CHICKEN GLOBIN mRNA AND GENES

A thesis submitted for
the Degree of
Doctor of Philosophy
in the
University of Adelaide
by

Robert Ian Richards, B.Sc.(Hons.)
Department of Biochemistry

April 1980

CONTENTS | Page

	Page
THESIS SUMMARY	i
STATEMENT	iii
ACKNOWLEDGMENTS	iv
ABBREVIATIONS	v
CHAPTER I - INTRODUCTION AND LITERATURE REVIEW	1
1. Introduction	1
2. Literature Review	4
A. Differentiation	4
B. Control of gene expression	5
i) Prokaryote	5
ii) Eukaryote	6
C. Eukaryote messenger RNA	10
D. Globin genes	13
i) Hemolgobin structure and function	13
ii) Classical genetic analysis	14
iii) Globin gene switching	16
E. Chicken erythroid tissue	17
i) Hemoglobins	17
ii) Tissue specific non-globin proteins	18
iii) Viral transformation and globin gene expression	20
CHAPTER II - MATERIALS AND METHODS	22
1. Materials	22
A. Chemicals and reagents	22
B. Enzymes	23
C. Specialized materials	24
i) Bacterial strains	24
ii) Nucleic acids	24
iii) Experimental animals	25

Page

D. Buffers and media 25

2. Methods 26

A. mRNA purification 26

i) Induction of anaemia 26

ii) Isolation of polysomes 26

iii) mRNA isolation 26

B. Oligo(dT) cellulose chromatography 27

C. Polyadenylation of RNA 28

D. Preparation of high molecular weight DNA 28

E. Synthesis of cDNA 29

i) Analytical cDNA synthesis 29

ii) Preparative cDNA synthesis 30

F. Second strand cDNA synthesis 30

G. Restriction endonuclease digestion 31

H. Routine gel electrophoresis 31

i) Polyacrylamide gel electrophoresis 31

ii) Agarose gel electrophoresis 32

I. Autoradiography 32

J. Electroelution 33

K. Construction of recombinant DNA 34

i) Blunt ending reactions 34

ii) Ligation of restriction endonuclease recognition sites 34

iii) Ligation of cDNA to plasmid DNA 36

L. Transformation and selection of recombinants 36

i) Transformation 36

ii) Selection of recombinants 37

M. Isolation of plasmid DNA 39

i) Amplification of plasmid DNA 39

Page

ii) Isolation of plasmid DNA 39

iii) Isolation of supercoiled DNA 40

N. Sequence analysis of DNA 41

i) End labelling DNA fragments 41

ii) Sequencing reactions 44

iii) Sequencing gels 47

O. Restriction endonuclease analysis of genomal DNA 48

i) Synthesis of probe 48

ii) Blot analysis 49

P. Containment facilities 50

CHAPTER III - ISOLATION OF GLOBIN mRNA AND cDNA CHARACTERIZATION 52

1. Introduction 52

2. Results 53

A. Chicken globin mRNA isolation 53

B. Polyadenylation of poly(A)$^{-}$ RNA 54

C. Synthesis of cDNA 54

D. Restriction endonuclease digestion 55

i) Single-stranded cDNA 55

ii) Double-stranded cDNA 56

3. Discussion 57

CHAPTER IV - MOLECULAR CLONING OF CHICKEN GLOBIN cDNA 58

1. Introduction 58

2. Results 58

A. Synthesis of blunt ended double-stranded cDNA 58

B. Ligation of *Hin*dIII recognition site 59

Page

C. Ligation of cDNA to pBR322 plasmid DNA 60

D. Transformation, selection and screening of recombinants 61

E. Sequence analysis of insert DNA 62

3. Discussion 63

CHAPTER V - SEQUENCE ANALYSIS OF β GLOBIN CODING cDNA CLONES 65

1. Introduction 65

2. Results 65

A. Sequencing strategy 66

B. 5' Terminal non-homology 68

C. Chicken β globin gene 70

3. Discussion 71

CHAPTER VI - SEQUENCE ANALYSIS OF α GLOBIN CODING cDNA CLONES 74

1. Introduction 74

2. Results 75

A. Sequence analysis of pCG_{α}-3 75

B. Comparison of cloned sequence restriction maps with restriction endonuclease digests of d.s. cDNA 77

C. Minor α globin coding mRNA species 77

D. Comparison of sequences 78

3. Discussion 80

CHAPTER VII - FINAL DISCUSSION 83

1. Introduction 83

2. Hemoglobin Switching in Chickens 83

3. Chicken Globin mRNA 85

A. 5' Untranslated region 85

Page

B. Coding region 86

C. 3' Untranslated region 86

4. Further Studies 87

BIBLIOGRAPHY 89

APPENDIX - PUBLICATIONS 99

SUMMARY

The work in this thesis concerns the characterization of globin mRNA from avian erythroid cells. The primary structure of the major mRNA species was deduced by sequence analysis of recombinant DNA clones containing cDNA inserts derived from purified polysomal chicken globin mRNA.

1. Anaemia was induced in 12 week old chickens by the injection of phenylhydrazine. Polysomes were isolated from the blood and the 9S RNA purified. Double-stranded DNA copies of the mRNA were synthesized by sequential reverse transcriptase reactions, with oligo-(dT) priming of the initial reaction. Double and single-stranded cDNA were subjected to cleavage with restriction endonucleases to characterize the sequences represented.
2. Full length double-stranded cDNA was inserted into the plasmid BR322 by the use of synthetic linker DNA containing the *Hin*dIII recognition sequence. Inserted DNA from the individual recombinants was purified from the plasmid DNA and subjected to sequence analysis by the chemical degradation method of Maxam and Gilbert.
3. The complete sequence of the longest inserts representing the two major mRNA transcripts was deduced.

   This work produced the following conclusions:-

   a) there are two major globin mRNA species present in the erythroid cells of phenylhydrazine induced anaemic chickens. The β coding species codes for a protein which is identical to that in non-anaemic adult blood. The α coding species codes

for a protein which is not like either of the α chains present in non-anaemic adult blood, and therefore must be induced by phenylhydrazine treatment,

b) the induced α coding mRNA codes for a protein which differs by 22 amino acids from that of α_A and by 61 amino acids from that of α_D, these two alpha chains being the normal components of adult chicken globin,

c) sequence analysis of several recombinants reveals that errors are made during the cloning process due to the necessity of the reverse transcriptase enzyme to use a loop structure of the first DNA strand as the primer for second strand synthesis. These "errors" are confined to the 5' end (with respect to the mRNA) of the inserts,

d) comparison of the 3' untranslated regions with mammalian globin mRNA sequences reveals homology which is consistent with chicken globin genes being relatively primitive and having origins close to the separation of globin into α and β chain types on the genealogical tree.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, it contains no material that has been previously published by any other person, except where due reference is made in the text.

Robert I. Richards

ACKNOWLEDGMENTS

I wish to express thanks to the following individuals who contributed in some way to the shape and form of my work and this thesis.

Dr. J.R.E. Wells: for his friendly supervision and encouragement during the course of this work.

Richard Harvey, Alan Robins, Dr. Bob Crawford, Paul Krieg and other members of Lab. 20A for their discussions, enthusiasm and assistance.

Professor W.H. Elliott: for permission to work in the Department of Biochemistry, University of Adelaide.

Dr. W.J. Rutter: for permission to visit and work in the Department of Biochemistry and Biophysics, University of California, San Francisco.

Drs. Pablo Valenzuela, Axel Ullrich, John Shine, Peter Seeburg, Barbara Cordell and Howard Goodman: for their assistance, advice and discussions.

For their assistance in the production of this thesis I would like to thank -

Jane Somerville: for excellent typing.

Lesley Crocker: for assistance with photography.

Dr. R.H. Symons: for critical reading.

William Mitchell: for title page artwork.

Finally, I wish to express thanks to my wife, Helene, my parents, a Commonwealth Postgraduate Research Award and the University of Adelaide Research and Publications Committee for financial assistance.

ABBREVIATIONS

A : adenine.

A_n : optical absorbance measured at wavelength *n* over a pathlength of 1 cm.

AMV reverse transcriptase : avian myeloblastosis virus RNA dependant DNA polymerase (EC 2.7.7.-).

b.p. : nucleotide base pairs (K.b.p. - kilo base pairs).

C : cytosine.

cDNA : DNA complementary to RNA (s.s. - single-stranded, d.s. - double-stranded).

Ci : Curie.

dATP : adenine deoxyribose-5'-triphosphate.

dCTP : cytosine deoxyribose-5'-triphosphate.

dGTP : guanine deoxyribose-5'-triphosphate.

DNA : deoxyribonucleic acid.

DNAse : deoxyribonuclease.

dTTP : thymine deoxyribose-5'-triphosphate.

EDTA : ethylenediaminetetracetic acid.

G : guanine.

Hb : hemoglobin.

HEPES : N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid.

mRNA : messenger RNA.

N : nucleotide.

oligo(dT) : oligo(thymidylic acid).

P : purine.

poly(A) : poly(adenylic acid).

RNA : ribonucleic acid.

RNAse : ribonuclease.

RNP : ribonucleoprotein.

S : sedimentation coefficient.

S_1 nuclease : (EC 3.1.4.21).

T : thymine.

TEMED : N,N,N',N'-tetramethylethylenediamine.

Tris : tris(hydroxymethyl)aminomethane.

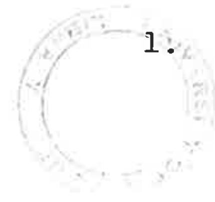
U : uracil.

U.V. : ultraviolet.

Y : pyrimidine.

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW



CHAPTER I - INTRODUCTION AND LITERATURE REVIEW

1. Introduction

The fundamental difference between genetic material and other molecules is the ability to dictate self replication.

Competition between genetic molecules, for the elements essential for survival and perpetuation, has led to the evolution of diverse species whose genetic materials are capable of coding for a wide variety of functions. These functions are directed towards providing an environment in which the genetic molecules can replicate (Dawkins, 1976).

In the case of single cell organisms, these functions are performed by that cell in competition with other individual cells. Multicellular organisms have their functions partitioned between the cells of that organism, which therefore have a symbiotic existence, competition remaining at the level of the individual. Only the germ cells of a higher eukaryote are capable of perpetuating the genetic material, the other cells of the organism performing an altruistic role. Red blood cells, for example, perform the function of distributing oxygen, an essential element for survival throughout the organism. In many species the genetic material is lost completely from these cells by enucleation, the cell becoming a circulating sack of oxygen carrying pigment. This thesis is ultimately concerned with the differentiation of cells to perform different functions in the higher eukaryote organism, the fundamental question being: how is it that only particular specific gene products are produced in the different cell types of an organism all of which arose from a single fertilized ovum and have

essentially identical genetic material? Clearly many processes are involved in differentiation, some of which are general in terms of the cells commitment to suppress most genetic information, and others which select the particular genes to be expressed in that cell type. Differentiation is also seen to be active at both cellular and molecular levels, with stem cells dictating the fate of progeny cells while molecules, such as hormones, regulate the expression of specified genes.

Work in this thesis is concerned with the control of specific gene expression at the molecular level, in particular with the family of genes which encode the peptide chains of haemoglobin, the major oxygen carrying pigment in higher eukaryote organisms. The principle aim was to determine the primary structure of messenger RNA transcripts present in adult chicken red blood cells active in gene expression. This work forms part of a study into the relationship between the structure of genetic material and its selective expression.

The basis for this approach is the belief that gene sequences contain information, in addition to that required as a code for a protein or functional RNA, which is concerned with the control of expression of that gene. These sequences, therefore, are related to the differentiation of the cell line in which a specific gene (or gene family) is expressed. Controlling sequences need not necessarily be confined to non-coding regions, for example, selection for a particular codon and, therefore, tRNA population may play a significant role at the level of translation of the mRNA sequence into protein.

Comparative analysis of gene sequences may be useful

in determining control regions of the gene which facilitate selective expression. Comparison of analogous genes in different species may reveal sequences conserved presumably as a consequence of the requirement for optimal control of expression. This analysis presumes the divergent evolution of non-essential sequences.

Another analysis, which complements this, is a study of genes which are not related and yet are expressed specifically, perhaps co-ordinately, in the one cell type. Again, common sequences in the respective genes would be candidates for control regions. This analysis presumes the convergent evolution of essential sequences.

Analysis of gene expression in the avian erythroid cell system, in this regard, affords two useful features:

(1) Avian globin genes are evolutionarily distinct from the extensively studied mammalian globin gene systems and, therefore, ideal for comparative analysis.

(2) Unique, possibly co-ordinated, expression of tissue specific non-globin genes, the nuclear proteins histone H_5 and high mobility group protein HMG-E (probably as a consequence of nucleus retention).[‡]

The remainder of this chapter is devoted to a review of the relevant literature, aimed at providing a background to the research presented in this thesis, the specific aims of which are as follows:

[‡]Both of these proteins will be discussed in more detail later in this chapter.

(1) Purification of sequences derived from the tissue specific gene transcripts expressed in adult chicken erythroid tissue.

(2) Derivation, using DNA sequence analysis, of the primary structure of these sequences.

(3) The use of purified gene transcript sequences as a probe for analysis and isolation of the chromosomal gene sequences.

2. Literature Review

A. Differentiation

Differentiation exhibits two striking features which any proposed model of the phenomenon must take into account:-

(1) Retention of potentially active genetic material

A possible method of selective expression could be the specific removal from the cell of genetic information not necessary for differentiation. That this is not the case was demonstrated in a spectacular experiment by Gurdon and Woodland (1970) which showed that a fully differentiated cell contained all the genetic information necessary to produce an entire individual. Examples of sequence amplification (ribosomal genes in *Xenopus laevis*, Brown and Blackler, 1972) or rearrangement (immunoglobulin genes, Hozumi and Tonegawa, 1976) tend to be the exception rather than the rule for selective expression.

(2) Differentiation is predetermined

At the cellular level it is apparent that stem cells predetermine the fate of daughter cells, at least to a certain extent. This is evident during hemoglobin HbA to HbC switching in sheep (cited in Neinhuis and Stamatoyannopoulos, 1978) where stem cells, which give rise to

only haemoglobin HbC synthesising progeny, appear in the bone marrow 24-36 hours before any accumulation of mRNA coding for the β chain of HbC. Therefore, the primary regulatory events leading to selective expression occur well before replication has ceased, and the specific regulatory controls are inherited by the daughter cells.

Differentiation occurs, therefore, when different regions of the genome are selectively expressed. There are several levels at which regulation is directed, although the fate of a differentiated cell, and therefore the specificity of the control of expression, is predetermined.

This thesis is particularly concerned with the synthesis of major gene products in a differentiated cell line, and although this can be viewed as being the end of differentiation rather than the means, differences between gene sequences appear to be of critical importance in the operation of selective expression mechanisms.

B. Control of gene expression

i) Prokaryote

The rapid replication of prokaryote genetic material and subsequent division enables it to rapidly acquire selective advantage in response to changes in environment. In these terms prokaryote evolution is much faster than that in eukaryotes whose replication time is orders of magnitude greater.

To facilitate adaptability, prokaryote genes exhibit precise and rapid mechanisms for selective expression. Being unicellular organisms they cannot partition functions, subsequently the entire genome is potentially responsive and so selective control is of necessity very

elaborate . Expression is controlled mainly at the level of transcription, the type of control related to the function of the gene products (Jacob and Monod, 1961). Genes involved in a common function are often clustered into an operon which imparts co-ordinate expression.

Because of the relatively small genome, ability to isolate mutants and the suitability to *in vivo* experiment a clear understanding of the molecular basis of gene control (Jacob and Monod, 1961) and the precise molecular interactions involved (Ogata and Gilbert, 1978; Keller and Calvo, 1979) now exist. Initiation of transcription occurs by specific binding of an RNA polymerase molecule to a nucleotide recognition sequence (promoter) which lies adjacent to the gene cluster. Control is mediated by protein-DNA interactions of repressor protein(s) with a precise nucleotide sequence (operator) that results in interference with the RNA polymerase molecule, preventing its progression and, therefore, inhibiting transcription (*trp* operon) or changing its affinity of binding to the promoter and thus stimulating synthesis (*lac* operon). Finer control is mediated by an attenuator sequence which effects premature termination of transcription in the case of the *trp*, *phe*, *his*, *thr* and *leu* operons in *E. coli* (Keller and Calvo, 1979).

Although there are obvious and extensive differences between prokaryotes and eukaryotes, an understanding of the mechanisms involved in prokaryote gene expression will undoubtedly facilitate that in eukaryotes, even if only through use as genetic tools.

ii) Eukaryote

In a differentiated eukaryotic cell the genetic

material exists in two states, the majority of the genome is silent in terms of gene expression, whereas tissue specific genes are actively expressed (Lewin, 1975). Attempts have been made to partition the genome into these active and inactive components, in the hope of purifying tissue specific genes and to identify some mechanism of selective expression.

Several methods were developed which divided chromatin on the basis of varying template activity with RNA polymerase and possibly different histone and non-histone protein content (Howk *et al.*, 1975). These properties were presumed to be related to transcriptional activity and, therefore, to selective gene expression. In cases where specific gene probes have been used no partitioning of active genes into "active" chromatin (euchromatin) or inactive genes into "inactive" chromatin (heterochromatin) was seen (Howk *et al.*, 1975; Krieg and Wells, 1976; Seidman and Cole, 1977; Itzhaki *et al.*, 1978).

An alternate approach to chromatin partitioning has indicated that a physical distinction does exist on the basis of susceptibility to certain nuclease attack (Weintraub and Groudine, 1976; Garel and Axel, 1976; Leibovitch and Harel, 1978). Studies on the fetal and adult globin genes in sheep suggest that selective and temporal sensitivity to mild pancreatic DNAse I digestion accompanies the expression of these genes (Young *et al.*, 1978) although it is not known whether the phenomenon is a cause or consequence of gene activity.

The physical state of an actively transcribed gene and how that state is selectively achieved remains

conjectural, although chemical analysis has recently given some insight into the components of a eukaryote gene.

The advent of recombinant DNA technology in the structural analysis of eukaryote genes has increased the awareness of the problems involved in eukaryote gene expression, but, with a few notable exceptions, has not increased the understanding of how these problems are overcome.

Eukaryotic genes are often clustered into families, in fact even in those cases where genes were thought to be unique (single copy per haploid genome), pseudogenes have been found in close proximity (Royal *et al.*, 1979). This clustering of related genes shows little similarity to the operons of prokaryotes as polycistronic-like RNA transcripts are only seldom found (ribosomal genes, Wellauer and Dawid, 1974; ACTH/β-endorphin precursor, Nakanishi *et al.*, 1979) and in certain cases the adjacent genes are expressed at different times during development (β-like globin genes, Weatherall and Clegg, 1979). Furthermore, where co-ordinate expression of two more genes is essential for the activity of a multi-component product, the genes are sometimes located on different chromosomes (e.g., globin genes of human, Deisseroth *et al.*, 1978, and chicken, Hughes *et al.*, 1979).

Perhaps the most surprising, and certainly unexpected, difference between prokaryote and eukaryote genes to date is the presence of intervening or intron sequences. These sequences occur in almost all types of gene sequences, notable exceptions being the histone genes (Levy *et al.*, 1979), and in identical locations in the analogous genes of different species. Their transcription

in the mRNA precursor (Tilgman *et al.*, 1978) necessitates a processing (splicing) step which is potentially involved in regulation of expression, but the extent of this contribution is undetermined. Breathnach *et al.* (1978) have shown that all intervening sequences (except those in yeast tRNA, O'Farrell *et al.*, 1978) have common or "concensus" sequences at their junction with exon sequences and so it is likely that a common processing pathway occurs for all mRNA sequences, thereby eliminating the possibility of differential processing as a selective step. Lerner *et al.* (1980) proposed that the small nuclear RNA, present as ribonucleoprotein particles (snRNPs), are intimately involved in the splicing step. The RNA sequence in snRNPs is capable of base pairing to the concensus junction sequences, imparting sensitivity to nuclease attack, carried out by the proteins associated with the snRNPs. This type of activity has a counterpart in RNase P in *E. coli* (Stark *et al.*, 1978), a ribonuclease which contains a short RNA molecule of similar, though not identical, function to that in the snRNPs.

A feature common to both prokaryote and eukaryote genes is a transcription promoter sequence. In prokaryotes the promoter is functionally defined as that region of DNA which is involved in the initiation of transcription by RNA polymerase. Most prokaryote promoters which have been sequenced have two common interacting sites, the Pribnow box (Pribnow, 1975) and the recognition site (see Calos, 1978 for references) both of which are based on model sequences.

A similar model or consensus sequence is situated in eukaryotes about 30 base pairs upstream from the initiation

of transcription (see Gannon *et al.*, 1979). This sequence is similar, but not identical, to the Pribnow box in that it involves an AT rich region flanked by GC rich regions. Regions of symmetry and homology have been found around the probable start to transcription by comparison of various gene sequences, but further studies are required to determine whether these are recognition sequences for specific regulatory proteins (Cochet *et al.*, 1979).

Restriction of control regions to the sequences preceding the gene transcript, the apparent case in prokaryotes, does not necessarily apply to eukaryote genes. Bogenhagen (cited in McKay, 1979) has found that the sequences important for expression of *Xenopus* 5S ribosomal RNA genes are intragenic, that is, the critical sequence of 30 bases for RNA polymerase III initiated transcription is found 50 bases inside the transcribed region of the gene. The extent of this phenomenon is unknown, but it opens the possibility that amino acid sequences, previously thought essential to the function of a peptide because of their conservation in the analogous genes of different species, could be retained as the consequence of a conserved nucleotide sequence control region within the coding region of the gene.

C. Eukaryote messenger RNA

A major difference between prokaryote and eukaryote cells is the partitioning of transcription and translation processes in eukaryotes by a nuclear membrane packaging the DNA into a nucleus. In prokaryotes these two processes are often concurrent events, degradation of the mRNA template commencing before transcription has terminated.

Eukaryote gene expression involves the transport of transcripts from the nucleus to the cytoplasm where translation occurs, the mRNA sequences consequently exhibit marked differences from their prokaryote counterparts.

Prokaryote mRNA sequences contain a purine rich ribosome binding site adjacent to the AUG initiation codon (Shine and Dalgarno, 1975), a counterpart to this sequence does not exist in eukaryotes (Baralle and Brownlee, 1978). Instead, the 5' terminus is post-transcriptionally modified by the addition of a 7-methyl-G "cap" structure (Furuichi *et al.*, 1975; Adams and Cory, 1975), the addition of which appears to be all that is essential for ribosome binding and subsequent efficient translation (Paterson and Rosenburg, 1979; Schroeder *et al.*, 1979). The proximity of the 7-methyl-G cap to the AUG initiation (i.e., the length of the 5' untranslated region) is critical (Rosenburg and Paterson, 1979) and in certain genes intervening sequences are spliced from the 5' untranslated region during mRNA maturation (Gannon *et al.*, 1979; Cochet *et al.*, 1979; Lomedico *et al.*, 1979). In these instances RNA splicing can been seen to be contributing toward the expression of the gene.

The 3' terminal region of eukaryote mRNA contains sequences in addition to those required to code for the protein, the exact function of which still remains unresolved. Most mRNAs have a post-transcriptionally added poly(A) tract which may be involved in stability of the RNA by imparting resistance to nuclease attack and has also been implicated in transport from the nucleus to the cytoplasm (Brawerman, 1976). A common AAUAAA sequence about 20 bases from the poly(A) tract may be involved in poly(A)

addition (Proudfoot and Brownlee, 1976). Regions of symmetry also exist as does a high degree of homology between analogous sequences in different species, but the functional significance remains speculative.

Konkel *et al.*, (1979) and Nishioka and Leder (1979) have shown that the major mutational change in a sequence is performed by insertion and deletion (rather than base change), and that the 3' untranslated regions in analogous genes have not altered significantly in length. Since insertion and deletion causes a change in the length of a sequence and some constraint on length is apparent in this region, the observed homology may be a reflection of this constraint, by disallowing the major method of mutation (insertion and deletion). Such an explanation implies functional significance to the 3' untranslated region dependent on its length, such as the proximity alignment of a bound protein, which is of sufficient importance for undesired mutations to be selected against. The co-ordinate increase in length of human α and β globin mRNA 3' untranslated regions compared to those of rabbit may reflect a change in humans of the organization of specific protein binding (Proudfoot, 1977; Proudfoot *et al.*, 1977).

Messenger RNA is isolated from polysomes in the form of a ribonucleoprotein (RNP) complex. In the case of globin sequences this RNP complex sediments at 20S compared to 9-10S for naked RNA. Therefore, it would appear that a significant amount of protein is bound *in vivo*. The specificity and organization of this protein is currently under investigation (see Preobrazhensky and Spirin, 1978, for review).

D. Globin genes

Hemoglobin is one of the most heavily researched of biological molecules, encompassing a wide variety of approaches from X-ray crystallography to population genetics. As such it is often the first approach used in the development of new research technologies and has earned the unfortunate title of 'model system'.

i) Hemoglobin structure and function

The major physiological role of hemoglobin is in the transport and distribution of oxygen. The oxygen molecules are reversibly attached to the iron atoms found at the centre of the heme groups. The peptide components of hemoglobin are therefore not directly involved in the functional activity of the molecule, instead providing the necessary structure for the reversible binding to occur (Kendrew *et al.*, 1961).

The functional molecule consists of four peptide sequences, two α-like and two β-like chains, each of which contains a heme group attached via the side chain of histidine residues. The biosynthetic pathway of heme is well understood as are the regulatory enzymes in the synthesis, δ-aminolevulinic acid synthetase and dehydrase, both of which are inhibited by the end product, heme (Lehninger, 1970).

Hemoglobin biosynthesis is under the direct control of hemin (the Fe^{+++} form of heme) which acts as a positive effector at the level of translation of globin peptide chains, by inhibition of the phosphorylation of the initiation factor, eIf-2. Phosphorylation of eIf-2 results in decreased binding of the Met-$tRNA_f$-GTP-eIF-2 complex to

40S ribosomal subunits, and presumably therefore to decreased initiation of translation (Tahara *et al.*, 1978). These fine controls are effective in the stoichiometric synthesis of hemoglobin components, but their relationship to differentiation of erythroid tissue at the level of gene transcription is undetermined.

As well as histidine residues, essential for heme binding, certain amino acids are conserved both in sequence and location in the hemoglobin peptide chains (Hunt and Dayhoff, 1976). These amino acids can be assigned to various functional requirements of the hemoglobin molecule such as heme contact, peptide chain contact (both $\alpha\beta$ and like chain), Bohr effect and 2,3-diphosphoglycerate binding (Perutz and Ten Eyck, 1972; Goodman *et al.*, 1975). Functions present in hemoglobin and not in ancestral globin chains have arisen through positive selection indicated by the evolutionary transition of monomeric globin to a tetramer, which has a sigmoidal oxygen equilibrium curve and, therefore, a more effective function.

Characteristic differences have developed between α and β chains. Alpha chains are 141 amino acid residues in length and their C-terminal amino acid is arginine, while β chains are 146 amino acid residues long and their C-terminus is histidine.

ii) Classical genetic analysis

The analysis of the genes of higher eukaryotes often suffers from the inability to isolate and study mutant genes. Studies on hemoglobin genes in man have to some extent overcome this problem through analysis of genetic variants in the population.

The genetic disorders involved in hemoglobin structure and synthesis can be divided into three categories (Weatherall and Clegg, 1979):

1. Structural hemoglobin variants - which arise from mutations in the coding region of the gene(s).
2. Thalassemias - which involve a reduced rate of synthesis of either of the α-like or β-like chains.
3. Hereditary persistence of fetal hemoglobin (HPFH)- where the switch from fetal to adult hemoglobin synthesis is affected.

In the first category α chain mutations affecting the termination codon, Hb Constant Spring and Hb Wayne, indicated that the mRNA sequence carried an untranslated region between the normal terminator codon and the poly(A) tract. Comparison of the 'read-through' C-terminal extensions of these two mutants enabled the accurate prediction of part of the nucleotide sequence, subsequently determined by direct analysis (Wilson *et al.*, 1977).

The translation of both Hb Constant Spring and Hb Wayne mRNAs is drastically reduced and the mRNA is unstable suggesting that a functional role of the 3' untranslated region is being interfered with due to translational 'read-through'. As a consequence of reduced synthesis these variants are phenotypically thalassemias. Most thalassemias and HPFH occur as a result of deletions, often as a consequence of unequal cross-over, in the β-like gene locus. In the cases of Hb Lepore and Hb Kenya, a fused gene is produced. These studies indicated the linkage of the β-like genes and their order in the locus (Weatherall

FIGURE I,1.

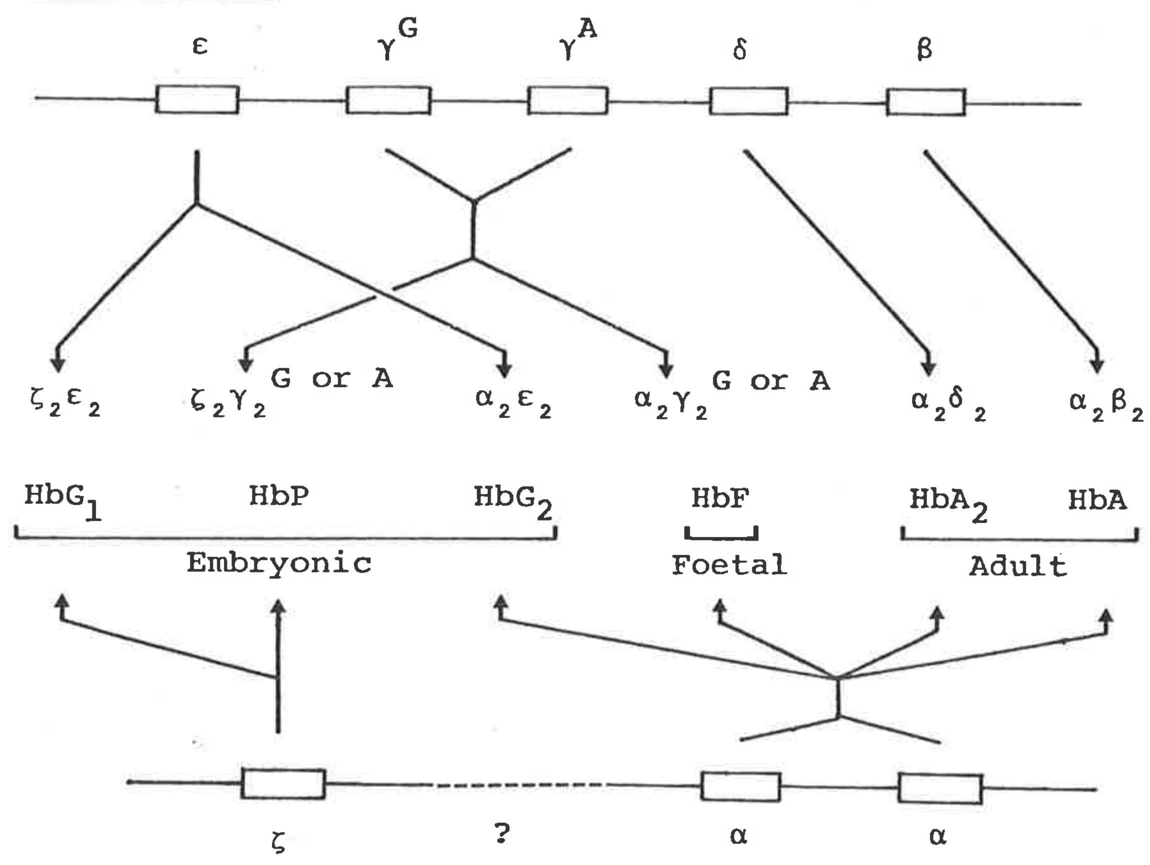
ORDER OF THE HUMAN β-LIKE GLOBIN GENE LOCUS

This figure shows the genes responsible for the hemoglobins expressed during human development and indicates the coincident transcriptional and developmental order of the genes in the β-like locus. Data is from Weatherall and Clegg (1979) with modifications from Dahl and Flavell (pers. comm.).

Abbreviations: HbG_1 - Hemoglobin Gower 1.
HbP - Hemoglobin Portland.
HbG_2 - Hemoglobin Gower 2.

Note: It is not known whether the ζ gene is linked to the α genes.

FIGURE I,1.



and Clegg, 1979).

iii) Globin gene switching

A curious and to date inexplicable feature of globin genes is their heterogeneity through development, the so-called gene switching. Different α-like and/or β-like chain components appear in the hemoglobin at the embryonic, fetal and adult stages of development, and at any one stage more than one chain type may be present. Furthermore, the locus of β-like chain genes in humans is transcriptionally ordered in the same sense as the genes are expressed in development (Dahl and Flavell, personal communication), see Figure I,1.

Although there are distinct chemical differences between the hemoglobins, there appears to be no functional difference between one type and the next (see Ingram, 1963). The change in synthesis often accompanies a change in site of erythropoiesis, the switch in phenotype being the result of the appearance of a new population of erythroblasts. This phenomenon was characterized in stress induced switching in sheep (HbA($\alpha_2\beta_2^A$) to HbC($\alpha_2\beta_2^C$)), where stem cells, committed to form HbC synthesising erythroblasts, appear in the bone marrow (Benz, cited in Nienhuis and Stamatoyannopoulos). These stem cells could arise from differentiation of some uncommitted cells in the bone marrow (Caplan and Ordahl, 1978) responding to the erythropoietin hormone, as has been found to be the case with estrogen and ovalbumin synthesis in non-target tissues of chicken (Tsai *et al.*, 1979).

The commitment to synthesise a particular hemoglobin type(s) is not stringent as primitive embryonic red cells of the mouse yolk sac synthesise not only fetal but

TABLE I,1.

