THE ISOLATION OF THE HUMAN GAMMA GLOBIN GENES

BY

RAYMOND WALTER MILLER DALGLEISH

a thesis submitted for the degree of

Doctor of Philosophy

in the

University of London

November, 1979.

Department of Biochemistry,

St Mary's Hospital Medical School,

London, W2 1PG.

RAYMOND WALTER MILLER DALGLEISH

THE ISOLATION OF THE HUMAN GAMMA GLOBIN GENES

A B S T R A C T

The investigation of the organisation and expression of eukaryiotic genes has been greatly aided by the development of recombinant DNA technology. In order to study the human gamma globin genes, a recombinant bacterial plasmid, containing a gamma globin cDNA sequence, was characterised and the construction of a recombinant plasmid containing a genomic gamma globin sequence, including both gamma globin genes, was undertaken.

The recombinant cDNA plasmid, pHγGl, was characterised by hybridisation, in solution, to human globin cDNA-αβγ and cDNA-αβ. This demonstrated that pHγGl could be used as a specific hybridisation probe for human gamma globin gene sequences.

Total human DNA was prepared and digested to completion with the restriction endonuclease Bgl II. The digested DNA was fractionated by electrophoresis in a Southern 'Gene Machine'. Gamma globin specific fractions were identified by Southern transfer hybridisation using radioactively labelled pHγGl DNA as probe. This DNA was fractionated further by reversed phase chromatography on RPC-5. Fractions were probed as before.

Single stranded homopolymer deoxynucleotide tails were added, using terminal transferase, to the termini of the purified DNA molecules. Complementary tails were added in an analogous manner to linearised pBR322 plasmid DNA. The tailed plasmid and human DNAs were allowed to anneal and the resulting recombinant molecules were used to transform E.coli χ1776. Identification of recombinant clones was attempted by hybridisation of radioactively labelled pHγGl DNA to lysed bacterial colonies on nitrocellulose filters. No positive scoring colonies were obtained.

In another set of experiments, Bgl II digested human DNA was fractionated by RPC-5 chromatography alone. Purified fractions of DNA were ligated into the Bam HI site of the plasmid pAT153 and the resulting molecules were used to transform E.coli HB101. Colonies of bacteria were screened as before. As yet no positive scoring colonies have been identified.

To my parents.

Acknowledgements

I would like to thank everyone at the Department of Biochemistry, St Mary's Hospital Medical School for their help. I especially thank Professor Bob Williamson and Dr. Peter Little for their advice and encouragement.

Thanks to Heather Grist for her understanding.

I am indebted to Lynne Sheasgreen and to Caroline Rogers for typing the script and to the M.R.C. for providing a student scholarship.

ABBREVIATIONS

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
hnRNA	heterogeneous nuclear RNA
tRNA	transfer RNA
cDNA	complementary DNA
A,C,G,T	Adenine, Cytosine, Guanine, Thymine (or, where appropriate; Adenosine...., or Adenylic acid....etc. prefix d denotes the deoxy form).
E.coli	*Escherichia coli*
Hb	haemoglobin
thal	thalassaemia
BSA	bovine serum albumin
EDTA	ethylenediaminetetraacetic acid
SDS	sodium dodecyl sulphate
PCA	perchloric acid
TCA	trichloroacetic acid
Tris	2-amino-2(hydroxymethyl)-propane-1,3-diol
Hepes	N-2-hydroxyethyl piperazine-N-2-ethane sulphonic acid
DEAE	diethylaminoethyl
RNase	ribonuclease
DNase	deoxyribonuclease
bp	base pair

ABBREVIATIONS continued

kb	kilobase pairs
dpm	disintegrations per minute
cpm	counts per minute
M	molar
w/v	weight per volume
v/v	volume per volume
g	gravity
rpm	revolutions per minute
HPLC	high pressure liquid chromatography
UV	ultra violet

CONTENTS

Abstract 2

Dedication 3

Acknowledgements 4

Abbreviations 5

Chapter 1 Introduction

1.1 Gene Expression In Eukaryotes 16

1.2 The Organisation Of DNA Sequences 17

1.3 The Organisation Of Specific Gene Sequences

(a) Total structural gene number 19

(b) Specific gene number 19

(c) Repeated genes 21

(d) Introns 22

1.4 The Human Globin Genes

(a) Haemoglobin 25

(b) Globin gene number 25

(c) The arrangement of the β-like globin genes 26

(d) The α-globin gene arrangement 27

(e) Chromosomal location of globin genes 28

(f) The thalassaemias 28

1.5 Recombinant DNA

(a) The Basic Techniques 32

(b) Vectors 34

(c) Safety 35

1.6 Objectives 37

Chapter 2 Materials And Methods

2.1 Materials 38

2.2 General Methods

(a) Agarose gel electrophoresis 41

(b) Gel staining and photography 41

(c) Gel filtration 41

(d) Ethanol precipitation of nucleic acid 42

(e) Restriction enzyme digestions 42

(f) Determination of radioactivity 43

2.3 Specific Methods

(a) Preparation of mRNA 44

(b) Preparation of cDNA 45

(c) Solution hybridisation 46

(d) Kinase labelling of RNA 47

(e) Plasmid DNA preparation 48

(f) Transfer hybridisation 50

(g) Labelling of plasmid DNA by nick translation 51

(h) Preparation of human DNA 51

(i) Preparation of Bgl II 52

(j) 'Gene Machine' 53

(k) RPC-5 chromatography 53

(l) Tailing of DNA 53

(m) Transformation of E.coli χ1776 54

(n) Transformation of E.coli HB101 54

(o) Preparation of replica filters for screening 55

(p) Grunstein-Hogness screening 56

(q) Ligation of DNA 56

(r) Phosphatase Treatment of vector DNA 57

(s) Preparation of single colony lysates 57

(t) Electron microscopy of DNA 57

Chapter 3 Results

3.1 Preparation Of Hybridisation Probes

(a) Purification of human foetal globin mRNA 59

(b) Synthesis of foetal globin cDNA 59

(c) Hybridisation of cDNA to template mRNA 61

(d) ^{32}P labelling of mRNA with polynucleotide kinase 62

3.2 <u>Characterisation Of pHγG1</u>

(a) Construction of recombinant plasmids containing human α-, β-, and γ-globin cDNA fragments 63

(b) Plasmid DNA preparation 63

(c) Verification of plasmid identification by solution hybridisation 64

3.3 <u>Mapping Of The γ-Globin Genes</u>

(a) Identification of a DNA fragment encoding both γ-globin genes 68

(b) Preparation of human DNA 68

(c) Digestion of human DNA with Bgl II 70

(d) Transfer hybridisation of Bgl II digested human DNA 72

(e) Verification of a single 13 kb γ-specific fragment by RPC-5 chromatography 72

3.4 <u>Purification Of Human DNA For Cloning</u>

(a) Introduction 80

(b) The preparation of Bgl II 80

(c) Strategy for the purification of the 13 kb Bgl II DNA fragment 81

(d) Fractionation of DNA on the Southern 'Gene Machine' 81

(e) RPC-5 chromatography of 'Gene Machine' purified DNA 83

(f) Estimation of the overall purification 83

3.5 <u>Cloning In The Host/Vector System *E.coli* χ1776/pBR322</u>

(a) Introduction 91

(b) The ligatability of Bgl II digested DNA 92

(c) Terminal transferase 94

(d) Tailing of pBR322 94

(e) Tailing of purified DNA 96

(f) Transformation of E.coli χ1776 98

(g) Annealing of tailed plasmid with insert and transformation 98

(h) Screening of recombinant clones 99

(i) Preparation and analysis of recombinant DNA 102

(j) Investigation of DNA for possible contaminants 102

3.6 Cloning In The Host/Vector System E.coli HB101/pAT153

(a) Introduction 106

(b) Restriction enzyme digestion of human and plasmid DNAs 106

(c) Fractionation of human DNA by RPC-5 chromatography 109

(d) Phosphatase treatment of linearised plasmid DNA and ligation 109

(e) Single colony lysates 112

(f) Ligation of DNA for cloning 114

Chapter 4 Discussion

4.1 The Preparation And Use Of Cloned cDNA Probes

(a) cDNA vs cloned cDNA 119

(b) The preparation and screening of the cDNA clones 120

(c) The use of the globin cDNA clones 121

4.2 The Cloning Of The γ-Globin Genes

(a) Mapping the γ-globin genes 124

(b) The preparation of Bgl II 124

(c) Purifying DNA for cloning 125

(d) The choice of host/vector systems 127

(e) Homopolymer tailing 129

(f) Cloning with tailed DNA 131

(g) Cloning by ligation 132

4.3 Cloned Globin DNA Sequences - Future Prospects

(a) Sequence comparisons and hybridisation studies 136

(b) Expression of cloned sequences 139

Bibliography 141

LIST OF FIGURES

Figure No.		Page
1	1.2% agarose/SDS gel electrophoresis of human globin mRNA	60
2	0.8% agarose gel electrophoresis of plasmid pHγGl DNA	65
3	Hybridisation of human globin cDNA-αβγ and cDNA-αβ to plasmid pHγGl DNA - titration curve	67
4	0.8% agarose gel electrophoresis of human DNA	69
5	0.8% agarose gel electrophoresis of Bgl II digested human DNA	71
6	Schematic representation of a Southern transfer	73
7	Autoradiogram of Southern transfer hybridisation of Bgl II digested human DNA	74
8	Elution profile of Bgl II digested human DNA from RPC-5	76
9a	0.8% agarose gel electrophoresis of RPC-5 fractionated Bgl II digested human DNA	77
9b	Autoradiogram of Southern tranfer hybridisation of RPC-5 fractionated Bgl II digested human DNA	79
10	Schematic representation of a section through a Southern 'Gene Machine'	82

Figure No.		Page
11	Autoradiogram of Southern transfer hybridisation of "Gene Machine" fractionated DNA	84
12	Autoradiogram of Southern transfer hybridisation of RPC-5 fractionation of DNA previously purified by "Gene Machine" fractionation	85
13	0.8% agarose gel electrophoresis of DNA fractionated by both "Gene Machine" and RPC-5	87
14	Autoradiogram of Southern transfer hybridisation of the agarose gel in figure 13	88
15	Optical density profiles of purified and unpurified DNA	89
16	0.8% agarose gel electrophoresis of ligated DNA samples	93
17	1% agarose gel electrophoresis of Hind III digested pBR322 DNA	95
18	Incorporation of ^{3}H-dGTP by terminal transferase into fraction 24 DNA	97
19	Grunstein - Hogness control screening	101
20	Grunstein - Hogness screening of recombinant clones	103
21	0.8% agarose gel electrophoresis of plasmid DNA from clones 4G6, 6E7 and 6F2	104
22	0.8% agarose gel electrophoresis of ligated Bgl II digested human DNA	105

Figure No.		Page
23	1% agarose gel electrophoresis of ligated Bam HI digested pAT153 DNA	108
24	0.8% agarose gel electrophoresis of RPC-5 fractionated Bgl II digested human DNA	110
25	Autoradiogram of Southern transfer hybridisation of the agarose gel in figure 24	111
26	0.8% agarose gel electrophoresis of ligated DNA	113
27	0.8% agarose gel electrophoresis of DNA from single colony bacterial lysates	115
28	0.8% agarose gel electrophoresis of ligated DNA	116

LIST OF TABLES

Table		Page
1	Summary of transformations of *E.coli* χ1776 with tailed vector and insert DNA	100
2	Summary of transformations of *E.coli* HB101 with vector and insert DNA ligated to each other at different ratios	118

Chapter 1

Introduction

1.1 Gene expression in eukaryotes

Despite intensive effort over the last two decades no coherent model has yet emerged for the control of the expression of eukaryotic genes. In contrast, the model of Jacob and Monod (1961) for prokaryotic gene expression is now well established and the finer points of control have been studied in detail for a number of systems.

The complexity of the processes involved in eukaryotic gene expression is evident. The processes themselves are sequential and include making the gene accessible for transcription, its transcription into a precursor RNA, the processing of the precursor RNA to give mRNA, the transport of the mature mRNA to the cytoplasm, the translation of the mRNA into protein and finally the post translational modification of the protein. Evidently, gene expression requires the co-ordination of a complex series of processes.

The emergence of a coherent model will only come about with an understanding of and an insight into these many diverse aspects of gene expression. Perhaps more than one model of control will be required to accommodate the different "classes of genes" such as as "housekeeping genes" and genes involved in terminal differentiation.

Many different approaches are being taken towards the establishment of a model and many have led to the establishment of new branches of biochemistry. The following discussion will attempt to summarise some of the advances both in our accumulated knowledge and in the associated technology.

1.2 The organisation of DNA sequences

The most direct approach to the problem of the control of gene expression in eukaryotes is to study the organisation of DNA both at a gross level and within specific genes and gene families.

The former approach produced the concept of different frequency classes within genomic DNA characterised by different rates of reassociation following denaturation (Britten and Kohne, 1968). The slow reassociating class includes sequences present in few copies in the genome as well as unique sequences present as a single copy (Laird, 1971). A number of studies (reviewed by Davidson and Britten, 1973), with both specific mRNA probes and with populations of mRNAs, have shown that the majority of structural genes, that is genes with a protein as an end product, fall into the slow reassociating class of DNA sequences, while a fraction of structural genes are present at higher multiplicity and would thus reassociate more quickly.

The intermediate reassociating class includes the genes for histones and ribosomal RNA, of which there are multiple copies. These genes are organised in tandemly repeated units, each unit consisting of one of each histone gene or of an 18S and 28S ribosomal gene as the case may be.

The fast reassociating class of DNA includes highly tandemly repetitive DNA sequences. These sequences consist of large numbers of simple repeated identical or similar elements (Walker *et al.*, 1970). They band separately from the bulk of the cellular DNA in CsCl gradients so are referred to as satellite sequences and constitute about 10% of the genome. This type of DNA is associated with constitutive heterochromatin (Yunis and Yasmineh, 1971) and is often concentrated close to

centromeres (Jones, 1970; Pardue and Gall, 1970) or chromosome ends (Hennig _et al._, 1970). It appears that these sequences are only transcribed to a very small extent, if at all (Walker, 1971). The fast reassociating class of DNA also includes palindromic sequences which were first identified because of their concentration independent mode of binding to hydroxylapatite following denaturation (Britten and Smith, 1970; Wilson and Thomas, 1973). This effect is due to the formation of intrastrand double stranded structures (Wilson and Thomas, 1974). These palindromic sequences are generally between 300 and 1 200 nucleotides in length, occur in groups of 2 to 4 and are widely distributed throughout the eukaryotic genome. Sequences flanking the palindromes contain both unique and repetitive DNA sequences (Davidson _et al._, 1973).

Unique and moderately repetitive DNA sequences have been demonstrated to be interspersed (Davidson and Britten, 1973; Chamberlin _et al._, 1975). This pattern has been observed in many species which suggests an underlying functional significance.

1.3 The organisation of specific gene sequences

(a) Total structural gene number

Another aspect of the eukaryotic genome that has received much attention in relation to gene control is the total number of structural genes. What percentage of the genome has coding potential for protein sequences? In mouse brain an estimate of the percentage of the genome that is transcribed is 10% (Hahn and Laird, 1971). Bantle and Hahn (1975) demonstrated that approximately a fifth of the transcribed RNA in mouse brain is processed to mature mRNA that will have coding potential. This portion is equal to 7.6% of the non-repeated DNA complexity.

In the human genome approximately 3% is thought to be expressed as protein and this represents about 7×10^4 structural genes (Bishop, 1974).

In sea urchins 28.5% of the sequence complexity of the genome is represented in hnRNA and of this less than 10% is processed to polysomal mRNA (Hough et al., 1975). The number of expressed structural genes is estimated to be 1.4×10^4. A similar estimate for HeLa cells is 3×10^4 (Bishop et al., 1974) and for mouse brain is 10^5 (Bantle and Hahn, 1976).

It is clear that only a very small portion of the eukaryotic genome is transcribed and of the transcription products only a small portion has coding potential. The non-coding portion may have controlling functions, be present for structural purposes for RNA processing or may be "nonsense" transcripts having no apparent function.

(b) Specific gene number

The number of copies of an individual structural gene can be ascertained using radiolabelled specific mRNAs, or their

cDNA copies, as gene-specific hybridisation probes. Genes analysed in this manner include those for mouse globin (Harrison et al., 1972; Harrison et al., 1974), duck globin (Bishop et al., 1972; Bishop and Rosbach, 1973), Bombyx mori fibroin (Suzuki et al., 1972; Gage and Manning, 1976), sea urchin histones (Kedes and Birnstiel, 1971; Weinberg et al., 1972) and chicken collagen (Frischauf et al, 1978). In each case, except for the histone genes, the copy number was found to be one or a few copies per structural gene.

The technique of probing genomic DNA to estimate gene number can indicate whether or not specific genes are either amplified during the ontogeny of specialised cells or deleted in cells that do not express that gene.

Packman et al. (1972) and Bishop et al.(1972) compared globin mRNA hybridisation to DNA extracted from haemoglobin synthesising erythrocyte precursor cells with that to DNA from non-erythroid cell types. No differences were observed. Suzuki et al. (1972) carried out similar experiments with Bombyx mori fibroin mRNA. They hybridised it both to DNA from whole larvae and to DNA from cells not synthesising fibroin. Again the results were identical, demonstrating no amplification for the genes studied. Kafatos (1972 a, b) argues that such an amplification is not necessary to explain the large accumulation of protein that occurs in these and other cell types.

In contrast, amplification of the gene for dihydrofolate reductase has been demonstrated in methotrexate resistant mouse sarcoma and lymphoma cells (Alt et al.,1978). Increased synthesis of dihydrofolate reductase with increased dosage of methotrexate is shown to be associated with gene amplification.

Kabat (1972) proposed that during erythroid differentiation

a looping out excision of globin genes takes place in erythroid stem cell populations, explaining the sequential expression of different globin genes in successive red cell populations. Mitchell and Williamson (1977) have shown this model to be invalid by demonstrating the full complement of γ-globin genes in human nucleated erythroid cells.

Bishop and Freeman (1973) have demonstrated the association of structural genes with neighbouring repetitive sequences. Duck globin genes are adjacent to sequences with a repetition frequency of about 20. A similar result has also been described for the rabbit globin genes (Flavell *et al.*, 1978 b). In both cases the repetitive sequences are to the 5' side of the gene.

Davidson *et al.* (1975) have demonstrated that the unique DNA in *Xenopus* is bounded by regions of repetitive DNA throughout the genome and that the pattern of association falls into a number of different classes.

These results are consistent with eukaryotic operon models proposed by a number of workers (Georgiev, 1969; Crick, 1971; Britten and Davidson, 1969).

(c) Repeated genes

Both histone and ribosomal genes are present in multiple copies in the eukaryotic genome.

In sea urchins the histone genes are reiterated about 300 to 1 000 times depending on the species (Weinberg *et al.*, 1972; Grunstein *et al.*, 1973). In contrast the histone genes in *Xenopus laevis* (Birnstiel *et al.*, 1975) and man (Wilson *et al.*, 1974) are reiterated only 10 to 20 times. In *Xenopus laevis* oocytes there are large stores of histones and histone mRNAs and this might explain how the large requirement for histones

during early embryonic development is met without gene multiplicity to the extent found in other genera.

Restriction endonuclease mapping of histone gene DNA with histone mRNAs as probes showed that the genes coding for the different histone proteins are interspersed in groups that are tandemly repeated (Weinberg *et al*., 1975; Schaffner *et al*., 1976). Other groups demonstrated that within tandem repeat units each of the histone are represented once, all the genes have the same orientation and each is separated from the next by a spacer region (Kedes *et al*., 1975; Homes *et al*., 1977; Schaffner *et al*., 1978).

The ribosomal and transfer RNA genes are in clustered repeats in all eukaryotes (Lewin, 1975). The repeats have been shown to consist of transcriptional units separated by non-transcribed spacer regions in *Xenopus* ribosomal genes (Wellauer *et al*., 1974; Wellauer and Reeder, 1975; Carroll and Brown, 1976 a, b). In *Drosophila* the ribosomal DNA was shown to have two distinct repeat sizes (Tartof and Dawid, 1976). It was later shown that the larger repeated size was caused in part by a 5 kb segment of DNA inserted into the coding region for the 28S rRNA (Glover and Hogness, 1977; White and Hogness, 1977). Wellauer and Dawid (1977) showed that the inserts are heterogeneous in length and that they are the major source of length heterogeneity within rDNA repeats.

(d) *Introns*

Introns interupt coding sequences in eukaryotic genes. Variously referred to as "introns", "insertions", "inserts" or "intervening sequences", they were first described in *Drosophila* rDNA. Ribosomal genes with and without introns were described

and those with introns were proposed to be "inactivated" genes. So far the evidence supports this hypothesis for ribosomal genes, although not, of course, for protein coding genes. During the period of this study introns have been described in many species.

The existence of introns has been demonstrated in the sequence for the large rRNA in certain strains of Tetrahymena (Wild and Gall, 1978), the mitochondrial rRNA genes of certain strains of yeast (Bos et al., 1978) and in chloroplast rRNA genes in Chlamydomonas (Rochaix and Malnoe, 1978). Introns have also been described in yeast tRNA genes (Goodman et al., 1977; Valenzuela et al., 1978; Etcheverry et al., 1979). The genes described so far all specify RNA.

The first description of an intron in a gene coding for a protein was for adenovirus, where a "leader" sequence is spliced to the 5' end of late mRNAs (Berget et al., 1977; Chow et al., 1977). In this case the intron lies not in the coding sequence but in the 5' non-coding region.

Introns were later described in higher eukaryotes: rabbit β-globin (Jeffreys and Flavell, 1977), mouse α and β-globin (Tilghman et al., 1978; Konkel et al., 1978), human δ and β-globin (Flavell et al., 1978 a; Mears et al., 1978), human γ-globin (Little et al., 1979; Tuan et al., 1979), chicken ovalbumin (Dugaiczyck et al., 1978; Garapin et al., 1978; Mandel et al., 1978) and mouse immunoglobulin (Brack et al., 1978; Seidman et al., 1978). The discovery of introns has resulted in much speculation as to their origin and function and the mechanisms whereby the RNA transcribed from genes with introns is spliced to give mature mRNA (Crick, 1979).

Studies of the intron-exon boundaries (the regions where splicing occurs) have been carried out for the ovalbumin gene

(Breathnach *et al.*, 1978; Kourilsky and Chambon, 1978). It has been shown that an intron sequence always starts with GU and ends with AG. This holds true for two globin introns (Van den Berg *et al.*, 1978; Konkel *et al.*, 1978), for the small intron in mouse λ light chain of immunoglobulin (Tonegawa *et al.*, 1978; Bernard *et al.*, 1978), and for several introns of SV40 (Ghosh *et al.*, 1979). Studies of the intron-exon junctions in a number of species has led to the proposal of a "consensus sequence" from which all junctions are evolved (Catterall *et al.*, 1978). However, while junction sequences may be related they are clearly not conserved exactly, even for two introns in a given gene.

Recently an RNA species with sequence complementary to the consensus sequence has been described (J. Steitz, personal communication). This may be involved in stabilising a hairpin structure necessary for the splicing out of introns.

1.4 The human globin genes

(a) Haemoglobin

Oxygen transport in the blood is carried out by haemoglobin, a tetrameric protein with prosthetic haem groups. The protein tetramer in all normal cases consists of two α-like and two β-like protein chains.

In adults the predominant haemoglobin is Hb A ($\alpha_2\beta_2$) but there is also a minor haemoglobin HbA$_2$($\alpha_2\delta_2$) which constitutes about 2 to 3% of the total haemoglobin. The δ chains are similar to the β chains, differing in only 10 amino acids.

During foetal development the major haemoglobin is HbF ($\alpha_2\gamma_2$). Two distinct γ chains can be detected: $^G\gamma$ with a glycine at amino acid 136 and $^A\gamma$ with an alanine at that position. Genetic evidence indicates that the two gamma-globin chains originate from different structural gene loci (Schroeder _et al._, 1968).

During the first eight weeks of embryonic life three other haemoglobins are found: Hb Gower 1 ($\zeta_2\varepsilon_2$), Hb Gower 2 ($\alpha_2\varepsilon_2$) (Huehns _et al._, 1961, 1964) and Hb Portland ($\zeta_2\gamma_2$) (Capp _et al._, 1967). The ε chain is related to the β like chains (β, δ and γ) and the ζ chain is related to α (Capp _et al._, 1970; Kamuzora and Lehmann, 1975).

(b) Globin gene number

Until recently data concerning globin gene number were derived solely from studies of patients with haemoglobinopathies or haemoglobin chain variants. Such studies suggested that there is a single gene locus per haploid chromosome for the β chain since in patients heterozygous for chain variants, 50% of chains are variant and 50% normal. In double heterozygotes for

β chain variants no HbA is detected.

Similar approaches for the α-globin genes have proved confusing but it became apparent that the normal complement of α-globin genes is 2 per haploid chromosome. This was shown by hybridisation studies of cDNA annealed to the cellular DNA of patients with HbH disease. The possibility that there was a multiple of four genes present per diploid chromosome was not formally ruled out (Kan et al., 1975a).

Similar hybridisation studies for β and δ-globin genes confirm the gene numbers determined by genetic studies (Williamson, 1976) i.e. there are single loci for β and δ.

Apart from the two γ-globin chains already described, a third γ-globin sequence is frequently found in some populations and is synthesised in large amounts in some cases of β^{O}-thalassaemia (Ricco et al., 1976). This γ chain variant, "$^{T}\gamma$", has a threonine residue at position 75. Hybridisation analysis with cDNA indicates two or three human γ-globin genes (Old et al., 1976; Mitchell and Williamson, 1977) which does not resolve the problem of whether or not "$^{T}\gamma$" is allelic with either $^{G}\gamma$ and $^{A}\gamma$. The numbers of ε and ζ-globin genes is unknown.

(c) The arrangement of the β-like globin genes

Linkage of the gene loci coding for the β-like globin chains has been established, in part, by study of the abnormal haemoglobins Hb Lepore and Hb anti-Lepore (see Lang and Lorkin, 1976) which are characterised by δβ and βδ protein fusion products respectively. They suggest close proximity of the δ and β loci in the human genome since the fusion products are assumed to have arisen by an unequal crossing over event involving the β and δ genes.

Hb Kenya is a similar type of protein chain fusion haemoglobin. The non α chain consists of the fusion of the N-terminal sequence of the $^{A}\gamma$ chain and the C-terminal region of the β-chain. This demonstrates the close linkage of the $^{A}\gamma$-globin gene with that for β-globin.

These studies have led to the proposal that the gene order is $^{G}\gamma^{A}\gamma\ \delta\beta$ or $^{A}\gamma\delta\beta^{G}\gamma$, with all genes transcribed in the same direction.

(d) The α-globin gene arrangement

In most instances a β chain variant amounts to 50% of β globin whereas an α chain variant can commonly give rise to 25%, 33% or 50% of the total α chains. This suggests duplication of the α gene, and Lehmann and Carrell (1968) suggested that the 25% figure arose by the variant being coded for by one of four α-globin genes which carried a mutation. This is consistent with studies of α thalassaemia (Kattamis and Lehmann, 1970a, b) and led to Lehmann (1970) proposing a model for α thalassaemia based on four α globin genes per diploid genome. The expression of Hb Constant Spring also supports the proposal of two α-globin genes per haploid chromosome.

In Negroes it was suggested that there may be no duplication of the α-globin genes, since neither HbH disease nor Hb Bart's hydrops fetalis are described in these populations (Lehmann, 1970). Later evidence was to show that the α-globin gene is duplicated in black populations (Bellevue et al., 1979; Dozy et al., 1979). Tolstoshev et al., (1977) investigated the number of α-globin genes in patients showing various levels of the α-globin variant HbJ Mexico. In all cases studied they found two α-globin genes per haploid chromosome.

(e) Chromosomal location of globin genes

By hybridising DNA from various mouse-human somatic cell hybrids to human globin cDNA, Deisseroth *et al.*, (1977, 1978) have localised the chromosomal position of the α-globin gene locus and the γγδβ-globin gene locus on chromosomes 16 and 11 respectively.

More recently the position within chromosome 11 of the γγδβ-globin gene locus has been determined by hybridisation studies. Mouse-human somatic cell hybrids were again used but in this case there were various different deletions of the human chromosome 11 that allowed localisation to a position close to the centromere (Gusella *et al.*, 1979).

(f) The thalassaemias

The thalassaemias are genetically inherited anaemias characterised by an imbalance in the synthesis of the adult globin chains of haemoglobin (Clegg and Weatherall, 1976; Weatherall, 1976; Forget, 1978; Weatherall and Clegg, 1979b).

In α-thalassaemia (α-thal) there is evidence that in most cases the α-globin genes are deleted (Ottolenghi *et al.*, 1974; Taylor *et al.*, 1974; Kan *et al.*, 1975). More recent evidence shows, however, that not all α-thals are the result of gene deletions and that in some cases the gene is present but non-functional (Orkin *et al.*, 1979a).

The β-thalassaemias (β-thals) are more complicated but can be divided into two general classes: β^+ and β^o-thal. In homozygote patients, the former is characterised by a reduction of and the latter by a complete absence of β-globin chain synthesis. There is general agreement that the reduction in chain synthesis in β^+-thal is due to a corresponding reduction in β-

globin mRNA sysnthesis (Benz and Forget, 1971; Nienhuis and Anderson, 1971; Housman *et al*., 1973; Kacian *et al*., 1973). The reduction in β-globin mRNA levels has not, as yet, been explained but is has been suggested that there may be a defect in hnRNA processing (Nienhuis *et al*., 1977).

The β^{o}-thals are heterogeneous in nature at the molecular level. Although characterised by the total absence of globin synthesis they show low variable levels of mRNA or total absence of mRNA in the cytoplasm. In some cases where the mRNA is present, it is shortened at the 3' end whereas in other cases it is full length but inactive (Benz *et al*., 1978; Old *et al*., 1978).

The β-globin gene has been shown to be present, by cDNA hybridisation analyses, in β^{o}-thals in the majority of cases studied (Tolstoshev *et al*., 1976).

Restriction endonuclease mapping has recently shown however that in three patients there is a deletion towards the 3' end of the β-globin gene (Flavell *et al*., 1979; Orkin *et al*., 1979b).

Comi *et al*. (1977) has shown that for one Italian case of β^{o}-thal normal levels of β-globin hnRNA are found but no cytoplasmic β-globin mRNA. This appears to be an example of a β^{o}-thal caused by a processing or transport defect.

The form of β^{o}-thal from the Ferrara region of Italy is of special interest. In some patients the β-globin mRNA is apparently intact but in reticulocyte lysates it does not act as a template for globin synthesis. Addition of cytoplasmic preparations from normal reticulocytes induces β-globin synthesis (Conconi *et al*., 1972). The addition of supernates from the reticulocytes of patients homozygous for $\delta\beta^{o}$-thal or

HPFH (hereditary persistance of foetal haemoglobin) does not induce synthesis. This suggests that some factor required for the expression of β-globin mRNA is coded for in the region common to the deletions in $\delta\beta^{o}$-thal and HPFH.

The $\delta\beta^{o}$-thals and HPFH may be considered together. Both are the result of deletions including at least portions of the δ and β-globin and, in some cases, the γ-globin genes. The deletion of the δ and β-globin genes in $^{G}\gamma^{A}\gamma$-$\delta\beta^{o}$-thal and in $^{G}\gamma^{A}\gamma$-HPFH has been demonstrated using cDNA hybridisation analysis (Kan _et al_., 1975b; Forget _et al_., 1976; Ottolenghi _et al_., 1976; Ramirez _et al_., 1976b; Orkin _et al_., 1978). The clinical severity of $\delta\beta^{o}$-thal and HPFH are markedly different. In HPFH the lack of δ and β-globin synthesis is compensated by synthesis of foetal γ-globin in adults, although chain synthesis is still somewhat imbalanced. The expression of the γ-globin genes is, in most cases, pancellular (Huisman _et al_., 1974, Wood _et al_., 1977a). Heterocellular HPFH is a type where foetal globin synthesis occurs only in a sub-population of cells and this may provide some insight into the mechanisms of the switch from γ to β-globin synthesis.

In $\delta\beta^{o}$-thal there is much less compensation by foetal globin synthesis for the loss of δ and β-globin syntheses. Since the γ and α-globin synthesis are highly imbalanced the result is a thalassaemic syndrome. It appears that the differences between the levels of expression of the γ globin chains in HPFH and $\delta\beta^{o}$-thal may be related to the different extents of the gene deletions. This provides a model for the study of control regions.

The Hb Lepores are clinically similar to the thalassaemias. The δβ-globin fusion product is synthesised very inefficiently

and this is shown to be associated with an unstable mRNA. This is also found to be the case in the anti-Lepore HbMiyada (Roberts et al., 1973). The δ globin mRNA is also unstable and so the instability of δβ and βδ hybrid mRNAs may be a result of having at least a portion of the δ-mRNA. In contrast, the mRNA coding for the γβ-globin fusion product in Hb Kenya has a normal stability compared with β and γ-globin mRNAs (Wood et al., 1976b).

More complete reviews of the molecular genetics of haemoglobin and the thalassaemias are provided by Weatherall and Clegg (1979a, b).

1.5 Recombinant DNA

(a) The basic techniques

In 1967 it was demonstrated that DNA ligases could join DNA molecules end to end. The next major discovery was that of Mandel and Higa (1970) who showed that phage λDNA could be introduced at high efficiency into E.coli that had been treated with calcium salts. This finding was exploited by Cohen et al. (1972) who applied the technique to plasmid DNA.

Phage λ is a bacteriophage of E.coli. The genetics of λ and the interaction between λ and different host strains of E.coli were well studied. Plasmids are a heterogeneous collection of circular DNA species that confer certain properties, including drug resistances, on bacteria that harbour them. The findings that both λ and plasmid DNAs could be introduced into E.coli paved the way for their use as cloning vectors.

It was demonstrated that the joining of two different plasmids resulted in a single recombinant plasmid that shared the antibiotic resistance properties of the two parental types (Cohen et al., 1973). A means whereby foreign DNA could be propagated in a bacterial plasmid was thus demonstrated.

In the early 1950s variability in the resistance of bacteria to viruses was described. Later studies on E.coli strains B and K led to the proposal of a model of restriction and modification to account for the observations (Arber and Dussoix, 1962; Dussoix and Arber, 1962). A restriction and modification system consists of an enzyme, coded for by the host bacterium, that will degrade DNA. The bacterium protects its own DNA from degradation by methylating it in a specific manner (modification). Foreign DNA entering the bacterium is degraded since it will not be protected by the specific

methylation.

This led in turn to the discovery of class II restriction endonucleases (restriction enzymes) which cleave specific DNA sequences (Smith and Wilcox, 1970; Kelly and Smith, 1970). The use of restriction enzymes by Danna and Nathans (1971) to fragment SV40 DNA led to the discovery of other restriction enzymes and refinements in their use.

