, 1980; Martin et al, 1980; A.J.Jeffreys, pers. comm.). The duplication event thus seems to be fairly recent in mammalian evolutionary history.

Some observations appear contradictory to this conclusion. Firstly, the sequence similarities between the two human γ-globin genes (Slightom et al, 1980) imply that the duplication event is even more recent than the primate study shows. Furthermore there is a polymorphic variation, detectable by its effect on a restriction enzyme site (Jeffreys, 1979), which is present in both γ-globin genes. These can be accounted for if information is transferred between the two genes during evolution by some unknown mechanism.

Although the rabbit β-like globin genes appear to be evolutionarily more primitive; there only being one adult-type gene, this is not the case in the mouse. Mice have a human-like adult β-globin pattern, (with a major and minor species), have duplicated α-globin genes and have two embryonic genes (Jahn et al, 1980). Duplication of globin genes would appear to be a frequent occurrence but not a necessary requirement in mammalian genomes.

ii) Linkage of Globin Genes.

Globin gene linkage has recently been shown both in cloned genes and in total DNA by Southern blotting. Constructing a restriction enzyme map around the genes using Southern blot hybridisation allowed gene linkage to be demonstrated between the human δ- and β-globin genes (Flavell et al, 1978b; Mears et al, 1978), the two γ-globin genes, $^{G}\gamma$- and $^{A}\gamma$ (Little et al, 1979), the γ-locus and the δ/β -locus (Bernards et al, 1979a) and the two α-globin genes (Orkin, 1979). All these linkages have been confirmed by gene cloning, and extended to include the embryonic globin genes of both the α- and β-families (Fritsch et al, 1980; Lauer et al, 1980; Maniatis et al, 1980).

The human β-globin linked family, therefore, consists of five functional genes. They are all transcribed in the same direction and are arranged in the order predicted from studies of the fusion polypeptides:

$$5' \quad \varepsilon \quad {}^{G}\gamma \quad {}^{A}\gamma \quad \delta \quad \beta \quad 3'.$$

The δ- and β-genes are about 7kb apart, the two γ-genes are 5kb apart, the ${}^{A}\gamma$ and δ-globin genes are separated by 15kb, and similarly the ε and ${}^{G}\gamma$-genes are 15kb apart.

The two human α-globin genes are 3.7kb apart, and are 9kb to the 3' side of a ζ-globin gene. This is itself 12kb from a second ζ-globin gene (Lauer et al, 1980). Although it is not known whether both the ζ-genes are expressed, this is believed to be the case as the 3' gene has been deleted in some forms of thalassaemia whilst still allowing some ζ-globin expression (Pressley et al, 1980). The sequence of the 3' gene suggests that it also is functional (Maniatis et al, 1980).

It is interesting that the genes are arranged in the order of their expression during development, and that the co-expressed genes are more closely linked than those excluded in expression.

Cloning has also demonstrated a similar type of linkage between β-globin genes in mouse and rabbit (Jahn et al, 1980; Hardison et al, 1979).

Recent work on the linkage of globin genes of the Xenopus species X.laevis and X.tropicalis have interesting implications for globin gene evolution (Jeffreys et al, 1980; Patient et al, 1980). Both mapping by Southern blotting and gene cloning showed that in X.laevis the α-major and β-major globin genes are linked, and are seperate from the linked α-minor and β-minor genes. X.tropicalis appears to have only one linked pair of α/β globin genes. The implication is that the primordial globin gene first duplicated to give a tandem repeat, which diverged to the α- and β-polypeptide structures. Sometime before mammalian evolution the genes were seperated by a chromosome translocation. X.laevis arose by tetraploidisation from the single globin-pair species.

b) Globin Gene Sequences and Intervening Sequences.

As there are more members in the β-globin families, and their developmental switches occur at more easily studied points, recent work has tended to concentrate rather more on this family.

Members of the β-globin gene family have been isolated from humans, rabbit, mouse, chicken, goat, sheep and Xenopus laevis (Blattner et al, 1978; Lawn et al, 1978; Ramirez et al, 1979; Proudfoot and Baralle, 1979; Fritsch et al, 1980; Van den Berg et al, 1978; Lacy et al, 1979; Tilghman et al, 1977; Tiemeier et al, 1978; Dodgson et al, 1979; Ginder et al, 1979; Robbins et al, 1979; Kretschmer et al, 1980; Patient et al). The β-globin genes have been "mapped" by Southern blot analysis from further species, including different mouse strains (Weaver et al, 1979) various primate species (A.J. Jeffreys, pers. comm.) and Xenopus tropicalis (Jeffreys et al, 1980)

Representatives from the less-studied α-globin gene family have been cloned from humans (Lauer et al, 1980), mouse (Leder et al, 1978; Blattner et al, 1978; Nishioka et al, 1980) and Xenopus laevis (Patient et al, 1980)

i) Intervening Sequences.

All these genes have two intervening sequences, and in the genes where they have been closely examined, they are in the same position. Several genes from both α- and β-globin families have been sequenced to reveal the exact position of the IVSs. The two IVSs are termed IVS-1 and IVS-2, and clearly have been present in the genes since before the divergence of the α- and β-globin families.

IVS-1 is located in all globin genes at approximately the codon for amino acid 30. There is a repeated sequence, so the placing is not exact, but if Chambon's rule is applied IVS-1 divides the codon between the second and third bases. Although the codon number differs from species to species and between the two families, it codes for a homologous arginine residue in all cases.

Tabulated below are the sizes of IVS-1 in genes characterised;

Human α-family	95bp	Lauer et al, 1980.
Human β-family	125-150bp	Maniatis *et al*, 1980.
Mouse α-gene	123bp	Nishioka *et al*, 1980.
Mouse β-family	116bp	Konkel *et al*, 1979.
Rabbit β-family	100-126bp	Van Ooyen *et al*, 1979; Hardison *et al*, 1979.
Chicken β-gene	c100bp	Dodgson *et al* 1979.
X.laevis α-gene	c150bp	
X.laevis β-gene	c185bp	Patient *et al* 1980.

The similarities of size of the intervening sequences are quite remarkable, especially in genes diverged for 500 million years, given the wide range of sizes of different IVSs in other genes. Whether this indicates that the sizes (but not the sequences) are a result of evolutionary conservation through necessity or is some indication of an evolutionary mechanism is not known. The same conservation does not apply across the board for IVS-2.

IVS-2 is also in a homologous position in all globin genes; between codons 99 and 100 in all α-globin genes, and between 104 and 105 in mammalian and 103 and 104 in *Xenopus* β-globin genes. In α-globin genes the IVS is small (in all cases larger than IVS-1 but not by much), whereas in β-globin genes the size of IVS-2 is up to 6 times larger than IVS-1. Again the data are tabulated below.

human α-genes	125bp	)
mouse α-gene	134bp	)
Xenopus α-gene	c310bp	)
human β-family	c700-900bp	) references as above
mouse β-family	628-625bp	)
rabbit β-gene	573bp	)
chicken β-gene	c800bp	)
Xenopus β-gene	c930bp.	)

Once again the significance, if any, of these observations is not apparent. It should be noted that the more evolutionarily distant species are less in the mammalian pattern. The *Xenopus* α-globin gene has almost a "β-like" large and small IVS.

Sequence comparisons between the intervening sequences of several β-globin genes have been performed. Two alleles of the rabbit adult β-globin have been sequenced and their intervening sequences are identical (see below). The two mouse adult β-globin genes, βmaj and βmin, were compared by Konkel *et al* (1979), who showed that although the coding sequences were 96% homologous (indicating 50 million years divergence), and the two IVS-1 had 97% homology, there was less homology in IVS-2. The 250bp at the 5' end of IVS-2 had 86% homology, but the rest of IVS-2 shared far less sequence. Van Ooyen *et al*, (1979) compared the major adult β-globin genes of rabbit and mouse. Although the coding potential of the genes is homologous at 81% of the positions, IVS-1 is only 53%, and IVS-2, 40% homologous.

Clearly the DNA sequences of the IVSs are not as conserved as the coding sequence, although there does seem to be some conservation of size and of limited regions of sequence. Signals required for the splicing of the primary transcripts must be preserved, but it is doubtful whether the extent of the conserved sequences reflect this. It is not possible to distinguish between conservation due to a functional requirement of the sequence and fortuitous conservation due merely to the mechanism of evolution or DNA mutagenesis.

ii) Coding Sequences.

The constraints which act to conserve the sequence of the globin proteins will, of course, also serve to maintain homology between the coding sequences of the different genes. As the third base of the codons is usually not essential to the meaning of the codon then evolution acting only on the protein should not conserve these third position bases. However, sequencing of two allelic β-globin genes from rabbit and the βmaj and βmin globin genes of mouse indicate that third position bases are conserved.

Van Ooyen et al (1979), and Hardison et al (1979) sequenced two alleles of rabbit β-globin. There are four amino acid differences between the two genes. All four differences are due to single base changes, and not only are there no silent, third position base changes but the two IVSs of both genes are identical. The mechanism by which this apparently highly unusual situation has arisen will be interesting to determine.

Similar conservation of coding sequence has occurred in the mouse major and minor β-globin genes (Konkel et al, 1979), in which there are 17 differences between the two genes' coding potential. Nine of these cause eight amino acid changes, and there are eight silent changes. The rate of mutation at third base positions is virtually the same as at the other two positions, which suggests that selection pressure is equally great on all bases in the mRNA, despite the rigours of protein conservation.

As discussed above, the variation between the IVSs of these genes is rather more than in the coding region. It is therefore likely that evolution is acting on the mRNA structure and function rather than the structure of the gene itself. The function of the mRNA which which is being preserved may be its three dimensional structure, its interaction with proteins or simply the availability of tRNA molecules in a particular cell type. A comparison of the 5' and 3' untranslated regions of the mRNAs shows that the 5' regions of the mouse maj and min mRNAs are 86% homologous, and the 3' regions are 32% homologous, compared to 96% in the translated region. This would suggest that it may be tRNA availability which is the important factor.

c)Globin Pseudogenes.

The term "pseudogene" was first coined by Jacq et al (1977), who found a truncated 5S rRNA gene in a cloned repeat unit of the *Xenopus* 5S gene family. This gene was believed to be inactive *in vivo*. Proudfoot and Maniatis (1980) have redefined the term so that it applies only to those genes with sufficient homology with expressed genes to be recognisably related, but with structural alterations which will prevent their expression.

Cloning of the globin gene families has revealed sequences which meet this definition interspersed among the functional genes. Two pseudogenes have been found in the human β-globin gene family (Fritsch *et al*, 1980); one is between the δ-globin and the $^{A}\gamma$-globin genes and the other is on the "far" (5') side of the ε-globin gene. An α-globin related pseudogene has been located in humans between the α- and ζ-gene pairs (Lauer *et al*, 1980). The rabbit β-globin locus is known to have one pseudogene (β2) between β1, the adult gene and β3, the foetal gene (Hardison *et al*, 1980). There is a pseudogene in an analogous position in the mouse β-locus, between the adult and the foetal genes (Jahn *et al*, 1980). Finally, a mouse α-globin pseudogene has been isolated, but its location with respect to the other mouse globin genes has not been ascertained (Nishioka and Leder, 1980; Vanin *et al*, 1980).

The sequences of the human β-pseudogenes have not yet been determined, but Proudfoot and Maniatis (1980) have sequenced the human α-globin pseudogene. The pseudogene has the putative trancriptional initiation signals at -70 (CCAAT) and -25 (CATAAG), although its transcriptional function *in vivo* or *in vitro* has not been determined. That the gene really is non-functional is demonstrated by the presence of small insertions and deletions which change the reading frame of the globin message. Furthermore neither IVS has the presumptive splicing signals; both have AG at the 3' end but neither have GT at the 5' junction.

The rabbit β2-pseudogene similarly has deletions which alter the reading frame so that it cannot code for a globin polypeptide. The gene has the "ATA box" at -25bp with respect to the 5' end of the mRNA, but is missing the "CCAAT box" at -70. Like the human α-pseudogene this gene has alterations at the junction of the IVS, and like the human pseudogene both the IVSs are normal at one end (in this case the 5' end) but defective at the other.

Whilst the human and rabbit pseudogenes have two intervening sequences in the expected positions, a mouse pseudogene sequenced by Nisioka and Leder (1980) and Vanin *et al* (1980) has no IVSs. The gene has the structure of a DNA copy of a mRNA; the absence of the intervening sequences does not alter the continuity of the coding sequences. There are, however, small deletions and insertions which render the message unintelligible as globin. Given that all globin genes examined have homologous IVSs, the best conclusion is that this gene has somehow had the intervening sequences deleted. It is possible that the deletion was mediated by the same mechanism which removes the intervening sequences from mRNA-precursors. It is unlikely that the pseudogene originated as a reverse transcript of the mRNA because the 5' flanking sequences, which are not in the mRNA are clearly in this pseudogene. A possibiblity, put forward by Leder *et al* (1980), is that the deletion was produced by a chance formation of a heteroduplex between an mRNA molecule and the gene (possibly during replication). Enzyme action could then convert the DNA strand to a copy of the RNA. Not all genes in which IVS-deletion has occurred will necessarily be pseudogenes; Lomedico *et al* (1979) describe an active rat insulin gene from which one IVS has been removed.

The TATAA sequence at -25 in functional genes appears in this pseudogene as TAGAA. Corden *et al* (1980) showed that changing the sequence in front of the conalbumin gene from TATAA to TACAA eliminated its function as a promoter of *in vitro* transcription. It is likely that this -globin pseudogene will be non-functional *in vitro* also. Vanin *et al* (1980) fail to detect any transcripts from either DNA strand *in vivo*. This makes rather fanciful their suggestion that the pseudogene might be transcribed in the opposite orientation to act as an "anti-transcript" to somehow regulate -globin gene expression.

Maniatis *et al* (1980) point out that the pseudogenes located so far all lie between functional clusters of genes, that is between the foetal and adult gene sets. They may serve an as yet unidentified function, or they may be relics of the continuing evolution of the genes. One possibility is that they act as blocks to prevent the information exchange which undoubtably occurs between adjacent, duplicated genes spreading to other genes in the cluster.

d)Repetitive DNA in the Globin Loci.

The repetitive nature of much of the DNA in the genome of higher eukaryotes has been studied for some time using a reannealling approach (reviewed by Lewin, 1974). A large proportion of the highly repeated DNA is interspersed with the unique sequence DNA (see Davidson and Britten, 1979). It was only with the advent of the new genetics that the arrangement of repetitive sequences with respect to well-characterised gene clusters could be studied. Although some studies have been made of other loci (insulin, for example, Bell et al, 1980) the β-globin gene families of rabbit and human have had the most intensive examination.

Using an imaginative adaptation of the Southern blot procedure, Flavell et al, (1978a) and Hoeijmakers-Van Dommelen et al (1980) initially described a highly repeated sequence to the 3' side of the rabbit β-globin gene. By means of a combination of conventional Southern blot hybridisation, a "two-dimensional" blot hybridisation and electron microscopy, Shen and Maniatis (1980) identified at least 20 repetitive elements distributed throughout the 49kb of the cloned rabbit β-globin gene family. Much of this repetitive DNA occurs as repeats or inverted repeats flanking individual genes, or groups of genes or the whole family. One family of repeats mainly occurs around the foetal/embryonic gene pair (β3/β4) and another family is mainly located around the adult gene (β1), although this family also flanks the whole locus. The elements of the families are repeated at least several hundred, and probably more, times throughout the genome.

Fritsch et al (1980) and Coggins et al (1980) identified one particular repeated unit occurring at six locations in the human β-globin gene family. The sequence occurs in pairs flanking the adult globin genes, the foetal gene pair, and to the 5' side of the embryonic globin gene. Elsewhere in the genome this sequence occurs many thousands of times. Houck et al (1979) showed that most of the highly repeated DNA which is interspersed with unique sequence DNA belong to one major 300bp family, the "Alu-family". It is likely that the repeated elements in the β-globin locus belong to this family. Duncan *et al* (1979) characterised three homologous regions around the γ-δ-β-globin genes, in approximately the same position as those described by Fritsch *et al*, two of which are transcribed *in vitro* by RNA polymerase III. These transcripts have been shown to be related in sequence to the Alu-family.

Jelenik *et al* (1980) draw together a large body of work which, it has become clear, all relates to the Alu family. In addition to the Alu-repeat being transcribed by RNA polymerase III *in vitro*, the small nuclear RNAs seen in all cells are also clearly related to it in sequence. The sequence is also seen in inverted repeats of RNA covalently attached to the hnRNA of chinese hamster ovary cells. Lerner *et al* (1980), and Rogers and Wall (1980) point out homologies between some small nuclear RNAs and the splice points involved in the removal of the intervening sequences from mammalian genes. They suggest that the snRNAs are part of the splicing mechanism. It is possible that each gene set has its own set of splicing RNA genes, which are activated when the gene set is activated and these might be seen as repetitive DNA elements.

On the other hand, Jelenik *et al* also point to homologies between the Alu-repeat and sequences at the origin of replication of four eukaryotic viruses. The repeats may function as replication origins in the mammalian genome, their presence in hnRNA being merely fortuitous, as the RNA polymerase II transcription unit might include the repeat.

The transcription by RNA polymerase III and the snRNAs may be part of the control or the mechanism of DNA replication.

e)The Old Genetics and the New Genetics.

The background supplied by classical genetic studies has resulted in the globin genes becoming a major focus for the new genetics. In the next section I describe the recent work on the molecular nature of the thalassaemia syndromes, and discuss some models which may apply to the regulation of globin gene expression and to the switches in gene expression which take place during development.

The new genetics is thus able to meet the old genetics in providing a rationalization of the previous observations. So, the so-called "C-value paradox", or why there is far more DNA in the genome than the genes require, can be at least rationalised, if not understood, in terms of intergenic DNA and intervening sequences. Similarly the observation of large RNA molecules in the nucleus, and of a large amount of nuclear-confined sequence can be viewed in terms of pre-spliced mRNA precursors. The wide range of repetitive elements are being characterised and possible functions ascribed to them. However, these new answers provide far more new questions themselves, as more and more hitherto unexpected aspects of gene structure and function come to light.

1.4 The Thalssaemia Syndromes.

The hereditary imbalances and deficiencies in globin synthesis are collectively known as the thalassaemia syndromes (Weatherall and Clegg, 1972; 1979). The availability of probes for the globin genes and their transcripts has led to a large research effort aimed at characterising and understanding the molecular nature of these syndromes. The aim is not only to understand the nature of genetic disease at the molecular level but also, it is hoped, that a study of defective genes will give insights into the way in which normal genes are expressed.

The disorders are broadly classified into α- and β-thalassaemias; both have uniquely interesting features, and both share some common principles.

a) Alpha-Thalassaemias.

There are four α-globin genes per diploid cell. Thus the observation that there seem to be four forms of α-thalassaemia can be rationalised by postulating that each form of increasing severity is due to the failure of expression from increasing numbers of genes. The most serious form of α-thalassaemia is hydrops foetalis, which is invariably perinatally fatal, and is characterised by a complete absence of α-globin synthesis.

Using solution hybridisation Ottolenghi et al (1974), Taylor et al (1974) and Kan et al (1975a) demonstrated that no α-globin genes were detectable in DNA from hydrops foetalis individuals and that hybridisation to DNA from patients with one active gene showed one quarter of the hybridisation to normal DNA. Thus it was suggested that the cause of α-thalassaemia was the deletion of one or both α-globin genes from the chromosome, and that these two chromosomes in different combinations with each other and with normal chromosomes cause the whole range of α-thalassaemia syndromes.

Southern blot hybridisation has since been used to examine the deletions. The chromosomes which had a single α-globin gene could be shown to have two types of deletion (Orkin _et al_, 1979a; Embury _et al_, 1979; Embury _et al_,quoted in Maniatis _et al_, 1980). One of these deletes 4.2kb, and includes the 5' α-globin gene. The other deletes 3.7kb by a fusion of the 5' end of one to the 3' end of the other α-gene. Some of the chromosomes which were lacking two α-globin genes give no hybridisable fragments, whilst others contain only the 3' part of an α-gene, which, of course, is not functional (Orkin _et al_, 1979; Orkin and Michelson 1980).

Analysis of large numbers of patients with this technique has allowed some conclusions to be drawn about the nature of the disease in different populations. The two-functional-gene form, α-thalassaemia 1 could be due to homozygosity for a one gene chromosome, or heterozygosity for a zero gene chromosome. Analysis of a number of Black α-thalassaemia-1 patients showed that the disease is almost invariably due to homozygosity for the one gene locus (Dozy _et al_, 1979; Higgs _et al_, 1979). The very low frequency of a wholly deleted α-locus means that hydrops foetalis, caused by two deleted chromosomes coming together, is virtually unknown. A similar situation is seen in Mediterranean populations, although the frequency of the zero-gene chromosome is a little higher (Kan, personal communication). By contrast, Asian α-thalassaemia-1 may be due to either a homozygous or a heterozygous gene combination, and consequently hydrops foetalis is less rare.

Both solution hybridisation (Kan et al, 1977) and Southern blotting (Orkin et al, 1979) have demonstrated that there are some α-thalassaemic loci which have no demonstrable deletion of DNA. This type of lesion, in which some very localised mutation has occurred to affect expression is rather more common at the β-locus and is discussed below.

b) Beta-thalassaemia.

As there is only one β-globin gene per haploid set, the characteristics of particular β-thalassaemic loci are better understood than the α-thalassaemic leisions. There are two broad groups into which it has been found convenient to place β-thalassaemias. The β^+-thalassaemias are characterised by low levels of β-globin in the red cells, wheras in β^o-thalassaemia there is a complete absence of the β-globin protein.

i) β^+-thalassaemia.

The reduced levels of normal β-globin in β^+-thalassaemia correlate with low levels of mRNA, which implies that the mRNA is functionally normal. Furthermore, the mRNA can be translated normally *in vitro* (Benz et al, 1978). Southern blot hybridisation shows that there is no deletion of the DNA flanking the β-globin gene (Flavell et al, 1979; Orkin et al, 1979b). The reduced mRNA level may be due to a reduced rate of transcription, to a reduced rate of processing or transport of the mRNA, or to an accelerated rate of cytoplasmic degradation.

Nienhuis et al (1977) showed that, in at least two out of three patients examined, although the cytoplasmic levels of β-globin mRNA were markedly lower than the level of α-globin mRNA in the cytoplasm, the concentrations of nuclear α and β-globin RNA were about equal. This points to the defect being at the post-transcriptional level. More recently Maquat et al (1980) and Kantor et al (1980) have examined the β-globin specific nuclear RNA from β^+-thalassaemic patients in the light of present knowledge of RNA splicing.

Maquat et al used pulse/chase experiments to follow the fate of the globin primary transcripts. In bone-marrow cells from patients with β^+-thalassaemia the counts were not chased into mRNA-sized molecules as quickly from β-globin transcripts as from α-globin transcripts in the same cells, or β-globin RNA in normal cells. Kantor et al found two different patterns of abnormal processing of β-globin mRNA in two different patients. In one patient the bone marrow cells appeared to accumulate a β-globin RNA which contained a partially removed IVS-2 transcript. It is believed that IVS-2 of mouse β-globin transcripts is removed in stages (Kinniburgh and Ross, 1979) and an RNA of the same size as the one accumulated in this patient is seen at low levels in normal cells. Thus this patient probably has a block in the normal splicing pathway. The other abnormal RNA seen in this study was a mRNA-sized molecule, but which still contained some IVS-2 sequence. Kantor et al suggest that this is an abnormal degradation product. It is interesting to note that these two patients show different abnormal patterns. There is obviously heterogeneity in the disease when it is examined in fine detail.

It is probably not true to say that all β^+-thalassaemia genes are characterised by abnormal processing. There are possibly some β-globin genes which have normal rates of processing but a reduced rate of transcription, but these have not yet been identified.

ii) β^o-thalassaemia.

Southern blotting reveals that in contrast to most α-thaassaemia gene, in most of the cases examined of β^o-thalassaemia, the β-globin gene is still present. However, three patients (all of Indian ancestry) were shown to have a 600bp deletion, which removes a part of IVS-2, the coding DNA adjacent to it, and 150bp of the 3' flanking sequences (Flavell et al, 1979; Orkin et al, 1979b; Orkin et al, 1980). (In all these individuals, only one of the two thalassaemic genes was of the deletion type). In the main, however the thalassaemic defects must involve very small sequence alterations or deletions operating from a long-range, where the DNA was not probed.

Old et al (1978) found three groups of β^o-thalassaemic defects. In red blood cells of the first type there is an absence of mRNA, in the second is RNA which is not translated and in the third there is a partial RNA transcript. This deliniation, however, is an oversimplification. Comi et al (1977) describe a gene which produces nuclear, but not cytoplasmic, β-globin RNA (see section 3.4a). Conconi and Del Senno (1974) suggest that there is a form of the disease characterised by the presence of β-globin mRNA which is inactive in vivo, but which is translated in vitro. As gene expression is a complex, multistage, process there are obviously many points at which defects can occur. Until the basis of gene expression is known more precisely it may not be possible to pinpoint a precise leision in many cases.

The precise molecular defect of one β^{o}-thalassaemia is known (Chang and Kan, 1979). By sequencing the β-globin mRNA, present at low levels in this patient, these workers detected a single base change at codon 17 which produced a nonsense codon. That this is the only defect was shown by translating the mRNA in vitro using a yeast suppressor tRNA to overcome the misplaced stop codon (Chang et al, 1979). The reduced level of mRNA seen in the blood cells is presumably due to degradation of the unused message.

c) Deletions in the β-Globin Locus: Long Range Effects.

A number of deletions have been found in the region of the β-globin gene family and most seem to have a long-range effect on the expression of other genes in the locus.

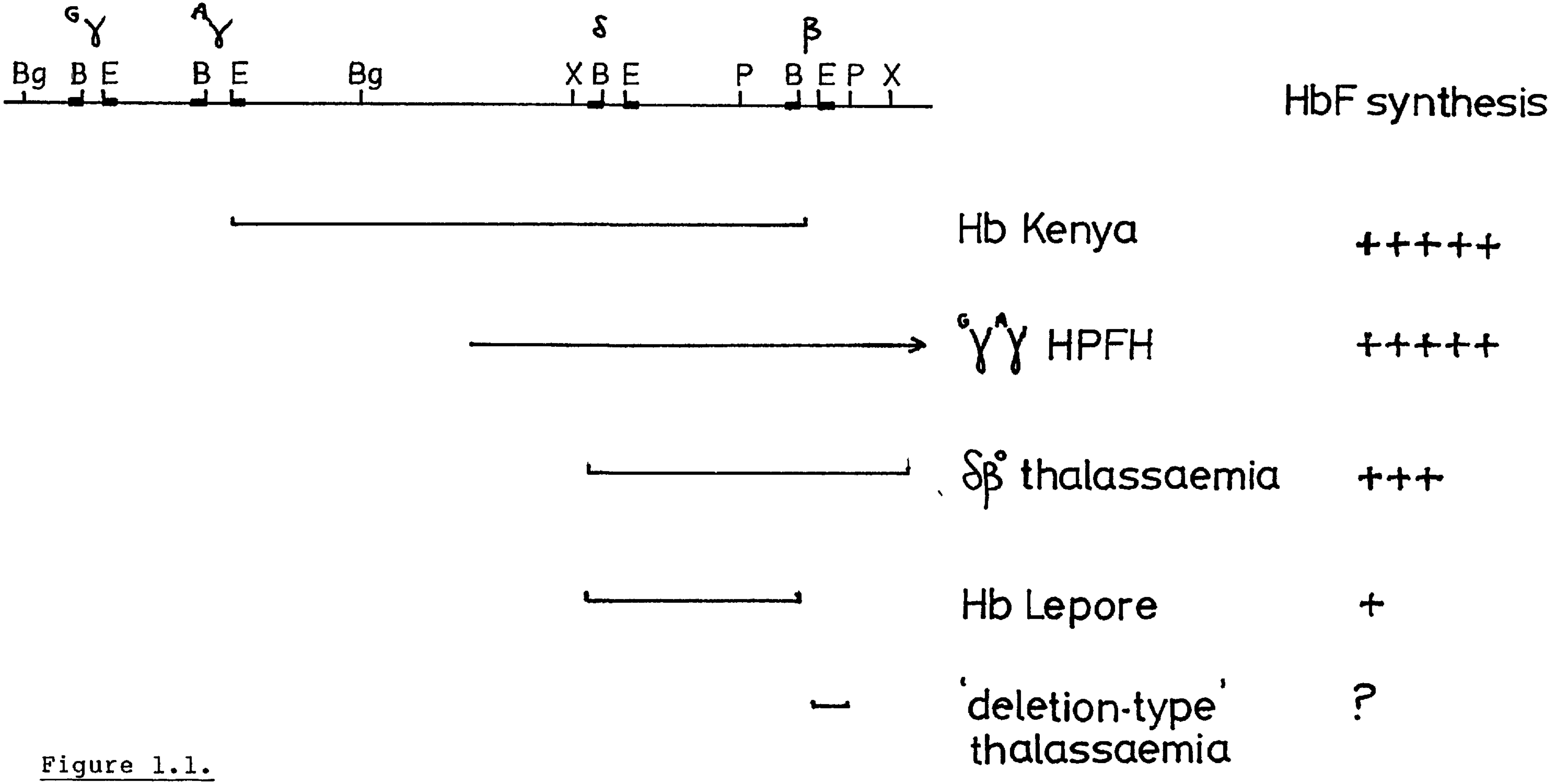
The switch from γ-globin to δ/β-globin gene expression around birth is a developmental switch which has long been considered tractable. The repression of β-globin synthesis is not absolute; all individuals express low levels of the protein to a variable degree, the extent of which may be inherited (Zago et al, 1979). There are several globin gene defects where the switch does not, to a greater or lesser extent, operate.

In hereditary persistence of foetal haemoglobin (HPFH) there is no synthesis of δ- or β-globin, but the defect is not pathological as the deficiency is made up by synthesis of γ-globin (Wood *et al*, 1979). Solution hybridisation, (Kan *et al*, 1975b; Forget *et al*, 1975) and later mapping of the genes, demonstrated that the in some forms of HPFH both the β and δ-globin genes are deleted, along with some of the sequences between the γ- and δ-genes (Fritsch *et al*, 1979; Tuan *et al*, 1979). This suggests that some of this DNA is involved in the to switch mechanism, and its deletion causes a malfunction of the switch.

Although no gene map of Hb Kenya is available, its structure (a fusion peptide with a C-terminal β-globin and an N-terminal γ-globin sequence) implies that it is a gene fusion, causing the deletion of the to intergenic DNA. There are no homozygous Hb Kenya individuals known, but the heterozygotes have an elevated level of $^{G}\gamma$- plus Hb Kenya (i.e. $^{A}\gamma$), comparable to HPFH heterozygotes. Knowledge of the protein structure alone led Huisman *et al* (1972) to suggest that the intergenic DNA was involved in the switch.

The deletion which causes $\delta\beta^{0}$-thalassaemia includes the β-gene but does not extend beyond the middle of the δ-gene (Fritsch *et al*, 1979; Ottolenghi *et al*, 1979; Bernards *et al*, 1979b). Heterozygotes for this defect have raised levels of HbF, but it is not sufficiently high in the homozygotes to compensate for the deficiency of β-globin synthesis. Similarly carriers of the δ/β fusion protein, Hb Lepore, in which the δ/β-intergenic DNA is deleted, have mildly raised HbF (Marrinucci *et al*, 1979). These data are summarised in Figure 1.1.

If a specific sequence acting as a switch is sought, then it is necessary to suggest that it lies between the δ-



$^{G}\gamma$
$^{A}\gamma$
δ
β
Bg B E B E Bg X B E P B E P X
HbF synthesis
Hb Kenya
+++++
$^{G}\gamma^{A}\gamma$ HPFH
+++++
$\delta\beta^{\circ}$ thalassaemia
+++
Hb Lepore
+
'deletion-type' thalassaemia
?

Figure 1.1.

Schematic representations of some deletions in the β-like globin gene region, and their effects on expression of the γ-globin genes. Levels of HbF synthesis are on an arbitary scale, and are by way of illustration only.

and β-genes. However, an explanation for the different levels of derepression of the γ-genes in the different disorders is needed.

The gradation of increasing derepression of γ-globin function, correlated with increasing size of deletion of DNA in the locus, leads to an alternative mechanism. One can postulate that the regulation of differential gene expression relies on the structure of the chromosomes, rather than the DNA sequence itself, and that a deletion of any DNA disrupts the fine tuning of gene control.

Van der Ploeg et al (1980) describe a γ/δ/β-thalassaemic locus in which the γ- and δ-genes are deleted, but the β-gene is still intact. The possibility of a second mutation which removes the β-gene function has not been eliminated, but it is probable that the β-gene is structurally normal, and the deletion of DNA 3kb away is exercising a long-range effect on its expression. The deletion includes the hypothetical "switch", but an effect due to chromosomal disruption may be more likely. This may be a general effect on the structure of the locus, or it may be due to the deletion of a specific chromosomal organiser region, which sets up chromosomal "domains" involved in co-expression of adjacent genes.

d) Perspectives.