CHICKEN HEMOGLOBIN PEPTIDE CHAINS

(Brown and Ingram, 1974)

Hemoglobin Types		α-like chain	β-like chain
[a]Early Embryonic	P	π,π'	ρ
	E	α^A	ε
	M	α^D	ε-like
[b]Late Embryonic -Early Adult	H	?	?
	A	α^A	β
Adult	D	α^D	β

a. Up to 14-16 days

b. 6-7 days to 50-90 days (Bruns and Ingram, 1973).

also adult globins from 13-14 days (Chui, cited in Neinhuis and Stamatoyannopoulos, 1978), and conversely early erythroblasts of the definitive cell series in chickens contain embryonic globins, in spite of the fact that later in development they produce only adult globins (Chapman and Tobin, 1979).

Switching of hemoglobin types in mammals, as a result of anaemic stress or normal development is mainly characterized by a change in the β-like chain component. Genes in the human β-like locus are transcriptionally and developmentally coincident (see Figure I,1), which is consistent with the hypothesis that differentiation in erythropoiesis involves firstly a "switching on" of pertinent regions of the genome, then selective and temporal expression of the genes in these loci. The phenotype of an erythroid cell would result from the phase of differentiation a cell was at when the force of differentiation was removed, presumably when released into the bloodstream (Baglioni, cited in Ingram, 1963). Severe anaemia in an adult would induce the formation of a cell line (e.g., HbC type in sheep) which is committed to premature release, or has limited reponse to the forces of differentiation.

E. Chicken erythroid tissue

Erythropoiesis in the chicken has been the subject of intensive study (Bruns and Ingram, 1973; Brown and Ingram, 1974). As in other species the peptide chain components of chicken hemoglobin alter throughout development.

i) Hemoglobins

Table I,1 from Brown and Ingram (1974) summarizes the chicken hemoglobin types and their α-like and β-like

chain components. From these immunological studies it is clear that there are at least 3 α-like chains (possibly five) and at least 3 β-like chains (possibly five). Although the changes in chain type differ, the chicken hemoglobins bare a striking resemblance to those in man (see Figure I,1) both in number and developmental switching.

ii) Tissue specific non-globin proteins

Avian red blood cells differ from those of mammals in that they retain their nucleus. Coincident with this retention, probably related to it, avian erythroid cells produce two tissue specific nuclear proteins, histone H_5 (Neelin and Butler, 1961) and high mobility group protein, HMG-E (Sterner *et al.*, 1978).

(a) Histone H_5

The histones H_{2A}, H_{2B}, H_3 and H_4, together with about 200 base pairs of DNA cluster in a repeating unit called a nucleosome (Kornberg, 1974). The H_1 histone proteins, a heterogenous group, are bound on the linker DNA between nucleosomes giving eukaryote chromatin its bead-on-string arrangement (Olins and Olins, 1974). The function of this structure and the role of the H_1 histones remain unclear. In avian erythroid tissues H_1 is to a great extent replaced by another histone, H_5, (Dick and Johns, 1969). The inactivity of the nuclei of these cells in RNA and DNA synthesis indicates that histone H_5 may function in suppressing these processes, condensing the nucleus into an inactive form. This argument is supported by the finding of H_5 in the red blood cells of other species which retain their nuclei (Goetz *et al.*,

1978). The unique synthesis of this histone during the maturation of the red blood cells (Appels and Wells, 1972) suggests that the control of gene expression may be linked to that of the globin genes. Recently histone H_5 mRNA has been purified (Molgaard *et al.*, 1980) and a comparison of the nucleotide sequence of the mRNA and gene with those coding for the chicken globins may provide some clue to the unique expression of these otherwise unrelated genes in erythroid tissue.

(b) High mobility group protein, HMG-E

Interest in the high mobility group (HMG) of non-histone chromosomal proteins (NHCP) was stimulated by the finding of their preferential release from chromatin by limited DNAase I digestion under conditions which selectively digest active genes (Vidali *et al.*, 1977). Detailed analysis of the primary structures of HMG proteins from a variety of tissues and species suggests that there is far too little specificity for these proteins to act as specific gene regulators (Watson *et al.*, 1977; Sterner *et al.*, 1978). Since these proteins are capable of interacting with DNA and H_1 histones (Shooter *et al.*, 1974) they may be involved in maintaining the structural configuration of an active gene.

Avian erythroid tissue contains a tissue specific high mobility group protein, HMG-E (Sterner *et al.*, 1978), and since HMG proteins can interact with histone H_1, the HMG-E may be involved in the same interaction with histone H_5, which replaces H_1 in this

tissue.

Like H_5, the gene for HMG-E is of interest since potentially it is co-ordinately expressed, and therefore controlled, with the chicken globin genes.

iii) Viral transformation and globin gene expression

The difficulty of *in vivo* experiments and inability to isolate mutants has seen retarded progress in the field of eukaryote gene control compared with that in prokaryotes. The use of transforming viruses (Therwath and Scherrer, 1978) and their temperature-sensitive mutants (Graf *et al.*, 1978; Keane *et al.*, 1979) which are capable of blocking differentiation promises a means of studying cellular differentiation at a molecular level.

Therwath and Scherrer (1978) and Graf *et al.* (1978) infected adult erythroid tissue with avian erythroblastosis virus (AEV) and isolated transformed cells. Therwath and Scherrer identified their transformed "erythroblasts" from peripheral blood by the continued presence of the tissue specific histone, H_5, (see I, 2, E , ii). They analysed the RNA of these cells and found that all sequences containing globin RNA were restricted to the nucleus. They concluded that the virus may affect differentiation by interfering with post-transcriptional controls, such as splicing. The finding that histone H_5 is still expressed in these cells would be consistent, if H_5 like the other histone genes previously characterized (Levy *et al.*, 1979) contains no intervening sequences and therefore its expression is not dependent upon splicing.

Graf *et al.* (1978) introduced the use of a temperature-sensitive mutant of AEV (ts34 AEV) in the analysis

of erythroid cell differentiation. Bone marrow cells transformed with ts34 AEV were capable of producing only a very limited amount of hemoglobin at 35°C (comparable to that of wt AEV erythroblasts) but at 41°C synthesis increased dramatically, whereas that in wt AEV erythroblasts showed only a marginal increase. This temperature-sensitive effect could be reversed by lowering the temperature to 35°C, and the authors concluded that an AEV encoded gene product is involved in the block of hemoglobin synthesis.

Transformation of pure cultures of primary mesenchymal cells by wt AEV and ts34 AEV and analysis of the cells and their products (Keane *et al.*, 1979) provided the most extensive study to date. Transformed cells were found blocked in the synthesis of globin peptide chains, heme and tissue specific histone H_5. These experiments differ from the previous two reports in that mesenchymal cells contain precursors for both the primitive and definitive cell series, which may respond differently to AEV transformation.

Further analysis of transformed cells and the use of a wider variety of experimental techniques, including specific nucleotide sequence probes and DNAse I sensitivity, should provide an insight into the molecular mechanisms of specific gene expression.

CHAPTER II

MATERIALS AND METHODS

CHAPTER II - MATERIALS AND METHODS

1. Materials

A. Chemicals and Reagents

All chemicals were of analytical reagent grade, or the highest available purity. The sourcesof the more important chemicals and reagents are listed.

Acrylamide : Sigma - twice recrystallized from $CHCl_3$.

Ampicillin : a gift from Beecham.

Bromophenol Blue : B.D.H. (Australia).

Caesium Chloride : Harshaw.

Chloramphenicol : a gift from Parke-Davis.

Diethyl Pyrocarbonate : Sigma.

Dimethyl Sulphate : Tokyo Kasei.

Dithiothreitol : Sigma.

Ethidium Bromide : Aerosol Industries.

Ficoll 400 : Pharmacia.

Formic Acid : Carlo Erba.

Heparin : Sigma.

HEPES : Sigma.

Hydrazine Hydrate : Tokyo Kasei.

Nitrocellulose Filters : Schleicher and Schuell.

N,N'-methylene bisacrylamide : Sigma.

Nucleoside triphosphates (ATP, dATP, dCTP, dGTP, dTTP) : Sigma.

α-^{32}P-dCTP : (50-350 Ci/mmole) a gift from Dr. R.H. Symons, (350-1500 Ci/mmole) New England Nuclear.

α-^{32}P-dGTP : (50-350 Ci/mmole) a gift from Dr. R.H. Symons.

PEI cellulose (POLYGRAM Cel 300 PEI/UV_{254}) : Macherey-Nagel.

Phenylhydrazine : Sigma.

Piperidine : B.D.H. (Australia).

^{32}P-orthophosphoric acid (in water) : New England Nuclear.

Polyvinyl pyrollidone : May and Baker.

Sephadex G-100 (coarse) : Pharmacia.

Spermine tetrahydrochloride : Calbiochem.

TEMED : Tokyo Kasei.

Tetracycline : a gift from Commonwealth Serum Laboratories.

Xylene Cyanol FF : Tokyo Kasei.

B. Enzymes

Enzymes used in this study were obtained from the sources listed.

Avian Myeloblastosis Virus RNA dependent DNA polymerase (Reverse Transcriptase) : was a gift to Dr. J.R.E. Wells from Dr. J.W. Beard.

Bacterial Alkaline Phosphatase (BAPF) : Worthington.

E. coli DNA Polymerase I : Boehringer-Mannheim.

E. coli DNA Polymerase I, Klenow fragment : Boehringer-Mannheim.

Glyceraldehyde-3-phosphate dehydrogenase (from rabbit muscle) : Calbiochem.

3-Phosphoglycerate kinase (from yeast) : Calbiochem.

Proteinase K : Boehringer-Mannheim.

Restriction Endonucleases : unless otherwise indicated all restriction endonucleases were purchased from New England Biolabs.

*Eco*RI - a gift from M. Betlach.

*Hae*III, *Alu*I - a gift from F. DeNoto.

*Hind*III - a gift from R.P. Harvey.

*Hpa*II - a gift from D. Clark-Walker.

*Hsu*I - a gift from E. Tischer.

Ribonuclease : Sigma.

S_1 nuclease : Miles.

T_4 DNA ligase : a gift from H. Heynecker.

T_4 polynucleotide kinase : Boehringer-Mannheim.

C. Specialized materials

i) Bacterial strains

The following derivatives of *E. coli* K12 were used for transformation and propagation of recombinant DNA in this study:

HB101 F^- *pro leu thi lac* Y $Str^r r_K^- m_K^-$ EndoI$^-$, *rec*A$^-$ (Boyer and Roulland-Dussoix, 1969), RRI F^- *pro leu thi lac* Y $Str^r r_K^- m_K^-$ (Bolivar *et al.*, 1977), χ1776 F^- *tau*A53 *dap*D8 *mer*A1 *sup*E42 Δ40 (*gal-uvr*B) λ^- *min*B2 *mal*A25 *thy*A57 *met*C65 Δ29 (*bio*H-*asd*) *cys*B2 *cyc*A1 *Hsd*R2 (R. Curtiss III) were all obtained from Dr. F. Bolivar.

The plasmid cloning vehicle pBR322 in HB101 was also obtained from Dr. F. Bolivar.

ii) Nucleic acids

Many purified nucleic acids and their derivatives were used in this study, the sources of which are indicated:

Calf thymus DNA primers (DNA digested with micrococcal nuclease to an average length of 12 nucleotides) : the gift of Dr. B. Cordell.

E. coli λ DNA (both native DNA and *Eco*RI digested, ^{32}P-labelled) : a gift of P. Seeburg.

(dCCAAGCTTGG) linker DNA : Collaborative Research.

E. coli tRNA : Sigma.

Fd coliphage RF DNA : a gift of P. Seeburg.

M. luteus DNA : Miles.

Oligo(dT)$_{12-15}$: P.L. Biochemicals.

Oligo(dT) cellulose (type T_3) : Collaborative Research.

Polyriboadenylic acid : P.L. Biochemicals.

iii) Experimental animals

White Leghorn Australorp chickens, 12 weeks old, were obtained from the Parafield Poultry Research Station of the Department of Agriculture, Parafield, South Australia.

D. Buffers and media

All buffers and media were prepared with glass distilled water and sterilized by autoclaving. Solutions containing labile chemicals were sterilized by filtration through Millipore apparatus or treatment with diethyl pyrocarbonate.

L Broth - 1% (w/v) Bacto-Tryptone, 0.5% (w/v) Yeast Extract, 1% (w/v) NaCl, pH 7.5 with NaOH.

χ Broth - L broth with 100 μg/ml diaminopimelic acid, 40 μg/ml thymine.

Agar Plates - Respective broth with 1.5% (w/v) Bacto-agar.

Denhardt's Solution - 0.2% (w/v) each of bovine serum albumin, ficoll and polyvinyl pyrrolidone (stored frozen, not sterilized).

2. Methods

A. mRNA purification

i) Induction of anaemia (Williams, 1970)

Animals used were White Leghorn Australorp hens, 12 weeks of age, obtained from Parafield Poultry Research Station of the Department of Agriculture, Parafield, South Australia. Hemolytic anaemia was induced by the subcutaneous injection of 2.5% (w/v) phenylhydrazine in 47.5% ethanol (v/v), pH 7.0. The course of injections (0.4 - 0.6 ml) lasted five days.

ii) Isolation of polysomes (Pemberton *et al.*, 1972)

On the sixth day blood was collected by heart puncture into NKM (0.15 M NaCl, 5 mM KCl, 2 mM $MgCl_2$) containing heparin (1 mg/ml). All further steps, unless otherwise stated, were carried out at 4°C.

Cells were washed twice in NKM, buffer and less dense white cells removed by aspiration, then lysed by the addition of 2 mM $MgCl_2$ and vortex mixing for 2 minutes. Tonicity was restored by the addition of an equal volume of TKM (10 mM Tris-HCl, pH 7.4, 0.2 M KCl, 2 mM $MgCl_2$) and cellular debris removed by centrifugation at 5,000 x g for 10 minutes. Polysomes were collected from the supernatant by centrifugation at 200,000 x g for 60 minutes through a pad of 50% (w/v) sucrose in ½ x TKM (5 mM Tris-HCl, pH 7.4, 0.1 M KCl, 1 mM $MgCl_2$).

iii) mRNA isolation (Pemberton *et al.*, 1972)

Polysomes were resuspended in TK (10 mM Tris-HCl, pH 7.4, 15 mM KCl) and dissociated into ribonucleoprotein (RNP) particles by the addition of half the volume of 1.0 M Tris-HCl, pH 9.0 followed by $^1/_{10}$ the volume of

0.3 M EDTA, pH 7.6. The 20S ribonucleoprotein particles were collected by fractionation of the dissociated polysomes through a linear 10-40% (w/v) sucrose gradient centrifuged at 210,000 x g for 16 hours. Fractionation of gradients was carried out on an ISCO density gradient fractionator with monitoring at A_{254}.

After ethanol precipitation and redissolving in 10 mM Tris-HCl, pH 9.0, 0.1% (w/v) sodium dodecyl sulphate containing 0.25 mg/ml Proteinase K, the 20S ribonucleoprotein particles were incubated at 37°C for 3 minutes, then loaded on to linear 10-40% (w/v) sucrose gradients in NET (10 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA). After centrifugation at 210,000 x g for 16 hours gradients were fractionated on an ISCO density gradient fractionator, monitoring at A_{254}, and the 9-10S RNA collected.

B. Oligo(dT) cellulose chromatography (Aviv and Leder, 1972)

Fractionation of RNA selecting for poly(A) containing species using oligo(dT) cellulose chromatography was carried out at room temperature.

Preswollen oligo(dT) cellulose (0.3 g) was mixed into a slurry with twice distilled water and poured into a glass column (1 cm x 10 cm) plugged with glass wool. The column was then washed with 0.1 M NaOH to remove fines and sterilize the column from ribonuclease activity.

The A_{254} of eluate was monitored using a Uvicord U.V. spectrophotometer. The column was equilibrated with hybridization buffer, H (0.5 M NaCl, 10 mM Tris-HCl, pH 7.5, 2 mM EDTA), and elution buffer, E (10 mM Tris-HCl, pH 7.5, 2 mM EDTA).

Prior to loading, the column was again equilibrated with H buffer, the sample dissolved and loaded in this buffer, and unbound material collected while washing with buffer H until the A_{254} trace returned to zero. Bound material was then eluted from the column using buffer E.

C. Polyadenylation of RNA

RNA was polyadenylated at its 3' terminus using terminal riboadenylate transferase activity isolated from *E. coli* (Sippel, 1973) and ATP. Polyadenylated RNA was separated from non-polyadenylated RNA and the incubation mixture by chromatography on oligo(dT) cellulose.

Polyadenylation was carried out in a 100 µl reaction containing 200 mM NaCl, 50 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 0.2 mM (^3H)ATP (100 µCi/µmol), 0.2 mg/ml RNA and 0.2 units (Sippel, 1973) enzyme, incubated at 37°C for 30 mins. Reactions were stopped by the addition of 400 µl of buffer H and chilling on ice.

D. Preparation of high molecular weight DNA (Gross-Bellard *et al.*, 1973)

Blood was collected by heart puncture from non-anaemic White Leghorn Australorp chickens. Cells were washed and lysed as described (II,2,A). The cell debris pellet containing intact nuclei was resuspended and washed twice in NKM, then added, in suspension, to a slowly stirring solution (10 mM NaCl, 10 mM Tris-HCl, pH 8.0, 10 mM EDTA, 0.5% sodium dodecyl sulphate), containing 100 µg/ml Proteinase K, dropwise over a 12 hour period at 37°C. A further 100 µg/ml of Proteinase K was added and the incubation continued for 12 hours. The reaction was stopped by the addition of

half the volume of phenol saturated with PSB (200 mM NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 50 mM β-mercaptoethanol) gentle shaking for 1 minute, the addition of half the volume of chloroform, shaking and centrifugation at 5,000 x g for 10 minutes to separate the phases. The aqueous phase was removed and dialysed against three changes of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA over 16 hours. The solution was then made 20 μg/ml in ribonuclease A, previously heat treated at 80°C for 20 mins to inactive DNase activity, and incubated at 37°C for 2 hours. Phenol, chloroform extraction and dialysis were performed as described above and the resultant DNA solution stored at 4°C.

E. Synthesis of cDNA (Verma *et al.*, 1972; Seeburg *et al.*, 1977a)

Two reactions were used for cDNA synthesis, one for analytical studies which was optimized for radioactive label incorporation and the second for preparative purposes, optimized for yield of cDNA. The major difference in the two reactions was the concentration of the radioactively labelled triphosphate which may have been rate-limiting.

i) Analytical cDNA synthesis

Reaction mixtures of 10 μl contained 50 mM Tris-HCl, pH 8.3, 20 mM KCl, 7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1.0 mM EDTA, 0.5 mM each of dGTP, dCTP and dTTP, 40 μM α-^{32}P-dATP (10 Ci/mmole), 20 μg/ml oligo$(dT)_{12-15}$, 100 μg/ml RNA and 3 units of AMV RNA-dependent DNA polymerase (reverse transcriptase). Incubation was at 42°C for 10 minutes and the reaction was stopped by the addition of 40 μl of ice cold water, and heating at 90°C for 2 minutes to melt RNA/DNA hybrids.

ii) Preparative cDNA synthesis

The reaction was carried out in a 0.5 ml incubation mixture containing 50 mM Tris-HCl, pH 8.3, 20 mM KCl, 7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.1 mM EDTA, 0.5 mM each of dGTP, dCTP and dTTP, 0.1 mM α-^{32}P-dATP (0.05 Ci/mmole), 30 μg/ml oligo$(dT)_{12-18}$, 150 μg/ml RNA and 50 units of AMV reverse transcriptase at 42°C for 30 minutes.

F. Second strand cDNA synthesis (Ullrich *et al.*, 1977)

The first cDNA strand reaction (II,2,E,ii) was stopped by addition of 300 μl of fresh 0.3 M NaOH, 1 mM EDTA. Alkaline hydrolysis of the RNA template was carried out at 37°C for 60 minutes. After neutralization, by addition of 50 μl of 1 M Tris-HCl, pH 7.3 and 70 μl of 1 N HCl, the cDNA was isolated by chromatography over Sephadex G-100 in a column (0.3 x 10 cm) equilibrated and run in 0.1 M NaCl, 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, then ethanol precipitated by addition of 0.05 volume of 5 M NH_4OAc, pH 7.5 and 2.5 volumes ethanol, at -80°C for 1 hour.

Second strand synthesis was carried out in a 40 μl reaction containing 50 mM Tris-HCl, pH 8.3, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM each of dGTP, dATP and dTTP, 200 μM α-^{32}P-dCTP (12.5 Ci/mmole), 2.5 μg cDNA and 60 units of AMV reverse transcriptase incubated at 37°C for 5 hours.

The reaction was stopped by the addition of EDTA, pH 8.0 to 20 mM and protein extracted by the addition of half the volume of phenol saturated in PSB, vortex mixing and the addition of half the volume of chloroform, vortex mixing and centrifugation in a Eppendorf microfuge for 2 minutes. To the aqueous phase 0.2 volume of loading buffer (50% (w/v) sucrose, 4 mM EDTA, 0.05% (w/v) bromophenol blue)

was added and the mixture fractionated over Sephadex G-100 as described previously.

G. Restriction endonuclease digestions

All restriction endonuclease digestions were performed according to the conditions of the supplier (New England Biolabs).

DNA was dissolved in water and 0.1 volume of 10 x *Hae*III buffer (66 mM Tris-HCl, pH 7.5, 66 mM $MgCl_2$, 66 mM β-mercaptoethanol) added. The concentrations of NaCl and KCl were adjusted to those recommended for the particular restriction endonuclease, as was the temperature of incubation (usually 37°C). Sufficient units of enzyme activity were added so as to give at least 5-fold over complete digestion, as judged by pilot experiments using λ or pBR322 DNA. Where enzyme preparations were found to be unstable, further aliquots were added during the incubation.

H. Routine gel electrophoresis

i) Polyacrylamide gel electrophoresis (Peacock and Dingham, 1967)

Electrophoresis of DNA was carried out on vertical 16 x 16 x 0.15 cm slab polyacrylamide gels of varying percentages, using a 19:1 ratio of acrylamide to N,N'-methylenebisacrylamide.

For 5% gels a 50 ml mixture containing 5 ml of 10 x TBE (0.5 M Tris-borate, pH 8.3, 10 mM EDTA), 12.5 ml of 20% acrylamide monomer (10 g acrylamide, 1 g N,N'-methylenebisacrylamide in 100 ml of water), 0.5 ml of 10% (w/v) ammonium persulphate and 25 μl TEMED was poured into a gel mould and allowed to polymerize.

Gel reservoir tanks contained approximately 250 ml of TBE buffer, and gels were pre-electrophoresed at 200 volts for 20-30 minutes.

Samples were dissolved in 20 μl of water, 5 μl of 5 x loading buffer (50% (w/v) sucrose, 4 mM EDTA, 0.05% (w/v) xylene cyanol FF, 0.05% (w/v) bromophenol blue; which included 0.5% Sarkosyl if protein was in the sample) and layered directly into the gel slots (1 or 1.5 cm wide). Gels were electrophoresed at 200 volts for 1½ hours or until dyes had moved the desired distance.

ii) Agarose gel electrophoresis

DNA was electrophoresed on vertical 16 x 16 x 0.3 cm slab agarose gels of varying percentages.

For 1% gels a 50 ml mixture containing 0.5 g agarose, 0.5 ml 10 x TEAS buffer (40 mM Tris-acetate, pH 8.15, 20 mM NaCl, 2 mM EDTA) and 45 ml of water was refluxed to dissolve agarose, poured into a gel mould and allowed to set at room temperature.

Gel reservoir tanks contained approximately 250 ml of TEAS buffer and electrophoresis was at 50 volts usually for 4 hours.

Samples were dissolved in 20 μl of water, 5 μl of loading buffer (25% (v/v) glycerol, 5% (w/v) sodium dodecyl sulphate, 25 mM EDTA, 0.05% (w/v) bromophenol blue) added and loaded directly into gel slots (0.4 or 0.9 cm).

I. Autoradiography

DNA, radioactively labelled with ^{32}P, which had been electrophoresed on polyacrylamide gels (II,2,H,i) was visualized by autoradiography.

Where DNA bands had low levels of radioactively (less

than 3 x 10^4 c.p.m.) polyacrylamide gels were dried, by heating to 65°C over a vacuum apparatus, on to a blotting paper (3MM) backing and autoradiographed at -80°C with a salt intensifying screen (Dupont Cronex Lighting Plus) and Kodak X-Omat X-ray film.

Where DNA bands were intended to be eluted from the gel, one of the glass plates was removed and plastic wrap (Saran or Gladwrap) was placed on the gel, two strips of cellutape adhered to the sizes and spotted with radioactive ink (to act as markers). A sheet of envelope packed X-Omat X-ray film was placed on the gel and autoradiographed for the desired length of time (10 minutes to 4 hours). After developing and fixing, the sheet of X-ray film was lined up with radioactive ink spots and needle holes pierced through the film and gel to mark the location of radioactive DNA bands.

Bands were cut from the gel with a scalpel and Cerenkov emission counted in a Packard Liquid Scintillation counter.

J. Electroelution

Slices of polyacrylamide gel containing DNA fragments were placed in 18/32 dialysis tubing (previously boiled in EDTA/$NaHCO_3$ then washed and boiled in water) with 0.8 ml of ½ x TBE buffer (II,2,H,i) sealed by clamping and placed in an electroelution chamber containing ∿500 ml of ½ x TBE, between the electrodes. A current of 50 mA (∿50 volts) was applied overnight to electrophorese the DNA out of the gel slice into the buffer.

The buffer was removed from the dialysis tubing, Cerenkov emission counted in a Packard Liquid Scintillation

counter and the DNA ethanol precipitated by addition of 0.05 volume of 5 M $(NH_4)_2SO_4$ and 2.5 volumes of ethanol at -80°C for 1 hour.

K. Construction of recombinant DNA

i) Blunt ending reactions (Seeburg *et al.*, 1977b)

The hairpin loop of double-stranded chicken globin cDNA was opened by use of the endonucleolytic activity of nuclease S_1 in a 50 µl reaction containing 0.3 M NaCl, 30 mM Na-acetate, pH 4.6, 4.5 mM $ZnCl_2$, 1.3 µg double-stranded cDNA (as judged by ^{32}P incorporation) and 125 units of S_1 nuclease incubated at 22°C for 30 minutes then 10°C for 15 minutes. The reaction was terminated by addition of 0.11 volume of 1 M Tris-HCl (pH 9.5). After addition of EDTA to 10 mM, the reaction mix was phenol extracted and chromatographed on Sephadex G-100 as described previously (II,2,F). Peak fractions containing double-stranded cDNA were pooled, 0.05 volume of 5 M NH_4OAc, pH 7.5 and 2.5 volumes of ethanol added, the solution vortexed and DNA precipitated at -80°C for 1 hour.

After centrifugation the precipitate was redissolved in water and incubated in a 25 µl reaction mixture containing 60 mM Tris-HCl, pH 7.5, 10 mM dithiothreitol, 8 mM $MgCl_2$, 2 mM each of dGTP, dCTP, dATP and dTTP and 1 unit of *E. coli* DNA polymerase I at 10°C for 15 minutes. EDTA was added to 10 mM and the reaction mix phenol extracted and ethanol precipitated as previously described.

ii) Ligation of restriction endonuclease recognition sites

Oligodeoxynucleotide linker (dCCAAGCTTGG) containing the recognition sequence for *Hin*dIII restriction

endonuclease was phosphorylated in a 10 µl "hot" reaction mixture containing 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 20 µM γ-^{32}P-ATP (1200 Ci/mmole), 25 pmoles of linker DNA and 5 units of T_4 polynucleotide kinase, or a 200 µl "cold" reaction mixture containing 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 0.15 mM ATP, 500 pmoles of linker DNA and 50 units of T_4 polynucleotide kinase, incubated at 37°C for 30 minutes, then stored frozen.

Linker DNA was ligated to the blunt ended cDNA in a 35 µl reaction mixture containing 2 µl "hot" and 3 µl "cold" kinased linker reaction mixtures, from above, and 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 0.3 mM ATP, 1.3 µg double-stranded cDNA (as judged by ^{32}P incorporation) and 0.5 units of T_4 DNA ligase at 10°C for 24 hours. This reaction mixture was then made 40 mM NaCl, 20 mM Tris-HCl, pH 7.5, 6.6 mM $MgCl_2$ and 6.6 mM dithiothreitol in a volume of 100 µl, 5 units of *Hsu*I restriction endonuclease (isoschizomer of *Hin*dIII) added and incubated at 37°C for 3 hours.

After addition of EDTA to 10 mM the reaction mixture was phenol extracted and ethanol precipitated.

cDNA was separated from cleaved linker DNA by electrophoresis on a 5% polyacrylamide gel in TBE buffer (II,2,H,i) and visualized by autoradiography for 1 hour at room temperature (II,2,I). Sections of the gel containing cDNA in the size range of 300 to 700 base pairs (as determined by co-electrophoresis of marker DNA *Hae*III cut pBR322, visualized by staining the gel with ethidium bromide) were cut out and the cDNA electroeluted (II,2,J) and ethanol precipitated.

iii) Ligation of cDNA to plasmid DNA (Ullrich *et al.*, 1977)

Plasmid DNA (pBR322) was linearized by cleavage with *Hsu*I restriction endonuclease (an isoschizomer of *Hin*dIII) in a 50 μl reaction containing 40 mM NaCl, 6.6 mM Tris-HCl, pH 7.5, 6.6 mM $MgCl_2$, 6.6 mM β-mercaptoethanol, 1 μg pBR322 DNA (supercoil form from CsCl gradients: II,2,M,iii) and 5 units of *Hsu*I restriction endonuclease at 37°C for 5 hours.

The 5' terminal phosphate groups were removed from the plasmid DNA by increasing the volume of the reaction to 100 μl with water, adding 0.025 volume of 1 M Tris-HCl, pH 8.0, 0.1 unit of bacterial alkaline phosphatase (previously dialyzed against 25 mM Tris-HCl, pH 8.0) and incubating at 65°C for 30 minutes. The reaction mixture was then made 10 mM EDTA, phenol extracted and ethanol precipitated as previously described (II,2,F).

Polyacrylamide gel fractionated cDNA (II,2,K,ii) was ligated to dephosphorylated *Hsu*I linearized pBR322 DNA in a 20 μl reaction containing 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM ATP, 25 ng of cDNA (as judged by ^{32}P incorporation), 50 ng of plasmid DNA and 0.5 units of T_4 DNA ligase at 10°C for 16 hours. The reaction was stopped by additon of EDTA to 10 mM, phenol extracted and ethanol precipitated (II,2,F).

L. Transformation and selection of recombinants (Ullrich *et al.*, 1977)

i) Transformation

An overnight culture of *E. coli* χ1776 was grown in 5 ml of χ1776 media (II,1,D), in a rotary shaker water-

bath at 37°C. The overnight culture was diluted to an A_{600} of approximately 0.05 with fresh χ1776 media, and grown to an A_{600} of between 0.3 and 0.4. Cells were pelleted and resuspended in 0.4 volumes of ice cold 10 mM NaCl. Unless otherwise indicated, all steps were carried out at 4°C. Cells were again pelleted and resuspended in 0.4 volumes of χ_f buffer (75 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5, 140 mM NaCl) and incubated on ice for 15 minutes. After pelleting cells were resuspended in $^1/_{100}$ volume of χ_f buffer.

The following experiments, where recombinant DNA was in contact with living organisms (*E. coli* χ1776), were performed by either Dr. Axel Ullrich or Dr. John Shine in the P2 facility at the Department of Biochemistry and Biophysics, University of California, San Francisco.

Recombinant DNA, constructed *in vitro* (II,2,K), dissolved in 50 μl of χ buffer, was added to 100 μl of resuspended cells and the suspension incubated on ice for 15 minutes. Cells were subjected to heat shock by warming the suspension to 37°C then incubating at 25°C for 4 min.

After a further incubation on ice for 30 minutes the cells were plated directly on to χ1776 agar plates (II,1,D) containing ampicillin (20 μg/ml) and grown at 32°C for 2 days.

Colonies were picked, using autoclaved toothpicks, on to χ1776 agar plates containing tetracycline (30 μg/ml) to test for sensitivity to tetracycline and χ1776 agar plates containing ampicillin (20 μg/ml) as a control and for temporary storage.

ii) Selection of recombinants

The manipulations, involving recombinant DNA in the presence of living organisms, were performed in the P2

facility by either Dr. Axel Ullrich or Dr. John Shine. Cells containing recombinant DNA were removed from the containment facility only after treatment with chloroform (to 0.1% (v/v).

Those single colonies, having a Ap^rTc^s phenotype, were grown overnight at 37°C in 4 ml of χ broth with 20 μg/ml ampicillin. To 1 ml of the culture glycerol (1 ml) was added and the mixture stored at -20°C. Cells from the remaining 3 ml of culture were chloroform treated, pelleted and resuspended in 100 μl of 25% sucrose, then 40 μl of 0.25 M EDTA, pH 8.0 was added. After addition of 20 μl of lysozyme (5 mg/ml) the mixture was incubated on ice for 15 minutes.

Plasmid DNA was released from the cells by addition of 160 μl of Triton Lysis Mix (0.1% (v/v) Triton X-100, 62.5 mM EDTA, pH 8.0, 20 mM Tris-HCl, pH 8.0) and gentle mixing. Cellular debris containing chromosomal DNA was pelleted by centrifugation for 15 minutes in an Eppendorf microfuge. The pellet was removed, 2 μl of ribonuclease A (1 mg/ml, previously heat treated at 80°C for 20 minutes to inactivate DNA'ase) added and the mixture incubated for 20 minutes at 37°C, after which 20 μl of proteinase K (1 mg/ml) was added and the incubation continued at 37°C for a further 20 minutes.

The reaction mix was phenol-chloroform extracted, as previously described (II,2,F) and a 30 μl aliquot electrophoresed on a 1% agarose gel (II,2,H,ii). The remaining solution was dialyzed against three changes of 0.3 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA over 16 hours, then ethanol precipitated by the addition of 2.5 volumes of ethanol,

mixing, and incubation at -80°C for 1 hour.