The use of restriction enzymes to cut DNA may be followed by the joining of the fragments using DNA ligase. Using this strategy, with the enzyme Eco RI, Cohen et al. (1973) constructed their recombinant plasmid. Restriction enzymes are reviewed by Roberts (1976). Another approach to the joining of DNA molecules was adopted by other groups of workers (Jackson et al., 1972; Lobban and Kaiser, 1973). Complementary homopolymer deoxynucleotide tails were added to the 3' termini of the different DNA molecules to be joined using calf thymus terminal transferase. The single stranded complementary tails were allowed to anneal and so a circular recombinant molecule wa s formed. Only circular molecules can transform the bacterial host so all transformants contain recombinant plasmids. Molecules unable to form circles do not transform.

Later still a third method was to be developed for the joining of DNA molecules. This method employs synthetic linkers which encode a restriction endonuclease recognition site (Bahl et al., 1976). The linkers may be ligated to DNA molecules then activated by digestion with the appropriate enzyme. The DNA molecule now has cohesive termini and may be inserted into the appropriate site in a cloning vector.

(b) Vectors

Although the transformation of E.coli was demonstrated with phage λ DNA before plasmid DNA the usefulness of λ as a cloning vector was not exploited initially. Most early work was carried out using plasmid vectors.

The most popular early plasmid used for genetic recombinant preparation was pSC101 (Cohen et al., 1973). This plasmid has a useful selection of unique restriction enzyme sites and confers tetracycline resistance on bacteria which are transformed by it. It has the disadvantages that it exists in only a few copies per bacterial cell and that this copy number cannot be increased by treating the cells with chloramphenicol (Clewell, 1972; Hershfield et al., 1974). Chloramphenicol allows the plasmid DNA, but not the bacterial DNA, to replicate so increasing the yield of plasmid. Advantage was taken of this phenomenon with plasmids that will undergo amplification such as Col EI which codes both for colicin EI and resistance to it. Derivatives of Col EI were constructed to provide more versatile vectors: pCR1 (Covey et al., 1976), pMB9 (Rodriguez et al., 1976) and pBR322 (Bolivar et al., 1977).

Three groups of workers independently constructed useful cloning vectors from phage λ by manipulation of restriction enzyme sites (Murray and Murray, 1974; Rambach and Tiollais, 1974; Thomas et al., 1974). These vectors involve the replacement of non-essential regions of the phage DNA by the DNA to be cloned. This places limits on the size of fragment that may be cloned as the recombinant molecule has to be approximately the same size as wild type λ. From the original vectors newer and more versatile derivatives have been constructed. From the vector of Thomas et al., (1974) the λgtWES.λB cloning system was

developed by Leder _et al_. (1977). Other λ cloning systems have been developed by Blattner _et al_. (1977) and Murray _et al_. (1977). All these newer systems incorporate biological safeguards such as deletions in non-essential genes and amber mutations.

Methods for the identification of specific recombinants _in situ_ on sheets of nitrocellulose were developed. These methods involve the growing of bacterial colonies on filters or the blotting of phage DNA from plaques. Replicas of colonies or plaques are easily obtained and the DNA in them may be fixed directly to the filter and this is then hybridised to a specific radioactive probe so that the required clone may be identified. The methods for bacteria and phage were developed by Grunstein and Hogness (1975) and Benton and Davies (1977) respectively.

Aspects of the methodologies and vectors are reviewed by Malcolm (1979), Brammer (1979) and Sherratt (1979).

(c) _Safety_

The first suggestion that genetic recombination experiments might pose new and novel dangers was made in a letter to the journal Science (Singer and Soll, 1973) following the Gordon Conference on Nucleic Acids in June 1973. The case was made for the setting up of a committee to make recommendations on the control of further work. The committee recommended that the National Institutes of Health should draw up guidelines for the work and establish an advisory committee and this was done in October 1974. At the Asilomar conference in February of the following year safety precautions were discussed and the bases of procedures for the safe handling of recombinant DNA were set down.

Meanwhile, in Britain, the Ashby committee published its report on the potential benefits and hazards of genetic engineering (Ashby, 1975). They recommended the setting up of a body to "draft a code of practice and to make recommendations on the establishment of a central advisory service."

The Williams working party was set up in August 1975 and reported the following year (Williams, 1976). This report has been the basis of safe working in all recombinant DNA experiments in this country. The Genetic Manipulation Advisory Group (GMAG) was set up to advise on the physical and biological containment appropriate for experiments notified to them. Four different physical containment categories were established and advice is given on the use of certain approved combinations of hosts and vectors.

In August 1978 Britain became the first country to introduce legislation requiring the notification of all genetic recombination experiments. This notification is to GMAG and to the Health and Safety Executive.

1.6 Objectives

The human globin genes provide an elegant system for the study of the sequential expression of genes under developmental control. The thalassaemias provide models for the study of the control mechanisms involved.

The realisation that DNA-DNA and RNA-DNA hybridisation could be used analytically for the study of single human genes coupled with advances in the *in vitro* manipulation of DNA has led to the ability to clone DNA sequences at will.

It was decided to, first, clone human globin mRNA derived DNA sequences in bacterial plasmids and then to use these cloned sequences as hybridisation probes to isolate a corresponding cloned genomic DNA sequence. The genomic sequence to be identified was that for the γ-globin genes.

Chapter 2

Materials and Methods

2.1 Materials

All chemicals were analar grade and supplied by B.D.H., Poole, England unless otherwise stated.

Isotopes

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

Buffers

Trizma base (Tris), HEPES and sodium cacodylate were obtained from Sigma Chemical Co., St. Louis, Missouri, U.S.A.

Chromatography

Oligo (dT) cellulose (grade T3) was obtained from Collaborative Research Inc., Waltham, Massachusetts, U.S.A. RPC-5 (Lot 6) was obtained from Miles Laboratories Inc., Kankakee, Illinois, U.S.A. Sephadex was obtained from Pharmacia, Upsalla, Sweden. Bio-Gel was obtained from Bio-Rad Laboratories, Richmond, California, U.S.A. DEAE cellulose (DE 52) was obtained from Whatman, Maidstone, England.

Enzymes

DNA polymerase I and DNase I were obtained from Boehringer, Mannheim, West Germany. TdT was a gift from H. Leibscher (Academy of Sciences, G.D.R.). DNA ligase, Eco RI and Hind III were obtained from Miles Laboratories. Lysozyme and nuclease S_1 were obtained from Sigma Chemical Co. RNases T_1 and T_2 were obtained from Calbiochem, San Diego, California, U.S.A. Bgl II and Bam HI were obtained from New England Biolabs, Beverly, Massachusetts,

U.S.A. Reverse transcriptase was a gift from Dr. J. Beard, through the Biological Carcinogenesis Branch, Viral Oncology Program, National Cancer Institute. T4 polynucleotide kinase and bacterial alkaline phosphatase were obtained from P-L Biochemicals, Milwaukee, Wisconsin, U.S.A.

Nucleic Acids

Phage λ DNA was obtained from Miles Laboratories. Salmon sperm DNA and deoxynucleotides were obtained from Sigma Chemical Co. Oligo (dT) was obtained from Collaborative Research Inc.

Electrophoresis

Agarose was obtained from Miles Laboratories. Ethidium bromide and Orange G were obtained from Sigma Chemical Co.

Photography and autoradiography

FP4 film, Phenisol and PQ Universal developers and Hypam fixer were obtained from Ilford, Basildon, England. Mach 2 intensifying screens and RX Medical X-ray film were obtained from Fuji Photo Co. Ltd., Tokyo, Japan.

Scintillation Counting

PPO and POPOP were obtained from Fisons, Loughborough, England. Triton X-100 was obtained from Koch-Light Laboratories Ltd., Colnbrook, England.

Filters

GF/A, GF/C, No. 1 and No. 17 were obtained from Whatman. Nitrocellulose filters were obtained from Sartorius, Gottingen, West Germany.

Transfer Hybridisations

Polyvinylpyrollidone and ficoll were obtained from Sigma Chemical Co.

Electron Microscopy

Copper grids were obtained from Polaron Equipment Ltd., Watford, England. Ammonium acetate and cytochrome C were obtained from Sigma Chemical Co. Parlodion was a gift from Dr. P. Rigby.

Bacterial Culture

Yeast extract, tryptone and brain heart infusion were obtained from Oxoid Ltd., Basingstoke, England. Agar was obtained from Difco Laboratories, Detroit, Michigan, U.S.A. Kanamycin, streptomycin, chloramphenicol, ampicillin, nalidixic acid, cycloserine and diaminopimelic acid were obtained from Sigma Chemical Co. Tetracycline was obtained from Lederle Laboratories, Gosport, England.

Miscellaneous

Heparin was obtained from Evans Medical, Liverpool, England, B.S.A. was obtained from Sigma Chemical Co. Rat liver ribonuclease inhibitor was a gift from G.D. Searle, High Wycombe, England. *E. coli* χ1776 was provided by Dr. R. Curtiss III.

2.2 General Methods

(a) Agarose gel electrophoresis

DNA and RNA samples were analysed by agarose slab gel electrophoresis. The appropriate percentage of agarose was dissolved with heating in E buffer (40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.5) for general use or in T buffer (40 mM Tris, 40 mM sodium acetate, 2 mM EDTA, 0.5 μg ml^{-1} ethidium bromide, pH 7.5) for gels for transfer hybridisations. For RNA samples 0.1% w/v SDS was included in both the gel and electrophoresis buffer. Gels were poured at 60°C, allowed to set, then the sample well formers removed and the tank chambers filled with the appropriate buffer. Samples were loaded onto gels in E buffer plus 2% (w/v) ficoll with a small amount of Orange G as a marker. Electrophoresis was for 5 min. at 20 V initially. Gels for transfer hybridisation were run at 20 - 30 V and others at 100 - 150 V for the required time.

(b) Gel staining and photography

Following electrophoresis gels were stained, if necessary, in E buffer containing 0.5 μg ml^{-1} ethidium bromide for 10 - 15 min. DNA or RNA was visualised with short U.V. light and photographed on Ilford FP4 film using a Polaroid MP-4 camera. Films were developed in PQ Universal developer and fixed in Hypam fixer.

(c) Gel filtration

Gel filtration using Sephadex G-50 or Bio-Gel A - 1.5 m was used routinely to desalt RNA and DNA to remove unreacted substrate following enzymatic syntheses. A pad of Dowex

chelating resin was often included in the column above and below the gel bed. This was to remove heavy metal ions which can reduce the solubility of nucleic acids especially cDNAs (P. Tolstoshev, personal communication). Columns were equilibrated with the appropriate buffer, the sample applied and the column washed with a constant flow of buffer. Fractions were collected and monitored and the excluded peak of nucleic acid pooled.

(d) Ethanol precipitation of nucleic acid

Both RNA and DNA were routinely precipitated by the addition of a $^{1}/10$ volume of 4 M NaCl and 2 volumes of ethanol. DNA eluted from RPC-5 was precipitated directly by the addition of 2 volumes of ethanol.

(e) Restriction enzyme digestions

All digestions of DNA with restriction enzymes were at $37^{o}C$ at a DNA concentration of 200 μg ml^{-1}. Digestions were carried out in the following buffers:-

Hind III: 6 mM Tris, 6 mM $MgCl_2$, 6 mM 2-mercapto ethanol, pH 7.8.

Eco RI : 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris, 0.5 mM dithiothreitol, pH 7.5.

Bam HI : 150 mM NaCl, 6 mM $MgCl_2$, 6 mM Tris, 100 μg ml^{-1} BSA, pH 7.9.

Bgl II : for New England Biolabs enzyme: 6 mM KCl, 10 mM Tris, 10 mM $MgCl_2$, 1 mM dithiothreitol, 200 μg ml^{-1} BSA, pH 7.4.

for enzyme prepared in this study: 50 mM Tris, 50 mM KCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, pH 7.9.

Digests were for 1 hour with 1 unit of enzyme per μg of DNA unless otherwise stated.

(f) Determination of radioactivity

^{3}H radioactivity was measured by liquid scintillation counting. Samples were dried on Whatman, GF/A or GF/C filters then counted in toluene based scintillator (TBS) (0.5% (w/v) 2,5 diphenyloxazole (PPO), 0.03% (w/v) 1,4-di-2-(5 phenyloxazolyl) benzene (POPOP) in toluene) or counted as an aqueous solution 10% (v/v) in TXBS (TBS: Triton X-100 in a ratio of 2:1). Efficiency of counting for ^{3}H was 30 - 35%. ^{32}P radioactivity was measured by Cerenkov radiation and the efficiency of counting was approximately 30%.

2.3 Specific Methods

(a) Preparation of mRNA

Blood Globin mRNA was prepared from umbilical cord blood of newborn infants or from the first 25 ml of blood removed from newborn infants or foetuses undergoing exchange transfusion. Blood was collected into ice cold 1 x SSC (SSC is 0.15 M NaCl, 0.015 M sodium citrate pH 7.0) containing 2.5 units ml^{-1} heparin, filtered through glass wool and centrifuged at 800 x g for 10 min at $4^{o}C$. The plasma and buffy coat were discarded and the remaining cells resuspended in fresh 1 x SSC followed by centrifugation as before. The washing was repeated twice more and the packed cells stored at $-70^{o}C$ until required.

Isolation of total RNA Frozen, packed cells were lysed by thawing in an equal volume of ANE (100 mM NaCl, 10 mM sodium acetate, 1 mM EDTA pH 6.0) + 2% (w/v) SDS and extracted extensively with half volumes of phenol/ chloroform/iso-amyl alcohol (50:50:1) until a clear interface was obtained. Extraction was by shaking the lysate with the extraction mix for 15 min followed by separation of the phases by centrifugation at 15 000 x g for 10 min. Nucleic acid was ethanol precipitated from the aqueous phase and the precipitate recovered by centrifugation at 15 000 x g for 10 min. DNA contamination was removed by treating the total RNA with DNase I. Incubation was for 1 hour at $0^{o}C$ in 60 mM NaCl, 10 mM $MgCl_2$, 50 mM Tris pH 7.0 with 5 μg ml^{-1} DNase I and 100 μg ml^{-1} heparin to inhibit contaminating ribonuclease activity. Total RNA was recovered by ethanol

precipitated following phenol/chloroform/iso-amyl alcohol extraction.

Oligo (dT) cellulose chromatography Globin mRNA was prepared from total RNA by oligo (dT) cellulose affinity chromatography essentially by the method of Aviv and Leder (1972). Oligo (dT) cellulose was suspended in water, washed with 0.1 M NaOH and finally equilibrated in NTSAR bufer (0.5 M NaCl, 10 mM Tris, 0.1% (w/v) N-lauroyl sarcosine pH 7.4). Total RNA samples were loaded in NTSAR buffer and the column extensively washed to remove non-specifically bound RNA. Poly A(+) RNA was eluted in TSAR buffer (NTSAR minus the NaCl), made 0.5 M in NaCl and re-cycled over the column. Two cycles of chromatography yielded mRNA approximately 90% pure.

(b) Preparation of cDNA

^{3}H-labelled globin complementary DNA (cDNA) was prepared using AMV reverse transcriptase on a globin mRNA template with an oligo (dT) primer according to the method of Harrison *et al.* (1972) as modified by Jackson *et al.* (1976). The mRNA template was de-salted on Sephadex G-50 prior to use. The incubation mix contained 500 µM dATP, 500 µM dGTP, 500 µM TTP, 40 µM ^{3}H-dCTP (22 - 27 Ci mmole^{-1}), 50 µg ml^{-1} actinomycin D, 50 µg ml^{-1} globin mRNA, 10 µg ml^{-1} oligo (dT), 10 units ml^{-1} rat liver ribonuclease inhibitor, 50 mM KCl, 5 mM $MgCl_2$, 10 mM dithiothreitol and 30 units ml^{-1} reverse transcriptase. Incubation was at 37°C for 2 hours then the reaction was stopped with the addition of EDTA to 10 mM. *E.coli* RNA was added to 50 µg ml^{-1} as carrier prior

to phenol/chloroform/iso-amyl alcohol extraction then the products were passed over Sephadex G-50 equilibrated with 50 mM NaCl, 5 mM HEPES pH 7.0. Fractions containing cDNA were pooled and treated with 0.3 M NaOH for 30 min at 60°C to hydrolyse the template RNA. Following neutralisation with HCl E.coli RNA was again added to 50 µg ml^{-1} and the cDNA precipitated with ethanol and recovered by centrifugation at 15 000 x g for 10 min. The cDNA was then de-salted on Sephadex G-50 and stored in sterile water at -20°C.

(c) Solution Hybridisation

Hybridisation Plasmid DNA was sheared by sonication to approximately 500 - 700 b.p. in length and hybridised to 0.2 ng of cDNA at increasing ratios of plasmid to cDNA. Hybridisations were in 2 µl of formamide hybridisation buffer (0.5 M NaCl, 25 mM HEPES, 10 mM EDTA, 50% (v/v) deionised formamide, pH 6.8) at 43°C in sealed siliconised glass capillary tubes following boiling for 5 min to denature the plasmid DNA. Incubation was for 60 hours. E.coli RNA (2.5 mg ml^{-1}) was included in the hybridisation as carrier.

Analysis of hybrids by S_1 nuclease Hybridisation reactions were flushed from capillaries with 250 µl of cold nuclease assay buffer (70 mM sodium acetate, 2.5 mM Zn SO_4, 0.14 M NaCl, pH 4.5) plus 14 µg ml^{-1} denatured calf thymus DNA and incubated with 100 µl of S_1 nuclease (30 units in nuclease assay buffer) at 37°C for 1 hour. A 100 µl aliquot was removed and counted to

determine the total radioactivity present. S_1 nuclease resistant material was precipitated from another 200 µl aliquot by adding 50 µl 6 N perchloric acid (PCA) and 50 µl 2 mg ml^{-1} BSA as carrier. PCA soluble radioactivity was measured in 200 µl of the supernatant following centrifugation at 1 500 x g for 15 min at $4^{o}C$ to pellet PCA precipitable material. The percentage of cDNA protected in the form of hybrids was calculated from the ratio of PCA soluble to total radioactivity.

(d) Kinase labelling of RNA

Cleavage Globin mRNA was labelled with ^{32}P using polynucleotide kinase essentially according to the method of Maziels (1976). The mRNA was de-salted then cleaved with NaOH to give smaller fragments. The mRNA was at 50 µg ml^{-1} and was cleaved in 0.1 M NaOH for 15 min at $0^{o}C$. Tris (pH 8.0) was added to 0.1 M and HCl added to give a neutral pH. The cleaved mRNA was ethanol precipitated, recovered, redissolved at 50 µg ml^{-1} in H_2O and stored until required at $-20^{o}C$.

Labelling RNA was labelled in an incubation containing 33 µg ml^{-1} cleaved RNA, 6.6 mM $MgCl_2$, 66 mM Tris, 13.3 mM mercapto ethanol and 50 units ml^{-1} T4 polynucleotide kinase, pH 7.6. Incubation was for 30 min at $37^{o}C$ then 3 volumes of SDS buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, 0.5% (w/v) SDS, pH 7.6) were added before passing over Sephadex G-50 equilibrated in SDS buffer. Labelled RNA was pooled and extracted once with phenol/chloroform/iso-amyl alcohol (50:50:1) after the addition of *E.coli* RNA to 50 µg ml^{-1} as

carrier. The labelled RNA was ethanol precipitated and stored in this form at -20°C until required.

(e) Plasmid DNA preparation

Growth of cultures 25 ml cultures of bacteria harbouring plasmids were grown, shaking, overnight at 37°C in L broth (5 gl^{-1} yeast extract, 10 gl^{-1} tryptone, 5 gl^{-1} NaCl and 1 gl^{-1} glucose, pH 7.2) supplemented with nutrients and antibiotics as follows:-

E.coli HB101	200 µg ml^{-1} streptomycin
E.coli χ1776	100 µg ml^{-1} diaminopimelic acid
	40 µg ml^{-1} thymidine and either
	50 µg ml^{-1} nalidixic acid or
	20 µg ml^{-1} cycloserine

pHγGI was selected with 25 µg ml^{-1} kanamycin and both pBR322 and pAT153 were selected with 50 µg ml^{-1} ampicillin.

2 l flasks containing 500 ml of broth were inoculated with 5 ml of the overnight culture and grown shaking at 37°C to mid log phase. Amplification of plasmids in E.coli HB101 was obtained by the addition of chloramphenicol to 200 µg ml^{-1} and allowing static incubation overnight. For the amplification of plasmids in E.coli χ1776 the concentration of chloramphenicol was reduced to 12.5 µg ml^{-1} and further incubation was for 5 hours only. Cells were killed by the addition of 5 ml of chloroform to each flask and shaking for 5 min. The bacteria were harvested by centrifugation in a Sorvall GSA angle rotor at 8 000 x g for 10 min at 4°C.

Cleared lysate preparation. Bacteria were washed in 25% (w/v) sucrose, 50 mM Tris, pH 8.0, pelleted and resuspended in a small volume of sucrose/tris. Lysis was achieved by treatment with fresh 450 µg ml^{-1} egg white lysozyme in 0.25 M Tris, pH8.0 on ice for 5 min followed by the addition of a 1/3 vol of 0.25 M EDTA pH 8.0 allowing the mix to sit on ice for a further 5 min. A 1/10 volume of lytic mix (2% (w/v) Triton X-100, 50 mM Tris, 60 mM EDTA, pH 8.0) was added and the mix left on ice for a further 20 min. Occasionally incubation at 37°C for 5 min was necessary to complete the lysis. Cell debris was removed by centrifugation at 25 000-30 000 rpm for 1 hour at 4° C in a swinging bucket rotor to give a cleared lysate.

Plasmid isolation. To the cleared lysate was added a 1/25 vol of 0.2 M K_2HPO_4 (pH 7.5) and this was then made 100% (w/v) in Cs Cl and a 1/10 volume of 10 mg ml^{-1} ethidium bromide was added. This gave a density of 1.56 g ml^{-1}. The solution was loaded into polyallomer tubes, overlaid with paraffin oil and centrifuged at 140 000 xg for 40 hours at 10°C in an angle rotor. Plasmid DNA was recovered by piercing the tubes with a 16 g needle sucking out the banded DNA by syringe under UV illumination. Ethidium bromide was removed by passing over Dowex resin (Bio-Rad AG 50W x 4) equilibrated with 10 mM Tris, 1 m EDTA, 200 mM NaCl, pH 8.0. CsCl was removed by dialysis against 10 mM Tris, 1 mM EDTA, pH 7.5 and the plasmid stored at -20°C. Occasionally a second CsCl/ethidium bromide centrifugation was necessary if the resolution of the bands was not adequate after the first centrifugation.

(f) Transfer hybridisation

Electrophoresis and transfer DNA samples for Southern transfer hybridisation analysis were electrophoresed as described in section 2.2 (a). Following electrophoresis, DNA in the gels was denatured by gently shaking the gel immersed in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 3 hours followed by shaking in neutralising solution (3 M NaCl, 0.5 M Tris, pH 5.4) again for 3 hours. DNA was transferred from gels to a single sheet of nitrocellulose by direct blotting in 20 x SSC essentially as described by Southern (1975). The sheet was washed in 2 x SSC for 10 min, following transfer, then baked at 80°C for 2 hours and either stored at 4°C or used immediately for hybridisation.

Pre-treatment and hybridisation All pre-treatments and hybridisations were at 65°C. Filters were immersed in 3 X SSC for ½ hour followed by Denhardt's soln (0.2% (w/v) ficoll, 0.2% (w/v) polyvinylpyrollidone, 0.2% (w/v) BSA) (Denhardt, 1966) in 3 x SSC for 3 hours followed by the same solution supplemented with 0.1% (w/v) SDS, 50 µg ml^{-1} sheared and denatured salmon sperm DNA and 10 µg ml^{-1} polyadenylic acid for 1 hour. Hybridisations were carried out in this latter supplemented Denhardt's solution in 3 x SSC at a probe concentration of 0.05 µg ml^{-1} in sealed perspex boxes for at least 40 hours. The plasmid probe, labelled by nick translation was added to the hybridisation solution following denaturation by boiling for 5 min.

Filter washing and autoradiography Following hybridisation the filters were washed 6 times each for 10 min at 65°C in Denhardt's solution in 3 x SSC supplemented with 0.1% (w/v) SDS and 50 µg ml^{-1} sheared and denatured salmon

sperm DNA. This was followed by two washes each of 30 min in 0.1 x SSC supplemented with DNA and SDS as before. Filters were dried in air, mounted on cardboard, wrapped in "cling film' and autoradiographed at -70°C using Fuji RX medical X-ray film (presensitised by partial fogging) and Fuji Mach 2 fast intensifying screens. X-ray films were developed in Ilford Phenisol developer and fixed in Hypam rapid fixer.

(g) Labelling of plasmid DNA by nick translation

DNA was labelled to high specific activity by the method of nick translation (Rigby *et al*.,1977). DNA at a concentration of 15 μg ml^{-1} was incubated at 15°C for 1½ hours in 120 units ml^{-1} DNA polymerase I, 170 pg ml^{-1} DNase I, 4 μM dATP, 4 μM dGTP, 2.6 μM α-^{32}P-dCTP, 2.6 μM α-^{32}P-TTP, 10 mM 2-mercapto ethanol, 5 mM $MgCl_2$ and 50 mM Tris, pH 7.5. The reaction was stopped by the addition of 2 volumes of stopmix (10 mM Tris, 12.5 mM EDTA, 0.5% (w/v) SDS, pH 7.5) and made 1 mg ml^{-1} in salmon sperm DNA prior to extraction with an equal volume of phenol. Unincorporated nucleotides were removed by passing over Bio-Gel A-1.5 m equilibrated in 3 x SSC. The specific activity was calculated from Cerenkov radiation and the labelled DNA was stored at -20°C until required.

(h) Preparation of human DNA

DNA was prepared from placentae cut into small pieces and homogenised in 0.15 M NaCl, 0.1 M EDTA, pH 10.5. Cells were lysed by the addition of SDS to 1% (w/v) and allowing to sit at room temperature for 15 min. The lysate was extracted once with an equal volume of phenol and twice with equal volumes of chloroform:octan-2-ol (24:1). The DNA was then ethanol precipitated and immediately spooled onto a

glass rod. This selects for high molecular weight DNA leaving low molecular weight DNA and most RNA behind. The DNA was redissolved in 10 mM Tris, 1 mM EDTA, pH 7.5 and treated with 50 μg ml^{-1} RNase T_1 or T_2 for 2 hours at 37°C. The DNA was phenol extracted, ethanol precipitated, recovered and stored in Tris/EDTA at -20°C.

(i) Preparation of Bgl II

Bacillus globigii was grown in brain heart infusion to stationary phase then harvested in a Sorvall GSA rotor at 8 000 x g for 10 min at 4°C. Two volumes of extraction buffer (10 mM Tris, 10 mM 2-mercapto ethanol, pH 7.9) were added and the cells resuspended. The cells were sonicated at setting 9 on a Dawe Sonicator for 10 x 1 min allowing cooling to below 10°C between treatments. Cell debris was removed by centrifugation in an angle rotor at 30 000 rpm for 1½ hours at 4°C. The cleared lysate was made 70% saturated in ammonium sulphate and stirred for 30 min at 4°C. The precipitate was collected by centrifugation at 16 000 x g for 10 min at 4°C, resuspended in Bio-Gel buffer (1 M NaCl, 10 mM Tris, 10 mM mercapto ethanol, pH 7.9) and passed over a column of Bio-Gel A-0.5 m equilibrated in the same buffer. Fractions were collected and assayed for Bgl I or Bgl II activity by incubation with λDNA as described in section 2.2 (e). Fractions active for both Bgl I and Bgl II were pooled and extensively dialysed against DEAE buffer (10 mM sodium phosphate, 10 mM mercapto ethanol, 0.1 mM EDTA, pH 7.4) and loaded onto a column of DEAE cellulose. A gradient of KCl from 0 to 0.5 M in DEAE buffer was applied to the column. Fractions were assayed as before. Fractions exhibiting Bgl I, Bgl II and Bgl I + Bgl II activity were pooled separately. The Bgl II fractions were dialysed agains DEAE buffer containing 50% glycerol. The Bgl I and Bgl I + Bgl II fractions were also retained and diluted with an equal volume of glycerol. All were stored at -20°C.

(j) Gene Machine

Total restricted DNA was fractionated on a Southern Gene Machine. DNA in electrophoresis buffer (40 mM sodium acetate, 10 mM Tris, 1 mM EDTA, pH 7.9) was loaded in 5% (w/v) sucrose with bromophenol blue as a marker into the sample well of a 2% (w/v) agarose Gene Machine gel. Electrophoresis was for 16 hours at 10 V after which the electrophoresis buffer was replaced. Further electrophoresis was at 20 V until all DNA was recovered.

(k) RPC-5 chromatography

The method used is essentially as described by Tilghman *et al*. (1977). RPC-5 powder was suspended in loading buffer (1.25 M sodium acetate, 50 mM Tris, 1 mM EDTA, pH 7.5). The slurry was ground to remove lumps then loaded into a HPLC column and packed at a pressure of 400 psi using a syringe pump. DNA in loading buffer was loaded onto the column using a Milton Roy Mini Pump. The column was washed with loading buffer then the DNA was eluted with a 1.6 to 1.8 M or 1.4 to 1.8 M sodium acetate gradient in 50 mM Tris, 1 mM EDTA, pH 7.5. Fractions were monitored by optical density at 260 nm. DNA was recovered by ethanol precipitation and stored in 10 mM Tris, 1 mM EDTA, pH 7.5 at -20°C.

(l) Tailing of DNA

Homopolymer tails were added to plasmid and human DNA with terminal polynucleotidyl transferase (TdT). DNA was incubated at 37°C at a terminus concentration of 5 nM in 1 mM $CoCl_2$, 0.1 M HEPES pH 7.0 with nucleotide and TdT concentrations and incubation times as described in 3.5 (d) and 3.5 (e). The length of tail was controlled by varying a combination of incubation time, nucleotide concentration and

TdT concentration as described.

Unincorporated nucleotides were removed by passing over Sephadex G-50 equilibrated in 50 mM NaCl, 2.5 mM Tris, pH 7.6. Tailed DNA fractions were pooled, lyophilised and resuspended in a ½ volume of distilled water for annealing. DNA was stored at -20°C until required.

(m) Transformation of *E. coli* χ1776

1 ml of an overnight culture of *E. coli* χ1776 was diluted 100 fold with fresh broth and grown to an optical density of 0.3 at 550 nm as described in section 2.3 (e). The culture was cooled on ice and the bacteria harvested and prepared for transformation according to the method of Norgard *et al.* (1978) as modified by H. Liebscher (personal communication). The bacteria were washed twice in 10 mM NaCl, resuspended in a 1/5 volume of Mn buffer (40 mM sodium acetate, 30 mM $CaCl_2$, 70 mM $MnCl_2$, 10 mM RbCl, pH 5.6) and kept on ice of 20 min. Bacteria were then pelleted resuspended in a 1/30 volume of Mn buffer and kept on ice until required.

2 volumes of bacteria were mixed with 1 volume of DNA and kept on ice for 1 hour. Aliquots were spread on pre-cooled agar plates containing selection antibiotics. Incubation was for 40 hours at 37°C.

(n) Transformation of *E. coli* HB101

The method used is essentially that of Cohen *et al.* (1972). *E. coli* HB101 was grown to an optical density of 0.6 at 600 nm, cooled on ice and harvested at 8 000 x g for 5 min at 4°C then resuspended in the original volume of ice

cold 80 mM $CaCl_2$ and kept on ice for 30 min. The bacteria were pelleted as before then resuspended in a 1/10 volume of 80 mM $CaCl_2$ and used directly for transformation. DNA, in 10 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM Tris, pH 7.5, at a concentration of 0.1 to 1 $\mu g\ ml^{-1}$ was added to 2 volumes of bacteria and kept on ice for 30 min. Incubation was next at 25° C for 15 min then at $37^{o}C$ for 30 min following the addition of 10 volumes of warm L broth. The bacteria were filtered under light vacuum onto nitrocellulose filters pre-sterilised by U.V. light. The filters were placed on L agar plates and incubated at 37° C for 2 hours then transferred to L agar plates containing selection antibiotics and incubated for 12 hours at 37° C. Usually 400 µl of bacteria were used for a 9 cm diameter filter to give a master filter.

(o) Preparation of replica filters for screening

Replica filters were prepared from master filters by direct blotting. The master filters were removed from their agar plates, placed on sterile filter paper and replicas made in duplicate by blotting filters, sterilised by U.V. light, onto the master filters keying the replicas by means of small holes in both the masters and replicas. The replica filters were pre-wetted on agar plates prior to blotting. The replicas were incubated for 5 hours at 37° C on L agar plates containing selection antibiotics and the master filters placed on fresh L agar plates with antibiotics and stored at 4° C. The replica filters were next transferred to L agar plates containing 200 $\mu g\ ml^{-1}$ chloramphenicol and incubated overnight at $37^{o}C$ to amplify the copy number of the plasmids.

(p) Grunstein-Hogness screening

This is a modified form of the technique of Grunstein and Hogness (1975). Bacterial colonies were lysed by placing the filters onto 0.75 ml puddles of 0.5 M NaOH on sheets of cling film for 2-3 min. The filters were dried in a Buchner funnel under light vacuum and the procedure repeated with fresh NaOH. The filters were transferred next to puddles of 1M Tris, pH 7.4 then to 1.5 M NaCl, 0.5 M Tris, pH 7.4 each for 2-3 min then dried as before. The filters were next transferred to 0.3 M NaCl, dried, then baked at 80°C for 2 hours.

Hybridisation of the radioactively labelled probe to the colonies was as for Southern transfer hybridisations described in 2.3 (f) except for the following modifications:

i) Pre-washing of the filters was for 3 hours in 5 x Denhardt's solution in 3 x SSC.

ii) Denatured pCR1 DNA was included at a concentration of 1 to 2 $\mu g\ ml^{-1}$ as competitor.

iii) The final two washes were in 0.2 x SSC rather than 0.1 x SSC.

(q) Ligation of DNA

DNA at a concentration of 2 to 20 $\mu g\ ml^{-1}$ was incubated in ligase buffer (4 mM $MgCl_2$, 1.2 mM EDTA, 4 mM NaCl, 0.5 mM ATP, 10 mM 2-mercapto ethanol, 50 $\mu g\ ml^{-1}$ BSA, 30 mM Tris, pH 8.0) with concentrations of DNA ligase as described in the text. Incubation was at 14° C for times between 15 and 64 hours. Fidelity of ligation was checked by re-digesting the DNA with the appropriate restriction enzyme.

(r) Phosphatase treatment of vector DNA

BamHI digested pAT153 DNA at a concentration of 20 $\mu g\ ml^{-1}$ in 50 mM Tris, 1 mM EDTA, pH 8.0 was treated with 0.02 units μg^{-1} DNA bacterial alkaline phosphatase for 30 min at 37° C. Following the incubation the reaction mix was extracted once with phenol/chloroform (1:1) and extensively extracted with ether. Completeness of phosphatase treatment was assayed by the inability to ligate the plasmid DNA as judged by agarose gel electrophoresis.