As our understanding of gene function improves, so it becomes clear that the defects in the various thalassaemia syndromes occur at many, if not all, of the stages beween DNA and protein.

Gene deletion is a very obvious means by which gene function may be lost. Deletions are commonly seen at the α-locus, eliminating one or both genes, but are somewhat rarer at the β-locus. The two types of single α-gene deletion can be explained in terms of unequal recombination involving two different regions of homology. These deleted loci are also seen when a recombinant λ-phage containing the β-genes are grown in bacterial cells (Lauer et al, 1980). This mechanism of gene deletion predicts that three α-globin genes should be generated on some chromosomes. Such triple-α loci are indeed found in both thalassaemic and non-thalassaemic populations (Goosens et al, 1980; Higgs et al, 1980).

Deletions in the β-globin family are much rarer, probably because there are no extensive regions of homology around the genes. It will be interesting to see if the break points of the deletions characterised do correlate with any of the sequences repeated within the β-locus.

The defects of the non-deletion β-thalassaemic genes are more easily studied than similarly malfunctional α-thalassaemic genes because a lack of expression from both β-genes is still compatible with life. Thus detection of β-globin RNA in a homozygous β^{o}-thalassaemic individual is definitive proof that a non-expressed gene is being transcribed, whereas this is impossible to assess in an α-thalassaemic patient. It is this relative ease of analysis which allows one to say that defects of β-globin gene function follow Murphy's Law; that is, as far as the population as a whole goes, what can go wrong will go wrong.

It is hoped that an understanding of the molecular lesions in the thalassaemias will lead to a better understanding of gene function. As described in section 1.1 the new genetics aims to determine the mechanics of gene expression by mutagenesis of genes in vitro. A fundamental drawback of this approach is implicit in the name sometimes given to the process; site-directed mutagenesis. Mutagenesis which is directed can only answer questions which are within our frame of reference. The defects in thalassaemia have not been directed, and therefore may indicate additional regions of importance in gene function. Two cases in point are the γ-δ-β-thalassaemia analysed by Van der Ploeg et al (1980) and the various gene-switching defects also described above. Both these types of defect act over a long range, and serve to draw our attention away from the type of control elements which prokaryotic molecular genetics has led us to expect.

These discoveries lead one to question the rationale of the main part of this thesis; that is, is the β-thalassaemic locus the same as the β-globin gene? Weatherall and Clegg (1972) bring together data of matings between normal (A/A) individuals and persons doubly heterozygous for β-thalassaemia and sickle cell disease. All children of these matings should, if the β-thalassaemia and β-globin genes are allelic, be either β-thalassaemia or sickle cell disease carriers. Out of 62 children only one clearly carried a chromosome which was the product of recombination so that the thalassaemia and the sickle cell determinants were brought into physical linkage. Leaving aside the fact that the β-thalassaemias are very heterogenous, it is clear that if the two genes are not truly allelic, they are very closely linked. The isolation of a β-globin gene and several thousand base pairs of its flanking sequences represents the "best-guess" strategy towards identifying the molecular defect preventing expression of the gene.

1.5 Interferon.

During the course of the work on globin gene isolation presented in this thesis an association with Searle Research led to a collaborative project. The commercial potential of interferon as a drug induced G.D. Searle to begin a project to isolate interferon genes with a view to obtaining a ready source of the protein. The experience aquired during the globin work was readily applied to this project. In the last section of the introduction I will briefly review the production of interferon, and describe recent findings of the structure of interferon genes.

Since the discovery of interferon in 1957 (Isaacs and Lindemann, 1957), public interest in the antiviral properties of the protein has been inconsistent and fuelled mainly by the media. The recent suggestions that interferons might act as anti-cancer drugs has, predictably, elicited the loudest and most unconsidered responses. Until recently the only interferon available in sufficient quantities for clinical trials was produced in Finland at the Central Public Health Laboratories by Dr. K.Cantell (Cantell, 1979). Interest is now at such a level that other groups are beginning to take up production. Burroughs-Wellcome and G.D.Searle are two major commercial concerns which are now involved in large-scale interferon production. Several clinical trials have been carried out to test the anti-viral properties of the drug (Merrigan *et al*, 1978) but few are without problems. No rigorous assessment of its anti-tumour action has been made.

Interferons show strong species specificity. Human interferon works best on human cells, although it does have activity in other primates. There are several different types of interferon produced in any species. The original nomenclature was to name each interferon according to the cell type from which it was made, but recently a new system has been proposed which does not refer to the cellular origins (Stewart et al, 1980). Both systems are defined here, and are used interchangeably.

a) Interferon Production.

Leukocytes are the best known source of interferon (called IFN-α under the new system). The Finnish group use human leukocytes from the buffy coats of blood, supplied by the Blood Transfusion Service, to produce interferon. The large-scale production techniques, although sub-optimal, still process a significant fraction (probably at least 10%) of all blood donations made in Finland.

Continuous culture lymphocyte lines produce lymphoblastoid interferon, which in the new nomenclature is also called IFN-α. The cell-line used, which has the best production levels, is the Namalwa line (Strander et al, 1974). Although the output from the cells is less than from buffy-coat leukocytes, the relative ease with which the cells are grown has prompted the Wellcome Research Laboratories to begin large scale production.

Fibroblast interferon (IFN-β) is induced in tissue-culture fibroblasts by chemical inducers. This is an advantage of this type of interferon production; IFN-β interferons must be induced by Sendai virus. Fibroblast interferon is antigenically different from leukocyte interferon (Havell et al, 1975). G.D.Searle are engaged in large scale production of this form of interferon to supply clinical trials, and have invested large amounts of capital in setting up production on an industrial scale.

Type II interferon, otherwise known as immune interferon or IFN-γ, is produced by leukocytes stimulated with mitogens or antigens. This type (or types) of interferon has attractive properties in that it seems to stimulate the immune system. It is a much more unstable protein than the type I interferons. There is no large scale production of this interferon currently underway.

Little is known about the differences in action or targets between the leukocyte and the fibroblast products. It is attractive to speculate that leukocyte interferon is a mobile, serum-bourne interferon whilst fibroblast is a locally-acting protein, in the same way that leukocytes are mobile and fibroblasts are localised.

Although all three type I interferons have been purified to homogeneity, the preparations currently used in clinical trials contain only 0.1% interferon. Some of the toxic side effects of the drug may be ascribable to the contaminating proteins. However, a recently developed technique should allow pure interferon to be made more easily. Using the methods developed by Kohler and Milstein (1975) a monoclonal antibody was made from a hybrid myeloma cell-line. The antibody was specific for lymphoblastoid interferon (Secher and Burke, 1980). When this antibody was bound to agarose in a column, an interferon preparation passed over the column was purified 5000-fold in a single step, with a recovery of 97%. The product had a specific activity of 1.8×10^8 U/mg, as compared to the estimate of 2.5×10^8 U/mg for pure lymphoblastoid interferon (Zoon et al, 1979).

Industrial-scale production of interferon is believed to be a commercial proposition. The question of the supply required to satisfy demand is not easily answered. Dose levels needed for effective treatment, and the range of diseases treatable with interferon are not known. If the speculated use against neoplastic diseases proves to be a reality, then the demand will be enormous. In order to be able to make large quantities of interferon easily, several large pharmaceutical companies, and several small venture-capital businesses are investigating the use of E.coli carrying interferon genes to manufacture interferon.

b) Synthesis of Interferon by E.coli.

Several groups have reported the isolation of E.coli clones carrying recombinant plasmids with interferon cDNA sequences. The ways by which this was done are described in section 1.1a. By further engineering of the sequences, to place the interferon gene under the control of an E.coli promoter, high levels of expression have been acheived. Nagata et al (1980a) reported that the leukocyte interferon gene placed in the ampicillin resistance gene of pBR322 produced only a few molecules of interferon per cell. On the other hand, Goedell et al (1980a) placed a leukocyte interferon cDNA adjacent to the promoter of the trpE gene, and could isolate mature interferon from the cells equivalent to several thousand molecules per cell. These bacterium produce 2.5x 10^8 U/l. As the annual production of Cantell's laboratory is 2x 10^{11} U. the potential for these E.coli production systems is obvious. A 2% pure preparation of the synthetic interferon was tested in squirrel monkeys, which it protected against infection by EMC virus.

The same group (Genentech) have made a fibroblast interferon cDNA producing strain from which can be isolated 8x 10^7 U/l of interferon (Goedell et al 1980b). Derynck et al (1980b) have also engineered expression of fibroblast interferon, but they obtain about 1000x less than the Genentech group.

One problem which was believed to be associated with the production of interferon by E.coli may not be serious. The absence of glycosyl groups from proteins made in bacteria was thought to possibly reduce the effectivness of the interferons. Success in in vivo trials of the interferon produced by Goedell et al (1980a) remove some doubts. The unmodified protein may of course be antigenic in man. However, Allen and Fantes (1980) find no evidence that IFN-α molecules really are glycosylated. The evidence is rather better that IFN-β is glycosylayed (Tan et al, 1979).

c) Interferon Gene Families.

By comparing the cDNA sequences Tanaguchi et al (1980b) showed that the fibroblast and leukocyte interferons are clearly related. They have probably arisen from a common ancestor by divergence over a period of between 500 and 1000 million years.

Data from protein sequencing, from DNA sequencing and from gene isolation indicate that the leukocyte interferons are made from a family of genes. Allen and Fantes (1980) found that purified lymphoblastoid interferon could be seperable into two fractions, which, on sequencing, revealed five different primary structures. One of these sequences alligns with the protein sequence predicted by the DNA sequence of a cloned cDNA made from leukocyte mRNA, determined by Mantei et al (1980). Streuli et al (in press, quoted in Allen and Fantes) have isolated another leukocyte interferon cDNA corresponding to another of the sequences of Allen and Fantes.

Therefore, leukocyte and lymphoblastoid interferons have the same composition, although the relative amounts of the different components may vary. The proteins form a family. At least some of the chromosomal genes which code for this family have been isolated by Nagata et al (1980b). They identify eight different restriction patterns among twelve leukocyte gene-containing recombinants. They determine the DNA sequence of the one corresponding to the cDNA sequenced by Mantei et al (1980). The information obtained by these analyses is discussed in section 5.4.

On the other hand, the amino acid sequences of all fibroblast interferon proteins and the sequences predicted from analysis of four different cDNAs show no heterogeneity (Tanaguchi et al, 1980; Derynck et al, 1980a; Goedell et al, 1980b; M.Houghton pers. comm.). There is one base difference between the DNA sequences reported by different groups. Derynck et al find codon 30 (tyr) to be TAC whereas Tanaguchi et al, Goedell et al and M.Houghton find TAT. This could be due to there being two fibroblast interferon genes, or to there being two alleles. It is probably not due to errors in sequencing, as the groups of Derynck *et al* and M. Houghton have now isolated another clone which has TAT at codon 30 (M.Houghton personal communication).

It appears that the fibroblast interferon gene is a single gene, or a member of a smaller family of expressed genes.

2. Materials and Methods.

2.1 Materials

a) Chemicals.

All chemicals were Analar grade and supplied by B.D.H., Poole, England, unless otherwise stated, and with the following exceptions. HEPES, Trizma base (Tris), sodium chloride and DEPC were obtained from Sigma Chemical Co., St. Louis, Missouri, U.S.A.

b) Isotopes.

Isotopes were obtained from the Radiochemical Centre, Amersham, England.

c) Chromatography.

Sephadex was supplied by Pharmacia, Upsalla, Sweden; Bio-Gel was from Bio-Rad Laboratories, Richmond, California, U.S.A. and RPC-5 was from Miles Laboratories Inc., Kankakee, Illinois, U.S.A. Agarose for electrophoresis was also supplied by Miles Laboratories Inc.

d) Southern Transfers.

Nitrocellulose filters were from Sartorius, Gottingen, West Germany, or from Schleicher and Schull, Dassel, West Germany. Ficoll was obtained from Pharmacia and polyvinyl pyrrolidone and bovine serum albumin from Sigma Chemical Co.

e) Photography and Autoradiography.

Gels were stained in ethidium bromide supplied by Sigma Chemical Co. FP4 film, Phenisol and PQ Universal developers and Hypam fixer were obtained from Ilford, Basildon, England. Rx Medical X-ray film and Mach 2 intensifying screens were from Fuji Photo Co. Ltd., Tokyo, Japan.

f)Nucleic Acids.

DNA from λ-phage was supplied by Boehringer Corporation (London) Ltd., Lewes, England. Herring sperm DNA and polyadenylic acid were from Sigma Chemical Co.

g)General Enzymes.

DNaseI was obtained through Worthington Biochemical Corp. Freehold, New Jersey, U.S.A. E.coli DNA polymerase I and proteinase K were from Boehringer; RNase and lysozyme were from Sigma Chemical Co. DNA ligase made from T4-infected E.coli was a gift from L. Woods, Imperial College, London.

h)Restriction Enzymes.

EcoRl was supplied by MRE, Porton Down, England. Some digests were performed using BamHI from New England Biolabs, Beverly, Massachusetts, U.S.A. SstI, KpnI, and TaqI were gifts from Dr.P. Rigby, Imperial College. HsuI, PstI, BglII, BamHI, HpaI, MspI, and XbaI were all purified in this laboratory by various workers under the supervision of Dr. J. Arrand.

i)Bacteriological Materials.

Yeast extract and tryptone were both from Oxoid Ltd., Basingstoke, England. BBL trypticase was from Becton, Dickinson and Co., Cockeysville, Maryland, U.S.A. Agar was supplied by Difco Laboratories, Detroit, Michigan, U.S.A. Kanamycin, ampicillin and chloramphenicol were obtained from Sigma Chemical Co.; tetracycline was from Lederle Laboratories, Gosport, England.

"Megaplates" are Nunc Bioassay dishes from Nunc, Roskilde, Denmark.

2.2 Bacterial Strains.

HB101 was used as host for both the previously constructed pHβG1 and the subcloning of Pad2 in pAT153. It was supplied by Dr. H. Boyer, and its genotype is:

strR, rB-, mB-, recA-, pro-, gal-. (Boyer and Roulland-Dussoix, 1969.)

LE392 was the host used for all λ-phage growth. It was supplied by Dr. P. Leder. It is the same as the strain ED8656, derived from ED803 (Wood, 1966), with the genotye:

rK-, mK+, recA-, supE, supF, gal-, met-.

Extracts used for in vitro packaging of phage chromosomes were prepared from strains BHB2671 and BHB2673, which are referred to in Collins and Hohn (1978) as:

N205(λimm434, cIts, b2, red3, Eam4, Sam7)λ and

N205(λimm434, cIts, b2, red3, Dam15, Sam7)λ.

and were a gift from B. Klein (University of Edinburgh).

The λ-phage vector NM788 was given by Dr. N. Murray (University of Edinburgh). It is a derivative of NM760, produced by introducing 3 amber mutations into the original strain which has deletions from wild-type of the att-red region, the cI gene and the nin5 region. The central, "silent", region has been replaced by a HindIII fragment of E.coli containing the trpE gene (Murray et al, 1977).

Subcloning was carried out using the plasmid pAT153, a gift from Prof. D. Sherratt (University of Glasgow), which is a deletion derivative of pBR322 (Twigg and Sherratt, 1980).

2.3 DNA Manipulation Methods.

a)Preparation of Human DNA.

DNA was prepared both from both human spleen samples and from blood using basically the same method. Spleens were first homogenised in 0.15M NaCl, 0.1M EDTA, pH 10.5 whereas blood samples were washed once by centrifuging and resuspending the pellet in isotonic saline and repelleting. The buffy coats containing the white cells were removed and added to a small volume of 0.15M NaCl, 0.1M EDTA, pH 10.5. The cells of either tissue preparation were then lysed by addition of an equal volume of "lysis mix". Lysis mix is 6% triisopropylnaphthalene (Eastman Kodak Ltd., Rochester, New York, U.S.A.), 8% butan-2-ol, and 3% SDS.

The cell extract was made to 0.5M sodium perchlorate before extracting with an equal volume of phenol equilibrated with 0.15M NaCl, 0.1M EDTA, pH 10.5. The aqueous phase was extracted with equal volumes of chloroform:octan-2-ol (24:1) repeatedly until the interphase was free of protein. The DNA was precipitated from the aqueous phase by addition of 2 volumes of ethanol and was spooled out. Spooling the DNA selects for high-molecular weight material and leaves most of the RNA behind.

The DNA was dissolved in 10mM Tris, 1mM EDTA, pH7.5. The solution was made to 2x SSC, and 50μg/ml RNase A was added. After incubation at 37° for 1h., the solution was made 0.1% SDS and digested with 10ug/ml proteinase K for 3-5h. at 37°. After further phenol and chloroform:octan-2-ol extractions the DNA was precipitated with ethanol, spooled and dissolved in 10mM Tris, 1mM EDTA, pH 7.5. DNA solutions were stored at -20°.

b)Restriction Enzyme Digestion of DNA.

Where possible, restriction enzyme digestion of human, plasmid or phage DNA was carried out at a DNA concentration of 100 to 200µg/ml. The buffer used varied according to the enzyme being used. In general the enzymes prepared in this laboratory or at Imperial College were used in 10mM Tris, pH 7.5, 6 mM MgCl2 and 6 mM β-mercaptoethanol. The EcoR1 from MRE was used in the same buffer supplemented with 0.1M NaCl and the BamHI suplied by New England Biolabs was used in this buffer with 0.15M NaCl. All digestions were performed in the presence of 50µg/ml nuclease-free bovine serum albumin (Bethesda Research Laboratories Inc., Rockville, Maryland, U.S.A.), and at 37°; with the exception of TaqI. TaqI digests were carried out at 65°.

c)Gel Electrophoresis.

In general both agarose and polyacrylamide gel electrophoreses, for the analysis of DNA, were carried out in E-buffer (40mM Tris, 20mM sodium acetate, 2mM EDTA, pH 7.5). DNA samples which were to be transferred to nitrocellulose filters were electrophoresed in T-buffer (40mM Tris, 40mM sodium acetate, 2mM EDTA, 0.5µg/ml ethidium bromide, pH 7.5). Gene-machine electrophoresis was carried out in 10mM Tris, 40mM sodium acetate, 1mM EDTA, pH 7.9.

Agarose gels were made by disolving the appropriate percentage of agarose powder by boiling in E-buffer, and poured into vertical or horizontal moulds. Acrylamide and bis-acrylamide in a ratio of 145:2 were cross-linked in E-buffer by the addition of 1 mg/ml ammonium persulphate and

0.5µl/ml of N,N,N',N',tetramethylethylenediamine.

Samples were loaded in 2% (w/v) ficoll plus a small amount of Orange G (Sigma) as a front marker. Electrophoresis was carried out in an electric field of 1 to 10 V/cm for from 1 to 15h.

After electrophoresis the gel was stained in E-buffer containing 0.5ug/ml ethidium bromide. The DNA was visualised with short-wavelength UV light and photographed on Ilford FP4 film using a Polaroid MP-4 camera with a red filter.

d)Southern Transfer and Hybridisation.

DNA was transferred from gels to nitrocellulose filters using the method of Southern (1975). In order for DNA to bind to nitrocellulose it must be single-stranded, and therefore must be denatured prior to transfer. The DNA was denatured by soaking the gel for 2-5h. in 0.5M NaOH, 1.5M NaCl, and the gel neutralised in 0.5M Tris, 1.5M NaCl, pH 7.5. The DNA was transferred by blotting through with 20x SSC onto a sheet of nitrocellulose for 15 to 60h. The nitrocellulose was washed for 10 minutes in 2x SSC and the DNA bound to the filter by baking at 80° for 2h.

Before hybridisation the filter was pretreated to eliminate its DNA-binding ability. All these treatments were performed at 65° and are essentially as described by Jeffreys and Flavell (1977). The filter was soaked for 30 minutes in 3x SSC followed by immersion for 3h. in 0.2% (w/v) ficoll, 0.2% (w/v) polyvinyl pyrrolidone, 0.2% (w/v) bovine serum albumin (this is 10x Denhardt's solution; Denhardt, (1966)) in 3x SSC. This was followed by 1h. in the same solution supplemented with 0.1% (w/v) SDS and 50µg/ml sheared and denatured herring sperm DNA. If the probe used contained poly d(A:T) tracts, this solution was further supplemented with 10μg/ml polyadenylic acid.

Hybridisation was carried out in this final solution, with the addition of the probe, labelled by nick-translation, at a concentration of 1 to 3 ng/ml of hybridisable sequence. The probe was denatured before hybridisation by heating to >95° for 5 minutes.

After hybridisation for 15 to 60h. the filters were washed several times at 65° in 3x SSC supplemented with 0.1% SDS and 50μg/ml sheared and denatured herring sperm DNA. The filters were finally washed with a solution of low salt concentration in order to reduce non-specific hybridisation. This was especially important if hybridisation to other, related, genes was to be minimised, or to reduce background hybridisation when a large excess of vector DNA was present in the hybridisation. The solution used was usually 0.1x SSC, 0.1% SDS, although higher salt concentrations were sometimes used, as noted in the text.

The filters finally were rinsed in 3x SSC, dried and autoradiographed at -70° using film presensitised by flashing and Mach 2 fast intensifying screens (Laskey and Mills, 1977).

e) Labelling of DNA by Nick-Translation.

To label DNA to high specific activity, nucleotides labelled at the α-position were incorporated into DNA by nick-translation (Rigby et al, 1977).

DNA at a concentration of 15ug/ml was incubated at 15° in a small volume containing 120 units/ml E.coli DNA polymerase I, 170pg/ml DNaseI, 4µM dATP, 4µM dGTP, 2.6 µM $\alpha^{32}P$ dCTP, 2.6 µM $\alpha^{32}P$ TTP (both of specific activity 400 Ci/mmol), 10 mM 2-mercaptoethanol, 5 mM MgC12, and 50 mM Tris, pH 7.5. After 90 to 120 minutes the reaction was stopped by diluting the mixture with 3x SSC, and addition of an equal volume of phenol. After phenol extraction the unincorporated nucleotides were removed by passing over Sephadex G-50 and the excluded peak pooled. The specific activity of the DNA was calculated by measuring Cherenkov radiation.

f) Ligation of DNA

Ligations were carried out at a DNA concentration of >100µg/ml to promote the formation of concatenates. The enzyme was not assayed precisely, but the quantity used (c.0.5 Weiss units) was that found empirically to be adequate. The reaction was performed in 20 mM Tris, pH 7.5, 10 mM $MgCl_2$, 20 mM 2-mercaptoethanol, 0.5 mM ATP, with 50µg/ml nuclease-free bovine serum albumin at room temperature for 6h.

g) Isolation of DNA from Agarose Gels.

Electroelution of DNA from agarose gel was performed by enclosing the fragments of agarose in a bag of dialysis tubing, which was completely filled with E-buffer.This was submerged under E-buffer and subjected to an electric field of 2 V/cm for 15 to 20 hours. The field was reversed for 15 minutes to remove DNA from the dialysis membrane and the E-buffer within the bag removed. The DNA was then phenol extracted, ether extracted and ethanol precipitated.

Sodium Perchlorate Elution (Chen and Thomas, 1980) was carried out by dissolving the gel fragment in 5 volumes of 6M sodium perchlorate in 30 mM Tris phosphate, pH 8, at 45° for 15 minutes. The DNA was removed from the solution by binding it to GF/C filters (Whatman, Maidstone, England) and washed with 6M sodium perchlorate, followed by propan-2-ol and ethanol. After drying, the DNA was eluted with 10 mM Tris, 1mM EDTA, pH 7.5 by centrifuging the solution through the filter. The DNA was, in general, concentrated by ethanol precipitation.

h) RPC-5 Chromatography.

DNA was fractionated on RPC-5 collumns as described by Tilghman *et al* (1977). The RPC-5 medium was packed into the column under a pressure of 400 psi in 1.25M sodium acetate, 50mM Tris, 1mM EDTA, pH 7.5. DNA was loaded onto the column in this same buffer using a Milton Roy Minipump, and the column washed with the buffer. DNA was eluted with a gradient of 1.6 to 1.8M sodium acetate in 50mM Tris, 1mM EDTA, pH 7.5. Fractions were monitored by optical density at 260 nm.

2.4 Bacteriological Methods.

Plasmids.

a) Plasmid Preparation.

Bacterial strains were grown in liquid culture in L broth. L broth is 5g/l yeast extract, 10g/l tryptone, 5g/l NaCl 1g/l glucose, pH 7.2. Strains containing different plasmids were grown in broth supplemented with antibiotics to select for the desired strain, as follows;

HB101 alone, 200ug/ml streptomycin

HB101+pHβG1, 25ug/ml kanamycin

HB101+R1-C,+Hβ1-S,+pAT153,+pAT153 subclones,
50μg/ml ampicillin.

For plasmid preparation, 50 ml cultures were grown overnight at 37° in L broth, and a fraction diluted 1 in 100 into 500ml of L broth. This culture was grown at 37° to log-phase; that is, an optical density of 0.8 at 650 nm. Chloramphenicol was added to 200ug/ml and the plasmids allowed to amplify overnight.

The bacteria were killed by the addition of 10ml/l chloroform and shaking for 5 minutes, and harvested by centrifugation in a Sorvall GSA rotor for 10 minutes at 8000x g at 4° . They were then washed in 25% (w/v) sucrose, 50mM Tris, pH 8, repelleted and resuspended in a small volume of the sucrose/tris solution. The cells were lysed by the addition firstly of 450μg/ml freshly dissolved lysozyme for 5 minutes on ice. The suspension was made 100mM Tris pH 8, 70mM EDTA and left for 5 minutes on ice. Finally the cells were lysed by the addition of Triton X-100 to make the suspension 0.2%. After 20 minutes on ice lysis was usually complete, as indicated by the suspension becoming very viscous. If lysis was not apparent, 5 minutes at 37° was usually sufficient complete it.

The cell debris and the chromosomal DNA were pelleted in a Beckman SW40 rotor at 150,000x g at 4°. The supernatant contained the plasmid DNA. This was phenol extracted several times and chloroform extracted before digesting with 50µg/ml RNase at room temperature for 1h. The plasmid was again phenol extracted and chloroform extracted.

The supernatant was made 8mM with K_2HPO4, pH 7.5, 1mg/ml ethidium bromide and 91% (w/v) with solid CsCl. The solution was cleared by centrifugation at 30,000x g for 30 minutes in the Beckman SW40, and centrifuged to equilibrium at 140,000x g in the SW40 for 40h. or the Sorvall TV-865B for 15h. The plasmid band was visualised under UV illumination and collected by piercing the tube with a syringe needle.

The ethidium bromide was removed by extraction with propan-2-ol and the DNA ethanol precipitated. The plasmid was then analysed by gel electrophoresis. If contamination with low molecular weight material was extensive, this was removed by sucrose gradient fractionation. Sucrose gradients were made from 10 to 40% in 10mM Tris, 1mM EDTA, pH 7.5, and centrifuged at 140,000x g for 100 minutes in the Sorvall TV-865B. Fractions were assayed for plasmid content by electrophoresis on a 1% agarose gel. The fractions free of contamination were pooled and precipitated with ethanol.

b) Transformation of HB101 with pAT153 Plasmids.

An overnight culture of HB101 was diluted 1 in 100 into L broth plus streptomycin and grown, with shaking, at 37° to an OD_{650} of 0.5. The cells were harvested by centrifugation at 6000x g in a Sorvall HB-4 rotor and resuspended in a volume of freshly made 80mM $CaCl_2$ equal to the original culture volume. After 15 minutes on ice the cells were concentrated 10 fold by pelleting again and respending in one tenth the original volume of 80mM $CaCl_2$, 15% glycerol. The cells were dispensed in 200µl aliquots.

The DNA to be introduced into the cells was added to this aliquot at a concentration of 0.5 to 2.5µg/ml in 100µl 10mM Tris, 1mM EDTA, pH 7.5. The cells were left in the DNA for 10 to 30 minutes on ice before freeze-shocking in a dry-ice/propan-2-ol (-55°) bath for 100 seconds. After thawing and leaving on ice for 5 minutes, the cells were heat-shocked at 42° for 60 seconds.

The cells were now transformed. The volume was made up to 1 ml by the addition of 700µl L broth. The efficiency of the transformation was assayed by plating different volumes of the suspension on L agar plates (L broth plus 1.5% agar) containing 50µg/ml ampicillin.

c)Colony Electrophoresis.

Analysis of the plasmid content of individual colonies was performed by lysis of the cells and agarose gel electrophoresis of their DNA.

A single colony was picked into 500µl L broth plus ampicillin, and grown for 2h. at 37 . Chloramphenicol was added to 170µg/ml and the cells left overnight to amplify the plasmid. The cells were harvested by centrifugation at about 15,000x g in an Eppendorf centrifuge, for 3 minutes, and resuspended in 20ul 50mM Tris, pH 8, 5mM EDTA. SDS was added to 0.75% and the cells digested with 700μg/ml proteinase K at 37° for 2h.

The lysate was loaded into the dry slot of an agarose gel and covered with molten agarose before electrophoresis.

Phage.

d) Preparation of Plating Cells.

The cells used to form the lawn on which the λ-phage are grown are prepared as follows. An overnight culture of the host bacterium (LE392 was used throughout this study) was diluted 1 in 100 into fresh L broth. The cells were grown for 2 hours before harvesting in a Sorvall HB4 rotor at 6000x g for 10 minutes. The cells were resuspended in 10mM $MgCl_2$, in the same volume as the original culture, and stored at 4° for several weeks.

e) Growth of Phage on Plates.

Phage were always stored and diluted in phage buffer. Phage buffer is 21.5mM KH_2PO_4, 49mM Na_2HPO_4, pH 7.15, 85mM NaCl, 1mM $MgSO_4$, 0.1mM $CaCl_2$, 0.001% gelatin. 100μl of the appropriate dilution of phage were added to 200μl of plating cells, and incubated at 37° for 15 minutes. 2.5ml of 0.7% agar in BBL broth (10g/l BBL trypticase peptone, 5g/l NaCl, 10mM $MgSO_4$), cooled to 45°, were added to the cells and the mixture quickly poured onto plates of 1% agar in BBL broth. The plates were incubated overnight at 37° to allow the lawn to grow.

f) Large Scale Growth of Phage.

Two methods were used for the large scale preparation of phage.

i)The m.o.i.=1 Method.

A single plaque from a plate was picked into 100ul phage buffer, and this was added to 200μl plating cells and poured in 0.7% agar onto 1.5% agar in L broth. After growth overnight the plate was confluent. The plate was overlaid with 5ml phage buffer and left at 4° overnight. The phage buffer was removed and titred for content of phage.

Host cells were grown in L broth, supplemented with 10mM $MgSO_4$ to an OD650 of 0.45 to 0.6 (equivalent to 2 to 3 x 10^8 cells/ml). Sufficient phage were added to the cells to give a multiplicity of infection (m.o.i.) of approximately 1. The culture was incubated statically for 15 minutes before shaking vigorously at 37° for between 6 and 20 hours, until the appearence of debris indicated that lysis had occurred.

The bacterial debris was removed by centrifuging at 6000x g for 10 minutes and the phage suspension was titred. The process was repeated until a sufficiently large preparation was made.

ii)The P.D.S. Method.

This method gave a lower yield. It was used, however, for the amplification of libraries where picking plaques is not a suitable method.

Host bacterium were grown to an OD_{650} of 0.45, from a dilution of 1 in 100 of an overnight culture into 100ml. The cells were then harvested as described above and resuspended in 5ml of phage buffer. This contained 2 x 10^{10} cells, to which were added 2 x 10^7 phage. After incubation at 37° for 15 minutes to preadsorb, the cells were diluted into 500ml of L broth plus 10mM $MgSO_4$ and incubated at 37° , with vigorous shaking, for 20h.

g) Preparation of Phage DNA.

DNA was prepared from phage in 500ml to 1 litre cultures. After cell lysis the remaining cells were killed by adding 5ml/l chloroform. 40g per litre NaCl were added, plus DNase and RNase to 1μg/ml. The culture was digested at room temperature for 1h. The cell debris was removed by centrifugation in a Sorvall GSA rotor at 16,000x g for 10 minutes. PEG 6000 was added to 10% and the solution kept at 4° for 5 to 15h. to precipitate the phage.

The phage precipitate was collected by centrifugation as above and redissolved in a small volume of phage buffer. Any cell debris still present was removed by centrifugation as above. Any bacterial DNA present was removed by digestion with 10μg/ml DNase. The phage were centrifuged to equilibrium on a 41.5% CsCl gradient, at 36,000 rpm in a Beckman SW40 for 40h or a Sorvall TV-865B for 15h. The phage banded in a tight, white band which was harvested with a Pasteur pipette and rebanded on a second CsCl gradient.