The plasmid DNA was pelleted, washed with ethanol, dried in a vacuum dessicator and resuspended in 20 μl of water. One tenth the volume of 10 x *Hae*III buffer (II,2,G) and 1 μl of restriction endonuclease *Hae*III (5 units) were added and the reaction mixture incubated at 37°C for 3 hours. The reaction was stopped by the addition of 5 μl of acrylamide gel loading buffer (with 0.5% (w/v) Sarkosyl) and the DNA electrophoresed on a 6% polyacrylamide gel in TBE (II,2,H,i).

M. Isolation of plasmid DNA

i) Amplification of plasmid DNA (Clewell, 1972)

A single colony of *E. coli* containing the plasmid vector pBR322, or recombinant derivative, was used to infect 5 ml of L broth (II,1,D) and grown overnight at 37°C with vigorous shaking. Antibiotic (ampicillin, 20 μg/ml or tetracycline, 30 μg/ml) was included in the overnight broth to maintain selective pressure for plasmid containing cells.

The overnight culture was diluted 100-fold with fresh broth (without antibiotic) and grown with aeration at 37°C to an A_{600} of 1.0, at which time chloramphenicol was added to a final concentration of 150 μg/ml, and incubation continued with aeration overnight.

Several drops of chloroform were added to kill bacteria (to 0.1% (v/v)) and the cells pelleted by centrifugation. Cell pellets were stored frozen.

ii) Isolation of plasmid DNA (Guerry *et al.*, 1973)

Cell pellets were thawed and resuspended in 10 ml of 25% (w/v) sucrose and 5 ml of 0.25 M EDTA, pH 8.0,

2.5 ml of lysozyme (5 mg/ml) added, and incubated on ice for 15 minutes. 15 ml of Triton lysis mix (II,2,L,ii) was added with gentle mixing and the mixture centrifuged at 90,000 x g for 30 minutes. The supernatant containing plasmid DNA was carefully decanted and treated with 20 μl of ribonuclease A and 250 μl proteinase K, phenol-chloroform extracted as previously described (II,2,L,ii), then dialyzed against three changes of 4 litres of 0.3 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, over 16 hours followed by ethanol precipitation with 2.5 volumes of ethanol at -80°C for 1 hour.

After centrifugation the pellet was washed with ethanol and redissolved in 2 ml of 0.3 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA (0.3 NET buffer) and 0.2 ml of polyacrylamide gel loading buffer (II,2,H,i, without Sarkosyl) added, then the mixture carefully layered on to the top of a 3 x 26 cm Sephadex G-100 column previously equilibrated and run in 0.3 NET buffer. Fractions (5 ml) were collected and the excluded volume peak of A_{254} containing plasmid DNA pooled and ethanol precipitated as above.

DNA yield was checked by measuring an aliquot, resuspended in 0.1 x TE buffer (1 mM Tris-HCl, pH 7.5, 0.1 mM EDTA), in a Zeiss spectrophotometer set at A_{254}, and also by gel electrophoresis on a 1% agarose gel (II,2,H,ii).

A further sample was cleaved with restriction endonuclease *Hae*III, and electrophoresed on a 5% polyacrylamide gel (II,2,H,i) to check fidelity of DNA during amplification.

iii) Isolation of supercoiled DNA

The DNA pellet from (II,2,M,ii) was resuspended

in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA (TE buffer) and CsCl added to a final concentration of 1 g/ml. One tenth the volume of ethidium bromide (10 mg/ml) was added and the mixture centrifuged at 200,000 x g at 15°C for 48 hours. Both bands, identified by fluorescence in U.V. light, were removed by side puncture with a syringe, extracted five times with isoamyl alcohol (saturated with 1 g/ml CsCl solution) to remove ethidium bromide, then dialyzed against TE (3 changes, overnight) then 0.1 x TE. The lower band contained supercoil DNA.

DNA was stored at 4°C in 0.1 x TE.

N. Sequence analysis of DNA

All DNA sequences were determined by the chemical degradation procedure of Maxam and Gilbert (1977, 1979).

i) End labelling DNA fragments

(a) Labelling 5' ends with γ-^{32}P-ATP and T_4 polynucleotide kinase

γ-^{32}P-ATP was prepared in an exchange reaction between ATP and ^{32}P catalyzed by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and 3-phosphoglycerate kinase (3-PGK)(Glynn and Chappel, 1964).

3-PGK, 20 mg, was dialyzed against two changes of 1 litre of 3.2 M ammonium sulphate, pH 8.0, 50 mM Tris-HCl, pH 8.0, 10 mM β-mercaptoethanol, 1 mM EDTA at 4°C for 16 hours. GAPDH, 100 mg, was dialyzed under identical conditions except that 0.1 mM nicotinamide adenine dinucleotide (NAD^+) was included in the dialysis buffer. Both enzymes were stored as a suspension at 4°C.

^{32}P, as orthophosphoric acid in water, was eva-

porated to dryness in a vacuum dessicator and re-dissolved to a final concentration of 0.5 mCi/μl in 50 mM Tris-HCl, pH 8.0, 2 mM reduced glutathione, 2 mM 3-phosphoglycerate, 0.2 mM ATP, 7 mM $MgCl_2$, 0.1 mM EDTA.

Exchange enzymes were prepared by centrifugation of 20 μl of dialyzed GAPDH and 10 μl of dialyzed 3-PGK, removal of the supernatant, careful washing of tube walls with water to remove ammonium ions (which inhibit T_4 polynucleotide kinase) and redissolving in 30 μl of water.

The exchange reaction was catalyzed by the addition of 1 μl of the exchange enzyme mix to the reaction mixture, and incubation at room temperature for 40 minutes, after which the reaction was stopped by the addition of 0.1 volume of 0.1 M EDTA, pH 8.0 and six times the volume of water followed by heating at 90°C for 2 minutes to inactivate the enzyme activity.

Fourteen times the reaction volume of ethanol was added and the mixture stored at -20°C. To test the extent of exchange, 1 μl of the final mixture was spotted on to a PEI cellulose strip (1.0 x 10 cm) and chromatographed in 0.75 M sodium phosphate, pH 3.5. The dried cellulose strip was autoradiographed for 30 seconds. The spot running with the solvent front was ^{32}P-orthophosphate, the spot at R_f ~0.6 was ^{32}P-ATP. 50-70% exchange of ^{32}P-orthophosphate, determined by cutting PEI-cellulose strip and counting radioactive spots in a scintillation spectrometer (Packard) was routinely achieved.

5' terminal phosphate groups were removed from

DNA (1-10 μg) in a 100 μl reaction mix containing 25 mM Tris-HCl, pH 8.0, 7 mM $MgCl_2$ and 0.2 units of bacterial alkaline phosphatase (BAPF) (previously dialyzed against 25 mM Tris-HCl, pH 8.0) at 65°C for 30 minutes. EDTA, to 5 mM, was added and the reaction mix phenol extracted and ethanol precipitated as previously described.

20 μl of γ-^{32}P-ATP (1000-1500 Ci/mmole, ∿200 μCi) was evaporated to dryness and dissolved in 60 mM Tris-HCl, pH 7.5, 9 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM spermidine. Dephosphorylated DNA was then dissolved in this mixture, 0.5 μl of T_4 polynucleotide kinase (5 units/μl) added and the reaction mixture incubated at 37°C for 45 minutes, phenol extracted and ethanol precipitated (II,2,F).

To isolate DNA labelled at only one end, either secondary restriction cleavage was performed and the products electrophoresed on a 5 or 6% polyacrylamide gel (II,2,H,i) or the DNA strands separated by heating, at 90°C for 2 min, in 40 μl of 30% (v/v) dimethylsulphoxide, 1 mM EDTA, pH 8.0, 0.05% xylene cyanol FF, 0.05% bromophenol blue and electrophoresis on a 5% polyacrylamide gel with a 50:1 acrylamide to bisacrylamide ratio (II,2,H,i) run at 100 volts in TBE.

Radioactively labelled DNA was cut out (II,2,I) and electroeluted (II,2,J) as previously described.

(b) Labelling 3' ends with α-^{32}P-dNTP and *E. coli* DNA polymerase I, Klenow fragment

DNA was radioactively labelled at 3' ends using the 3' → 5' exonuclease and DNA dependent DNA poly-

merase activities of the Klenow fragment of DNA polymerase I with α-^{32}P-deoxynucleoside triphosphates. Using selected labelled and non-labelled triphosphates, the DNA fragment, if generated by the cleavage of two restriction endonucleases with different specificities, could be selectively labelled at one end.

DNA (1 μg) was dissolved in a 20 μl reaction mixture containing 60 mM Tris-HCl, pH 7.5, 8 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM cold dNTP, 1 μM α-^{32}P-dNTP (250 Ci/mmole) and 1.5 units of *E. coli* DNA polymerase I, Klenow fragment, incubated at 30°C for 30 minutes, 5 μl of polyacrylamide gel loading buffer (II,2,H,i with 0.5% Sarkosyl) added and electrophoresed on a 6% polyacrylamide gel in TBE buffer (II,2,H,i). After autoradiography, ^{32}P-labelled DNA was electro-eluted from the gel (II,2,I and J).

ii) Sequencing reactions

Four reactions were used in the analysis of end-labelled DNA to determine its sequence, specific for purine (P), guanine (G), pyrimidine (Y) or cytosine (C) bases. Adenine and thymine were identified by their presence in the purine and pyrimidine reactions, respectively, and absence in the guanine and cytosine specific reactions. During the course of these studies the reaction conditions used varied, only the initial and final methods will be detailed.

(a) Base modification (initial)

End labelled DNA was dissolved in 30 μl of water and divided into four aliquots G^{o} (5 μl), P^{o} (10 μl), Y^{o} (10 μl), C^{o} (5 μl). 300 μl of cacodylate buffer (50 mM sodium cacodylate, pH 8.0, 10 mM $MgCl_2$,

0.1 mM EDTA) was added to G^{o} and P^{o}, 15 μl of water to Y^{o} and 20 μl 5 M NaCl to C^{o}, followed by 1 μl carrier DNA (4 mg/ml *Eco*RI digested chicken DNA; II,2,O) to each reaction mix. Dimethyl sulphate (2 μl) was added to G^{o} and P^{o}, the reaction mixtures incubated at 20°C for 8 and 20 minutes respectively, 50 μl of DMS stop mix (3 M sodium acetate, pH 6.0, 2.5 M β-mercaptoethanol, 100 mM magnesium acetate, 1 mM EDTA, 0.1 mg/ml *E. coli* tRNA) added and nucleic acids precipitated by addition of 1 ml of ethanol and incubation at -70°C for 15 minutes.

Y^{o} and C^{o} reactions were initiated by the addition of 30 μl of hydrazine and incubated at 21°C for 20 and 30 minutes respectively, after which 300 μl of hydrazine stop mix (0.3 M sodium acetate, pH 6.0, 10 mM magnesium acetate, 0.1 mM EDTA, 25 μg/ml *E. coli* tRNA) was added and ethanol precipitated as above.

After centrifugation, DNA was redissolved in 300 μl 0.3 M sodium acetate, pH 6.0, and ethanol precipitated again to remove dimethyl sulphate and hydrazine. Samples were again centrifuged and washed with 1 ml of ethanol to remove salt, followed by evaporation to dryness under vacuum.

(b) Base removal and strand scission (initial)

The P^{o} sample was redissolved in 20 μl of water and 5 μl of 0.5 N HCl added followed by incubation on ice for 2 hours (vortex mixing every 15 minutes) and ethanol precipitation by addition of 0.3 ml of sodium acetate, pH 6.0, and 1 ml of ethanol, -70°C, 1 hour. After centrifugation and washing with 1 ml of ethanol,

the DNA was redissolved in 10 μl of freshly prepared 0.1 M NaOH, 1 mM EDTA, sealed in a glass capillary tube and incubated at 90°C for 30 minutes. The reaction mix was then added to 10 μl of urea dye mix (60% (w/v) urea, 0.1% (w/v) bromophenol blue, 0.15% (w/v) xylene cyanol FF).

The G^{o}, Y^{o} and C^{o} samples were dissolved in 20 μl of freshly prepared 1 M piperidine, sealed in silicon coated glass capillary tubes and incubated at 90°C for 60 minutes. The contents were washed from the capillary tubes with an additional 25 μl of water and evaporated to dryness under vacuum. A further 25 μl of water was added and again evaporated to dryness, to remove residual piperidine. DNA was redissolved in 10 μl of freshly prepared 0.1 M NaOH, 1 mM EDTA and 10 μl of urea dye mix added.

(c) Base modification (revised)

End labelled DNA was dissolved in 30 μl of water and divided into four aliquots G^{1} (5 μl), P^{1} (10 μl), Y^{1} (10 μl), C^{1} (5 μl). 200 μl of cacodylate buffer and 1 μl of dimethyl sulphate were added to G^{1} and the reaction mixture incubated at 21°C for 6 to 8 minutes depending on length of sequence analyzed. The reaction was stopped by the addition of 50 μl of G^{1} stop mix (3 M sodium acetate, pH 6.0, 2.5 M β-mercaptoethanol 1 mM EDTA, 0.1 mg/ml *E. coli* tRNA), 1 ml of ethanol added and the DNA precipitated at -70°C for 1 hour.

25 μl of formic acid was added to the P^{1} sample and incubated at 21°C for 10 minutes, followed by the addition of P^{1} stop mix (0.3 M sodium acetate, pH 6.0,

0.1 mM EDTA, 25 μl/ml *E. coli* tRNA) and 1 ml of ethanol at -70°C for 1 hour.

15 μl water and 20 μl 5 M NaCl were added to Y^1 and C^1 respectively. 30 μl hydrazine was added to both, with incubation at 21°C for 20 minutes (Y^1) or 30 minutes (C^1), followed by the addition of 300 μl of either Y^1 stop mix (0.3 M NaCl, 0.1 mM EDTA, pH 8.0, 25 μg/ml *E. coli* tRNA) or C^1 stop mix (0.1 mM EDTA, pH 8.0, 25 μg/ml *E. coli* tRNA) and 1 ml of ethanol at -70°C for 1 hour.

After centrifugation all samples were reprecipitated by the addition of 300 μl 0.3 M sodium acetate, pH 6.0 and 1 ml of ethanol at -70°C for 1 hour, then centrifuged, washed with 1 ml of ethanol and evaporated to dryness under vacuum.

(d) Base removal and strand scission (revised)

All samples were redissolved in 25 μl of freshly prepared 1 M piperidine, heated at 90°C for 15 minutes (in Eppendorf microfuge tubes) then evaporated to dryness under vacuum. After the addition of 25 μl of water samples again evaporated to dryness then dissolved in formamide loading buffer (90% (v/v) deionized formamide, 0.01% (w/v) bromophenol blue, 0.01% (w/v) xylene cyanol FF, 0.1 mM EDTA, pH 8.0), either 4 μl or multiples thereof.

iii) Sequencing gels

Products of the chemical degradation sequencing reactions were separated by electrophoresis on polyacrylamide gels which include 8.3 M urea as a denaturant. Initially 1.5 mm thick gels were used, but these were later replaced

by 0.5 mm thick gels giving increased resolution.

For 20% gels a 100 ml mixture containing 20 ml of 10 x TBE buffer (II,2,H,i), 19 g of acrylamide, 1 g of N,N'-methylenebisacrylamide, 42 g of urea, 1 ml of 10% (w/v) ammonium persulphate was filtered and degassed, 20 µl TEMED added and poured into a 30 x 40 cm gel mould, and allowed to polymerize for at least four hours.

Reservoir tanks contained approximately 2 litres of 2 x TBE buffer (no urea) and the gels were pre-electrophoresed for 1 to 4 hours.

Samples were heated to 90°C for 2 minutes then chilled on ice before loading. Debris and urea were removed from sample wells prior to loading by flushing with electrophoresis buffer (2 x TBE) from a syringe.

10% and 8% gels, with correspondingly reduced amounts of acrylamide and BIS, were constructed in the same manner. All gels were run at 1000-1200 volts. Gels were autoradiographed as described (II,2,I) in an Ilford autoradiography cassette.

Two or more autoradiographs were taken of each gel due to differential exposure from each of the four sets of reaction products. Because of this, the number of gels autoradiographed and the difficulty in reproducing autoradiographs photographically, only autoradiographs of gels which illustrate a point were included in this thesis. All derived sequences were verified by Dr. J.R.E. Wells.

O. Restriction endonuclease analysis of genomal DNA

i) Synthesis of probe (Taylor *et al.*, 1976)

Radioactively labelled DNA probe was synthesised in a 100 µl reaction mixture containing 50 mM Tris-HCl,

pH 8.3, 10 mM $MgCl_2$, 20 mM β-mercaptoethanol, 5 mg/ml calf thymus DNA (previously digested with microccocal nuclease to 8 to 15 base lengths, preboiled for 2 minutes, then chilled on ice), 0.1 mM each of dTTP, dGTP and dATP, 1 μM dCTP (350 Ci/mmole) and 2 μg/ml cloned DNA (isolated from recombinant plasmid, preboiled for 2 minutes, then chilled on ice). 10 μl of AMV reverse transcriptase (30 units) was added, the reaction catalyzed at 42°C for 60 minutes, then phenol-chloroform extracted and chromatographed over Sephadex G-100 (II,2,F). The excluded volume peak of radioactivity, containing radioactively labelled DNA, was pooled and ethanol precipitated.

ii) Blot analysis (Southern, 1975)

Chicken genomal DNA isolated from the nuclei of red blood cells (II,2,D) was digested to completion with restriction endonuclease *Eco*RI (II,2,G). Totallity of digestion was verified by inclusion of λ DNA in a sample of the final digestion mixture.

After phenol-chloroform extraction and ethanol precipitation (II,2,F) the DNA was resuspended in 40 μl of water and 10 μl of agarose gel loading buffer added, loaded on to a 1% agarose gel (5 mm thick) and electrophoresed at 50 mA (II,2,H,ii).

The gel was then washed in 0.5 M NaOH for 1 hour, 1 M Tris-HCl, pH 7.0 for 1 hour and placed on to continuously wet, with 6 x SSC, Whatman 3 MM paper. The DNA was transferred on to 0.1 μm nitrocellulose filter by blotting overnight.

DNA was baked on to the nitrocellulose filter at 80°C under vacuum for 8 hours. The filter was then incubated at 42°C, overnight in a sealed plastic bag containing

15 ml of preannealing mix (50 mM HEPES, pH 7.0, 10 x Denhardt's solution, 3 x SSC, 54 μg/ml *Eco*RI digested, boiled, *Micrococcus luteus* DNA, 40 μg/ml yeast tRNA, 40 μg/ml poly-(ribo)A, 50% (v/v) formamide), after which the radioactively labelled probe DNA (8 x 10^6 c.p.m., boiled then chilled on ice) was added to the annealing mix, the bag resealed and incubation continued at 42°C for 48 hours.

The nitrocellulose filter was removed from the plastic bag and incubated at 21°C for 60 minutes in 4 x SSC, 1 x Denhardt's solution, 0.2% (w/v) sodium dodecyl sulphate. This was followed by a 60 minute wash at 68°C in 0.4 x SSC, 0.2% sodium dodecyl sulphate. The filter was rinsed in 2 x SSC to remove sodium dodecyl sulphate, air dried and autoradiographed with an intensifying screen at -70°C for two weeks.

P. Containment facilities

All work involving recombinant DNA in viable organisms described in this thesis was carried out under P_2 (NIH guidelines, U.S.A.) or C_3, C_1 (ASCORD guidelines, Australia) containment facilities in accordance with the respective guidelines. Initially *E. coli* χ1776 was used as the host strain, then when stringency of containment was relaxed (ASCORD guidelines revision) and approval granted, recombinant DNA was grown in the *E. coli* RRI strain (Rodriguez *et al.*, 1977).

At all times experiments were monitored and approved by local biohazards committees.

When the work described in this thesis, involving recombinant DNA in viable organisms, was initiated no member of the University of Adelaide was authorized to work in any containment facility. Dr. John Shine and Dr. Axel Ullrich,

both of whom were authorized to conduct such experiments, kindly agreed to perform the manipulations detailed in section II,2,L which required the use of a containment facility.

When this work was continued at Adelaide University under authorization of the Academy of Sciences Committee on Recombinant DNA, the experiments were performed in a C_1 or C_3 facility (depending on requirements of the experiment) at the Department of Biochemistry, University of Adelaide.

CHAPTER III

ISOLATION OF GLOBIN mRNA AND cDNA CHARACTERIZATION

CHAPTER III - ISOLATION OF GLOBIN mRNA AND cDNA CHARACTERIZATION

1. Introduction

The ability to isolate RNA transcripts has been crucial to the structural analysis of eukaryote genes, particularly in those cases where useful mutants are not available. Since the RNA carries much of the information encoded in a gene, sequence analysis of mRNA has enabled some insight into the possible control regions important for that gene's expression (Proudfoot and Brownlee, 1976). The homology of the RNA transcript with its gene and lack of homology with the rest of the genome enabled the use of radioactive probes derived from the RNA transcript as experimental tools in studying the structure of the gene.

This chapter briefly discusses aspects of the purification of chicken globin mRNA relevant to the thesis, and characterization of *in vitro* synthesised cDNA by restriction endonuclease digestion.

Messenger RNA is only a minor cellular component representing less than 5% of the total RNA of a cell (Lewin, 1975). Because of this, a variety of methods have evolved enabling the enrichment and purification of messenger RNA. The cells actively expressing a particular gene may constitute only a minor portion of the tissue and so enrichment of cell type is often the first step in mRNA preparation. This may be accomplished using physical properties of the cells, such as differential density (Ullrich *et al.*, 1977), by the establishment of a stable tumour cell line in which the gene is active (Villa-Komaroff *et al.*, 1978) or by hormonal (O'Malley and Means, 1974) or chemical (Pemberton

FIGURE III,1.

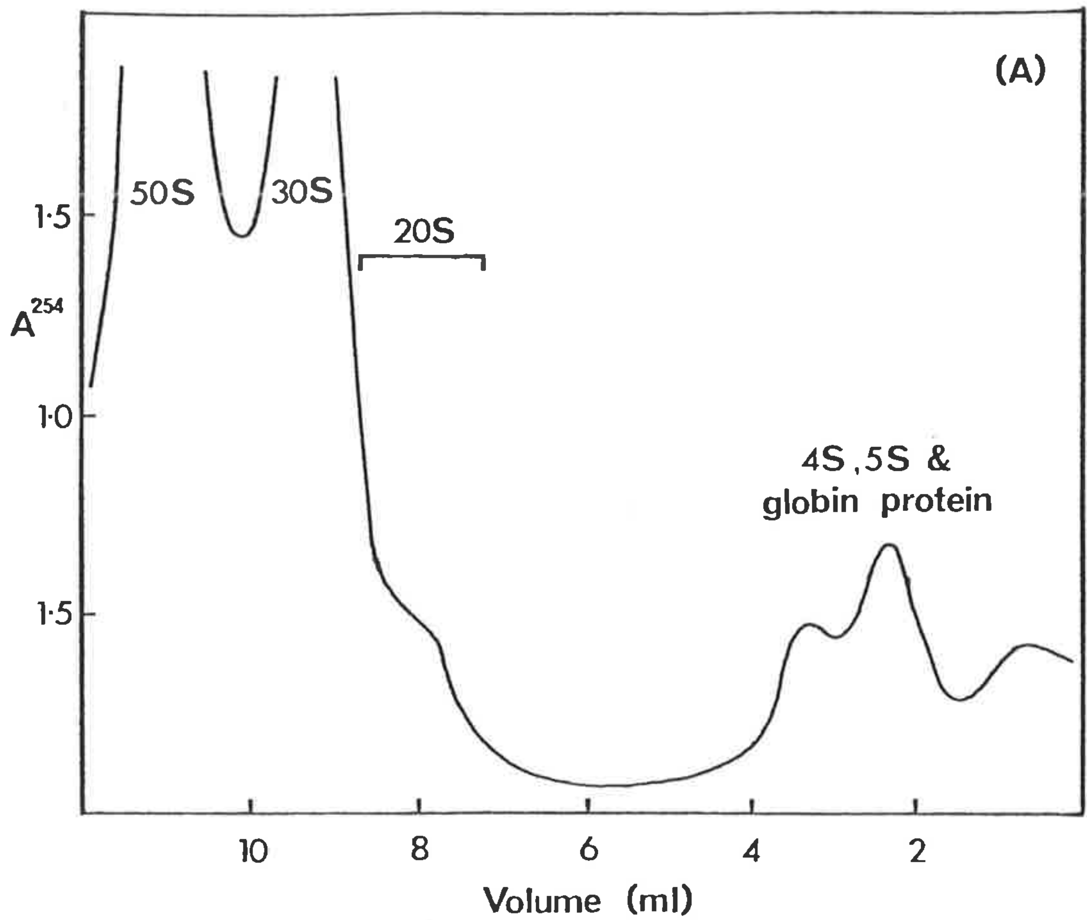
CHICKEN GLOBIN mRNA PURIFICATION

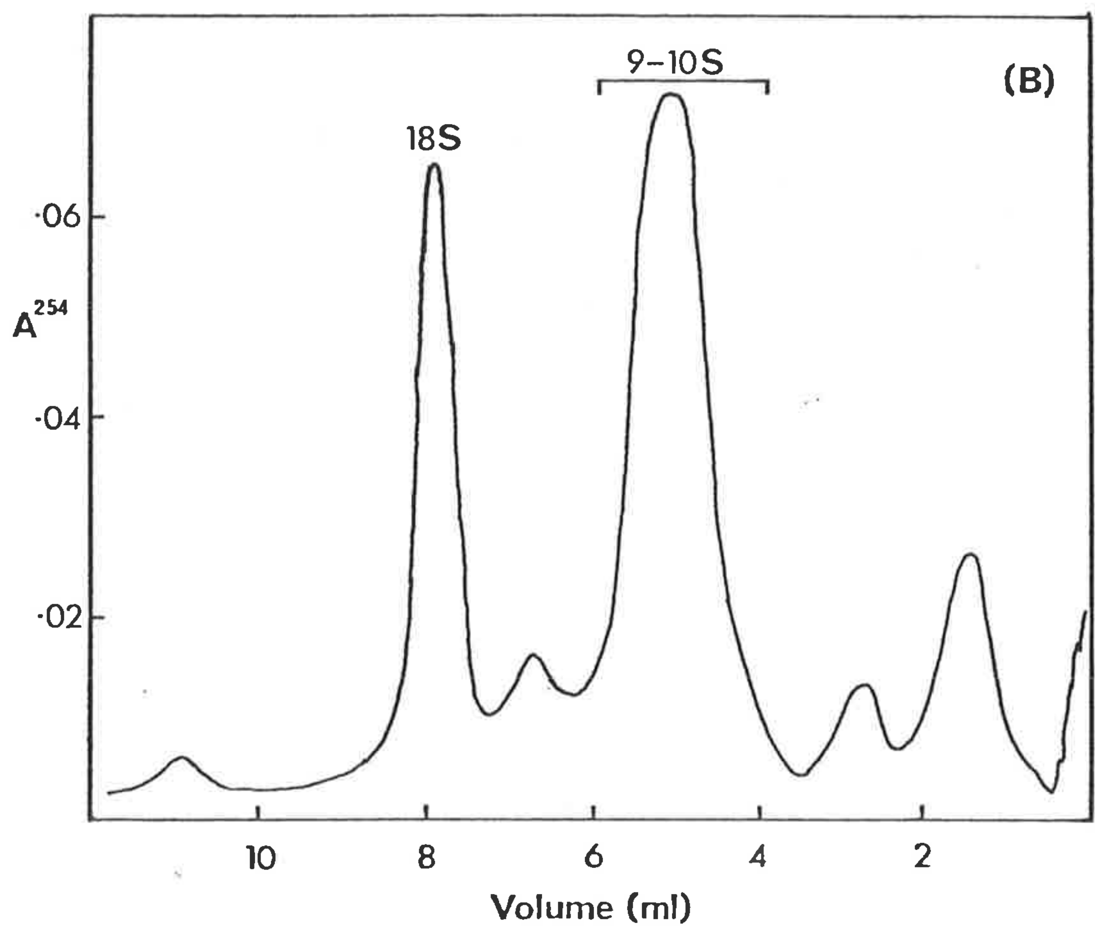
(a) Isolation of 20S RNP particles

Polysomes were isolated from chicken red blood cells as described in section II,2,A and dissociated with EDTA. Centrifugation was carried out at 4°C for 16 hours in a Beckman SW41 rotor, on 10% - 40% (w/v) sucrose gradients, at 210,000 x g. The 20S shoulder was collected and ethanol precipitated. Sedimentation was from right to left.

(b) Isolation of 9-10S RNA

20S RNPs were redissolved in 10 mM Tris-HCl, pH 9.0, 1 mM EDTA, 0.1% S.D.S., 10 μg/ml proteinase K, heated at 37°C for 3 minutes then centrifuged on 10% - 40% (w/v) sucrose gradients (as above). RNA sedimenting at 9-10S was collected as globin mRNA and ethanol precipitated. Sedimentation was from right to left.


(A)
50S
30S
20S
4S ,5S &
globin protein
1·5
1·0
1·5
A254
10
8
6
4
2
Volume (ml)


(B)
9–10S
18S
·06
·04
·02
A254
10
8
6
4
2
Volume (ml)

et al., 1972) induction of cell proliferation and/or gene expression.

The methods of mRNA extraction and purification vary, often as a consequence of levels of endogenous ribonuclease activity (Chirgwin *et al.*, 1979). Use is often made of the 3' poly(A) tract common to most eukaryote mRNA sequences by affinity chromatography on oligo(dT) cellulose (Aviv and Leder, 1972). Where a mixture of mRNA species occurs in the one cell type, these can often be separated on the basis of size by sucrose gradient centrifugation or polyacrylamide gel electrophoresis (Forget *et al.*, 1975). This approach was found to be ineffective in separating chicken globin mRNA into α and β chain coding species suitable for sequence analysis‡, and so alternative approaches were sought, including restriction endonuclease digestion analysis and recombinant DNA cloning. ‡(A. Robins, pers. comm.)

2. Results

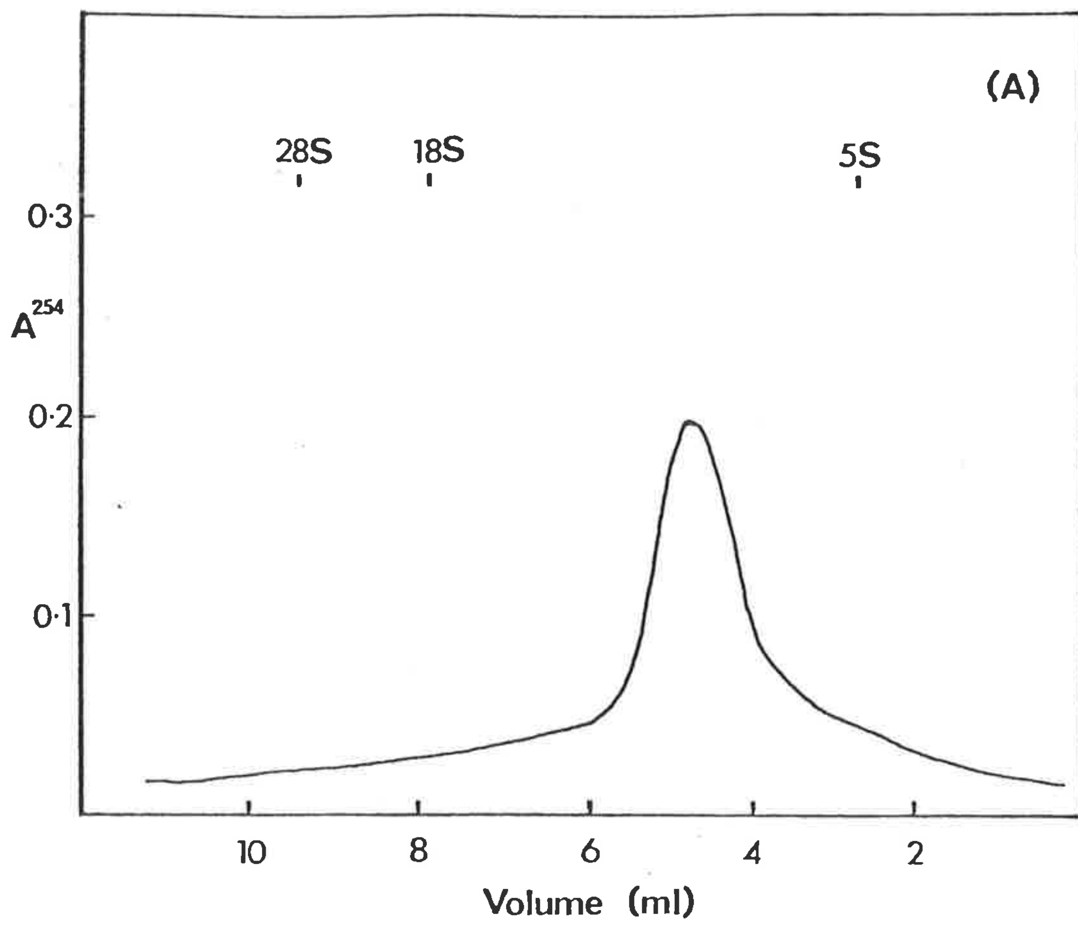
A. Chicken globin mRNA isolation

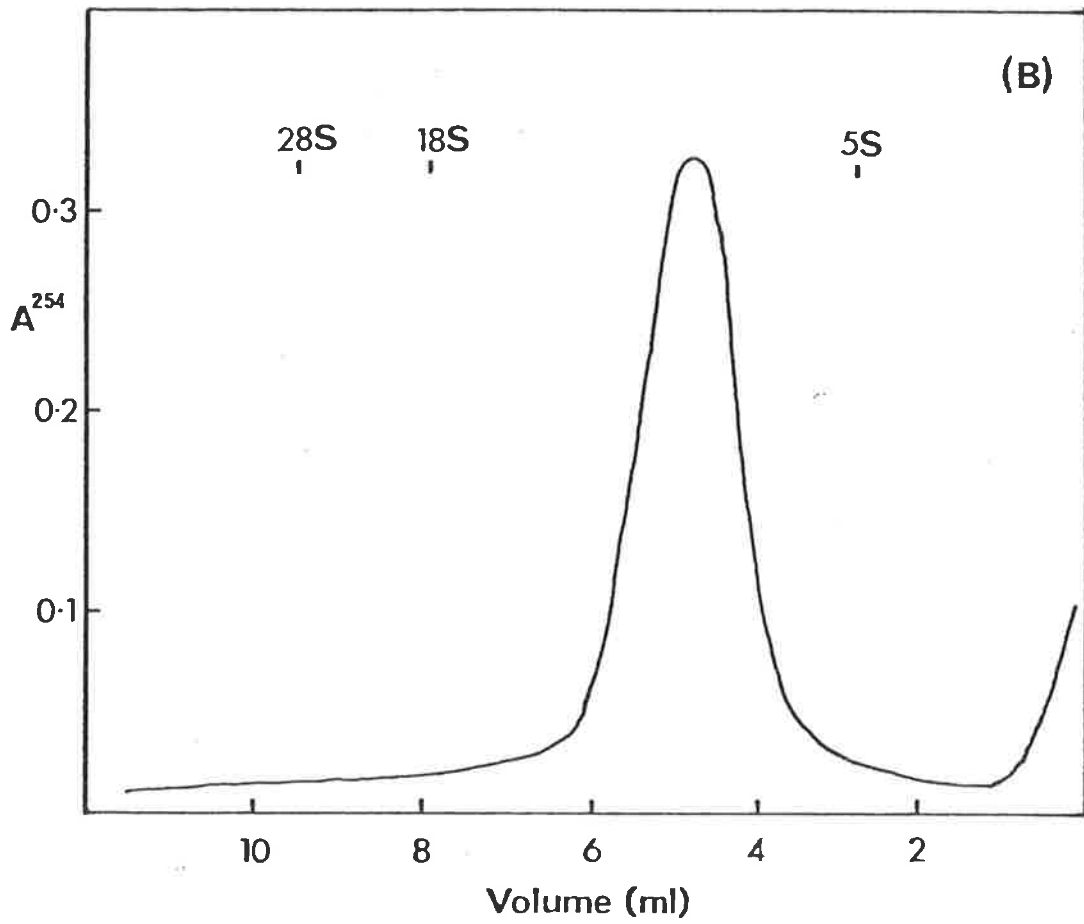
Over 90% of the protein synthesised by chicken red blood cells is hemoglobin (Williams, 1970). To increase the level of cells actively synthesising hemoglobin, hemolytic anaemia was induced in chickens by the injection of phenylhydrazine. The mRNA purification method described in section II,2,A is essentially that of Pemberton *et al.*, (1972) for duck globin mRNA. After dissociation with EDTA, chicken red blood cell polysomes were centrifuged on 10% - 40% (w/v) sucrose gradients and 20S ribonucleoprotein particles (RNPs) collected (Figure III,1,a). Protein was removed from RNPs by incubation with proteinase K in a solution containing sodium dodecyl sulphate and the 9-10S

FIGURE III,2.