(s) Preparation of single colony lysates

Colonies of bacteria were grown, until about 1 mm in diameter, on nitrocellulose filters and the plasmid copy number amplified by incubating the filters overnight, on L agar plates containing 200 $\mu g\ ml^{-1}$ chloramphenicol. Single colonies were picked into 20 µl of 50 mM Tris, 5 mM EDTA, pH 7.6 and suspended by vortexing. 1.5 µl of 10% (w/v) SDS and 50 µl of proteinase K (1mg ml^{-1} in 50 mM Tris, 5 mM EDTA, pH 7.6) were added and the mix incubated at 37° C for 2 hours. The sample was next split in half and to one half was added 50 ng of supercoiled pAT153 DNA. 20 µl of molten 0.8% (w/v) agarose was added to the samples prior to loading onto a 0.8% (w/v) agarose gel in the usual manner. Electrophoresis was at 10 V for approximately 6 hours.

(t) Electron microscopy of DNA

Samples were prepared essentially according to the method of Davis et al. (1971). DNA was prepared at a concentration of 20 to 100 ng ml^{-1} in 0.5 M ammonium acetate, 0.1 mg ml^{-1} cytochrome C and 50 µl of this was placed, as a

bubble, on a piece of parafilm and allowed to stand for 1 to 2 min. A parlodion coated electron microscope grid was touched to the surface of the bubble for 2 sec then to a bubble of 0.25 M ammonium acetate for 5 sec. Grids were stained in uranyl acetate (50 μM uranyl acetate, 50 μM HCl in 90% (v/v) ethanol for 10 sec and allowed to dry in air. Grids were viewed in a Philips EM 300 electron microscope.

Chapter 3

Results

3.1 Preparation of Hybridisation Probes

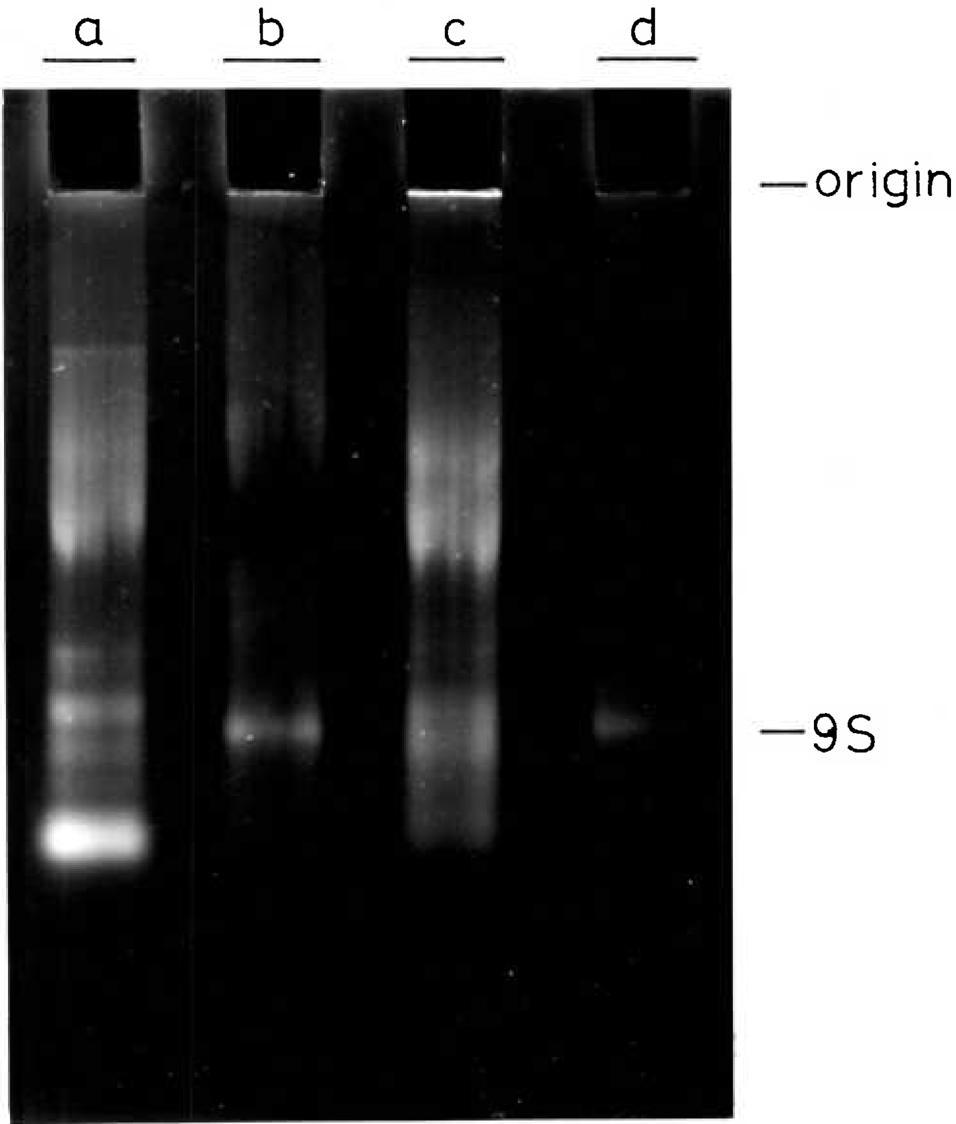
(a) Purification of human foetal globin mRNA

Poly A(+) RNA was prepared as described in 2.3 (a) either from neonatal exchange transfusion or umbilical cord blood (mRNA-αβγ). Figure 1 shows an SDS/agarose gel profile of the RNA following two cycles of purification on oligo (dT) cellulose. The 9S mRNA although substantially purified is slightly contaminated with 18S and 28S rRNA. For the preparation of globin cDNA further purification was not deemed necessary since the likelihood of ribosomal sequences being transcribed into cDNA was low and of little consequence since the cDNA was for screening purposes and not for cloning. From 100 ml of packed red blood cells, consisting of about 20% reticulocytes, 80 μg of mRNA was obtained running as a single band on SDS/agarose gels. The RNA was desalted by passing over Sephadex G-50.

(b) Synthesis of foetal globin cDNA

A.M.V. reverse transcriptase (RNA dependent DNA polymerase) was used to synthesise cDNA on a mRNA template using oligo (dT) as a primer. Since the cDNA was to be used as a hybridisation probe it had to be of the highest possible specific activity. This was achieved by using the single radioactive precursor undiluted in the reaction. About 30 μg of mRNA-αβγ was incubated in the reaction described in 2.3 (b). The reaction products were phenol/chloroform extracted, passed over Sephadex G-50 then treated with alkali to hydrolyse the template mRNA. The yield was calculated by measuring the radioactivity of a small aliquot and was usually of the order of 2 - 3 μg. No size analysis of

Figure 1



the cDNA was carried out but the reaction conditions are known to give full length transcripts as well as shorter ones. (M. Courtney, personal communication). The cDNA prepared had a specific activity of approximately 10^7 c.p.m. μg^{-1}.

(c) Hybridisation of cDNA to template mRNA

The transcription products may be tested for integrity by hybridising the cDNA back to its template mRNA. The hybridisation was carried out in the form of a titration curve where increasing amounts of mRNA are hybridised to a fixed amount of cDNA. The extent of hybridisation was calculated by determining the S_1 nuclease resistant portion of the cDNA. (Young *et al*., 1974). S_1 nuclease from *Aspergillus oryzae* is specific for single stranded nucleic acids and will make acid soluble any radioactive counts from the cDNA not in a mRNA/cDNA hybrid. Each hybridisation was incubated for a time such that a Cot of 0.05 - 0.1 was achieved which is 20 times that required for complete hybridisation at a mRNA/cDNA ratio of 1. Young *et al*. (1974) showed that this is the ratio at which hybridisation proceeds most slowly. An ideal hybridisation curve for mRNA/cDNA would show a very steep slope reaching a maximum at a mRNA/cDNA ratio of 1 with a plateau of 100% for all greater ratios. In practice the slope becomes a curve because not all transcripts are full length so affecting the hybridisation rate and the plateau value reached is only usually 80 - 85% because a fraction of the cDNA is incapable of forming a hybrid. (Harrison *et al*., 1974). Titration curves for the cDNA-αβγ were constructed by others confirming the suitability for use as a hybridisation probe.

(d) ^{32}P labelling of mRNA with polynucleotide kinase

The enzyme polynucleotide kinase isolated from T4 infected E.coli B catalyses the transfer of the γ-phosphate group from ATP to the 5' hydroxyl group of double or single stranded nucleic acid (Richardson, 1965). If the ATP is labelled with ^{32}P in the γ position then the nucleic acid becomes radioactively labelled. The cap on the 5' end of mRNA means that no free 5' hydroxyl exists and so labelling is not possible. Even if the 5' ends of the mRNA were available for labelling the resulting specific activity would be inadequate since only one ^{32}P would be added per mRNA molecule. Prior to labelling the mRNA was subjected to mild alkali hydrolysis to give RNA fragments of approximately 100 bases in length. This allowed the incorporation of a greater amount of ^{32}P per mRNA molecule. The labelling was carried out as described in 2.3 (d). The reaction products were phenol/ chloroform extracted and passed over Sephadex G-50. The labelled mRNA had approximately 3×10^6 c.p.m. μg^{-1}. The labelled mRNA fragments were used as hybridisation probes in the method of Grunstein and Hogness, (1975) for the screening of recombinant DNA clones by P. Little and Ch. Coutelle in this laboratory. Both human mRNA-αβγ and mRNA-αβ were labelled in this manner. The mRNA-αβ was prepared, by other workers, from adults with raised reticulocyte counts.

3.2 Characterisation of pHγG1

(a) Construction of recombinant plasmids containing human α-, β- and γ-globin cDNA fragments

In collaboration with another laboratory, this laboratory constructed recombinant plasmids from double stranded cDNA prepared on a purified template of foetal globin mRNA. The double stranded cDNA was 'tailed' into the plasmid pCR1 and the recombinant DNA used to transform E. coli HB101. Colonies of bacteria were screened by the method of Grunstein and Hogness, (1975) using ^{32}P labelled mRNA probes and a number of colonies were identified as containing either α-, β- or γ-globin cDNA inserts in the plasmids that they carried. Details of the construction and identification are described in the attached paper (Little et al., 1978).

(b) Plasmid DNA preparation

A cleared lysate from a culture of an E. coli clone tentatively identified as containing γ-globin cDNA sequences was prepared by P. Little and Ch. Coutelle and provided for further experiments to confirm the assignation. The lysate was prepared at the category III level of physical containment at Imperial College, London as advised by the Genetic Manipulation Advisory Group (G.M.A.G.) from a clone designated γ49. The bacteria were grown in the presence of 25 μg ml^{-1} kanamycin to select for the plasmid pCR1 which carries a kanamycin resistance determinant. The double stranded cDNA had been inserted into the plasmid at a point remote from the determinant hence selection was possible using kanamycin. When the bacteria had reached mid-log phase chloramphenicol was added to 200 μg ml^{-1} to amplify the copy number of the plasmid per bacterium. (Clewell

et al., 1972). The cells were killed by the addition of chloroform (Weissmann and Boll, 1977) and lysed as described in 2.3 (e). Cell debris was removed by centrifugation and plasmid DNA prepared from the cleared lysate by CsCl/ethidium bromide buoyant density centrifugation (Radloff et al., 1967). The basis of this purification technique is that plasmid DNA exists in the bacterium in a supercoiled form which binds less ethidium bromide than the bulk E.coli DNA which is not supercoiled as isolated. The binding of ethidium bromide results in a decrease in buoyant density and as a result the plasmid, which has bound less ethidium bromide, is separated from other DNA in the CsCl buoyant density gradient. It is possible to prepare between 0.5 and 1 mg of plasmid DNA from 2 l of culture though yields are variable and sometimes much lower than this. Following removal of the ethidium bromide and CsCl the plasmid was usually slightly degraded. Figure 2 shows a 0.8% agarose gel of plasmid prepared by this method. The upper and lower prominent bands in track (a) represent open-circular and supercoiled forms of the plasmid respectively. The faint band between them represents the linear form of the plasmid.

(c) Verification of plasmid identification by solution hybridisation

A clone of bacteria had been tentatively identified as containing γ-globin cDNA sequences. It was decided to verify this by hybridisation of plasmid DNA in solution to cDNA-αβγ and cDNA-αβ. If γ-globin cDNA sequences were present in the plasmid DNA these would protect the γ component of cDNA-αβγ but none of cDNA-αβ in the S_1 nuclease assay. γ49 plasmid DNA was sheared by sonication to a size of approximately 500-700

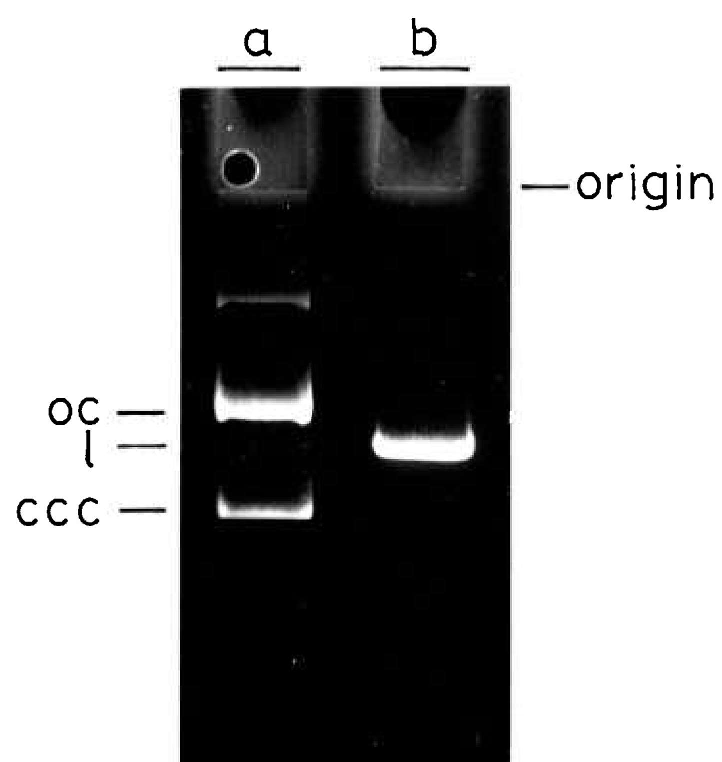
Figure 2.

0.8% agarose gel electrophoresis of plasmid pHγGl DNA.

Track a: 0.5 µg of pHγGl DNA as isolated by CsCl/ ethidium bromide buoyant density centrifugation. The covalently closed circular, open circular and linear forms are marked ccc, oc and l respectively.

Track b: 0.5 µg of Eco RI digested pHγGl DNA.

Figure 2

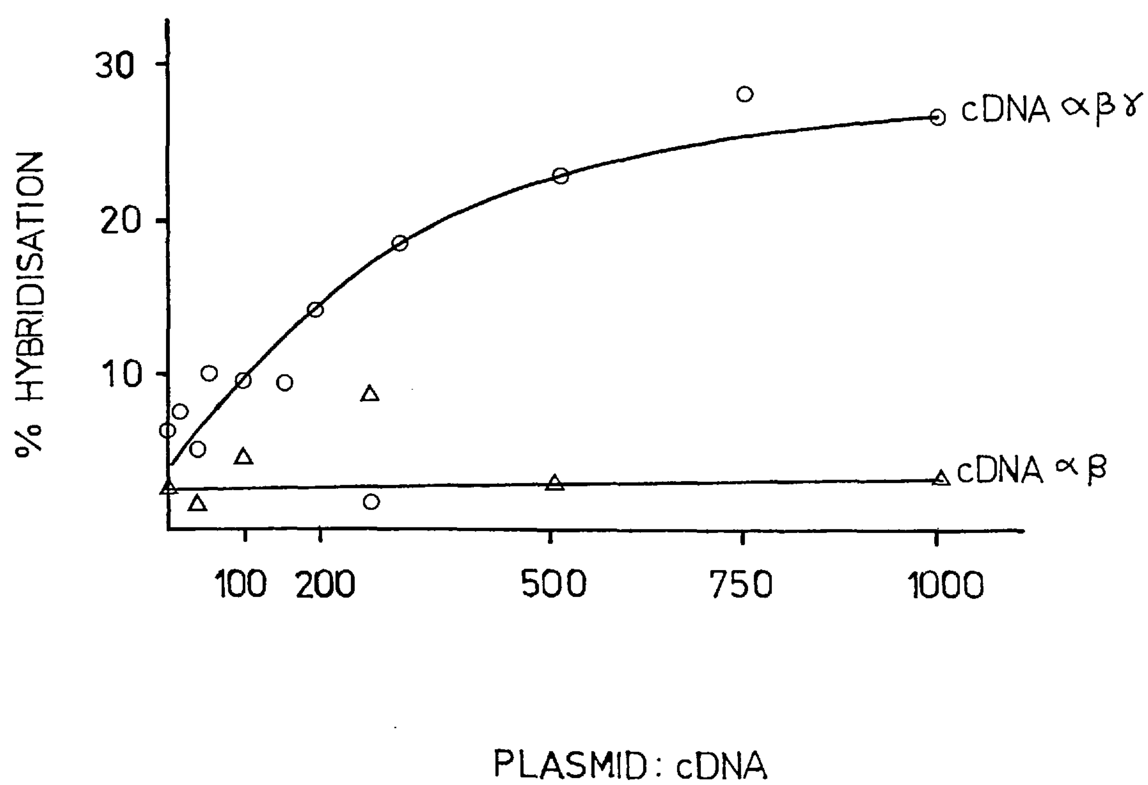


base pairs and hybridised at increasing concentrations to 0.2 ng of cDNA-αβγ or cDNA-αβ. Hybridisation was in formamide hybridisation buffer as described in 2.3 (c). The percentage of hybridisation was measured by S_1 nuclease assay as previously described and the results are shown in figure 3. As expected γ49 DNA hybridises to cDNA-αβγ but not to cDNA-αβ. The background level of hybridisation of γ49 DNA to cDNA-αβ is explained by the fact that about 5% of cDNA preparations is S_1 nuclease resistant. Hybridisation of γ49 DNA to cDNA-αβγ gives a plateau value of around 25%. 32% of mRNA-αβγ is γ-specific (Old *et al.*, 1976) and so 32% of cDNA-αβγ would be expected to be γ-specific. In any preparation of cDNA hybridised fully to its template mRNA there is a 15-20% S_1 nuclease sensitive component. Taking both these facts into account the plateau value of 25% hybridisation is that which would be expected and shows that the inserted γ-cDNA sequence is near to full length. The presence of γ-cDNA sequences in clone γ49 was thus confirmed. Further confirmation was later obtained by the sequencing of a small part of γ-cDNA sequence (Little *et al.*, 1978) and comparing this with the known amino acid sequence for γ-globin (Dayhoff, 1972). The clone γ49 was redesignated pHγGl.

Figure 3.

Hybridisation of human globin cDNA-αβ and cDNA-αβγ to plasmid pHγG1 DNA - titration curve.
0.2 ng aliquots of ^{3}H-cDNA-αβγ or ^{3}H-cDNA-αβ were hybridised to completion with varying amounts of pHγG1 DNA in the presence of carrier E.coli ribosomal RNA. Hybridisations were assayed by digestion with S_1 nuclease as described in the text.

Figure 3



3.3 Mapping of the γ-Globin Genes

(a) Identification of a DNA fragment encoding both γ globin genes

The technique of transfer hybridisation of Southern (1975) allows a physical map to be determined for any number of restriction enzyme sites, within the resolution of the technique, around and within a sequence for which a hybridisation probe is available. The derivation of the physical map of the δβ-globin gene locus is described in detail in the work of Flavell et al. (1978). The physical map around the γ-globin gene locus has also been reported (Little et al., 1979). Recently the linkage of γ, δ and β genes has been formally established (Fritsch et al., 1979; Bernards et al., 1979). In a Bgl II digest of human DNA a single fragment 13 kb in length is found to hybridise to the γ-specific probe pHγG1.

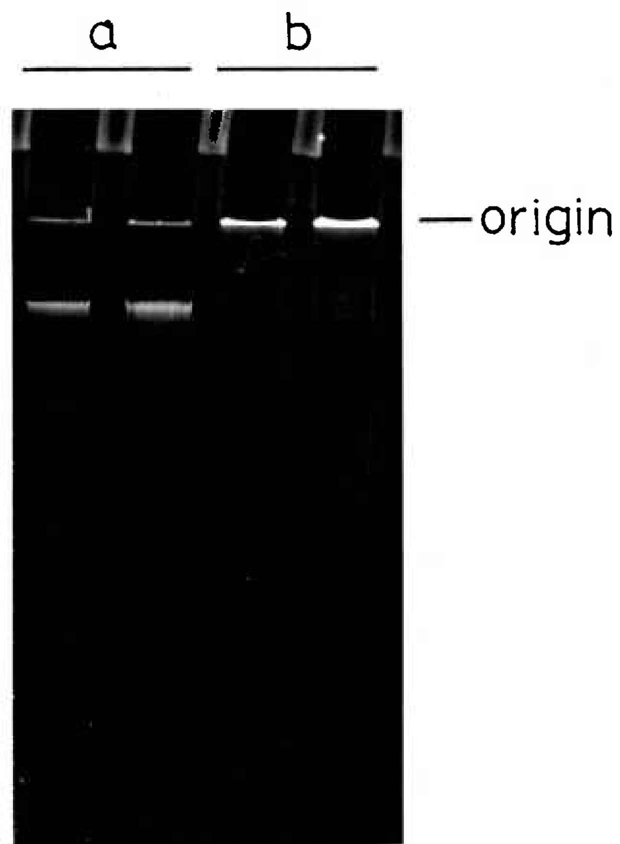
(b) Preparation of human DNA

In order to verify the last fact and to carry out further investigations including the cloning of the 13 kb fragment a large amount of DNA was required. It was thought preferable that the DNA be from a single individual because of the possibility of restriction enzyme site polymorphisms. Placentae are a good source of large amounts of DNA. The placenta of a new-born female infant of Welsh parentage was obtained from St. Mary's Hospital soon after delivery and DNA prepared immediately. The placental tissue is fibrous and it was found to be easiest to prepare DNA from the maternal side of the organ. The DNA prepared as described in 2.3 (h) was of high molecular weight as judged by electrophoresis in a 0.8% agarose gel. Figure 4 shows samples of the isolated DNA run native

Figure 4.

0.8% agarose gel electrophoresis of human DNA.
Tracks a: 0.1 and 0.2 μg of native human DNA.
Tracks b: 0.5 and 1.0 μg human DNA denatured with 0.14M NaOH prior to electrophoresis.

Figure 4



and denatured. The denatured DNA does not run at a markedly different mobility indicating that it has few single stranded nicks. DNA prepared from placentae that are not fresh is often degraded and on electrophoresis shows bands corresponding to nucleosome repeats. (P. Little, personal communication).

(c) Digestion of human DNA with Bgl II

Approximately 2 mg of human DNA was digested to completion with Bgl II. The completeness of digestion was monitored by removing a small aliquot from the main incubation and adding λDNA to a final concentration of 40 μg ml^{-1}. The λDNA is digested along with the human DNA and when an aliquot from the monitoring reaction is electrophoresed in an agarose gel the λ/Bgl II digestion pattern is seen superimposed on that of the digested human DNA. The band pattern for a complete digestion of λDNA by Bgl II is known and if this pattern is found superimposed on the human DNA digestion then the human DNA can be assumed to be digested to completion since the monitoring digest contained 20% more DNA. This is shown in Figure 5. The digestion was carried out as described in 2.2 (e) with an incubation time in 100% excess of that calculated to give complete digestion. A further aliquot was removed from the main incubation and to it was added 14 fold more enzyme and digestion continued for a further ½ hour. This represented approximately a further 2½ fold over digestion. The pattern of digestion of the human DNA remained unchanged as shown in Figure 5. The reaction in the main incubation, which had been stored on ice, was stopped by the addition of EDTA and the reaction products phenol/chloroform extracted then ethanol precipitated and finally re-dissolved for further use.

Figure 5.

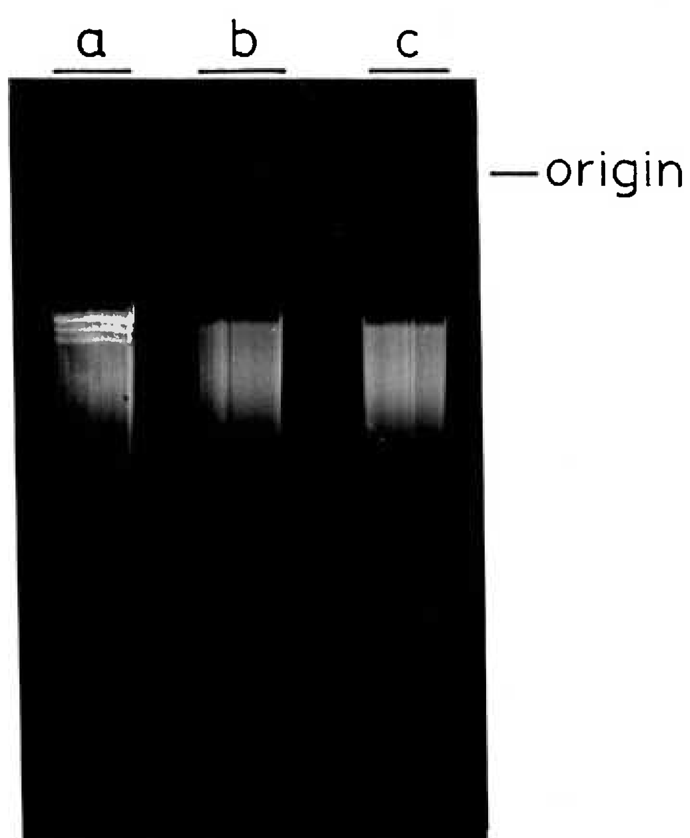
0.8% agarose gel electrophoresis of Bgl II digested human DNA.

Track a: Checking digest containing 3 μg of human DNA and 0.6 μg of λ DNA. The pattern of bands superimposed on the background smear of DNA is that of λ DNA digested with Bgl II.

Track b: 3 μg of Bgl II digested human DNA from the main digestion incubation.

Track c: As track b but with DNA digested for a further 30 min with 14 fold more enzyme.

Figure 5



(d) Transfer hybridisation of Bgl II digested human DNA

The Bgl II digested DNA was run in a 0.8% agarose gel in T buffer as described in 2.2 (a). The higher ionic strength of this buffer as compared to E buffer reputedly gives a better separation of DNA fragments. Following electrophoresis the DNA in the gel is denatured by immersing the gel in alkali followed by immersion in a neutralising bath. DNA is transferred to a nitrocellulose filter sheet by direct blotting from the gel. Figure 6 shows a schematic representation. When pHγGl DNA, labelled with ^{32}P by the method of nick translation, is hybridised in solution to the DNA bound to the nitrocellulose a single band is detected on the nitrocellulose by autoradiography as shown in Figure 7. The single DNA fragment containing all γ-globin DNA sequences was thus confirmed.

The nick translation method of labelling DNA to high specific activity is described in 2.3 (g). DNA is incubated in the presence of DNA polymerase I from E.coli, DNase I and the deoxynucleotides dCTP, TTP, dGTP and dATP. The DNase I nicks the DNA to be labelled and the DNA polymerase I, which has an associated 5' → 3' exonuclease activity, removes a number of nucleotides beyond the nick which are then replaced with nucleotides labelled with ^{32}P in the α position in this case.

The detection of radioactivity by autoradiography of the hybridised nitrocellulose sheet was enhanced with the use of fast intensifying screens at $-70^{o}C$ as described in Lasky and Mills, (1977).

(e) Verification of a single 13 kb γ specific DNA fragment by RPC-5 chromatography

The possibility of two different co-migrating γ-specific

Figure 6.

Schematic representation of a Southern transfer. The paper towels draw 20 x SSC through the gel and the DNA is carried from the gel to the sheet of nitrocellulose. Transfers are carried out at $4^{o}C$.

Figure 6

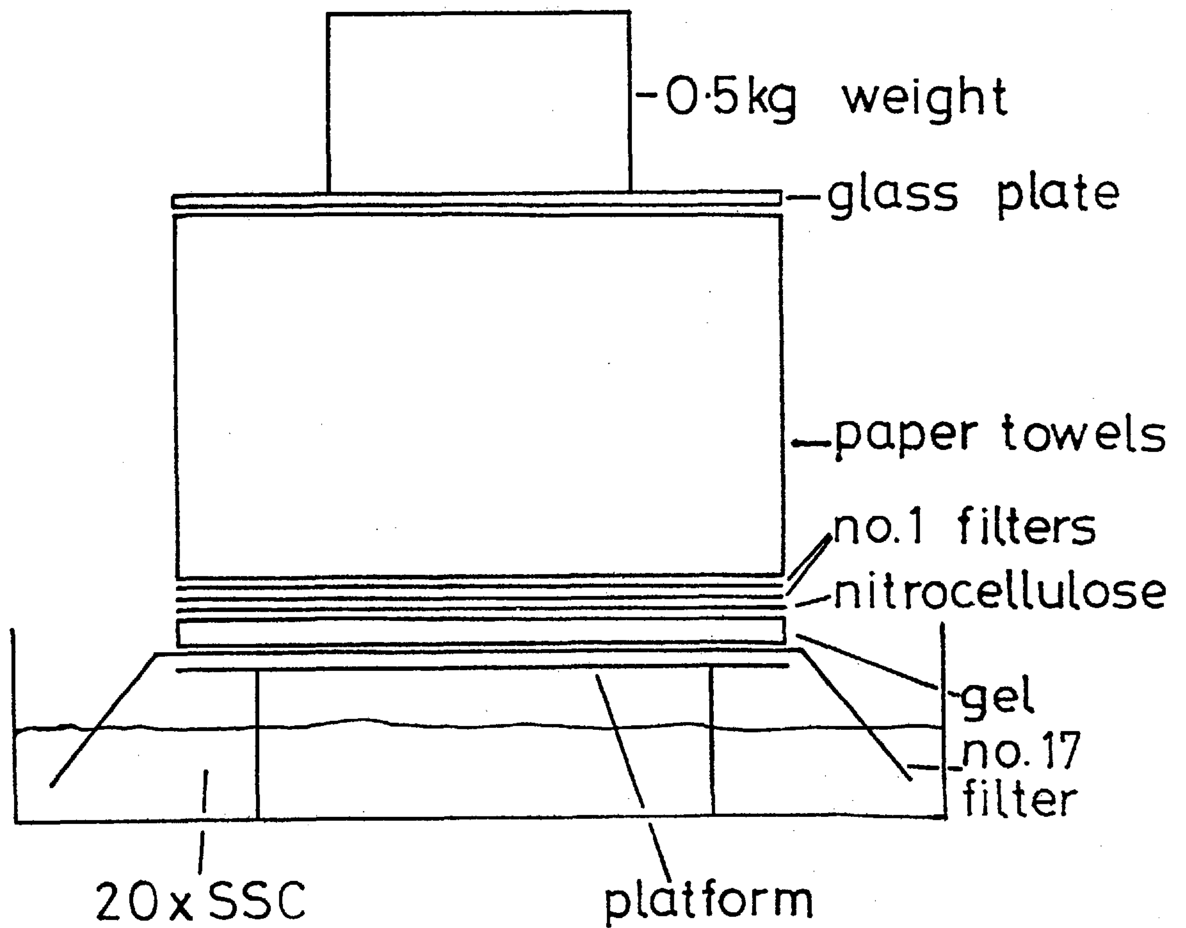
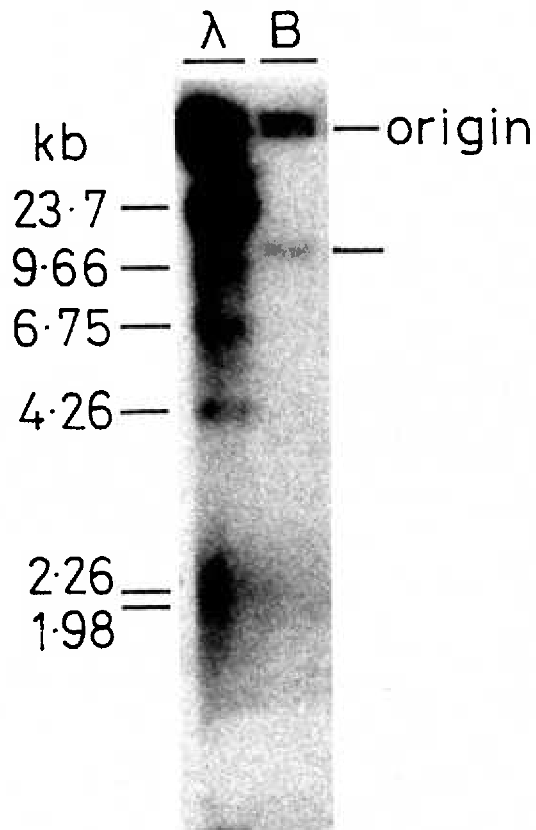


Figure 7.

Autoradiogram of Southern transfer hybridisation of Bgl II digested human DNA.
20 μg of Bgl II digested human DNA (track B) and 0.6 μg of Hind III digested λ DNA (track λ) were electrophoresed in a 0.8% agarose gel. DNA was transferred to nitrocellulose and hybridised to ^{32}P labelled pHγGl DNA. Hybridisation was detected by autoradiography. A bar marks the position of the γ specific Bgl II DNA fragment. Under the conditions used, pHγGl DNA cross hybridises with the λ DNA fragments. The size of the λ/Hind III bands is given in kb.

Figure 7



DNA fragments is not formally excluded in the last experiment. The resolution of agarose gels of the percentage used is not sufficient to separate two DNA fragments of a similar size in this size range. There was also the possibility that each of the γ gene sequences resided in separate 13 kb DNA fragments. These possibilities were excluded by the use of reversed phase chromatography on RPC-5. RPC-5 has the property of separating DNA fragments of the same size according to other parameters (Hardies and Wells, 1976). These include G+C content and the distribution of bases within the fragments of DNA. RPC-5 was developed as a medium for the separation of tRNAs and consists of polychlorotrifluoro-ethylene resin (Plaskon CTFE 2300) coated with a trialkylmethylammonium chloride (Adogen 464) (Pearson _et al._, 1971). RPC-5 is suspended in DNA loading buffer and packed under pressure in a high pressure chromatography column. DNA is applied to the column in loading buffer (1.25 M sodium acetate) with a Milton Roy Mini-Pump. All DNA binds to the column under these conditions. Although chemically inert the base resin appears to play some role in the binding of the DNA since the substitution of the resin by a similar type can alter the binding characteristics. DNA is eluted with a continuous gradient of sodium acetate and the output from the column monitored by optical density at 260 nm. 1 mg of Bgl II digested DNA was fractionated on an RPC-5 column of 5 ml bed volume using a 400 ml gradient from 1.4 to 1.8 M. The elution profile from the column is shown in Figure 8. DNA eluted from the RPC-5 was run on a 0.8% agarose gel for transfer hybridisation. Figure 9a shows the gel profile of the DNA. This confirms that RPC-5 fractionates partly on size (Hardies and Wells, 1976). The transfer hybridisation was carried out as previously described

Figure 8.

Elution profile of Bgl II digested human DNA from RPC-5.

1 mg of Bgl II digested human DNA was eluted from a 5 ml bed volume RPC-5 column with a sodium acetate gradient from 1.4 to 1.8 M. Elution was monitored as transmission at 260 nm.