Following the centrifugation, the CsCl was removed by dialysis against 25mM NaCl, 10mM Tris, pH 8, 1mM $MgSO_4$. The solution was made to 2% SDS, 10mM EDTA and heated to 65° for 15 minutes. Proteinase K was added to 50μg/ml and the solution digested for 1h at 37° . The DNA was extracted several times with phenol equilibrated with 0.5M Tris, pH 8, and finally extracted with chloroform/octan-2-ol (24:1) before ethanol precipitation.

The DNA pellet was dissolved in 10mM Tris, pH 7.5, 1mM EDTA and stored at 4°.

h) Small-Scale DNA Preparation from Phage.

Analysis of the DNA content of single plaques was carried out by a method essentially of Cameron *et al* (1977).

A single plaque was picked into phage buffer and preadsorbed to 200µl plating cells before mixing with 2.5ml 0.7% agarose in BBL broth plus 10mM $MgSO_4$ and pouring onto 90mm plates of 1.5% agarose in L broth. The plates were overlaid with 5ml 10mM Tris, pH 8, 10mM EDTA and left for 20h at 4° or 4h at 37°. The supernatant was removed and made 50mM EDTA, pH 8.5, 20mM Tris, 0.4% SDS, and diethylpyrocarbonate added to 0.002%. Samples were heated for 30 minutes at 65°, cooled on ice and made 1M with potassium acetate. The solution was kept on ice for 1h and then centrifuged for 10 minutes at 15,000x g.

The supernatant was phenol extracted and ethanol precipitated before being dissolved in 1/10 the original volume of 10mM Tris, pH 7.5, 1mM EDTA.

i) Preparation of Extracts for In Vitro Packaging of Phage Chromosomes.

The method is as descibed in Collins and Hohn (1978). The two strains, BHB2671 and BHB2673 were grown separately on 1.5% agar in L broth, to give a dense growth. The cells were scraped off and suspended in L broth. These cells were diluted into 250ml each of L broth to an OD_{650} of 0.13 to 0.15. The cells were grown at 32° through one generation, to an OD_{650} of 0.3. The lysogens in each strain was induced by incubating for 15 minutes at 45 , and the cells subsequently grown at 37° for 3h.

The two strains were mixed, harvested by centrifugation in a Sorvall GSA rotor at 8000x g for 10 minutes, and resuspended in an equal volume of M9. M9 is 49mM Na HPO , 21.5mM KH_4PO , pH 7.15, 8.5mM NaCl, 2mM NH_4Cl, 1mM $MgSO_4$, 0.1mM $CaCl_2$, 0.4% glucose. The cells were then irradiated in 200ml batches in a tray at a distance of 40cm from a "Mineralight S-68" UV source, for 3.5 minutes. This was to inactivate the genomes of the lysogenic phage in the extract, to prevent their recombination with the exogenously added phage DNA. The irradiation had been calibrated by D. Westaway to give a reduction in titre of endogenous phage of 10^{-4}

Following irradiation the cells were concentrated by successive centrifugation. The cells in M9 were centrifuged at 8000x g in the Sorvall GSA rotor for 10 minutes, and resuspended in 1/50 the original volume in concentration buffer. Concentration buffer is 40mM Tris, pH 8, 10mM NaN_2, 10mM $MgCl_2$, 10mM putrescine dihydrochloride, 10mM spermidine hydrochloride, 0.1% 2-mercaptoethanol, 7% dimethylsulphoxide. The cells were centrifuged again, in the Sorvall HB-4 rotor for 10 minutes at 16,000x g, and resuspended in 1/400 the original volume. Finally the cells were centrifuged at about 15,000x g in an Eppendorf centrifuge for 1 minute. 1/5000 volume of concentration buffer was added and the cells dispensed in 40µl aliquots into Eppendorf tubes and snap-frozen in liquid nitrogen. The extracts were stored at -70°.

j)In vitro Packaging of Phage Chromosomes.

Before use the extracts were transferred to liquid nitrogen. They were then thawed on ice. ATP was added to 8mM and the DNA to be packaged was added up to a maximum of 1µg per 40µl aliquot. The extract was incubated at 37° for 45 minutes. The reaction was stopped by adding a second aliquot of extract, which had been incubated with 20ng DNaseI for 10 minutes. The extract was then incubated for a further 45 minutes, and finally mixed with 500µl of phage buffer and a drop of chloroform. The phage were titred and stored at 4°.

k) Plaque Hybridisation.

The method is essentially that of Benton and Davis (1977). Plates which were to be used for plaque hybridisation were made with 1% agarose, with a phage-containing top layer of 0.7% agarose. After growth the plates were chilled at 4°. The nitrocellulose filters, to which the DNA was to be transferred, were soaked in 3x SSC and blotted dry.

DNA was transferred from each plate in duplicate by placing the two filters, in succession, in contact with the plates for 2 minutes for the first transfer, or 3 minutes for the second. The filter was carefully removed, denatured on a pad of 0.5M NaOH, 1.5M NaCl for 2 minutes, and neutalised in 0.5M Tris, pH 7.5, 1.5M NaCl. After blotting dry the filters were baked for 2h at 80° and subsequently hybridised as described in 2.3d.

3. Strategies and Tactics For The Isolation of Human Genes.

3.1 Introduction.

a) Host/Vector Systems.

The extrachromosomal elements which can be manipulated to act as vectors for exogenous DNA can be basically divided into two groups; plasmids or phage. As the use of "disabled" host/vector systems permits a lowering of the physical containment category for the experiment, several disabled systems have been developed.

The plasmid vector which rapidly became the most-used was a ColE1 derivative, pBR322, and its relatives, as they contain two drug-resistance markers, for tetracycline and ampicillin resistance. Either of these can be used to select for transformants whilst the other may be used to detect the presence of an inserted DNA segment by the loss of the resistance. The *E. coli* strain χ1776, developed by Curtiss and coworkers (1977), can be used as a host for pBR322, and is a system recognised as disabled by the U.K. Genetic Manipulation Advisory Group (GMAG). This strain requires thymine, or thymidine, plus diaminopimelic acid, is sensitive to bile salts and is resistant to a number of bacteriophage with which a recombinant plasmid could recombine.

χ1776, however, is notoriously difficult to grow and its use has recently been superceded by the use of the disabled plasmid pAT153 (Twigg and Sherrat 1980). This is a derivative of pBR322 which has had its origin of conjugal transfer (*nif*) deleted and therefore cannot be mobilised. It may therefore be used with any rec A- strain of *E.coli* K12 and these systems are then regarded as disabled. A second advantage of this vector is that the deletion in the *nif* region allows the plasmid to grow to a higher copy number and so give a higher yield.

However, whilst they have often been used for the isolation of double-stranded cDNA molecules, plasmids have rarely been used to isolate chromosomally derived genes from higher eukaryotes. The main reasons for this are that the transformation efficiency of plasmids, espacially those bearing large inserts, is low compared to bacteriophage packaged *in vitro*, and the screening of large numbers of plasmid-containing colonies has not proved as simple as screening of λ-phage.

The rapid screening method of Benton and Davis (1977) and the *in vitro* encapsidation methods of Collins and Hohn (1978) and Sternberg *et al* (1977) have made lambda phage vectors the method of choice. Of these the Charon series (Blattner *et al*, 1977), λgtWES (see Tilghman *et al*, 1977) and NM788 (Murray *et al*, 1977) are the most useful as they are replacement vectors, that is the non-essential central region can be removed from the two "arms". Together the arms are too small to produce viable phage, only those arms enclosing an inserted DNA fragment will be packaged into phage particles and thus the background of non-recombinant phage is minimised. Current GMAG guidelines permit the use of these vectors with any rec A- strain of *E.coli* as a disabled system.

There are only two enzymes which can be used to produce fragments with cohesive termini complementary to the ends of the arms of the commonly used vectors; EcoR1 for λgtWES and Charon 4a, and Hind III (or Hsu I) for NM788. The size requirement imposed by packaging into the phage heads means that there is a strict size range of DNA which can be inserted into each vector. For NM788 and λgtWES this is from less than 1kb to about 14kb, whilst for Charon 4a it is from 8 to 21kb.

b) Gene Enrichment Methods.

One problem inherent in identifying a desired recombinant clone is that of dealing with large numbers of recombinants, needed to ensure that the fragment is present in the collection. It is useful to reduce the number involved by enriching the DNA used for the fragment of interest. This approach is only of use if the restriction enzyme "map" around the gene is known, and fragments suitable for cloning, which contain the regions of interest, are available.

Although a growing number of gene-enrichment methods are being developed, there are only two which have wide usage. Fractionation of the DNA fragments according to size by gel electrophoresis, whether on a large scale (on a "gene-machine") or a small scale, has been used by several workers to enrich DNA for a particular fragment prior to cloning. (For example; Cordell et al, 1979, rat insulin; Van den Berg et al, 1978, rabbit β-globin; Proudfoot and Baralle, 1979 human ε-globin.)

RPC-5 chromatography, first described as a method for fractionating tRNA molecules, has also been used to fractionate DNA. The principle by which this fractionation is acheived is not known, but it is not purely on the basis of size. This method of fractionation produces an enriched fraction which contains a wide range of fragment sizes, and as such may not be useful for clonong immediately. However, Gorin and Tilghman (1980) used this method to enrich for the mouse α-fetoprotein gene prior to cloning.

If these two methods are used in tandem the gene enrichment produced is obviously multiplied. Using a tandem fractionation, Tilghman et al (1977) enriched for the mouse β-globin gene 500-fold, and so had only to screen 2,000 phage to identify a "positive" clone., as compared to around 1,000,000 which would contain one copy of the gene if total DNA were used.

Of course, if the enzyme used to cut the DNA cuts within the gene or region of the genome of interest, then this region must be enriched and cloned in fragments, or a "library" approach used as described below.

c) Synthesis of Gene Libraries.

The screening techniques which are now in use have enabled large numbers of recombinant phage to be screened with relative ease, and have made the library approach to gene cloning more feasible. In general terms this means the synthesis of a large number of recombinants so that there is a high probability that the whole of the genome is included in the collection.

A simple approach, whereby a total restriction digest of the genome is "shotgunned" into the bacteriophage will not meet this requirement, as there will be fragments too large and too small to be packaged into phage particles. A partial digestion with the enzyme producing the correct cohesive ends will result in the small fragments becoming "clonable" as a result of their being still joined to adjacent fragments. A mixture of DNA fragments partially and fully digested with EcoR1 or HindIII will provide an almost complete library; the only fragments not represented will be those which are too large to be inserted into phage after even a total digestion.

Although this method is adequate for most genes, in order to produce the most complete library a wide range of under-digestions need to be performed on any one sample. A more satisfactory method is that developed by Maniatis et al (1978) which uses two restriction enzymes which cut DNA frequently (Hae III and Alu I) to produce, after partial digestion, an overlapping collection of fragments which should span almost all of the genome. These fragments are inserted into the vector chromosome by the addition of synthetic EcoR1 cohesive ends, after the EcoR1 sites in the fragments themselves have been inactivated by "EcoR1 methylase".

The partial digestion approach to the synthesis of libraries has two other advantages. By inserting larger fragments into the vector the whole genome is contained in fewer recombinants, and so fewer need be screened in order to isolate a particular gene. Furthermore, because overlapping fragments are used, a set of recombinant bacteriphage can be isolated which will cover a large region of the genome surrounding the gene. The overlapping fragments can be used to "walk along the chromosome" and adjacent or nearby DNA sequences isolated.

The primary aim of this thesis was to isolate the β-globin genes from a patient with β^0-thalassaemia. Using the β-globin gene as a model, methods for gene enrichment and for the production of recombinant phage were developed. These methods were applied to the isolation of the β-thalassaemic gene, and also to the isolation of the gene for human fibroblast interferon.

3.2 Developing the Technology.

a) The Gene Machine.

"Gene machine" is a term applied to preparative agarose gel electrophoresis systems. Two such apparatus have been described in the literature. The one used by Tilghman *et al* (1977) has a slab of agarose and the sample is run horizontally. The other was designed by Southern (1979) and was the one used in this study.

The gel is cast in a cylindrical mould, with the loading slot a circular groove, concentric to the gel's perimeter. Electrophoresis takes place towards the centre, where the anode, enclosed in a tube of dialysis membrane, is located. The basic design of the apparatus is illusrated in Figure 3.1. An electromechanical timer, at preset intervals, reverses the current for 30 seconds to free the DNA which has bound to the dialysis tubing. A pump then operates to collect the DNA now in the central buffer compartment, and the cycle begins again.

Placental DNA prepared from a normal individual was used to test the fractionation. 6 mg of Eco R1 digested DNA was electrophoresed on the gene machine using Orange G as the tracking dye. The gel was run at 20v and fractions collected every 30 minutes, starting after the Orange G was observed to have eluted. Alternate fractions were run on analytical agarose gels in order to estimate their molecular weight. 154 fractions were collected, which included fragments from 0.46 to 7.2 kilobases in length. Figure 3.2 illustrates the fractionation obtained. The fractions were pooled in pairs

GENE MACHINE

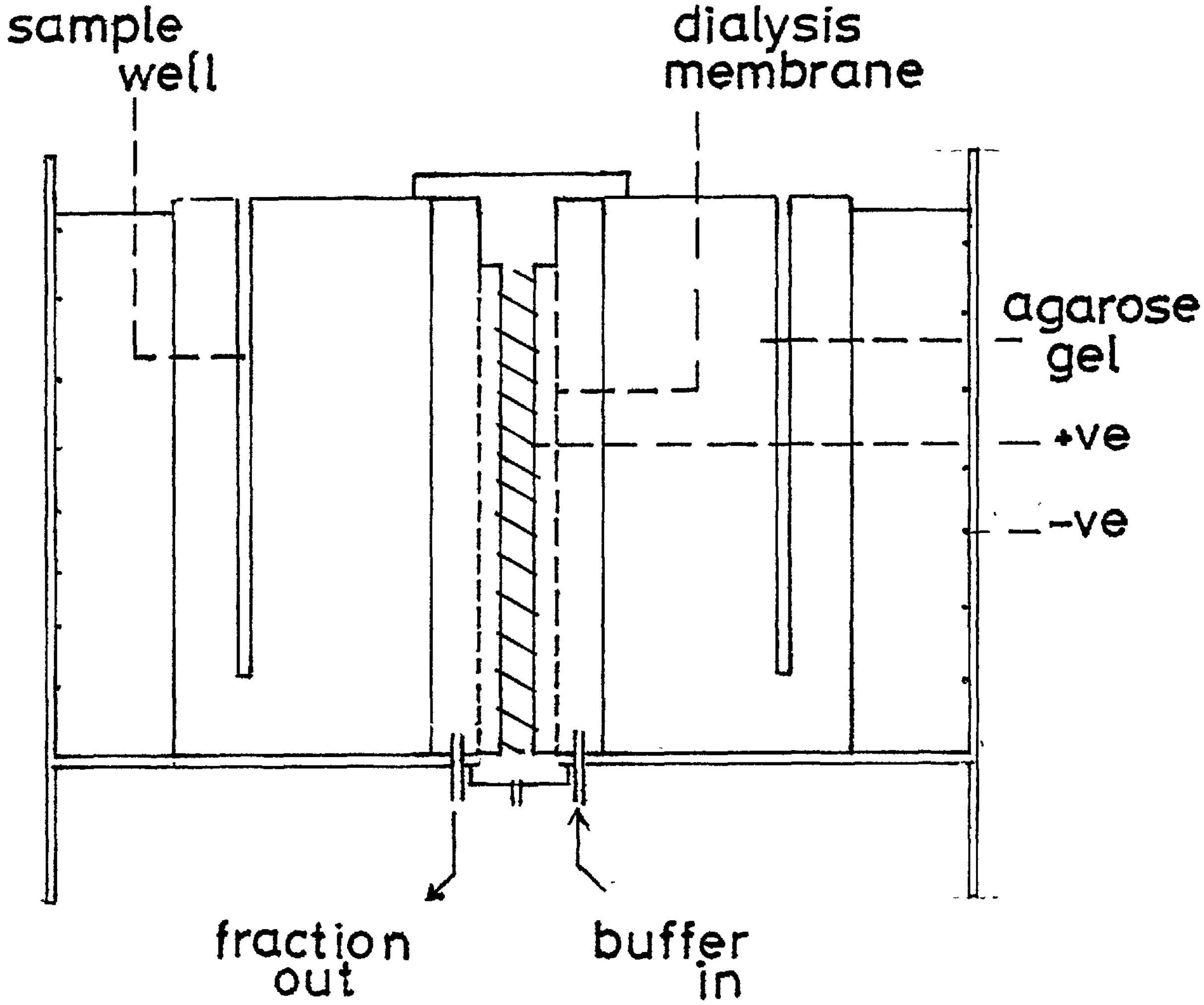

sample well
dialysis membrane
agarose gel
+ve
-ve
fraction out
buffer in

Figure 3.1.

Schematic represenation of the Southern Gene Machine. The view is of a cross-section through the cylindrical gel. The sample well is a circular slot, concentric with the circumference of the gel.

Details of the operation of the gene machine are in the text.

and approximately 2µg of each pool from 2.5 to 7.2 kb were electrophoresed and a blot of the gel was hybridised to radioactive pHβG1. An autoradiograph of the transfer is shown in Figure 3.3 The modal sizes of the hybridising fractions were estimated to be 5.7 - 6.5kb (pool 1) and 3.4 - 3.6 kb (pool 2), which correspond to the size estimates made of the 3' and 5' ends respectively of the β-globin gene (Flavell et al, 1978; Lawn et al, 1978).

The larger fragment, as can be seen, is spread over more fractions than the smaller. This is to be expected, as a larger fragment migrates more slowly on electrophoresis, and so will be collected over a longer period of time, and thus over more fractions.

Although the fractionation produced by the gene machine is good, the apparatus has a number of disadvantages. There are several pumps and other mechanical parts which may be unreliable over the long period (several days) of electrophoresis. A large amount of starting material is required; of the order of milligrams. The large number of fractions produced is cumbersome and their analysis difficult as the concentration of DNA may fall to less than 1µg/ml.

b) Small Scale Preparative Gel Electrophoresis.

The high effeciency with which recombined phage chromosomes can be packaged into viable virus particles *in vitro* means that only a few micrograms of DNA enriched for the fragment to be cloned is sufficient to produce enough recombinants to include all the fragments. This quantity can be obtained by eluting the DNA from the region of an agarose

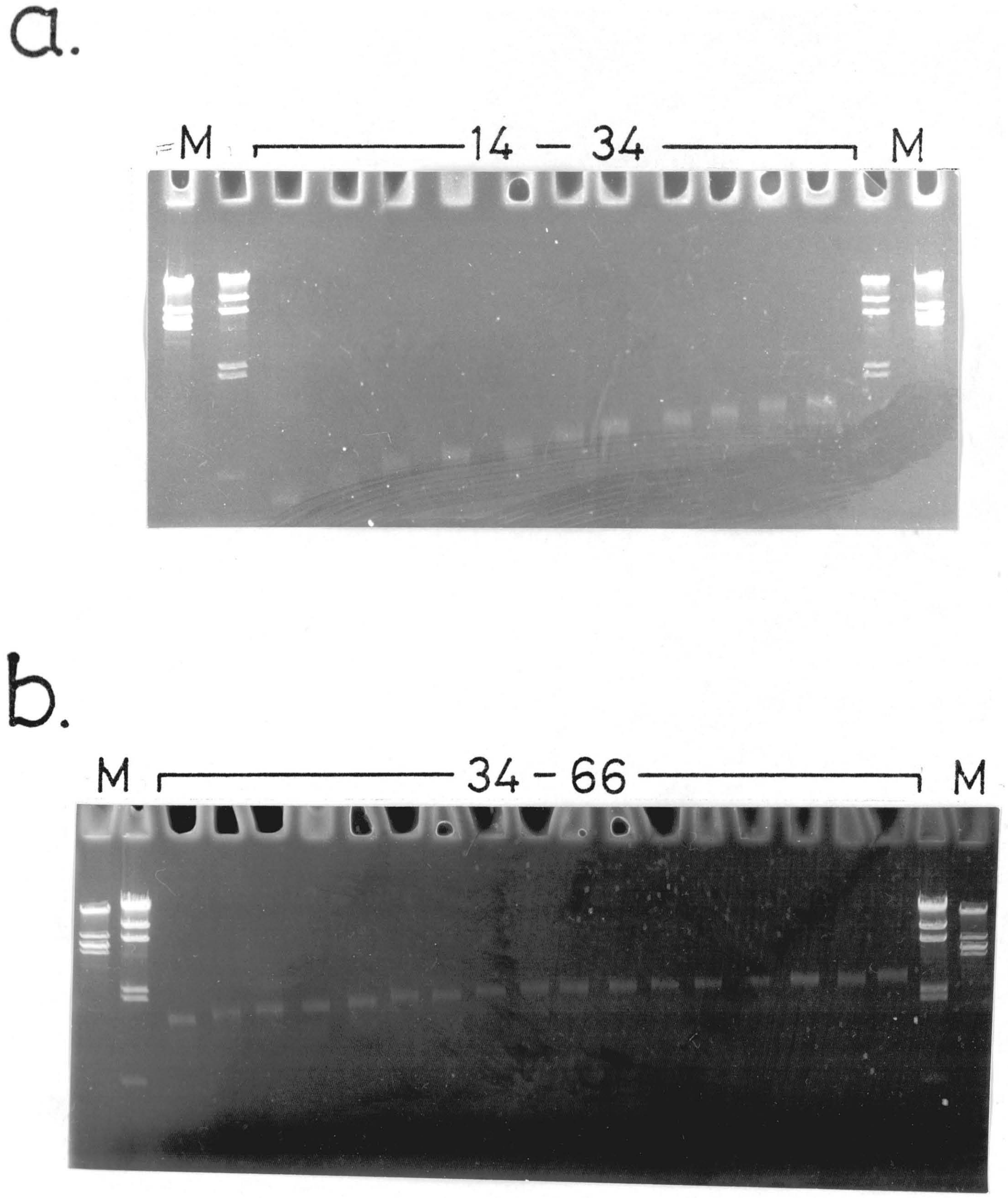


Figure 3.2.

Agarose gel electrophoresis of fractions collected during the operation of the gene machine. The DNA undergoing fractionation was normal DNA digested with EcoR1. 20ul aliquots of alternate fractions were loaded directly onto the gel.

M: markers, λ-DNA digested with HsuI (inner tracks) and with EcoE1 (outer tracks).

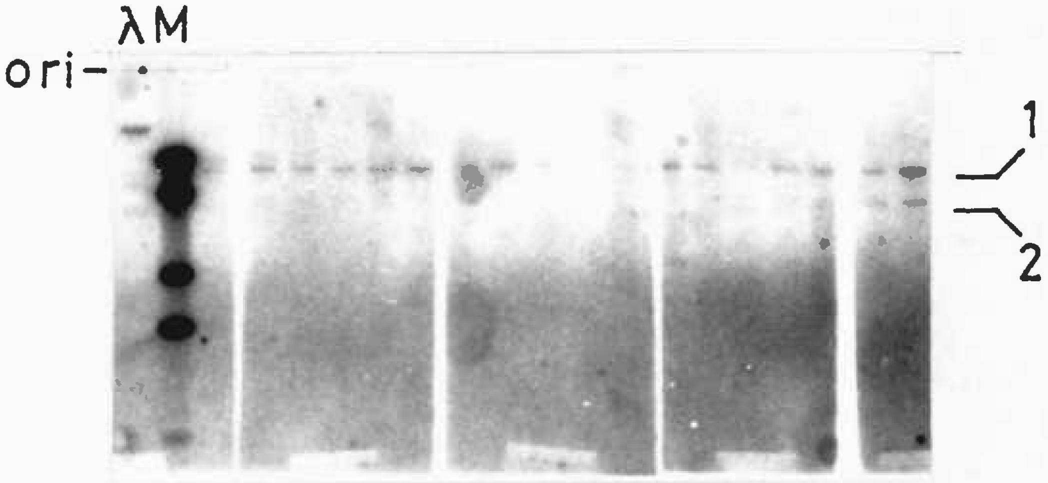


Figure 3.3.

Southern blot hybridisation to gene machine fractions. Each fraction was pooled pairwise, and three of these pools loaded onto each slot of the gel before electrophoresis. Adjacent pools were loaded into adjacent slots. Therefore, slot 1 contained pools 1, 19 and 37 and so on. The filters were hybridised with nick-translated pHβG1, a β-globin cDNA-containing plasmid.

1 and 2 indicate those fractions hybridising to the probe, and are derived from the 5' and 3' ends of the β-globin gene respectively.

M: pHβG1 digested with BglI (sizes are, 6.6, 4.4, 1.82 and 1.0kb)

λ: λ-DNA digested with EcoR1.

gel which has been ascertained to contain the size fraction of interest.

Any damage done to a DNA fragment during its preparation is a potential source of mutation when it is repaired by the bacterial cell after cloning. Thus a very gentle method of isolation from agarose gels was developed.

10µg of HsuI digested DNA from a patient with homozygous β°-thalassaemia was loaded in each of five slots on a 1% horizontal agarose gel, and electrophoresed at 2V/cm for 15 hours. λ-DNA cut with HsuI was coelectrophoresed in an adjacent track as a marker. This track was removed, stained with ethidium bromide and the bands visualised with UV light. Using these size markers as a guide, the region of the human DNA tracks spanning fragments between 6.75 and 9.5 kb were cut out. The preparative gel was subsequently stained in ethidium bromide and photographed (Figure 3.4).

The piece of agarose was sealed into a dialysis bag, with 200µl of E-buffer, submerged under E-buffer, and subjected to a current of 100mA. After 24h. the current was reversed for 15 min. to remove the DNA from the dialysis membrane and the E-buffer in the dialysis bag, containing the DNA, was collected. After phenol extraction and ethanol precipitation an aliquot of the fraction was run on a 1% analytical agarose gel (Figure 3.5a).

Degradation of the DNA, as can be seen from the gel, is minimal. A fraction of this DNA was tested to determine if the ends of the fragments were able to be ligated and thus able to recombine in vitro. The DNA was treated with T4-ligase for 5h. at room temperature and elecrophoresed on a

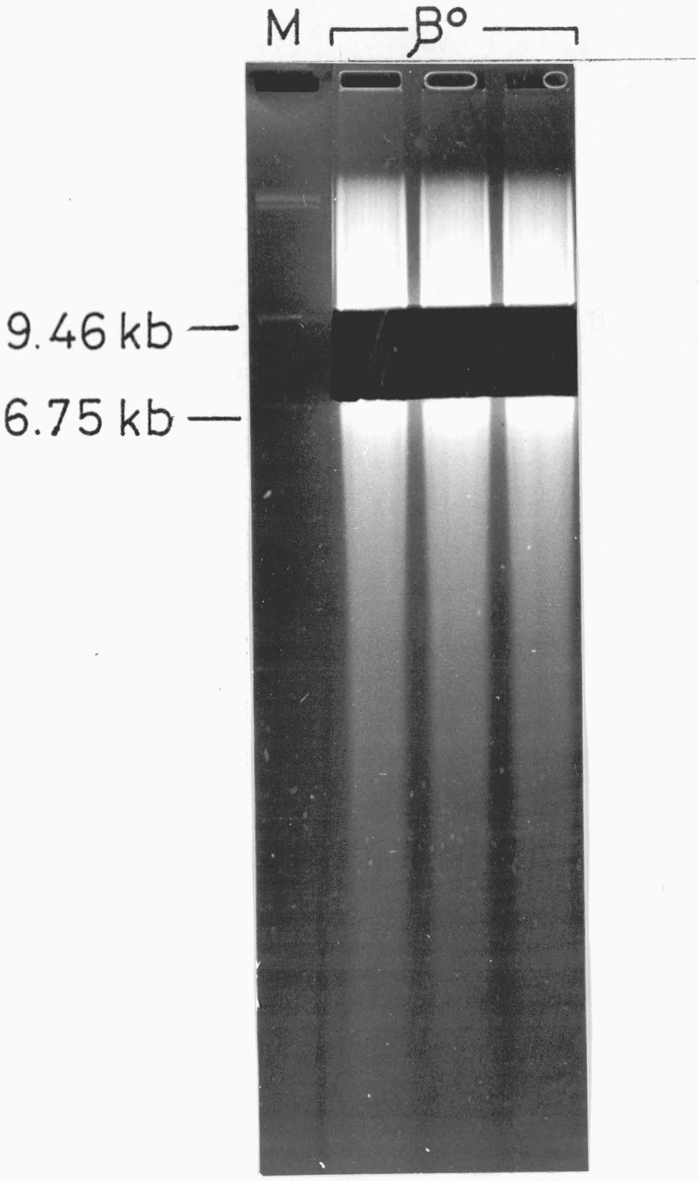


Figure 3.4.

Small-scale preparative agarose gel electrophoresis. HsuI-digested DNA from a patient with homozygous β°-thalassaemia electrophoresed on a 1% agarose gel. The marker track was stained seperately to determine the position of fragments between 9.46 and 6.75kb. After removal of this DNA the rest of the gel was stained and photographed under UV illumination.

M: λ-DNA digested with HsuI.

0.7% agarose gel to assay for ligation of the fragments. A photograph of the ethidium bromide stained gel is in Figure 3.5b. A proportion of the DNA has been ligated to form concatenates. The DNA which has not changed its mobility will consist of two fractions, the fragments which have circularised and the fragments which have not ligated. Thus, at least some of this enriched DNA is capable of recombining in vitro.

c) RPC-5 Chromatography.

2 mg of HsuI digested DNA, from a patient with β^{o}-thalassaemia, was bound to an RPC-5 column, and eluted with a 70 ml gradient of 1.5 to 1.7 M sodium acetate in 50mM Tris HCl pH 7.5 1mM EDTA, as described in 2.3h. 40 fractions were collected and concentrated by ethanol precipitation. An aliquot of each fraction was run on a 1% agarose gel (Figure 3.6) which was transferred to nitrocellulose and hybridised with ^{32}P labelled pHβ G1 to identify the β-globin gene containing fractions. An autoradiograph of the filter is shown in Figure 3.7.

The fractions enriched for the β-globin gene are indicated in this figure. By densitometric comparison of the stained DNA tracks, it was estimated that 1.49μg of fraction 28 were transferred from the gel. A comparison of the hybridisation to fraction 28 and to 10μg and 2μg of unfractionated DNA allows an estimate of 6.7 to 7.3-fold for the enrichment with respect to the β-globin gene.

d) Partial Digestion with EcoR1.

As discussed in 3.1c, partial digestion of human DNA

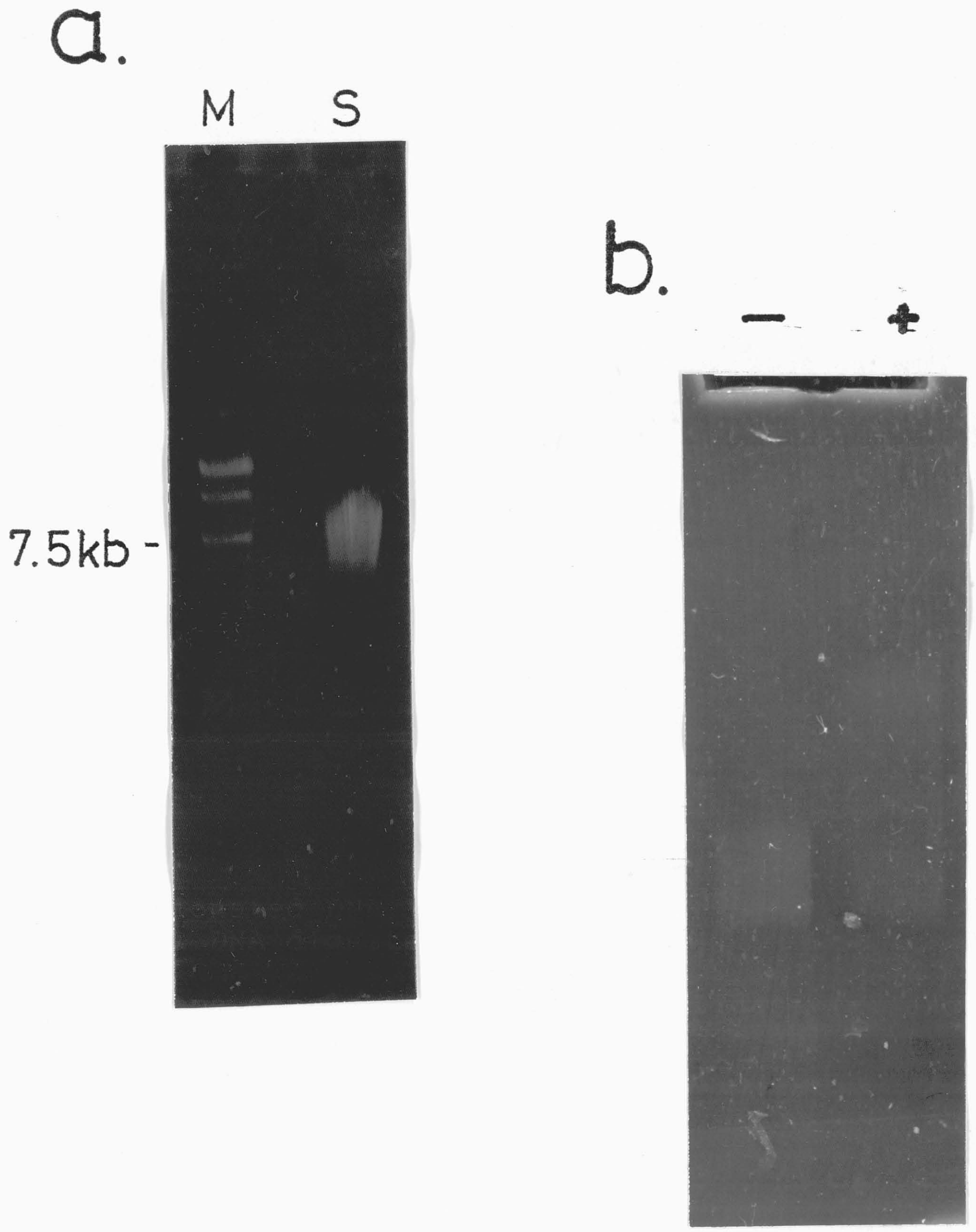


Figure 3.5.

a) DNA prepared on the agarose gel shown in Figure 3.4, re-electrophoresed on a 1% analytical agarose gel.