SUCROSE GRADIENT CENTRIFUGATION OF *IN VITRO* POLYADENYLATED RNA

9-10S globin mRNA which did not bind to oligo(dT) cellulose (Aviv and Leder, 1972) and therefore designated poly(A)$^{-}$ RNA, was subjected to polyadenylation *in vitro* as described in section II,2,C. The products of this reaction were again chromatographed on oligo(dT) cellulose. The bound (A) and unbound (B) RNA was heated at 65°C for 2 min, then subjected to centrifugation on 10% - 40% (w/v) sucrose gradients at 210,000 xg for 16 hours at 4°C. 9-10S RNA was collected and ethanol precipitated. Ribosomal RNA markers were centrifuged on a separate gradient. Sedimentation was from right to left.


(A)
28S
18S
5S
0·3
A254
0·2
0·1
10
8
6
4
2
Volume (ml)


(B)
28S
18S
5S
0·3
A254
0·2
0·1
10
8
6
4
2
Volume (ml)

RNA collected from a further centrifugation on 10% - 40% (w/v) sucrose gradients (Figure III,1,b). Translation of mRNA, isolated in this manner, under optimal conditions in a wheat embryo cell free translation system, showed globin (59% α and 41% β chains) as the sole product (Scott, 1975).

B. Polyadenylation of poly(A)$^-$ RNA

In an effort to enrich for poly(A)$^-$ RNA (and presumably therefore H_5 mRNA; Scott and Wells, 1976), the 9-10S RNA was chromatographed on oligo(dT) cellulose (Aviv and Leder, 1972) as described in section II,2,B.

In a series of experiments 20-30% of the 9-10S RNA was found not to bind to oligo(dT) cellulose. Sucrose gradient centrifugation of the 'bound' poly(A)$^+$ and 'unbound' poly(A)$^-$ RNA showed that the RNA was still intact, sedimenting at 9-10S (data not shown). The unbound RNA, poly(A)$^-$, was incubated with terminal riboadenylate transferase and ATP as described in section II,2,C and the products again chromatographed on oligo(dT) cellulose. The 'bound' and 'unbound' fractions were again centrifuged on sucrose gradients and found to be intact (Figures III,2,a and b). Approximately 20% of the *in vitro* polyadenylated poly(A)$^-$ RNA was bound to the oligo(dT) cellulose column.

C. Synthesis of cDNA

cDNA was synthesised from the RNA template using oligo-(dT) priming on the natural or the *in vitro* added 3' poly(A) tract. The cDNA was judged to be a full-length copy by centrifugation in sucrose gradients at 9-10S (data not shown) and by electrophoresis on polyacrylamide gels (II,2,H,i) when compared to rabbit globin cDNA (synthesised from rabbit

globin mRNA purified using the same conditions as those for chicken globin mRNA) which has been shown to copy the mRNA template to full-length (Efstratiadis *et al.*, 1975), see Figure III,3. Also indicated by electrophoresis is a number of partial synthesis products which may result from difficulty of the reverse transcriptase enzyme to copy hairpin structures in the chicken mRNA sequences (Efstratiadis *et al.*, 1976).

D. Restriction endonuclease digestion

Since oligo(dT) primed cDNA is always slightly heterogeneous in length, due to variable initiation (and termination) of the reverse transcriptase reactions, only fragments arising from restriction endonuclease digestion at two or more sites in a predominant cDNA species will give rise to a discrete band. Therefore, bands indicate the length of internal digestion products of the major species in the cDNA.

i) Single-stranded cDNA

In an effort to characterize sequences present and determine the feasibility of sequencing cDNA restriction fragments (Seeburg *et al.*, 1977a) aliquots of chicken globin cDNA (of high specific activity, II,2,E,i) were digested with restriction enzymes capable of digesting single-stranded DNA (Horiuchi and Zinder, 1975). The major advantages of this over the use of double-stranded cDNA were that only one reverse transcriptase synthesis was needed, therefore higher yield obtained from mRNA template. Secondly, that single-stranded DNA has only one 5' terminal phosphate group and, therefore, DNA radioactively labelled at this site would not need further digestion or treatment to be suitable for sequence analysis (Maxam and Gilbert, 1977).

FIGURE III,3.

*Hae*III DIGESTION OF SINGLE-STRANDED cDNA

Aliquots of ^{32}P-labelled s.s. cDNA (5×10^4 to 10^5 c.p.m.) were subjected to *Hae*III restriction endonuclease digestion (II,2,G), 0.2 volumes of acrylamide gel loading buffer (with sarkosyl) added and the DNA electrophoresed on a 30 cm long 5% polyacrylamide gel (II,2,H,i). DNA was visualized by autoradiography.

Track A. *Hae*III digestion of rabbit globin cDNA for 4 hours (5 units of enzyme).

B. Undigested rabbit globin cDNA.

C. *Hae*III digestion of chicken globin cDNA for 4 hours (5 units of enzyme).

D. As for track C with an additional 5 units of *Hae*III for a further 4 hours.

E. cDNA from track D with another 5 units of *Hae*III added, and incubation for 24 hours.

F. Undigested chicken globin cDNA.

G. Another preparation of chicken globin cDNA under the same *Hae*III digestion conditions as track E.

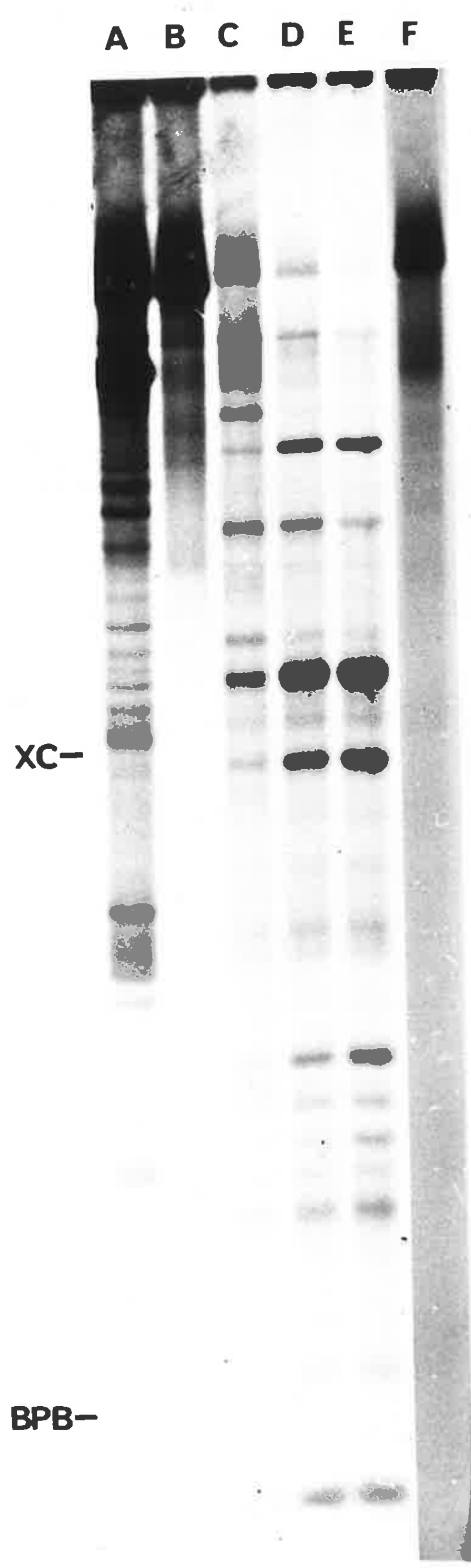
H. *Hae*III digestion (5 units, 24 hours) of cDNA from *in vitro* polyadenylated poly(A)$^-$ 9-10S mRNA (section III,2,B).

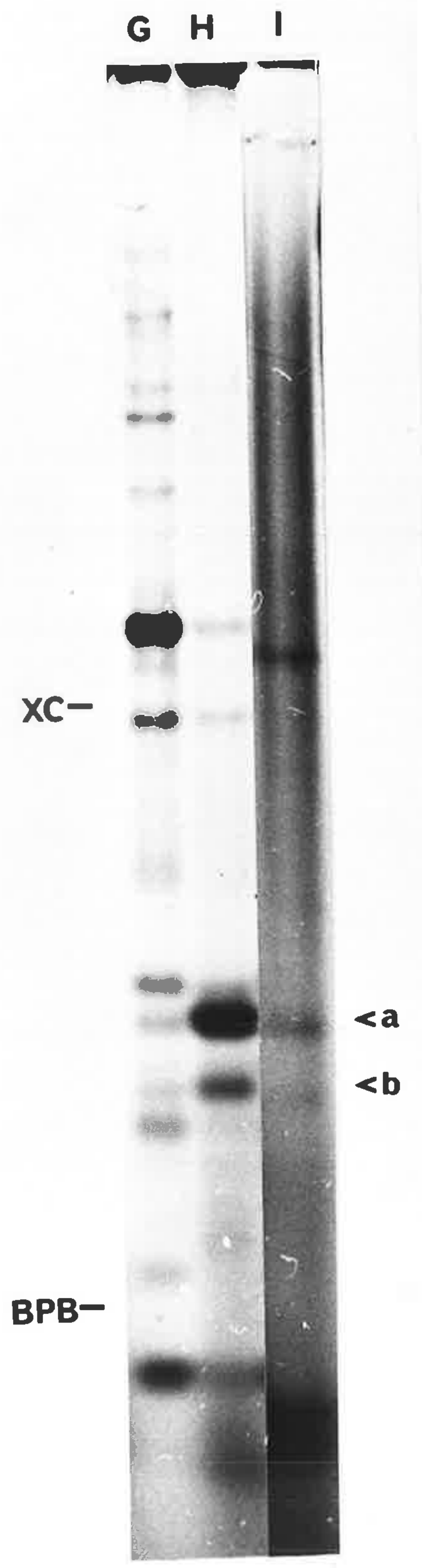
I. *Hae*III digestion (5 units, 24 hours) of cDNA from *in vitro* polyadenylated 28S chicken ribosomal RNA.

Tracker dyes: XC - xylene cyanol FF

BPB - bromophenol blue.

(a) and (b) indicate *Hae*III digest fragments common to tracks H and I.


A
B
C
D
E
F
XC
BPB


G
H
I
XC
BPB
<a
<b

Only one enzyme, restriction endonuclease *Hae*III, was found to cleave the single-stranded cDNA (others assayed were *Hpa*II and *Hha*I) and digestion with this enzyme was almost always partial, see Figure III,3. This difficulty may have been due to a lower specific activity of the enzyme for single-stranded DNA, as opposed to double-stranded, however repeated digestion did not markedly affect the banding pattern. The same was true for *Hae*III digested rabbit globin single-stranded cDNA, the size fragments of which were known (Seeburg *et al.*, 1977a), see Figure III,3 tract A. Additional bands found in *Hae*III digested cDNA from *in vitro* polyadenylated poly(A)$^-$ RNA (section III,2,B) were found to co-electrophorese with *Hae*III cDNA digestion products from *in vitro* polyadenylated 28S ribosomal RNA (see Figure III,3, fragments (a) and (b)) and therefore this approach to purification of histone H_5 cDNA was abandoned, as poly(A)$^-$ RNA from the globin mRNA preparation was therefore mainly ribosomal RNA breakdown products.

Because of the variability of digestion, this approach was not pursued as a means of sequencing the cDNA.

ii) Double-stranded cDNA

Doubled-stranded cDNA was synthesised by sequential reverse transcriptase reactions (Ullrich *et al.*, 1977) using poly(A)$^+$ 9-10S chicken globin mRNA (III,2,B) as template and the reaction conditions described in sections II,2,E,ii and II,2,F. After S_1 nuclease (II,2,K,i) digestion aliquots of this double-stranded cDNA, containing approximately 3 x 10^4 c.p.m. of ^{32}P radioactivity, were digested with restriction endonucleases (II,2,G) and electrophoresed on polyacrylamide gels (II,2,H,i). Autoradiographs of

FIGURE III,4.

RESTRICTION ENDONUCLEASE DIGESTION OF CHICKEN GLOBIN DOUBLE-STRANDED cDNA

Aliquots of S_1 nuclease digested ^{32}P-labelled (3×10^4 c.p.m.) cDNA (II,2,K,i) were incubated with each enzyme for 1 hour, electrophoresed on a 6% polyacrylamide gel (II,2,H,i) and autoradiographed. Molecular weight markers are Fd phage RF DNA *Hae*III fragments (Beck *et al.*, 1978) radioactively labelled with γ-^{32}P-ATP and T_4 polynucleotide kinase.

*Alu*I and *Sst*I enzymes were later found to contain contaminating non-specific exonuclease activity (F. de Noto, pers. comm.).

In addition, *Eco*RI, *Bam*HI and *Hin*dIII were found not to cleave the cDNA (data not shown).

Track A. *Bgl*I

B. *Pvu*II

C. *Xba*I

D. *Mbo*II

E. *Ava*II

F. *Hin*cII

G. *Hin*fI

H. *Sal*I

I. *Hae*III Fd RF DNA.

See Figure III,5 for additional restriction endonuclease digestion data.

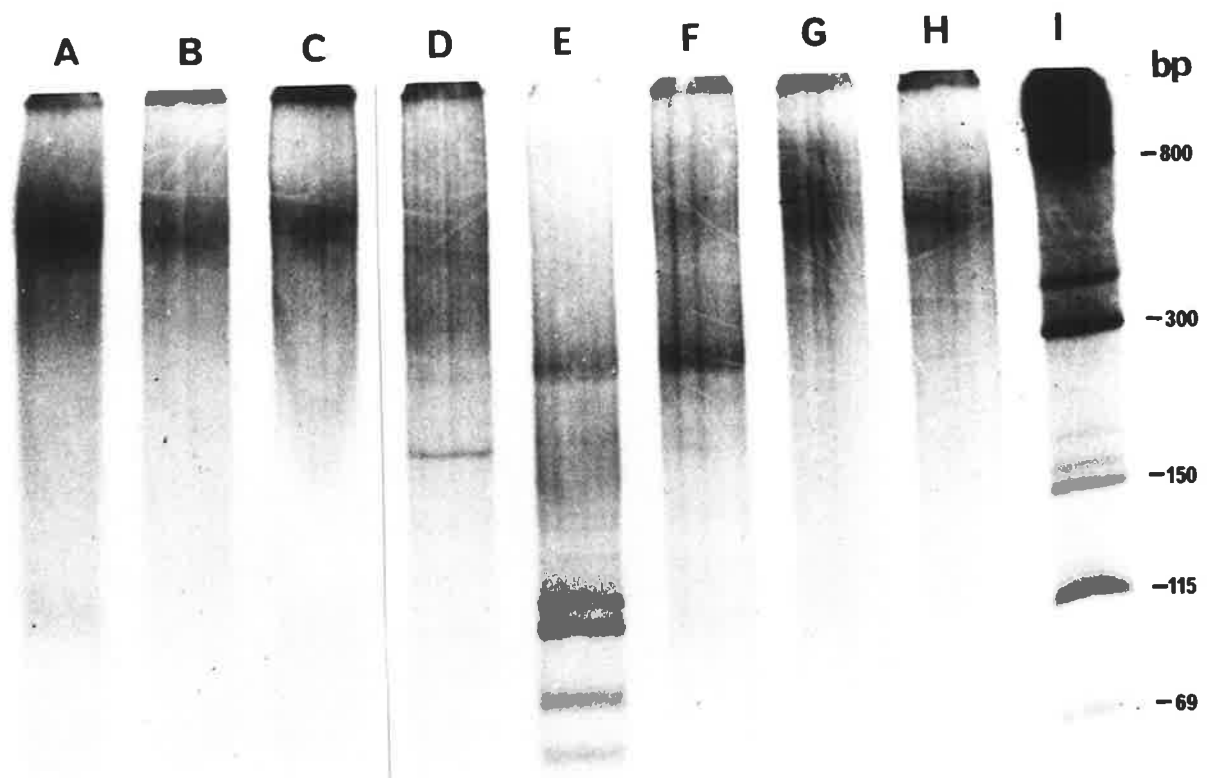

A
B
C
D
E
F
G
H
I
bp
800
300
150
115
69

FIGURE III,5.

RESTRICTION ENDONUCLEASE DIGESTION OF CHICKEN GLOBIN DOUBLE-STRANDED cDNA

Details are as described in the legend to Fig. III,4.

Track A. *Hae*III

B. *Hpa*II

C. *Hha*I

D. *Hpa*I

E. *Xma*I

F. *Alu*I

G. *Kpn*I

H. *Sst*I

I. *Pst*I

J. *Hae*III Fd RF DNA.

Restriction endonuclease digestion of cDNA data is summarized in Table III,1.

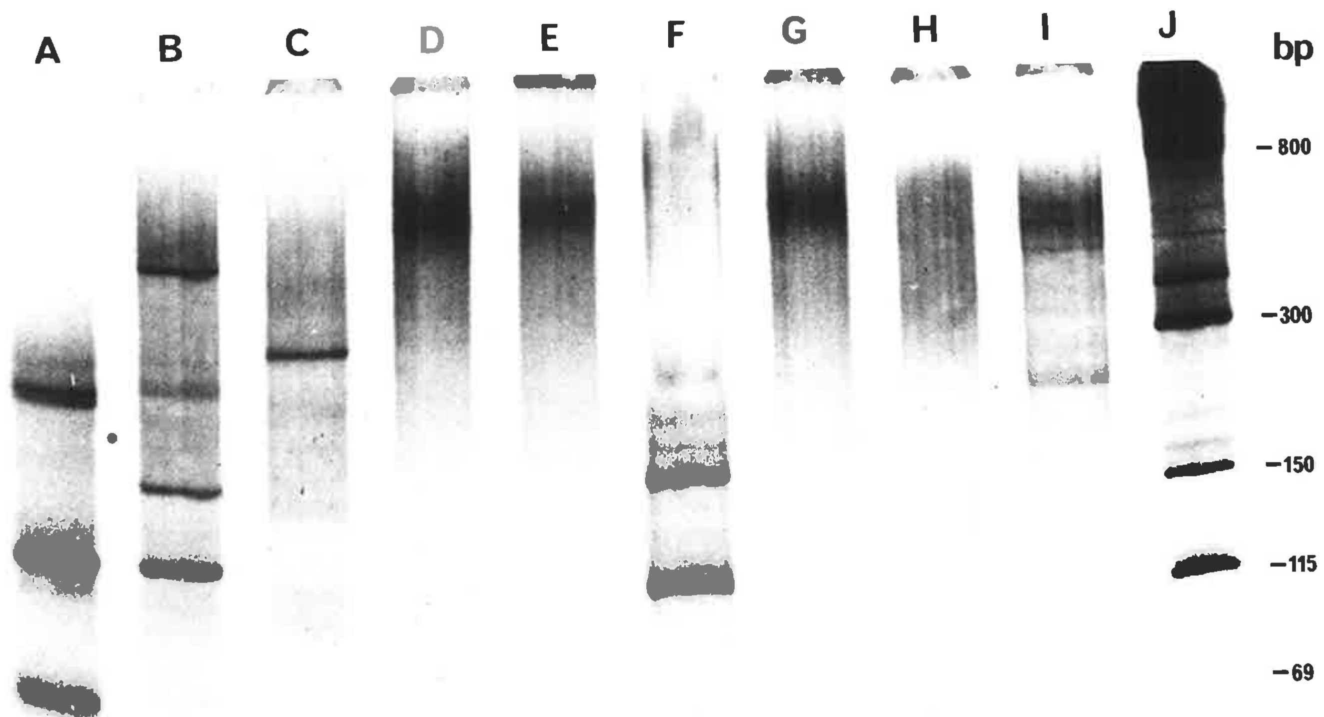

A
B
C
D
E
F
G
H
I
J
bp
800
300
150
115
69

TABLE III,1.

APPROXIMATE LENGTHS OF MAJOR d.s.cDNA RESTRICTION ENDONUCLEASE DIGEST FRAGMENTS

Restriction Endonuclease	Fragments generated (in base pairs)
*Hae*III	260, 250, 125, 70
*Hpa*II	450, 170, 125
*Hha*I	320
*Alu*I	170, 120
*Ava*II	115, 100, 80, 65
*Mbo*II	210 (minor band)

Data is from Figures III,4 and III,5 and is deduced by comparison with molecular weight markers (*Hae*III digested Fd phage RF DNA, Beck *et al.*, 1978).

these gels are shown in Figures III,4 and III,5.

*Hae*III digestion confirmed several of the major bands seen in single-stranded cDNA digests (Figure III,3). The approximate lengths of digestion products are summarized in Table III,1.

3. Discussion

Chicken globin mRNA was purified as poly(A) containing 9-10S RNA from the red blood cell polysomes of chickens with induced hemolytic anaemia. Being polysomal this mRNA was active in translation *in vivo* and therefore represents the hemoglobin phenotypes expressed by these cells.

Single and double-stranded cDNA synthesised *in vitro* using AMV reverse transcriptase were judged to be full-length by acrylamide gel electrophoresis against DNA of known length.

Restriction endonuclease digestion of single and double-stranded chicken globin cDNA enabled the characterization of the population of mRNA sequences, from which it was derived, by visualization using autoradiography of the predominant sequences present.

CHAPTER IV

MOLECULAR CLONING OF CHICKEN GLOBIN cDNA

CHAPTER IV - MOLECULAR CLONING OF CHICKEN GLOBIN cDNA

1. Introduction

The use of sequence specific restriction endonucleases in the analysis of DNA has introduced a new technology to this field of research, that of recombinant DNA.

The method is based on two main factors:

(1) The ability to join the DNA (genetic material) from two different organisms.

(2) The fact that prokaryote cells only replicate one species of vector molecule (either drug resistance conferring plasmid or bacteriophage) per cell.

The first point allows the joining (or ligation) of "foreign" DNA to a prokaryote vector DNA and the second allows the purification (by isolation of a single colony) and amplification of the hybrid vector (or recombinant) DNA.

Recombinant DNA technology therefore overcomes two of the major problems associated with sequencing of a eukaryote mRNA, by enabling the purification of a single sequence and in sufficient quantity to make the analysis possible.

Figure IV,1 outlines the scheme used in construction of the recombinant DNA containing sequences derived from chicken globin mRNA. The remainder of this chapter is devoted to the description of this construction and the identification of the cloned sequences.

2. Results

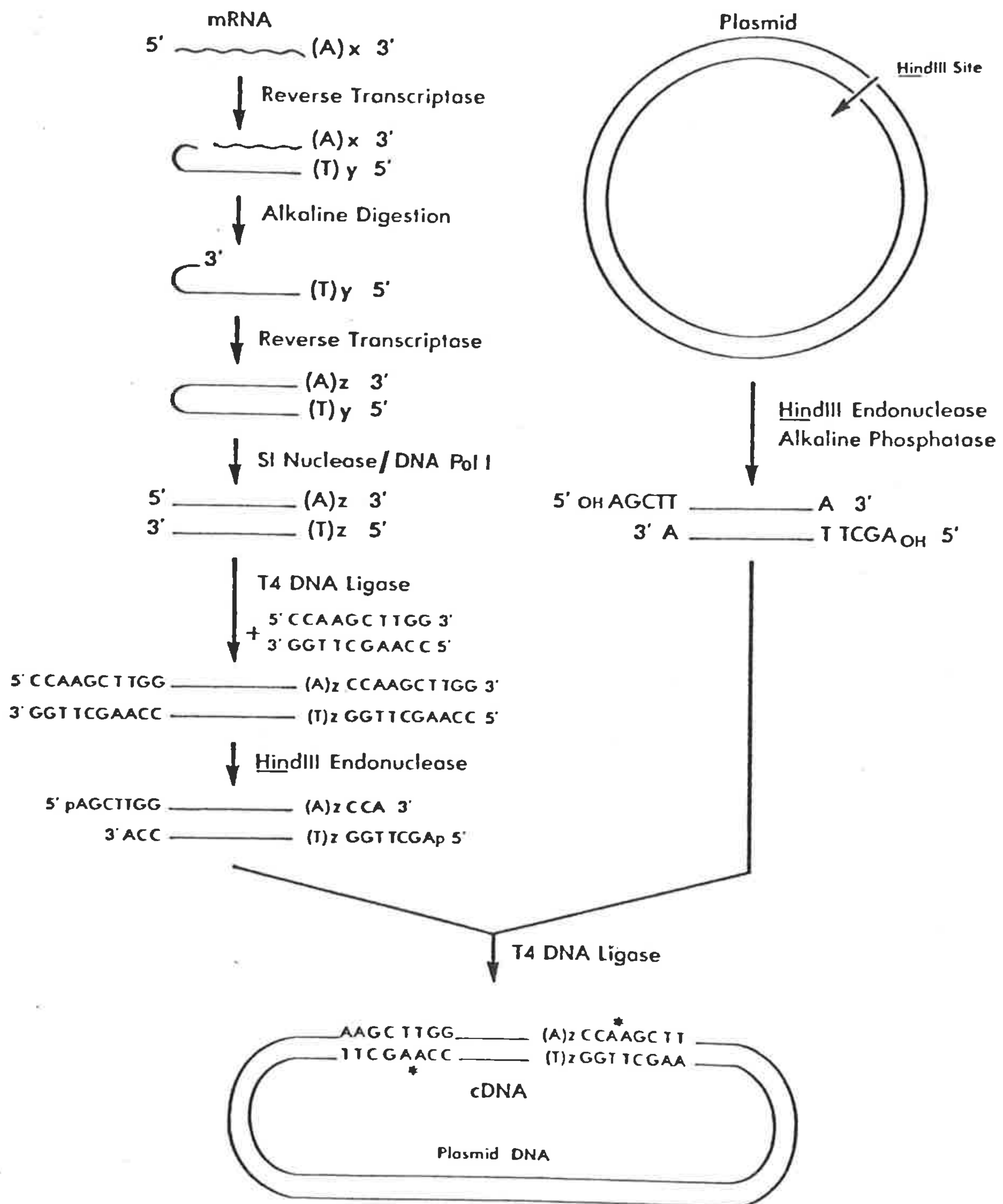
A. Synthesis of blunt ended double-stranded cDNA

Double-stranded cDNA was synthesised from 2.5 μg of single-stranded chicken globin cDNA (II,2,E,ii) using

FIGURE IV,1.

CONSTRUCTION OF RECOMBINANT DNA

Schematic diagram for insertion of cDNA into bacterial plasmids with the use of chemically synthesized restriction site linkers. The asterisks in the recombinant plasmid indicate the position where a phosphodiester bond was not formed because of the absence of a 5' terminal phosphate. (This figure is modified from Ullrich *et al.*, 1977.)



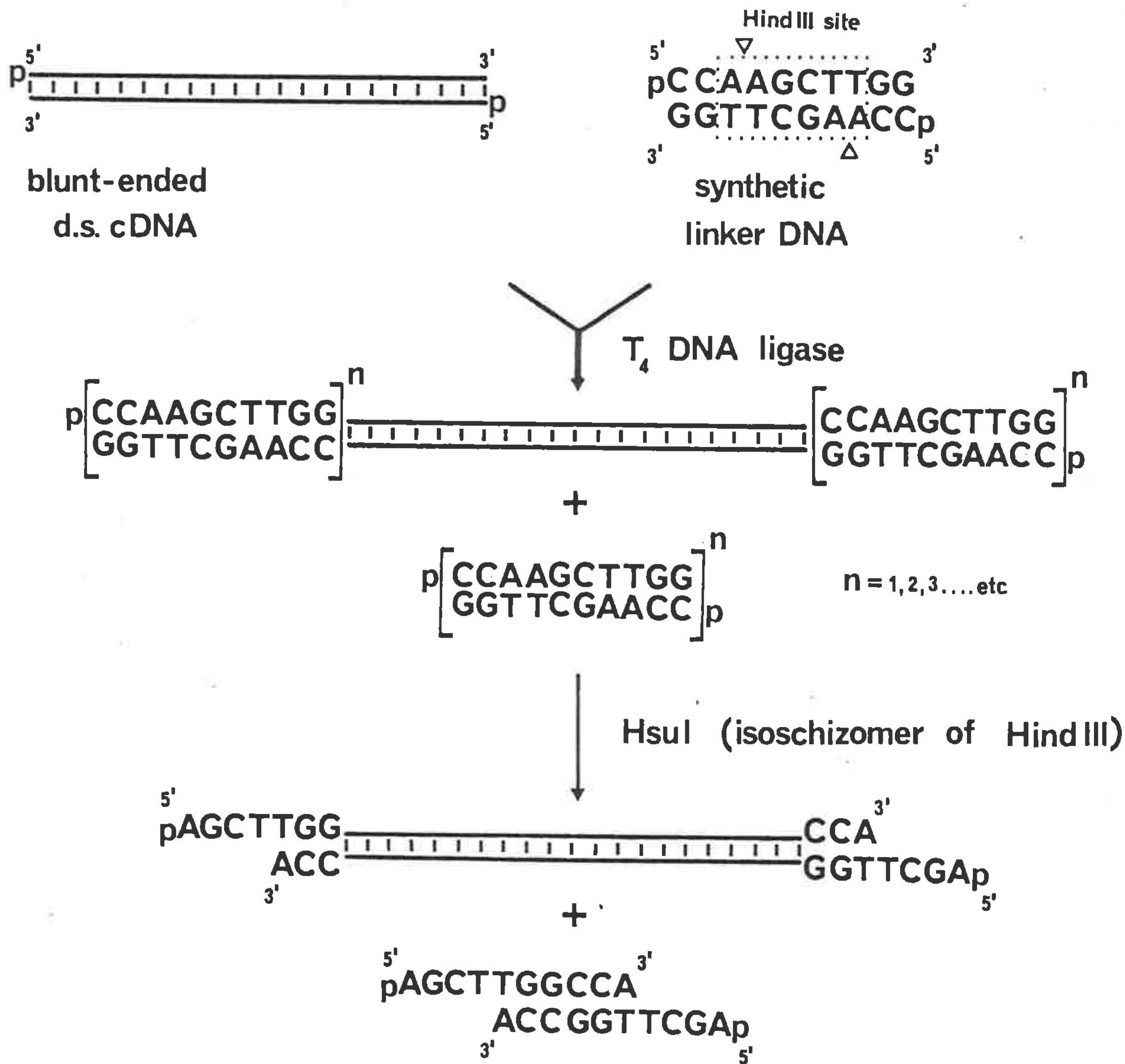
mRNA
5'
(A)x 3'
Reverse Transcriptase
(A)x 3'
(T) y 5'
Alkaline Digestion
3'
(T)y 5'
Reverse Transcriptase
(A)z 3'
(T) y 5'
SI Nuclease/DNA Pol I
5'
(A)z 3'
3'
(T)z 5'
T4 DNA Ligase
5' CCAAGCTTGG 3'
3' GGTTCGAACC 5'
5' CCAAGCTTGG
(A)z CCAAGCTTGG 3'
3' GGTTCGAACC
(T)z GGTTCGAACC 5'
HindIII Endonuclease
5' pAGCTTGG
(A)z CCA 3'
3' ACC
(T)z GGTTCGAp 5'
Plasmid
HindIII Site
HindIII Endonuclease
Alkaline Phosphatase
5' OH AGCTT
A 3'
3' A
T TCGA OH 5'
T4 DNA Ligase
AAGCTTGG
(A)z CCAAGCTT
TTCGAACC
(T)z GGTTCGAA
cDNA
Plasmid DNA

FIGURE IV,2.