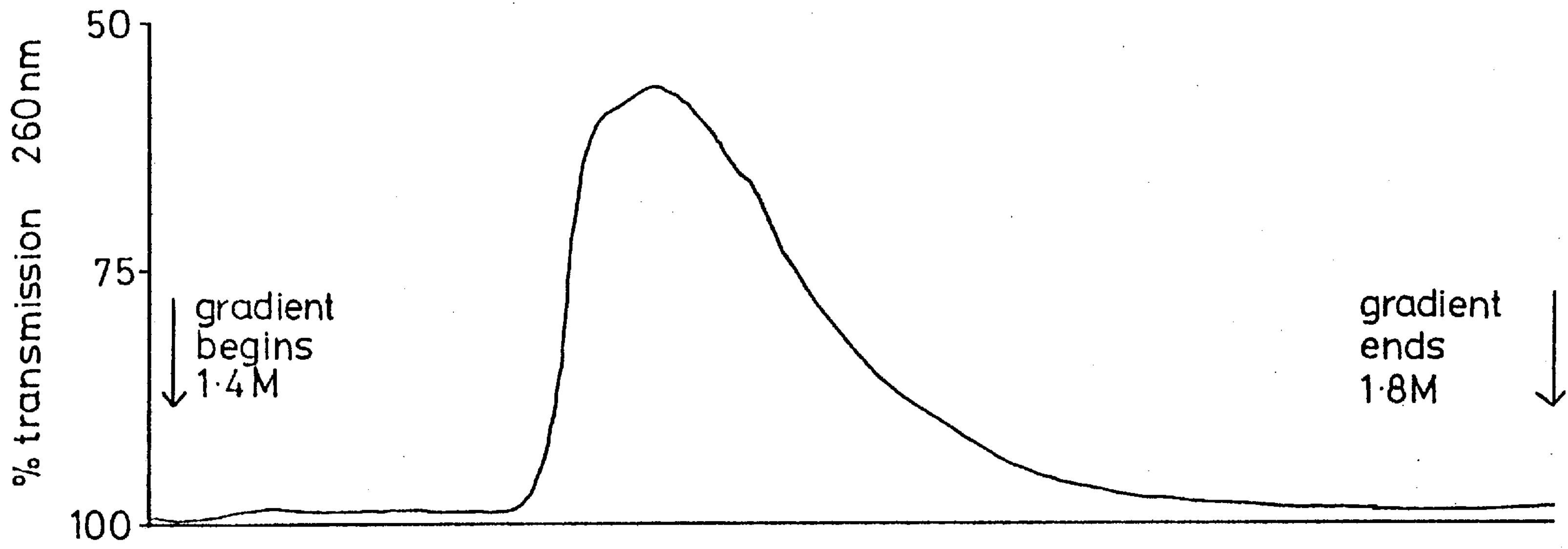


Figure 8

Figure 9a.

0.8% agarose gel electrophoresis of RPC-5 fractionated Bgl II digested human DNA. Fractions are loaded left to right as eluted from the column and are shown under the bars (the blank tracks do not contain DNA). λ and λgt WES.λB DNAs digested with Eco RI are shown as size markers in tracks λ and W respectively. Unfractionated Bgl II digested human DNA is in tracks T.

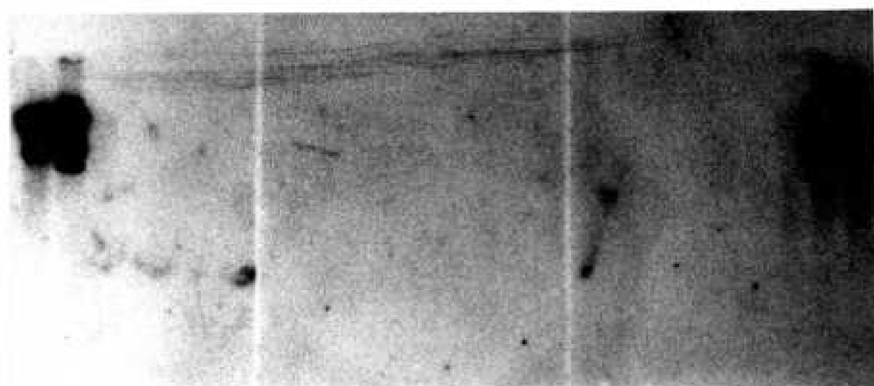
Figure 9a

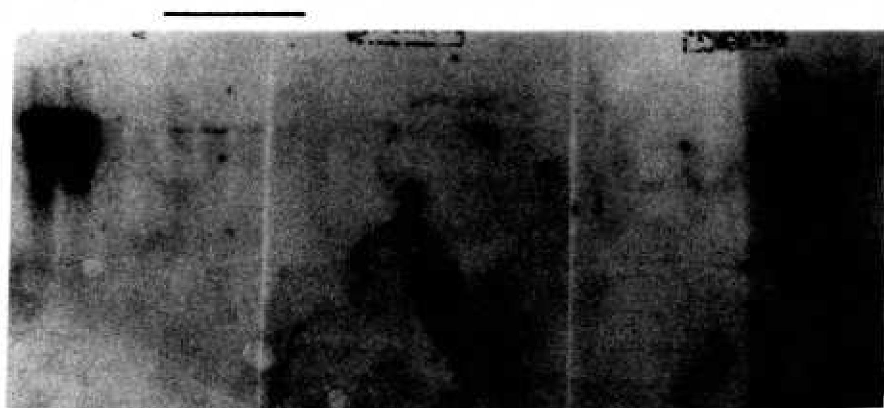
(Fig. 9b). The γ specific Bgl II fragment elutes from the RPC-5 in fractions 24-28. The elution of hybridising fragments at only one point in the gradient provides additional confirmation that the γ-specific sequences are contained in a single Bgl II digestion fragment of DNA 13 kb in length.

Figure 9b.

Autoradiogram of Southern transfer hybridisation of RPC-5 fractionated Bgl II digested human DNA. DNA from the gel in figure 9a was transferred to nitrocellulose and hybridised to ^{32}P labelled pHγG1 DNA. The bar indicates fractions 24 to 28 which hybridise with the probe.

Figure 9b





3.4 Purification of Human DNA for Cloning

(a) Introduction

To purify the 13 kb Bgl II sufficiently for cloning required a large amount of Bgl II digested DNA. It was proposed to digest 50 mg of DNA so requiring tens of thousands of units of the enzyme. The cost was prohibitive for the purchase of enzyme of a quality suitable for cloning. It was decided to purify the enzyme Bgl II.

(b) The preparation of Bgl II

Bacillus globigii has two class II restriction endonucleases designated Bgl I and Bgl II (Wilson and Young, 1976). The Bgl I and Bgl II were partially purified by gel filtration then separated from each other by ion exchange chromatography as described in 2.3 (i). A cleared lysate from 35 g of B.globigii cells was passed over a Bio-Gel A 0.5 m column of 470 ml bed volume. The column was washed through with buffer and fractions of approximately 10 ml collected. Fractions were assayed for enzyme activity as described in section 2.2 (e) by their ability to digest λDNA to give the appropriate pattern. Fractions 13 - 30 were pooled for further purification. Bgl I and Bgl II activities were separated by ion exchange chromatography on DEAE cellulose. The elution conditions used gave an overlap of activities as well as pure Bgl I and Bgl II. Restriction endonucleases are often contaminated with non-specific nucleases after partial purification. To check for this type of contamination λDNA was overdigested 20 fold without exhibiting non-specific degradation. Bgl II free of other nuclease contamination was pooled and assayed for activity. A fixed amount of λDNA was digested with increasing amounts of Bgl II until total

digestion was achieved and the activity calculated. The activity was 4 units μl^{-1} - a unit being defined as the amount of enzyme that will digest 1 μg of λDNA in 1 hour. The yield was approximately 80 000 units.

(c) Strategy for the purification of the 13 kb Bgl II DNA fragment

For the cloning system to be used a high degree of preliminary purification of the DNA sequence to be cloned is desirable. No single method of fractionation that was available was sufficient in the degree of purification afforded. It was decided to use a combination of preparative gel electrophoresis (the Southern "Gene Machine") and RPC-5 chromatography. The capacity of the "Gene Machine" is large compared to RPC-5 and was used as a first purification step. Size fractions from the Gene Machine containing γ-globin DNA sequences were further fractionated by RPC-5 chromatography.

50 mg of human DNA prepared as in 3.3 (b) was digested with 25 000 units of Bgl II for 4 hours at 37°C and the completeness of digestion monitored as previously described. Following phenol/chloroform extraction and ethanol precipitation of the digested products 42 mg of DNA were recovered.

(d) Fractionation of DNA on the Southern Gene Machine

The Southern "Gene Machine" (Southern, 1979) is a large cylindrical agarose gel electrophoresis system. Samples of DNA are electrophoresed from the outside to the centre of the gel where size fractions are collected on a dialysis membrane. A schematic diagram is shown in Figure 10. At regular intervals the DNA is eluted from the membrane and collected.

Figure 10.

Schematic representation of a section through a southern 'Gene Machine'.

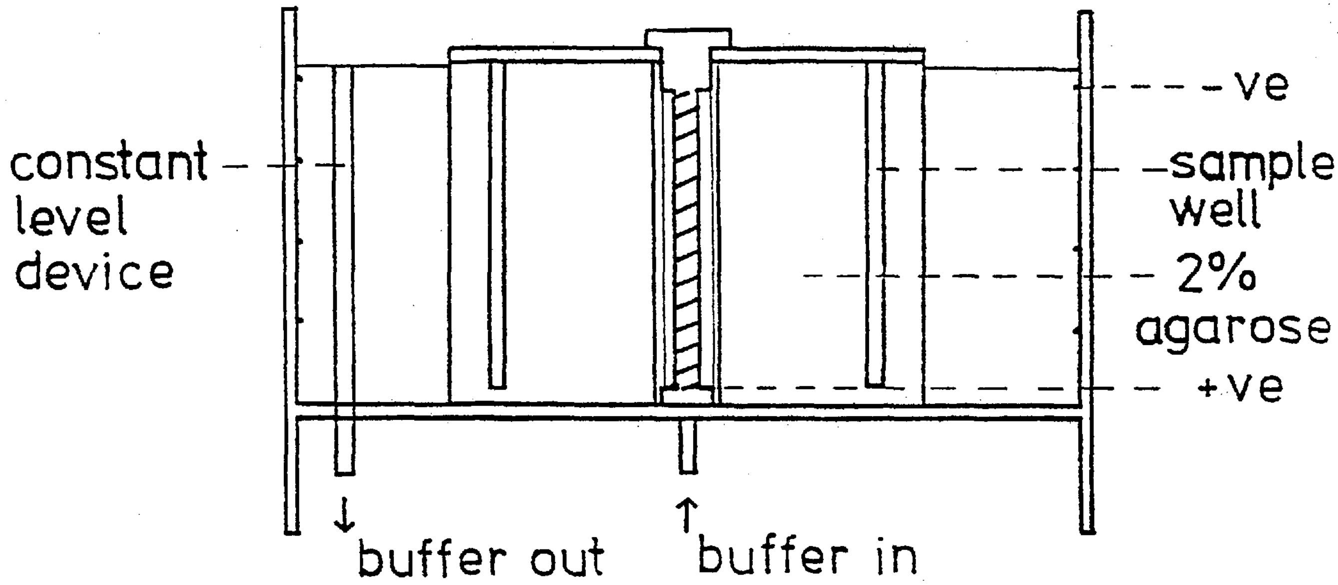
Top: general view

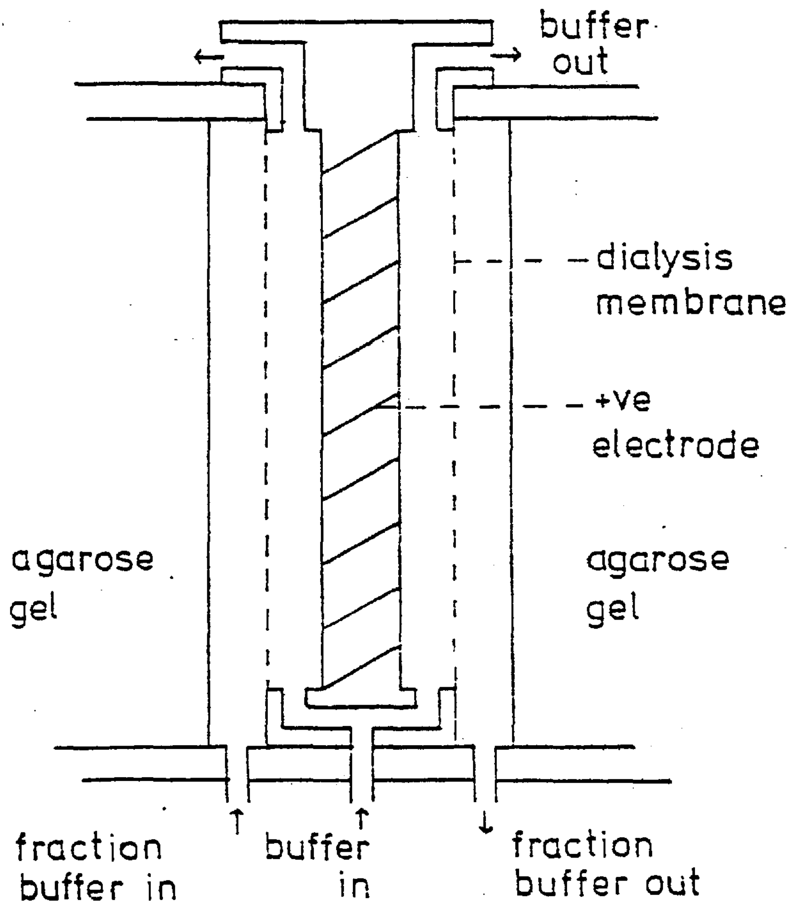
Bottom: detailed view of the anode

DNA is electrophoresed from the sample well to the dialysis membrane. At 30 min intervals the current is reversed and the DNA moves from the membrane into the buffer between the gel and the membrane. This buffer is collected then replaced with fresh buffer and electrophoresis is continued until the next fraction is to be collected. Buffer is circulated between the anode and cathode by a pump and all functions are controlled automatically.

Figure 10

THE GENE MACHINE





Fractionation is slow but accurate, dividing the DNA into hundreds of individual fractions by size. The "Gene Machine" electrophoresis was carried out on 40 mg of Bgl II digested DNA in collaboration with M. Doel and E. Cook of the Biochemistry Research Section, G.D. Searle Ltd., High Wycombe. Approximately 300 20 ml fractions were obtained eluting the DNA every 30 min. Fractions containing DNA fragments around 13 kb in length, as determined by agarose gel electrophoresis, were screened by transfer hybridisation with pHγGl. Fractions 241 to 288 were screened pooling the fractions in pairs i.e. 241 + 242, 243 + 244 etc. for convenience. The autoradiogram in Figure 11 shows that the γ probe hybridises to fractions 265 - 274. These fractions were pooled and prepared for RPC-5 chromatography.

(e) RPC-5 chromatography of gene machine purified DNA

Approximately 400 μg of γ specific DNA were recovered from the gene machine and fractionated on RPC-5. A 2 ml bed volume column was used and the DNA was eluted with a 40 ml gradient from 1.6 to 1.8 M sodium acetate. 47 fractions were collected from the column and were screened by transfer hybridisation. Figure 12 shows the autoradiogram showing that fractions 23 - 26 score highly positive with the γ probe. A repeat of this experiment showed that fractions 21, 22 and 27 also contained γ DNA sequences (data not shown).

(f) Estimation of the overall purification

A check of the relative purities of fractions 21 - 27 was necessary. 0.5, 0.1 and 0.02 μg of DNA from each of the fractions were probed along with 10, 15 and 20 μg of unpurified

Figure 11.

Autoradiogram of southern transfer hybridisation of 'Gene Machine' fractionated DNA.
DNA from fractions 241 to 288 from the 'Gene Machine' was electrophoresed in a 0.8% agarose gel and transferred to nitrocellulose for hybridisation with ^{32}P labelled pHγGl DNA. Hybridisation of the probe to fractions 265 to 274 is indicated by the bar.

Figure 11

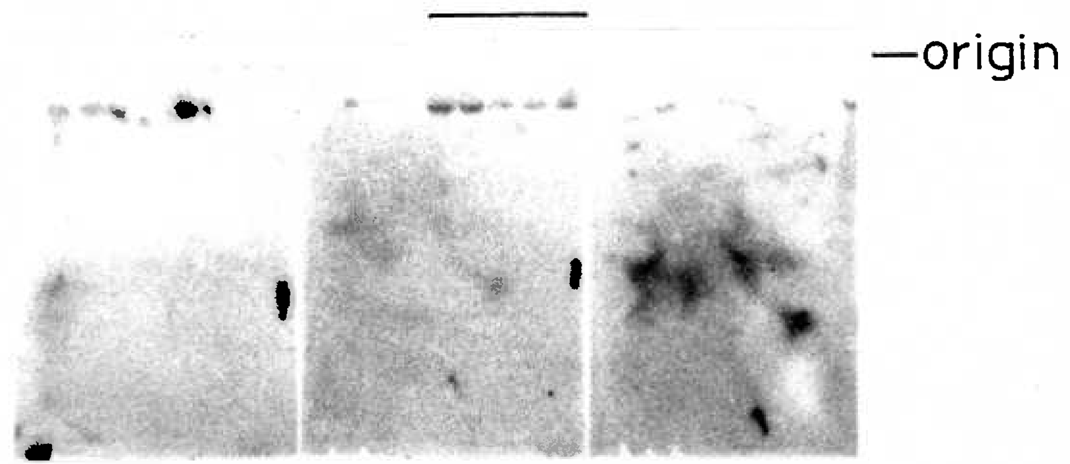
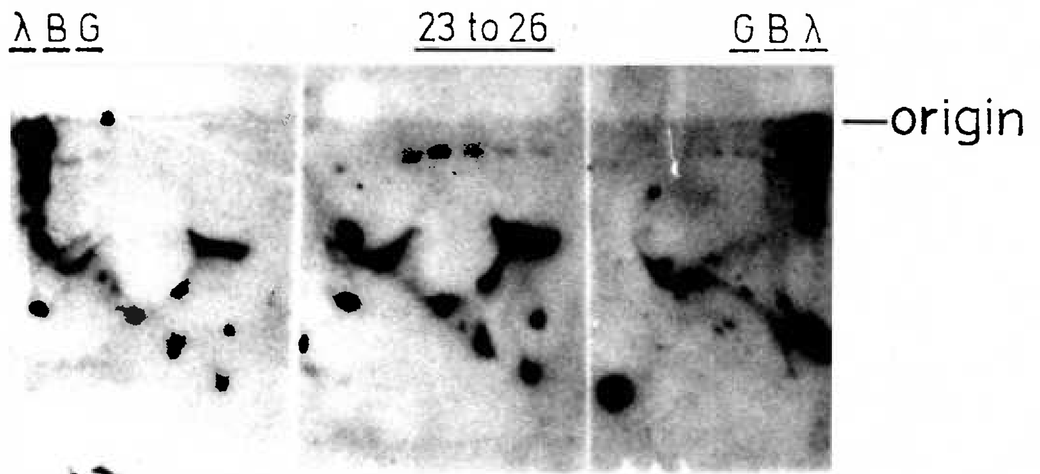
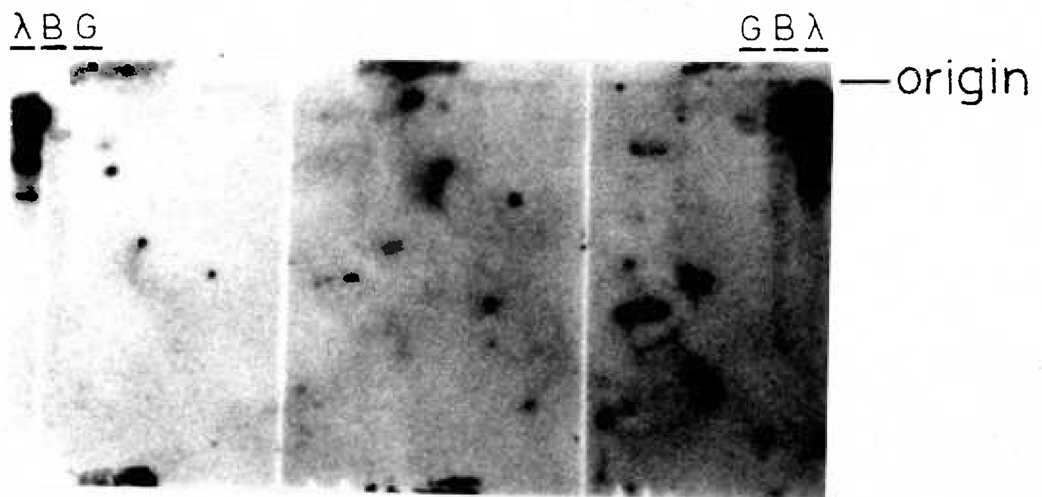


Figure 12.

Autoradiogram of Southern transfer hybridisation of RPC-5 fractionation of DNA previously purified by 'Gene Machine' fractionation.
Of the 47 fractions collected from the RPC-5 column, aliquots from those containing DNA were electrophoresed in a 0.8% agarose gel and transferred to nitrocellulose for hybridisation to ^{32}P labelled pHγGl DNA. Hybridisation of the probe to fractions 23 to 26 is indicated.

Figure 12



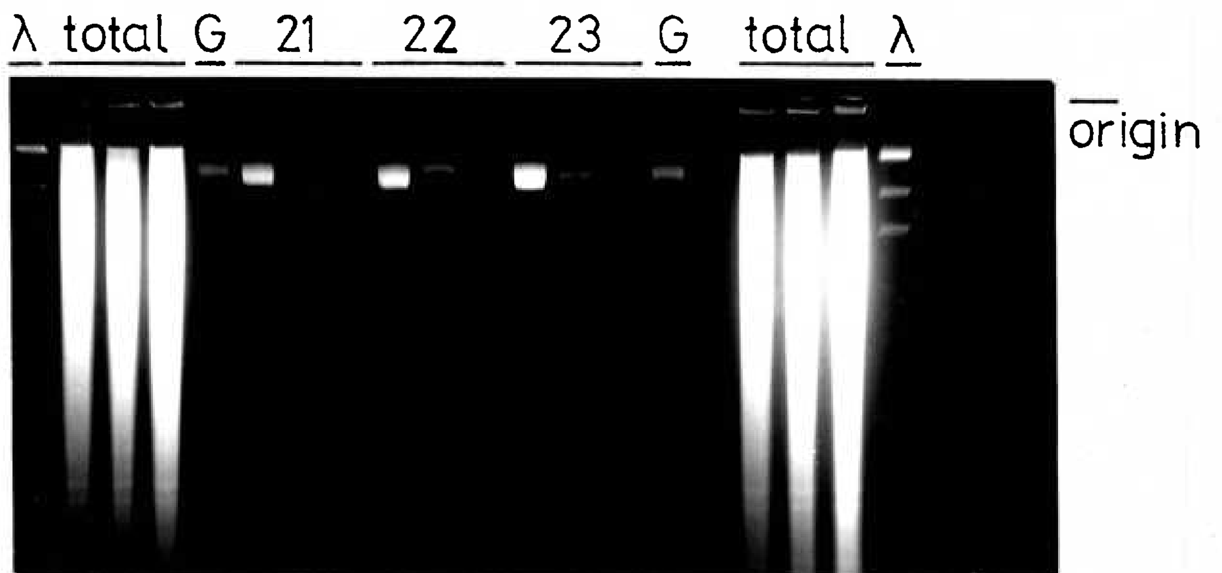


DNA in a transfer hybridisation. Comparing known amounts of DNA from each of the fractions in this way allowed a semi-quantitative comparison of their purification. Intensities of bands on autoradiograms from transfer hybridisations are not strictly quantitative but at least a comparison can be made. The agarose gel is shown in Figure 13 and the autoradiogram is shown in Figure 14. Fractions 22 and 23 were clearly the most highly enriched for γ globin DNA sequences. For fraction 22 an estimation of the enrichment was made. A densitometer scan was made of the photographic negative of the agarose gel used for the transfer hybridisation. The tracks containing 20 μg of unpurified DNA and 0.5 μg of purified fraction 22 were scanned. Fortuitously the traces coincided for the two scans. A representation of the scans is shown in Figure 15. The areas under the curves, which are proportional to the amount of DNA present in the gel tracks, were in the ratio of 1:14.3. This means that 7% of the human genome is represented in this size class of DNA. (The fact that the fraction 22 had been purified on RPC-5 as well as by taking a size cut is of no consequence in this calculation). The human haploid genome has a complexity of 3×10^9 base pairs. This means that 2.1×10^8 base pairs are in the size cut taken from the gene machine i.e. approximately 15 000 fragments each 13 kb in length are present in a Bgl II digest of the human genome. Comparing the intensity of bands in the autoradiogram in Figure 13 it can be seen that those of 0.02 μg of fraction 22 and 0.2 μg of DNA purified by the gene machine alone are of the same order. This represents a further purification of around 10 fold. This means that within the purified DNA the fragment to be cloned will be approximately 1 in 1500.

Figure 13.

0.8% agarose gel electrophoresis of DNA fractionated by both 'Gene Machine' and RPC-5.
Sets of tracks marked 21 to 27 contain, from left to right, 0.5, 0.1 and 0.02 μg of fractionated DNA.
The sets of tracks marked total contain, from left to right, 10, 15 and 20 μg of unfractionated Bgl II digested human DNA.
The tracks indicated by G contain 0.2 μg of DNA fractionated by the 'Gene Machine' alone and the tracks marked λ contain 0.4 μg of a (partial) Hind III digestion of λ DNA.

Figure 13



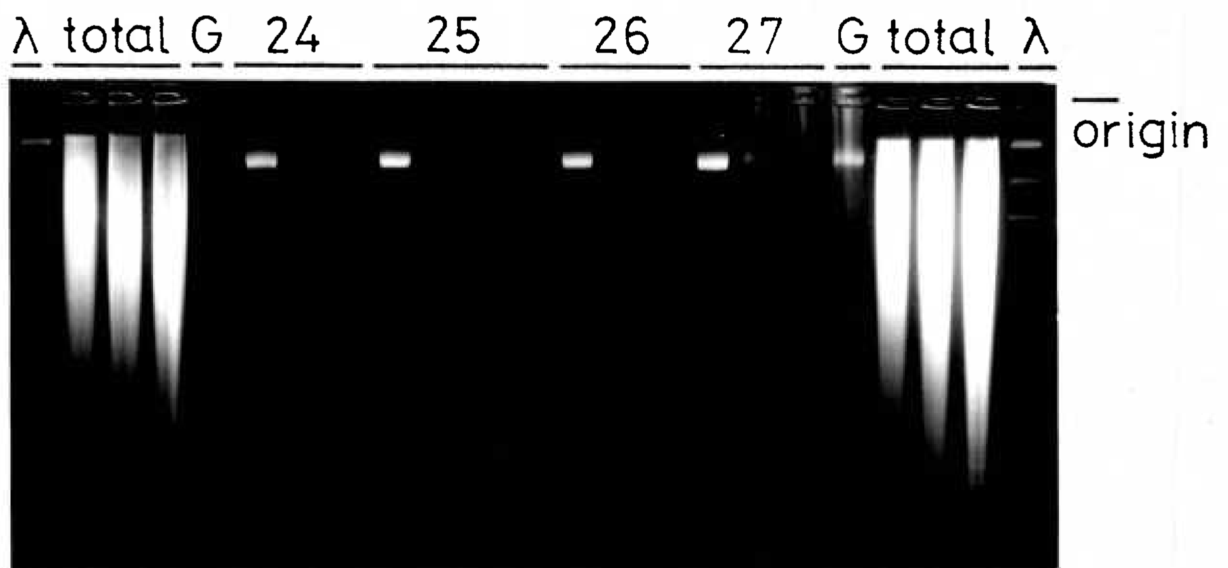


Figure 14.

Autoradiogram of Southern transfer hybridisation of the agarose gel in figure 13.
DNA was transferred from the agarose gel in figure 13 to nitrocellulose and was hybridised to ^{32}P labelled pHγGl DNA. Tracks are marked as in figure 13.

Figure 14

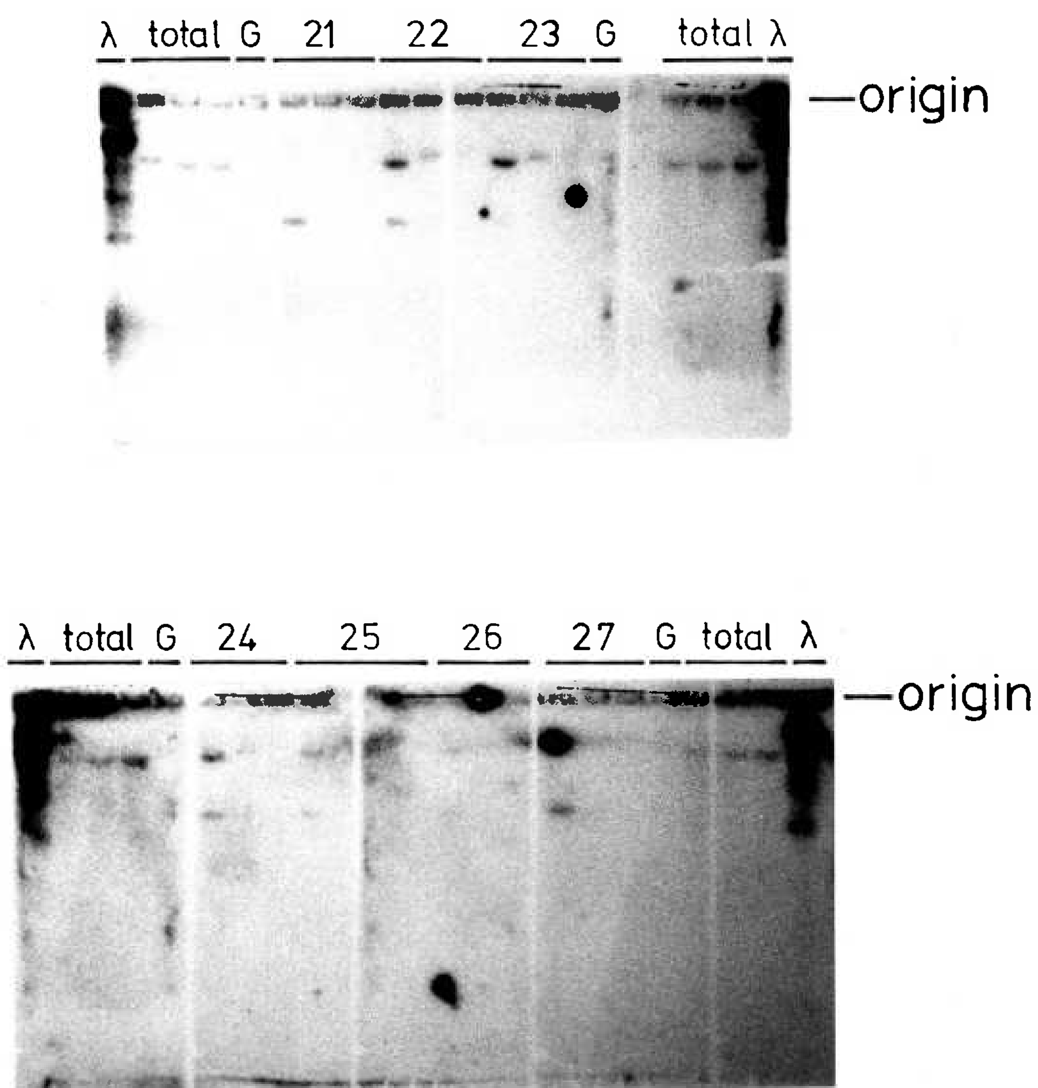


Figure 15.

Optical density profiles of purified and unpurified DNA.

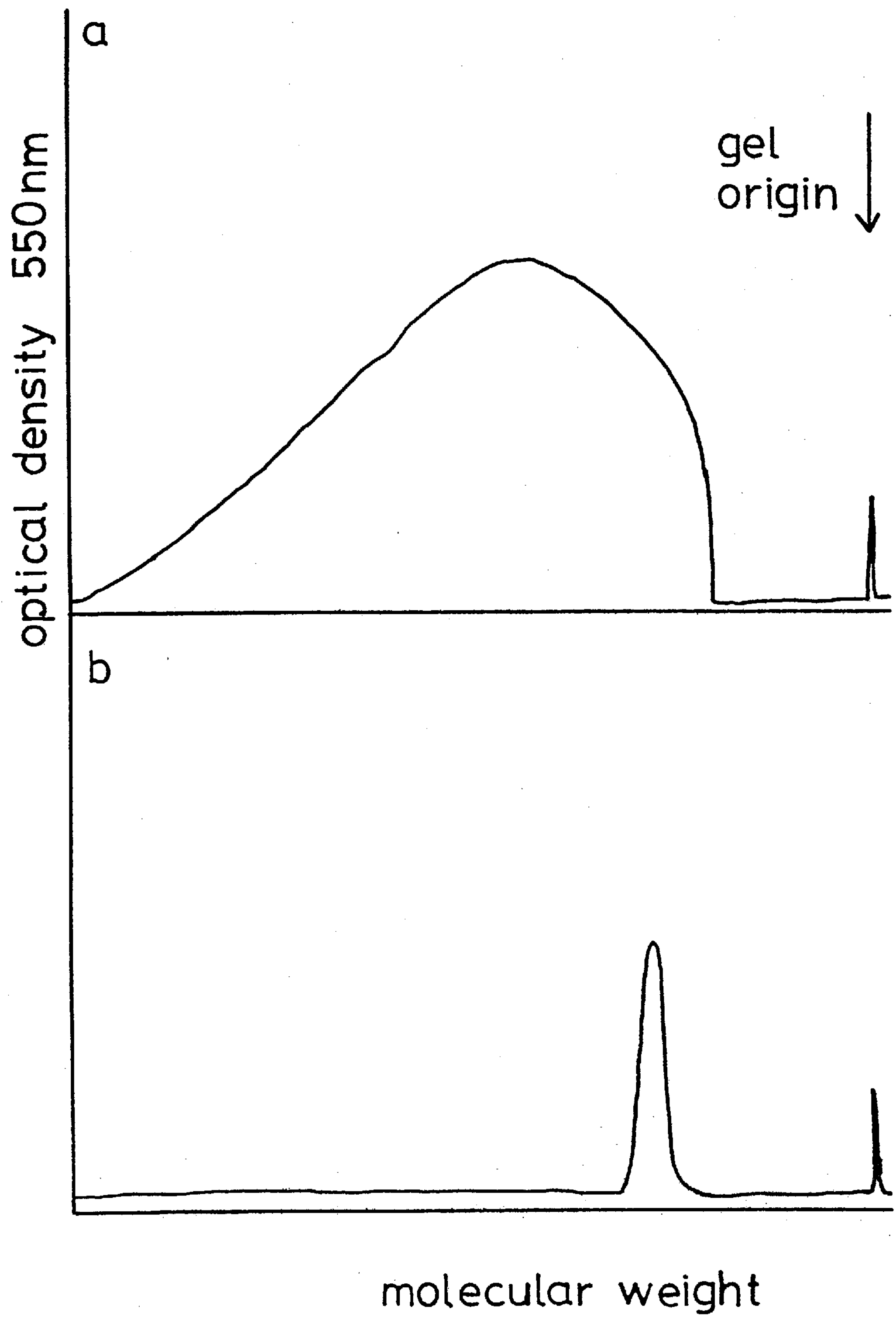
The optical density at 550 nm was measured for tracks on the photographic negative of figure 13.