S: fractionated DNA.

M: λ-Padl DNA digested with HsuI.

b) The same sample treated with T4 ligase (+) and untreated (-). Analysis is on a 0.7% agarose gel.

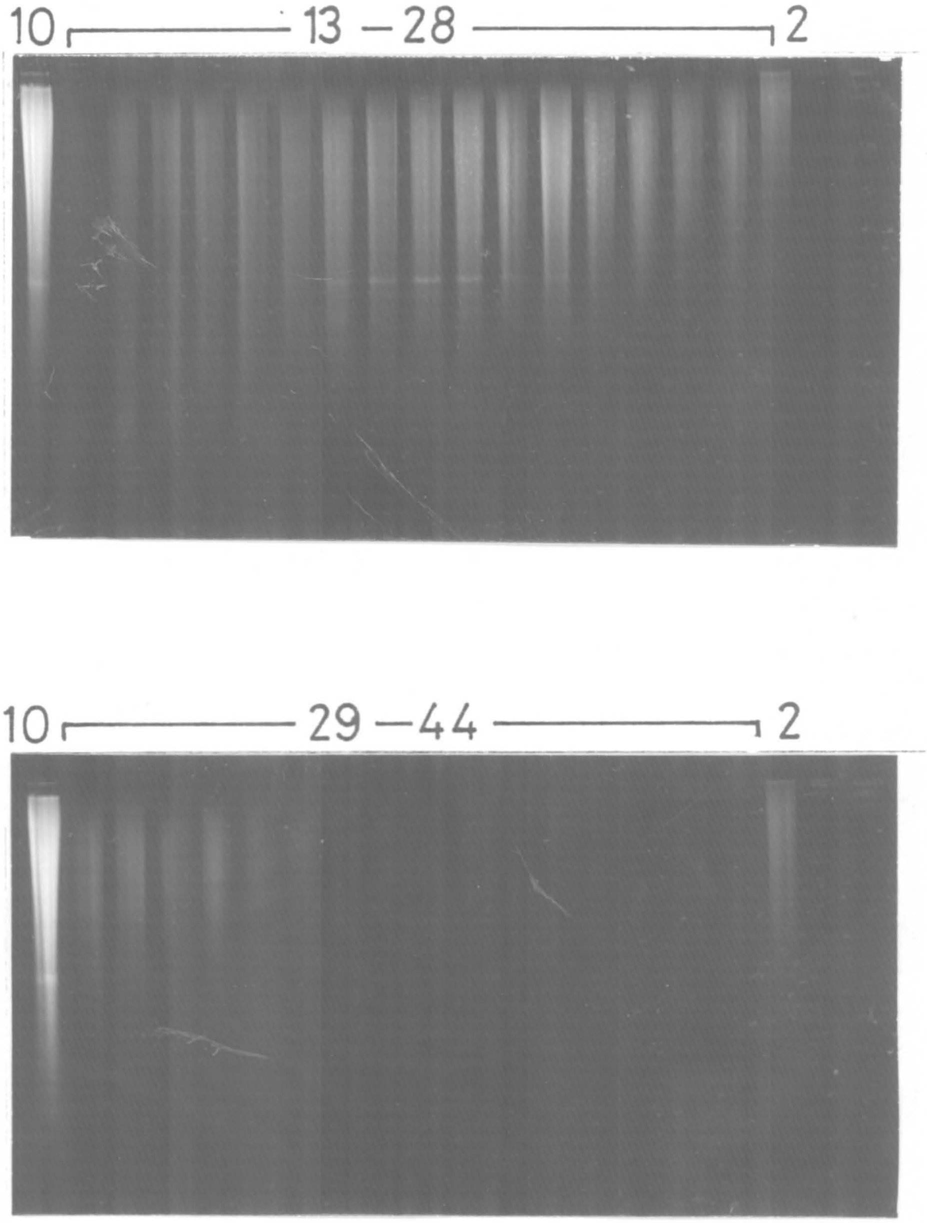
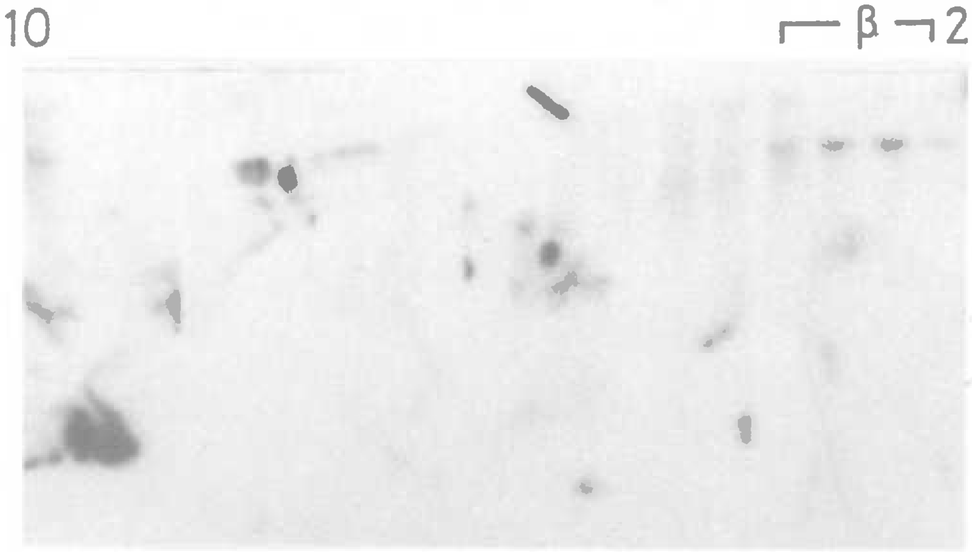


Figure 3.6.

RPC-5 fractionation of the DNA from a homozygous β°-thalassaemic patient, digested with HsuI.

Aliqouts of each fraction were electrophoresed on a 1% agarose gel.

10 and 2 are 10ug and 2ug of total, unfractionated DNA.



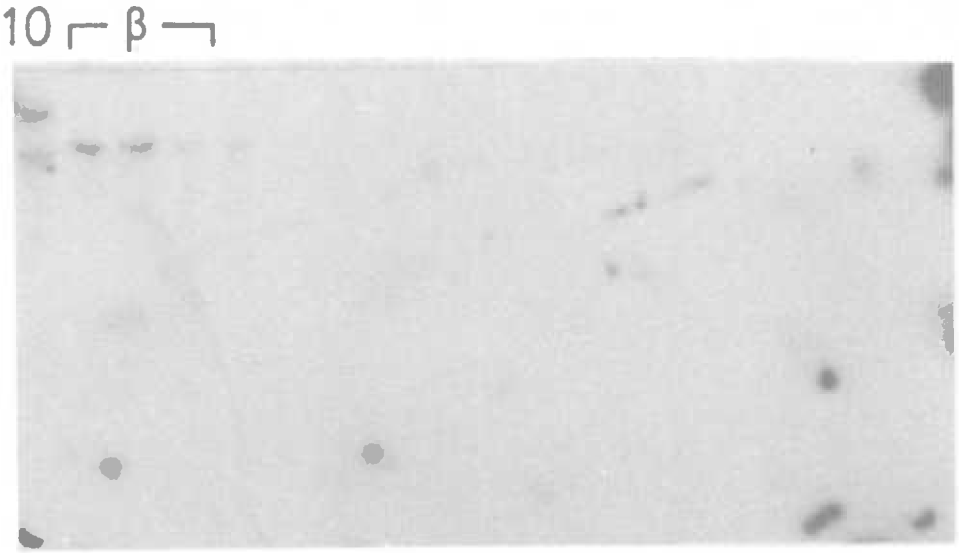


Figure 3.7.

Southern blot hybridisation of the DNA electrophoresed and shown in Figure 3.6. The filters were hybridised with pHβG1.

10 and 2 again refer to 10ug and 2ug of total DNA.

The tracks marked β contain the β-globin specific band. The band of higher molecular weight which is visible is the δ-globin band.

with EcoRl can be used to generate a library of overlapping DNA fragments. As it is known that EcoRl cuts the β-globin gene in the coding sequence, this was again used as a model to test the method.

One unit of EcoRl digests to completion 1μg of DNA in 1 hour. The time course of the reaction follows Michaelis-Menton kinetics, so that the digestion is rapid initially and tends towards a plateau. If a partial digestion is carried out with the full amount of enzyme for only a fraction of the time it is difficult to calculate the degree of under-digestion acheived. On the other hand, to digest the DNA with a fraction of the amount of enzyme required for a complete digest, for the full time, will give a partial digest which can be assessed accurately in terms of the fraction of the total EcoRl sites which have been cleaved. In this experiment 1000μg of DNA from a normal individual was digested for 1h. with either 300 units or 500 units of EcoRl. These were regarded as 30% and 50% digestions, respectively.

After phenol extraction, chloroform/octanol extraction and ethanol precipitation, the products of the two digestions were fractionated by sedimentation through 10-30% sucrose gradients. The fractions were analysed by agarose gel electrophoresis and those fractions with a modal fragment size of between 8 and 15 kb were pooled and refractionated until very little material was visible at less than 8kb.

To check that the large fragments were derived from partial digests of smaller fragments, a part of each digest was further treated with EcoRl to restrict them to completion. These samples were run on a 1% agarose gel

(Figure 3.8) and transferred to a nitrocellulose filter before hybridisation with nick-translated pHβG1 (Figure 3.9).

The autoradiograph shows that fully digested fragments are still clearly present even after 30% digestion, but that these can be removed by sucrose gradient sedimentation. The high molecular weight fraction can be seen to be hybridising to the β-globin probe, and specific β-globin fragments are produced by further digestion of this fraction.

The DNA so fractionated could be used to produce a "library" of human genes by cloning in λgtWES, at least as far as fragments of the size of the β-globin gene are concerned.

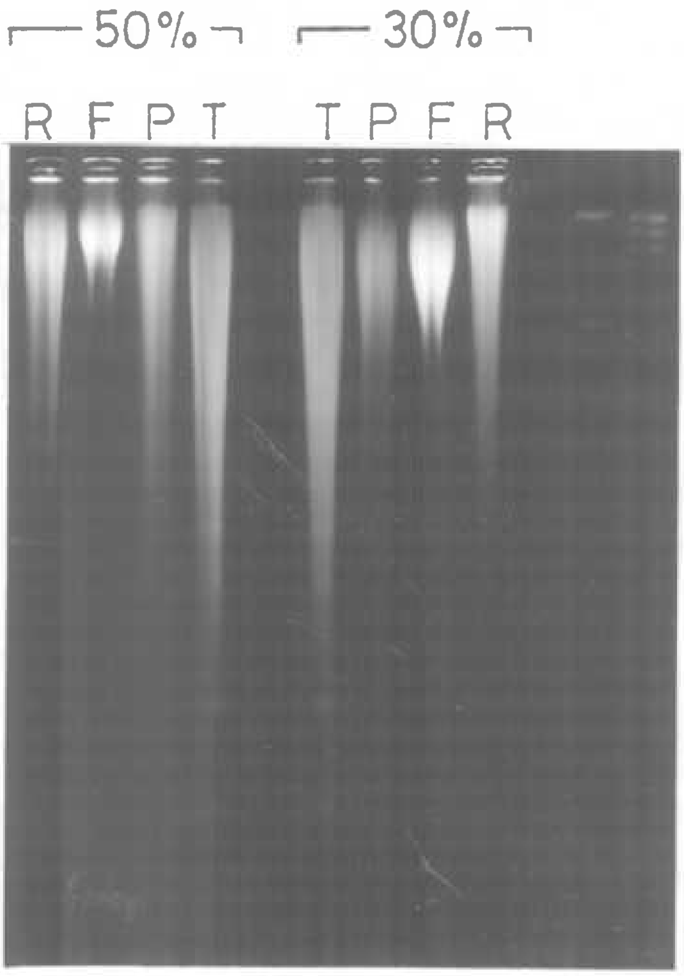


Figure 3.8.

Analysis of partial digestion of DNA with EcoRl.

30% and 50% refer to the extent of digestion, as indicated in the text.

T: total digestion.

P: partial digestion.

F: partially-digested DNA fractionated on sucrose gradients.

R: Fractionated DNA redigested with EcoRl.

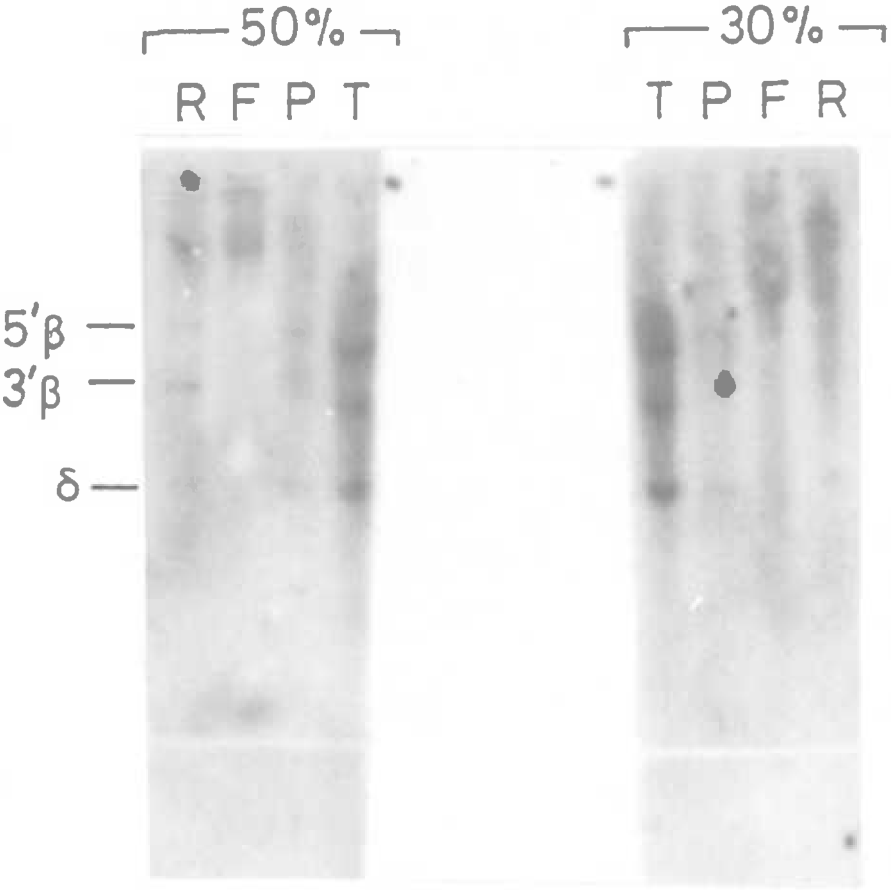

50%
R F P T
30%
T P F R
5′β
3′β
δ

Figure 3.9.

Southern blot hybridisation of the DNA on the agarose gel in Figure 3.8. The probe was nick-translated pHβG1.

Abbreviations as in Figure 3.8.

3.3 Manipulating the Vector.

a) DNA Preparation.

DNA was prepared as described in 2.4g from the λ-strain NM788, using a stock derived from a single plaque. The biological competence of the DNA was tested by packaging into phage particles *in vitro* and assaying the yield of viable viruses. The yield was compared to the yield from DNA of previously confirmed biological activity, and was identical, (in this case, 1.96 x 10^7 p.f.u./μg DNA).

b) Isolation of the "Arms".

80μg of vector DNA was digested with HsuI. As a low level of undigested DNA will give a large number of parental (background) phage, compared to the number of recombinant phage produced, the DNA was digested with a 4-fold excess of enzyme. The DNA was assayed for undigested phage chromosomes by *in vitro* encapsidation. The proportion of intact phage chromosomes was 0.01% of the DNA. To verify that the DNA remained in a potentially functional form, 0.5μg of the digest was treated with DNA ligase for 5h. at room temperature and packaged *in vitro*. The yield of virus particles was approximately 10% of undigested DNA, (or a 1000x increase over the unligated fragments), indicating that the fragments could still form viable phage.

The theoretical yield of correctly-assembled phage chromosomes in this ligation reaction is 11%. Any left arm can be ligated via its HsuI site to any of three fragments; another left arm, a right arm or the correct central fragment. Assuming this to be a random process, then one-third of these dimers will be able, by joining to a right arm, form a functional chromosome. This latter process again will occur in about one third of cases. Hence the observed yield of 10% closely approximates the theoretical maximum, indicating that the HsuI-generated termini are still intact.

The central fragment of the HsuI-digested phage chromosome was removed from the arms by centrifugation on a 10-40% sucrose gradient in 50mM Tris, pH 7.5, 1mM EDTA, 1M NaCl. Aliquots of each fraction were run on a 1% agarose gel (Figure 3.10) and fractions containing arms essentially free of the central fragment were pooled, diluted with an equal volume of water to reduce the salt concentration, and ethanol precipitated. In some cases the fractionation was repeated to further reduce the relative concentration of the central fragment.

These pools were tested for the presence of the central fragment by ligation at room temperature for 5h. and packaging in vitro. The yields of phage indicate that the first-cycle arms have a packaging efficiency of 3.5% intact DNA or about one-third of cleaved and religated fragments. After a second fractionation this was reduced ten-fold.

c) In Vitro Recombination of Human DNA with NM788.

In order to test the ability of the arms to recombine in vitro, 0.65μg of total HsuI digested human DNA was added to 0.56μg of NM788 arms and treated with DNA ligase at room temperature for 5h. The molar ratio of ends of human DNA to vector DNA was not calculated, because of the wide range of human DNA fragment size, but there is a large excess of hunan fragments. NM788 arms alone were ligated in parallel and both ligations were packaged in vitro and assayed for phage.

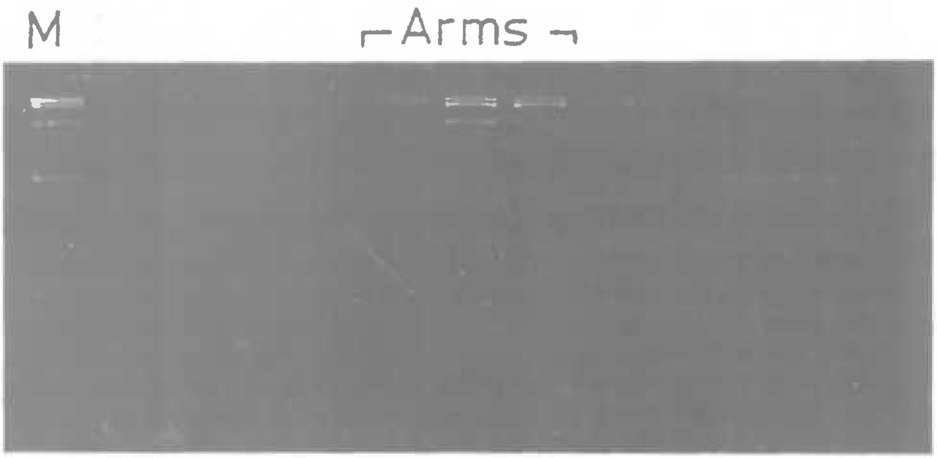


Figure 3.10.

Agarose gel electrophoresis of aliquots of fractions obtained by sucrose gradient centrifugation of HsuI digested NM788.

Fractions marked "arms" were pooled and used for *in vitro* recombination with human DNA.

M: HsuI-digested NM788 DNA.

Addition of the HsuI-cleaved human DNA; a suitable substrate for insertion between the arms, resulted in a 1.85x increase in viable phage over background, which implied that 47% of these phage were human DNA-containing recombinants.

Ten plaques were picked and a small-scale DNA preparation made from each. This DNA was digested with HsuI and electrophoresed on a 1% agarose gel to reveal the inserted DNA fragments. Figure 3.11 shows this analysis and reveals that only one out of the ten was, in fact, a non-recombinant phage (track 7). This might sugest that 90% of these phage are recombinants. Most of the recombinant phage have multiple inserted sequences.

The fragments of human DNA in the recombinant chromosome in track 2 appear to be at a molar concentration much less than the concentration of the arms in the same track. This may be due to the phage being a mixture of clones, not separated when the plaque from which the DNA was made was picked (although efforts were made to avoid it). Alternatively, it may have been a single clone, containing a large inserted sequence which undergoes spontaneous deletions during propagation, generating the series of bands seen in the track. This phenomenon has been observed previously in phage containing repeated DNA sequences, includind the human α-globin genes (Lauer _et_ _al_, 1980) and human satellite DNA (Cooke and Hindley, 1979).

The difference between the two estimates of the proportion of recombinant phage derived above is probably due to the fact that a comparison of the yield of viruses from arms ligated with and without human DNA does not take into

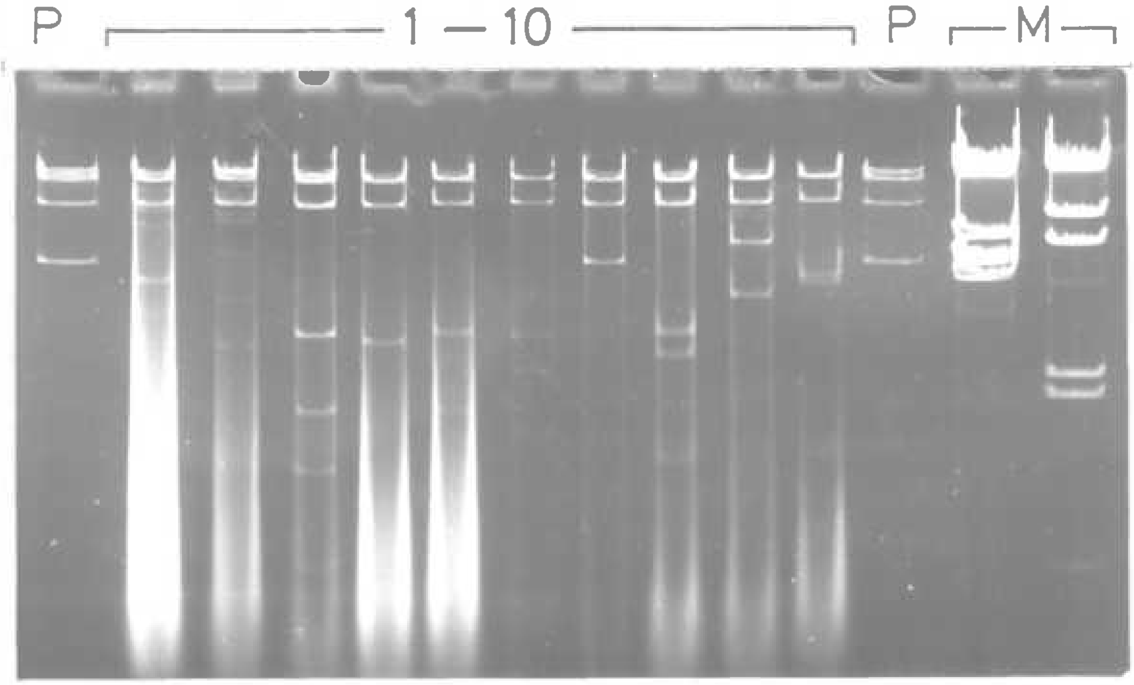


Figure 3.11.

Agarose gel electrophoresis of HsuI digested DNA, isolated from phage produced by in vitro recombination with human DNA.

P: parental phage DNA (NM788), digested with HsuI.

M: λ-DNA digested with EcoR1 (left) and HsuI (right).

account any competition between the human DNA and the central phage-derived fragment. In this experiment there was a vast excess of human DNA which would effectively compete with the central fragment. In any case it is apparent that the more accurate estimate of the proportion of recombinant phage is obtained by direct analysis, using small-scale DNA preparation.

3.4 Isolation of a β-Globin Gene From a Patient With Homozygous β^0-Thalassaemia.

a) The Patient.

The DNA used in this study was made from the spleen of a patient with homozygous β^0-thalassaemia. The patient, a young male, has been previously studied by Dr. Sergio Ottolenghi and his colleagues at the Institute of General Pathology, and the Paediatric Clinic, at the University of Milan. In their work, a cDNA probe, made against the mRNA from the reticulocytes of a patient with HbH disease, (and therefore enriched for β^0-globin sequences), was hybridised to the DNA from this patient in cDNA excess. The hybridisation was indistinguishable from a hybridisation to normal DNA. They concluded that the β-globin genes were present in this patient, although non-functional (Comi et al, 1977). In the same study, the cytoplasmic and nuclear RNA from this and one other patient were hybridised to the cDNA probe. In neither patient could hybridisation be detected above background to cytoplasmic RNA, implying an absence of mRNA. However, in the patient studied here, there was a significant level of RNA in the nucleus hybridising to the β-globin probe. In the other patient there was none. The hybridisable RNA amounted, in some measurements, to about 50% with respect to the α-globin RNA.

A model can be formulated for the molecular pathology of the disease in this patient. Thalassaemia is known to be a disease of heterogenous pathology. The three patients in whom a deleted β-globin gene appears to cause β^0-thalassaemia are all heterozygous for the deleted gene, their other chromosomes carrying a non-deletion thalassaemic gene. It is reasonable to believe that the patient studied here may be heterozygous for two, different, β°-thalassaemia genes. With this in mind, the simplest model has one of the two genes transcriptionally inactive, whilst the other is transcribed at a normal rate, resulting in 50% normal levels of nuclear RNA. The transcripts from this gene do not reach the cytoplasm because of some defect in processing or transport.

Other models are, of course, possibly applicable; the experiments described here are intended to begin to test these hypotheses and to attempt to define the molecular leision(s) in this patient.

b) Gross Structure of the β-Globin Gene.

The conclusions drawn in the earlier study were confirmed and extended using the Southern blot technique and a cloned β-globin gene as a probe.

The patient had previously been therapeutically splenectomised, and DNA was prepared from a portion of the spleen as described in 2.3a. This DNA, and DNA made from the white blood cells of a haematologically normal individual, were digested with HsuI, MspI, BglII and XbaI, and elecrophoresed in parallel on a 1% agarose gel. After transfer to a nitrocellulose sheet the DNA was hybridised with 32-P labelled b-globin DNA. An autoradiograph of the filter is shown in Figure 3.12a. Figure 3.12b shows an identical experiment digesting with EcoR1 and PstI. The probe in 3.12a was a cloned PstI fragment of chromosomal DNA, containing the β-globin gene (Hβ1-S, a gift from Dr.T.

Maniatis), and in 3.12b was a cloned β-globin cDNA (Little et al, 1978).

As can be seen there is no visible difference between the two DNA samples. The HsuI digestion of the thalassaemic DNA is not complete, and there is a region of hybridising material near the top of the gel. Incomplete digestion of this DNA with HsuI was a persistent problem. Using the chromosomal gene as a probe (Figure 3.12a), only the β-globin-specific bands can be seen. As the δ-globin fragments are not strongly hybridised, the probe is homologous to them over only a fraction of its length and no extensive sequence homology flanks the two genes. Furthermore, no other hybridisation is seen to bands in any track; the DNA in the 4.4kb fragment is unique throughout the genome.

The sites for the enzymes used in this experiment lie up to 9kb to the 5' and 4kb to the 3' side of the β-blobin gene. Figure 3.13 shows the arrangement of these sites, taken from Flavell et al (1978) and Van der Ploeg and Flavell (1980). Two of the enzymes (XbaI and MspI) give fragments which include the δ-globin gene. Thus, in this 13kb region of the δ/β globin gene locus there is no detectable deletion, insertion or rearrangement of DNA sequence with respect to the normal gene. If the thalassaemic leision(s) is in this region then it is not due to a large deletion of DNA with respect to the normal gene. If the precise location of the genetic defect can be pinpointed then, as it is presumably due to only a small sequence change, a functionally important structure will be identifiable.

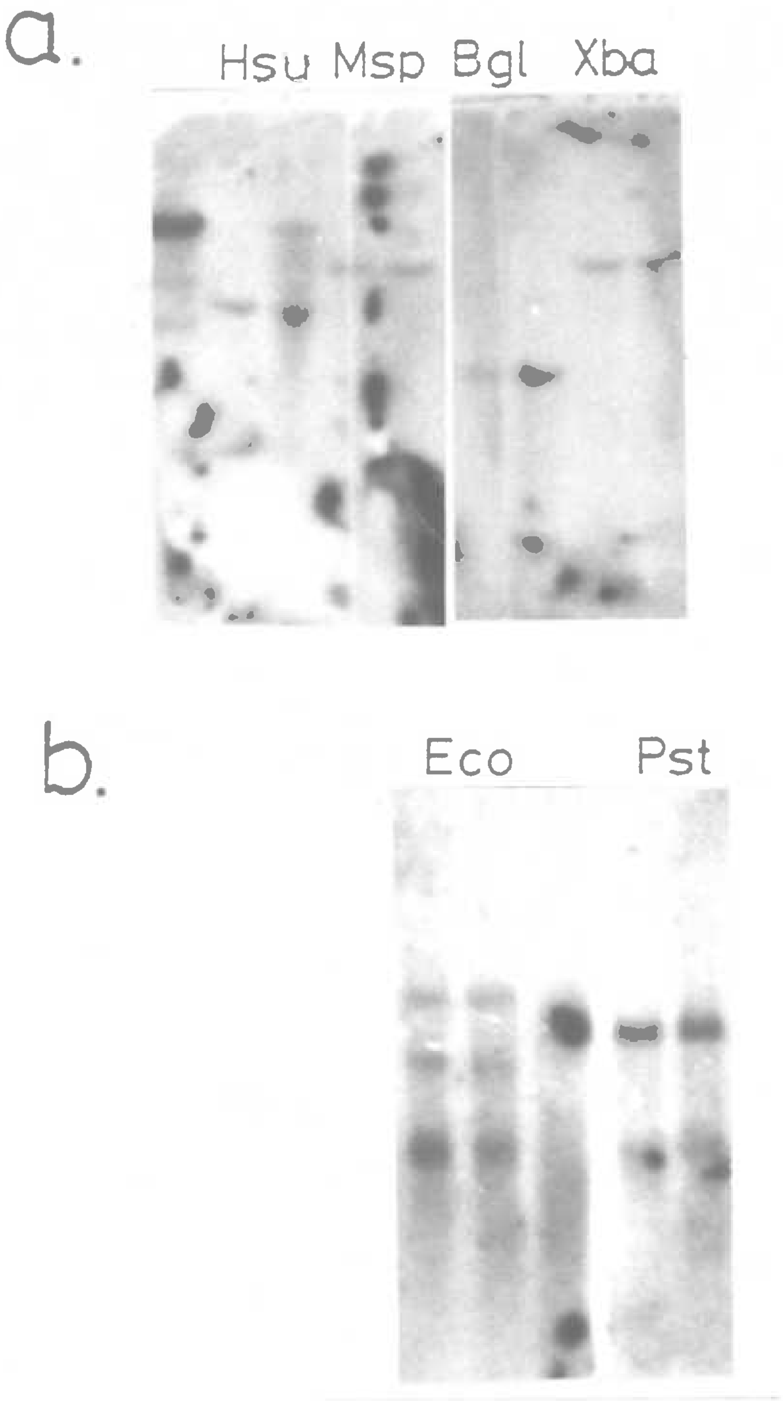


Figure 3.12.

Southern blot hybridisation to DNA from the patient referred to in the text, and to normal DNA.

The two DNA samples were electrophoresed in adjacent tracks. In these autoradiographs the normal DNA is on the left, and the thalassaemic DNA on the right.

a) DNA cut with HsuI, MspI, BglII, XbaI. The filters were hybridised with nick-translated H 1-S, which contains the 4.4kb fragment of the human β-globin gene.

b) DNA cut with EcoR1 and PstI, and probed with pHβG1.