LIGATION OF *Hin*dIII RECOGNITION SEQUENCE

TO DOUBLE-STRANDED cDNA

Schematic diagram of the products generated during *in vitro* addition of synthetic linker DNA to cDNA sequences.



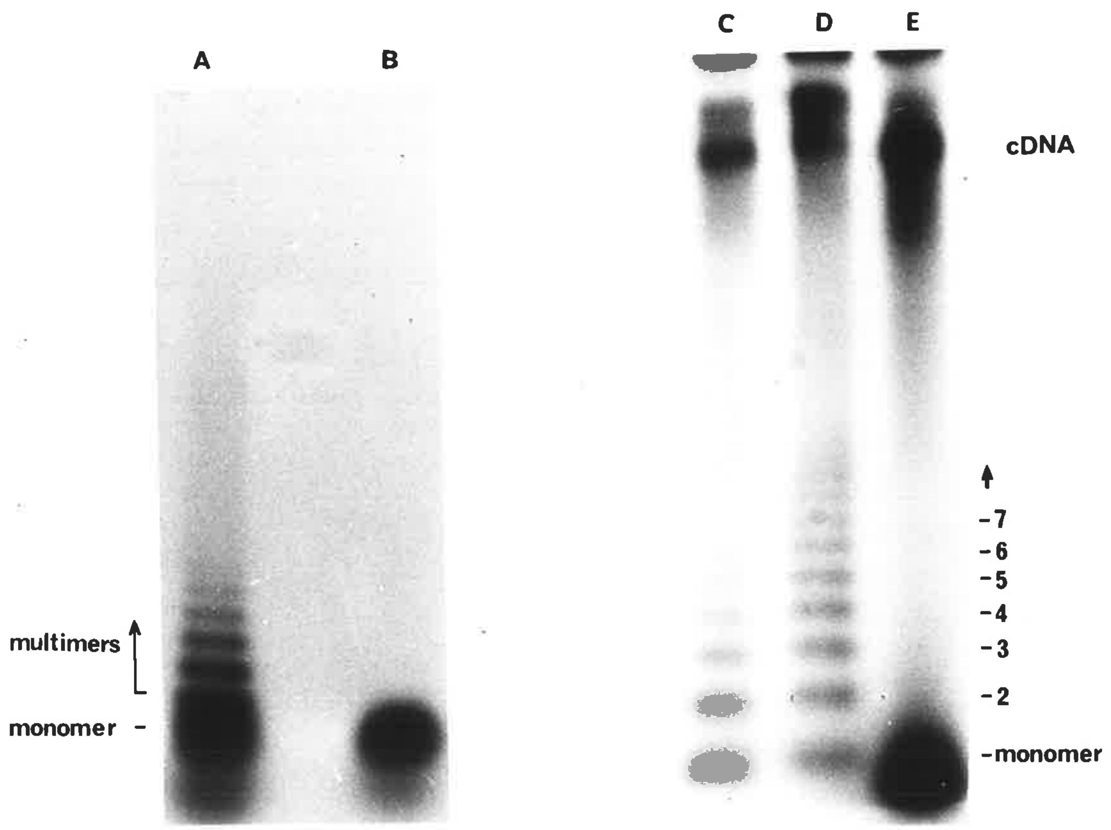
Hind III site
pCCAAGCTTGG
GGTTCGAACCp
blunt-ended
d.s. cDNA
synthetic
linker DNA
T4 DNA ligase
pCCAAGCTTGG
GGTTCGAACC
n
CCAAGCTTGG
GGTTCGAACCp
+
pCCAAGCTTGG
GGTTCGAACCp
n = 1, 2, 3....etc
HsuI (isoschizomer of Hind III)
pAGCTTGG
ACC
CCA
GGTTCGAp
+
pAGCTTGGCCA
ACCGGTTCGAp

FIGURE IV,3.

LIGATION OF SYNTHETIC LINKER DNA TO CHICKEN GLOBIN d.s.cDNA

Alquots of ligation and *Hsu*I restriction endonuclease reactions (II,2,K,ii) were electrophoresed on 6% polyacrylamide gels (II,2,H,i) and autoradiographed for 16 hours.

Tracks A and B are aliquots from a 1 hour control ligation reaction which contained no cDNA, before (A) and after (B) *Hsu*I digestion. Tracks C and D are aliquots of the ligation reaction containing cDNA after 4 hours (C) and 24 hours (D) incubation. Track E is an aliquot of the DNA from track D after overnight digestion with *Hsu*I. (See Fig. IV,2 for details of reaction products.) The ladder of bands is due to multimers of the synthetic linker DNA.


A
B
multimers
monomer
C
D
E
cDNA
-7
-6
-5
-4
-3
-2
-monomer

reverse transcriptase by the method of Ullrich *et al.* (1977) as described in section II,2,F. By incorporation of radioactivity 10% copy was achieved which resulted in 0.5 μg of double-stranded cDNA. To make this cDNA suitable for ligation to a plasmid vector, the single-stranded and hairpin loop DNA was digested with S_1 nuclease (Maniatis *et al.*, 1976). As only the second strand of this cDNA was radioactively labelled, the 95% recovery of radioactivity indicated 0.475 μg of double-stranded cDNA. Figures III,4 and III,5 show that this cDNA was mainly full-length copy.

S_1 nuclease does not always give flush ended DNA and as this was required for efficient ligation of linker DNA, the double-stranded cDNA was subjected to *E. coli* DNA polymerase I treatment (Seeburg *et al.*, 1977b) as described in section II,2,K,i to repair overlapping 5' ends and to hydrolyse overlapping 3' ends.

B. Ligation of *Hin*dIII recognition site

In order that the cDNA could be inserted into (and removed from) the *Hin*dIII restriction endonuclease site of the plasmid vector, pBR322, synthetic linker DNA encoding the *Hin*dIII recognition sequence was ligated to its termini (Fig. IV,2). *Hin*dIII was previously found not to cleave the double-stranded cDNA (data not shown) and therefore full-length cDNA should be inserted.

The ligation reaction (II,2,K,ii) was followed by the laddering of linker DNA into multimers (Figure IV,3). An increase in the molecular weight of the cDNA is also evident in this autoradiograph, probably due to formation of cDNA dimers as well as ligation of linker DNA to the cDNA.

*Hsu*I (an isoschizomer of *Hin*dIII) was used to cleave

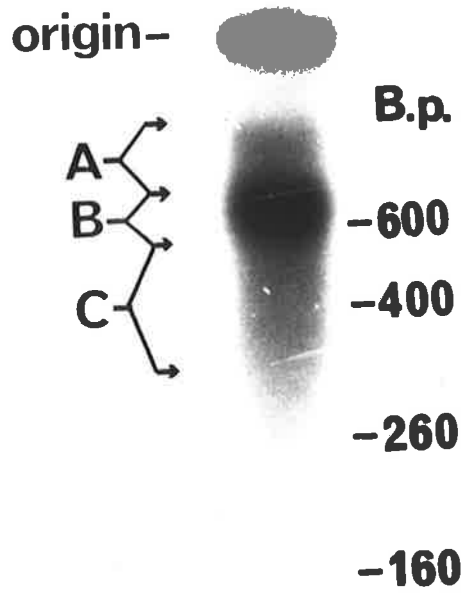
FIGURE IV,4.

PREPARATIVE ELECTROPHORESIS OF 'STICKY ENDED' d.s.cDNA

'Sticky ended' d.s.cDNA was electrophoresed on a 6% polyacrylamide gel (II,2,H,i) to separate monomeric linker DNA, and fractionate the cDNA on the basis of size. Molecular weight markers (shown in nucleotide base pairs, Bp) of *Hae*III cut pBR322 DNA (Sutcliffe, 1978) were run in a parallel track and visualized by ultraviolet light after ethidium bromide staining.

The cDNA was electroeluted from slices of the gel, after autoradiography (II,2,I and J) and recovery measured by incorporation of ^{32}P-radioactivity.

Slice	A.	110 ng
	B.	121 ng
	C.	97 ng.



linker >

the synthetic linker DNA to monomer form, leaving *Hin*dIII 'sticky ends', suitable for ligation into the *Hin*dIII site of pBR322, on the ends of the cDNA (see Figure IV, 1 and 2). This enzyme was used instead of *Hin*dIII, as preparations of *Hin*dIII often contain contaminating *Hin*dII activity (for which a cleavage site was found to exist in the cDNA, *Hinc*II and *Hin*dII being isoschizomers, see Figure III,4) and *Hemophillus suis*, from which *Hsu*I was purified, does not contain an isoschizomer of *Hin*dII (E. Tischer, pers. comm.).

'Sticky ended' cDNA was separated from monomeric linker DNA by electrophoresis on polyacrylamide gels (II,2,H,i) as seen in Figure IV,4 followed by electroelution (II,2,J) of gel slices. Recovery of cDNA was judged by ^{32}P radioactivity.

C. Ligation of cDNA to pBR322 plasmid DNA

The plasmid vector DNA was treated with bacterial alkaline phosphatase, after linearization with *Hsu*I, to remove 5' terminal phosphate groups. As at least one 5' terminal phosphate group is necessary for the ligation activity of T_4 DNA ligase, this treatment effectively reduces the number of parental recombinants, in favour of those which have foreign DNA sequences (which have 5' terminal phosphates) inserted into the *Hin*dIII site (Ullrich *et al.*, 1977).

25 ng of fractionated 'sticky ended' ds cDNA was ligated to 50 ng of dephosphorylated *Hsu*I cut pBR322 DNA (Bolivar *et al.*, 1977). The high molar ratio of cDNA to plasmid DNA also served to decrease the number of parental recombinants, and in some cases more than one cDNA sequence was inserted into a single plasmid *Hin*dIII site.

FIGURE IV,5.

AGAROSE GEL ELECTROPHORESIS OF 'MINISCREEN' PLASMID DNA

Plasmid DNA prepared by 'miniscreening' Ap^rTc^s colonies (II,2,L,ii) was electrophoresed on a 1% agarose gel (II,2,H, ii). The DNA was visualized by ultraviolet light after ethidium bromide staining.

Tracks D, I and M are pBR322 DNA, all other tracks are from different recombinant clones. All appeared to contain higher molecular weight DNA than pBR322, and therefore contained additional, presumably inserted, sequences.

Ch - *E. coli* chromosomal DNA

CS - closed circular form of plasmid DNA

SC - super-coiled form of plasmid DNA.

Note: Track K was later found to have two inserted sequences. Only 16 of the 30 colonies screened are shown.

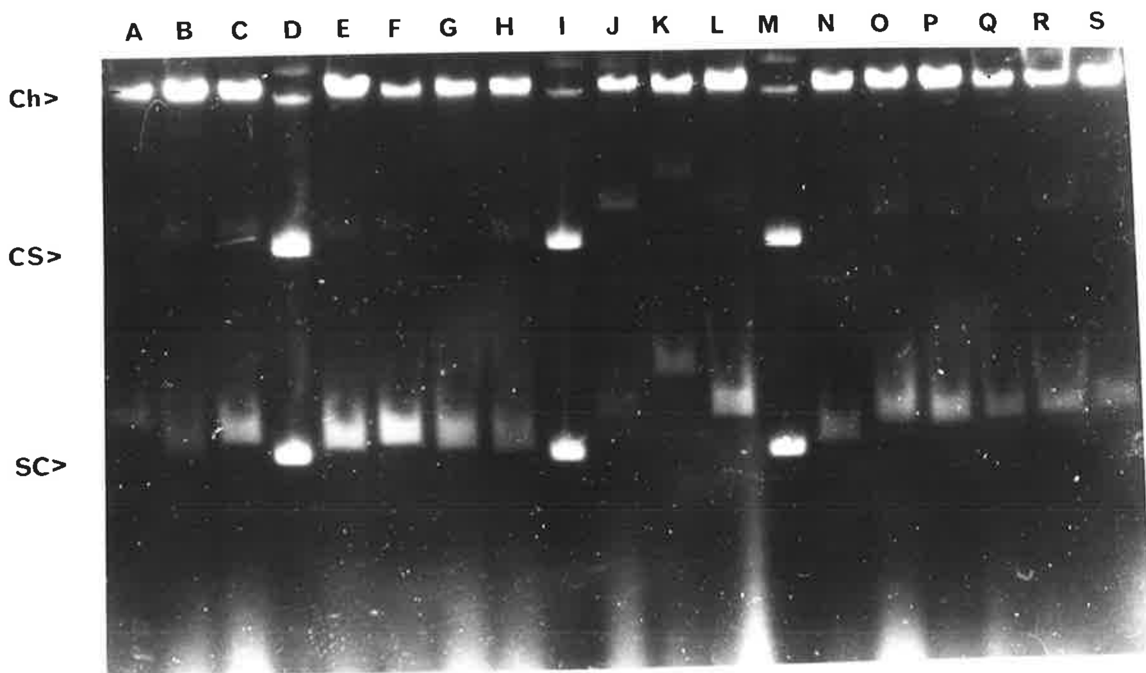

A B C D E F G H I J K L M N O P Q R S
Ch>
CS>
SC>

FIGURE IV,6.

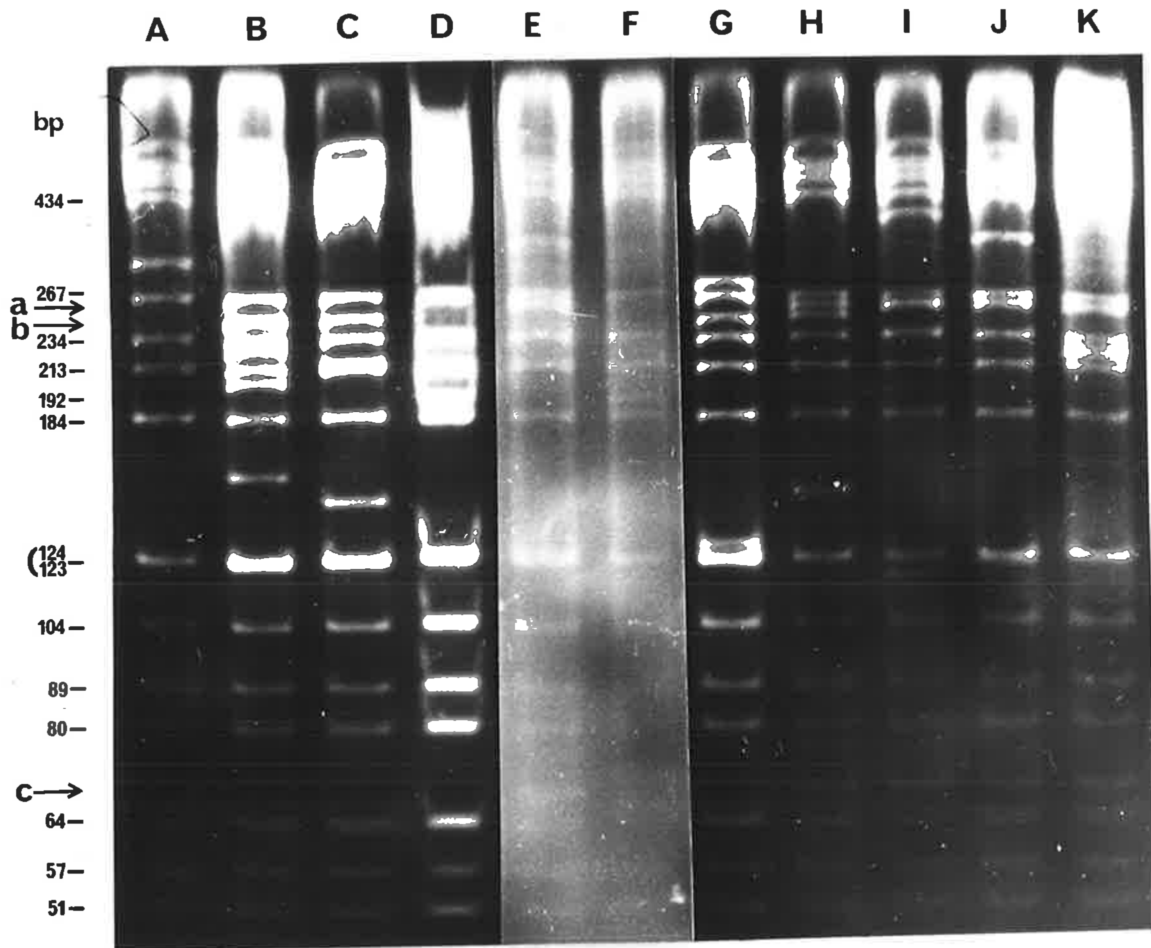
POLYACRYLAMIDE GEL ELECTROPHORESIS OF *Hae*III CUT 'MINISCREEN' PLASMID DNA

Plasmid DNA (II,2,L,ii) from Ap^rTc^s colonies digested with *Hae*III restriction endonuclease was electrophoresed on a 5% polyacrylamide gel (II,2,H,i) and the DNA visualized by ultraviolet light after ethidium bromide staining. Track D is *Hae*III digested pBR322 DNA as markers (Sutcliffe, 1978), all other tracks are from different recombinant clones.

The 192 base pair fragment contains the single *Hin*dIII restriction site and is therefore absent from *Hae*III digests of clones with inserts in this site. The plasmid in track F retained this 192 base pair fragment, had no additional *Hae*III bands to that of pBR322 and therefore contained no inserted DNA. All other tracks show additional bands to that of *Hae*III cut pBR322 DNA and the loss of the 192 base pair band. Five tracks, E, H, I, J and K, contained common additional bands (a) and (c) and were later shown to contain sequences coding for chicken β globin. Tracks B, C and G all contained the common additional band (b) (and were later found to have a fragment which comigrated with the 123 and 124 base pair bands of *Hae*III cut pBR322 DNA) and were later shown to contain sequences coding for chicken α globin.

These four bands a, b, c and the 124 base pair comigrant coincided precisely with the major *Hae*III digestion products of chicken globin d.s.cDNA (Fig. III,5).

Note: Only 10 of the 30 colonies screened are shown here.


A
B
C
D
E
F
G
H
I
J
K
bp
434
a
267
b
234
213
192
184
124
123
104
89
80
c
64
57
51

D. Transformation, selection and screening of recombinants

As detailed in section II,2,P, initial experiments involving recombinant DNA in the presence of viable organisms, and therefore requiring physical containment, were performed by either Dr. J. Shine or Dr. A. Ullrich.

A total of 30 ng of recombinant DNA (ligated cDNA, from each of fractions A, B and C shown in Figure IV,4, and plasmid DNA) was used to transform 100 μl of resuspended *E. coli* χ1776 cells (II,2,L,i). Thirty colonies with $Ap^{r}Tc^{s}$ phenotype were subjected to 'mini-screening' (II,2,L, ii) and aliquots of the resultant plasmid DNA electrophoresed on 1% agarose gels (II,2,H,ii) or *Hae*III restriction endonuclease digested and electrophoresed on 6% polyacrylamide gels (II,2,H,i). The results of this screening are shown in Figures IV,5 and IV,6.

*Hae*III digests of many hybrid plasmids revealed the presence of either of the two *Hae*III cDNA bands of approximately 260 base pairs each (see Figure III,5, track A) which suggested that these two fragments ((a) and (b) in Figure IV,6) and the cDNA clones which contained them, were derived from two different major cDNA species.

Since *E. coli* χ1776 is not easily amplified (II,2,M,i), plasmid DNA in these cells was isolated from 200 ml overnight cultures as described in section II,2,M,ii. Once recombinant DNA sequences were identified, by sequence analysis (II,2,N), and permission from ASCORD was granted, these plasmid DNAs were transferred to the *E. coli* K12 derivative RRI (Bolivar *et al.*, 1977) and the DNA isolated from amplified cultures (II,2,M).

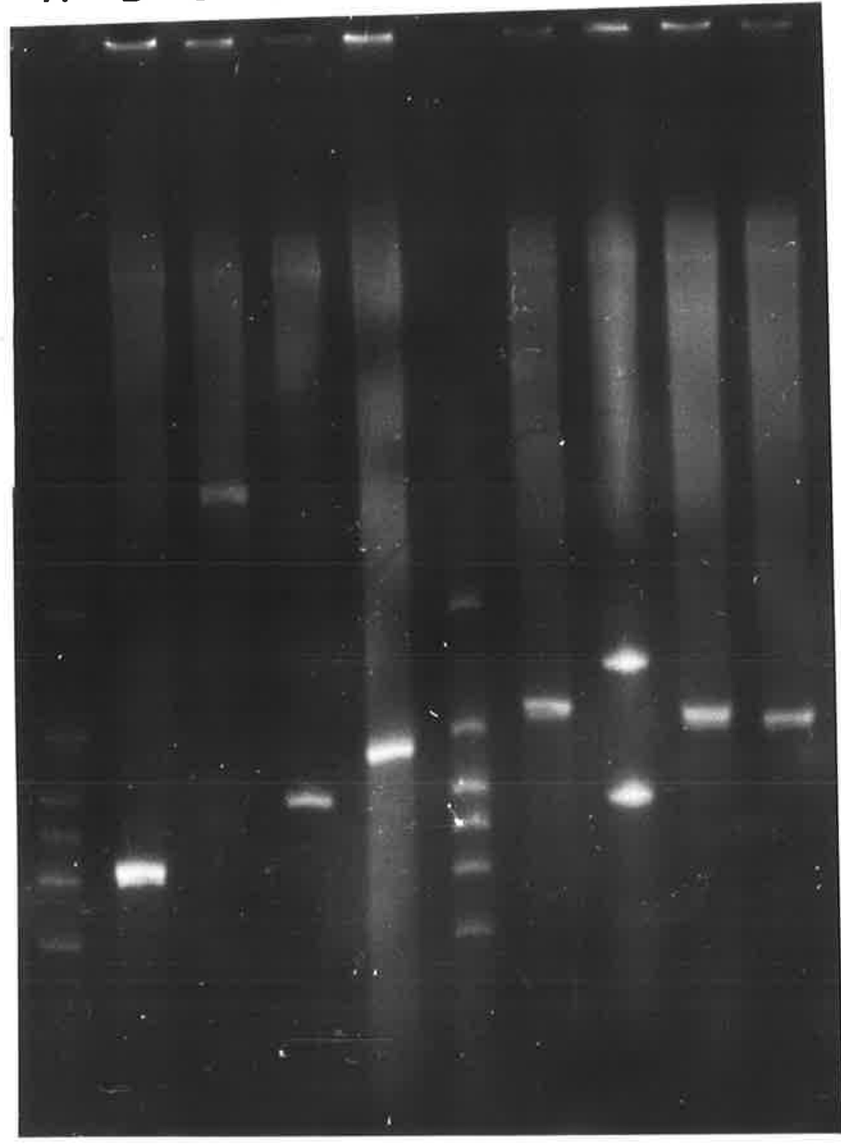
Eight clones, with obvious inserts (Figures IV,5 and

FIGURE IV,7.

POLYACRYLAMIDE GEL ELECTROPHORESIS OF *Hsu*I EXCISED 'INSERT' DNA FOR SIZE DETERMINATION

Aliquots of *Hsu*I cut plasmid DNA from $Tc^{s}Ap^{r}$ colonies were electrophoresed on a 4% polyacrylamide gel (II,2,H,i) to determine the approximate length of inserted sequences. Tracks A and F contain *Hae*III digested pGH-1 DNA (a pBR322 derivative recombinant containing a 670 base pair *Hae*III fragment, a gift of P. Seeburg) as molecular weight markers, in base pairs (Sutcliffe, 1978 and P. Seeburg pers. comm.). Tracks C and H show two inserted sequences, C has no *Hin*dIII linker separating the two inserts, track H does. The linearized pBR322 DNA does not run as a tight band on these gels, but as a smear down the gel. The reason for this is not known.

The DNA was visualized by ultraviolet light after staining with ethidium bromide.


A B C D E F G H I J
bp
670
587
540
504
458
434

IV,6), were otherwise randomly selected to be analyzed in greater detail. Larger preparations (from 200 ml cultures) of the plasmid from these clones was isolated and subjected to *Hsu*I cleavage. Aliquots of these digestion reactions were electrophoresed on a 4% polyacrylamide gel (II,2,H,i) to determine the size of the insert DNA (Figure IV,7). Two clones were found to have two inserted cDNA sequences each, in one case (Figure IV,7, track C) the inserts were not linked by a *Hin*dIII restriction site (synthetic linker DNA) consequently characterization of this clone was abandoned. The other eight inserts (Figure IV,7, tracks B, D, E, G, H, I and J) ranged in size from 470 to about 630 base pairs. Polyacrylamide gel electrophoresis of cDNA, shown in Figure IV,4, although effective in removing monomer linker sequences, was ineffective in size fractionation of cDNA as the 470 base pair insert (Figure IV,7, track B) came from fraction A of Figure IV,4, whilst the 580 base pair insert (Figure IV,7, track J) came from fraction C of Figure IV,4.

E. Sequence analysis of insert DNA

Insert DNA (2 μg) was isolated from those recombinant plasmids thought to contain globin sequences by *Hsu*I digestion and electrophoresis on 3 mm thick polyacrylamide gels (II,2, H,i) followed by electroelution (II,2,J).

The insert DNA was labelled at the 5' termini by a reaction with γ-^{32}P-ATP catalyzed by T_4 polynucleotide kinase (II,2,N,i,a). After heating at 65°C for 20 minutes to inactivate the kinase enzyme, the DNA was cleaved with *Hae*III restriction endonuclease, electrophoresed on 6% polyacrylamide gels and the radioactively labelled DNA bands located by autoradiography and electroeluted.

FIGURE IV,8.

SEQUENCING GEL OF pCG_{α}-1

This figure shows the derived sequence of part of a $Tc^{s}Ap^{r}$ clone which contained the *Hae*III fragment (b) (Fig. IV,6; track C). The four reactions are specific for:

G - guanine

P - purine

Y - pyrimidine

C - cytosine.

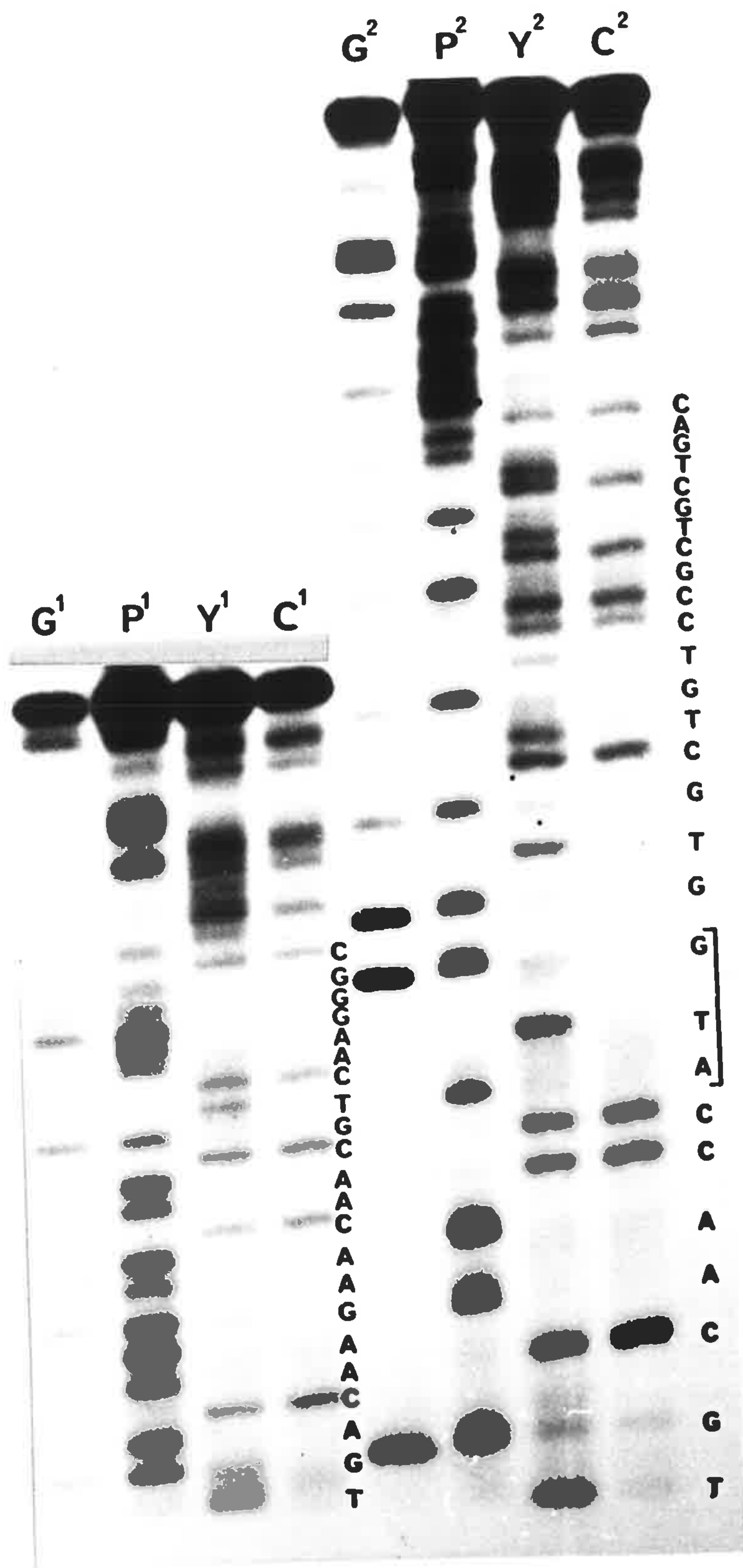
Subscript numbers indicate the order of loading of the reaction products. The second set (2) was loaded when xylene cyanol tracker dye in the first set (1) had traversed half the length of the gel (20 cm). In the second set the first 10 bases including 7 of the synthetic linker have been run off of the gel.

ATG initiation codon is bracketed.

The sequencing gel was 20% polyacrylamide.

Migration of DNA bands was judged by comigration of tracker dyes. On 20% polyacrylamide-urea sequencing gels xylene cyanol comigrates with DNA of 28 bases in length, bromophenol blue with DNA 10 bases long (Maxam and Gilbert, 1977).

In the autoradiograph shown here the bromophenol blue dye from the second loading had run just off the gel.



G1 P1 Y1 C1
G2 P2 Y2 C2
C G G G A C T G C A A C A A G A A C A G T
C A G T C G T C G C C T G T C G T G G T A C C A A C G T

FIGURE IV,9.

SEQUENCING GEL OF pCG_{β}-1

This figure shows the derived sequence of part of a $Tc^{s}Ap^{r}$ clone which contained the *Hae*III fragments (a) and (c) (Fig. IV,6; track J).

Details are as in Fig. IV,8.