Profile a: 20 µg of unpurified Bgl II digested human DNA.

Profile b: 0.5 µg of DNA from purified fraction 22.

The peak height in profile b coincides with the trace of profile a.

Figure 15



The overall enrichment of the DNA is of the order of 150 fold (14.3 x 10) which is verified by comparing the intensities of the bands in the autoradiogram in Figure 13 in the tracks of 10 µg unpurified DNA and 0.1 µg of fraction 22. The level of purification now made feasible the screening of transformed clones of E.coli for a clone containing the 13 kb Bgl II restriction fragment.

3.5 Cloning in the Host/Vector System E.coli χ1776/pBR322

(a) Introduction

The guidelines controlling genetic manipulation in this country (Williams, 1976) require certain levels of physical and biological containment dependent on the nature of the experiment to be carried out. An application was made to the Genetic Manipulation Advisory Group (GMAG) in March, 1978. They advised that to clone human genomic DNA in a disabled host/vector system, Category III level of physical containment was appropriate. Category III physical containment consists of a facility set aside solely for genetic manipulation with controlled access of personnel. The facility is maintained at negative air pressure with access through a simple air lock. All air exhausted from the facility passes through HEPA filters and all liquid waste is treated before release through the normal waste systems. Contaminated equipment leaves the facility by way of a double-ended autoclave. All aerosol producing manipulations are carried out in a Porton type cabinet. The Category III facility in the Biochemistry Department at Imperial College, London was used.

The host/vector system used was E.coli χ1776/pBR322. The disablement of this system resides in the host χ1776 (Curtiss et al., 1976) which requires thymine or thymidine and diaminopimelic acid for survival, is sensitive to bile salts and resistant to a number of different promiscuous bacteriophage. The plasmid vector pBR322 (Bolivar et al., 1977) is a derivative of the plasmid pMB1. It has single sites for a number of restriction enzymes and carries markers in the form of tetracycline and ampicillin resistance determinants. It has the added desirable feature that it is mobilised only very

poorly by conjugative plasmids. The proposed method of insertion of the purified DNA into the plasmid was by ligation into the single Bam HI site. The restriction enzymes Bgl II and Bam HI produce the same cohesive ends (Pirotta, 1976) so if pBR322 is cleaved to the linear form with Bam HI then Bgl II digested DNA may be inserted at the Bam HI site. The insertion involves the annealing of the cohesive ends intermolecularly and the closing of the single strand nicks with DNA ligase. The Bam HI digested pBR322 is prevented from recircularising without including an inserted DNA sequence by pretreatment with a phosphatase to remove the 5' phosphates required by DNA ligase to close single stranded nicks.

(b) Ligatability of Bgl II digested DNA

The ability to ligate Bgl II digested DNA was investigated. Both unpurified DNA and purified fraction 22 DNA were treated with DNA ligase as described in 2.3 (q). A portion of this DNA was re-digested with Bgl II. The starting DNA, the ligated DNA and the re-digested DNA were each electrophoresed in a 0.8% agarose gel. The results are shown in Figure 16a. Neither the unpurified DNA nor the fraction 22 DNA were ligated to any extent. This indicates that the inability to ligate the purified DNA was not as a consequence of the purification techniques used. The conditions of ligation were checked by digesting λDNA with Eco RI and Bgl II and treating the products with DNA ligase as before then redigesting the DNA with the appropriate enzyme. The results are shown in Figure 16b. The Eco RI digested λDNA when treated with ligase moves to a position of higher molecular weight. On digestion with Eco RI the normal pattern returns. Bgl II digested λDNA was not ligated. The results suggest that

Figure 16.

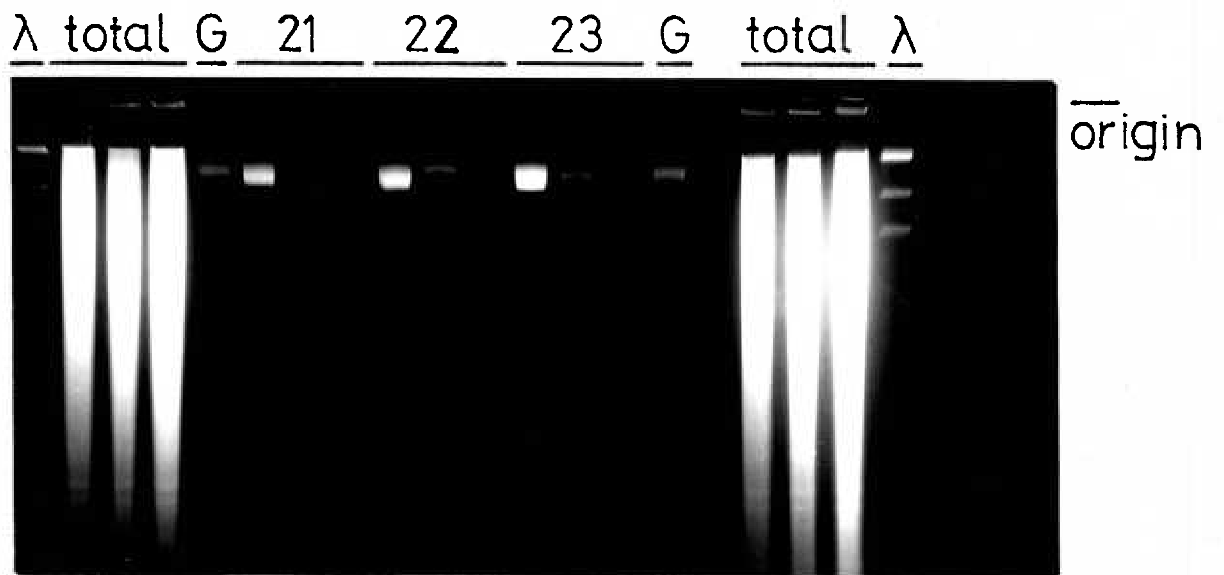
0.8% agarose gel electrophoresis of ligated DNA samples.

From left to right each set of tracks shows the starting DNA, that DNA treated with DNA ligase and the ligated DNA redigested with the appropriate restriction enzyme.

a: The total tracks contain 5 μg of Bgl II digested human DNA and tracks 22 contain 0.1 μg of purified fraction 22 DNA.

b: The λBgl II tracks contain 0.25 μg of Bgl II digested λ DNA and the λRI tracks 0.25 μg of Eco RI digested λ DNA.

Figure 13



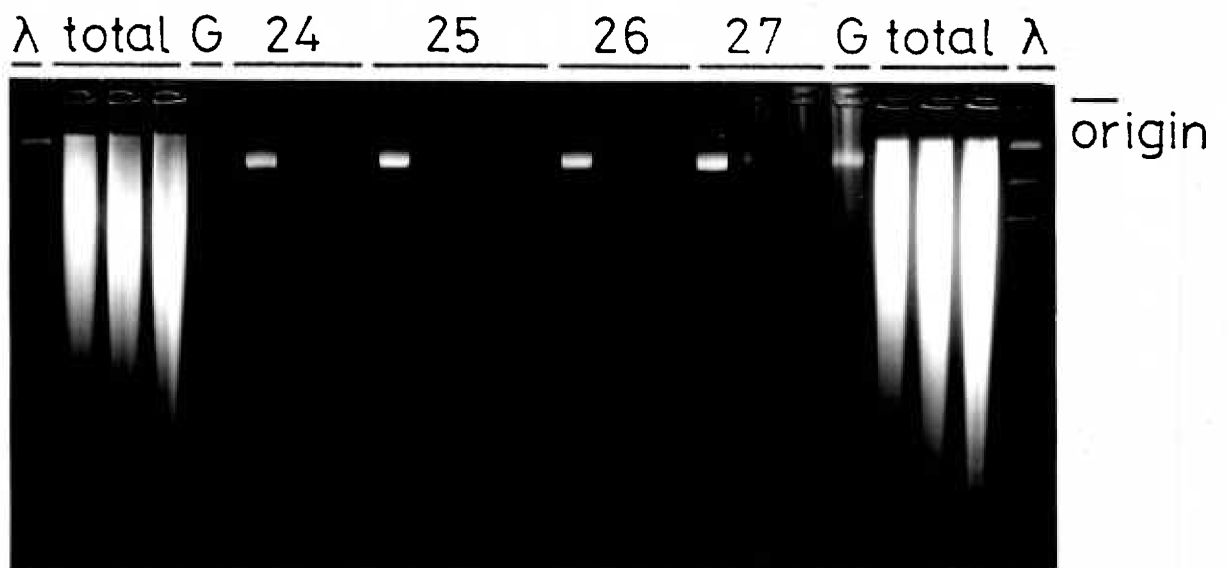


Figure 14

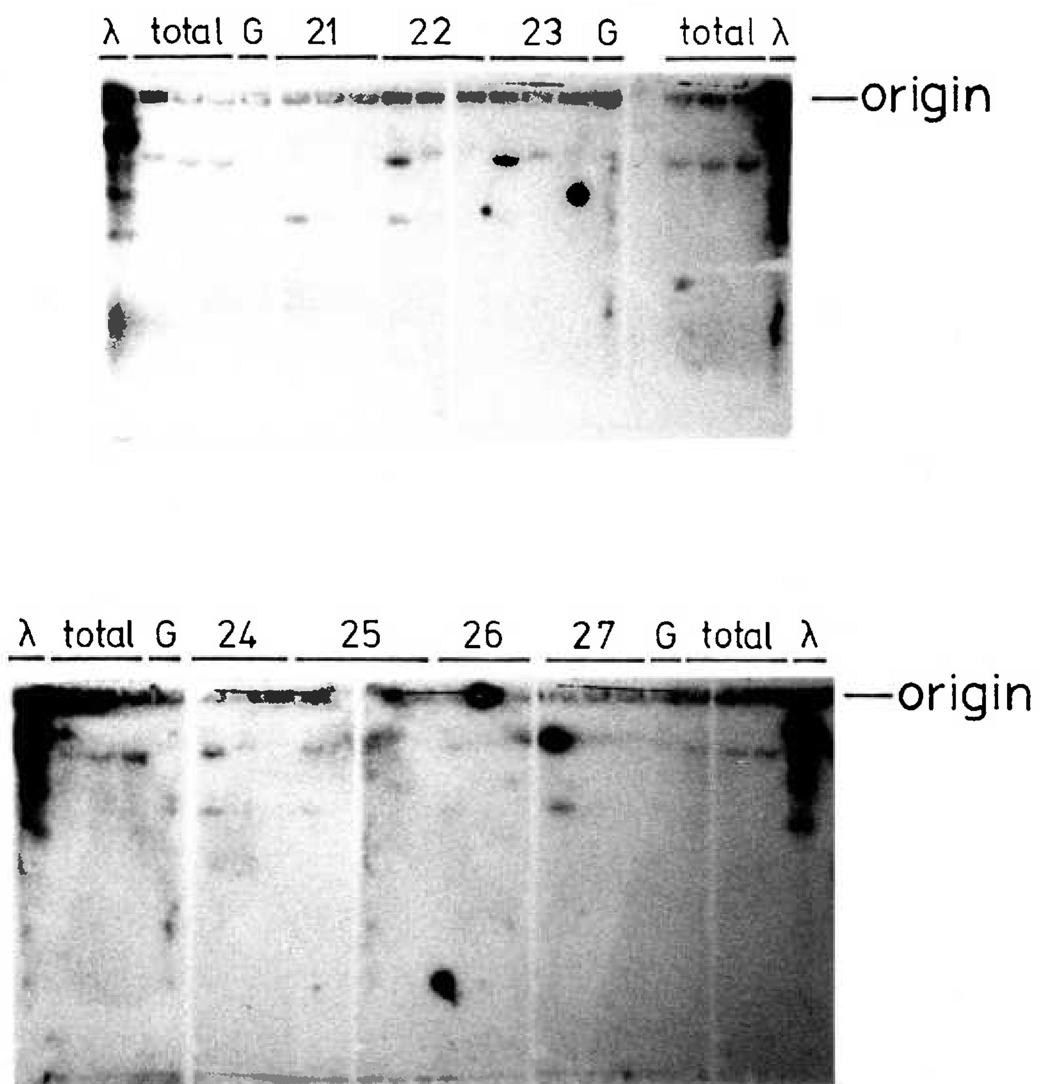


Figure 15.

Optical density profiles of purified and unpurified DNA.

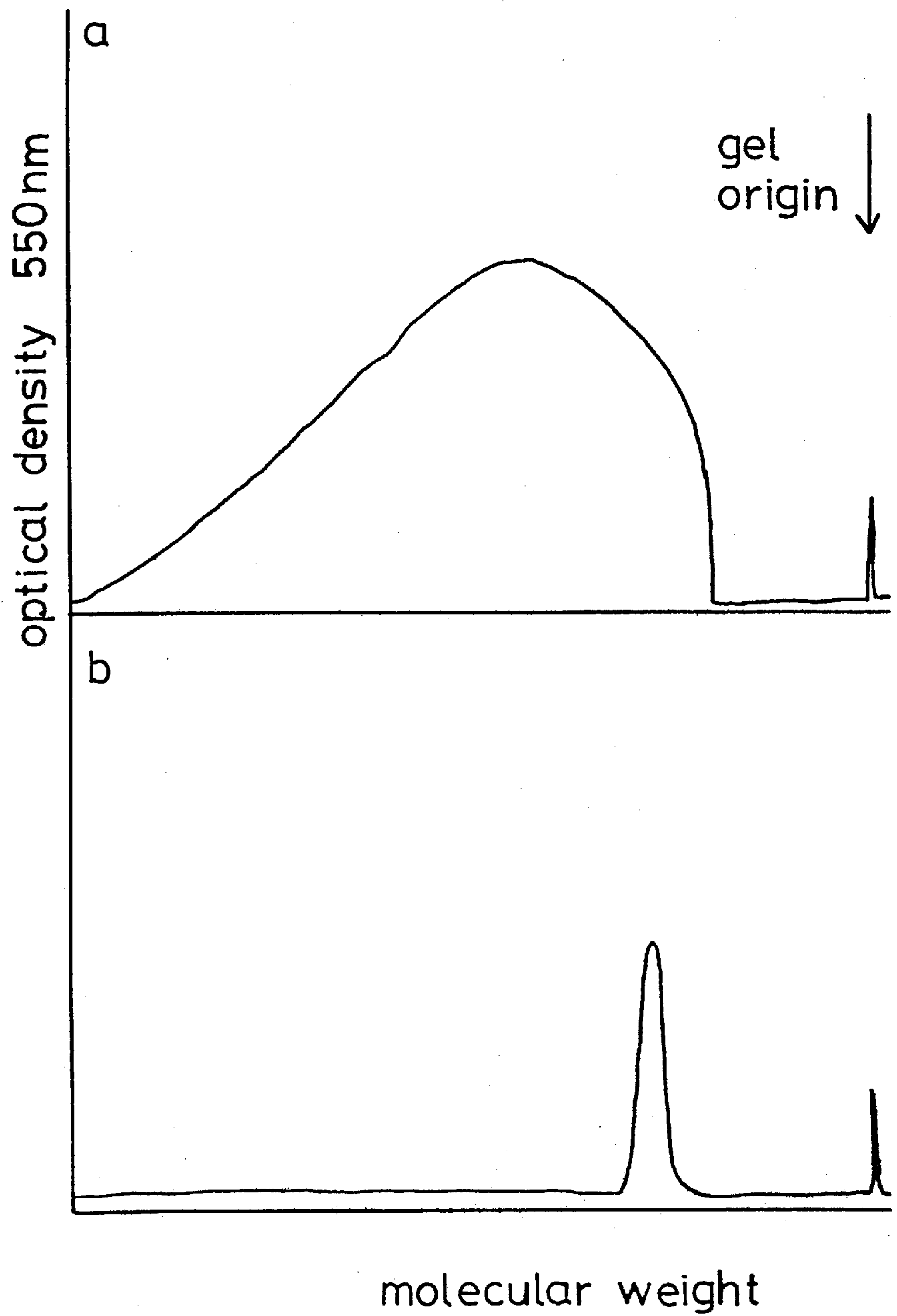
The optical density at 550 nm was measured for tracks on the photographic negative of figure 13.

Profile a: 20 µg of unpurified Bgl II digested human DNA.

Profile b: 0.5 µg of DNA from purified fraction 22.

The peak height in profile b coincides with the trace of profile a.

Figure 15



The overall enrichment of the DNA is of the order of 150 fold (14.3 x 10) which is verified by comparing the intensities of the bands in the autoradiogram in Figure 13 in the tracks of 10 μg unpurified DNA and 0.1 μg of fraction 22. The level of purification now made feasible the screening of transformed clones of E.coli for a clone containing the 13 kb Bgl II restriction fragment.

3.5 Cloning in the Host/Vector System E.coli χ1776/pBR322

(a) Introduction

The guidelines controlling genetic manipulation in this country (Williams, 1976) require certain levels of physical and biological containment dependent on the nature of the experiment to be carried out. An application was made to the Genetic Manipulation Advisory Group (GMAG) in March, 1978. They advised that to clone human genomic DNA in a disabled host/vector system, Category III level of physical containment was appropriate. Category III physical containment consists of a facility set aside solely for genetic manipulation with controlled access of personnel. The facility is maintained at negative air pressure with access through a simple air lock. All air exhausted from the facility passes through HEPA filters and all liquid waste is treated before release through the normal waste systems. Contaminated equipment leaves the facility by way of a double-ended autoclave. All aerosol producing manipulations are carried out in a Porton type cabinet. The Category III facility in the Biochemistry Department at Imperial College, London was used.

The host/vector system used was *E.coli* χ1776/pBR322. The disablement of this system resides in the host χ1776 (Curtiss *et al.*, 1976) which requires thymine or thymidine and diaminopimelic acid for survival, is sensitive to bile salts and resistant to a number of different promiscuous bacteriophage. The plasmid vector pBR322 (Bolivar *et al.*, 1977) is a derivative of the plasmid pMB1. It has single sites for a number of restriction enzymes and carries markers in the form of tetracycline and ampicillin resistance determinants. It has the added desirable feature that it is mobilised only very

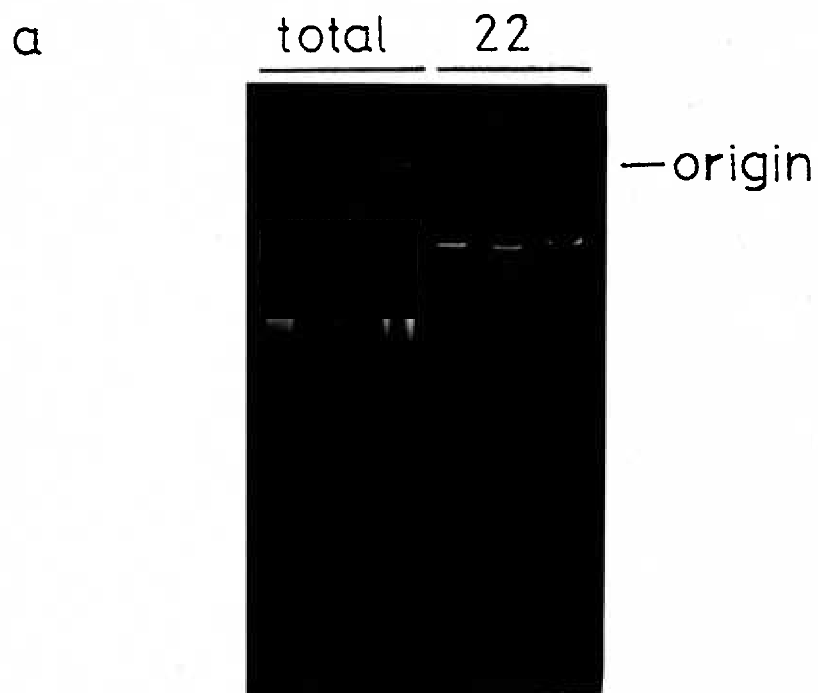
poorly by conjugative plasmids. The proposed method of insertion of the purified DNA into the plasmid was by ligation into the single Bam HI site. The restriction enzymes Bgl II and Bam HI produce the same cohesive ends (Pirotta, 1976) so if pBR322 is cleaved to the linear form with Bam HI then Bgl II digested DNA may be inserted at the Bam HI site. The insertion involves the annealing of the cohesive ends intermolecularly and the closing of the single strand nicks with DNA ligase. The Bam HI digested pBR322 is prevented from recircularising without including an inserted DNA sequence by pretreatment with a phosphatase to remove the 5' phosphates required by DNA ligase to close single stranded nicks.

(b) Ligatability of Bgl II digested DNA

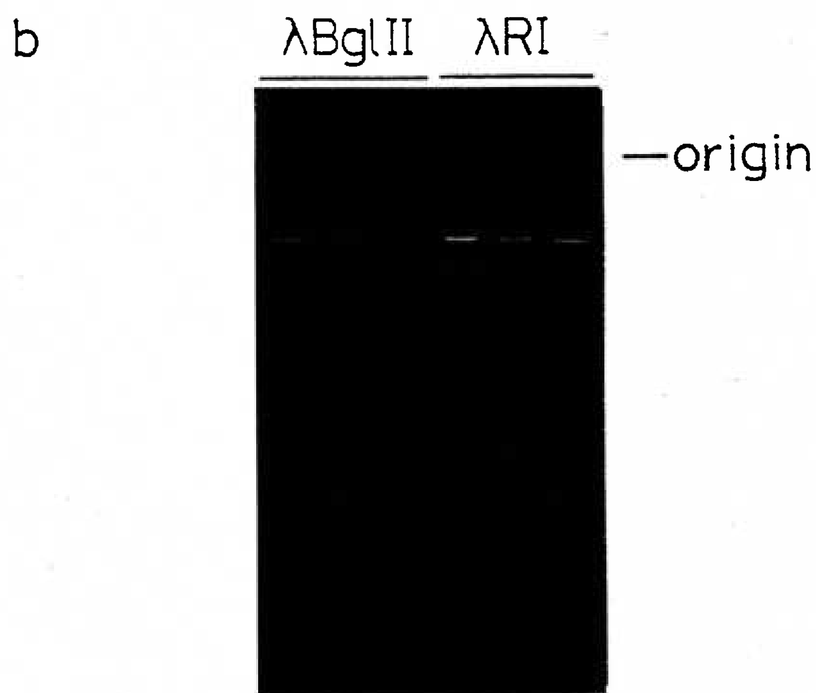
The ability to ligate Bgl II digested DNA was investigated. Both unpurified DNA and purified fraction 22 DNA were treated with DNA ligase as described in 2.3 (q). A portion of this DNA was re-digested with Bgl II. The starting DNA, the ligated DNA and the re-digested DNA were each electrophoresed in a 0.8% agarose gel. The results are shown in Figure 16a. Neither the unpurified DNA nor the fraction 22 DNA were ligated to any extent. This indicates that the inability to ligate the purified DNA was not as a consequence of the purification techniques used. The conditions of ligation were checked by digesting λDNA with Eco RI and Bgl II and treating the products with DNA ligase as before then redigesting the DNA with the appropriate enzyme. The results are shown in Figure 16b. The Eco RI digested λDNA when treated with ligase moves to a position of higher molecular weight. On digestion with Eco RI the normal pattern returns. Bgl II digested λDNA was not ligated. The results suggest that

Figure 16

a



b



the Bgl II is contaminated with phosphatase activity or slight exonuclease activity.

An alternative method for the insertion of the purified DNA into the plasmid vector was attempted. Homopolymer tails may be added enzymatically to the termini of DNA molecules: if the vector and insert DNAs are tailed with complementary deoxynucleotides they may be annealed together to form recombinant circularised molecules for transformation.

(c) Terminal transferase

Terminal deoxynucleotidyl transferase (TdT) isolated from calf thymus was first used by Jackson et al., (1972) to add homopolymer tails to the termini of DNA molecules. TdT catalyses the addition of nucleotides and deoxynucleotides to the 3' termini of single stranded DNA in the presence of Mg^{2+} cations. If Mg^{2+} is replaced by Co^{2+} then the enzyme will add tails to 3' termini of double stranded DNA whether "flush" or staggerred with respect to the 5' terminus (Roychoudhury et al., 1976). The majority of restriction endonucleases, including Bgl II and Hind III produce 5' extensions and so tailing was carried out in the presence of Co^{2+} cations.

(d) Tailing of pBR322

30 µg of pBR322 DNA entirely in the supercoiled and open circular form as judged by agarose gel electrophoresis was digested as described in 2.2 (e) with Hind III which makes a single cut. Figure 17 shows the DNA before and after digestion with Hind III. The digested DNA can be seen to be entirely in the linear form.

Tailing conditions previously established were used

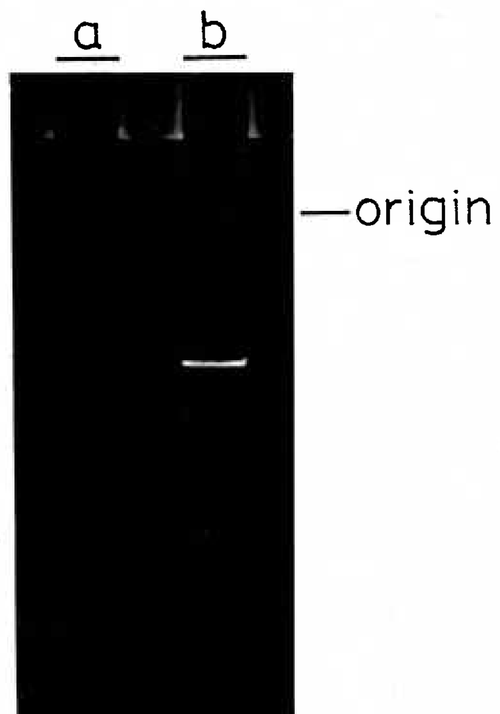
Figure 17.

1% agarose gel electrophoresis of Hind III digested pBR322 DNA.

Track a: 0.2 µg of pBR322 DNA. The upper and lower bands the open circular and covalently closed circular forms respectively of the plasmid.

Track b: 0.4 µg of Hind III digested pBR322 DNA.

Figure 17



(H. Liebscher, personal communication). 3.25 μg of Hind III digested pBR322 were incubated as detailed in 2.3 (1) in 20 μM ^{3}H-dCTP (22 Ci mmol^{-1}) with 25 units of TdT at a DNA terminus concentration of 5 nM for 12 min at 37° C. From an aliquot of the reaction removed at 12 min the tail length was calculated as the number of deoxynucleotides added by measuring the radioactivity of acid precipitable material. The tail length was calculated to be 24 nucleotides. Unincorporated nucleotides were removed by passing over Sephadex G-50 after phenol/chloroform extraction. The Sephadex was equilibrated in a buffer of half the concentration of that in which the tailed DNA was to be annealed. This allowed the recovered DNA to be concentrated by lyophilisation followed by dissolving in half the former volume of distilled water. This tailed plasmid DNA was designated pBR322. dC-24. A further 3.25 μg of Hind III digested DNA was incubated as before with 40 μM ^{3}H-dCTP and 50 units of TdT. The tail length calculated was 80 nucleotides and the tailed DNA was designated pBR322.dC-80.

(e) Tailing of purified DNA

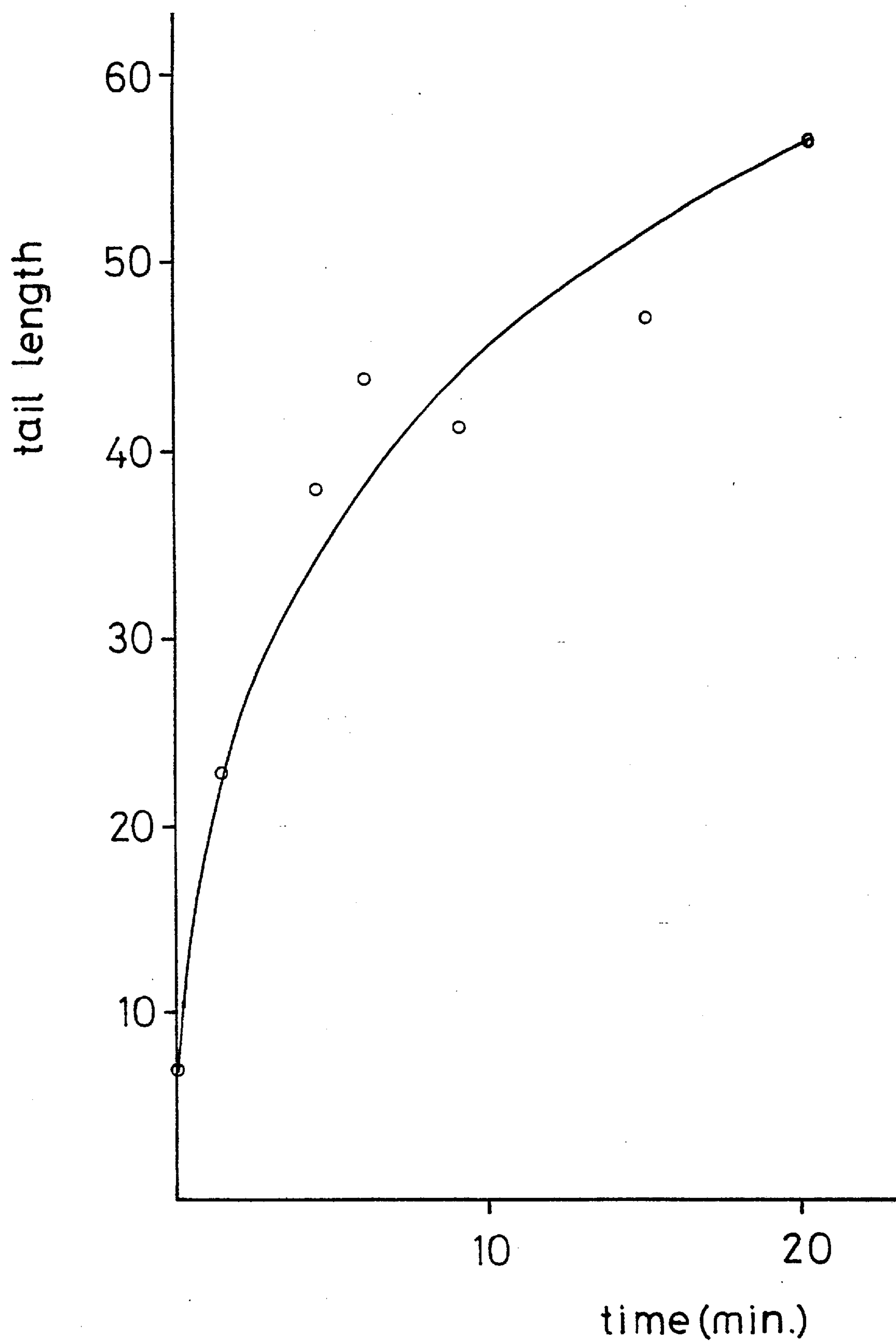
Fractions 22, 23 and 24 of the purified DNA were to be tailed in an analogous manner to the linearised plasmid. Small scale analytical reactions were carried out for the tailing of each fraction of DNA. The concentration of ^{3}H-dGTP required to tail the DNA was higher than that required for the ^{3}H-dCTP to tail pBR322. Figure 18 shows the incorporation of ^{3}H-dGTP vs. time for fraction 24 DNA at a ^{3}H-dGTP concentration of 500 μM. Incorporation is plotted as tail length in nucleotides in a reaction containing 0.6 μg DNA and 2.5 units of TdT. Analytical curves were constructed for each DNA

Figure 18.

Incorporation of ^{3}H-dGTP by terminal transferase into fraction 24 DNA.

From an incubation containing 0.6 μg of DNA, 0.05 μg aliquots were removed at the times shown and the DNA was precipitated with TCA. Aliquots were diluted with water to 100 μl and to each was added 30 μl of 1 mg ml^{-1} BSA and 150 μl of 10% (w/v) TCA plus 1% (w/v) sodium pyrophosphate. After 30 min on ice this was filtered through Whatman GF/A or GF/C filters which were washed prior to scintillation counting in TBS. Incorporation is plotted as tail length, measured in nucleotides, vs. time.

Figure 18



fraction and times chosen to give tail lengths of approximately 50 nucleotides. The actual tail lengths calculated for fractions 22, 23 and 24 following tailing were 57, 48 and 47 respectively. The tailed purified DNAs were designated 22.dG, 23.dG and 24.dG. Each was concentrated before use as described for the tailed plasmid DNA.

(f) Transformation of *E.coli* χ1776

Recombinant DNA is introduced into *E.coli* by pre-treating log phase cells with $CaCl_2$ (Cohen *et al.*, 1972) then incubating the cells with the DNA in $CaCl_2$. The $CaCl_2$ alters the cell surface to facilitate the uptake of DNA. The standard method however gives a low transfection efficiency with *E.coli* χ1776, and an improved method of transformation giving a higher efficiency has been developed (Norgard *et al.*, 1978). A modification by H. Liebscher of this method was used (personal communication). Cells were incubated with pBR322 and then plated out on L agar plates containing ampicillin. The efficiency of transformation in three experiments were 6×10^5, 2×10^5 and 1.5×10^5 transformants μg^{-1} of DNA using 0.1 ng of DNA per plate. Efficiency of this order was adequate for the cloning to be carried out.

(g) Annealing of tailed plasmid with insert and transformation

pBR322.dC-24 and pBR322.dC-80 were annealed to an equimolar amount of 22.dG. The incubation was in DNA dilution buffer at a plasmid concentration of 200 ng ml^{-1} for 2 hr at 65^oC, following which the mix was allowed to cool to room temperature over a period of 5 hours. The conditions allow the annealing of the complementary tails so that circles are formed.

E.coli χ1776 was transformed as described in 2.3 (m) with the annealed DNA at Category III containment and the results are shown in Table 1. The efficiency of transformation was approximately 3 - 10 fold reduced as compared with transformation with pBR322.

(h) Screening of recombinant clones

Approximately 1500 clones were screened by the method of Grunstein and Hogness, (1975). Colonies of bacteria are grown on sheets of nitrocellulose then lysed and the DNA fixed to the nitrocellulose. A ^{32}P labelled probe is hybridised and the radioactivity in positive colonies detected by autoradiography. Colonies were picked by sterile toothpick from the agar plates into L broth + ampicillin in the wells of standard micro-titre dishes; 1 colony per well. The dishes were incubated overnight and replicas made onto nitrocellulose filters by means of a comb of 48 pins. The pins were set in an array 8 x 6 allowing the replication of the well contents of half of a dish onto a 9 cm circle of nitrocelluse laid onto the surface of an L agar plate. The filters were incubated overnight and the colonies of bacteria lysed and fixed to the filters as described in 2.3.(p). ^{32}P labelled pHγG1 was hybridised to the filters in essentially the same way as for transfer hybridisations.

It was possible that the pCR1 portion of pHγG1 would hybridise to the pBR322 portion of the new recombinants so giving false positives because of the relatedness of the two plasmids. This was shown not to be the case by hybridising pHγG1 to E.coli χ1776/pB322 and to E.coli χ1776/pHγG1. The results are shown in Figure 19. The labelled pHγG1 hybridises only weakly to colonies with pBR322 as compared with the

Table 1.