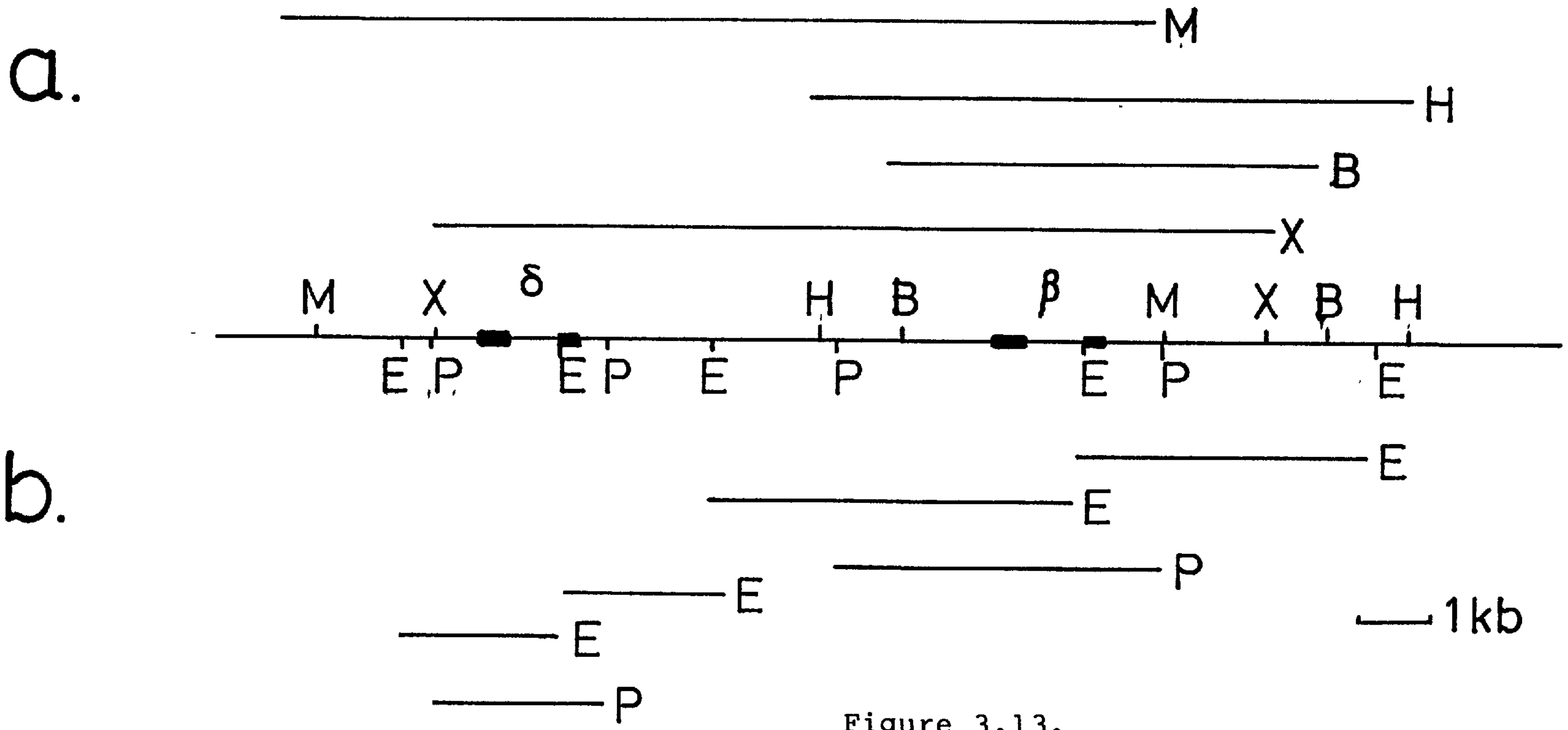

a.
M
H
B
X
M
X
δ
H
B
β
M
X
B
H
E
P
E
P
E
P
E
P
E
b.
E
E
P
E
E
P
1kb

Figure 3.13.

Restriction enzyme map around the δ/β-globin locus.

The bars indicate fragments visualised in Figure 3.12, a and b refer to parts a and b of that figure.

M: MspI, H: HsuI, B: BglII, X: XbaI, E: EcoRl, P: PstI.

c) Construction of a β^{o}-Thalassaemic Gene-Containing Recombinant

In order to study the globin genes from this patient in more detail, a recombinant phage was isolated which contained the HsuI fragment containing the β-globin gene. Figure 3.13 shows that this is a 7.5kb fragment, enclosing the gene with approximately equal amounts of DNA on each side. Bearing in mind the provisos in Section 1.4c, it is assumed that the β-thalasaemic leision is in or very close to the β-globin gene. The β-globin gene from a thalassaemic patient is called the "β--thalassaemic gene" in this thesis.

Using the methods developed in 3.2 and 3.3, a collection of recombinant phage was constructed, consisting of the vector NM788 and a fraction of DNA from the patient, enriched for the β-globin gene.

100μg of HsuI digested DNA from the patient was electrophoresed in ten tracks on a 1% agarose gel. By comparison with co-electrophoresed size markers the region of the gel containing DNA fragments between 6.5 and 9kb was removed, and the DNA electroeluted from it, as discussed in 3.2b. The recovery of DNA was estimated by running a portion on a gel and comparing the intensity of staining with a marker of known concentration. The recovery was estimated to be approximately 7μg. The DNA was ligated to itself, and analysed by agarose gel electrophoresis to test its ability to recombine in vitro.

1μg of purified NM788 arms were mixed with 0.28μg of the enriched DNA fraction, and treated with T4-ligase for 5h. at room temperature. Three identical reactions were carried out. This DNA and ligated arms alone were packaged in vitro into phage particles. The yield of viable phage was increased at least ten-fold by addition of the human DNA to the phage arms. The phage titre varied from 40,000 to 80,000 in each reaction.

Small-scale DNA preparations were made from 12 clones, digested with HsuI, and electrophoresed on a 1% agarose gel, which is shown in Figure 3.14. The samples are flanked by markers of HsuI-cut NM788 and λ-Pad1 (a recombinant derived from NM788 containing the HsuI β-globin fragment from a patient with β^+-thalassaemia; made by D.Westaway). The gel indicates that 5 of the 12 were non-recombinant. Of the remaining seven, another two contain two small fragments, suggesting that these derived from fragments of DNA containing an HsuI site. These must have been products of an incomplete digestion. The HsuI-digested DNA used in this experiment is the same as was hybridised in Figure 3.12a, in which the incomplete nature of the digest is clear.

It was thus estimated that about 50-60% of the phage were recombinants, although only 30-40% were "useful" in that they carried a DNA fragment of the desired size. Approximately 180,000 phage were prepared, of which about 65,000 were of interest. Using the formula of Clarke and Carbon (1976), the probability of a collection of recombinants containing a particular sequence can be calculated.

This formula is; $P = 1-(1-f)^N$

where P is the probability, N is the number of recombinants, and f is the

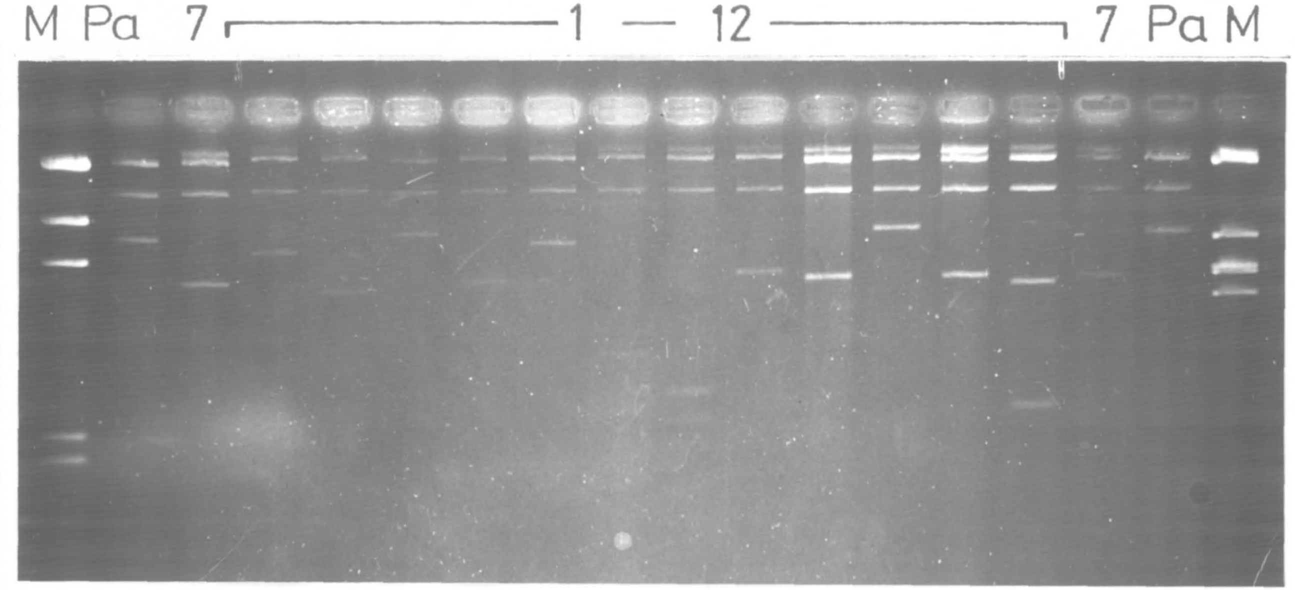


Figure 3.14.

Agarose gel electrophoresis of HsuI-digested DNA isolated from phage made by *in vitro* recombination with human DNA, as described in the text.

M: markers, λ-DNA digested with HsuI (left) and EcoRl (right).
Pa: λ-Padl DNA digested with HsuI.
7: λ-NM788 DNA digested with HsuI.

fraction of the genome contained in each clone.

If it is assumed that the enrichment of the globin gene was about 5x, and the human genome is about 2 x 10 kb, then f is 0.0000188. The probability that the 90,000 recombinants prepared here contain one clone "positive" for the β°-thalassaemic gene is about 70%.

d) Identification and Isolation of a β°-Thalassaemic Gene-Containing Recombinant.

The three preparations of recombinant phage, described above, were screened by the Benton and Davis plaque hybridisation method (1977).

Each of the preparations was separately grown on a lawn of LE392 on a "megaplate" as described in section 2.4k. To provide a control in the subsequent screening, a spot of 1µl of the δ/β -globin recombinant, HβG2, containing 10 to 20 phage, was added to the lawn in two corners of the plate. After growth the plates were chilled overnight at 4° . Two nitrocellulose filters were placed in contact with the lawn in succession, the first for two minutes and the second for three. These picked up duplicate imprints of the plaques from each plate. The filters were treated as described in Materials and Methods to denature the DNA, baked at 80° for two hours, and prepared for hybridisation. The probe used for hybridisation was a plasmid containing the PstI fragment of the chromosomal β-globin gene, as used in 3.4b. The plasmid was labelled by nick-translation to 5×10^{7} dpm/ug and used at a concentration of 2ng/ml. It had been previously observed that probe concentrations significantly higher than this gave an inferior signal-to-noise ratio.

After hybridising for 36h the filters were washed repeatedly in 3xSSC/10xDenhardt's solution, and autoradiographed overnight. On one plate a spot of hybridisation was seen, and is shown in Figure 3.15. The control phage can also be seen on the autoradiograph. These serve to indicate the extent of hybridisation to known positive plaques, and to allow the two duplicate filters to be alligned with a reasonable degree of accuracy. This allignment shows that the positive plaque seen is in the same position on each filter.

Several plaques in the region of this hybridisation were identified by aligning the filter to the plate, and these were picked with a sterile pipette tip into 150µl of phage buffer. These phage were then plated on a bacterial lawn on 90mm Petri dishes, using 1, 10 and 20µl of the suspension. Duplicate filter impressions were taken from these plates, treated as before, hybridised and washed. Autoradiographs of one set of these filters are shown in Figure 3.16. Hybridisation was seen to about one third of the plaques, which agrees roughly with the number of plaques picked from the megaplate. A positive plaque, well separated from its neighbours, was picked from the 1µl plate and a small-scale DNA preparation was made from it.

A fraction of this DNA was digested with HsuI and electrophoresed as shown in Figure 3.17. The fragment inserted into this recombinant, as can be seen, is of the same size as the fragment in Pad1, which contains a β^+-thalassaemic fragment of the size expected of the β^0-thalassemic gene. There is no detectable size difference when the two fragments are run together in the same track.

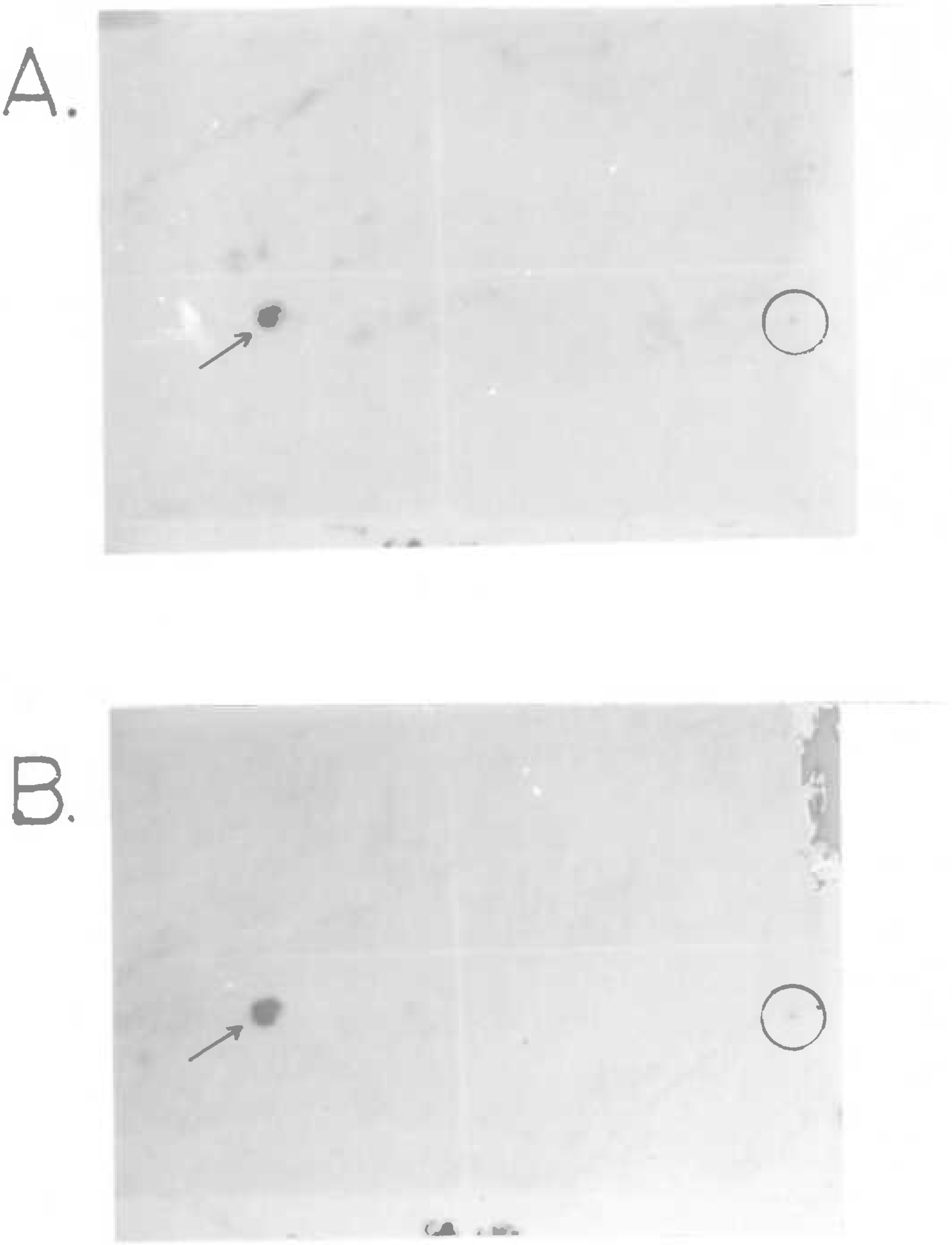


Figure 3.15.

Plaque hybridisation analysis of recombinant phage. A and B are duplicate filters, taken from the same plate.

The strong hybridisation (arrowed) is to control plaques of λ-HβG1 added to the plate. Hybridisation to the phage subsequently purified is circled.

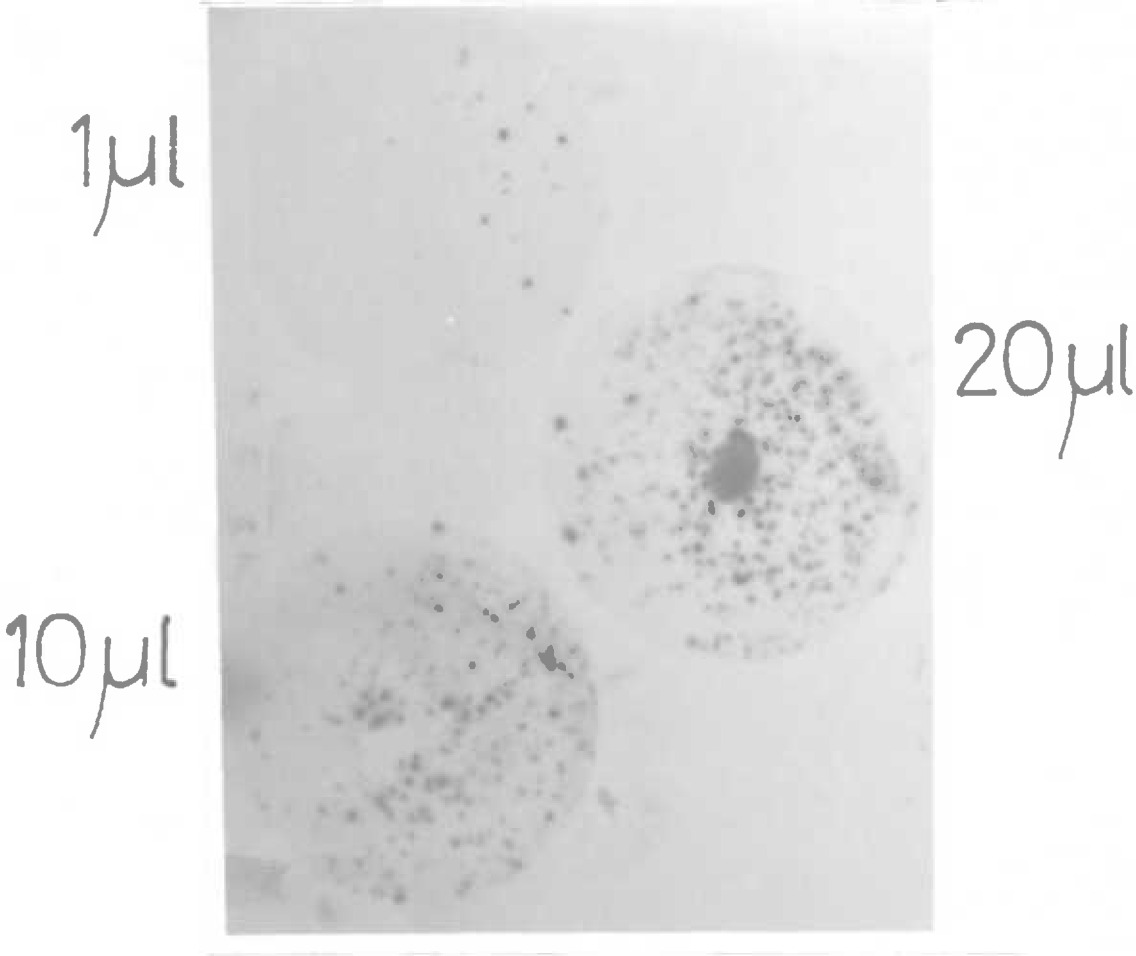


Figure 3.16.

Hybridisation to plaques obtained from the area circled in Figure 3.15 and replated. Different volumes of the phage suspension were used, as indicated.

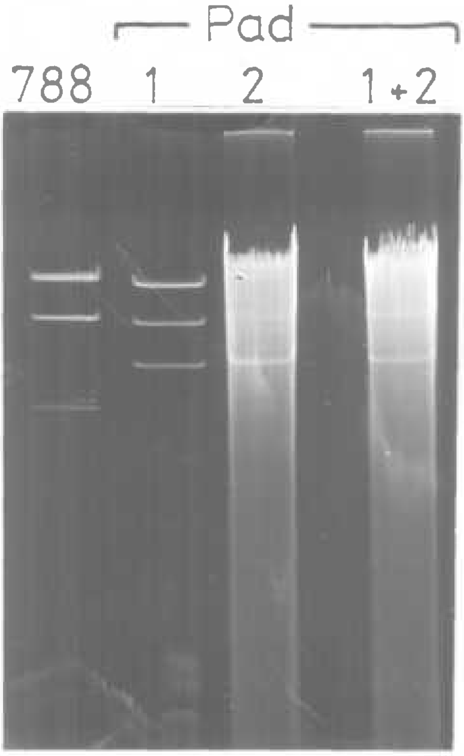


Figure 3,17.

Gel electrophoresis of HsuI-digested DNA isolated from a positively hybridising phage of Figure 3.16.

788: λ -NM788 DNA)
1: λ -Pad1 DNA) all digested with HsuI.
2: λ -Pad2 DNA)
1+2: Pad1 + Pad2 DNA)

Mapping data in Section 4.1 indicates that this phage does not have the same restriction map as Pad1. It is not, therefore, an accidental reisolation of the same recombinant. This phage was designated λ-NM788/β°/Pad2. A large scale DNA preparation was made from the clone, and used in subsequent studies.

3.5 Isolation of the Gene for Fibroblast Interferon (IFN-β) From a Human Gene Library.

In recent years, libraries of genes from several different organisms have been prepared, using the methods outlined in 3.1c. Maniatis and his coworkers (1978) have described the construction of, among others, a human gene library. The availability of these libraries has meant that the isolation of any normal gene for which a hybridisation probe exists is relatively easy.

Workers at Searle Research have made a fibroblast interferon cDNA recombinant plasmid. This cloned cDNA was used as a probe to screen the human gene library made by Maniatis *et al*.

a) Screening the Library.

The sample of the library, as originally supllied by Dr.Maniatis, contained only about 5×10^6 recombinants. It was therefore amplified prior to screening using the PDS method, as described in 2.4fii. 3×10^6 phage were mixed with 2.5×10^9 cells of LE392 and preadsorbed before diluting into 500ml L-broth and incubating at 37°. After 18h. growth the cell-debris was cleared and the phage suspension titred, which showed a 500x amplification.

Six "megaplates" were used to screen 360,000 recombinants. Duplicate nitrocellulose filter impressions were hybridised in two sets with a fibroblast interferon probe. This probe was a 700bp HindIII fragment from a cloned fibroblast interferon cDNA, eluted from an acrylamide gel at Searle Research and nick-translated with 32-P nucleotide triphosphates. The screening revealed seven positive clones on three out of the six plates; the other three had none, even after much longer autoradiography to expose any weak signals. Figure 3.18 shows the autoradiographs of the three plates with positive hybridisation. The distribution of the hybridising clones is certainly odd and no explanation, other than chance, is readily available.

Plaques from the region of the megaplates corresponding to the hybridisation were picked with a sterile tip into 100ul of phage buffer and replated at a low density. Impressions of these plates were taken onto nitrocellulose and hybridised with the interferon cDNA. Six of the seven plates gave positive signals on these filters; the phage on the seventh, which was negative, were replated at a higher density. This plate, on rescreening, showed positive hybridisation.

Plaques which were identified as having hybridised in the second round of screening were picked into phage buffer and DNA prepared from these clones by the method of Cameron et al (1977). A small number of phage from each clone were grown again on plates, rescreened by hybridisation with the nick-translated cDNA probe to ascertain that the plaques were positive, and a single plaque picked for each recombinant.

Six independently identified clones were analysed by EcoRl digestion and agarose gel electrophoresis. Figure 3.19 is a photograph of five of the recombinants. The two "arms" of the Charon 4a vector used in the construction of the library can be seen near the top of each track. The inserted human DNA is the several bands of lower molecular weight,

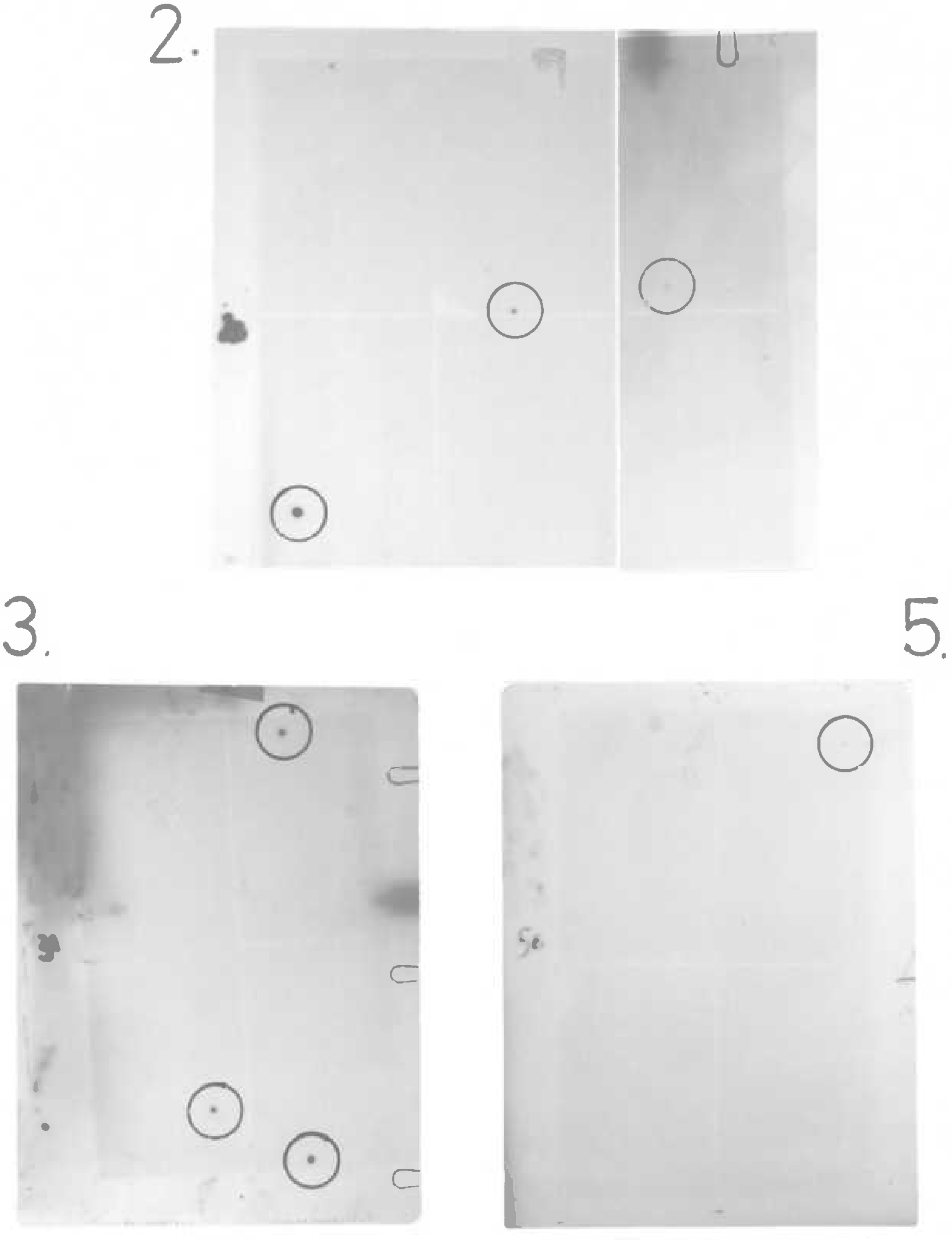


Figure 3.18.

Plaque hybridisation to recombinants from the human gene library, using the fibroblast interferon probe. The seven positive plaques are circled. All these spots also appeared on the duplicate filters.

plus the band migrating very close to, and slightly slower than, the smaller arm. It can be seen that each track is apparently identical, except for additional bands in 2.1 and 3.1. These additional bands must be due to a contaminating phage, as the sums of the inserted fragment sizes in these two tracks are about 30kb and 35 kb, respectively, which is greater than the capacity of Charon 4a. The contaminants were eliminated by repicking a positive plaque.

The DNA of the digests of clones 2.2, 3.1, 3.2, 3.3 and 5.1 was transferred to nitrocellulose and hybridised with the interferon gene probe. The fragment of approximately 1.5kb was shown, after autoradiography, to hybridise strongly in each of the tracks (Figure 3.20).

As all the clones examined appeared to be identical, only one, 3.1, was grown on a large scale. DNA was prepared from this clone and used in subsequent structural analyses.

b)Problems of Amplification.

The initial phage population of three million recombinants had been derived from an original library of 600,000 independent clones. According to the equation of Clarke and Carbon (1976), this library had a probability of over 99% of containing the entire human genome (assuming that it was, in fact, a random collection). It should be noted that the library is not in this ideal state; Fritsch et al (1980) found it to be lacking the γ-globin genes. If the amplification stages were all faithful in preserving the complexity of the library, then the final stock of recombinant phage should contain almost all the human genes. The probability is over 90% that this collection contains

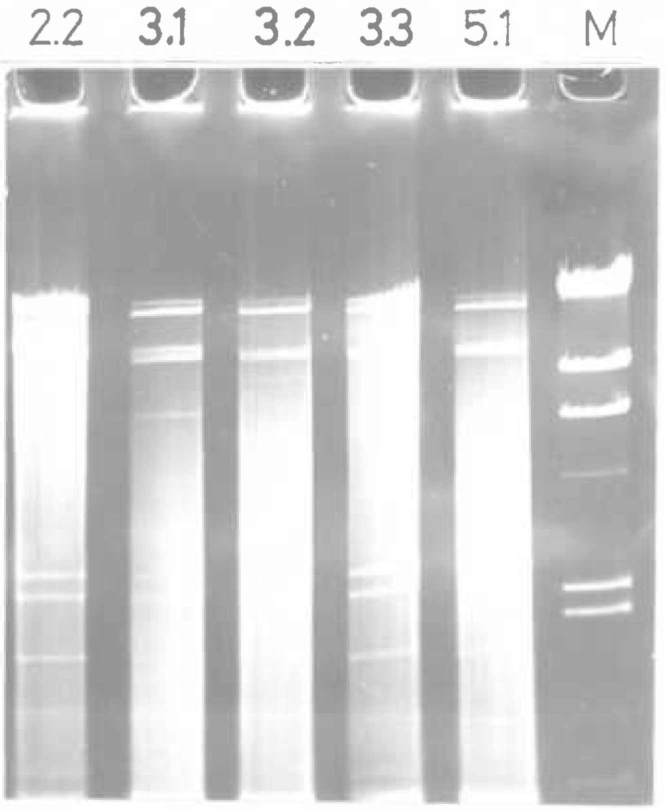


Figure 3.19.

Agarose gel electrophoresis of EcoR1-digested DNA isolated from plaque purified phage, identified in Figure 3.18.

M: λ-DNA digested with HsuI.

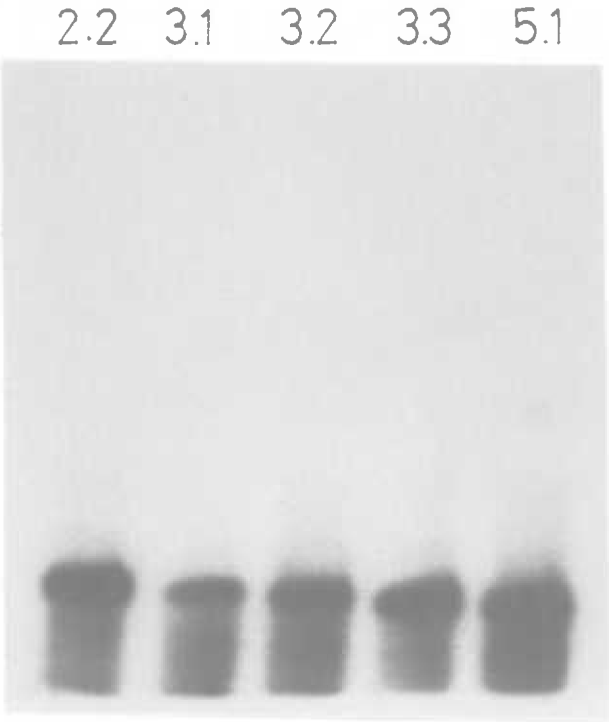


Figure 3.20.

Southern blot hybridisation to the DNA electrophoresed on the agarose gel shown in Figure 3.19.

The probe used was the fibroblast interferon cDNA-containing plasmid.

two, independently derived, copies of each gene; and is over 50% that it contains over nine copies.