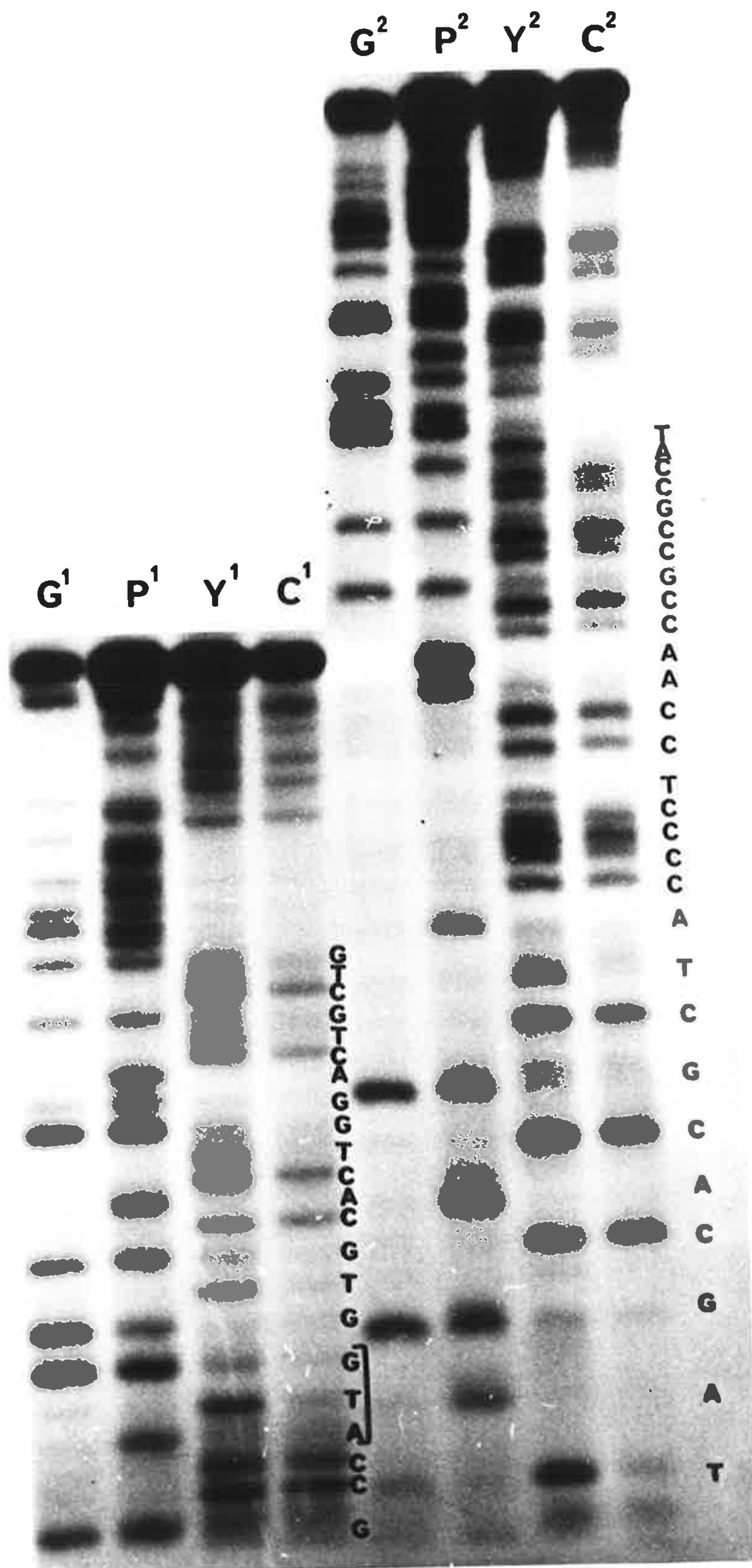
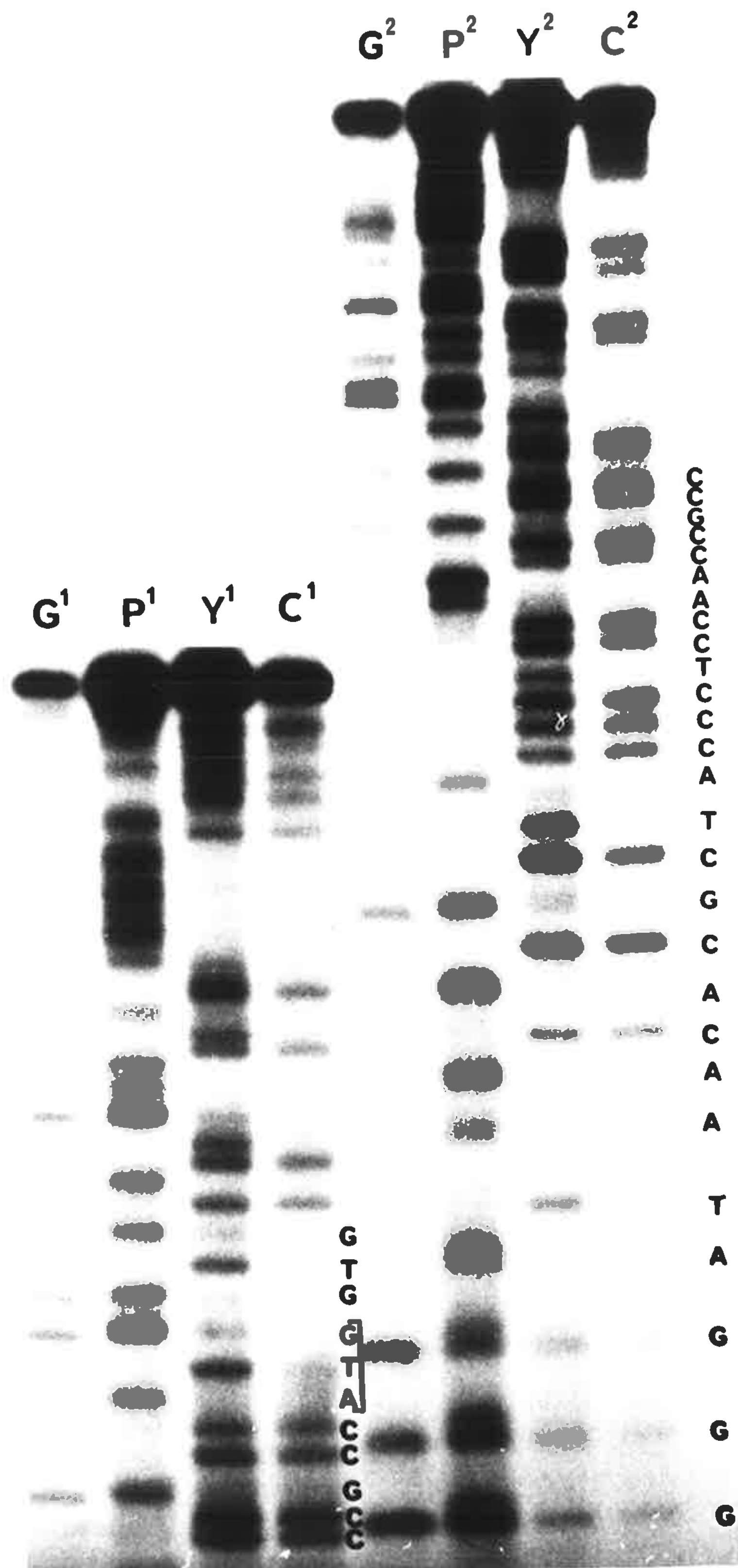

G2 P2 Y2 C2
G1 P1 Y1 C1
GTCGTCAGGTCACGTGGTACCG
TACCGCCGCCAACCTCCCCATCGCACGAT

FIGURE IV,10.

SEQUENCING GEL OF pCG_{β}-2

This figure shows the derived sequence of part of a $Tc^{s}Ap^{r}$ clone which contained the *Hae*III fragments (a) and (c) (Fig. IV,6; track K).

Details are as in Fig. IV,8.



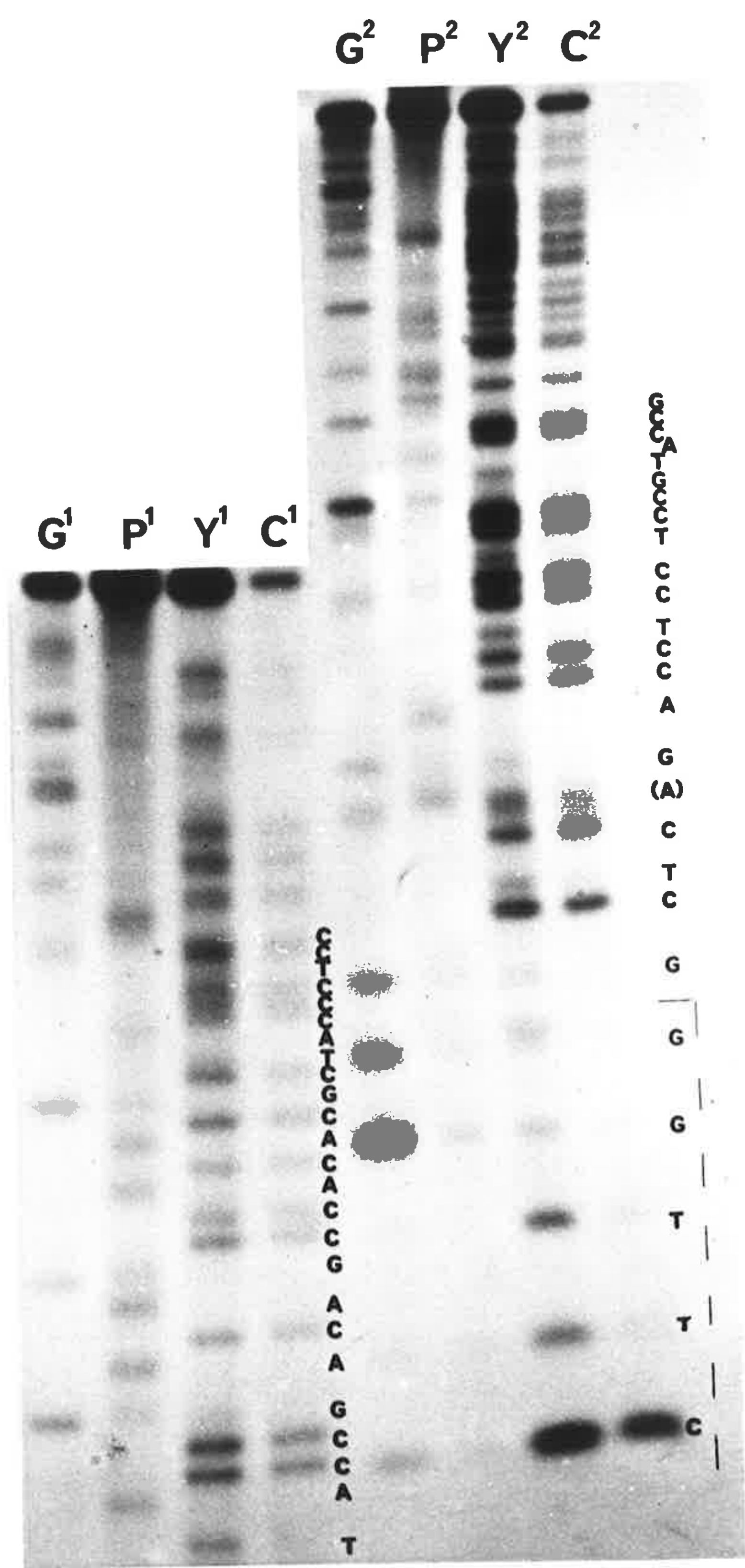
G1 P1 Y1 C1
G2 P2 Y2 C2
G T G G T A C C G C C
C C G C C A A C C T C C C A A T C G C A C A A T A G G G

FIGURE IV,11.

SEQUENCING GEL OF pCG_{β}-3

This figure shows the derived sequence of part of a $Tc^{s}Ap^{r}$ clone which contained the *Hae*III fragments (a) and (c) (Fig. IV,6; track H) and was the larger of two inserts in the one recombinant plasmid (Fig. IV,7; track H).

Details are as in Fig. IV,8 except that only the first two bases of the synthetic linker DNA were run off of the gel. The dashed bracket indicates bases derived from the synthetic linker DNA, the A residue in brackets was determined from another sequencing gel (not shown). The original autoradiograph of this gel could be read well past the ATG initiation codon and the derived sequence had a coding potential for chicken β globin (see Fig. IV,13).



G1 P1 Y1 C1
G2 P2 Y2 C2

FIGURE IV,12.

CODING POTENTIAL OF DERIVED SEQUENCES

(ALPHA CODING CLONES)

The coding potential of sequences derived from cDNA inserts is shown. The bases corresponding to the AUG initiation codon are indicated by a solid box. Sequences from pCG_{α}-1 and pCG_{α}-3 differ from the α_A sequence of Matsuda *et al.* (1977) at amino acid position 4, which is *asn* in α_A, indicated by the dotted box. All other nucleotide sequences shown here have a coding potential which is in agreement with that required by the α_A globin chain amino acid sequence, and were designated alpha coding clones on this basis.

Note: The original sequencing gel autoradiograph shown in Fig. IV,8 could be read considerably further than is indicated in this figure.

FIGURE IV,12.

pCG_{α}-1 (Fig. IV,8)

val leu ser ala ala asp lys asn asn

[linker] NNNTGCAACC [ATG] GTG CTG TCC GCT GCT GAC AAG AAC AAC

10

val lys gly ile phe thr lys ile ala gly

GTC AAG GGC ATC TTC ACC AAA ATC GCC GGC C: *Hae*III site

pCG_{α}-2 (sequencing gel not shown)

50

tyr phe pro his phe asp leu ser his gly ser

[linker] N NNC TAC TTC CCC CAC TTC GAT CTG TCA CAC GGC TCC

ala

GCT C ⟶ *Alu*I

pCG_{α}-3 (sequencing gel not shown)

val leu ser ala ala asp lys

[linker] NNNCGGGTGCAACC [ATG] GTG CTG TCC GCT GCT GAC AAG

10

asn asn val lys gly ile

AAC AAC GTC AAG GGC ATC ⟶ *Hae*III.

FIGURE IV,13.

CODING POTENTIAL OF DERIVED SEQUENCES
(BETA CODING CLONES)

The coding potential of sequences derived from cDNA inserts is shown. The bases corresponding to the AUG initiation codon are indicated by a solid box. The coding region nucleotide sequences of the β coding inserts, shown here, were consistent with those required for the chicken β globin amino acid sequence (Matsuda *et al.*, 1973). Non-homologous bases (underlined) were found in the 5' terminal sequences (with respect to the mRNA sequence) of pCG_{β}-1 and pCG_{β}-2 when compared to pCG_{β}-3 (as is seen in the sequencing gels, Figs. IV,9 to IV,11).

Note: The original sequencing gel autoradiographs shown in Figs. IV,9 to IV,11 could be read considerably further than is indicated in these figures.

FIGURE IV,13.

pCG_{β}-1 (Fig. IV,9)

val his trp thr

linker NNNTAGCACGCTACCCCTCCAACCGCCGCC ATG GTG CAC TGG ACT

10
ala glu glu lys gln leu ile thr gly

GCT GAG GAG AAG CAG CTC ATC ACC GGC C

*Hae*III site

pCG_{β}-2 (Fig. IV,10)

val his trp

linker NNNGGGATAACACGCTACCCTCCAACCGCCGCC ATG GTG CAC TGG

10
thr ala glu glu lys gln leu ile thr gly

ACT GCT GAG GAG AAG CAG CTC ATC ACC GGC C

*Hae*III site

pCG_{β}-3 (Fig. IV,11)

CTTGG GCTCAGACCTCCTCCGTACCGACAGCCACACGCTACCCTCCAACCGCCGCC

linker

val his trp thr ala glu

ATG GTG CAC TGG ACT GCT GAG ⟶ *Hae*III.

End-labelled DNA was then subjected to the sequencing reactions of Maxam and Gilbert (1977) as described in section II,2,N, and electrophoresed on 20% polyacrylamide-urea sequencing gels (Figures IV,8 to IV,11). In these gels poly(dT) sequences, originating from mRNA poly(A) tracts and presumptive 3' untranslated sequences, were found but could not be read as far as the termination codon. Therefore only sequences corresponding to the 5' end of the mRNA, which included sequences from the coding regions, could be used to identify the coding potential of the insert DNA.

The assignment of the derived sequences to α and β globin coding mRNA sequences is shown in Figs. IV,12 & IV,13. In each of the sequencing gels the first ten bases (including seven of the synthetic linker DNA) of the second loading were run off the gel so that the sequences could be read further (with the exception of pCG_{β}-3). From this rapid analysis three insert sequences were shown to code for alpha globin chains and three inserts for beta globin chains. On this basis these clones were designated pCG_{α}-1 to pCG_{α}-3 and pCG_{β}-1 to pCG_{β}-3.

3. DISCUSSION

Using recombinant DNA techniques the cDNA sequences derived from chicken globin mRNA were purified and amplified in the plasmid vector pBR322 in *E. coli* host bacteria. Insertion of the cDNA into the plasmid vector by use of synthetic linker DNA encoding the *Hin*dIII restriction endonuclease recognition site enabled the efficient excision of insert DNA, in a form which was easily end-labelled for DNA sequence analysis using the Maxam and Gilbert (1977) method.

The other major technique of recombinant DNA construc-

tion, based on the annealing of homopolymer extensions of cDNA and vector DNA, usually cleaved at the *Pst*I site (Bolivar *et al.*, 1977), does not allow for such rapid analysis, as DNA excised using the *Pst*I enzyme is very difficult to end-label due to inaccessible 5' termini (see section V,2,A).

DNA sequences derived from the chemical degradation method were found to have a coding potential for alpha or beta chicken globin chains. Alpha clones pCG_{α}-1 and pCG_{α}-3 showed a consistent nucleotide sequence difference from that required for the α_A amino acid sequence at residue position 4, as determined by Matsuda *et al.* (1971). The further analysis of these and other alpha coding clones and their difference in coding potential from that required by the α_A amino acid sequence is the subject of chapter VI.

The amino acid coding potential of nucleotide sequences derived from the beta coding clones was in complete agreement with the amino acid sequence established by Matsuda *et al.* (1973). Nucleotide sequence differences were found between the different inserts restricted to a few bases at the 5' end of each insert (with respect to the mRNA sequence). The characterization of this 5' terminal heterogeneity and complete sequence analysis of the longest beta coding insert, pCG_{β}-3, are described in chapter V.

CHAPTER V

SEQUENCE ANALYSIS OF β GLOBIN CODING cDNA CLONES

CHAPTER V - SEQUENCE ANALYSIS OF β GLOBIN CODING cDNA CLONES

1. Introduction

The primary aim of the work in this chapter was to establish as much of the chicken β globin mRNA sequence as possible, by sequence analysis of cDNA clones so as to allow a comparative study with the β globin mRNA sequences of rabbit and human (Kafatos *et al.*, 1977). These mammalian sequences had shown a high degree of homology which may have only been a result of their limited evolutionary divergence. Comparative analysis with the β sequence of chicken, a more distant evolutionary species, may reveal sequences important to expression of globin genes through their conservation between these species.

The sequence analysis of the 5' terminus (with respect to mRNA sequence) of several β globin coding clones was undertaken to determine the origin of the non-homologous sequences described in chapter IV.

Finally, radioactively labelled β globin coding insert DNA was used as a probe in a restriction endonuclease digest of chicken genomal DNA to determine its usefulness as a probe for the adult chicken β globin gene.

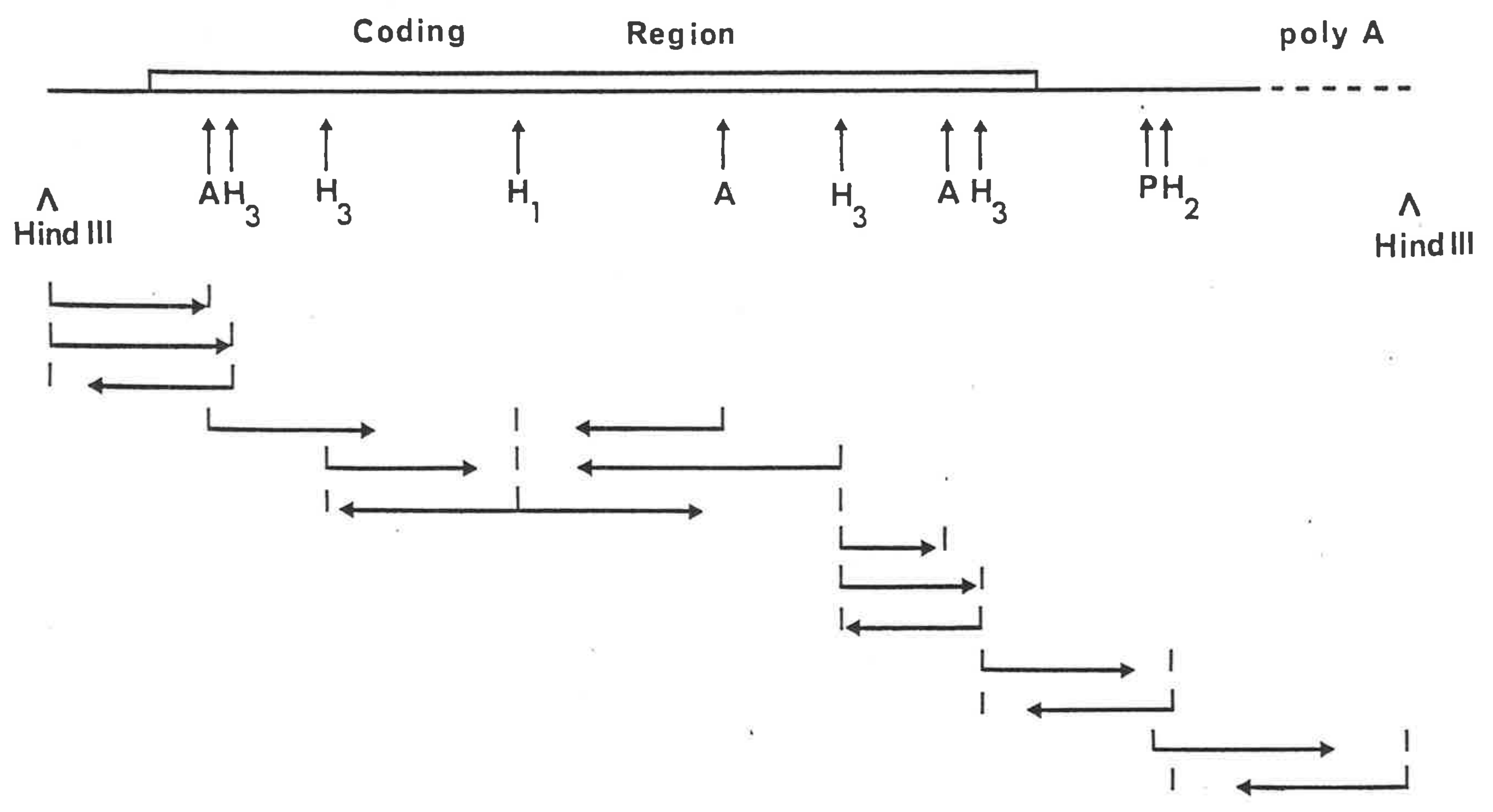
2. Results

The longest β globin coding cDNA insert, pCG_{β}-3, was chosen for complete sequence analysis (Figure IV,7, track H). As this insert was one of two in the same *Hin*dIII site of the one plasmid vector, the first step was to isolate and reinsert the sequence as a single copy into the *Hin*dIII site of pBR322 (using methods described in section II,2,K and L). This recloning was carried out to avoid potential excision

FIGURE V,1.

SEQUENCING STRATEGY OF pCG_{β}-3

Only those restriction sites used in the analysis are shown (A - *Alu*I, H_1 - *Hha*I, H_2 - *Hpa*II, H_3 - *Hae*III, P - *Pst*I). Arrows indicate the direction and extent of derived sequences. Λ indicates attachment of synthetic linker DNA encoding *Hin*dIII recognition site.



Coding Region
poly A
A H3
H3
H1
A
H3
A H3
P H2
Hind III
Hind III

of a repeated sequence from the plasmid DNA *in vivo*.

Plasmid DNA was purified from a six litre overnight culture of *E. coli* χ1776 containing the reconstructed pBR322 derivative. The insert DNA was purified by polyacrylamide gel electrophoresis and electroelution (section II,2,J), after excision by *Hsu*I restriction endonuclease from the plasmid DNA.

A. Sequencing strategy

The chemical degradation method of DNA sequencing (Maxam and Gilbert, 1977) was used for all sequence analysis.

The sequencing strategy depended to a large extent upon suitable restriction endonucleases available at the time. Many commercial preparations contained low levels of non-specific exonuclease activity which made them unsuitable for sequence analysis. The cleavage patterns determined by digestion of the chicken globin ds cDNA with *Hae*III, *Hpa*II, *Hha*I and *Alu*I were particularly useful in establishing the sequencing approach shown in Figure V,1, as each of these enzymes was shown to cleave the major cDNA species present. The basic approach used was to cleave the insert DNA with one of these four restriction enzymes, 5' end – label the resultant fragments with T_4 polynucleotide kinase and γ-^{32}P-ATP, then cleave these end-labelled fragments with another of the four enzymes. Fragments labelled at only one end, and therefore suitable for sequence analysis, were then purified by polyacrylamide gel electrophoresis and electro-elution (II,2,J).

For two regions of the sequence, in particular, this approach was insufficient and alternative techniques were used. The first of these was strand separation using

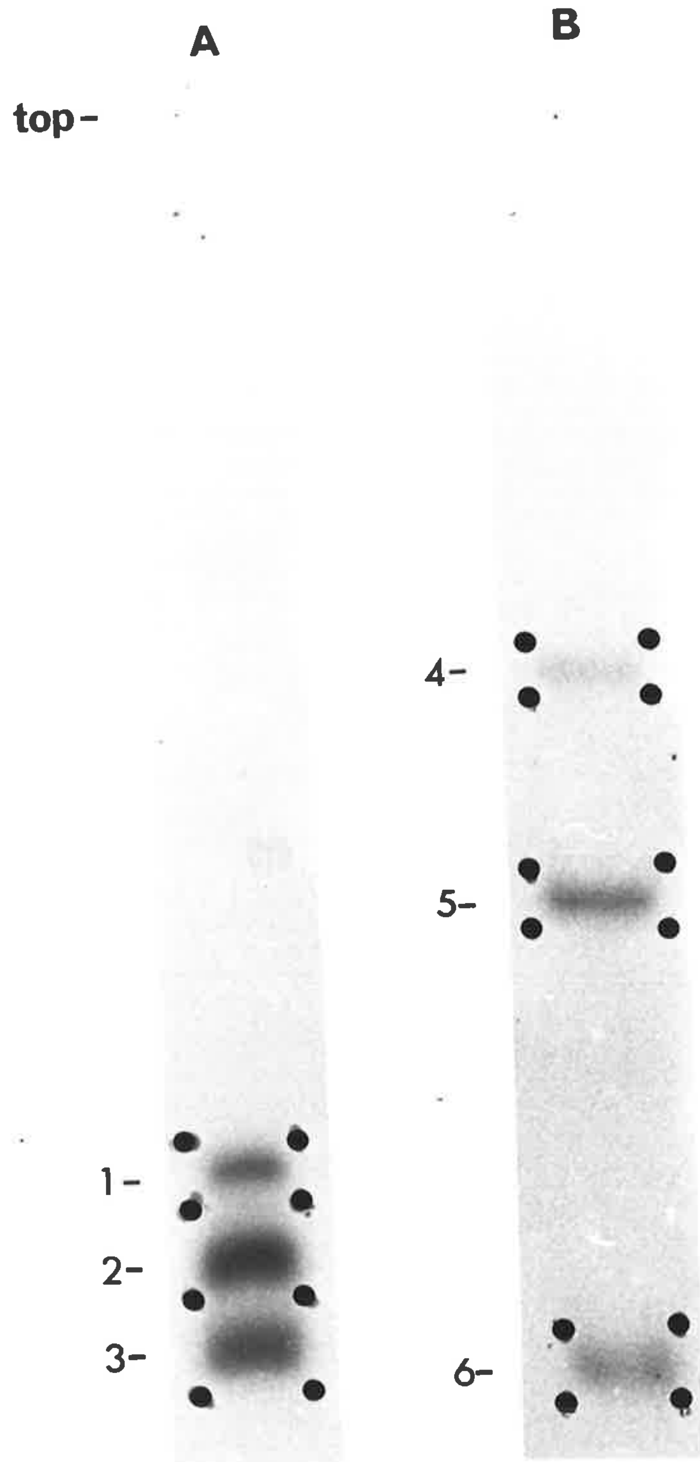
FIGURE V,2.

STRAND SEPARATION OF END LABELLED DNA

Use was made of strand separation techniques (Maxam and Gilbert, 1977) to generate DNA fragments, labelled at only one end, suitable for sequence analysis.

Insert DNA from pCG_{β}-3 was digested with restriction endonuclease *Hae*III, end-labelled with T_4 polynucleotide kinase and γ-^{32}P-ATP (II,2,N,i) and the labelled DNA fragments separated on a 5% polyacrylamide gel (II,2,H,i). Two radioactively labelled fragments, A (69 base pairs in length) and B (170 base pairs in length) were isolated from gel slices by electroelution (II,2,J). These isolated fragments were then denatured by heating in 0.3 M NaOH, 1 mM EDTA at 90°C for 2 min and electrophoresed on a non-denaturing gel (II,2,H,i) until the double-stranded form of the fragment should be about 5 cm from the bottom of the gel (20 cm long). After autoradiography, shown here, ^{32}P-labelled DNA fragments were electroeluted from gel slices (dots indicate corners of gel slice).

In both cases three radioactively labelled fragments were found (1 to 6) which were subjected to sequence analysis. Fragments 1 and 6 were the original double-stranded forms of fragments A and B respectively, the other four fragments were the separated single strands of fragments A and B (sequencing gels are not shown).


A
B
top-
1-
2-
3-
4-
5-
6-

sodium hydroxide solution as a denaturant. A short sequence of 16 bases, from amino acids *leu* 133 to *ala* 138 was inaccessible due to the location of restriction sites (and exhausted supplies of *Alu*I and *Hpa*II restriction enzymes). The insert DNA was cut with *Hae*III, labelled with γ-^{32}P-ATP and T_4 polynucleotide kinase, and the 69 base pair fragment isolated from a 5% polyacrylamide gel. This DNA was heated in 0.3 M NaOH, 1 mM EDTA at 90°C for 2 minutes, then chilled on ice and loaded on to a 5% polyacrylamide gel with a 50:1 acrylamide to bisacrylamide ratio. After autoradiography (Figure V,2) the separated single-stranded DNA fragments were electroeluted from gel slices and subjected to sequencing reactions.

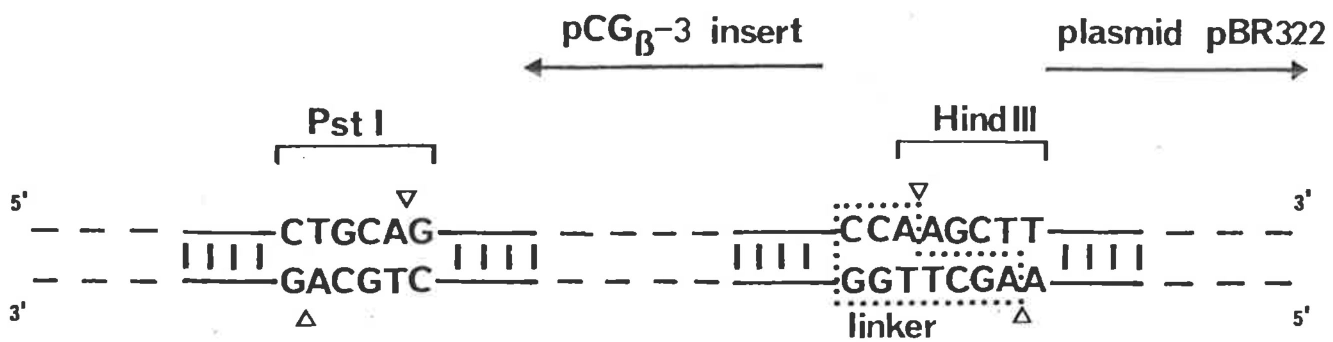
The 3' end *Hae*III fragment, with respect to the mRNA sequence, of ∿177 base pairs containing ∿70 base pairs of poly(A/T) was also strand separated in a parallel track (see Figure V,2) for sequence analysis. Separation of the single-stranded forms of both fragments was effective, but more so in the case of the 177 base pair fragment, probably due to the extensive homopolymer sequences.

The second difficult region was the 45 bases of 3' untranslated region between the *Hpa*II site and the poly(A) tract. Since there was no verification of the sequence by coding potential, the *Hpa*II site had to be sequenced across to ensure that two or more *Hpa*II sites did not exist very close together in the sequence. Use was made of a *Pst*I site centred 10 bases to the 5' side (with respect to the mRNA sequence) of the *Hpa*II site. Because of the 3' single-strand protrusion left by the *Pst*I digestion, these sites are notoriously difficult to label using γ-^{32}P-ATP

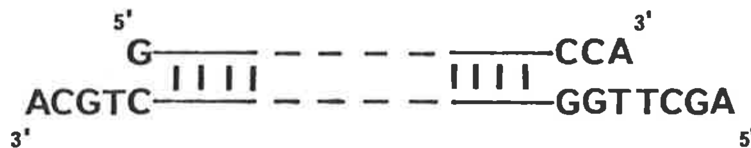
FIGURE V,3.

LABELLING OF pCG_{β}-3 *Pst*I SITE

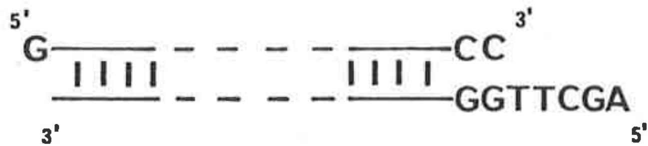
Schematic diagram of specific labelling of the *Pst*I site with α-^{32}P-dCTP and the Klenow fragment of *E. coli* DNA polymerase I. Asterisk indicates site of ^{32}P phosphate. Dotted box indicates the bases derived from the synthetic linker DNA sequence. Triangles indicate the sites of cleavage of the restriction endonucleases.



(a) Pst I, Hind III digestion



(b) Klenow fragment of DNA Pol I (3'→5' exonuclease activity)



(c) Klenow fragment of DNA Pol I (DNA Polymerase activity) α-^{32}P*dCTP, dATP

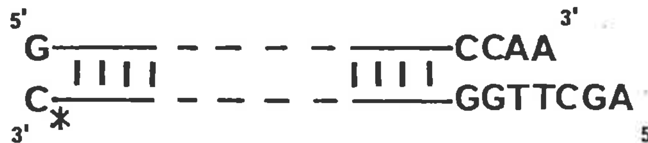


FIGURE V,4.

SEQUENCING GEL OF THE *Pst*I - *Hin*dIII pCG$_{\beta}$-3 FRAGMENT

This figure shows the derived sequence from the selectively labelled *Pst*I - *Hin*dIII pCG$_{\beta}$-3 fragment (see Fig. V,3). The four reactions are specific for:

G - guanine

P - purine

Y - pyrimidine

C - cytosine.

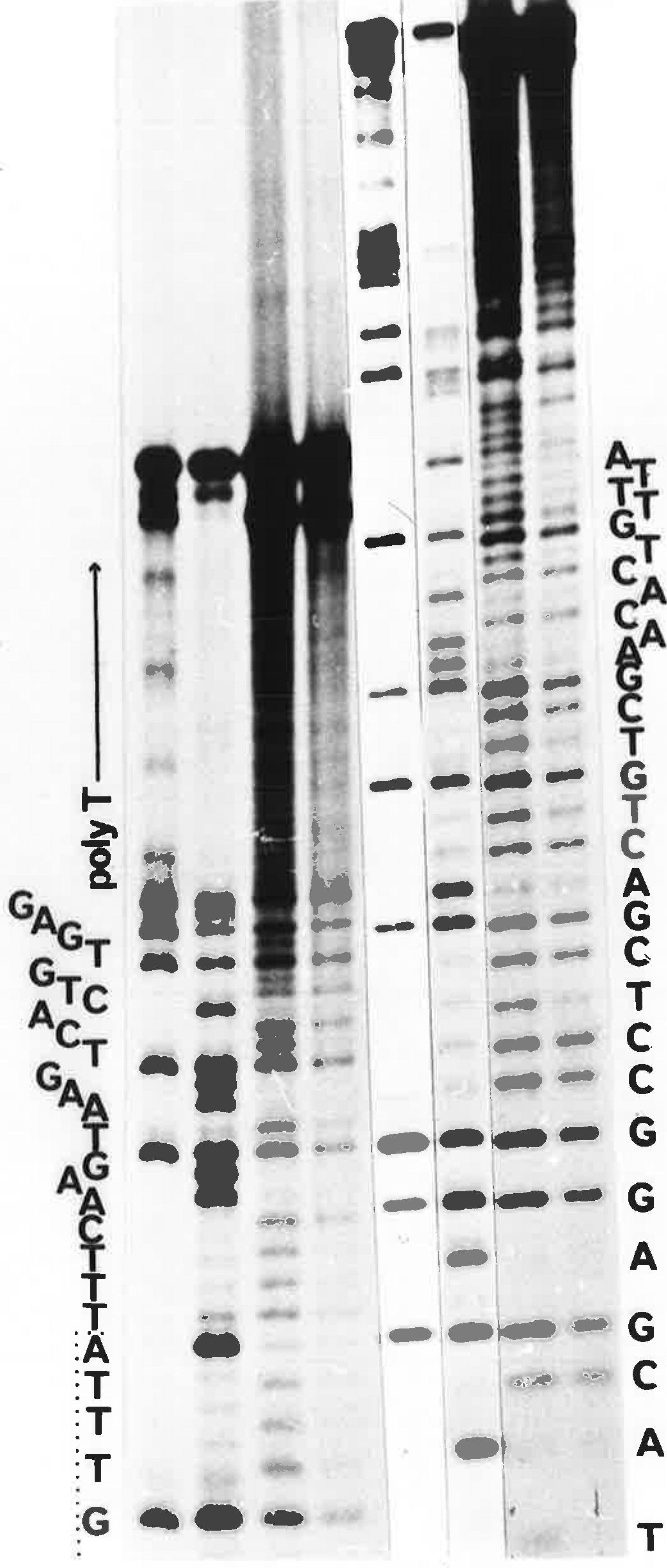
Subscript numbers indicate the order of loading of the reaction products. The second set (2) was loaded when xylene cyanol tracker dye in the first set (1) had traversed half the length of the gel (20 cm). In the second set the first three bases have been run off the gel, which was 20% polyacrylamide.