Summary of transformations of *E.coli* χ1776 with tailed vector and insert DNA.
Efficiency is expressed as the number of transformants per μg of vector DNA.

Table 1

1st. transformation

	plasmid (ng)	transformants	efficiency
pRB322.dC-24/22.dG	15	502	3.3×10^4
pBR322.dC-80/22.dG	15	56	3.7×10^3

efficiency with supercoiled pBR322: 1.1×10^5

2nd. transformation

	plasmid (ng)	transformants	efficiency
pBR322.dC-24/22.dG	15	445	3×10^4
pBR322.dC-80/22.dG	15	232	1.5×10^4

efficiency with supercoiled pBR322: 2.4×10^5

3rd. transformation

	plasmid (ng)	transformants	efficiency
pBR322.dC-24/22.dG	30	582	1.9×10^4

efficiency with supercoiled pBR322: not calculated

All control plates without DNA had no transformants

Total number of transformants: 1817

Figure 19.

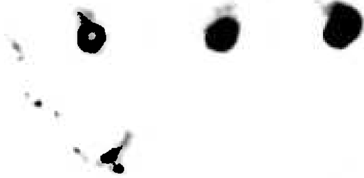
Grunstein-Hogness control screening.

^{32}P labelled pHγGl DNA was hybridised to lysed colonies of E.coli χ1776 harbouring either pBR322 or pHγGl on circles of nitrocellulose. The colonies were arranged in two arrays each 5x5. The strong hybridisation is to the colonies containing pHγGl. Hybridisation to colonies containing pBR322 is variable but much weaker.

Figure 19

G/H χ1776/pBR322
χ1776/pHγG1

pHγG1 PROBE



homologous DNA. Even so, denatured pCR1 was included in the hybridisations as competitor. Figure 20 shows autoradiograms from the screening of recombinant clones. A number of positive scoring colonies can be seen and of these 12 were picked and rescreened twice. On both occasions the screening was negative for all clones. The possibility of a technical error was excluded by the inclusion of colonies containing pHγG1 which all scored positive.

(i) Preparation and analysis of recombinant DNA

It was decided to investigate further three of the clones that had scored positive in the first screen. 500 ml cultures of clones designated 4G6, 6F2 and 6E7 were grown and cleared lysates of the bacteria prepared at Category III containment as previously described. Restriction enzyme digestions were carried out on the lysates following a single phenol/chloroform extraction. Figure 21 shows a 0.8% agarose gel of Hind III digests of the three lysates and pBR322 as well as the four DNAs undigested. It shows that in 4G6, 6E7 and 6F2 there is no inserted DNA sequence and that the Hind III site is intact even though the site should have been blocked by the addition of tails. Even though the Hind III digestion of 4G6 (track b) is partial the linear form can be clearly seen.

(j) Investigation of DNA for possible contaminants

When either tailed pBR322 DNA and tailed fraction 22 DNA were used individually to transform E.coli HB101 no transformants were obtained. This indicated that there was no contaminating circular form plasmid DNA in either of the DNA stocks used for the cloning procedures in 3.5(g).

Figure 20.

Grunstein-Hogness screening of recombinant clones. ^{32}P labelled pHγGl DNA was hybridised to lysed colonies of recombinant *E.coli* χ1776 on circles of nitrocellulose. The colonies are in 6 arrays each 6x8. The position of clones 4G6, 6E7 and 6F2 are indicated by the letters G,E and F respectively.

Figure 20

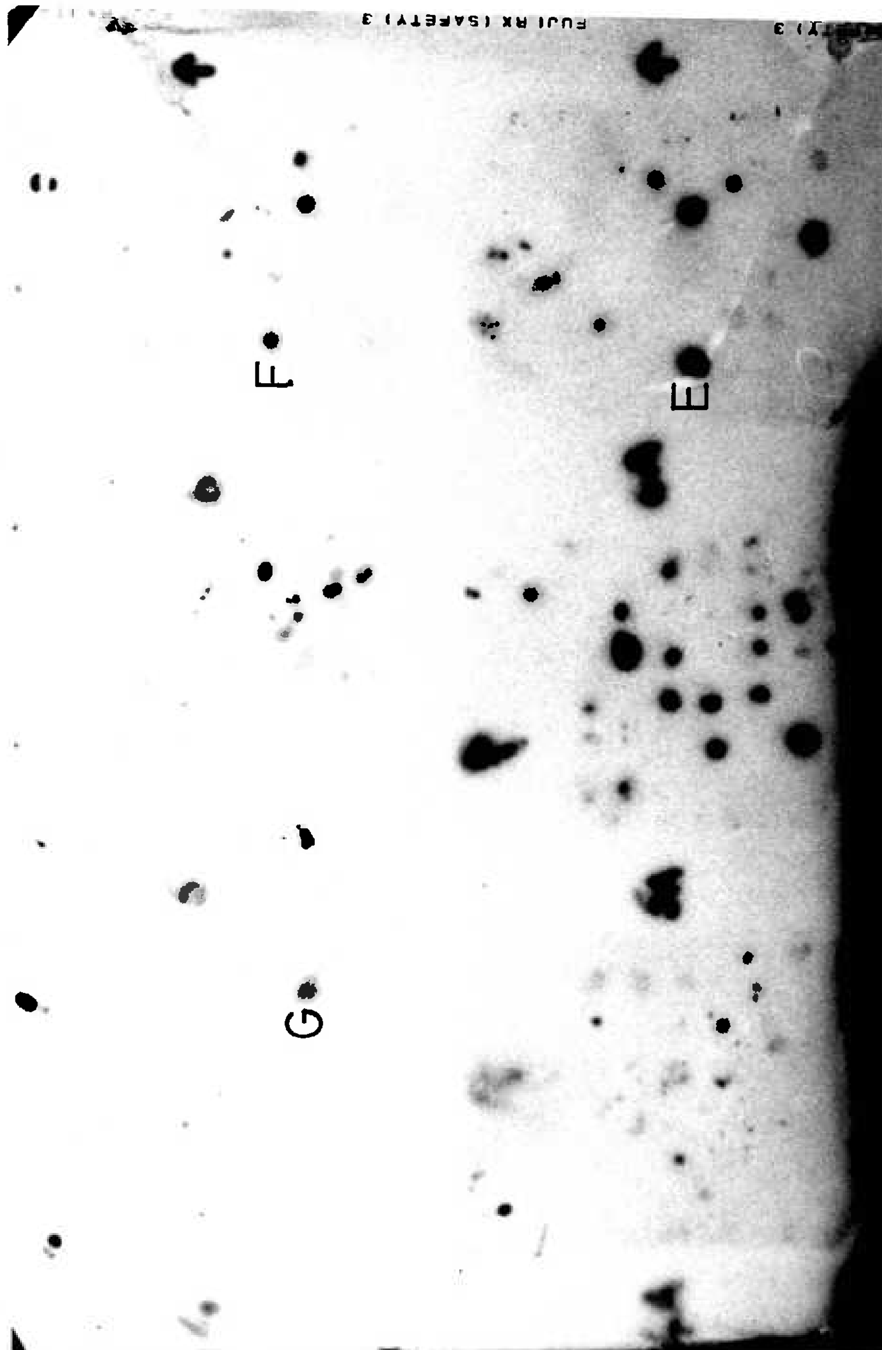


Figure 21.

0.8% agarose gel electrophoresis of plasmid DNA from clones 4G6, 6E7 and 6F2.

pBR322, 4G6, 6E7 and 6F2 plasmid DNAs were digested with Hind III and were electrophoresed in a 0.8% agarose gel. Each pair of tracks shows the undigested and digested DNA.

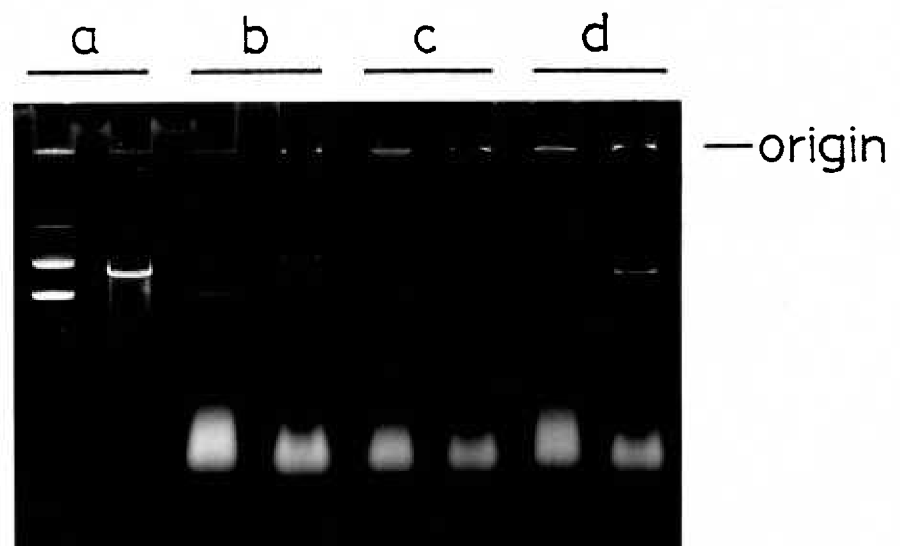
Track a: pBR322

Track b: 4G6

Track c: 6E7

Track d: 6F2

Figure 21



The tailed fraction 22 DNA was viewed in the electron mircroscope as described in 2.3 (t) to see if any low level of contamination with vector sized DNA existed. The DNA was found to be extensively degraded and no longer of one size class. Cytochrome C is often contaminated with randomly sized DNA fragments in commercial preparations (P. Little, personal communication) and so could give rise to the apparent degradation. This possibility was excluded by preparing a sample grid without DNA. No DNA was seen on this grid. Tailed fractions 23 and 24 were also found to be degraded and unlikely to be suitable for the formation of recombinant molecules. It was decided to attempt the cloning again by the method originally proposed: insertion of Bgl II digested DNA, by ligation, into the Bam HI site of a derivative of pBR322.

3.6 Cloning in the host/vestor system E.coli HB101/pAT153

(a) Introduction

In 1979, after revision of the GMAG guidelines, the proposed cloning experiments were first down-categorised to Category II* and, at a later date, Category I* containment was advised for the work. New host/vector systems were also approved by GMAG as being disabled. A derivative of the plasmid pBR322 was developed such that it was not possible to mobilise the plasmid to another host *in vivo*. This derivative, pAT153 (Twigg and Sherratt, 1979) used in conjunction with the $recA^-$ host E.coli HB101 (Boyer and Roulland-Dussoix, 1969) satisfies the biological containment requirements. Using this host/vector system it was proposed to clone the 13 kb gamma specific Bgl II fragment of human DNA by ligating it into the Bam HI site in pAT153.

(b) Restriction enzyme digestion of human and plasmid DNAs

Human DNA was prepared as previously described in 3.3 (b) and pAT153 DNA prepared as described for other plasmids in 3.2 (b). 1.5 mg of human DNA was digested with Bgl II purchased from New England Biolabs Inc. Completeness of digestion was monitored by including an internal λDNA marker in an aliquot from the main digestion. Following phenol/chloroform and ether extractions an aliquot of the digested DNA was incubated with DNA ligase as described in 2.3 (q). Figure 22 shows the digested DNA before and after treatment with DNA ligase. This clearly indicates the integrity of the cohesive termini following digestion with Bgl II. pAT153 DNA was digested to completion with Bam HI. Figure 23 shows the linearised pAT153 DNA before and after incubation with DNA ligase. Again the integrity of the

Figure 22.

0.8% agarose gel electrophoresis of ligated Bgl II digested human DNA.

Track a: 3 μg of Bgl II digested human DNA treated with DNA ligase.

Track b: 3 μg of Bgl II digested human DNA.

Figure 22

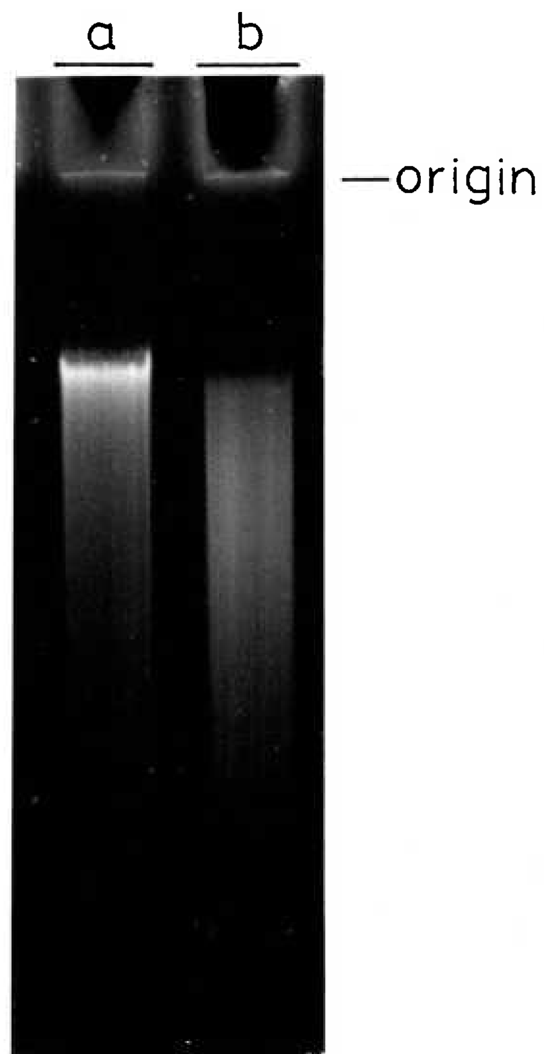


Figure 23.

1% agarose gel electrophoresis of ligated Bam HI digested pAT153 DNA.

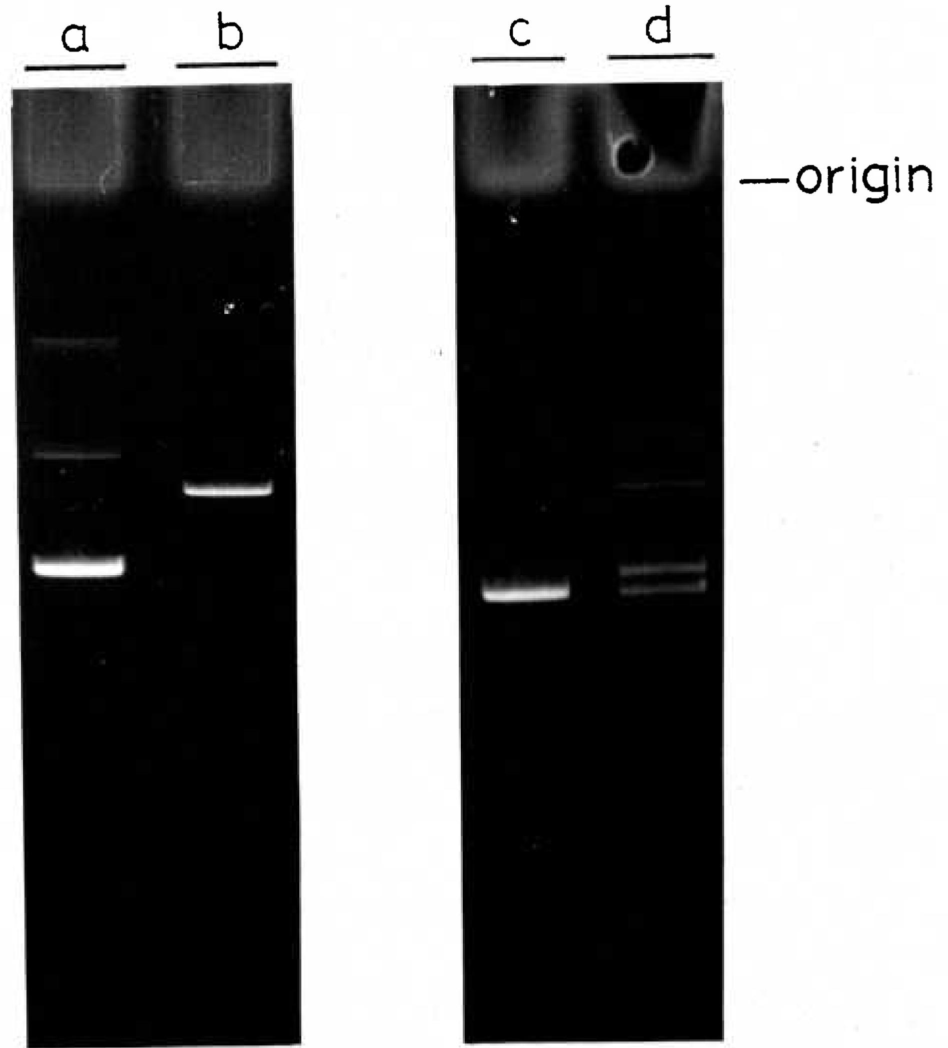
Track a: 0.3 μg of undigested pAT153 DNA.

Track b: 0.1 μg of Bam HI digested pAT153 DNA.

Track c: 0.2 μg of Bam HI digested pAT153 DNA.

Track d: As track c but with DNA treated with DNA ligase.

Figure 23



cohesive termini is demonstrated. The multiple forms of the plasmid after ligation are expected since conditions were not optimised for the formation of circles. It has been shown that the digested plasmid and human DNAs can be ligated each to themselves and so it should be possible to clone Bgl II fragments into Bam HI site in the plasmid using these preparations of DNA.

(c) Fractionation of human DNA by RPC-5 chromatography

1 mg of Bgl II digested DNA was fractionated by RPC-5 chromatography as described in 2.3 (k) using a 1.4 to 1.8 M sodium acetate gradient to elute the DNA. Forty five fractions were collected and ethanol precipitated. The precipitates were suspended in a constant volume of 10 mM Tris pH 7.4 and a constant sized aliquot from each fraction was run on a 1% agarose gel. in T buffer for a Southern blot to be carried out. Tracks containing unfractionated DNA were included and this is shown in Figure 24. The DNA in the gel was blotted onto a sheet of nitrocellulose following denaturation as described in 2.3 (f). Figure 25 shows the autoradiogram following hybridisation with nick translated pHγGl. Fractions 10 to 14 hybridise with the probe and of these fraction 11 hybridises most strongly. It was decided to clone DNA fragments from this fraction.

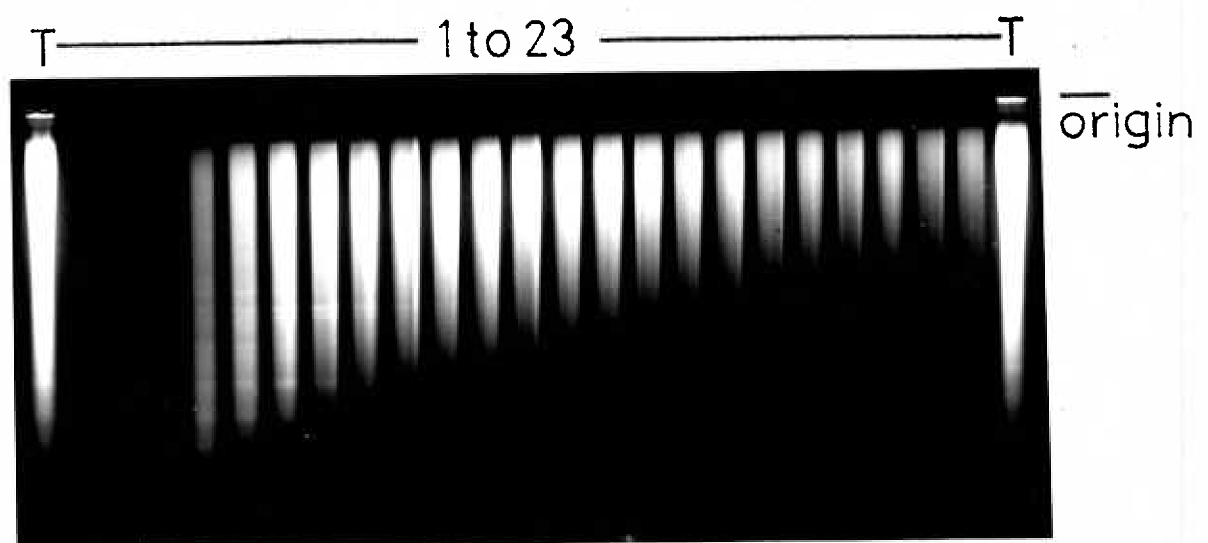
(d) Phosphatase treatment of linearised plasmid DNA and ligation

Attempting to ligate fragments of DNA into a restriction site in a plasmid could result in the reformation of parental molecules as well as recombinants. The parental molecules will transform the host bacterium more efficiently than the recombinant molecules and so re-circularisation of the vector must

Figure 24.

0.8% agarose gel electrophoresis of RPC-5 fractionated Bgl II digested human DNA. DNA from fractions 1 to 45 from the RPC-5 column was electrophoresed in the tracks indicated. The tracks T contain 20 µg of the unfractionated DNA.

Figure 24



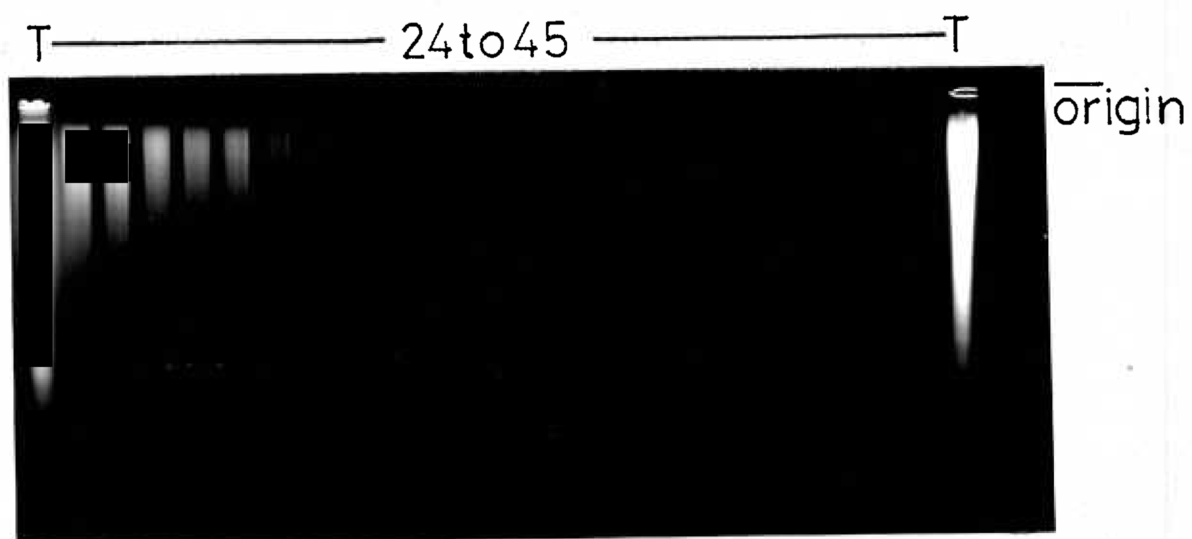
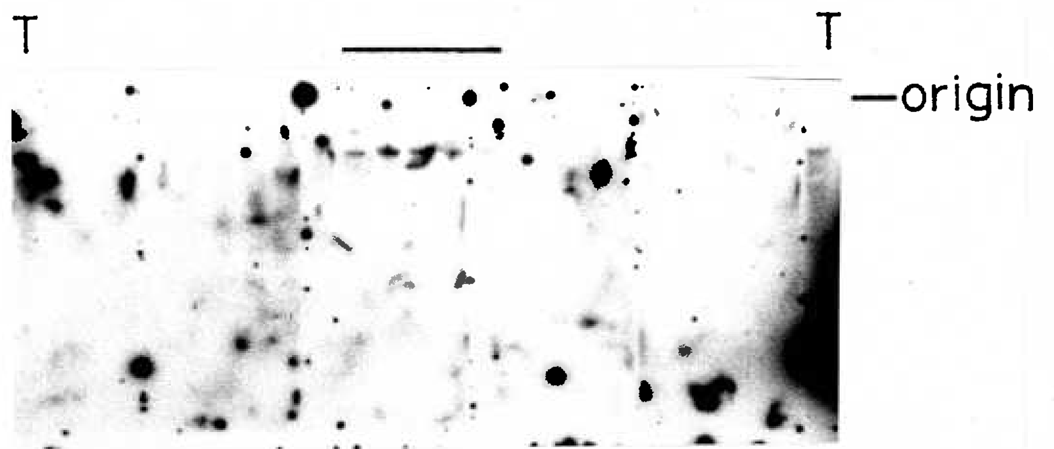
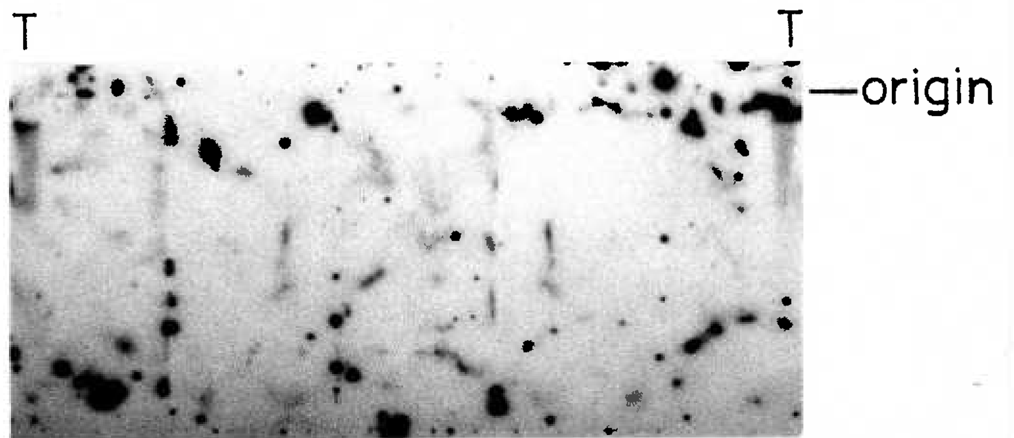


Figure 25.

Autoradiogram of Southern transfer hybridisation of the agarose gel in figure 24.
DNA was transferred from the agarose gel in figure 24 to nitrocellulose and was hybridised to ^{32}P labelled pHγGl DNA. The bar indicates hybridisation to fractions 10 to 14. Tracks T are as in figure 24.

Figure 25





be prevented to avoid a high background of non-recombinant clones.

The ligation reaction requires the presence of a phosphate group on the 5' terminus which is to be ligated to a 3' terminus. If the phosphates are enzymatically removed from the 5' termini of the linearised plasmid then the formation of circles is only possible with the inclusion of a piece of foreign DNA. In this case, the ligation reaction only ligates one strand at the junction between the plasmid and inserted DNA, but this is sufficient for the recombinant molecule to transform the host bacterium.

The linearised pAT153 DNA was incubated with bacterial alkaline phosphatase as described in 2.3 (r). Figure 26 shows 0.1 μg of the phosphatase treated vector DNA before and after ligation. Track d shows 0.1 μg of the phosphatase vector ligated to 0.5 μg of Bgl II digested human DNA. The intensity of the vector DNA band is markedly reduced compared to that in track c indicating that it ligates to the human DNA fragments. Microdensitometer scan of the photographic negative indicated that 60% of the plasmid DNA had ligated to the human DNA. This indicates that none of the vector DNA can ligate to itself after phosphatase treatment and that greater than 50% of it can ligate to inserted DNA.

10 ng of the ligated vector in track c and 10 ng of the vector ligated to 50 ng of Bgl II digested human DNA were used to transform E.coli HB101. This gave 2 and 153 transformants respectively indicating that there is a background of 1 to 2% non-recombinant transformants.

(e) Single colony lysates

To demonstrate directly that the recombinants did indeed

Figure 26.

0.8% agarose gel electrophoresis of ligated DNA.

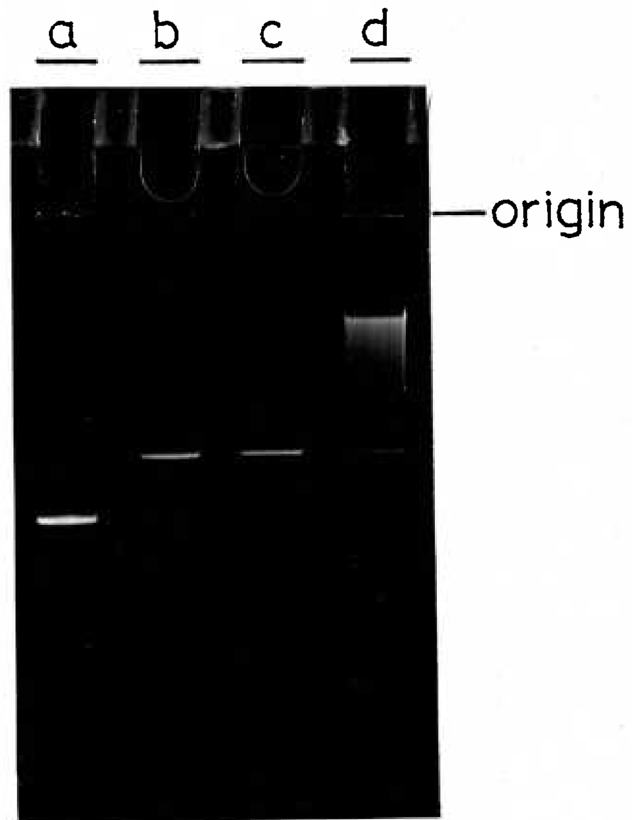
Track a: 0.3 μg of pAT153 DNA.

Track b: 0.1 μg of phosphatase treated Bam HI digested pAT153 DNA.

Track c: 0.1 μg of DNA as in track b treated with DNA ligase.

Track d: 0.1 μg of phosphatase treated Bam HI digested pAT153 DNA ligated with 0.5 μg of Bgl II digested human DNA.

Figure 26



contain inserted human sequences, lysates were prepared from single colonies of bacteria and analysed on agarose gels as described in 2.3 (s). 5 colonies were picked separately into a lysis mix containing SDS and proteinase K and incubated for 2 hours. The samples were split in two and to one half was added 50 ng of marker supercoiled pAT153 DNA. Figure 27 shows that in three cases the plasmid has picked up an insert whereas in the other two there appears to be no inserted sequence. The size of the inserted fragments in the three recombinant clones could not be determined as supercoiled DNA size markers were not available.

(f) Ligation of DNA for cloning

0.1 µg of linearised phosphatase-treated pAT153 was incubated with 0.25 µg of fraction 11 RPC-5 fractionated human DNA under different ligation conditions. In both cases 20 units of ligase μg^{-1} of vector were used but the DNA concentration and incubation time altered. For the incubation time of 16 hrs the vector DNA concentration was 5 $\mu g\ ml^{-1}$ and for the incubation time of 64 hours the vector DNA concentration was 15 $\mu g\ ml^{-1}$.

10 ng of ligated vector DNA from these incubations gave 11 and 113 transformants respectively when used to transform E.coli HB101 at an efficiency of transformation of greater than $10^6\ \mu g^{-1}$ with supercoiled pAT153. Figure 28 shows an agarose gel electrophoresis profile of the products of the two ligation reactions. Clearly the 64 hour ligation time increases the extent of ligation to the vector so reducing the intensity of the vector band in track c. Together these results show that the 64 hour incubation time gives a greater extent of ligation and a greater number of recombinants.

Figure 27.

0.8% agarose gel electrophoresis of DNA from single colony bacterial lysates.
The samples are run in pairs with the right hand sample containing 50 ng of supercoiled pAT153 DNA. The position of this DNA is indicated.

Figure 27

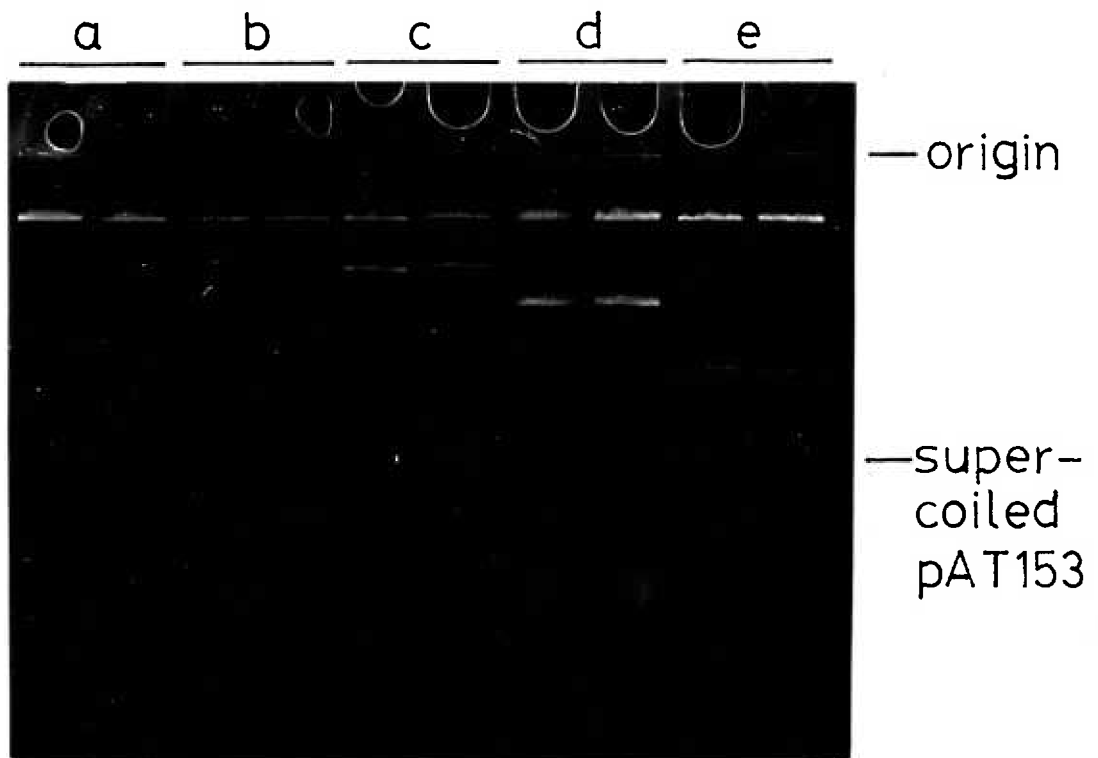


Figure 28.

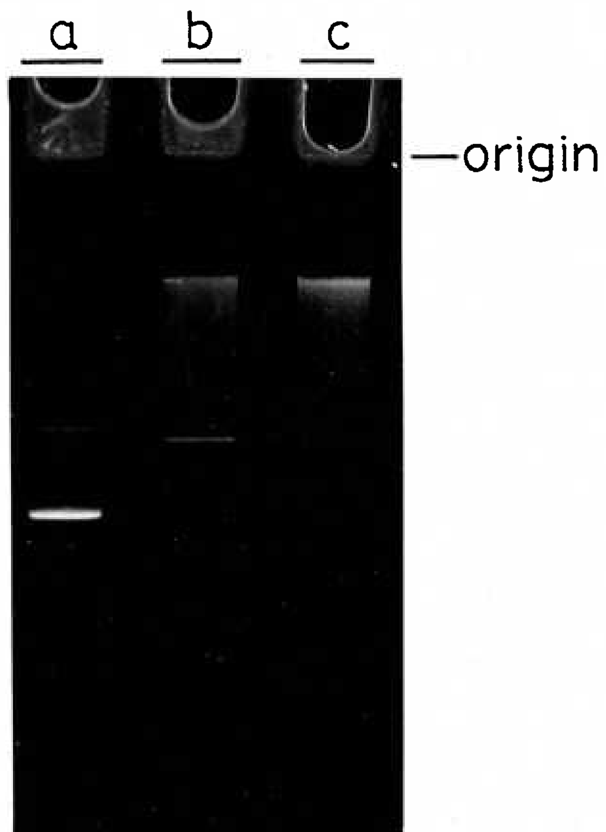
0.8% agarose gel electrophoresis of ligated DNA.