After screening 360,000 members of this library, seven clones were identified; at least six of which are identical. As the method used to generate the fragments for the library should produce quasi-random, overlapping, fragments, it is highly unlikely that two clones will contain the same fragments. It must therefore be concluded that the six recombinants examined here are, in fact, siblings from a single clone from the original library stock.

The likeliest explanation for this observation is that the amplification steps do not give equal amplification of all the recombinants. Patient _et al_, (1980) have observed a similar phenomenon when growing recombinant phage containing _Xenopus_ DNA. Obviously care must be taken if libraries are to be continually amplified, rather than constructed anew.

4. Structure and Function of a Cloned β^o-thalassaemic Gene.

The clone designated λ-NM788/β^o/Pad2 (abbreviated to Pad2) was digested with various enzymes to compare its structure with λ-NM788/β^+/Pad1, which is another recombinant containing the HsuI β-globin gene fragment, and which was made and propagated in this laboratory.

The HsuI human DNA fragment was then subcloned into pAT153 and further restriction enzyme mapping carried out to determine if any deletions, insertions or gross rearrangements with respect to the normal β-globin gene could be detected.

Finally, the β^o-thalassaemic gene was transcribed *in vitro* and its function compared to a normal β-globin gene.

4.1. Structural Studies.

a) Structure of Pad2.

Samples of DNA from Pad2 and the vector, NM788, were each digested with HsuI, EcoR1, BamHI, and the combinations EcoR1+HsuI and EcoR1+BamHI. Each sample was electrophoresed on a 1% agarose gel, stained in ethidium bromide and photographed in UV light. The DNA was transferred to nitrocellulose and hybridised with a nick-translated β-globin cDNA probe, pHβG1. Both the photograph and autoradiograph from this gel are in Figure 4.1.

Fragments originating from the β-globin gene were determined by comparison with the adjacent NM788 track. Their sizes are listed in Table 4.1 and those fragments hybridising to the cDNA probe are marked with an asterisk. Reference to

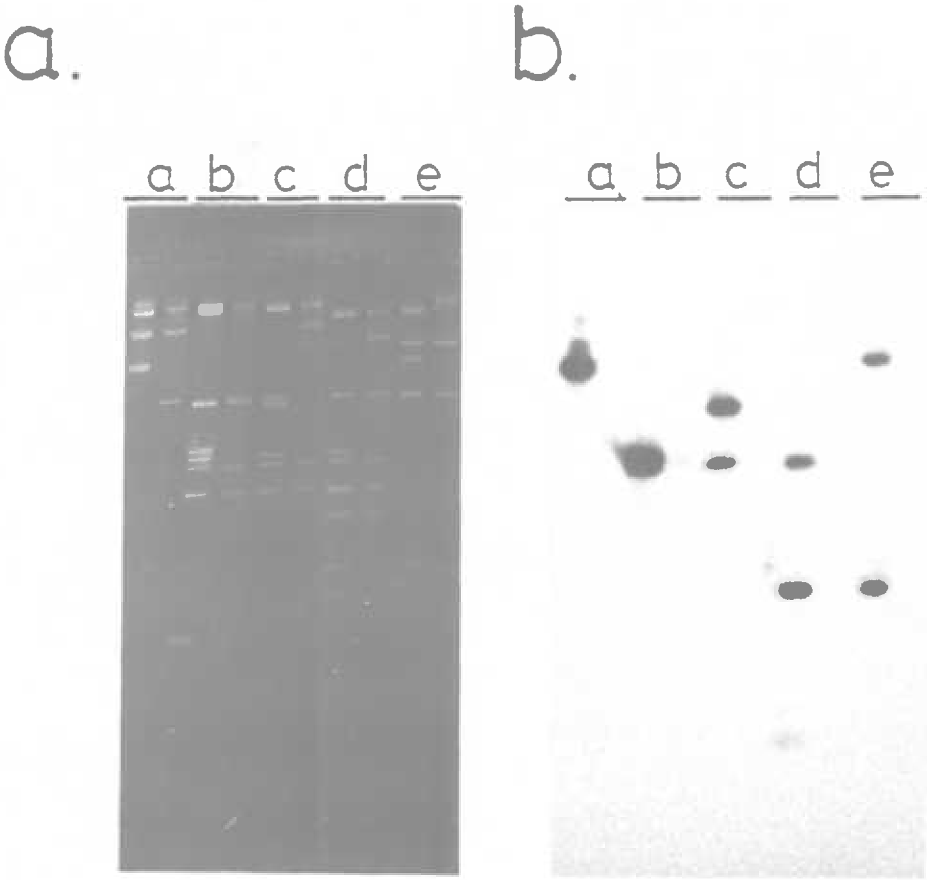


Figure 4.1.

Agarose gel electrophoresis (a) and Southern blot hybridisations (b) to DNA from NM788/β°/Pad2 and from NM788.

The probe used was nick-translated pHβG1.

In each pair of tracks Pad2 is on the left, NM788 on the right. Enzymes used were: a, HsuI; b, HsuI+EcoR1; c, EcoR1; d, EcoR1+BamHI; e, BamHI.

TABLE 4.1

HsuI	8.2kb *
HsuI+EcoR1	3.95kb *
	3.7kb *
EcoR1	5.8kb
	5.7kb *
	3.85kb *
EcoR1+BamHI	3.8kb *
	3.2kb
	2.03kb *
	1.7kb *
BamHI	8.8kb *
	2.1kb *

Table 4.1.

Sizes of the fragments of Pad2 derived wholly or partly from human DNA. These fragments are identified by comparison of the Pad2 tracks with the adjacent NM788 track in Figure 4.1. Those fragments which hybridise with pHβG1 are marked with an asterisk (*).

There is a systematic overestimation of fragment sizes when they are compared to the known sizes of globin gene fragments. This is probably due to "edge-effects" causing the marker tracks to migrate abnormally. The rudimentary map in Figure 4.2 can nonetheless be deduced from this data.

the restriction enzyme map of NM788, and to the structure of the normal β-globin gene (Lawn et al, 1978; D. Westaway, personal communication) allowed a rudimentary map to be assembled from the data (Figure 4.2)

It is interesting to note that the EcoR1+BamHI IVS-2-containing fragment is visualised after Southern blot hybridisation (albeit at a greatly reduced intensity) in Figure 4.1. The extent of sequence complementary to globin cDNA in this fragment is only 66bp. The limit of the technique has not been reached even at this low level of hybridisation.

The orientation of the cloned human DNA fragment in Pad2 is the reverse of the orientation of the fragment in Pad1. This confirms that Pad2 is a new isolate, and is not a reisolation of Pad1, which could conceivably have occurred by accidental contamination.

b) Subcloning of the β^{o}-thalasaemic Gene into pAT153.

Digestion of Pad2 with HsuI produces 3 fragments. Two of these are the arms of the vector which have a HsuI site at one end, and at the other a lambda phage cohesive terminus. The third fragment is the human DNA, and it is the only one flanked by two HsuI sites. Thus, only this fragment from a total HsuI digest of Pad2 will recombine in vitro with a plasmid vector.

The HsuI site in pAT153 is near the 5' end of the plasmid gene conferring resistance to tetracycline. Cloning into this HsuI site allows double selection to be carried out. The presence of the plasmid is detected by ampicillin resistance, and the presence of DNA inserted into the HsuI

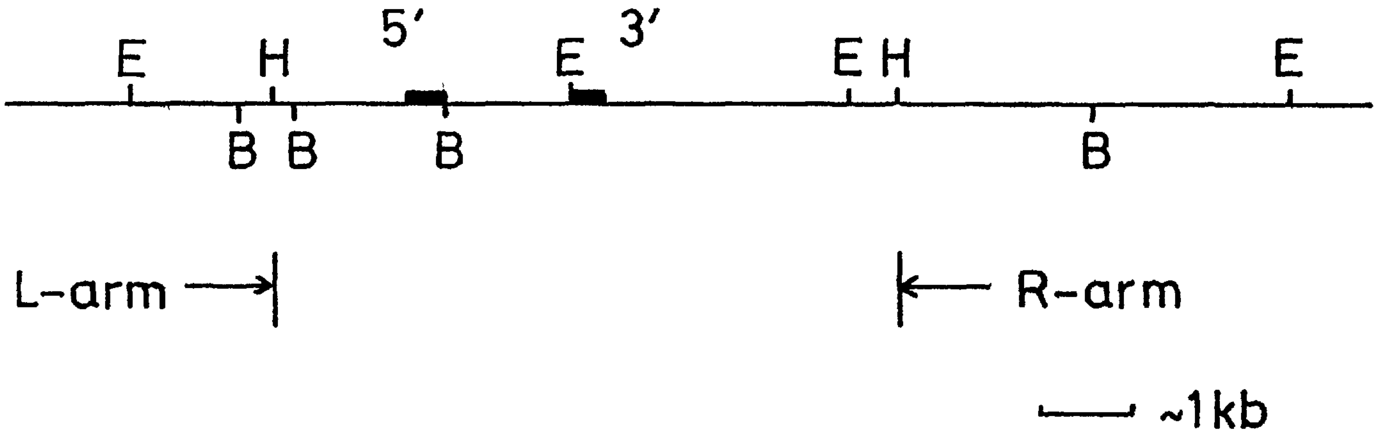


Figure 4.2.

Rudimentary restriction enzyme map of Pad2, derived from the data in Figure 4.1 and Table 4.1, with reference to the known maps of the β-globin gene and of NM788.

The positions of the EcoRl and BamHI sites in the vector nearest to the inserted human DNA are shown. The position of these sites indicates that the orientation of the gene with respect to the vector is as shown.

Only those sites detected by this analysis are shown. A more comprehensive map is in Figure 4.5.

B, BamHI; E, EcoRl; H, HsuI.

site by tetracycline sensitivity.

An HsuI digest of Pad2 (1μg) was mixed with 0.25μg of HsuI digested pAT153 and T4 ligase added. The mixture was incubated for 5 hours at room temperature before adding to transformation competent cells,as described in Materials and Methods. Approximately 840 ampicillin resistant colonies were obtained. This was 10% of the number obtained when the same amount of HsuI-cut plasmid DNA was ligated and transformed in the absence of Pad2 DNA. The difference may be due in part to the fact that only one third of the Pad2 fagments ligated to the plasmid will be capable of circularisation, and thus of completing the reaction, as only one third have two HsuI termini.

Forty-eight ampicillin resistant clones were streaked in duplicate onto L-plates containing ampicillin or tetracycline. All the clones grew on ampicillin. Two clones did not grow at all on tetracycline, and seven showed only very weak growth. Control colonies which were known to contain an inserted DNA fragment in the HsuI site also showed a small amount of growth on tetracycline. It appears that insertion into the HsuI site does not always completely eliminate tetracycline resistance. It is possible that some of the seven "leaky" clones contain a DNA fagment inserted into the HsuI site, although two of these which were chosen at random did not contain a large insert when examined on gel electrophoresis.

Two fully tetracycline sensitive clones and one fully resistant clone were picked, grown in 0.5ml of L-broth for two hours and amplified overnight in 170μg/ml chloramphenicol

The cells were harvested and treated with proteinase K. The cell lysates were loaded onto an agarose gel and the DNA was electrophoresed, using as markers, pAT153 and a subclone of the HsuI fragment containing the β^{+}-thalassaemic gene from Padl. Figure 4.3 is a photograph of the gel. It can be seen that both tetracycline sensitive clones contain inserted DNA of about the same size as the Hsu fragment from Padl.

One of these clones was selected and plasmid prepared from a one litre culture of the cells. It will be shown below that it contains the β-globin specific HsuI fragment, of about 7.5kb in length, and it was therefore designated β^{o}/7.5.

c) Restriction Enzyme Mapping of β^{o}/7.5.

The plasmid identified above was digested with a range of restriction enzymes to generate fragments which could be directly compared with the fragments produced by digestion of subclones of the normal β-globin gene isolated by Lawn et al (1978). These subclones are named R1-C and R1-D, and are, respectively, the 5' and 3' EcoR1 fragments of the gene. Fragments of about 2kb were regarded as sufficiently small to allow detection of deletions of the order of 100bp. Smaller fragments obviously allow detection of smaller deletions. In our experience, samples in adjacent, parallel tracks on gel electrophoresis do not always migrate identically. Therefore, in addition to running the digests of the two genes alone, they were also mixed and coelectrophoresed. This will allow the identification of any band-splitting, which is a much more sensitive method of detecting deletions.

The 5' end of the β^{o}-thalassaemic gene was examined

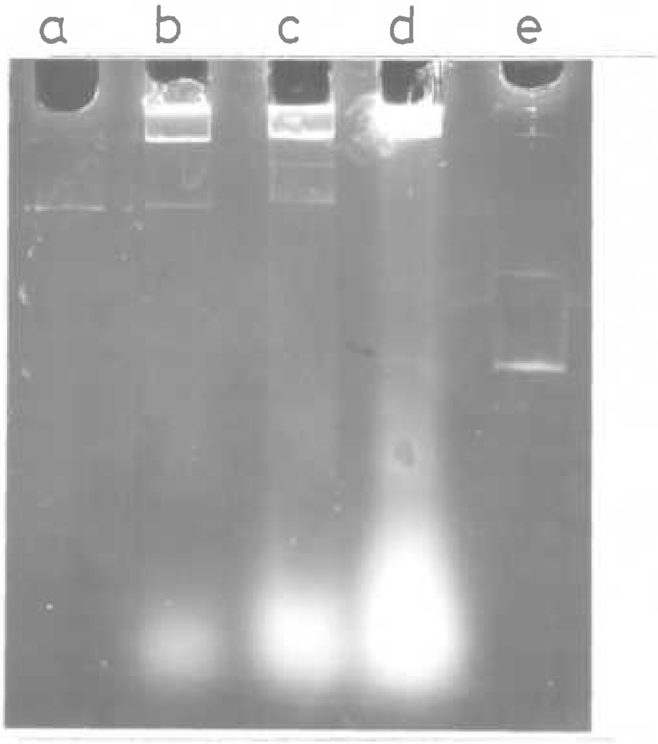


Figure 4.3.

"Colony electrophoresis" of DNA in clones isolated during subcloning of the HsuI fragment of Pad2.

a, 7.5/β^{+} DNA (derived from Pad1).

b,c, both tetracycline sensitive, ampicillin resistant colonies.

d, tetracycline and ampicillin resistant colony.

e, pAT153 DNA.

using BamHI+EcoR1 and BamHI+HpaI digests of both it and the normal β-gene. The fragments were separated by 1.5% agarose gel electrophoresis. The 3' ends of the genes were compared after EcoR1+PstI and EcoR1+BglII digests were electrophoresed on 1.5% agarose and 5% acrylamide gels. Figure 4.4 shows these gels. The fragments indicated are illustrated on the map of the normal β-globin gene in Figure 4.5. As can be seen, no deletion or insertion is detectable in the coding region, in the intervening sequences or up to 2.2kb upstream and 3.4kb downstream of the gene.

It is not possible to correlate the loss of function of the gene with any gross abnormality in its structure. In order to find any difference between this thalassaemic gene and the normal gene a more sensitive method must be used. The best method would be to determine the DNA sequence of the defective gene, and to compare it to the sequence of a normal gene. However, any difference found is not necessarily functionally significant, as it may be merely a polymorphic variant of the normal sequence. Furthermore, as the patient was thought to be doubly heterozygous for two different thalassaemic genes with different molecular phenotypes (section 3.4a) it was not known whether to look for a "transcriptional" or a "processing" defect in the gene. Therefore, it was thought that a an assay for the function of the cloned genes should be attempted, and the transcription of the thalassaemic gene studied in vitro.

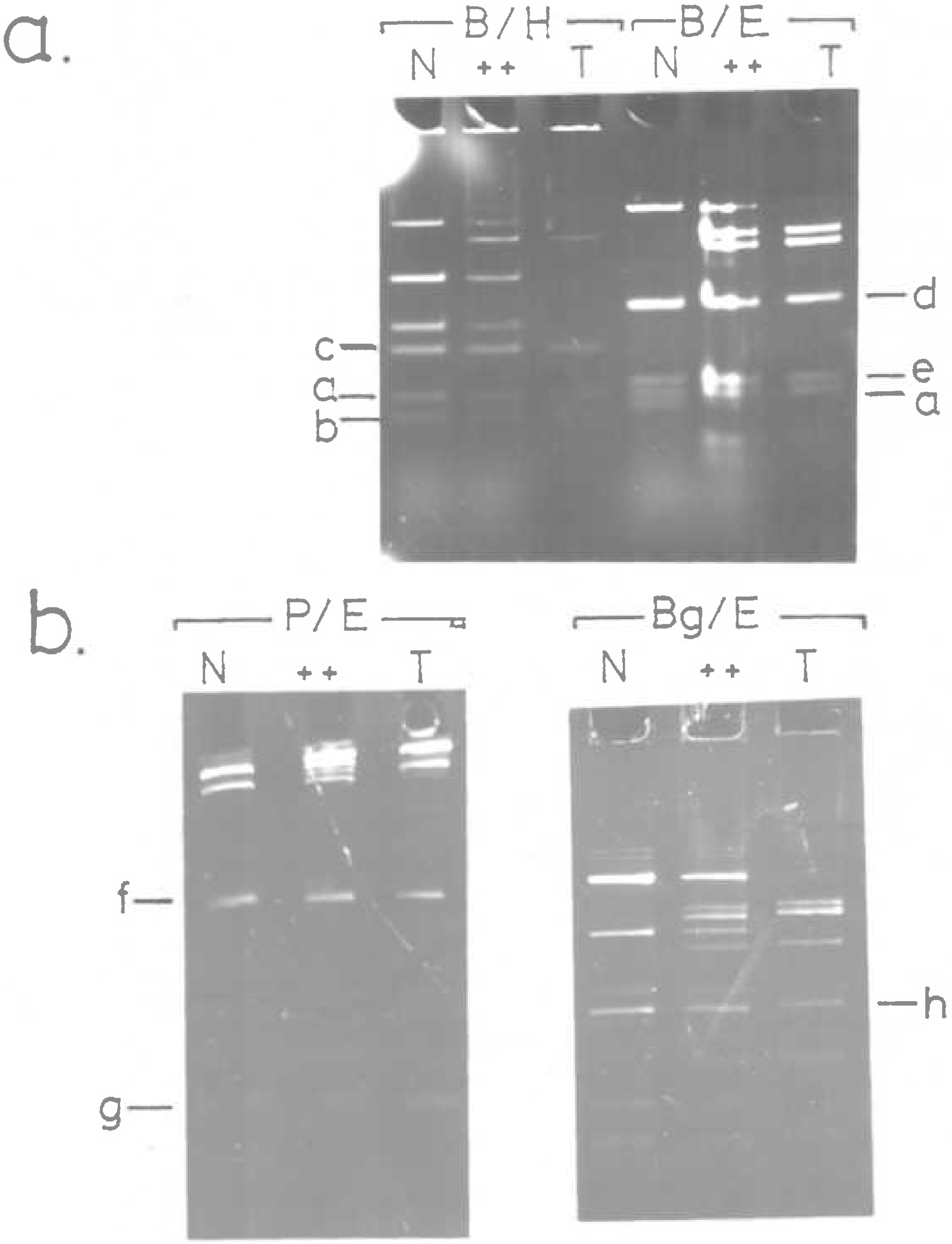


Figure 4.4.

Agarose gel electrophoresis of normal and β^0-thalassaemic β-globin genes.

a, analysis of the 5' end, B/H: BamHI+HpaI
B/E: BamHI+EcoR1.

b, analysis of the 3' end, P/E: PstI+EcoR1
Bg/E: BglII+EcoR1.

On each photograph, N is the normal gene, T, the thalassaemic gene and ++ is both genes mixed.

Bands labelled a-h are compared fragments, indicated in Figure 4.5.

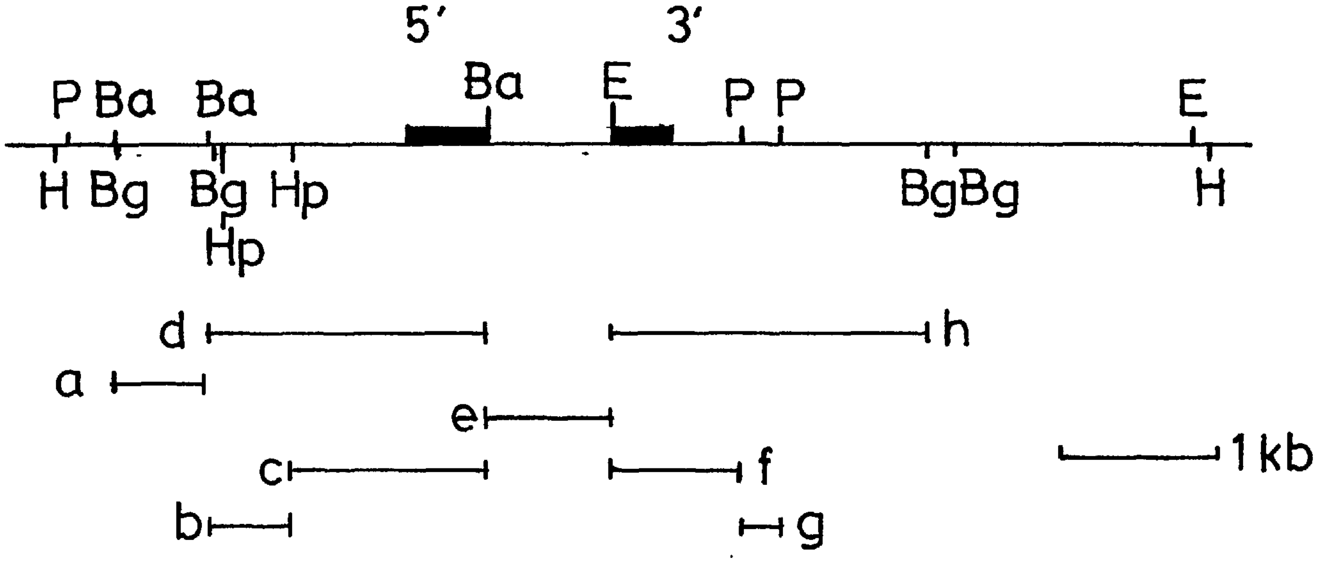

5′
3′
P
Ba
Ba
Ba
E
P
P
E
H
Bg
Bg
Hp
Hp
BgBg
H
d
h
a
e
c
f
1kb
b
g

Figure 4.5.

Restriction enzyme map of the HsuI fragment containing the β-globin gene. Data are from Lawn *et al* (1978) and D. Westaway (personal communication).

Bars marked a-h indicate the restriction fragments which are compared directly in Figure 4.4.

H, HsuI; Ba, BamHI; Bg, BglII; E, EcoRl; Hp, HpaI; P, PstI.

4.2 In Vitro Transcription of Normal and Thalassaemic Globin Genes.

Two methods for *in vitro* transcription of exogenous genes by RNA polymerase II have recently been described. The one described by Weil *et al* (1980) utilizes exogenously added calf thymus RNA polymerase II and a supernatant of KB cell cytoplasm. Manley *et al* (1980) use the endogenous RNA polymerase II in whole-cell extracts of HeLa cells to transcribe the genes. Neither method is able to splice or to terminate the transcripts. The templates used, therefore, are truncated; they are cut by a restriction enzyme at a site located within the gene at a known distance from the first base of the transcript.

Both the methods of Weil *et al* and Manley *et al* were used to investigate the transcription of the cloned bo-thalassaemic gene.

Extracts of HeLa and KB cells were prepared in this laboratory by the method of Manley *et al* (1980). These were used to transcribe $\beta^{0}/7.5$ and a cloned normal β-globin gene (Hβ1-S). The plasmids were digested with BamHI to produce a template for run-off transcription of about 475 nucleotides. The transcripts were labelled with α-^{32}P-GTP or UTP, electrophoresed on 5% acrylamide gels, and autoradiographed. Unfortunately, although transcripts of the size expected if correct initiation was occurring were visible, there was a large number of other transcription products. These presumably originated from the cloned human DNA, as transcription of the vector (pBR322 or pAT153) DNA only resulted in two major bands. The results obtained in this laboratory do not convincingly show transcription from normal or abnormal genes.

Much more convincing data have been obtained by Rhyaiza de Freund, working as a visitor in the laboratory of Dr.P. Chambon, using the β -thalassaemic gene isolated by me and described above. In the experiments shown here, subclones of the 1.9kb 5' BamHI fragment of both normal and the thalassaemic β-globin genes, in pAT153, were used as templates. Both were digested with AccI, which cleaves 283bp from the 5' end of the gene. The DNA, at a concentration of 10ug/ml was transcribed by calf thymus RNA polymerase II added to an S-100 supernatant of HeLa cell cytoplasm, as described by Weil et al (1979).

Figure 4.6 shows the products of this transcription electrophoresed on a 5% acrylamide, 8M urea sequencing gel. Both normal and thalassaemic genes give rise to a very strong transcript of 280 nucleotides. Both genes are apparently transcribed with equal efficiency.

As the system is well-characterised in Dr. Chambon's laboratory (see for example Wasylyk et al, 1980; Corden et al, 1980) we may be confident that it accurately transcribes exogenous eukaryotic genes. These results indicate that there is no difference in the in vitro transcriptional function of the β^o-thalassaemic gene, compared to a normal β-globin gene.

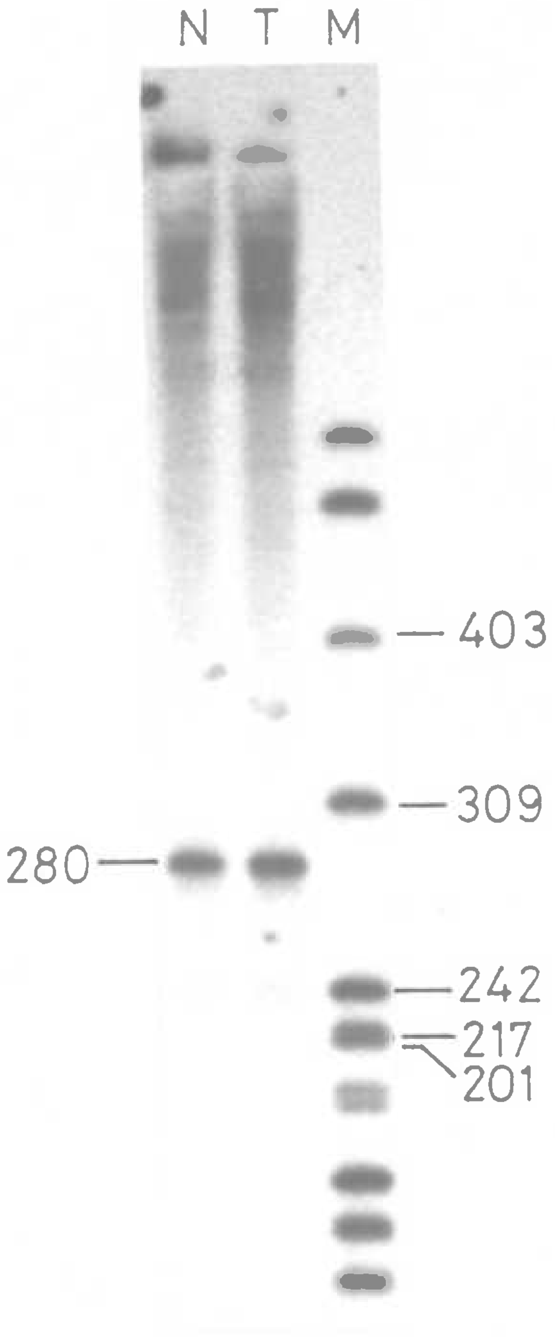


Figure 4.6.

In vitro transcription of normal and thalassaemic β-globin genes. The transcripts have been electrophoresed on a 5% acrylamide sequencing gel and visualised by autoradiography.

N: transcript of normal β-globin gene.

T: transcript of β°-thalassaemic β-globin gene.

M: markers; end-labelled MspI-digested pBR322.

Sizes are marked in nucleotides.

4.3 Discussion.

The absence of gross structural differences between the normal β-globin gene and the β^{o}-thalassaemic gene isolated here confirms and extends the findings of Flavell *et al* (1979) and the experiments in Section 3.4b. Other β^{o}-thalassaemic genes have been isolated and shown to be structurally identical with the normal gene, even at the DNA sequence level (R.A. Flavell, personal communication).

Interpretation of the *in vitro* transcriptional data presented here is problematic. The hypothesis put forward in Section 3.4 suggests that there are two phenotypically different thalassaemic genes carried by the patient studied. One of these would be transcribed but not processed, whilst the other is not transcribed. If the transcriptional systems described here accurately mimic *in vivo* transcription then the gene studied here would be classed as a transcribed (but presumably non-processed) gene.

However, it appears that transcription *in vitro* is more permissive than *in vivo*. Function of the gene *in vitro* depends only on the presence of the TATA box. It is clear that more than this is required *in vivo*. The β-globin gene in the patient with γ-δ-β-thalassaemia descibed by Van der Ploeg *et al* (1980) is non-functional, but is probably structurally normal for several thousand bases to the 5' side of the gene. Other β^{o}-thalassaemic genes have been isolated, and all show normal transcriptional activity *in vitro* (H. Dahl and R.A. Flavell, personal communication), although the transcriptional status of the genes *in vivo* has not been rigourously assessed.

The transcriptional assays carried out here indicate that the TATA box is almost certainly present in front of the gene. Assessment of whether or not this cloned gene is transcibed *in vivo*, and whether or not any transcript is spliced must await more sophisticated assays.

5. Structural Analysis of the Fibroblast Interferon Gene.

DNA was prepared from the recombinant phage designated 3.1 in section 3.5b. Based on the hybridisation data the clone was tentatively identified as containing a gene for fibroblast interferon (IFN-β). A map of restriction enzyme cleavage sites in this clone was constructed in order to confirm the tentative identification and to provide a basis for sequence and other analyses. The clone was redesignated λ.H.IFNβ/Pad3.

5.1 Construction of a Nearest-Site Map.

a) Deduction of the Map.

A series of digests of the DNA was performed, using various restriction enzymes, both singly and in pairs. The digested DNA was electrophoresed on a 1% agarose gel and photographed under UV-light before transfer to nitrocellulose. The filters were hybridised with nick-translated fibroblast interferon cDNA. Autoradiographs of the filters are shown in Figure 5.1. Bands marked with an arrow are products of incomplete digestion or of contaminating enzyme activity. The sizes of the hybridizing fragments were estimated by comparison with coelectrophoresed markers of known molecular weight, and are tabulated in Table 5.1

It should be noted initially that BglII and PstI both produce two interferon-specific bands of hybridisation. It is known from sequence studies of fibroblast interferon cDNA that the coding sequence of the gene contains both a BglII and a PstI site. No other enzyme used here cuts the known sequence of the probe, and none produce two bands (Figure

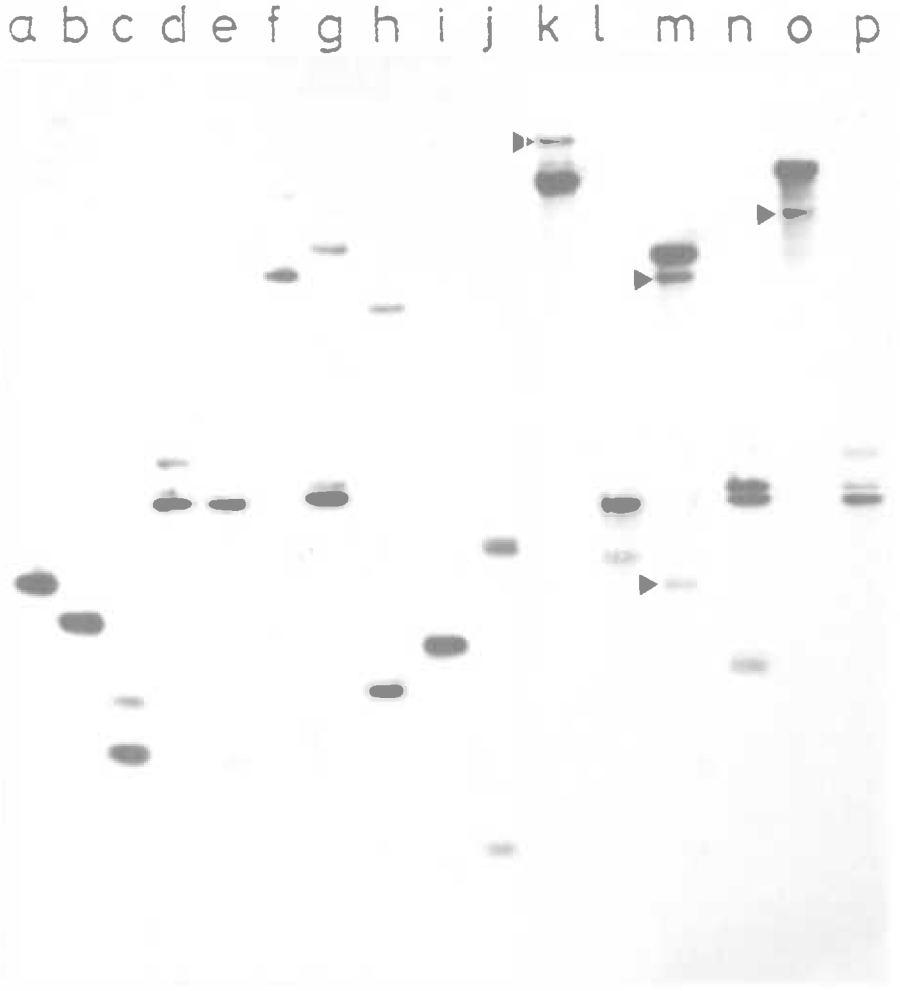


Figure 5.1.