Hydrazine, in the Y and C reactions (Y° and C° in section II,2,N,ii) was found to have cleaved the DNA at guanine residues. This did not greatly impair reading of the sequence and was later reduced by the use of modified sequencing reactions (Busslinger *et al.*, 1979).

The poly(dT) tract, coresponding to the poly(A) tract of the mRNA was seen to contain 4 G residues, which may be an artefact of the gel, but could arise from slippage of the oligo(dT) primer during cDNA synthesis (Seeburg *et al.*, 1977b).

The sequence shown here was verified by sequence analysis, in the opposite direction, of end-labelled pCG$_{\beta}$-3 insert DNA cut with *Hae*III restriction endonuclease (not shown).

The dotted bracket in the first set indicates those bases which are also read, as the last five bases, in the second set.



G1 P1 Y1 C1 G2 P2 Y2 C2
poly T
GAGT
GTC
ACT
GAA
T
AG
A
C
T
T
T
A
T
T
T
G
AT
TT
GT
CA
CA
A
GC
T
G
T
C
A
G
C
T
C
C
G
G
A
G
C
A
T

FIGURE V,5.

COMPLETE NUCLEOTIDE SEQUENCE OF THE mRNA

CORRESPONDING TO pCG_{β}-3

The sequences of other β coding cDNA inserts are included to show 5' terminal non-homologous bases (underlined).

A - adenine

G - guanine

C - cytosine

U - uracil

N - undetermined nucleotide.

Only one β coding cDNA insert, pCG_{β}-5, did not show 5' terminal non-homologous bases.

5'

start

pCG_{β}-3 GCUCAGACCUCCUCCGUACCGACAGCCACACGCUACCC UCCAACCGCCGCC AUG

pCG_{β}-2 NNNGGGAUAACACGCUACCC UCCAACCGCCGCC AUG

pCG_{β}-1 NNNUAGCACGCUACCCCUCCAACCGCCGCC AUG

pCG_{β}-6 GAGACACGCUACCC UCCAACCGCCGCC AUG

10 20

val his trp thr ala glu glu lys gln leu ile thr gly leu trp gly lys val asn val

GUG CAC UGG ACU GCU GAG GAG AAG CAG CUC AUC ACC GGC CUC UGG GGC AAG GUC AAU GUG

pCG_{β}-4 GUG

30 40

ala glu cys gly ala glu ala leu ala arg leu leu ile val tyr pro trp thr gln arg

GCC GAA UGU GGG GCC GAA GCC CUG GCC AGG CUG CUG AUC GUC UAC CCC UGG ACC CAG AGG

GCU GUC GGU GGG GCC GAA GCC CUG GCC AGG

pCG_{β}-5 GU GGG GCC

50 60

phe phe ala ser phe gly asn leu ser ser pro thr ala ile leu gly asn pro met val

UUC UUU GCG UCC UUU GGG AAC CUC UCC AGC CCC ACU GCC AUC CUU GGC AAC CCC AUG GUC

70 80

arg ala his gly lys lys val leu thr ser phe gly asp ala val lys asn leu asp asn

CGC GCC CAC GGC AAG AAA GUG CUC ACC UCC UUU GGG GAU GCU GUG AAG AAC CUG GAC AAC

90 100

ile lys asn thr phe ser gln leu ser glu leu his cys asp lys leu his val asp pro

AUC AAG AAC ACC UUC UCC CAA CUG UCC GAA CUG CAU UGU GAC AAG CUG CAU GUG GAC CCC

110 120

glu asn phe arg leu leu gly asp ile leu ile ile val leu ala ala his phe ser lys

GAG AAC UUC AGG CUC CUG GGU GAC AUC CUC AUC AUU GUC CUG GCC GCC CAC UUC AGC AAG

130 140

asp phe thr pro glu cys gln ala ala trp gln lys leu val arg val val ala his ala

GAC UUC ACU CCU GAA UGC CAG GCU GCC UGG CAG AAG CUG GUC CGC GUG GUG GCC CAU GCC

stop

leu ala arg lys tyr his

CUG GCU CGC AAG UAC CAC UAA GCACCAGCACCAAAGAUCACGGAGCACCUACAACCAUUGCAUGCACCU

GCAGAAAUGCUCCGGAGCUGACAGCUUGUGACAAAUAAAGUUCAUUCAGUGACACUC poly(A) 3'

and polynucleotide kinase, which requires an accessible 5' end. After several unsuccessful attempts at labelling this site using the kinase method, an alternate labelling method was devised (see Figure V,3). This method was based on the observations of Donelson and Wu (1972) and involved the use of the two enzyme activities of the Klenow fragment of *E. coli* DNA polymerase I to catalyze an exchange reaction between α-^{32}P-dCTP and the first base paired, 3' terminal C residue, as described in Figure V,3. Use was made of the specificity of this reaction to label only the *Pst*I end of the *Pst*I - *Hin*dIII DNA fragment by including unlabelled dATP to block the dCTP exchange at the *Hin*dIII end. This was effective because the rate of the polymerase reaction was faster than that of the 3' → 5' exonuclease. The labelled fragment was isolated by acrylamide gel electrophoresis and electroelution and subjected to sequencing reactions (Figure V,4).

Figure V,1 illustrates the direction and extent of derived sequences. Where possible, both strands of DNA (or the one strand in both directions) were sequenced and where this proved difficult one strand was sequenced a number of times. Four bases at amino acids *arg* 61 and *ala* 62 were not directly sequenced but their base composition inferred from the existence of the *Hha*I restriction site (GCGC). There was complete agreement, in this case, between the coding potential of pCG$_\beta$-3 (Figure V,5) and the adult chicken β globin amino acid sequence established by Matsuda *et al.* (1973).

B. <u>5' Terminal non-homology</u>

To determine the origin of the non-homologous bases seen in pCG$_\beta$-1 and pCG$_\beta$-2 (Figure IV,13) other β globin

FIGURE V,6.

AUTORADIOGRAPH OF SEQUENCING GEL SHOWING

5' TERMINAL SEQUENCE (WITH RESPECT TO mRNA SEQUENCE)

OF pCG_{β}-4

The relevant sequence of pCG_{β}-3 is included to show non-homology (indicated by dashed horizontal lines). The solid bracket indicates the base sequence derived from the synthetic linker DNA. The dotted bracket indicates bases read in both sets of gel loading. The arrows mark sites where C residues are absent from the sequence due to methylation of *Eco*RII sites by *E. coli*.

The four reactions are specific for:

G - guanine

P - purine

Y - pyrimidine

C - cytosine.

Subscript numbers indicate the order of loading of the reaction products. The second set (2) was loaded when xylene cyanol tracker dye in the first set (1) had traversed half the length of the 20% polyacrylamide gel.

G^2 P^2 Y^2 C^2

G^1 P^1 Y^1 C^1

AGTCGTCGGA CG G TCCG AA GCCG

pCG_{β}-4

GCCGGGTGGCTGTCGGTGGGTT

pCG_{β}-3

GCCGGGTGTAAGCCGGTG

coding inserts were sequenced at their 5' termini (with respect to the mRNA sequence). The logic of this approach was that if the non-homology was due to allelic variants in the 5' untranslated region sequence, then the different cloned sequences should fall into groups with consistent differences in the sequence of this part of the 5' untranslated region. If the non-homology was due to "errors" introduced during the construction of the recombinant DNA then more variable sequences would be found. Where a cloned sequence had its 5' terminus (with respect to the mRNA sequence) within the coding region then this 5' non-homology could result in an incorrect coding potential.

The relevant terminal sequences of five β globin coding inserts are included in Figure V,5 along with the pCG_{β}-3 sequence to illustrate this 5' non-homology, which was present in all but one case, pCG_{β}-5. These inserts terminate at common regions of the mRNA sequence suggesting secondary structure is important in the self-priming second strand reverse transcriptase reaction (see Figure IV,1). The non-homologous bases are confirned to the terminal 20 bases or less of each insert. This suggests that incorrect bases are inserted by reverse transcriptase during the "loop" formation or, more likely, arose during the repair process with *E. coli* DNA polymerase I in the blunt-ending reaction (Figure IV,1). It follows that several bases at the 5' end of pCG_{β}-3 may be incorrect. Confirmation that these errors were due to *in vitro* reactions was obtained by inspection of the 5' terminal sequence of the β globin coding clone pCG_{β}-4 (Figures V,5 and 6) which terminates within the coding sequence. A potential model for the generation of "errors"

FIGURE V,7.

POTENTIAL MECHANISM FOR THE GENERATION OF "ERRORS" AT THE 5' END OF cDNA

Limited S_1 nuclease cleavage of the hairpin-loop generated during synthesis of double-stranded cDNA results in a duplex molecule terminating in two non-paired strands. The unpaired 3' end of the strand complementary to the 5' terminus of mRNA is then removed by the 3'-5' exonuclease activity of DNA polymerase I and then resynthesised using the other unpaired strand as template. This results in the incorporation of "incorrect" sequences since this portion of the cDNA is derived from sequences further towards the 5' end of the mRNA.

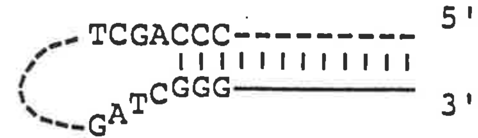
The sequences shown here are a hypothetical example and do not represent sequences derived from either α or β globin cDNA clones.

FIGURE V,7.

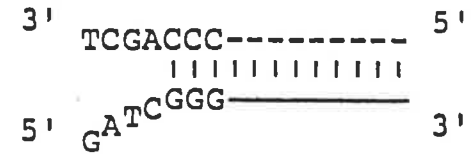
mRNA — 5' CCCGAUC———AGCUGGG——— 3'

single-stranded cDNA — 3' GGGCTAG-------TCGACCC--------- 5'

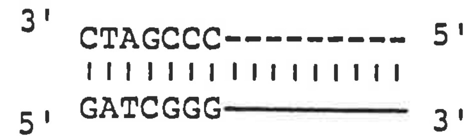
double-stranded "hairpin" cDNA



limited S_1 nuclease digestion



DNA polymerase I repair



deduced mRNA sequence

correct mRNA sequence

FIGURE V,8.

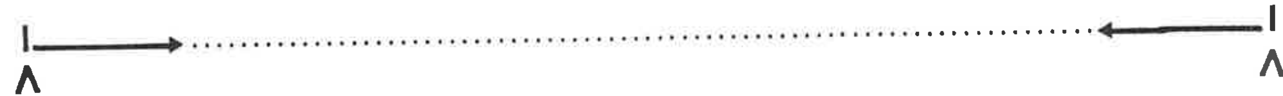
DERIVED SEQUENCES OF β GLOBIN CODING cDNA CLONES

This figure illustrates the extent of chicken β globin mRNA sequences cloned in the six β globin cDNA clones characterized. Solid lines indicate the extent, and arrows the direction, of derived sequences from each clone. The sequencing strategy of pCG_{β}-3 is illustrated in Fig. V,1. The internal sequences of pCG_{β}-4 insert DNA were obtained from end-labelled fragments generated by either *Hha*I digestion, end-labelling with γ-^{32}P-ATP and T_4 polynucleotide kinase and *Hae*III digestion, or *Hae*III digestion, end-labelling with γ-^{32}P-ATP and T_4 polynucleotide kinase and strand separation as illustrated in Fig. V,2. All other sequences were obtained from end-labelled fragments generated by end-labelling isolated insert DNA with γ-^{32}P-ATP and T_4 polynucleotide kinase and digestion with either *Hae*III or *Hha*I restriction endonucleases.

Vertical bars indicate the site of end-labelling with γ-^{32}P-ATP and T_4 polynucleotide kinase. ∧ indicates the site of synthetic linker DNA (encoding *Hin*dIII site) attachment to cDNA sequence. Dotted lines indicate that the sequences were present (by size of insert DNA and restriction digest fragments) but their sequences not derived.

β-globin mRNA

pCG_{β}-1

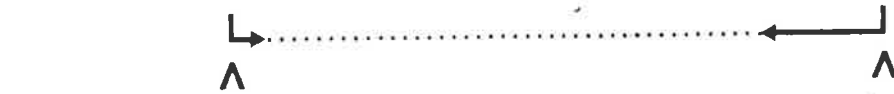


pCG_{β}-2

pCG_{β}-3

pCG_{β}-4

pCG_{β}-5



pCG_{β}-6

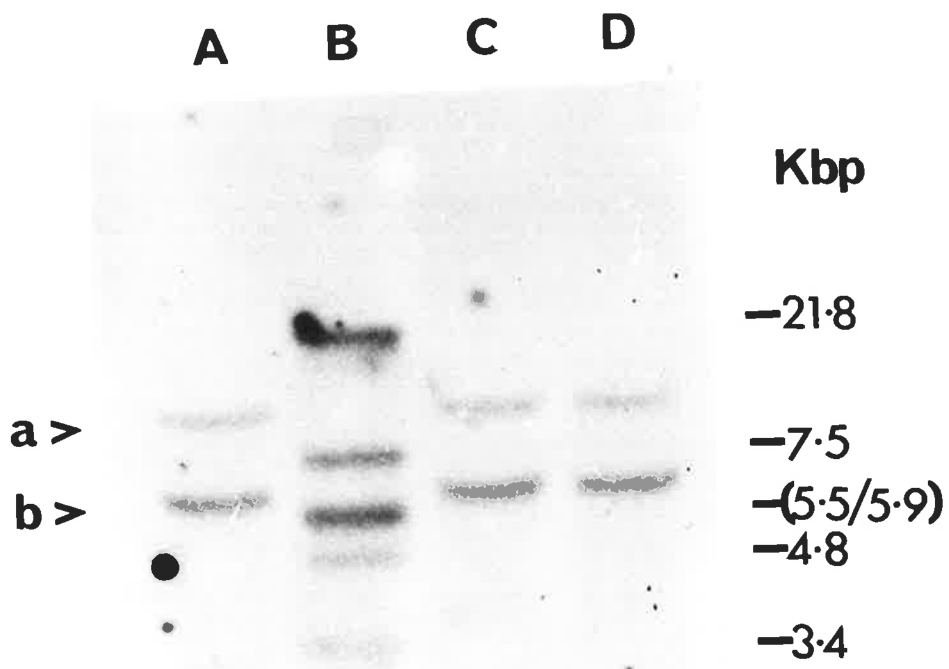
FIGURE V,9.

SOUTHERN BLOT OF *Eco*RI CUT CHICKEN GENOMAL DNA

WITH pCG$_{\beta}$-3 PROBE

Chicken genomal DNA was digested to completion with restriction endonuclease *Eco*RI and electrophoresed on a 1% agarose gel (II,2,O,ii). DNA fragments were transferred from the gel to a nitrocellulose filter. After annealing to the filter β globin hybridizing sequences were detected by hybridization with radioactively labelled pCG$_{\beta}$-3 insert DNA (II,2,O,i).

Tracks A, C and D are *Eco*RI cut chicken genomal DNA (20 μg each). Track B is *Eco*RI cut λ DNA, radioactively labelled (a gift of P. Seeburg) as molecular weight markers (in kilobase pairs, Kbp). In each of the chicken DNA tracks two bands (a) and (b) were detected.


A
B
C
D
Kbp
21·8
7·5
(5·5/5·9)
4·8
3·4
a >
b >

is illustrated in Figure V,7.

Figure V,8 shows the extent of β globin mRNA sequences cloned in the six β globin cDNA clones characterized, and also the extent of the derived sequences from each of these clones. With the exception of the 5' end non-homologous bases (Figure V,5), all derived sequences in analogous regions were identical.

C. <u>Chicken β globin gene</u>

A major potential use of cloned cDNA is as a probe for the gene sequence from which it was derived, the advantages over mRNA being the same as those for sequence analysis, the purity and quantity of the cloned cDNA.

Insert DNA was used as a template for randomly primed, radioactively labelled copy DNA by AMV reverse transcriptase using a variation of the method described by Taylor *et al.* (1976). This radioactively labelled probe was then hybridized to *Eco*RI digested chicken chromosomal DNA annealed to a nitrocellulose filter after electrophoresis on a 1% agarose gel (see II,2,O for methods). Bacteriophage λ DNA, *Eco*RI cut, and radioactively labelled by nick-translation (Maniatis *et al.*, 1975), a gift of Dr. P. Seeburg, was included in the agarose gel electrophoresis as molecular weight markers.

Figure V,9 shows two bands of *Eco*RI cut chicken genomal DNA hybridized to the β globin coding probe. The approximate molecular weights of these bands, by comparison with the molecular weight markers in kilobase pairs (Kbp) were 9.5 and 6. These findings were in agreement with later studies by Engel and Dodgson (1978) and Hughes *et al.* (1979) who suggested that the 6 Kbp band contains the adult

TABLE V,1.

CODON UTILIZATION OF CHICKEN β GLOBIN mRNA

	U		C		A		G		
U	Phe	3	Ser	–	Tyr	–	Cys	2	U
	Phe	5	Ser	5	Tyr	2	Cys	1	C
	Leu	–	Ser	–	Term	1	Term	–	A
	Leu	–	Ser	–	Term	–	Trp	4	G
C	Leu	1	Pro	1	His	3	Arg	–	U
	Leu	6	Pro	4	His	4	Arg	3	C
	Leu	–	Pro	–	GluN	1	Arg	–	A
	Leu	11	Pro	–	GluN	4	Arg	–	G
A	Ile	1	Thr	3	AspN	1	Ser	–	U
	Ile	6	Thr	4	AspN	6	Ser	2	C
	Ile	–	Thr	–	Lys	1	Arg	–	A
	Met	1	Thr	–	Lys	9	Arg	3	G
G	Val	–	Ala	4	Asp	1	Gly	1	U
	Val	5	Ala	11	Asp	5	Gly	4	C
	Val	–	Ala	–	Glu	4	Gly	–	A
	Val	7	Ala	1	Glu	3	Gly	3	G

The table shows the preference for codons ending in G or C, and the discrimination against those containing C-G.

β globin gene and the 9.5 Kbp fragment contains an embryonic β-like sequence. The most direct means of verifying this assignment would be by sequence analysis of cloned genomal DNA containing these fragments. This line of work was attempted using radioactively labelled pCG_{β}-3 as a probe for the gene in cloning experiments involving plasmid derivatives called cosmids (Collins and Bruning, 1978). This method of genomal DNA cloning had been successful in the isolation and characterization of the chicken ovalbumin gene region (Royal *et al.*, 1979). These experiments were unsuccessful due mainly to difficulties with *in vitro* packaging of the cosmid recombinant derivatives into bacteriophage λ heads (Sternberg *et al.*, 1977) and therefore will not be reported in this thesis.

3. Discussion

The nucleotide sequence of chicken β globin mRNA, as deduced from the sequence of cloned cDNA, has a relatively high GC content (57% compared with 41% for the genome, Sinclair and Brown, 1971). This is particularly evident in the redundant bases within the coding region in which 49% of codons are NNC and 30% NNG (Table V,1). This selection in the mRNA sequence for a high GC content (also found in growth hormone mRNA, Martial *et al.*, 1979 and chorionic somatomammotropin mRNA, Shine *et al.*, 1977) may result in a more stable overall secondary structure with a high degree of nuclease resistance. In addition, the stable secondary structure may be an essential feature of processing of pre-cursor mRNA. In one possible conformation of chicken β globin mRNA (data not shown) derived by comparison with that proposed by Philipp *et al.* (1978) for rabbit β globin mRNA,

FIGURE V,10.

THE 5' UNTRANSLATED REGION

Comparison of the 5' untranslated regions of chicken, rabbit (Baralle, 1977a) and human (Baralle, 1977b) β globin mRNA. Boxes show base changes between the three sequences. Sequences have been aligned to show maximum homology.

Bases are numbered from the first base of the coding region.

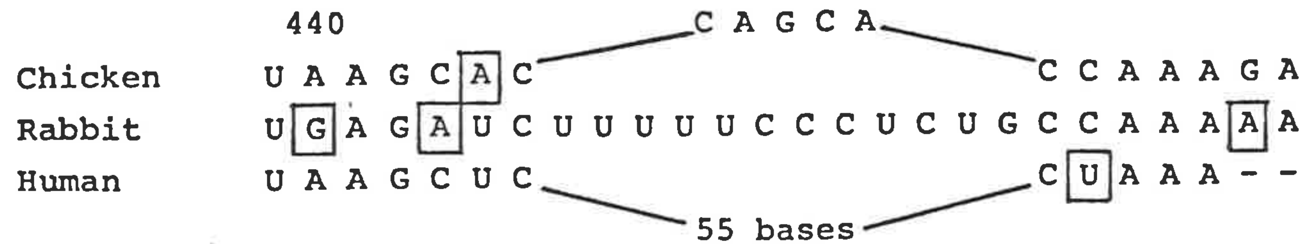
```
               -50                 -40                 -30
Rabbit         G C U U U U G A C A C A A C U G U G U U U - - - - - -
Human          G C U U C U G A C A C A A C U G U G U U C - - - - - -
Chicken        G C U - C A G A C - C U C C U C C G U A C C G A C A
                         -50                 -40

                           -20                 -10               -1
- - - A C U U G C A A U C C C C C A A A A C A G A C A G A A U G
- - - A C U A G C A A - - C C U C - A A A C A G A C A C C A U G
G C C A C A C G C U A - C C C U C - C A A C C G C C G C C A U G
-30                 -20                 -10                  -1
```

FIGURE V,11.

THE 3' UNTRANSLATED REGION

Homology in the 3' untranslated regions of chicken, rabbit (Proudfoot, 1977 and Efstratiadis *et al.*, 1977) and human (Proudfoot, 1977) β globin mRNA. Boxes show base changes between the three sequences. Sequences have been aligned to show maximum homology.

Bases are numbered from the first base of the coding region.



460 480

U C A C G G A G C A C C U A C A A C C A U U G C A U G C A C C U
U U A U G G G G - - - - - A C A U C - A U - G - A A G C C C C U
- C - U G G G G G - - - - A U A U U - A U - G - A A G G G C C U

500 520

- G C A G A A A U G C U C C G G A G C U G A C A G C U U G U G A
U G - A G C - A U - C - - - - - - - - U G A C - - - U U C U G G
U G - A G C - A U - C - - - - - - - - U G - G A - - U U C U G C

C A A A U A A A G U U C A U U C A G U G A C A C U - C poly(A)
C U A A U A A A G G A A A U U U A U U U U C A U U G C poly(A)
C U A A U A A A A A A C A U U U A U U U U C A U U G C poly(A)

many of the G and C residues in the third position of codons are involved in hydrogen bonding. Despite the high GC content there is a relatively low frequency of the C-G doublet, which is characteristic of eukaryote DNA (Russell *et al.*, 1976) (19 C-G compared with 43 G-C).

Kafatos *et al.* (1977) have carried out an extensive analysis of the homology between rabbit and human β globin mRNA sequences. Comparison of the chicken sequence with those derived from rabbit (Baralle, 1977a; Proudfoot, 1977; Efstratiadis *et al.*, 1977) and human (Proudfoot, 1977; Baralle, 1977b and Marotta *et al.*, 1977) reveals some interesting features. In pCG_{β}-3, of the 51 bases in the 5' untranslated region (AUG not included) there are at most 25 bases homologous with rabbit and 30 bases homologous with human (Figure V,10). No more than four contiguous bases are found to be homologous in any part of this region suggesting that if a ribosome binding site exists, analogous to that in prokaryotes (Shine and Dalgarno, 1975) then there is little selective pressure to maintain the specificity of this sequence. A similar conclusion has been previously arrived at by a comparison of 5' untranslated regions from a variety of mRNAs (Baralle and Brownlee, 1978). In the case of both human and rabbit β globin mRNA there are six more bases (not shown in Figure V,10) before the 7meG cap which suggests, as is expected from the method of double-stranded cDNA synthesis, that pCG_{β}-3 does not contain the complete 5'-untranslated region of chicken β globin mRNA.

Comparison of respective 3'-untranslated regions shows a similar degree of homology as described for the 5' ends (Figure V,11). Of the 108 bases in the chicken sequence,

57 are homologous with rabbit and 52 with human. For rabbit and human β globin mRNAs there is a region immediately after the termination codon of complete divergence and the same is true for the chicken sequence. This is followed by a region of homology, which contains 24 deletions (or insertions) and 10 base changes (chicken compared with rabbit), up to the highly conserved AAUAAA sequence (Proudfoot and Brownlee, 1976). From this sequence to the poly(A) tract there is only one deletion but 8 base changes (both in rabbit and human) suggesting that the length is more critical than the sequence.

Comparison of the nucleotide sequences of chicken and rabbit β chain coding regions (data not shown) shows 120 base changes. Of these, 66 are involved in an amino acid change while the remaining 54 conserve the protein sequence. The 72% nucleotide sequence homology between the chicken and rabbit globin coding regions is significantly higher than the homology between the non-coding regions (49% 5' end, 54% 3' end); a result consistent with selection at the amino acid level being a significant factor in the maintenance of nucleotide sequence.

CHAPTER VI

SEQUENCE ANALYSIS OF α GLOBIN CODING cDNA CLONES

CHAPTER VI - SEQUENCE ANALYSIS OF α GLOBIN CODING cDNA CLONES

1. Introduction

Work described in this chapter was initially aimed at the determination of the sequence of chicken alpha globin mRNA from cloned cDNA, for a comparative analysis with the mammalian alpha globin mRNA sequences. A comparison with the chicken beta globin mRNA sequence was also envisaged because Goodman *et al.* (1975), by analyzing globin amino acid sequences, had determined a genealogical tree for the coding region of globin genes which suggested that the chicken globin genes were relatively primitive and lay close to the separation of globin into alpha and beta chain types.

The sequence analysis of several alpha coding cDNA clones was undertaken when it was determined that none of the derived sequences contained a coding potential for either of the normal adult α globin chains (α_A, Matsuda *et al.*, 1971, or α_D, Takei *et al.*, 1975).

Restriction endonuclease digestion of cDNA, as described in chapter III, has been useful in determining the number and complexity of sequences present arising from a mixture of mRNA species (Seeburg *et al.*, 1977a). To show that the α and β sequences derived by cDNA clone analysis were the representatives of the polysomal mRNA, and therefore the phenotypic globins, a comparison was made of the restriction endonuclease digestion patterns predicted from individual α and β cDNA clones with those derived from the total cDNA.

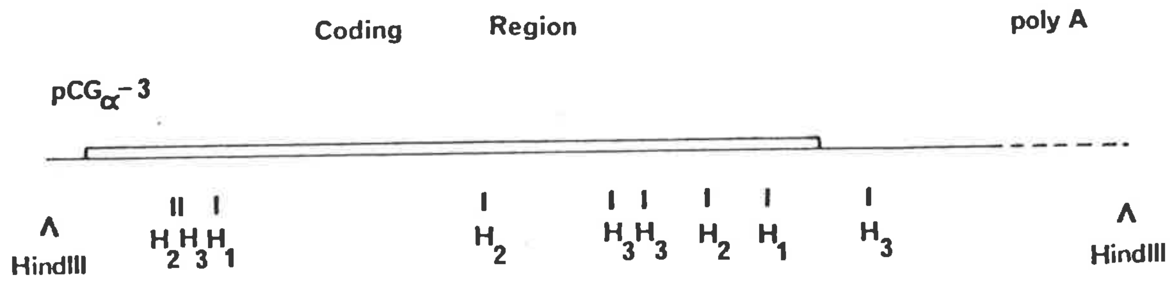
FIGURE VI,1.

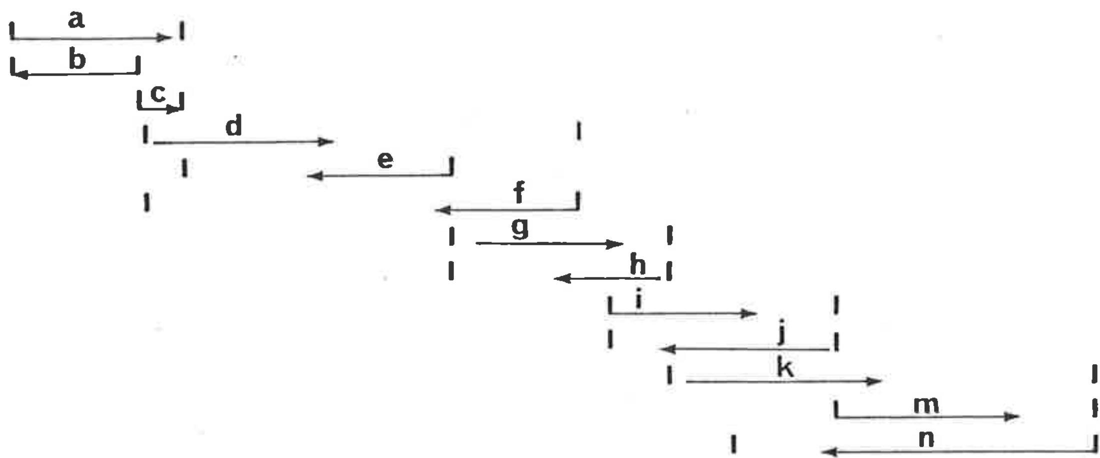
SEQUENCING STRATEGY FOR ALPHA CODING DNA SEQUENCES

Only those restriction sites used in the DNA sequence analysis are shown (H_1 - *Hha*I, H_2 - *Hpa*II, H_3 - *Hae*III). Arrows indicate the direction and extent of derived sequences. Λ denotes site of *Hin*dIII linker DNA attachment on termini of cloned DNA sequences and indicates the extent of the cloned sequence with respect to the mRNA. Restriction fragments were labelled for sequence analysis by incubation with T_4 polynucleotide kinase and γ-^{32}P-ATP. Strand separation in dimethyl sulphoxide followed by electrophoresis in non-denaturing acrylamide gels was carried out as described by Maxam and Gilbert (1979) and detailed in section II,2,N,i.

Sequences were derived from end-labelled fragments generated as follows:

Sequence		Generation of Labelled DNA Fragment
a,n	-	end-labelling pCG_α-3 insert/*Hha*I cleavage.
b	-	*Hpa*II cleavage pCG_α-3 plasmid/end-labelling/*Hin*dIII cleavage.
c,e	-	*Hpa*II cleavage pCG_α-3 insert/end-labelling/*Hha*I cleavage.
d,f,i,j,m	-	*Hae*III cleavage pCG_α-3 insert/end-labelling/strand-separation.
g,h,k	-	*Hpa*II cleavage pCG_α-3 insert/end-labelling/strand-separation.


Coding
Region
poly A
pCGα-3
HindIII
H 2
H 3
H 1
H2
H3
H3
H2
H1
H3
HindIII


a
b
c
d
e
f
g
h
i
j
k
m
n

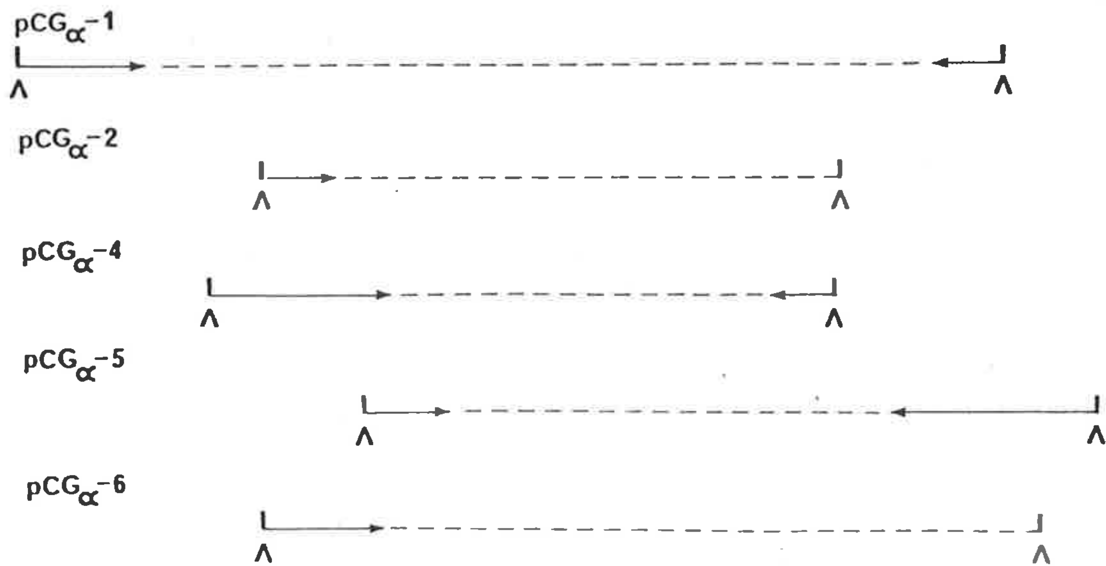

pCGα-1
pCGα-2
pCGα-4
pCGα-5
pCGα-6

FIGURE VI,2.