Track a: 0.3 µg of pAT153 DNA.

Track b: 0.05 µg of phosphatase treated Bam HI digested pAT153 DNA ligated with 0.125 µg of Bgl II digested human DNA for 16 hr at a DNA concentration of 5 µg ml $^{-1}$.

Track c: As track b but ligated for 64 hr at a DNA concentration of 15 µg ml^{-1}.

Figure 28



A series of ligation incubations were next set up with 20 ng of vector DNA and all other conditions as for previous 64 hour incubations. The ratio of inserted DNA to vector DNA was varied. The conditions and results are summarised in Table 2. After ligation each sample was split 1:9 to give 2 ng and 18 ng of vector ligated to insert respectively. Each was diluted to 100 μg with transformation buffer for E.coli HB101 and used for transformation in the usual manner.

Further ligations were carried out at an insert : vector weight ratio of 1:1. 1 μg of vector DNA was ligated with 1 μg of insert DNA for 64 hours with 40 units of ligase at a vector DNA concentration of 2 $\mu g\ ml^{-1}$. This DNA was used to transform E.coli HB101 and approximately 20 000 transformants were obtained at an efficiency of approximately $10^6\ \mu g^{-1}$ with supercoiled vector DNA.

The transformants were screened in duplicate as described in 2.3 (o) and 2.3 (p) with ^{32}P labelled pHγG1 as probe. No positive colonies were obtained.

Table 2.

Summary of transformations of *E.coli* HB101 with vector and insert DNA ligated to each other at different ratios.

Table 2

insert:vector (ng:ng)	number of transformants 2 ng vector	18 ng vector
5.1	8	12
2.5:1	8	7
1:1	6	18
0.5:1	10	9
0.1:1	0	1

Efficiency of transformation with supercoiled pAT153 = 6×10^4 μg^{-1}

Chapter 4

Discussion

4.1 The preparation and use of cloned cDNA probes

(a) cDNA vs cloned cDNA

The use of cDNA, prepared from mRNA, as a hybridisation probe has proved to be of great value in the past in the analysis both of the normal human globin genes and the globin genes in the various thalassaemias (Ottolenghi *et al*., 1974; Taylor *et al*., 1974; Lanyon *et al*., 1975; Ottolenghi *et al*., 1975; Kan *et al*., 1975; Tolstoshev *et al*., 1976; Ottolenghi *et al*., 1976; Ramirez *et al*., 1976a; Ramirez *et al*., 1976b; Old *et al*., 1976; Temple *et al*., 1977; Mitchell and Williamson, 1977; Godet *et al*., 1977). These studies were all possible despite the fact that α and β specific cDNAs were only 80 to 90% pure (Tolstoshev *et* al., 1976) and tedious to prepare. The preparation of α or β specific cDNAs relied on a continuing source of blood from patients with globin chain imbalances. With the advent of recombinant DNA techniques the possibility of cloning α, β and γ globin specific cDNAs became apparent. It was proposed to prepare double stranded cDNA, insert it into a bacterial plasmid and propagate it in *E.coli*. This would provide probes for globin DNA sequences that were 100% pure, the preparation of which would not rely on the sometimes erratic availability of pathological samples.

(b) The preparation and screening of the cDNA clones

Foetal globin mRNA was purified from total human reticulocyte RNA by oligo (dT) cellulose chromatography followed by sucrose gradient rate sedimentation. From this mRNA, cDNA was prepared using reverse transcriptase. The cDNA was double stranded with *E.coli* DNA polymerase I and the 3' termini extended

with poly (dA) homopolymer tails following treatment with S_1 nuclease. The plasmid pCR1 was cleaved at the single Eco RI site and the 3' termini were extended with poly (T) homopolymer tails. The tailed cDNA and tailed plasmid DNA were allowed to anneal and were used to transform E.coli HB101. Of the 282 transformants obtained, 58 were tentaively identified as containing γ-globin cDNA sequences, 15 as β-globin cDNA sequences and 15 as α-globin cDNA sequences. These assignations were by colony hybridisation using partially purified chain-specific globin mRNA probes (Grunstein and Hogness, 1975). Full details of the construction of the plasmids are given in Little et al. (1978).

A clone was tentatively identified as γ-globin cDNA by its differential hybridisation to mRNA-αβ and mRNA-αβγ which had been labelled in vitro. The conditions used for the alkali hydrolysis prior to labelling produce fairly large RNA fragments (about 100 bases). A higher specific activity could have been obtained if the mRNA had been hydrolysed further. This would have had the effect however of reducing the probe length even further below the size, optimum for hybridisation kinetics, of 400 bases (Rigby et al., 1977). The specific activity of approximately 10^7 dpm μg^{-1} was found to be adequate and this was achieved by using the labelled ^{32}P source carrier free. This method of preparing hybridisation probes for Grunstein-Hogness screening has also been adopted by other workers (Liebscher et al., 1979).

The preparation of the cDNA probes for the characterisation of the recombinant clone is described in 3.1 (b). The mRNA template was estimated to be 80 to 90% pure as judged by SDS/agarose gel electrophoresis. The synthesis of cDNA was according to the

method of Jackson et al. (1976) which favours long transcripts. Earlier methods yielded only partial transcripts (Young et al., 1974) but improvements have been made. Efstratiadis et al. (1975) and Kacian and Myers (1976) have reported that increased precursor nucleotide concentrations increase the average length of the cDNA. Monahan et al. (1976) showed that a short incubation time (5 min) and an increased incubation temperature (46°C) also yielded long transcripts. Buell et al. (1978) reported that optimum incubation conditions varied from one enzyme batch to another. The reaction conditions used followed the recommendations of Efstratiadis et al. (1975) and included 500 μM precursor nucleotides. Ribonuclease inhibitor was included to ensure the integrity of the mRNA template during the incubation thus favouring full length transcripts.

The hybridisation of cDNA-αβ and cDNA-αβγ to γ49DNA as described in 3.2 (c) verified that the plasmid did contain a γ-globin cDNA sequence. The 25% level of hybridisation indicated that the insert was near to full length and this was later confirmed, by other workers, by the direct excision of the inserted sequence from the plasmid (Little et al., 1978). Plasmid DNA was incubated under conditions where the homopolymer (dA) and (T) regions were preferentially melted. Only these regions were digested by S_1 nuclease, and the inserted sequence was judged to be 500 base pairs in length following agarose gel electrophoresis to separate the inserted sequence from the plasmid DNA.

(c) The use of the globin cDNA clones

The cloning operation which produced pHγG1 also yielded pHαG1 and pHβG1, α and β globin specific cDNA plasmids

respectively. The three plasmids have proved most useful in the determination of the physical maps around the globin gene loci. The plasmid pHβG1 has been used to map the δβ locus (Flavell et al., 1978) and pHγG1 has been used to determine the structure of the γ gene locus (Little et al., 1979a). A combination of the use of both plasmids as probes has lead to the linking of the two maps to give a map for the γγδβ locus (Bernards et al., 1979). The map of the α-gene locus has been determined in other laboratories using cDNA probes (Surrey et al., 1978; Orkin, 1978). The map has been confirmed in this laboratory using the α globin cDNA plasmid JW101 prepared by Wilson et al. (1978). A recombinant plasmid containing γ-globin cDNA sequences has also been constructed by Humphries et al. (1978).

As well as being used to establish the physical map around the globin genes, the plasmids have been used to study the genes in some of the thalassaemias, the genes of patients with haemoglobin chain variants and to study polymorphisms in restriction sites. Physical maps for the β-globin genes in patients with β thalassaemia have been derived with pHβG1 (Flavell et al., 1979). Analysis of the γ-globin genes, with pHγG1, in patients with $HbF_{Sardinia}$ have shown that the gene coding for the globin variant is allelic with one of the known γ-globin genes (Little et al., 1979b). Studies of restriction site polymorphisms around the β like globin genes have been carried out using the plasmids pHβG1 and pHγG1 (Jeffreys, 1979).

The construction of recombinant plasmids provides a readily available and reproducible source of gene specific probe. Plasmid probes are versatile and lend themselves easily to the preparation of probes for specific regions of the gene,

such as 5' or 3' specific probes.

There is the disadvantage however that strand specific probes can only be prepared with difficulty from the double stranded plasmid DNA. Where strand specific probes are required cloning in a single stranded vector becomes an attractive possibility. The phage M13 is presently being developed as such a vector.

The use of a cloned cDNA probe does not automatically eliminate the cross hybridisation problems encountered with cDNA probes. Although the problems are substantially reduced, care has to be taken to ensure an adequate stringency of hybridisation to provide unequivocal results. The plasmids pHγGl and pHβGl cross hybridise with ε-globin gene sequences as well as pHγGl cross hybridising with β-globin sequences and pHβGl cross hybridising with γ globin sequences.

4.2 The cloning of the γ-globin genes

(a) Mapping the γ-globin genes

Preliminary work indicated that the human γ-globin genes were contained in a single Bgl II restriction fragment (R. Flavell, personal communication). This was verified in the experiments described in 3.3 (d). Results published later were also in agreement with this and gave the fragment length as 13 kb (Little *et al*., 1979a) or 13.5 kb (Tuan *et al*., 1979). The former value is used in this study.

No formal proof existed that the γ-globin genes did not exist in two separate 13 kb fragments. This possibility was excluded by the use of RPC-5 chromatography which is able to separate DNA fragments of exactly the same length according to other parameters. These parameters include base composition and the distribution of bases, particularly the clustering of As and Ts, within a fragment (Hardies and Well, 1976; Larson *et al*., 1979; Patient *et al*., 1979). The experiments in 3.3 (e) indicate that there is a single 13 kb DNA fragment encoding the γ-globin genes. It is not possible to exclude the possibility that the entire locus contained in the 13 kb fragment is duplicated. Other evidence demonstrates, however, that there are not four γ-globin genes per haploid chromosome (Old *et al*., 1976; Ramirez *et al*., 1976b; Mitchell and Williamson, 1977).

(b) The preparation of Bgl II

Preliminary attempts were made to purify Bgl II by means of a single step method described by Bickle *et al*., (1977). This method gave low yields of rather impure enzyme and so standard methods of gel filtration and ion exchange chromatography were employed instead. The enzyme prepared by this

latter method, as described in 3.4 (b), appeared initially to be of adequate purity for the proposed study. The results described in 3.5 (b) indicate, however, that there was either exonuclease or phosphatase contamination present in the enzyme. Each would result in the inability to ligate DNA fragments generated with this enzyme preparation. Fractions eluted from the DEAE - cellulose column should have been assayed for the ability to generate DNA fragments that could be ligated. Alternatively the enzyme could have been further purified by phosphocellulose chromatography. Had the Bgl II been further purified, advantage could have been taken of the ability to clone Bgl II DNA fragments by ligating them into the Bam HI site of the plasmid pBR322.

(c) Purifying DNA for cloning

The purification strategy for this study was to first fractionate, by preparative agarose gel electrophoresis, 40 mg of Bgl II digested DNA. This was carried out on a "Gene Machine" (Southern,1979) which has the advantage of large capacity over other types of gel electrophoresis size fractionation. The method of Tabak and Flavell (1978) for the recovery of DNA from agarose gels gives equally good fractionation but the system has only a relatively low capacity. This is unfortunate since the method is considerably more convenient than the use of the "Gene Machine" which is prone to a number of technical problems. Poor fitting of the dialysis membrane at the anode can result in loss of DNA and non-central placement of the anode causes the DNA to elute in diffuse size fractions. The "Mighty Machine" (Polsky *et al.*, 1978) circumvents some of the problems by using a slab gel and flat membrane. Tube gels for the

preparative fractionation of both DNA and RNA have also been described (Hagen, 1979; Bostian et al., 1979).

Other size fractionation techniques such as sucrose gradients are unlikely to give a comparable purification. Sucrose gradients are useful for the removal of small DNA fragments; these, when ligated into a plasmid vector, give rise to recombinant molecules which transform host cells at a higher efficiency than plasmids containing large inserts. This can be a problem when attempting to clone large DNA fragments (Sherratt, 1979). In general, sucrose gradients suffer from the disadvantage of low capacity.

The fractions from the "Gene Machine" that hybridised with pHγG1 were pooled. 400 μg of DNA was recovered and fractionated by RPC-5 chromatography. Because the separation is based on a number of properties rather than a single one it has found favour as a purification technique in cloning (Tilghman et al., 1978 and Woo et al., 1978). The fractionation profile of DNA on RPC-5 is quite unlike that of size fractionation techniques and so the two methods can be used together to advantage. Alternative novel methods of fractionation have been developed with base-pair specific dyes polymerised into acrylamide gels (Bunemann and Muller, 1978). The fractionation however is not as good in our hands as with RPC-5 (R. Elles, personal communication). Recently a method has been developed for the fractionation of DNA using countercurrent two-phase partition chromatography (Sutherland and Ito, 1978). The general usefulness of this technique is doubtful due to the low degree of purification that can be achieved (Elles and Sutherland, 1979). Other methods exist for the purification of specific DNA sequences but are mostly applicable to single stranded DNA. Some of the methods are

reviewed in the attached paper (Courtney, Dalgleish and Williamson, 1979).

RPC-5 remains the most versatile fractionation medium for both double and single stranded DNA (Wells *et al.*, 1979) and often may be used for a single fractionation step (Tiemeier *et al.*, 1977).

Fractionation of the DNA by gel electrophoresis followed by RPC-5 chromatography provided a sufficient degree of purification for the screening of recombinant clones to be practical, using a plasmid vector cloning system with the methodology available in mid-1978. The development of screening methods, where upwards of 2 000 colonies could be screened on a 90 mm nitrocellulose filter, reduced the required degree of DNA puri-ficiation. This is reflected in the methods of purification described in 3.6 and carried out in 1979.

(d) The choice of host/vector systems

When the proposed cloning experiments were notified to GMAG in March 1978 the advice given was that the work should be carried out at the Category III level of physical containment with a disabled host/vector system. The containment laboratory is described in 3.5 (a) and little variation, except in detail, exists between different Category III laboratories.

A number of alternative host/vector systems were available. The choice of vector lay between derivatives of phage λ and bacterial plasmids. No λ vectors for cloning fragments of DNA generated by Bgl II or Bam HI existed until recently. This compelled the use of a plasmid vector. Of the plasmid vectors available, pBR322 was most versatile because of its small size, sites for inserting DNA and high copy number after amplification. No biological containment (disablement) is provided as such by

pBR322. This is conferred by using the plasmid in conjunction with the host *E.coli* χ1776. The use of *E.coli* χ1776 poses many problems not the least of which is simply growing the bacteria. Special preparation techniques have had to be developed for its transformation (Norgard *et al.*, 1978; D. Hanahan, unpublished observations). The longer generation time, compared with strains such as HB101, make it inconvenient to use as transformants cannot be detected until about 40 hours after transformation.

E.coli χ1776/pBR322 was used in 3.5 for cloning but at a later date the regulations governing genetic recombination experiments were changed and also new and improved host/vector systems were developed. The reduction to Category I* for physical containment made the manipulations easier and the acceptance of the new host/vector systems as being disabled reduced or eliminated many of the problems associated with the use of *E.coli* χ1776.

New plasmid vectors were constructed that did not require a disabled host for biological containment: pACYC184 (Chang and Cohen, 1978) and pAT153 (Twigg and Sherratt, 1979). The biological containment of these systems resides in the inability of the vector plasmid to undergo mobilisation by another plasmid and in this way transfer from its host to another bacterium.

A Bam HI λ insertion vector has been developed (B. Klein, personal communication) but it cannot accommodate DNA fragments as large as 13 kb.

It is now possible to insert DNA into a λ vector by homopolymer tailing. Previously this was not possible as the tailing reactions made the resulting λDNA non-infective by standard transformation procedures. The advent of the packaging of λDNA *in vitro* circumvents these problems (Hohn and Hohn, 1974; Hohn

and Murray, 1977). The λ DNA is first circularised by incubation with DNA ligase, then linearised by digestion with the appropriate restriction enzyme. A central fragment unnecessary for phage function is separated and discarded, and the ligated vector arms have homopolymer tails added to their 3' termini. A DNA fragment with complementary tails may now be annealed with the tailed vector. Incubation of the recombinant molecules with the λ packaging proteins regenerates the λ cohesive ends which are essential for infection, and the DNA is packaged into a virus particle.

Another possible cloning vector is a cosmid (Collins and Hohn, 1978). This is a plasmid type vector that has the λ cos cohesive termini. Recombinant molecules of the correct size can thus be packaged into λ phage heads in vitro as already described. No Bam HI or Bgl II cosmid vector is available at present for the cloning of fragments of DNA 13 kb in length though there is one, designated pJC75-78, that would be suitable for larger fragments (Collins and Hohn, 1978).

Of the vectors described pBR322, pAT153 or pACYC184 used in conjunction with a suitable host strain of E.coli provide, at this time, the simplest and most convenient, if not the most versatile and elegant, of cloning systems. Part of this study was carried out using E.coli χ1776/pBR322 and the later experiments using E.coli HB101/pAT153.

(e) Homopolymer tailing

For reasons already discussed it was necessary to insert the purified DNA into pBR322 by means of homopolymer tails in the first cloning experiments described in 3.5. Conditions had already been established in this laboratory for the addition of

poly (dC) and poly (dG) tracts to 3' termini so G and C tailing rather than A and T tailing was adopted. The polymerisation of the tails is described in 3.5 (d) and 3.5 (e). The concentration of dGTP required to tail the inserted DNA was higher than that of the dCTP required to tail the linearised vector. This observation is consistent with those of other workers (D. Woods, personal communication).

The calculation of tail length was in each case based on the incorporation of radioactive counts into acid precipitable material. Knowing the starting concentration of both termini and substrate nucleotide it is possible to calculate tail length. The calculation is based on the premise that the incorporation of labelled deoxynucleotide occurs only at the termini. In practice the DNA will contain some single stranded nicks which will also act as a primer for the tailing reaction (Jackson et al., 1972). The pBR322 DNA used in this study contained a portion of material with single stranded nicks as can be seen in Figure 17, track (a). While it is likely that this resulted in the calculated tail length being high, it is not possible to estimate the degree of error in the calculation.

It has been observed that, under certain conditions, terminal transferase will catalyse the polymersation of homopolymer tracts in the absence of a primer. These tracts would be acid precipitable and would result in an artifially high estimation of tail length. To minimise this effect the reactions were carried out in cacodylate buffer for short incubation times (Kato et al, 1967).

During this study it was found that plasmid DNA prepared by the acidic phenol extraction method of Zasloff et al. (1978) was a poor primer for the tailing reaction compared with plasmid DNA prepared by CsCl/ethidium bromide buoyant density

centrifugation. This effect is probably because the former method does not remove the Triton X-100 detergent used early in the purification of the DNA. This highlights the very fastidious nature of the terminal transferase.

(f) Cloning with tailed DNA

The tailed vector and insert DNAs used in 3.5 (g) were allowed to anneal by incubating at 65°C for 2 hours following which the mix was allowed to cool to room temperature over a period of 5 hours. The DNA was at a concentration of 200 ng ml^{-1} with an equimolar ratio of insert to vector. These conditions have been demonstrated to give recombinant molecules (Ch. Coutelle, personal communication). Little information exists for the optimum annealing conditions for G and C tails. Chang et al. (1978) annealed the DNA at a much lower concentration and ratio of insert to vector was not given. In their case the inserted sequence was much shorter than 13 kb (about 1.5 kb).

Neither the tailed insert DNA or tailed plasmid DNA gave transformants with E.coli, when used alone, with amounts that would allow the detection of transformants at a level of 1% relative to supercoiled vector DNA. When annealed, the two DNAs gave the transformants shown in Table 1.

Of the 1817 transformants obtained, approximately 1500 were screened as described in 3.5 (h). Of these, more than ten gave strong positive scores. In two subsequent screens twelve of these gave negative scores. Three of the twelve, 4G6, 6E7 and 6F2 were analysed and found to have intact Hind III sites in their plasmids and no inserted sequences.

An explanation for this is that they arose from uncut plasmid vector DNA that was not detected in the control experiments.

Transformants arising from newly constructed recombinant molecules grow more slowly than those from non-recombinant molecules (own unpublished observations) and so there is an initial selection for non-recombinant clones. All further growth of clones was under antibiotic resistance selection. If recombinant clones grew more slowly because of the insert in the plasmid then there would again be selection for non-recombinants. The fact that 4G6, 6E7 and 6F2 were not recombinant clones is not inconsistent with the experimental procedures used.

(g) Cloning by ligation

The original method proposed for the formation of recombinants is potentially less fraught with technical problems. The enzymes used, Bgl II and Bam HI, were demonstrated to give DNA fragments that were ligatable. The only other required precaution was to treat the Bam HI digested vector DNA with phosphatase and show that this DNA could be ligated to insert DNA which had not been treated with phosphatase, but not recircularised by ligation in the absence of added DNA.

Following phosphatase treatment of the linearised vector it was shown that upon ligation there was a background of approximately 1% compared with supercoiled DNA in a transformation assay. Greater than 50% of this vector was capable of ligating to other DNA sequences with the appropriate cohesive termini.

With hindsight it would have been preferable to have further reduced the background by about a factor of 10. The

reason for this is that the number of transformants with recombinant DNA was only approximately 2% of that with supercoiled DNA. This means that only half of the transformants obtained were recombinants, which is consistent with the observation in 3.6 (e) that 2 lysates from 5 showed no inserted sequence in the plasmid. Thus twice as many transformants have to be screened than would have otherwisé been necessary.

The effect of varying the ratio of insert:vector was not as expected. It was expected that there would be an optimum ratio for the formation of recombinants. Either side of this ratio the formation of concatenates of either vector or insert would be favoured and this would be manifest in a reduction in the number of transformants. Table 2 shows that if there is an optimum it is very wide. Certainly the efficiency of transformation is markedly reduced at an insert:vector ratio of 0.1. In the absence of any indication of a true optimum, a ratio of 1:1 was used in all further studies.

The results in Table 2 indicate that approximately the same numbers of transformants were obtained whether 2 ng or 18 ng of DNA were used in the transformations. When larger amounts of DNA were used to produce the 20 000 transformants that were screened this effect was not observed. The first result is consistent with the observations of other workers (D. Sherratt, personal communication). The reasons why the effect is not observed with larger amounts of DNA are not apparent.

An equation exists to describe the DNA concentration at which the formation of circles and concatenates is balanced for a single participating species in a ligation reaction (V. Sgaramella, personal communication). This does not cover the situation, however, where two species participate and only

one is able to ligate to itself. The best approach to problems of optima is empirical and this is the approach that is being investigated by interested commercial organisations.

Approximately 20 000 transformants were screened without finding a clone that hybridised with pHγG1. Perhaps up to 50% of these were non-recombinants and so only 10 000 recombinants were screened.

Estimations of the number of transformants that have to be screened to find a positive colony can be made. There is an equation to calculate the number of transformants required to represent the entire genome of an organism with any given probability (Clarke and Carbon, 1976). The equation is:

$N = \frac{\ln (1 - P)}{\ln (1 - f)}$ where N is the number of transformants

P is the probability factor

f is the fraction of the genome represented by each cloned fragment (assuming an average size).

The haploid human genome is 3×10^9 bp. Statistically, the average size of a cloned fragment will be 4.1 kb (i.e. the Bgl II site will occur once in 4^6 bases). The value for f becomes 1.4×10^{-6} (i.e. $\frac{4.1 \times 10^3}{3 \times 10^9}$). For a 99% probability of total representation of the genome, and hence the cloning of the γ specific Bgl II fragment, the equation becomes:

$$N = \frac{\ln (1 - 0.99)}{\ln (1 - [1.4 \times 10^{-6}])} = 3.3 \times 10^6.$$

This number of transformants would have to be screened if no purification of the DNA was carried out prior to cloning RPC-5 chromatography affords about 10-fold purification and so the number of transformants to be screened becomes 3.3×10^5.

For a 90% probability the number of transformants to be screened becomes 1.6×10^5.

These calculations indicate that many more transformants would have to be screened to ensure the isolation of the γ-specific Bgl II fragment with any certainty.

Two accounts of the cloning of γ-globin genomic DNA sequences have been published. Blattner _et_ _al_. (1978) and Smithies _et_ _al_. (1978) have cloned the 2.7 kb Eco RI encoding the 3' end of the $^{A}\gamma$-globin gene in the λ vector Charon 3A. The entire γ-globin genes locus has been cloned by insertion of Eco RI partial digestions of DNA into the λ vector Charon 4A (Ramirez _et_ _al_., 1979).

4.3 Cloned globin DNA sequences - future prospects

(a) Sequence comparisons and hybridisation studies

It is apparent that the cloning of DNA sequences will play a central role in the elucidation of gene control mechanisms.

The most direct approach to the problem is to compare cloned DNA sequences from both normal individuals and those with some recognisable defect. This may be of limited value as there may be differences in the DNA sequences that are not associated with the defect. Sequence polymorphisms have been demonstrated in the globin genes of normal individuals (Jeffreys, 1979). Such comparisons within coding regions would only be conclusive if they identified base change that resulted in a changed amino acid or, say, the introduction of a stop codon. A defect of this type has recently been identified by comparing mRNA sequences from a normal patient and a β^{O}-thal patient. A base change had occurred that introduced a stop codon at amino acid 17 (Chang and Kan, 1979). Such an analysis could equally well be performed on cloned genomic DNA sequences.

For sequences outside the coding region such comparisons are not always valid since base changes may be phenotypically silent. It would not be possible to distinguish easily whether or not a given base change was the cause of the phenotypic change.

Advantage may be taken of base changes in other ways. A single base change may either create or destroy a restriction endonuclease recognition site and this phenotypically silent mutation may in turn be genetically linked, either functionally or by chance, with a phenotypically recognisable mutation such as sickle cell anaemia. It has been shown that alteration of the pattern of Hpa I restriction endonuclease sites has a close

correlation with the sickle cell allele of β-globin (Kan and Dozy, 1979) and this has been used for ante-natal diagnosis of the sickle cell gene in a foetus at risk (Kan and Dozy, 1978). It has also been demonstrated that there are polymorphisms of Hsu I restriction sites associated with the β^{o}-thal gene in some Greek-Cypriots (P. Little, personal communication).

It should be noted that these studies were possible without recourse to the cloning of genomic DNA though they did utilise cloned cDNA sequences.

The implications of the discovery of these polymorphisms linked to globin gene defects are far reaching. It becomes theoretically possible to probe for a given gene defect by way of a linked DNA sequence polymorphism even if the defective gene itself cannot be probed directly. Thus, the exact nature of the gene defect responsible for a given clinical condition need not be identified to allow diagnosis.

This has applications in ante-natal diagnosis. Nowadays amniocentesis poses little risk to foetus or mother and provides amniotic fluid cells from which DNA can be readily prepared. Using this DNA, diagnoses of the type outlined above could be carried out.

A library of cloned genomic sequences could conceivably provide probes for a large range of clinical conditions.

Gross deletions in DNA sequences are, in general, better suited for comparative studies. In the first instance these studies need not require the cloning of genomic sequences though fine mapping would require cloned sequences. Fritsch et al. (1979) have analysed the DNA from patients with HPFH and $\delta\beta^{o}$-thal. They have shown that the extent of the deletions in these conditions correlates with the expression of the γ-globin genes in the adult patient. This analysis, by Southern transfer

hybridisation, used a cloned DNA sequence from between the $^{A}\gamma$ and δ genes as a probe. The results indicate a region to the 5' side of the δ gene that may control expression of the γ genes. The authors admit that they cannot exclude, however, the possibility that the γ genes are dereppressed by another linked mutation that was not detected in the analysis.

Studies of cloned sequences around the β-like locus of these patients would locate precisely the extent of the deletions and could perhaps shed light on the mechanism of the switch from foetal to adult globin synthesis.

Cloning will also provide probes that will allow a more precise study of hnRNA processing. Intron regions are transcribed along with exons to give a nuclear precursor to mRNA and so the intron must be removed during the maturation process. The β^{o}-thal patient studied by Comi et al. (1977) appears to be defective in β-globin hnRNA processing or transport and may provide a good system for the study of processing. Analysis of the genomic DNA sequences around the intron-exon junctions may shed light on the defect. The cloning of the β-globin gene from this patient is currently being undertaken in this laboratory.

The development of a technique for the transfer hybridisation of RNA, in an analogous manner to that for DNA, has simplified the analysis of mRNA precursors (Alwine et al., 1977). A precursor to γ-globin mRNA has recently been identified, in preparations from human foetal liver, using this technique (Courtney and Williamson, 1979). The precursor is polyadenylated and is probably the transcribed RNA prior to exision of both intron sequences. A full analysis of the precursor awaits the availability of cloned probes for the intron regions and

and the regions immediately flanking the gene. The analyses carried out were with pHγG1 as the probe.

(b) Expression of cloned sequences

In eukaryotes transcription is accomplished by three RNA polymerases. *In vitro* transcription of exogenous DNA has been demonstrated with polymerase III but not as yet with any fidelity, with polymerase II, which transcribes unique gene sequences including the globin genes.

Greater success has been achieved with the expression *in vivo* of recombinants containing eukaryotic sequences. Globin genomic DNA sequences from mouse have been cloned in the monkey virus SV40 and these sequences expressed in infected monkey cells in culture (Hamer and Leder, 1979). The expression of rabbit globin genomic DNA sequences has been described in mouse thymidine kinase$^-$ (tk$^-$) cells which are transformed with the sequences joined to the cloned herpes simplex virus tk gene (Mantei *et al.*, 1979) or physically mixed but not chemically joined to the tk gene (Wigler *et al.*,1979). In all three cases the globin genes were transcribed and in the first case β-globin was detected.

The SV40 expression system has, so far, yielded the most information but is limited by the size range of sequences that may be cloned and expressed. In certain respects the tk co-transformation system (Wigler *et al.*, 1979) offers greater promise for *in vivo* investigations.

Site specific mutagenesis *in vitro* (Hutchinson *et al.*, 1978; Shortle and Nathans, 1978; Muller *et al.*, 1978) allows the opportunity to study the effect of base changes in cloned DNA sequences. Mutated globin gene sequences could be prepared

and the effect of the mutation studied *in vivo*. This might provide further information on the control of gene expression.

The study of cloned globin genes in human cells ought to be possible. Transformation of tk^- HeLa cells with the cloned herpes tk gene has been accomplished and co-transformation should be also possible. Expression, in an homologous system, of cloned eukaryotic sequences would hopefully provide great insights into the control of gene expression.

Co-transformation of cells in culture may, one day, have a therapeutic use. Erythroid stem cells from thalassaemic patients could be transformed with a functional copy of the gene in which the patient has a defect. Thus the thalassaemic gene is supplemented by a normal gene and the cells are returned to the marrow where they will hopefully proliferate and reduce the effects of thalassaemia.

A clone of the γ-globin Bgl II genomic fragment would provide a series of hybridisation probes to allow the detailed study of both gene structure and RNA processing. The cloned fragment could be studied in whole or in part in expression systems to elucidate the control mechanisms that govern the expression of the $^{G}\gamma$ and $^{A}\gamma$ genes and the switch from foetal to adult globin synthesis.

BIBLIOGRAPHY

Alt, F.W., Kellens, R.E., Bertino, J.R. and Schimke, R.T. (1978). J. Biol. Chem. 253, 1357-1370.

Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5350-5354.

Arber, W. and Dussoix, D. (1962). J. Mol. Biol. 5, 18-36.

Ashby, Lord (1974). Report of the Working Party on the Experimental Manipulation of the Genetic Composition of Microorganisms. Cmnd. 5880, H.M.S.O., London.

Aviv, H. and Leder, P. (1972). Proc. Natl. Acad. Sci. U.S.A. 69 1408-1412.

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

Bantle, J.A. and Hahn, W.E.(1976). Cell 8, 139-150.

Bellevue, R. Dosik, H. and Rieder, R.F. (1979). Brit. J. Haem. 41, 193-202.

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

Benz, E.J. and Forget, B.G. (1971). J. Clin. Invest. 50, 2755-2760.

Benz, E.J., Forget, B.G., Hillman, D.G., Cohen-Solal, M., Pritchard, J., Cavallesco, C., Prensky, W. and Housman, D. (1978). Cell 14, 299-312.

Berget, S.M., Moore, C. and Sharp, P.A. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 3171-3175.

Bernard, O., Mozumi, N. and Tonegawa, S. (1978). Cell 15, 1133-1144.

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

Bickle, T.A., Pirrotta, V. and Imber, R. (1977). Nucl. Acids Res. 4, 2561-2572.

Birnstiel, M.L., Gross, K., Schaffner, N. and Telford, J. (1975). Febs. Proc. 38, 3-24.

Bishop, J.O. (1974). Cell 2, 81-86.

Bishop, J.O. and Freeman, K.B. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 707-716.

Bishop, J.O., Pemberton, R. and Baglioni, C. (1972). Nature New Biol. 235, 231-234.

Bishop, J.O. and Rosbach, M. (1973). Nature New Biol. 241, 204-207.

Bishop, J.O., Rosbach, M. and Evans, D. (1974). J. Mol. Biol. 85, 75-86.

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, 1279-1284.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, 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.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977). Gene 2, 95-113.

Bollum, F.J. (1967). Methods in Enzymology, L. Grossman and K. Moldave eds. vol. 21, p. 350. (Academic Press, New York).

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

Bostian, K.A., Lee, R.C. and Halvorson, H.O. (1979). Anal. Biochem. 95, 174-182.

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

Brack, C., Hirama, M., Lenhard-Schuller, R. and Tonegawa, S. (1978). Cell 15, 1-14.

Brammer, W.J. (1979). Biochem. Soc. Symp. 44, 13-27.

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

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

Britten, R.J. and Kohne, D.E. (1968). Science 161, 529-540.

Britten, R.J. and Smith, J., (1970). Carnegie Trust Washington Year Book, 68, 378.

Buell, G.N., Wickens, M.P., Payvar, F. and Schimke, R.T. (1978). J. Biol. Chem. 253, 2471-2482.

Bunemann, H. and Muller, W. (1978). Nucl. Acids Res. 5, 1059-1074.

Capp, G.L., Rigas, D.A. and Jones, R.T. (1967). Science 157, 65-66

Capp, G.L., Rigas, D.A. and Jones, R.T. (1970). Nature 228, 278-28

Carroll, D. and Brown, D.D. (1976a). Cell 7, 467-475.

Carroll, D. and Brown, D.D. (1976b). Cell 7, 477-486.