Southern blot hybridisation to H.IFNβ/Pad3.

Probe used was an IFNβ cDNA cloned in pAT153.

a, EcoRl; b, EcoRl+HsuI; c, EcoRl+BglII; d, BglII; e, HsuI+BglII; f, HsuI; g, PstI; h, HsuI+PstI, i, EcoRl+PstI; j, BglII+PstI; k, XbaI; l, BglII+XbaI; m, SstI; n, BglII+SstI; o, KpnI; p, KpnI+BglII.

Bands marked ▶ are products of partial digestion or of contaminating enzyme activity.

TABLE 5.1

DIGEST	FRAGMENT SIZE
EcoR1	1.55
HsuI	6.8
BglII	2.25 + 2.6
PstI	7.8 + 2.35
KpnI	13.6
SstI	7.1
XbaI	12.8
EcoR1+HsuI	1.25
EcoR1+BglII	0.9 + 0.65
EcoR1+PstI	1.25 + 0.34
BglII+HsuI	2.2 + 0.66
BglII+PstI	0.38 + 1.9 (doublet)
BglII+KpnI	2.25 + 2.65
BglII+SstI	2.3 + 1.1
BglII+XbaI	2.25 + 1.8
HsuI+PstI	5.6 + 0.98

Table 5.1.

Sizes of fragments of λH.IFNβ/Pad3 hybridising to the IFN cDNA probe, seen in Figure 5.1. Sizes were calculated by comparison with restriction enzyme digests of λ-DNA and are given in kilobase pairs.

5.1). These data are consistent with the gene being that for fibroblast interferon.

The data in Table 5.1 permit a map to be drawn of the nearest site of cleavage on each side of the gene, for each enzyme. The EcoR1 fragment containing the gene is 1.55kb. A double digest of EcoR1+BglII produces two fragments of 0.66 and 0.9; thus the EcoR1 fragment contains a single BglII site, 0.9kb from one end and within the gene itself. As the HsuI+BglII digest produces a band of 0.66kb and EcoR1+HsuI gives a fragment of 1.25kb, then one HsuI site must be in the larger EcoR1+BglII fragment; 0.3kb from the EcoR1 site. The positioning of one HsuI site allows the other to be placed 6.8kb away, on the other side of the gene. As one of the two BglII fragments (2.2kb) is not cut by HsuI, then the site producing this fragment must be within the HsuI fragment. The other site must be 2.6kb to the other side of the intragenic site.

Both PstI fragments are reduced in size when an HsuI+PstI double digest is examined. Thus the larger PstI fragment must be colinear with the larger HsuI fragment, and similarly the smaller fragments are both on the same side of the gene. This can also be expressed as the cleavage of the HsuI fragment at the intragenic PstI site. Therefore, one PstI site is 2.2kb (7.8-5.6) outside the HsuI site; and the other sites lie one within the gene, and one on the other side, 2.35kb away. The postion of the PstI site with respect to the BglII site is accurately assessed by a BglII+PstI digest, in which there is a faint band of 0.38kb, corresponding to the distance between the sites within the gene.

The other two bands predicted from the map to be produced by this double digest are both approximately 1.9kb, and hybridisation to a doublet of about 2kb is in fact seen.

The smaller BglII fragment is not cut by SstI, but the larger is; which places one SstI site 1.1kb from the intragenic BglII site, and therefore the other must be 7.1kb from that. Likewise the XbaI site which cuts the smaller BglII fragment is 1.8kb from the intragenic site, and the second XbaI site is 14kb from this.

The map of restriction enzyme sites incorporating all this information is in Figure 5.2.

b) Orientation of the Map.

The map can be alligned with respect to the direction of transcription. The sequence data indicates that the PstI site is to the 5' side (in terms of the mRNA) of the BglII site. This orientation is therefore indicated in Figure 5.2. Additional evidence is provided by the fact that the bands produced by PstI, alone or in combination, which are 5' of that site are more intense than those to the 3' side; whilst the fragments 5' of the BglII site are more strongly hybridised than those 3' of it. The cloned cDNA used as a probe extends from the initiation codon of the protein to the 3' end of the mRNA. There is only 140bp of hybridisable material to the 5' side of the PstI site, and only 200bp to the 3' side of the BglII site, compared to around 500bp on the opposite sides of both sites. The difference in band intensity due to this difference in hybridisable sequence can be clearly seen in Figure 5.1.

c) Positioning the Map Within λH.IFNβ/Pad3.

The gene, as mapped above, was positioned with respect

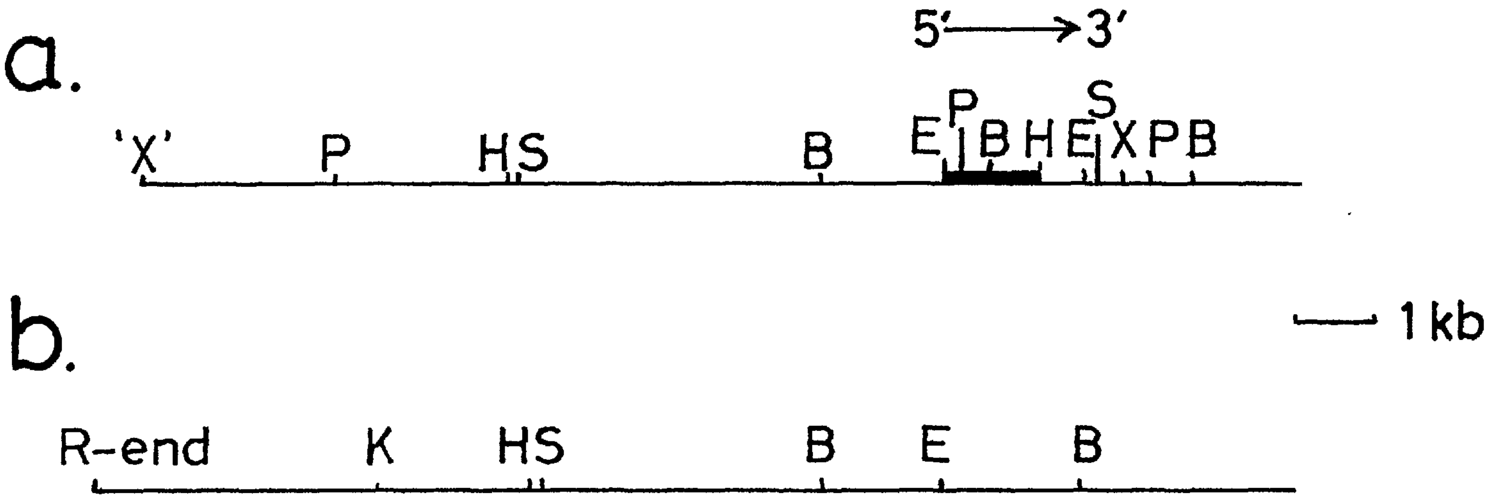

a.
5'→3'
'X'
P
HS
B
E
P
B
H
E
S
X
P
B
b.
1kb
R-end
K
HS
B
E
B

Figure 5.2.

Nearest site map of the restriction enzyme sites around the interferon gene in λH.IFNβ/Pad3, and of the "right arm" of Charon4A.

a, λH.IFNβ/Pad3

b, R-end of Charon 4A.

B, BglII; E, EcoRl; H, HsuI; K, KpnI; P, PstI; S, SstI; X, XbaI; 'X', apparent XbaI site.

to the restriction enzyme sites in the vector.

Data obtained using the Southern blot technique on total human DNA by M.Houghton at Searle Research (manuscript submitted) and by A.Brand at St. Mary's (unpublished) indicate that the fragments produced by EcoR1 and HsuI (or its isoschizomer HindIII) are somewhat larger than those produced from the recombinant. The possibility that all the sites mapped to the 5' side of the PstI site were derived from the vector was therefore examined.

Data provided by F.Blattner gives the position of all the sites mapped above (with the exception of PstI) in both the left and right arms of the Charon 4a vector. Comparison of the Figure 5.2 with the data shows that all the sites 5' of the gene are consistent, within a 3 to 7% error, with them being the so-called "right-arm" of the vector. There are no XbaI sites in this arm; the size of XbaI fragment in Table 5.1 is, within experimental error, the size of the fragment from the mapped site to the end of the phage chromosome.

The gene-containing EcoR1 fragment therefore is adjacent to the smaller of the two vector EcoR1 fragments.

5.2 Arrangement of the Cloned EcoR1 Fragments.

All the EcoR1 sites in the cloned human DNA were located by examining the fragments produced by EcoR1, SstI and KpnI; used singly and in combination. The fragments were separated on an agarose gel, a photograph of which is in Figure 5.3. Table 5.2 lists the sizes of the fragments visible, after excluding fragments due wholly to cleavage of vector DNA. The EcoR1 fragments are all wholly derived from

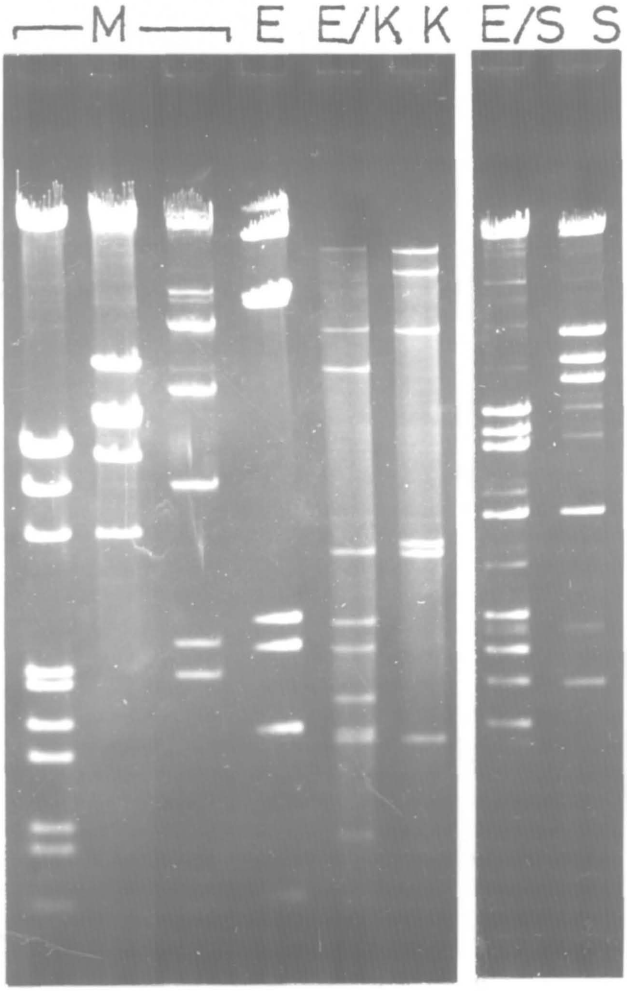


Figure 5.3.

Agarose gel electrophoresis of restriction enzyme digests of λH.IFNβ/Pad3.

E, EcoRl; E/K, EcoRl+KpnI; K, KpnI; E/S, EcoRl+SstI; S, SstI.

M: markers, λ-DNA digested with, left to right, HsuI+EcoRl; EcoRl; HsuI.

TABLE 5.2

DIGEST	FRAGMENTS
EcoRl	10.8 2.4 2.2 1.55 0.6
KpnI	13.6 9.2 3.3
EcoRl+KpnI	9.3 2.35 2.18 1.82 1.55 0.59
SstI	19 8.9 7.85 7.1
EcoRl+SstI	5.85 5.35 2.35 2.1 1.55

Table 5.2.

Fragments of λH.IFNβ/Pad3 wholly or partly containing human DNA, and visible on the agarose gel in Figure 5.3. Fragments of the vector alone were identified by comparison with data supplied by F. Blattner, and are not listed.

human DNA. The resolution in Figure 5.3 between the 10.8kb human fragment and the right-arm of the vector is not good; the gel was overloaded in order to visualise the smaller fragments. The size of the cloned segment of human DNA is therefore the sum of these fragments, which is 17.55kb. Several additional bands are seen in DNA digested by SstI. These are thought to be due to contamination of the enzyme preparation with SstII. The minor (submolar) bands were ignored in compiling Table 5.2.

KpnI generates three fragments wholly or partly derived from human DNA. The 13.6kb KpnI fragment contains the gene, and links the human DNA to the right-arm of the vector. From the known position of the KpnI site in the vector arm, the "nearest-site" within the human DNA can be positioned. The 9.2kb KpnI fragment is not cut by EcoR1 in a KpnI+EcoR1 double digest, and it must therefore be enclosed within the largest (10.8kb) EcoR1 fragment. The only other KpnI fragment (3.3kb) is cut by EcoR1 and must therefore extend from the vector into the human DNA. As it is known that this does not extend from the right-arm (which is adjacent to the coding segment of the gene) then the fragment can be placed between the known KpnI site in the left-arm to a site in the human DNA. The relative positions of all three Kpn fragments are now known. The distance between the outermost sites allows an estimate of the size of the cloned DNA segment; this is calculated as 17.85kb, which is in remarkable agreement with the estimate from the EcoR1 fragments. The position of fragments and sites within the human DNA are referred to in terms of the orientation of the interferon gene; 5' or 3' as appropriate

A preparative 1% agarose gel was used to electrophorese 10ug of a KpnI digest of the recombinant. The fragments of 13.6kb and 3.3kb were isolated, using the sodium perchlorate method (Section 2.3g), recut with EcoR1, and reelectrophoresed on a 1% agarose gel. Figure 5.4 is a photograph of the gel, stained with ethidium bromide and illuminated with UV-light. EcoR1, KpnI and EcoR1+KpnI digests of the whole recombinant were loaded in adjacent tracks. The 3' (3.3kb) KpnI fragment contains the 2.4kb EcoR1 fragment (plus a vector-derived 0.9 kb fragment). The 5' (13.6kb) KpnI fragment contains a 7.2kb vector fragment and the 2.2 and 1.55kb human fragments. Although not clearly seen, it is assumed that the 0.6kb fragment is also in this track. Subsequent data confirm that this must be the case.

Closer examination of the KpnI+EcoR1 fragments (Figure 5.4) suggest that the 2.2kb EcoR1 fragment is slightly reduced in size after KpnI digestion. Thus the KpnI site is slightly to the 3' side of the penultimate EcoR1 site. As EcoR1 does not cut the 9.2kb KpnI fragment, then the 10.8 kb EcoR1 fragment must extend from the site just mapped.

All the EcoR1 fragments can thus be positioned using the Kpn digests, with the exception of the 2.4 and the 0.6kb fragments. Digestion with SstI orders these. The "nearest-site" map places an SstI site a few hundred bases to the 3' side of the 1.55kb fragment. Agarose gel electrophoresis suggests that it is not the 2.4 kb fragment which is cut by SstI. Confirmation that it is, indeed, the 0.6kb fragment which contains the SstI site is obtained by electrophoresing

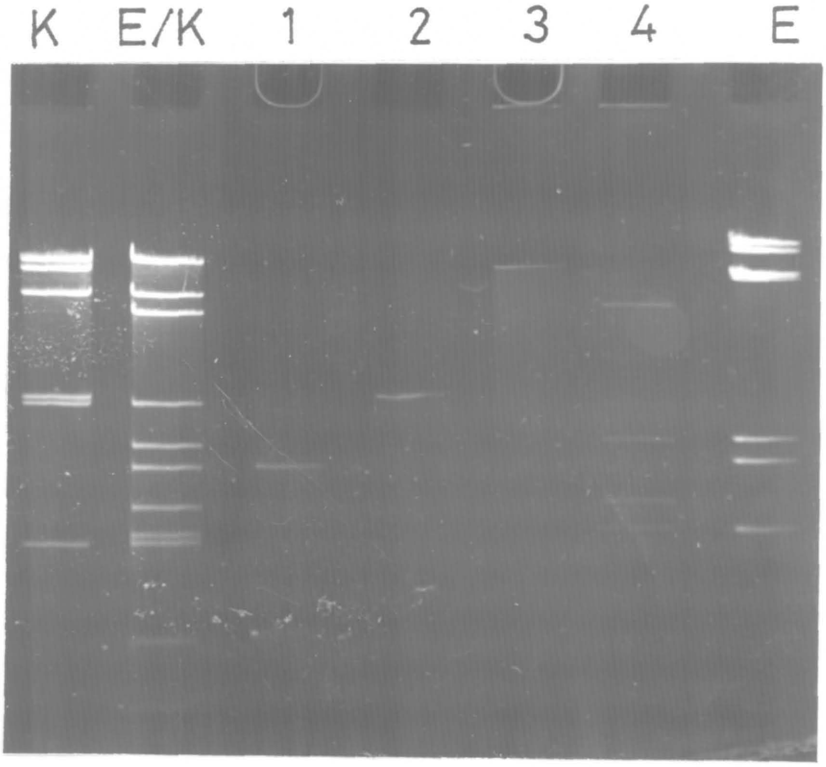


Figure 5.4.

Agarose gel electrophoresis of isolated KpnI fragments of λH.IFNβ/Pad3.

K, KpnI; E/K, EcoR1+KpnI; E, EcoR1.

1, isolated 3.3kb KpnI fragment, recut with EcoR1.

2, isoalted 3.3kb KpnI fragment.

3, isolated 13.6kb KpnI fragment.

4, isolated 13.6kb KpnI fragment, recut with EcoR1.

an EcoR1 and an SstI+EcoR1 digest on a 5% acrylamide gel, which is shown in Figure 5.5a. The 600bp EcoR1 fragment is cut by SstI into two fragments of about 400 and 200bp in length. All the EcoR1 fragments are located within the map, which is shown in Figure 5.6.

The remaining SstI sites are located in the map as follows. The largest fragment of the SstI digest derives from the left-arm of the vector plus a small amount of human DNA (there are no SstI sites in the left arm). The 2.2kb EcoR1 fragment is reduced very slightly in size by SstI (Figure 5.5b), and a 100bp fragment is seen on the acrylamide gel in Figure 5.5b. Therefore the 3' SstI site is mapped about 0.1kb inside the 5' end of the 2.2kb EcoR1 fragment. The other site produces two fragments in an EcoR1+SstI double digest, of 5.85 and 5.35kb, from the largest EcoR1 fragment. With a little ambiguity, an SstI site can be placed near the centre of this fragment.

5.3 Structure of the Fibroblast Interferon Gene.

In interpreting the structure of the fibroblast interferon gene, it is important to know whether the whole gene is contained within the recombinant described here. In other words; how much DNA is there to the 5' side of the restriction enzyme sites known to be in the coding sequence of the gene?

The size of the BglII+EcoR1 fragment which contains the 5' DNA is estimated to be 650bp from the agarose gel electrophoresis shown in Figure 5.1. No better estimate of this size

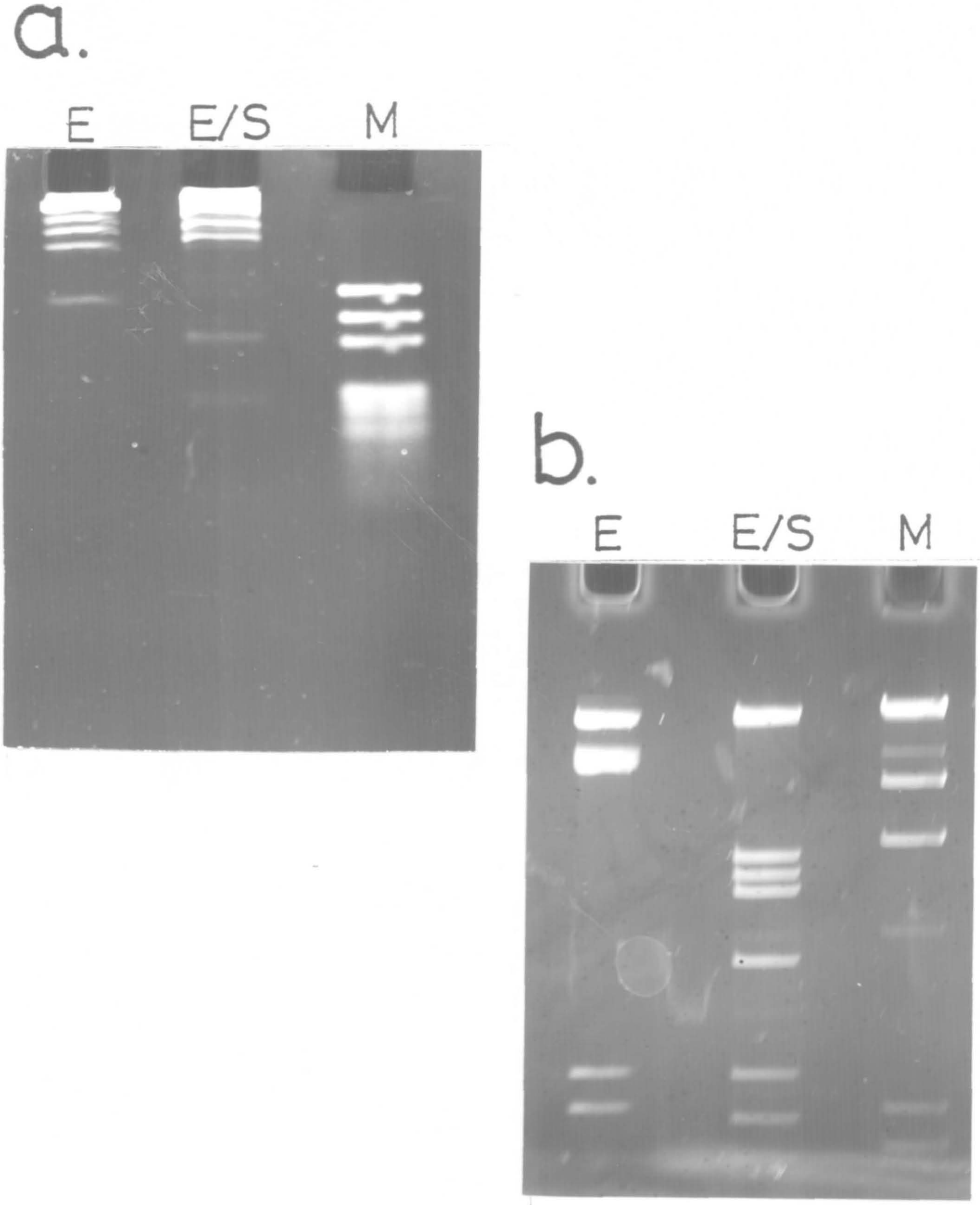


Figure 5.5.

Gel electrophoresis of restriction enzyme digests of H.IFNβ/Pad3.

a, acrylamide gel; E, EcoRl; E/S, EcoRl+SstI; M, MspI-digested pBR322.

b, agarose gel; E, EcoRl; E/S, EcoRl+SstI; M, λ-DNA digested with HsuI.

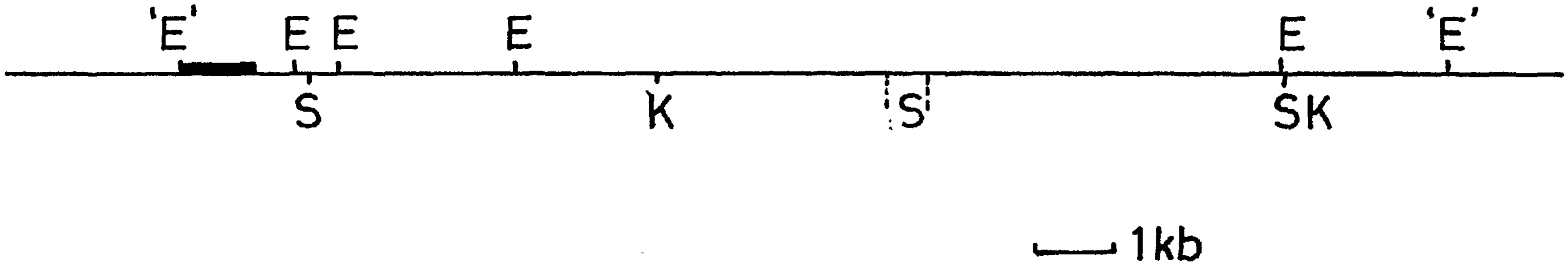


Figure 5.6.

Restriction site map of λH.IFNβ/Pad3.

E, EcoRl; K, KpnI; S, SstI; 'E', synthetic EcoRl sites added during manufacture of recombinants.

DNA which hybridises to IFNβ-cDNA is marked in a heavier line.

was obtained when acrylamide gel electrophoresis was performed on the isolated BglII fragment recut with EcoR1. As the position of the BglII site, determined by sequencing, is 640bp from the 5' terminus of the mRNA, the EcoR1 site must lie either in an intervening sequence or very close to the 5' end of the gene.

The distance beween the BglII and the PstI sites as shown in these experiments is about 380bp. In the mRNA sequence determined by Houghton et al (1980) the sites are 360 bp apart. Thus, in this region of the gene, at least, ther do not seem to be any intervening sequences.

Confirmation of both these surmises has been obtained by work at Searle Research (M. Houghton, I.J. Jackson et al, manuscript submitted). In this work the sequence of the IFNβ sequence in λH.IFNβ/Pad3 was determined. At the 5' end of the gene are 22 base pairs between the start of the mRNA and a sequence of 12bp which corresponds to the EcoR1 linker used by Maniatis et al during construction of the gene library. The EcoR1 linker is attached to the sequence of a cleaved HaeIII site, at which the DNA was digest prior to cloning.

Sequence analysis of the whole of the coding region of the gene indicates that not only are there no intervening sequences between the BglII and PstI sites, but there are none in the whole region of the gene which gives rise to the mRNA.

5.4 Discussion.

The data here show that the recombinant isolated (λH.IFNβ/Pad3) probably contains the whole of the coding region of the gene with very little DNA to the 5' side but about 17kb to the 3' side. Thus the cloned gene is likely to be missing the putative transcriptional control elements which have been described to the 5' side of most eukaryotic genes. Because of the reasons detailed in 3.5c, clones which have these sequences will have to be isolated from another stock of a prepared library, or isolated from a "tailor-made" collection of recombinants.

No cross-hybridising sequences are detected in the 17kb of DNA to the 3' side of the gene. As the stringency of washing given to the filters after hybridisation is rather low (3xSSC at 65°) it can be said that there are no genes or pseudogenes which could be regarded as belonging to the same family of genes as the fibroblast interferon gene in the region of the genome examined. The failure to detect any other cross-hybridising recombinants during the screening of the recombinants is not good evidence for the absence of other genes for the reasons outlined in 3.5c. However, M. Houghton _et al_ (manuscript submitted) detected no other cross-hybridising genes at a stringency as low as 5xSSC at 65° This observation is a notable contrast to the work of Nagata _et al_ (1980b) on the leukocyte interferon gene. These workers were able to isolate 8 different, but closely homologous, genes from a human library, and concluded that the leukocyte interferon genes form an extensive family, as do several mammalian genes characterised to date.

The work described here, in conjunction with the the results from Searle Research, indicate an absence of intervening sequence in the gene. This is in accord with the data of Nagata _et al_, who find that, in the leukocyte interferon gene which they examine, there are no intervening sequences. As the two genes are clearly evolutionarily related (Tanaguchi et al, 1980) perhaps this is not too surprising.

The interferon genes are the first genes other than histones to be shown not to require splicing of the primary transcript. They are the first polyadenylated mRNAs for which this has been demonstrated. This may cast light on the function (if any) of intervening sequences or of the poly(A) tails of mRNA.

The only known non-polyadenylated mRNAs code for histones, which are known not to have intervening sequences, at least in sea-urchin (Schaffner et al, 1978). This observation, coupled with the fact that polyadenylation probably occurs before splicing (Abelson, 1979) led to a suggestion that the function of the poly(A) tail was to act as a guide for the splicing mechanism. The presence of poly(A) tails and the absence of intervening sequences from interferon genes is an indication that this is not necessarily the case; although it might be argued that the poly(A) tail is a vestigial structure, and the genes have lost their intervening sequences.

Another proposed function for intervening sequences is that they disrupt the continuity of homology between related genes, and therefore reduce the frequency of gene loss by unequal crossover (Konkel et al, 1979). This would certainly be a rationalisation for the absence of intervening sequences from the fibroblast interferon gene, as it apparently has no relatives with which to cross-over. It may only be applicable to the leukocyte genes if the members of the family are not closely linked on the chromosomes, and Nagata _et al_ give evidence that at least some of them are.

The suggestion that intervening sequences play a role in promoting the evolution of genes (Gilbert, 1978) is not reflected on by the interferon gene results. The genes may simply have "stepped out" of the evolutionary system, or they perhaps may be regarded as a basic evolutionary unit, with only a single function.

What must be true for the interferon genes is that there is at least one less stage between transcription of the gene and translation of the mRNA. This may mean that the expression of the genes is more rapid than it may otherwise have been. This, of course, would be important to a cell which requires a rapid response to a viral infection. The histone genes similarly seem to be designed to meet a sudden increase in demand for their products during chromosome replication. The histone genes go somewhat further in this role, as they are in multiple copies (30-40 in humans; Melli _et al_, 1977). Neither the leukocyte nor fibroblast interferon genes are repeated to the same degree.

6. Conclusions.

In this final section I will briefly go over what I feel are the significant points described and discussed in the preceeding chapters and I will end by indicating one direction in which future experiments may proceed.

6.1 Gene Structure and IVS Function.

The finding that the interferon genes have no intervening sequences is a significant one. That both the leukocyte and fibroblast interferon genes have no IVSs indicates that these genes have been without them for more than 500 million years. It is likely that they have never had them. These genes cannot be assigned to a "specialised" class, as the histone genes have been. There was no *a priori* reason to expect that the genes should differ from what has become the norm in gene structure.

Most generalisations made about the presence of IVSs have not proved to be valid. Early speculation that intervening sequences were only present in the highly specialised or "superabundant" class of genes was incorrect. The gene for mouse dihydrofolate reductase, which is a "housekeeping" enzyme, has at least five IVSs.

It has been suggested that IVSs serve to interupt related genes with regions of non-homology to minimise recombination between them (Konkel *et al*, 1979), but several examples are known of genes which do not have relatives but do have IVSs. The leukocyte interferon genes are now an example of genes which are a family, some of which are physically linked, which do not have intervening sequences.

It appears that genes which have intervening sequences always have them, even in the most evolutionarily distant orgaisms. There are cases known where genes and pseudogenes have lost IVSs (Nishioka and Leder, 1980; Vanin et al, 1980; Lomedico et al, 1980), but none known where they appear to have gained them. Intervening sequences should thus not be regarded as independent entities, but rather as the space between specific coding regions. Crick (1979) pointed out that once a cell has developed a mechanism for removing intervening sequences from an RNA transcript, the evolutionary potential of the genome is vastly increased. That evolution can occur by shuffling of genetic domains (which may or may not be equivalent to protein domains) is an attractive proposal. There is then no need for the IVSs to be removed from the genome, and with no strong selecion pressure in its favour, the precision required to remove them without affecting gene function probably will allow remval only very rarely.

To ask, in general terms, what is the "function" of intervening sequences misses the point, as originally they were unused pieces of DNA between cotranscribed coding domains. This is not to say, however, that they still serve no function. At least 1,000 million years of evolution have no doubt exploited many of the possible functional uses which IVSs might acquire.

Therefore, those genes which have IVss which have begun to diverge, will recombine less, which will accelerate their divergence. Those genes for which maintenance of sequence homogeneity is acceptable, or even desirable, such as the human $^{G}\gamma$ and $^{A}\gamma$-globin genes will have very closely related IVSs. The immunoglobulin genes utilise the genetic domains to shuffle the protein domains. The shuffling is used to generate diversity by rearranging coding units (Sakano et al, 1979; Weigert et al, 1980; Early et al, 1980), to allow developmental switching of protein domains (Coleclough et al, 1980), and to activate one allele to the exclusion of the other (Joho and Weissman, 1980). Differential splicing of the same pre-mRNA produces two mRNAs coding for immunoglobulins with identical N-termini but different C-termini (Liu et al, 1980). In other genes, the requirement that the primary transcript be spliced adds a further step between gene and protein at which control can be, and no doubt in some cases is, exercised.

These functions, however, should be viewed as bonuses. The various functions of the intervening sequences are not the reasons for their original presence, and possibly not even for their maintenance. Those genes which were not, as it were, "born" with intervening sequences must get by without them, as the leukocyte and fibroblast interferon genes show they can.