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

Chamberlin, M.E., Britten, R.J. and Davidson, E.H. (1975). J. Mol. Biol. 96, 317-333.

Chang, A.C.Y. and Cohen, S.N. (1978). J. Bact. 134, 1141-1156.

Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T. and Cohen, S.N. (1978). Nature 275, 617-624.

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

Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977). Cell 12, 1-8.

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

Clegg, J.B. and Weatherall, D.J. (1976). Brit. Med. Bull. 32, 262-269.

Clewell, D.B. (1972). J. Bact. 110, 667-673.

Clewell, D.B. and Helinski, D.R. (1972). J. Bact. 110, 1135-1146.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W. and Helling, R.B., (1973). Proc. Natl. Acad. Sci. U.S.A. 70, 3240-3244.

Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972). Proc. Natl. Acad. Sci. U.S.A. 69, 2110-2114.

Collins, J. and Hohn, B. (1978). Proc. Natl. 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., Rowley, P.T., del Senno, L., Pontremoli, S. and Volpato, S. (1972). Nature New Biol. 238, 83-87.

Courtney, M., Dalgleish, R. and Williamson, R. (1979). Proceedings of 12th FEBS Meeting, Dresden, 1978. S. Rosenthal et al. eds. (Pergamon Press, Oxford).

Courtney, M. and Williamson, R. (1979). Nucl. Acids Res. 7, 1121-1130.

Covey, C., Richardson, D. and Carbon, J. (1976). Mol. Gen. Genet. 145, 155-158.

Crick, F.H. (1971). Nature 234, 25-27.

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

Curtiss, R. III, Pereira, A., Hsu, J., Hull, C., Bark, E., Maturin, Sr. J., Goldschmidt, R., Moody, R., Inoue, M. and Alexander, L. (1976). Micro 76 - The annual report, Department of Microbiology, University of Alabama in Birmingham, 96-106.

Danna, K. and Nathans, D. (1971). Proc. Natl. Acad. Sci. U.S.A. 68, 2913-2917.

Davidson, E.H. and Britten, R.J. (1973). Quart. Rev. Biol. 48, 565-613.

Davidson, E.H., Galau, G.A., Angerer, R.C. and Britten, R.J. (1975 Chromosoma 51, 253-259.

Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973). J. Mol. Biol. 77, 1-23.

Davis, R.W., Simon, M. and Davidson, N. (1971). Methods in Enzymology, L. Grossman and K. Moldave eds. vol. 20, pp. 413-428. (Academic Press, New York).

Dayhoff, M.O. (1972). Atlas of Protein Sequence and Structure, 5. (National Biomedical Research Foundation, Washington D.C.).

Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. and Ruddle, F.H. (1978). Proc. Natl. Acad. Sci. U.S.A. 75, 1456-1460.

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W.R., Ruddle, F., Lawrence, J., Creagen, R. and Kucher Lapati, R. (1977). Cell 12, 205-218.

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

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.

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

Dussoix, D. and Arber, W. (1962). J. Mol. Biol. 5, 37-49.

Efstratiadis, A., Maniatis, T. Kafatos, F.C., Jeffrey, A. and Vournakis, J.N. (1975). Cell 4, 367-378.

Elles, R.G. and Sutherland, I.A. (1979). Biochem. Soc. Trans. (in press).

Etcheverry, T., Colby, D. and Guthrie, C. (1979). Cell 18, 11-26.

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

Flavell, R.A., Jeffreys, A.J. and Grosveld, G.C. (1978b). 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. (1978a). Cell, 15, 25-41.

Forget, B.G. (1978). Trends in Biochem. Sci. 3, 86-90.

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.

Frischauf, A.M., Lehrach, H., Rosner, C. and Boedtker, H. (1978). Biochemistry 17, 3243-3249.

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

Gage, L.P. and Manning, R.F. (1976). J. Mol. Biol. 101, 327-348.

Georgiev, G.P. (1969). J. Theoret. Biol. 25, 473-490.

Ghosh, P.K., Reddy, V.B., Swinscose, J., Lebowitz, P. and Weissmann, S.M. (1978). J. Mol. Biol. 126, 813-846.

Glover, D.M. and Hogness, D.S. (1977). Cell 10, 167-176.

Godet, J., Verdier, G., Nigon, V., Belhani, M., Richard, F., Colonna, P., Mitchell, J., Williamson, R. and Tolstoshev, P. (1977). Blood 50, 463-470.

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

Grunstein, M. and Hogness, D.S. (1975). Proc. Natl. Acad. Sci. U.S.A. 72, 3961-3965.

Grunstein, M., Schedl, P. and Kedes, L. (1973). Molecular Cytogenetics, B. Hamkalo and J. Papaconstantinou eds. pp. 115-122. (Plenum Press, New York).

Gusella, J., Varsanyi-Breiner, A., Kao, F-T., Jones, C., Puck, T.T., Keys, C., Orkin, S. and Housman, D. (1979). Proc. Natl. Acad. Sci. U.S.A. 76, 5239-5243.

Hagen, F.S. (1979). Anal. Biochem. 93, 299-305.

Hahn, W.E. and Laird, C.D. (1971). Science 173, 158-161.

Hamer, D.H. and Leder, D. (1979). Nature 281, 35-40.

Hardies, S.C. and Wells, R.D. (1976). Proc. Natl. Acad. Sci. U.S.A 73, 3117-3121.

Harrison, P.R., Birnie, G.D., Hell, A., Humphries, S., Young, B.D. and Paul, J. (1974). J. Mol. Biol. 84, 539- 554.

Harrison, P.R., Hell, A., Birnie, G.D. and Paul, J. (1972). Nature 239, 219-221.

Hennig, W., Hennig, I. and Stein, H. (1970). Chromosoma 32, 31.

Hershfield, V., Boyer, H.W., Yanofsky, C., Lovett, N., and Helinski, D.R. (1974). Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3263.

Hohn, B. and Hohn, T. (1974). Proc. Natl. Acad. Sci. U.S.A. 71, 2372-2376.

Hohn, B. and Murray, K. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3263.

Holmes, D.S., Cohn, R.H., Kedes, L.H. and Davidson, N. (1977). Biochemistry 16, 1504-1512.

Hough, B.R., Smith, M.J., Britten, R.J. and Davidson, E.H. (1975). Cell 5, 291-302.

Housman, D., Forget, B.G., Skoultchi, A. and Benz, E.J. (1973). Proc. Natl. Acad. Sci. U.S.A. 70, 1809-1813.

Huehns, E.R., Dance, N., Beaven, G.H., Hecht, F. and Motulsky, A.G. (1964). Cold Spring Harbor Symp. Quant. Biol. 19, 327-331.

Huehns, E.R., Flynn, F.V., Butler, E.A. and Beaven, G.H. (1961). Nature 189, 496-497.

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

Humphries, P., Coggins, L.W., Old, R.W. Mitchell, G.J., Coleclough C. and Paul, J. (1978). Mol. Gen. Genet. 165, 65-71.

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

Jackson, D.A., Symons, R.H. and Berg, P. (1972). Proc. Natl. Acad. Sci. U.S.A. 69, 2904-2909.

Jackson, J.F, Tolstoshev, P., Williamson, R. and Hendrick, D. (1976). Nucl. Acids Res. 3, 2019-2026.

Jacob, F. and Monod, J. (1961) J. Mol. Biol. 3, 318-356.

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

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

Jones, K.W. (1970). Nature 225, 912-915.

Kabat, D. (1972). Science 175, 134-140.

Kacian, D.L., Gambino, R., Dow, L.W., Grossbard, E., Natta, C., Ramirez, F., Spiegelman, S., Marks, P.A. and Bank, A. (1973). Proc. Natl. Acad. Sci. U.S.A. 70, 1886-1890.

Kacian, D.L. and Myers, J.C. (1976). Proc. Natl. Acad. Sci. U.S.A. 73, 2191-2195.

Kafatos, F.C. (1972a). Acta Endrocrinologica Suppl. 168, 319-345.

Kafatos, F.C. (1972b). Cur. Topics in Develop. Biol. 7, 125-191.

Kamuzora, H. and Lehmann, H. (1975). Nature 256, 511-513.

Kan, Y.W. and Dozy, A.M. (1978). Lancet (ii), 910-912.

Kan, Y.W. and Dozy, A.M. (1979). Proc. Natl. Acad. Sci. U.S.A. 75, 5631-5635.

Kan, Y.W., Dozy, A.M., Varmus, H.E., Taylor, J.M., Holland, J.P., Lie-Injo, L.E., 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.

Kato, K-I., Goncalves, J., Houts, G.E. and Bollum, F.J. (1967). J. Biol. Chem. 242, 2780-2789.

Kattamis, C. and Lehmann, H. (1970a). Human Heredity 20, 156-164.

Kattamis, C. and Lehmann, H. (1970b). Lancet (ii), 653-637.

Kedes, L.H. and Birnstiel, M.L. (1971). Nature New Biol. 230, 165-169.

Kedes, L.H., Cohn, R., Lowry, J., Chang, A.C.Y. and Cohen, S.N. (1975). Cell 6, 359-370.

Kelly, T.J. and Smith, H.O. (1970). J. Mol. Biol. 51, 393-409.

Konkel, D.A., Tilghman, S.M. and Leder, P. (1978). Cell 15, 1125-1132.

Kourilsky, P. and Chambon, P. (1978). Trends Biochem. Sci. 3, 244-247.

Laird, C.D. (1970). Chromosoma 32, 378-406.

Lang, A. and Lorkin, P.A. (1976). Brit. Med. Bull. 32, 239-245.

Lanyon, W.G., Ottolenghi, S. and Williamson, R. (1975). Proc. Natl. Acad. Sci. U.S.A. 72, 258-262.

Larson, J.E., Hardies, S.C. Patient, R.K. and Wells, R.D. (1979). J. Biol. Chem. 254, 5535-5541.

Lasky, R.A. and Mills, A.D. (1977). FEBS Lett. 82, 314-316.

Leder, P., Tiemeier, D. and Enquist, L. (1977). Science 196, 175-177.

Lehmann, H. (1970). Lancet (ii), 78-80.

Lehmann, H. and Carrell, R.W. (1968). Brit. Med. J. 4, 748-750.

Lehmann, H. and Lang, A. (1974). Ann. New York Acad. Sci. 232, 152-158.

Leibscher, D-H., Coutelle, C., Rapoport, T. Rosenthal, S. and Williamson, R. (1979). Gene (in press).

Lewin, B. (1975). Cell 4, 11-19.

Little, P., Curtis, P., Coutelle, Ch., Van den Berg, J., Dalgleish, R., Malcolm, S., Courtney, M., Westaway, D. and Williamson, R. (1978) Nature 273, 640-643.

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

Little, P.F.R., Williamson, R., Annison, G., Flavell, R.A., de Boer, E., Bernini, L.F., Ottolenghi, S., Saglio, G. and Mazza, U. (1979b). Nature 282, 316-318.

Lobban, P.E. and Kaiser, A.D. (1973). J. Mol. Biol. 78, 453-471.

Malcolm, A.D.B. (1979). Biochem. Soc. Symp. 44, 1-12.

Mandel, J.L., Breathnach, R., Gerlinger, P., Le Meur, M. Gannon, F. and Chambon, P. (1978). Cell 14, 641-653.

Mandel, M. and Higa, A. (1970). J. Mol. Biol. 53, 159-162.

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

Maziels, N. (1976). Cell 9, 431-438.

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

Mitchell, G.J. and Williamson, R. (1977). Nucl. Acids Res. 6, 3557-3562.

Monahan, J.J., Harris, S.E., Woo, S.L.C., Robberson, D.L. and O'Malley, B.W. (1976). Biochemistry 15, 223-233.

Morrow, J.F., Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Goodman, H.M. and Helling, R.B. (1974). Proc. Natl. Acad. Sci. U.S.A. 71, 1743-1747.

Muller, W., Weber, H., Meyer,F. and Weissmann, C. (1978). J. Mol. Biol. 124, 343-358.

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

Murray, N.E. and Murray, K. (1974). Nature 251, 476-481.

Nienhuis, A.W. and Anderson, W.F. (1971). J. Clin. Invest. 50, 2458-2460.

Neinhuis, A.W., Turner, P. and Benz, E.J. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 3960-3964.

Norgard, M.V., Keem, K. and Monahan, J.J. (1978). Gene 3, 279-292.

Old, J., Clegg, J.B., Weatherall, D.J., Ottolenghi, S., Comi, P., Giglioni, B., Mitchell, J., Tolstoshev, P. and Williamson, R. (1976). Cell 8, 13-28.

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. Natl. Acad. Sci. U.S.A. 75, 5950-5954.

Orkin, S.H., Alter, B.P., Altay, C., Mahoney, M.J. Lazarus, H., Hobbins, J.C. and Nathan, D.G. (1978). New Eng. J. Med. 299, 166-172.

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

Ottolenghi, S., Comi, P., Giglioni, B., Tolstoshev, P., Lanyon, W.G., Mitchell, G.J., Williamson, R., Russo, G., Musumeci, S., Schiliro, G., Tsistrakis, G.A., Charache, S., Wood, W.G, Clegg, J.B. and Weatherall, D.J. (1976). Cell 9, 71-80.

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

Ottolenghi, S., Lanyon, W.G., Williamson, R., Weatherall, D.J., Clegg, J.B. and Pitcher, C.S. (1975). Proc. Natl. Acad. Sci. U.S.A. 72, 2294-2299.

Packman, S., Aviv, H., Ross, J. and Leder, P. (1972). Biochem. Biophys. Res. Commun. 49, 813-819.

Pardue, M.L. and Gall, J.G. (1970). Science 168, 1356-1358.

Patient, R.K., Hardies, S.C., Larson, J.E., Inman, R.B., Maquat, L.E. and Wells, R.D. (1979). J. Biol. Chem. 254, 5548-5554.

Pearson, R.L., Weiss, J.F. and Kelmers, A.D. (1971). Biochim. Biophys. Acta, 228, 770-774.

Pirotta, V. (1976). Nucl. Acids Res. 3, 1747-1760.

Polsky,F. , Edgell, M.A., Seidman, J.G. and Leder, P. (1978). Anal. Biochem. 87, 397-410.

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

Rambach, A. and Tiollais, P. (1974). Proc. Natl. Acad. Sci. U.S.A. 71, 3927-3930.

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

Ramirez, F., O'Donnell, J.V., Marks, P.A., Bank, A., Musumeci, S., Schiliro, G., Pizzarelli, G., Russo, G., Luppis, B. and Gambino, R. (1976a). Nature 263, 471-475.

Ramirez, F., O'Donnel, J.V., Natta, C. and Bank, A. (1976b). J. Clin. Invest. 58, 1475-1481.

Ricco, G., Mazza, U., Turi, R.M., Pick, P.G., Camaschella, C., Saglio, G. and Bernini, L.F. (1976). Hum. Genet. 32, 305-31

Richardson, C.C. (1965). Proc. Natl. Acad. Sci. U.S.A. 54, 158-165

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

Roberts, A.V., Clegg, J.B., Weatherall, D.J. and Ohta, Y. (1973). Nature New Biol. 245, 23-24.

Roberts, R.J. (1976). C.R.C. Crit. Rev. Biochem. 4 123-164.

Rochaix, J.D. and Malnoe, P. (1978). Cell 15, 661-670.

Rodriguez, R.L., Bolivar, F., Goodman, H.M., Boyer, H.W. and Betlach, M. (1976). ICN-UCLA Symp. Mol. Cell Biol. 5, 471-47

Roychoudhury, R., Jay, E. and Wu, R. (1976). Nucl. Acids Res. 3, 101-116.

Schaffner, W., Gross, K., Telford, J. and Birnstiel, M.L. (1976) Cell 8, 471-478.

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

Schroeder, W.A., Huisman, T.H.J., Shelton, R., Shelton, J.B., Kleihauer, E.F., Dozy, A.M., Robberson, B. (1968). Proc. Natl. Acad. Sci. U.S.A. 60, 537-544.

Seidman, J.G., Leder, A., Nau, M., Norman, B. and Leder, P. (1978) Science 202, 11-17.

Sherratt, D.J. (1979). Biochem. Soc. Symp. 44, 29-36.

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

Singer, M. and Soll, D. (1973). Science 181, 1114.

Smith, H.O. and Wilcox, K.W. (1970). J. Mol. Biol. 51 379-391.

Smithies, O., Blechl, A.E., Denniston-Thompson, K., Newell, N., Richards, J.E., Slightom, J.L., Tucker, P.W. and Blattner, F.R. (1978). Science 202, 1284-1289.

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

Southern, E.M. (1979). Anal. Biochem. (in press).

Surrey, S., Chambers, J.S., Muni, D. and Schwartz, E. (1978). Biochem. Biophys. Res. Comm. 83, 1125-1131.

Sutherland, I.A. and Ito, Y. (1978). J. High Res. Chrom. and Chrom. Comm. 3, 171-172.

Suzuki, Y., Gage, L.P. and Brown, D.D. (1972). J. Mol. Biol. 70, 637-656.

Tabak, H.F. and Flavell, R.A. (1978). Nucl. Acids Res. 5, 2321-2332.

Tartof, K.D. and Dawid, I.B. (1976). Nature 263, 27-30.

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

Temple, G.F., Chang, J.C. and Kan, Y.W. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 3047-3051.

Thomas, M., Cameron, J.R. and Davis, R.W. (1974). Proc. Natl. Acad. Sci. U.S.A. 71, 4579-4583.

Tiemeier, D.C., Tilghman, S.M., Leder, P. (1977). Gene 2, 173-191.

Tilghman, S.M., Tiemeier, D.C., Peterlin, B.M., Sullivan, M., Maizel, J.V., Leder, P. (1978). Proc. Natl. Acad. Sci. U.S.A 75, 725-729.

Tilghman, S.M., Tiemeier, D.C., Polsky, F., Edgell, M.H., Seidman, J.G., Leder, A., Enquist, L.W., Norman, B. and Leder, P. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 4406-4410.

Tolstohev, P., Mitchell, J., Lanyon, G., Williamson, R., Ottolenghi, S., Comi, P., Giglioni, B., Masera, G., Modell, B., Weatherall, D.J. and Clegg, J.B. (1976). Nature 260, 95-98.

Tolstoshev, P., Williamson, R., Eskdale, J., Verdier, G, Godet, J., Nigon, V., Trabuchet, G. and Benabadji, M. (1977). Eur. J. Biochem. 78, 161-165.

Tonegawa, S., Maxam, A.M., Tizard, O., Bernard, W., and Gilbert, W. (1978).Proc. Natl. Acad. Sci. U.S.A. 75, 1485-1489.

Tuan, D., Biro, P.A., de Riel, J.K., Lazarus, H. and Forget, B.G. (1979). Nucl. Acids Res. 6, 2519-2544.

Twigg, A. and Sherratt, D. (1979). Nature (in press).

Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. and Rutter, W.J. (1978). Proc. Natl. 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 Weissmann, C. (1978). Nature 276, 37-44.

Walker, P.M.B. (1971). Prog. Biophys. Mol. Biol. 23, 145.

Walker, P.M.B., Flamm, W.G. and McLaren, A. (1970). Handbook of Molecular Cytology, p. 53. (North Holland, Amsterdam).

Weatherall, D.J. (1976). Johns Hopkins Med. J. 139, 205-210.

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

Weatherall, D.J. and Clegg, J.B. (1979b). The Thalassaemia Syndromes, third edition (Blackwell Scientific Publications, Oxford).

Weinberg, E.S., Birnstiel, M.L., Purdom, I.F. and Williamson, R. (1972). Nature 240, 225-228.

Weinberg, E.W., Overton, G.C., Shutt, R.H. and Reeder, R.H. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 4815-4819.

Weissmann, C. and Boll, W. (1976). Nature 261, 428-429.

Wellauer, P.K. and Dawid, I.B. (1977). Cell 10, 177-192.

Wellauer, P.K. and Reeder, R.H. (1975). J. Mol. Biol. 94, 151-161.

Wellauer, P.K., Reeder, R.H., Carroll, D., Brown, D.D., Deutch, A. Higashinakagawa, T. and Dawid, I.B. (1974). Proc. Natl. Acad. Sci. U.S.A. 71, 2823-2827.

Wells, R.D., Hardies, S.C., Horn, G.T., Klein, B., Larson, J.E., Neuendorf, S.K., Panayotatos, N.,Patient, R.K. and Selsing, E. (1979). Methods in Enzymology (in press).

White, R.L. and Hogness, D.S (1977). Cell 10, 177-192.

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.

Wild, M.A. and Gall, J.G. (1978). J. Cell. Biol. 79, 144a.

Williams, R.E.O. (1976). Report of the Working Party on the Practice of Genetic Manipulation, Cmnd. 6600, H.M.S.O., London.

Williamson, R. (1976). Brit. Med. Bull. 32, 246-250.

Wilson, D.A. and Thomas, C.A. (1973). Biochem. Biophys, Acta 331, 333-340.

Wilson, D.A. and Thomas, C.A. (1974). J. Mol. Biol. 84, 115-144.

Wilson, G.A. and Young, E.E. (1976). Microbiology, D. Schlessinger ed., pp. 350-357. (American Society for Microbiology, Washington, D.C.).

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.

Wilson, M.C., Melli, M. and Birnstiel, M.L. (1974). Biochem. Biophys. Res. Commun. 61, 354-358.

Woo, S.L.C., Dugaiczyk, A., Tsai, M-J., Lai, E.C., Caterall, J.F. and O'Malley, B.W. (1978). Proc. Natl. Acad. Sci. U.S.A. 75, 3688-3692.

Wood, W.G., Clegg, J.B. and Weatherall, D.J. (1977a). Progress in Hematology, E.B. Brown ed., vol. X, pp. 43-90. (Grune and Stratton, New York).

Wood, W.G., Clegg, J.B., Weatherall, D.J., Gyde, O.H.B., Obeid, D.A., Tarlow, M.J. and Hewitt, S. (1977b). J. Med. Genet. 14, 237-244.

Young, B.D., Harrison, P.R., Gilmour, R.S., Birnie, G.D., Hell, A. Humphries, S. and Paul, J. (1974). J. Mol. Biol. 84, 555-568

Yunis, J.J. and Yasmineh, W.G. (1971). Science 174, 1200-1209.

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

(Reprinted from Nature, Vol. 273, No. 5664, pp. 640–643, June 22, 1978)
© Macmillan Journals Ltd., 1978

Isolation and partial sequence of recombinant plasmids containing human α-, β- and γ-globin cDNA fragments

P. Little*, P. Curtis†, Ch. Coutelle*, J. Van Den Berg†, R. Dalgleish*, S. Malcolm*, M. Courtney*, D. Westaway* & R. Williamson*

*Department of Biochemistry, St. Mary's Hospital Medical School, University of London, London W2, UK
†Institute for Molecular Biology I, University of Zurich, Honggerberg, 8093 Zurich, Switzerland

Human globin cDNA-derived recombinants with plasmid pCR1 have been prepared for use as specific hybridisation probes and for the partial sequencing of α-, β- and γ-globin genes.

TO analyse the molecular biology of developmental and pathological systems in humans, one must investigate not only the expression of single genes, but also that of families of genes. One of the most interesting and most easily studied human gene families specifies the β-like globin chains. The gene arrangement is known, and the four β-like genes are thought to be closely linked[1]. There is a switch during foetal and neonatal development from γ-globin synthesis to β-globin synthesis, while the synthesis of δ-globin remains at a low level throughout[2]. There is evidence of different half lives for the otherwise similar δ and β-globin mRNA (ref. 3). Finally, there is a wide spectrum of genetic disease, the thalassaemias, in which the expression of the globin genes is affected, and which have been shown to be due to lesions at any one of several levels: gene deletion, lack of transcription, faulty post-transcriptional processing or translation defects[4–8]. A similar α-gene family exists, although developmental switches in this case are less well understood[9]. There is also regulated expression of the amounts of α-globin and β-globin, a useful model for coordinate gene expression at the molecular level[10].

Preparation and characterisation of chimaeric plasmids

To study the problems listed above, it is essential to prepare nucleic acid probes which are specific for each globin gene. So far, this has been done using either mRNA or cDNA. Preparation of total globin mRNA, and thus its complementary copy, is no longer difficult, but purifying a sequence complementary to only one of the globin genes poses problems, particularly if more than a very small amount of probe is required[11,12]. To meet the needs for complete purity and large amounts of these probes, we have prepared chimaeric plasmids in which a double-stranded DNA copy of human globin foetal mRNA has formed a recombinant with plasmid pCR1 (ref. 13).

The plasmids that have been isolated have been characterised by 'Grunstein–Hogness' hybridisation on a filter to purified polynucleotide kinase-labelled globin mRNA (refs 14, 15), then by solution hybridisation in plasmid excess to chain-specific cDNAs. The plasmids have also been characterised by restriction mapping, and in two cases the globin cDNA length confirmed by excision and direct measurement.

Plasmids were selected containing α-, β- and γ-globin gene sequences and in each case the plasmid selected contained a cDNA-derived sequence greater than half the length of the mRNA. The availability of this family of human globin gene sequences should facilitate detailed analysis of normal globin gene arrangement and the causes of dysfunction, particularly in thalassaemia.

Human globin messenger RNA was prepared using two cycles of oligo dT cellulose affinity chromatography (Collaborative Research, grade T3) using 'normal' adult reticulocytes (mRNA (α, β)), neonatal exchange transfusion blood (mRNA (α, β, γ)), and blood from a previously studied double heterozygote $\beta_0/\delta\beta_0$ thalassaemic with no mRNA (β) in circulating red blood cells (mRNA (α, γ)) (ref. 16). Messenger RNA for clone identification was used without further purification.

A sample of mRNA (α, γ, β) was used to prepare double-stranded cDNA for insertion into plasmids[17–21]. This mRNA was first prepared by affinity chromatography, as above, treated with ribonuclease-free DNase (Worthington) and then centrifuged on a sucrose gradient; only material sedimenting at 9S was collected. Complementary DNA was prepared using RNA-dependent DNA polymerase. The cDNA was treated with alkali to remove RNA, neutralised and used as a template for the synthesis of a second, complementary strand with RNA-dependent DNA polymerase. The double-strand product was treated with S1 nuclease, elongated with terminal transferase, and a poly dA sequence was added to the 3′ termini of the double-stranded cDNA using the method of Maniatis[19,20].

Plasmid identification and hybridisation to purified globin cDNA

Plasmid pCR1 was purified and cleaved at the single *Eco*RI restriction site, and the 3′ termini were extended with poly dT sequences using terminal transferase[20]. 100 ng of tailed plasmid were mixed with 2 ng of tailed double-stranded cDNA, annealed, and used to transfect *Escherichia coli* HB101. The molar ratio of plasmid to insert was 2.3:1. The number of recombinant clones obtained in three experiments, as judged by kanamycin resistance, was 37, 70 and 175. These were picked and plated out in arrays and sufficient cells grown to allow transfer to several replica Millipore HAWP filters. All the transfections and growth of recombinant clones before positive identification were carried out under UK category 3 biohazard containment conditions in the Department of Biochemistry, Imperial College, as advised by the UK Genetic Manipulation Advisory Group[22].

Replica Millipore filters were prepared for hybridisation to appropriate (^{32}P)-labelled mRNAs essentially by the method of Grunstein and Hogness[14]. The mRNAs were cleaved into fragments approximately 100 nucleotides long by brief exposure to alkali before terminal labelling with polynucleotide kinase[15]. Each filter was hybridised to 2.5×10^6 d.p.m. of (^{32}P)-mRNA in a volume of 1.5 ml in 50% formamide hybridisation buffer for 18 h at 42°C.

After autoradiography of the filters, a set of clones hybridising strongly with (^{32}P)-mRNA (α, γ, β) and not scoring with (^{32}P)-mRNA (α, β) was identified as being most likely to contain γ-globin cDNA inserts. Another set scoring with all three probes mRNA (α, γ, β), mRNA (α, β), and mRNA (α, γ) was identified as most likely to contain α-globin inserts. Small

cleaved, AAGCTT) (ref. 26) cleave the α-globin cDNA insert, and do not cleave the β- or γ-globin cDNA inserts. *Eco*RI cleaves both the β- and γ-globin cDNAs at a single site, and does not cleave the α-globin cDNA insert.

It is possible to derive restriction site maps of the inserted fragments by comparison of fragment patterns generated from chimaeric and non-chimaeric plasmids. Figure 2 shows this approach using *Hpa*II. Track *b* is the fragment pattern for pCR1 with this enzyme. A single fragment at 790 base pairs is removed by digestion with *Eco*RI (track *c*) to generate a new band at 690 base pairs. This indicates that the single *Eco*RI site is 100 base pairs from one end of a 790-base pair *Hpa*II fragment. The α plasmid gives a pattern (track *e*) with two new bands at 1,000 base pairs and 440 base pairs (this band runs at the same mobility as a pCR1 fragment, but its presence is proved by microdensitometry). This pattern may be interpreted by assuming that there is an *Hpa*II site within the inserted DNA giving rise to the two new fragments. Thus, the 1,000-base pair fragment contains 690 base pairs of pCR1 and 310 base pairs of inserted DNA, the 440 fragment contains 100 base pairs of plasmid and 340 base pairs of inserted DNA. This enables the map in Fig. 3 to be derived. The γ plasmid gives a pattern with one new 1,500-base pair fragment (Fig. 2, track *h*). This consists of 790 base pairs of pCR1 plus 710 base pairs of inserted DNA. There is no internal *Hpa*II site. Subsequent digestion with *Eco*RI breaks this new fragment to give two new bands (Fig. 2, track *i*) and the location of the *Eco*RI site is established in a fashion directly analogous to the α *Hpa*II site. The β plasmid gave a pattern of fragments that was very similar to the pHγG1: a new band at 1,550 base pairs was visible, indicating an inserted DNA fragment of 760 base pairs. On digestion with *Eco*RI, this fragment was split to give new bands at 1,000 base pairs and 550 base pairs (data not presented).

Similar experiments were also carried out using *Hha*I to give the final maps shown in Fig. 3. There is ambiguity in the α *Hha*I map where the possible location of a second *Hha*I site remains obscure. The single *Eco*RI site in pHβG1 and pHγG1 allows a tentative statement on the orientation of the inserted DNA. The cDNAβ sequence is known[27] and has a single *Eco*RI site toward the 3′-end at the amino acid 121–122. Thus, the asymmetric distribution of DNA to the left and right suggests that the cDNA is inserted as shown in Fig. 3. Similar arguments apply to pHγG1. Inspection of the protein sequence indicates only one possible *Eco*RI site in the coding sequence, again at 121–122. What has not been established from the restriction map alone is the absolute orientation of the *Hpa*II/*Hha*I fragments within the whole plasmid.

The size of DNA which has been inserted into a plasmid using poly dA-poly dT tailing, as in this case, can also be estimated by excising the insert using S1 nuclease in conditions where only the poly dA-poly dT regions melt[28]. This was done for the γ-globin plasmid pHγG1 and the results are shown in Fig. 4. The size of the inserted sequence is 500 nucleotide pairs. As this does not include the dA and dT sequences, it is in excellent agreement with the restriction data. Similar experiments for pHβG1 indicate an inserted sequence of 540 base pairs (data not given).

All three plasmids were further characterised by direct sequence analysis around a restriction site in the coding region of the inserted DNA. The results (Fig. 5) confirm the assignments made from the hybridisation data and further differentiate between pHβG1 and pHγG1. Sequencing was carried out according to the method of Maxam and Gilbert[32] around the *Hin*dIII site in the pHαG1 inserted sequence and around the *Eco*RI site in the pHβG1 and pHγG1 inserted sequences.

Sequencing of the α plasmid shows that out of the five possible *Hin*dIII sites predicted for the human α globin protein sequence, the cDNA is cleaved at a site spanning the amino acids 90–91. The nucleotide sequences derived from pHβG1 and pHγG1 confirm that the *Eco*RI site lies at amino acids 120–121, and in addition, the nucleotide sequence for the β globin agrees fully with sequences previously published[27,34]. The sequencing studies also confirm the proposed orientation of each inserted sequence relative to known pCR1 restriction sites.

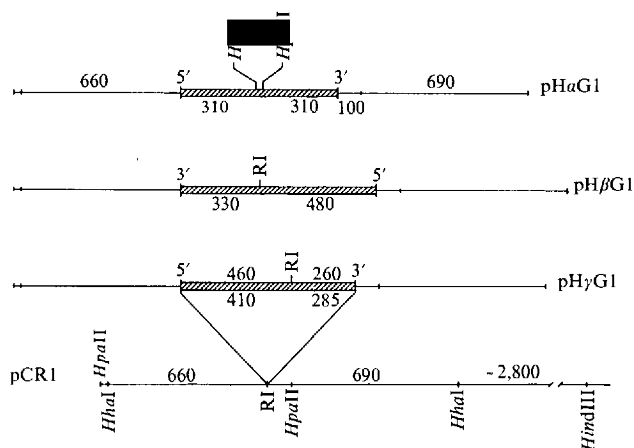


Fig. 3 Restriction map around the site of insertion of plasmids pCR1, pHαG1, pHβG1 and pHγG1. [RI] indicates the two halves of the *Eco*R1 site of insertion. Approximate map distances are given in base pairs. The *Hpa*II and *Hha*I sites in pHαG1 can be calculated to be about 30 base pairs apart, but their order relative to other restriction sites cannot be established from the restriction data presented in Fig. 2. The orientation of the inserted DNA in pHαG1 is deduced from the nucleotide sequence shown in Fig. 5. The two sets of map distances presented for pHγG1 are derived from experiments excising the inserted DNA from pCR1 with *Hpa*II (above) and *Hha*I (below). The inserted DNA sequences in each plasmid (including the poly dAdT tracks) are signified by shading.

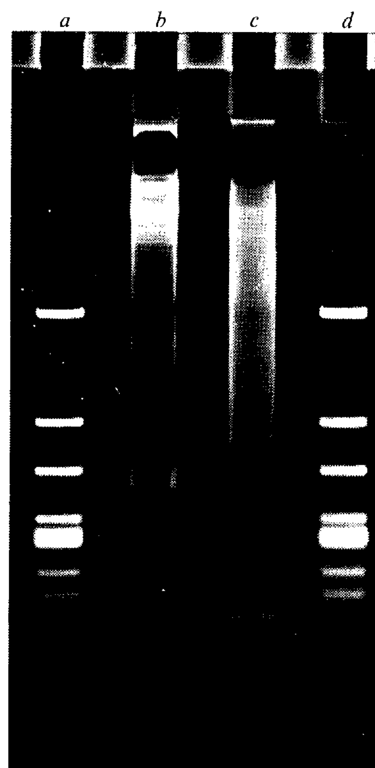


Fig. 4 Excision of the cDNA sequences from pHγG1 with S1 nuclease. Tracks *a* and *d*, *Hae*III digested SV40 DNA. Track *b*, 1.0 μg pHγG1 DNA digested with 24 units S1 nuclease (Sigma) in 20 μl of 45% formamide, 30 mM NaOAc, 1 mM $ZnSO_4$, 0.2 M NaCl, *p*H 4.5, for 30 min at 55°C. Track *c*, 1.0 μg of pCR1 digested as *b*. The whole reaction mix is run on a 2.0% Agarose gel as described in the legend to Fig. 5. The prominent band seen in *b* is the inserted cDNA.