6.2 In Vitro Analysis of Gene Expression.

That the β^{o}-thalassaemic and normal β-globin genes are transcribed with equal efficiency in vitro is not too surprising, in view of the known transcriptional state of at least one of the genes in vivo. Further studies on this patient should include a better analysis of the RNA molecules in red cell precursors, isolation of more clones containing the β-globin gene which will perhaps allow the identification of both alleles, and the insertion of the gene isolated here into fibroblasts to compare its splicing pathways with normal β-globin splicing.

It is likely that many thalassaemic globin genes which are not transcribed in vivo will be found to be efficiently transcribed in vitro. Furthermore, the sequence of at least one β^{o}-thalassaemic gene is the same as the normal gene for several hundreds of base pairs upstream of the site of initiation of transcription (R.A. Flavell, personal communication).

The drawback of the methods of in vitro analysis of gene function is that they work without many of the cellular control mechanisms which operate in vivo. An extract of fibroblast cells should not be expected to transcribe normal globin genes at all, unless the normal controls of gene expression are somehow circumvented. Defective genes, which are not normally transcribed, function efficiently in these systems. The methods of both Weil et al (1979) and Manley et al (1980) were originally developed using the adenovirus late gene as a model template. This is a gene which is transcribed in tissue culture fibroblasts under normal cicumstances, and it is transcribed more efficiently than any other gene examined to date (B. Wasylyk, personal communication). However, although it is a "high-output" gene in vivo, it is not normally expressed until after early gene expression and

DNA replication has occurred. The adenovirus early genes are transcribed very inefficiently *in vitro*.

The relative transcriptional efficiency of chicken genes *in vitro* does not reflect their transcription *in vivo*. The conalbumin gene is a better *in vitro* template than the ovalbumin gene, but it is transcribed at a lower level in hormone-stimulated tissue (Wasylyk *et al*, 1980). The β-globin genes transcribed here direct about the same amount of RNA synthesis *in vitro* as the ovalbumin gene (R. de Freund, personal communication).

The absence of transcriptional termination when using these two methods is possibly due to the absence or malfunction of one or a few proteins. This may prove to be readily tractable. Splicing of the primary transcript, on the other hand, may be a more complex function which could require that the RNA, for example, is packaged into RNP particles to form a specific 3-D structure on which the splicing mechanisms can operate. Producing a good *in vitro* pre-mRNA splicing system, and dissection of its mechanism, may be difficult.

Even the initiation process does not, it seems, mimic the initiation seen when a gene is introduced into a fibroblast cell. The results of Dierks *et al* (1980) indicate that the presence of the TATA box, plus 50bp upstream, will direct a low proportion of correctly initiated transcripts, but that rather more DNA upstream of this is required to increase this proportion. The requirement for more DNA 5' of the gene *in vivo* than *in vitro* is, in this case, not a developmental requirement (in the sense that the deletions of the $\gamma/\delta/\beta$-globin gene region indicate) as the gene is a globin gene in a fibroblast cell line.

Recent work on the differential control of 5S RNA synthesis in Xenopus oocytes and somatic cells has nevertheless indicated that in vitro methods may allow some aspects of developmental biology to be studied. Extracts of oocyte nuclei will transcribe 5S genes, whereas extracts of the nucleus of the mature egg will not. The egg extract will, however, transcribe tRNA genes. The failure of the egg extract to use 5S DNA as a template is due solely to the absence of a single protein factor, which has been purified and shown to bind to the 5S genes (Engelke et al, 1980). The factor, therefore, has the potential for controlling 5S gene expression independently of tRNA synthesis.

Pelham and Brown (1980) have shown that the same factor also binds to 5S RNA. A feedback control mechanism can be postulated whereby a fixed concentration of the protein will direct the synthesis of 5S RNA to allow it to accumulate to a level at which the concentration of RNA molecules is sufficient to bind out all the protein. If the 5S RNA is removed into ribosomes, the protein will bring up the level of RNA again to the steady state. In order to change the steady state level of 5S RNA one must postulate that the level of the active protein is changed.

If this factor, or any other factor, is found to be involved in the regulation of 5S synthesis during oogenesis or embryogenesis, the problem of control is not solved, but is merely deferred to a different gene. The principle whereby this problem is never solved, merely transferred to another focus has been recognised since Roman times; "sed quis custodiet ipsos custodes" (Juvenal; Satires 3, 347).

A study of cell biology as well as of molecular biology will be required if we are to break into the chain and to understand the basis of developmental biology. This is not how a globin gene is turned on, for that is in the nature of a bone marrow cell, but is what determines that a cell will become part of the bone marrow in the first place.

6.3 Future Problems.

I have discussed, both above and in Section 1, the various means by which isolated gene function can be studied. In this final section I hope to show that these methods, although undeniably powerful, are nonetheless limited in scope. Manipulating DNA *in vitro* or by using tissue culture cells can provide information about the way single genes or families of genes are regulated. However, understanding the way that an organism develops during embryogenesis is a prime aim of biology, and understanding the genetics of the process is a major part of this understanding. It is doubtful whether the strictly reductionist approach of the new genetics will show how genes work in development.

Differentiation of cells, it is true, can be accounted for by differential gene expression. However, all cells have, in addition to their rigid genomic programme, a "memory" of their history manifested by their protein complement and by contact with neighbouring cells. Activation of a particular gene at a particular time may only produce the required developmental effect if it occurs in the correct historical context. It seems likely that taking apart cells in a quest for the way in which genes work will destroy this context.

The reductionist approach is sometimes defended as the logical consequence of a determinist philosophy. However, whilst it can be argued that the whole is no more than the sum of its parts, each part is so interrelated with the others that a full understanding of any one part is not possible in the absence of the interrelationships.

For example, although it is known that Lesch-Nyhan's syndrome is associated with HGPRT deficiency, a study of HGPRT-deficient tissue culture cells will not indicate how the deficiency results in the behavioural patterns of the syndrome. A more general case is that of chromosome complement. Most tissue culture cells are aneuploid, but grow quite normally. On the other hand, most embryos with an abnormal karyote are not viable, or have some phenotypic abnormality at birth. It seems inconceivable that experiments solely rooted in reductionism will indicate how these defects arise.

What other approaches can then be used? Methods are available by which the genotype of a cell can be altered. These include DNA transformation (Wigler et al, 1979), the use of eukaryotic viral vectors (Goff and Berg, 1979) and microinjection (Mertz and Gurdon, 1977). If these cells can be introduced into whole animals, then the interaction of the new genotype with other normal or other altered cells can be examined.

Mouse embryos have proven to be remarkably manipulable.

Mosaic tertaparental and hexaparental mice have been created by mixing the cells from two or three 4 or 8 cell embryos (see Markert and Petters, 1978). Haematopoeitic cells from the livers of normal mouse foetuses have been introduced into the placenta of potentially anaemic foetuses, and these cells become incorporated into the bone marrow of the adult mouse (Fleschman and Mintz, 1979). These experiments have presaged an important breakthrough in molecular genetics, by Ruddle and his coworkers.

These workers have reported the introduction by microinjection of the thymidine kinase gene from herpes simplex virus into a single fertilised egg. They have reimplanted this egg into the uterus of a surrogate mother, where it developed to term. After birth, the mouse was found to contain the TK gene incorporated into its genome. They have not yet examined the expression of the gene. This method allows the examination of the expression and interaction of at least small groups of genes.

If at some point it becomes possible to introduce tissue culture cells into whole animals, whether by mosaic formation or nuclear transplantation, then the whole reassembly process from gene to animal will be available in a way which can be assayed, if not at this time controlled.

When we are able not only to alter the structure of genes _in vitro_, but also to change the structure of chromosomes _in vivo_, and be able to direct the outcome, we will finally deserve to call ourselves "genetic engineers".

......Yes, it sure has been a long, hard climb,
Train wheels running through the back of my memory,
When I ran on a hilltop following a pack of wild geese.
Someday, everything is going to be smooth, like a rhapsody,
When I paint my masterpiece.

Bob Dylan, 1971.

REFERENCES.

Abelson, J. (1979). Ann. Revs. Biochem. 48, 1035-1065.

Alestrom, P., Akusjarvi, G., Perricaudet, M., Mathews, M.B., Klessig, D.F., and Pettersson, U. (1980). Cell. 19, 671-681.

Allen, G. and Fantes, K.H. (1980). Nature, 287, 408-411.

Baglioni, C. (1962). Proc. Nat. Acad. Sci. U.S.A. 48, 1880-1885.

Bell, G.I., Pichet, R. and Rutter, W.J. (1980). Nucl. Acids Res. 8, 4091-4109.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980). Nucl. Acids Res. 8, 127-142.

Benton, W.D. and Davis, R.W. (1977). Science 196, 180-182.

Benz, E.J., Forget, B.G., Hillman, D.G., Cohen-Solal, M., Pritchard, J., Cavellesco, C., Prensky, W. and Housman D. (1978). Cell 14, 199-312.

Bernards, R., Little, P.F.R., Annison, G., Williamson, R., and Flavell, R.A. (1979a). Proc. Nat. Acad. Sci. U.S.A. 76, 4827-4831.

Bernards, R., Kooter, J.M., and Flavell, R.A. (1979b). Gene 6, 265-280.

Birkenmeier, E.H., Brown, D.D., and Jordan, E. (1978). Cell 15, 1077-1086.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thomson, K., Faber, H.E., Furlong, L-A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., and Smithies, O. (1977). Science 197, 161-169.

Blattner, F.R., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Richards, J.E., Slightom, J.L., Tucker, P.W., and Smithies, O. (1978). Science 202, 1279-1283.

Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980). Cell 19, 27-35.

Bos, J.L., Heyting, C., Borst, P., Arnberg, A.C. and Van Bruggen, E.F.J. (1978). Nature 275, 336-337.

Boyer, H.W. and Roulland-Dussoix, D. (1969). J. Mol. Biol. 41, 459-472.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 4853-4857.

Breathnach, R., Mantei, N. and Chambon, P. (1980). Proc Nat. Acad. Sci. U.S.A. 77, 740-744.

Brack, C. and Tonegawa, S. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5652-5656.

Cameron, J.R., Philippsen, P. and Davis, R.W. (1977). Nucl. Acids Res. 4, 1429-1448.

Cantell, K. (1979). Interferon 1, ed. Gresser, I., Pp 2-28. (Academic Press, London).

Catterall, J.F., O'Malley, B.W., Robertson, M.A., Staden, R., Tanaka, Y. and Brownlee, G.G. (1978). Nature 275, 510-513.

Chang, J.C. and Kan Y.W. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 2886-2889.

Chang, J.C., Temple, G.F., Trecartin, R.F. and Kan, Y.W. (1979). Nature 281, 602.

Chen, C.W. and Thomas C.A. (1980). Anal. Biochem. 101, 339-341.

Clarke, L. and Carbon, J. (1976). Cell 9, 91-99.

Clarkson, S.G., Smith, H.O., Schaffner, W., Gross, K.W. and Birnstiel, M.L. (1976). Nucl. Acids Res. 3, 2617-2632.

Coggins, L.W., Grindlay, G.J., Vass, J.K., Slater, A.A., Montague, P., Stinson, M.A. and Paul, J. (1980). Nucl. Acids Res. 8, 3319-3333.

Coleclough, C., Cooper, D. and Perry, R.P. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 1422-1426.

Collins, J. and Hohn, B. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 4242-4246.

Comi, P., Giglioni, B., Barbarano, L., Ottolenghi, S., Williamson, R., Novakova, M. and Masera, G. (1977). Eur. J. Biochem. 79, 617-622.

Conconi, F. and Del Senno, L. (1974). Ann. N.Y. Acad. Sci. 232, 54.

Cooke, H.J. and Hindley, J. (1979). Nucl. Acids Res. 6, 3177-3197.

Cordell, B., Bell, G., Tischer, E., Denoto, F.M., Ullrich, A., Pictet, R., Rutter, W.J. and Goodman, H.M. (1979). Cell 18 , 533-543.

Corden, J. Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980). Science 209, 1406-1414.

Craig, E.A., McCarthy, B.J. and Wadsworth, S.C. (1979). Cell 16, 575-588.

Crick, F. (1979) Science 204, 264-271.

Curtiss, R., Pereira, A., Hsu, J., Hull, C., Bark, E., Maturin, J., Goldschmidt, R., Moody, R., Inoue, M. and Alexander, L. (1976). Annual Report of the Dept. of Microbiology, Univ. of Alabama in Birmingham, 96-106.

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

Denhardt, D.T. (1966). Biochem. Biophys. Res. Comm. 23, 641-646.

Derynck, R., Content, J., DeClerq, E., Volckaert, G., Tavernier, J., Devos, R. and Fiers, W. (1980a). Nature 285, 542-547.

Derynck, R., Remaut, E., Saman, E., Stanssens, P., De Clerq, E., Content, J. and Fiers, W. (1980b). Nature 287, 193-197.

Dierks, P., Van Ooyen, A., Meyer, F., Mantei, N. and Weissman, C. (1980). Experientia 36, 742.

Dodgson, J.B., Strommer, J. and Engel, J.D. (1979). Cell 17, 879-887.

Dozy, A.M., Kan, Y.W., Embury, S.H., Mentzer, W.C. Wang, W.C., Lubin, B., Davis, J.R. and Koenig, H.M. (1979). Nature 280, 605-607.

Dugiczyk, A., Woo, S.L.C., Lai, E.C., Mace, M.L., McReynolds, L. and O'Malley, B. (1978) Nature 274, 328-333.

Duncan, C., Biro, P.A., Chowdary, P.V., Elder, J.T., Wang, R.R.C., Forget, B.G., De Riel, J.K. and Weissman, S.M. (1980). Proc. Nat. Acad. Sci. U.S.A. 76,5095-5099.

Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) Cell 19 981-992.

Embury, S., Lebo, R., Dozy, A.M. and Kan, Y.W. (1979). J. Clin. Invest. 63, 1307-1310.

Flavell, R.A., Jeffreys, A.J., and Grosveld, G.C. (1978a). Cold Spring Harbor Symp. Quant. Biol. 42, 1003-1010.

Flavell, R.A., Kooter, J.M., De Boer, E., Little, P.F.R. and Williamson, R. (1978b). Cell 15, 25-41.

Flavell, R.A., Bernards, R., Kooter, J.M., DeBoer, E., Little, P.F.R., Annison, G. and Williamson, R. (1979). Nucl. Acids Res. 6, 2749-2760.

Flessman, R.A. and Mintz, B. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 5736-5740.

Forget, B.G., Hillman, D.G., Lazarus, H., Barell, E.F., Benz, E.J., Caskey, C.T., Huisman, T.H.J., Schroeder, W.A. and Housman, D. (1976). Cell 7, 323-329.

Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1979). Nature 279, 598-603.

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

Gallwitz, D. and Seidal, R. (1980) Nucl. Acids Res. 8, 1043-1059.

Gilbert, W. (1978) Nature 271, 501.

Ginder, G.D., Wood, W.I. and Felsenfeld, G. (1979) J. Biol. Chem. 254, 8099-8102.

Goeddel, D.V., Yelverton, E., Ullrich, A., Heyneker, H.L., Miozzari, G., Holmes, W., Seeburg, P.H., Dull, T., May, L., Stebbing, N., Crea, R., Shuichiro, M., McCandliss, R., Sloma, A., Tabor, J.M., Gross, M., Familletti, P.C. and Pestka, S. (1980a). Nature 287, 411-416.

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

Goff, S.P. and Berg, P. (1979). J. Mol. Biol. 133, 359-382.

Goodman, H.M., Olson, M.V. and Hall, B.D. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5453-5457.

Goosens, M., Dozy, A.M., Embury, S., Zacharides, Z., Hadjiminas, M., Stamatoyannopoulos, G. and Kan, Y.W. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 518-521.

Gorin, M.B. and Tilghman, S.A. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 1351-1355.

Grosscheld, R. and Birnstiel, M.L. (1980). Proc. Nat. Acad. Sci. U.S.A. 77. 1432-1436.

Hamer, D.H., Smith, K.D., Boyer, S.H. and Leder, P. (1979a). Cell 17, 725-735.

Hamer, D.H. and Leder, P. (1979b). Cell 17, 737-747.

Hardison, R.C., Butler, E.T., Lacy, E., Maniatis, T., Rosenthal, N. and Efstradiadis, A. (1979). Cell 18, 1285-1297.

Havell, E.A., Berman, B., Ogburn, C.A., Berg, K., Paneker, K. and Vilcek, J. (1975). Proc. Nat. Acad. Sci. U.S.A. 72, 2185-2186.

Higgs, D.R., Old, J.M., Clegg, J.B., Pressley, L., Hunt, D.M., Weatherall, D.J. and Sergeant, G.R. (1979). Lancet ii, 272-276.

Higgs, D.R., Old, J.M., Pressley, L., Clegg, J.B. and Weatherall, D.J. (1980). Nature 284, 632-635.

Hoeijmakers-Van Dommelen, H.A.M., Grosveld, G.C., De Boer, E., Flavell, R.A., Varley, J.M. and Jeffreys A.J. (1980). J. Mol. Biol. 140, 531-547.

Houck, C.M., Rinehart, F.P. and Schmidt, C.W. (1979). J. Mol. Biol. 132, 289-306.

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.

Hsaio, C-L. and Carbon, J. (1979.) Proc. Nat. Acad. Sci. U.S.A. 76, 3829-3833.

Huisman, T.H.J., Wrightsone, R.N., Wilson, J.B., Schroeder, W.A. and Kendall, A.G. (1972). Arch. Biochem. Biophys. 153, 850-853.

Huisman, T.H.J., Schroeder, W.A., Efremov, G.D., Duma, H., Meadenovski, B., Hyman, C.B., Rachmilewitz, E.A., Bouver, N., Miller, A., Brodie, A.R., Shelton, J.R. and Appel, G. (1974). Ann. N.Y. Acad. Sci. 232, 107-124.

Hutchinson, C.A., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P. and Smith, M. (1978). J. Biol. Chem. 253, 6551-6560.

Isaacs, A. and Lindenmann, J. (1957). Proc. Roy. Soc. B 147, 258-267.

Jacq, C., Miller, J.R. and Brownlee, G.G. (1977). Cell 12, 109-120.

Jahn, C.L., Hutchinson, C.A., Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F. and Edgell, M.H. (1980). Cell 21, 149-158.

Jeffreys, A.J. and Flavell, R.A. (1977a). Cell 12, 429-439.

Jeffreys, A.J. and Flavell, R.A. (1977b). Cell 12, 1097-1108.

Jeffreys, A.J. (1979). Cell 18, 1-10.

Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M. and Williams, J.G. (1980). Cell 21, 555-564.

Jelenik, W.G., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.N., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980). Proc. Acad. Sci. U.S.A. 77, 1398-1402.

Joho, R. and Weissmann, I.L. (1980). Nature 284, 179-181.

Jones, C.W. and Kafatos, F.C. (1980) Nature 284, 635-638.

Kan, Y.W., Dozy, A., Varmus, H., Taylor, J., Holland, J., Lie-Injo, L., Ganesan, J. and Todd, D. (1975a). Nature 255, 255-256.

Kan, Y.W., Holland, J.P., Dozy, A.M., Charache, S. and Kazazian, H.H. (1975b). Nature 258, 162-163.

Kan, Y.W., Dozy, A.M., Trecartin, R. and Todd, D. (1977). New Engl. J. Med. 297, 1081-1084.

Kantor, J.A., Turner, P.H. and Nienhuis, A.W. (1980). Cell 21, 149-158.

Kedes, L.H., Chang, A.C.Y., Houseman, D. and Cohen, L.H. (1975). Nature 255, 533-538.

Kohler, G. and Milstein, C. (1975) Nature 256, 495-497.

Konkel, D.A., Maizel, J.V. and Leder, P. (1979). Cell 18, 865-873.

Korn, L.J. and Brown, D.D. (1978). Cell 15, 1145-1156.

Knapp, G., Beckmann, J.S., Johnson, P.F., Fuhrman, S.A. and Abelson, J. (1978) Cell 14, 221-236.

Knapp, G., Ogden, R.C., Peebles, C.L. and Abelson, J. (1979) Cell 18, 37-45.

Kretschmer, P.J., Kaufman, R.E., Coon, H.C., Chen, M-J. Y., Geist, G.E. and Nienhuis, A.W. (1980). J. Biol. Chem. 255, 3212-3220.

Lacy, E., Hardison, R.C., Quon, D. and Maniatis, T. (1979) Cell 18,, 1273-1283.

Laskey, R.A. and Mills, A.D. (1977) FEBS Letts 82, 314-316.

Lauer, J., Shen, C-K. J. and Maniatis, T. (1980) Cell 20, 119-130.

Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978). Cell 15, 1157-1174.

Leder, A., Miller, H.I., Hamer, D.H., Seidman, J.G., Norman, B., Sullivan, M. and Ledr, P. (1978) Proc. Acad. Sci. U.S.A. 75, 6187-6191.

Leder, P., Hansen, J.N., Konkel, D., Leder, A., Nishioka, Y. and Talkington, C. (1980). Science 209, 1336-1342.

Lerner, M., Boyle, J., Mount, S., Wolin, S. and Steitz, J. (1980). Nature 283, 220-224.

Lewin, B. (1974). Gene Expression, 2, John Wiley & Sons, London.

Little, P.F.R., Curtis, P., Coutelle, C., Van Den Berg, J., Dalgleish, R., Malcolm, S., Courtney, M., Weataway, D.A. and Williamson. (1978) Nature 273, 640-643.

Little, P.F.R., Flavell, R.A., Kooter, J.M., Annison, G. and Williamson, R. (1979) Nature 278, 227-231.

Liu, C-P., Tucker, P.W., Mushinski, J.F. and Blattner, F.R. (1980). Science 209, 1348-1353.

Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979). Cell 18, 545-558.

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

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978). Cell 15, 687-701.

Maniatis, T., Fritsch, E.F., Lauer, J. and Lawn, R.M. (1980) Annual Rev. Genet. in press.

Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L., (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 3855-3859.

Mantei, N., Boll, W. and Weissman, C. (1979). Nature 281, 40-46.

Mantei, N., Schwartzstein, M., Srueli, M., Pancin, S., Nagata, S. and Weissmann, C. (1980). Gene 10, 1-10.

Maquat, L.E., Kinnibyrgh, A.J., Beach, L.R., Hinig, G.R., Lazerson, J., Ershler, W.B. and Ross, J. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 4287-4291.

Markert, C.L. and Petters, R.M. (1978). Science 202, 56-58.

Martin, S.L., Zimmer, E.A., Kan, Y.W. and Wilson, A.G. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 3563-3566.

Maxam, A. and Gilbert, W. (1978). Proc. Nat. Acad. Sci. U.S.A. 74, 560-563.

McDonald, R.J., Crerar, M.M., Swain, W.F., Pictet, R.L. Thomas, G. and Rutter, W.J. (1980). Nature 287, 117-122.

McReynolds, L.A., Manahan, J.J., Bendure, P.W., Woo, S.L.C., Paddock, G.V., Salser, W., Dorson, J., Moses, R.E. O'Malley, B.W. (1977). J. Biol. Chem. 252, 1840-1843.

Mears, J.G., Ramirez, F., Leibowitz, D. and Bank, A. (1978). Cell 15, 15-23.

Melli, M., Spinelli, G. and Wilson, M. (1977). Experientia 33, 825.

Merrigan, T.C., Rand, K.H., Pollard, R.B., Abdallah, P.S., Jordan, G.W. and Fried, R.P. (1978). New Engl. J. Med. 298, 981-987.

Mertz, J.E. and Gurdon, J.B. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 1502-1505.

Murray, N.E., Brammar, W.J. and Murray, K. (1977). Mol. and Gen. Genet. 150, 53-61.

Nagata, S., Taira, H., Hall, A., Johnsrud, L., Strueli, M., Ecsodi, J., Boll, W., Cantell, K. and Weissman, C. (1980a). Nature 284, 316-320.

Nagata, S., Mantei, N. and Weissmann, C. (1980b). Nature 287, 401-408.

Nasmyth, K.A. and Tatchell, K. (1980). Cell 19, 753-764.

Ng, S.Y., Parker, C.S. and Roeder, R.G. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 136-140.

Nienhuis, A.W., Turner, P. and Benz, E.J. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 3969-3964.

Nishioka, Y., Leder, A. and Leder, P. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 2806-2809.

Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N. and Schimke, R.T. (1980). Cell 19, 355-364.

O'Hare, K., Breathnach, R., Benoist, C. and Chambon, P. (1979). Nucl. Acids Res. 7, 321-334.

Ohshima, S. and Suzuki, Y. (1977). Proc.Nat. Acad. Sci. U.S.A. 74, 5363-6366.

Old, J.M., Proudfoot, N.J., Wood, W.G., Longley, J.I., Clegg, J.B. and Weatherall, D.J. (1978). Cell 14, 289-298.

Orkin, S.H. (1978). Proc Nat. Acad. Sci. U.S.A. 75, 5950-5954.

Orkin, S.H., Old, J.M., Weatherall, D.J. and Nathan, D.G. (1979a). Proc. Nat. Acad. Sci. U.S.A. 76, 2400-2404.

Orkin, S.H., Old, J.M., Lazarus, H., Altay, C., Gurgey, A., Weatherall, D.J. and Nathan, D.G. (1979b). Cell 17, 33-42.

Orkin, S.H. and Michelson, A. (1980). Nature 286, %38-540.

Orkin, S.H., Koloder, R., Michelson, A. and Husson, R. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 3558-3562.

Ottolenghi, S., Lanyon, W.G., Paul, J., Williamson, R., Weatherall, D.J., Clegg, J.B., Pritchard, J., Pootrakul, S. and Wong, H.B. (1974). Nature 251, 389-392.

Ottolenghi, S., Giglioni, B., Comi, P., Gianni, A.M., Polli, E., Acquaye, C.T.A., Oldhom, J.H. and Masera, G. (1979) Nature 278, 654-679.

Patient, R.K., Elkington, J.A., Kay, R.M. and Williams, J.G. (1980). Cell 21, 565-574.

Pellicer, A. Wigler, M. Axel, R. and Silverstein, S. (1978). Cell 14, 133-141.

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

Pressley, L., Higgs, D., Clegg, J.B. and Weatherall, D.J. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 3586-3589.

Pribnow, D. (1975). J. Mol. Biol. 99, 419-443.

Proudfoot, N.J. and Baralle, F. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 5435-5439.

Proudfoot, N.J. and Maniatis, T. (1980). Cell 21, 537-547.

Ramirez, F., Burns, A.L., Mears, J.G., Spence, S., Starkman, D. and Bank, A. (1979). Nucl. Acids Res. 7, 1147-1162.

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

Robbins, J., Rosteck, P., Haynes, J.R., Freyer, G., Cleary, M.L., Kalter, H.D., Smith, K. and Lingrel, J.B. (1979) J. Biol. Chem. 254, 6187-6195.

Roberts, J.L., Seeburg, P.H., Shine, J., Herbert, E., Baxter, J.D. and Goodman, H.M. (1979). Proc Nat. Acad. Sci. U.S.A. 76, 2153-2157.

Rogers, J. and Wall, R. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 1877-1879.

Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M-J. and O'Malley, B.W. (1978). Cell 15, 671-685.

Rosenberg, M. and Court, D. (1979). Annual Rev. Genet. 13, 319-355.

Ross, J. and Knecht, A. (1978). J. Mol. Biol. 119, 1-34.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J.L., Lemur, M., Bregegegre, F., Gannon, F., LePennec, J.P., Chambon, P. and Kourilsky, P. (1979). Nature 279, 125-132.

Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A. Maki, R., Wall, R. and Tonegawa, S. (1979). Nature 277, 627-633.

Sanger, F. and Coulson, A.R. (1975). J. Mol. Biol. 94, 441-448.

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

Scherer, S. and Davis, R.W. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 4951-4955.

Schreier, P.H. and Cortese, R. (1979). J. Mol. Biol. 129, 169-172.

Secher, D.S. and Burke, D.C. (1980). Nature 285, 446-450.

Schaffner, W. Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978). Cell 14, 655-671.

Shen, C-K.J. and Maniatis, T. (1980). Cell 19, 379-391.

Shortle, D. and Nathans, D. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 2170-2174.

Slightom, J., Blechl, A. and Smithies, O. (1980). Cell 21, 627-638.

Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980). Cell 19, 13-26.

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

Southern, E.M. (1979). Anal. Biochem. 100, 304-318.

Sternberg, N., Tiemeier, D. and Enquist, L.W. (1977). Gene 1, 255-318.

Stewart, W.E. et al (1980) Nature 286, 110.

Strander, H., Mogensen, K.E. and Cantell, K. (1974). J. Clin. Microbiol. 116-117.

Taniguchi, T. Sakai, M., Fujii-Kuriyama, Y., Muramatsu, M., Kobayashi, S. and Sudo, T. (1979a). Proc. Japan. Acad. 55, Ser. B. 464-469.

Taniguchi, T., Mantei, N., Schwartztein, M., Nagata, S., Muramatsu, M. and Weissman, C. (1980). Nature 285, 547-549.

Taylor, J.M., Dozy, A.M., Kan, Y.W., Varmus, H.E., Lie-Injo, L.E., Ganesan, J. and Todd, D. (1974). Nature 251, 392-393.

Tiemeier, D.C., Tilghman, S.M., Polsky, F.I., Seidman, J.G., Leder, A., Edgell, M.H. and Leder, P. (1978). Cell 14, 237-245.

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J. and Leder P. (1978a). Proc. Nat. Acad. Sci. U.S.A. 75, 725-728.

Tilghman, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P. and Weissman, C. (1978b). Proc. Nat. Acad. Sci. U.S.A. 75, 1309-1313.

Tobin, S.L., Zulauf, E., Sanchez, F., Craig, E.A. and McCarthy, B.J. (1980). Cell 19, 121-131.

Tsai, S.Y., Roop, D.R., Stumph, W.E., Tsai, M-J. and O'Malley, B.W. (1980). Biochem. 19, 1755-1761.

Tsujimoto, Y. and Suzuki, Y. (1979). Cell 18, 581-600.

Tuan, D., Biro, P.A., de Riel, J.K., Lazarus, H. and Forget, B.G. (1979). Nucl. Acids Res. 6, 2519-1544.

Twigg, A.J. and Sherrat, D. (1980). Nature 283, 216-218.

Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. and Rutter, W.J. (1978). Proc. Nat, Acad. Sci. U.S.A. 75, 190-194.

Van den Berg, J., van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R.A. and Weissman, C. (1978). Nature 276, 37-44.

Van der Ploeg, L.H.T., Konnings, A., Oort, M., Roos, D., Bernini, L. and Flavell, R.A. (1980). Nature 283, 637-642.

Van der Ploeg, L.H.T. and Flavell, R.A. (1980). Cell 19, 947-958.

Van Ooyen, A., van den Berg, J., Mantei, N. and Weissman, C. (1979). Science 206, 337-344.

Vanin, E.E. Goldberg, G.I., Tucker, P.W. and Smithies, O. (1980). Nature 286, 222-226.

Wahli, W., Dawid, I.B., Wyler, T., Weber, R. and Ryffel, G.U. (1980). Cell 20, 107-117.

Wall, R., Gilmore-Herbert, M., Higuchi, R., Komaromy, M., Paddock, G., Strommer, J. and Salser, W. (1978). Nucl. Acids Res. 5, 3113-3129.

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

Wasylyk, B., Kedinger,C., Corden, J., Brison, O. and Chambon, P. (1980). Nature 285, 367-340.

Weatherall, D.J. and Clegg, J.B. (1972). The Thalassaemia Syndromes, 2nd Edition. Blackwell Scientific.

Weatherall, D.J. and Clegg, J.B. (1979). Cell 16, 467-479.

Weaver, S., Haigwood, N.L., Hutchinson, C.A. and Edgell, M.H. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 1385-1389.

Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J. and Hood, L. (1980) Nature 283, 497-499.

Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979). Cell 18, 469-484.

Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Cell 16, 777-785.

Williams, J.G. and Lloyd, M.M. (1979). J. Mol. Biol. 129, 19-35.

Wilson, J.T., Wilson, L.B., de Riel, J.K., Villa-Komaroff, L., Efstratiadis, A., Forget, B.G. and Weissman, S.M. (1978). Nucl. Acids Res. 5, 563-581.

Wood, W.B. (1966). J. Mol. Biol. 16, 118-133.

Wu, G. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 2175-2178.

Zago, M.A., Wood, W.G., Clegg, J.B., Weatherall, D.J., O'Sullivan, M. and Gunson, H. (1979) Blood 53, 977-986.

Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. and Wilson, A.C. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 2158-2162.

Zoon, K.C., Smith, M.E., Bridgen, P.J., zur Nedden, D. and Anfinson, C.B. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 5601-5605.

Pelham, H.R.B. and Brown, D.D. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 4170-4174.

Tan, Y.H., Barakat, F., Berthold, W., Smith-Johannsen, H. and Tan, C. (1979) J. Biol. Chem. 254, 8067-8073.