STRAND SEPARATION OF *Hae*III AND *Hpa*II END-LABELLED pCG$_{\alpha}$-3 FRAGMENTS

Use was made of strand separation techniques (Maxam and Gilbert, 1979) to generate DNA fragments, labelled at only one end, suitable for sequence analysis.

Insert DNA from pCG$_{\alpha}$-3 was digested with restriction endonuclease *Hae*III or *Hpa*II and end-labelled with T_4 polynucleotide kinase and γ-^{32}P-ATP. The labelled DNA was heated in 30% (v/v) dimethyl sulphoxide at 90°C for 2 min and the denatured strands separated on a 5% polyacrylamide gel with 50:1 acrylamide to bisacrylamide ratio (II,2,N,i). End-labelled DNA fragments were electroeluted from gel slices and subjected to sequence analysis.

Tract A. An aliquot (3%) of *Hpa*II cut, end-labelled pCG$_{\alpha}$-3 loaded, but not heated, in dimethyl sulphoxide.

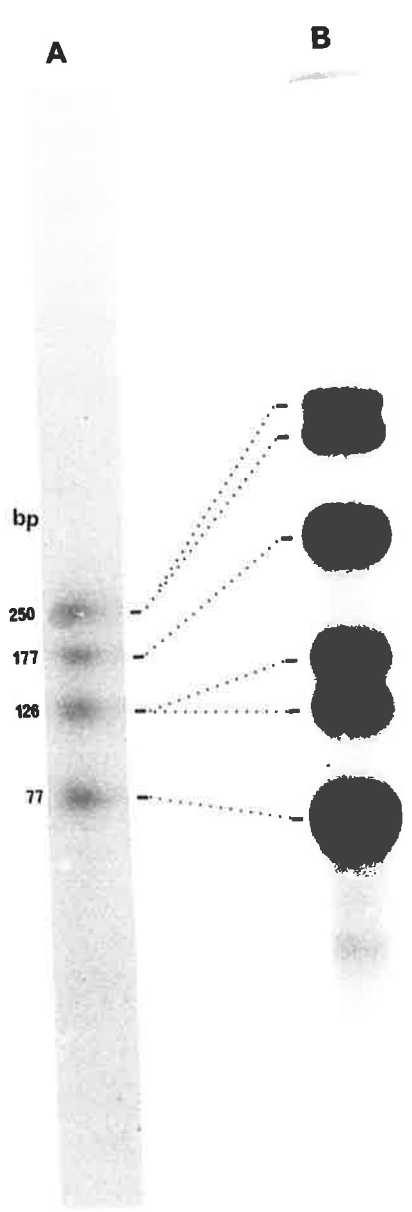
B. An aliquot (97%) of *Hpa*II cut, end-labelled pCG$_{\alpha}$-3 heated in dimethyl sulphoxide.

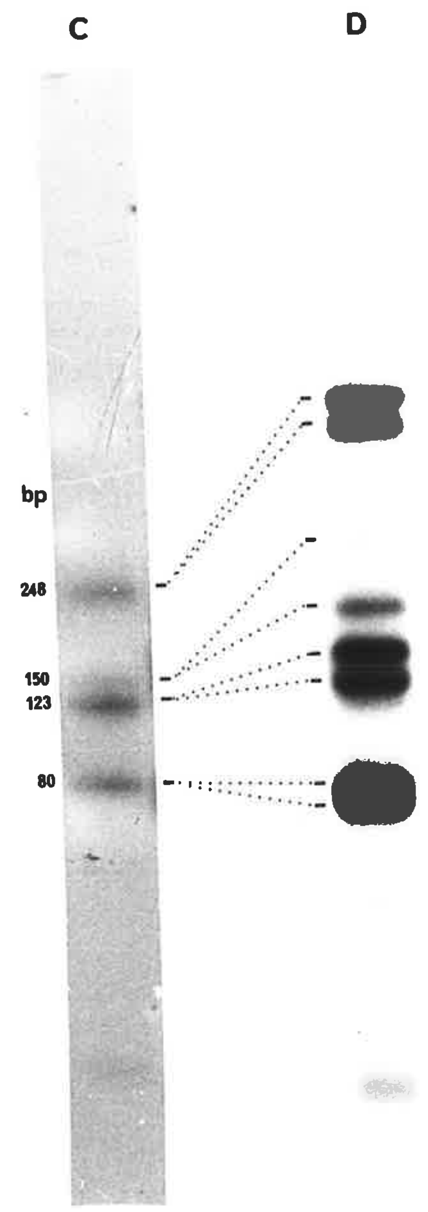
C. An aliquot (3%) of *Hae*III cut, end-labelled pCG$_{\alpha}$-3 loaded, but not heated, in dimethyl sulphoxide.

D. An aliquot (97%) of *Hae*III cut, end-labelled pCG$_{\alpha}$-3 heated in dimethyl sulphoxide.

Values indicate the size, in base pairs (bp) of double-stranded DNA fragments, and dotted lines indicate the separation of these fragments into the single-stranded fragments in tracks B and D.

The single-stranded fragments derived from the *Hpa*II digest 177 bp and 77 bp double-stranded fragments comigrated (see dotted lines) and therefore could not be sequenced. These sequences were determined as shown in Fig. VI,1.


A
B
bp
250
177
126
77


C
D
bp
248
150
123
80

2. Results

A. Sequence analysis of pCG_{α}-3

The longest alpha globin coding insert, pCG_{α}-3, (Figure IV,7, track E) which includes sequences from the 3' poly(A) tract to 15 bases on the 5' side of the AUG initiation codon, was completely sequenced by the strategy shown in Figure VI,1. End-labelled fragments for sequence analysis were derived mainly from *Hae*III or *Hpa*II digestion of isolated pCG_{α}-3 insert DNA, radioactive labelling with T_4 polynucleotide kinase and γ-^{32}P-ATP, strand separation in dimethyl sulphoxide and isolation by non-denaturing gel electrophoresis (II,2,N,i,a). This method was effective in all but one region of the sequence, due to the separated strands of the 177 base pair *Hpa*II fragment (from *ala* 18 to *ala* 77) co-electrophoresing on the strand separation gel (see Figure VI,2). The sequence to the 5' side of the *Hpa*II site at *ala* 77 was derived from a labelled fragment generated by *Hpa*II digestion of pCG_{α}-3 insert DNA, labelling with T_4 polynucleotide kinase and γ-^{32}P-ATP and digestion of the labelled DNA with restriction endonuclease *Hha*I. Sequencing reactions used were modified from those used in chapter V as described by Maxam and Gilbert (1979), Busslinger *et al.* (1979) and G. Peterson (pers. comm.) and are detailed in section II,2,N,ii. Where possible both strands were sequenced and where this proved difficult one strand was sequenced several times.

The use of strand separation to generate end-labelled DNA fragments was particularly advantageous in the sequence conformation as both strands of the one sequence were accessible from the one labelling reaction. In one

FIGURE VI,3.

COMPLETE NUCLEOTIDE SEQUENCE OF THE mRNA

CORRESPONDING TO pCG_{α}-3

The coding potential of the derived α_S mRNA is shown. The end sequence of pCG_{α}-5 is included to show 5' terminal heterogeneity generated during *E. coli* DNA polymerase I blunt ending reaction, as discussed in Chapter V. Vertical bars indicate non-homologous bases.

Triangles indicate amino acids which differ between α_S and α_A globin chains (see Figure VI,4).

pCG_{α}-3

5' start

A G C A C G G G U G C A A C C AUG

10 20

val leu ser ala ala asp lys asn asn val lys gly ile phe thr lys ile ala gly his

GUG CUG UCC GCU GCU GAC AAG AAC AAC GUC AAG GGC AUC UUC ACC AAA AUC GCC GGC CAU

30 40

ala glu glu tyr gly ala glu thr leu glu arg met phe thr thr tyr pro pro thr lys

GCU GAG GAG UAU GGC GCC GAG ACC CUG GAA AGG AUG UUC ACC ACC UAC CCC CCA ACC AAG

50 60

thr tyr phe pro his phe asp leu ser his gly ser ala gln ile lys gly his gly lys

ACC UAC UUC CCC CAC UUC GAU CUG UCA CAC GGC UCC GCU CAG AUC AAG GGG CAC GGC AAG

III II

pCG_{α}-5 C UCG AUC

70 80

lys val val ala ala leu ile glu ala ala asn his ile asp asp ile ala gly thr leu

AAG GUA GUG GCU GCC UUG AUC GAG GCU GCC AAC CAC AUU GAU GAC AUC GCC GGC ACC CUC

AAC CUA GUG GCU GCC UUG AUC GAG GCU GCC

90 100

ser lys leu ser asp leu his ala his lys leu arg val asp pro val asn phe lys leu

UCC AAG CUC AGC GAC CUC CAU GCC CAC AAG CUC CGC GUG GAC CCU GUC AAC UUC AAA CUC

110 120

leu gly gln cys phe leu val val val ala ile his his pro ala ala leu thr pro glu

CUG GGC CAA UGC UUC CUG GUG GUG GUG GCC AUC CAC CAC CCU GCU GCC CUG ACC CCG GAG

130 140

val his ala ser leu asp lys phe leu cys ala val gly thr val leu thr ala lys tyr

GUC CAU GCU UCC CUG GAC AAG UUC UUG UGC GCC GUG GGC ACU GUG CUG ACC GCC AAG UAC

arg stop

CGU UAA GACGGCACGGUGGCUAGAGCUGGGGCCAACCCAUCGCCAGCCCUCCGACAGCGAGCAGCCAAAUGAGAU

3'

GAAAUAAAAUCUGUUGCAUUUGUGCUCC poly(A)

FIGURE VI,4.

AMINO ACID DIFFERENCES BETWEEN α_S AND α_A

Coding potential of derived α_S mRNA is compared with the codon requirements of the α_A chain amino acids (Matsuda *et al.*, 1971) at those sites which differ between the two amino acid sequences. This comparison indicates between 32 and 35 (depending on codons utilized for *leu* and *ser*) required base changes in nucleotide sequence of the respective mRNA's. (|) indicates required base change, A - adenine, G - guanine, C - cytosine, U - uracil, Y - pyrimidine, P - purine, N - any of A, G, C or U.

Amino acid residue No.	4	34	35	36	38	63	64	66	67	68	70	71	73	77	79	109	110	111	113	116	118	120
α_S	*ala*	*thr*	*thr*	*tyr*	*pro*	*val*	*ala*	*leu*	*ile*	*glu*	*ala*	*asn*	*ile*	*ala*	*thr*	*val*	*ala*	*ile*	*his*	*ala*	*thr*	*glu*
Nucleic acid sequence α_S	GCU	ACC	ACC	UAC	CCA	GUG	GCU	UUG	AUC	GAG	GCC	AAC	AUU	GCC	ACC	GUG	GCC	AUC	CAC	GCC	ACC	GAG
Nucleic acid sequence α_A	AAY	AUY	GGN	UUY	ACN	GCN	CUN	AUY	ACN	AAY	AUY	GAP	GCN	UCN	GCN	CUN	GUN	GCN	CUN	GAP	GCN	AAP
		A					U Y	A			A			AGY		U Y			U Y			
α_A	*asn*	*ile*	*gly*	*phe*	*thr*	*ala*	*leu*	*ile*	*thr*	*asn*	*ile*	*glu*	*ala*	*ser*	*ala*	*leu*	*val*	*ala*	*leu*	*glu*	*ala*	*lys*

instance (not shown) this was particularly useful as hairpin compression in the sequencing gel made sequencing of both strands a necessity.

The modified reactions shortened the manipulation time and reduced levels of hydrazine cleavage at G residues (Busslinger *et al.*, 1979). The use of thinner sequencing gels (0.5 mm as opposed to 1.5 mm) increased the resolution of bands, and therefore the length of sequence read from one sequencing run.

Terminal sequences of all alpha globin coding clones were derived from fragments isolated after T_4 polynucleotide kinase, γ-^{32}P-ATP labelling of isolated insert DNA and subsequent cleavage with either *Hae*III or *Hha*I restriction endonuclease.

With the exception of 5 bases at the 5' terminus of pCG_α-5 (with respect to mRNA sequence) the sequences derived from all alpha globin coding clones (Figure VI,1) were identical to pCG_α-3.

The coding potential of pCG_α-3 (Figure VI,3) differs from that required by α_A (Matsuda *et al.*, 1971) by 22 amino acids (see Figure VI,4) and α_D (Takei *et al.*, 1975) by 61 amino acids. In all cases where the encoded amino acid sequence of pCG_α-3 differed from that of α_A the nucleotide sequence was verified by sequence analysis of both strands, or derivation of the relevant sequence in another alpha globin coding cDNA clone (Figure VI,1).

Cummings *et al.* (1978) reported the nucleotide sequence, from alpha globin coding chicken cDNA clones, between amino acids 60 to 83. This sequence was identical to that found in pCG_α-3 and strongly supported the hypothesis

TABLE VI,1.

ASSIGNMENT OF FRAGMENTS GENERATED FROM RESTRICTION ENDONUCLEASE DIGESTION OF CHICKEN GLOBIN DOUBLE-STRANDED cDNA TO SEQUENCES OF cDNA CLONES pCG_{α}-3 AND pCG_{β}-3

Restriction Endonuclease	Assignment of cDNA fragment	
	pCG_{α}-3	pCG_{β}-3
*Hae*III	248, 123	258, 69
*Hpa*II	177, 126	466
*Hha*I	315	+
*Ava*II	81	117, 105, 67
**Mbo*II	+	+
*Hinc*II	+	-
*Pst*I	-	+

Restriction fragments were assigned on the basis of their length (in base pairs), derived from relative mobility in acrylamide gel electrophoresis (see Figs. III,4 and III,5 and Table III,1). + denotes one cleavage, - denotes no cleavage. Values indicate the length of internal cleavage fragments in base pairs.

In addition, *Bgl*I, *Pvu*II, *Xba*I, *Hin*fI, *Sal*I, *Hpa*I, *Xma*I, *Kpn*I, *Eco*RI, *Bam*HI and *Hin*dIII sites were not found in cloned sequences, nor did these enzymes cleave cDNA.

**Mbo*II digestion generates a 210 base pair fragment (Fig. III, 4) which is unaccounted for and may therefore be due to a minor cDNA species.

that a new alpha globin coding mRNA was expressed in the red blood cells of chickens with phenylhydrazine induced hemolytic anaemia. The *Hae*III restriction map of pCG$_{\alpha}$-3, however, was at variance with that reported by Cummings *et al.* (1978). The assignment of a *Hae*III site at amino acids 109/110 in the mRNA sequence (which is the case in pCG$_{\alpha}$-3) was verified by *Hae*III restriction digest of the cDNA (Figure III,5, track A).

B. <u>Comparison of cloned sequence restriction maps with restriction endonuclease digests of ds cDNA</u>

Table VI,1 summarizes the results of the assignment of cDNA restriction digest fragments (from Chapter III) to the restriction maps of either pCG$_{\alpha}$-3 or pCG$_{\beta}$-3. Only one minor fragment of approximately 210 base pairs from *Mbo*II digestion could not be directly assigned and may have arisen from a minor alpha globin coding mRNA which was shown to be present in chicken globin mRNA (prepared as described in II,2,A) by *in vitro* translation studies (Knochel *et al.*, 1976 and A. Robins, pers. comm.). In addition, for those enzymes which did not cleave the cDNA, no site could be found in either pCG$_{\alpha}$-3 or pCG$_{\beta}$-3.

These results were consistent with the synthesis of a new major alpha globin chain in chickens as a result of anaemic stress, and this chain was therefore designated α_S.

C. <u>Minor α globin coding mRNA species?</u>

A search was made for a cloned cDNA sequence with a distinct restriction endonuclease map from either α_S or β globin mRNA in the hope of finding a sequence coding for the minor alpha globin mRNA species identified by *in vitro*

FIGURE VI,5.

SEQUENCES DERIVED FROM pCG_{α}-7

Insert DNA was isolated from pCG_{α}-7 and end-labelled with T_4 polynucleotide kinase and γ-^{32}P-ATP (II,2,N,i). After *Hae*III digestion and electrophoresis on a polyacrylamide gel, two fragments were isolated (A) and (B) and subjected to sequence analysis.

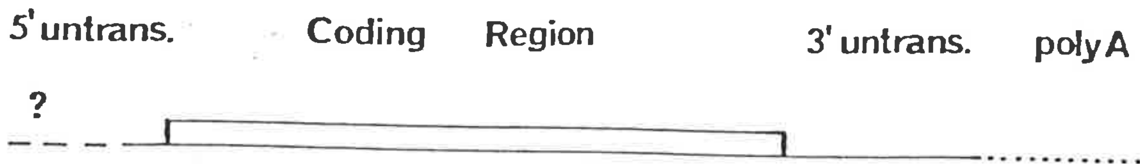
Figure A shows the location and direction of the derived sequences, with respect to α_S mRNA, and the location of synthetic linker (**Λ**).

Figure B shows the derived sequences from each fragment. The derived sequences agree exactly with those of pCG_{α}-3 in analogous regions except for 4 of the 5' terminal bases of pCG_{α}-3, with respect to the mRNA sequence, suggesting the inclusion of incorrect bases, as discussed in Chapter V. The relevant sequence of pCG_{α}-3 is included to indicate non-homologous bases. The bases derived from the AUG initiation codon are indicated by a solid box as are those from the ubiquitous AAUAAA hexanucleotide in the 3' untranslated region (Proudfoot and Brownlee, 1976). The location of the synthetic linker DNA sequence is indicated by solid boxes. N indicates an undetermined base.

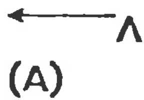
FIGURE VI,5.

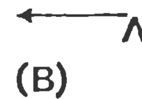
A

α_s mRNA



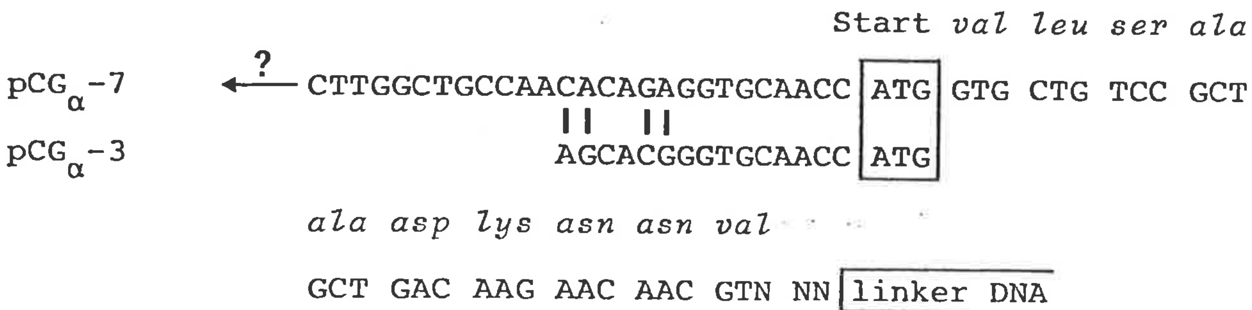
pCG_α-7





B

Fragment (A)



Fragment (B)

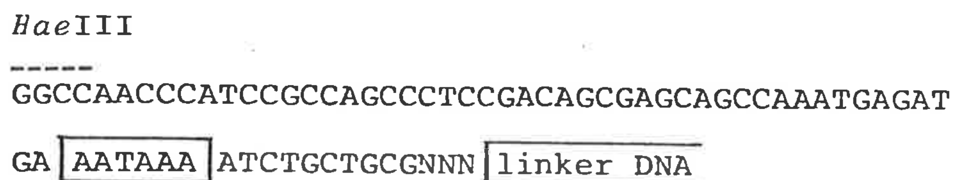


FIGURE VI,6.

SEQUENCE HOMOLOGY IN THE 3' UNTRANSLATED REGIONS OF CHICKEN α_S AND β GLOBIN mRNA

Sequences have been aligned to show maximum homology. Vertical bars indicate base changes, dots a deletion or insertion. Solid horizontal lines indicate the extent of palindrome sequences, dashed horizontal lines the bases involved in the palindromes. Triangles show the centre of symmetry of palindromes. Over the entire length of the 3' untranslated regions there are 21 base changes and 28 deletions or insertions resulting in a length difference of 8 bases (total lengths for α_S and β are 100 and 108 bases respectively, including the UAA termination codon). In both sequences 22 bases are involved in the palindrome, in the β sequence this is over a region of 28 bases, in the α_S sequence 33 bases. 11 of the 21 base changes are localized in the palindromic regions and 10 of these changes are involved in a change of composition of the respective palindromes.

```
       term.
β     UAAGCACCAGCACCAAAGAUCACGGAGCACCUA··CAA·CCAUUGC·AUG
α     UAAG·AC·GGCACGGUGGCU·A··GAGCUGGGGCCCAACCCAUCGCCA·G

CACCUGCAGAAAUGCUCCGGAGCUGACAGCUUGUGAC·AAAUAAA····GUU·C
C·CCUCC·GACA·GC····GAGCAGCCAAA·UGAGAUGAAAUAAAAUCUGUUGC

AUUCAGUGACACUC·-poly(A)
AUUU·GUG···CUCC-poly(A)
```

translation. One clone (later designated pCG_{α}-7) showed three additional bands on *Hae*III digestion of the hybrid plasmid to a similar digest of pBR322 DNA alone. One of these bands, therefore, represented an internal *Hae*III fragment of the pCG_{α}-7 insert (the other two containing pCG_{α}-7 and pBR322 sequences joined by *Hin*dIII linker DNA) and as none of the three bands corresponded to any of the internal *Hae*III fragments of pCG_{α}-3 or pCG_{β}-3 this clone was thought to represent a different mRNA sequence. Insert DNA was isolated from this recombinant plasmid end-labelled with T_4 polynucleotide kinase, cleaved with *Hae*III restriction endonuclease and the isolated labelled DNA fragments subjected to sequence analysis.

Both derived sequences corresponded to pCG_{α}-3, however the direction of the derived sequences indicated that the difference in *Hae*III restriction pattern was due to the insertion of two cDNA sequences into the one *Hin*dIII site of pBR322, without a synthetic *Hin*dIII linker DNA sequence separating them. This approach to finding alpha globin coding sequences, different from that of α_S mRNA, was therefore abandoned. This analysis did, however, extend the derived sequence at the 5' end of α_S mRNA and suggests that four bases in this region of pCG_{α}-3 are incorrect (Figure VI,5).

D. Comparison of sequences

Figure VI,6 shows sequence homology between the 3' untranslated sequences of chicken α_S and β globin mRNA. The unusual feature was the extent of homology of these regions for the different chains. Previously such homology had only been found between sequences of the same chain

type, such as rabbit and human β globin mRNA 3' untranslated regions (Proudfoot, 1977), whereas rabbit α and β globin mRNA show very little homology in this region (Proudfoot *et al.*, 1977). The retention of 3' untranslated sequences between chicken α_S and β globin mRNAs confirms the evolutionary relationship established for the globin genes in the coding region by amino acid sequence analysis (Goodman *et al.*, 1975).

Translation studies (Kronenburg *et al.*, 1979) have shown the rabbit beta globin mRNA 3' untranslated region to be non-essential for expression *in vitro*, however, studies on human globin gene variants suggest that alteration of 3' untranslated sequences markedly affects *in vivo* expression (Weatherall and Clegg, 1979). The functional role of this region therefore remains in doubt, except that a fundamental role in translation is unlikely.

A second feature to emerge from chicken globin mRNA studies was the conservation of a palindrome in the 3' untranslated region, despite extensive base changes between α_S and β globin genes. The data is shown in Figure VI,6. The conservation of site is remarkable in view of the fact that 11 of the 21 base changes between chicken α_S and β globin mRNA in this region are involved in a change of composition of the palindromes. The function of such a structure remains speculative, although the finding of Bogenhagen and Brown (cited in McKay, 1979), that control regions can be intragenic, raises the possibility that this region may be important to the expression of these genes.

Regulatory sequences have been proposed for the 3' end of sea urchin genes (Busslinger *et al.*, 1979) although

in this case the sequences involved have dyad symmetry. No analogous palindromic sequences to those found in chicken could be found in any of the mammalian globin gene sequences (Proudfoot, 1977; Proudfoot *et al.*, 1977; Poon *et al.*, 1978; Konkel *et al.*, 1979 and Nishioka and Leder, 1979).

3. Discussion

Translation *in vitro* of mRNA derived from the red blood cells of anaemic chickens and separation of the products on either the triton-polyacrylamide gel system of Borun *et al.* (1977) (A. Robins, pers. comm.) or carboxymethyl cellulose (Knochel *et al.*, 1976) suggests that two alpha and one beta globin chain coding mRNA species are present, similar to the normal peptide chain components of adult chicken hemoglobins A and D. The observations reported here indicate that hemolytic anaemia in chickens affects the expression of globin genes by inducing the synthesis of an alpha chain not normally seen in adult chicken hemoglobin, at the expense of expression of the normal major alpha globin species, α_A. No clones of the minor alpha component, with characteristics of α_D, have yet been found. This may be due to inefficient synthesis of α_D-like cDNA, as has been reported for rabbit globin mRNA (Efstratiadis *et al.*, 1976). However, the fact that equivalent numbers of α_S and β globin cDNA clones were obtained from randomly selected recombinants suggests a non-preferential reverse transcriptase reaction.

The molecular cloning and sequence analysis of the 210 base pair *Mbo*II cDNA fragment may reveal sequences of the minor alpha coding species, and would certainly identify the origin of this fragment.

TABLE VI,2.

CODON UTILIZATION OF CHICKEN α_S GLOBIN mRNA

Nucleotide Base of Codon									
1	2								3
	U		C		A		G		
U	Phe	—	Ser	—	Tyr	1	Cys	—	U
	Phe	7	Ser	4	Tyr	3	Cys	2	C
	Leu	—	Ser	1	Term	1	Term	—	A
	Leu	2	Ser	—	Term	—	Trp	—	G
C	Leu	—	Pro	2	His	3	Arg	1	U
	Leu	5	Pro	2	His	7	Arg	1	C
	Leu	—	Pro	1	GluN	1	Arg	—	A
	Leu	8	Pro	1	GluN	1	Arg	—	G
A	Ile	1	Thr	1	AspN	—	Ser	—	U
	Ile	6	Thr	9	AspN	4	Ser	1	C
	Ile	—	Thr	—	Lys	2	Arg	—	A
	Met	1	Thr	—	Lys	10	Arg	1	G
G	Val	—	Ala	8	Asp	2	Gly	—	U
	Val	3	Ala	10	Asp	5	Gly	8	C
	Val	1	Ala	—	Glu	1	Gly	—	A
	Val	8	Ala	—	Glu	5	Gly	1	G

A preference is indicated for codons ending in C or G (54% NNC, 27% NNG) as is the case for chicken β globin mRNA (Chapter V)(49% NNC, 30% NNG).

Sequence analysis of alpha globin coding cDNA clones revealed that the 5' terminal sequence of each clone (with respect to the mRNA sequence) was identical to the corresponding sequence in pCG$_\alpha$-3, with the exception of pCG$_\alpha$-5. The non-homologous region at the 5' end of pCG$_\alpha$-5 (see Figure VI,3) was proposed to be generated during the blunt-ending reaction with *E. coli* DNA polymerase I (see Chapter V) and its existence in α_S globin coding clones as well as β globin coding clones suggests that the phenomenon is not sequence specific.

Restriction cleavage analysis of double-stranded cDNA (chapter III) indicates that α_S and β globin mRNA are the major species represented in the cDNA, and confirms the fidelity of these sequences throughout cloning in *E. coli* χ1776.

The nucleotide sequence of α_S globin mRNA, as deduced from the sequence of pCG$_\alpha$-3, shows distinct similarity to other eukaryote mRNA sequences. The predominant features being a high GC content (59%) as exemplified by codon utilization (Table VI,2, 54% NNC, 27% NNG), the infrequent occurrence of the C-G doublet (23 C-G compared with 44 G-C) and the ubiquitous hexanucleotide AAUAAA in the 3' untranslated region (Proudfoot and Brownlee, 1976). All these features are common to the β globin mRNA sequence (chapter V).

The coding potential of pCG$_\alpha$-3 shows 22 amino acid changes from the α_A sequence reported by Matsuda *et al.* (1971). This requires between 32 and 35 nucleotide changes depending on serine and leucine codon utilization (Figure VI,4). In view of the strong codon selection of eukaryotes and this small required difference in nucleotide sequence,

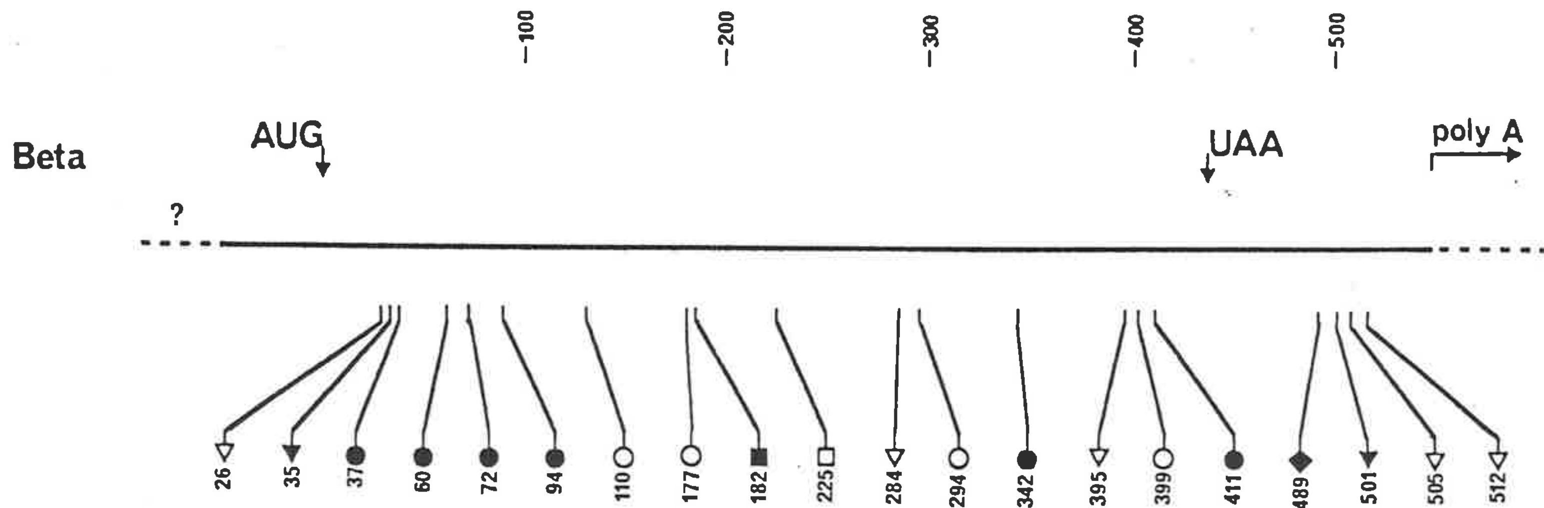
FIGURE VI,7.

MAPS OF THE SITES OF RESTRICTION ENDONUCLEASES IN THE DEDUCED SEQUENCES OF CHICKEN α_S AND β GLOBIN mRNA

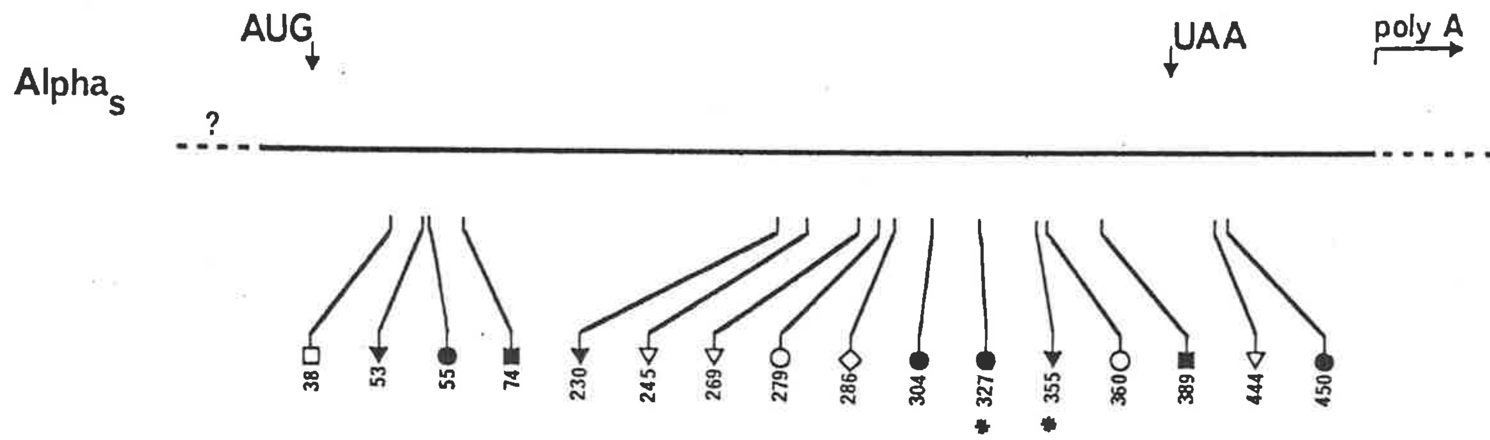
Bases are numbered from the first base of the coding region. The sites of the initiation (AUG) and termination (UAA) codons are included to indicate the extremities of the coding region.

*Hpa*II	– ▼
*Hae*III	– ●
*Hha*I	– ■
*Alu*I	– ▽
*Hind*II	– ◇
*Ava*II	– ○
*Mbo*II	– □
*Pst*I	– ◆

Numbering of restriction endonuclease sites indicates the location of the first base (with respect to the 5' end of mRNA) of the recognition site of the specified restriction endonuclease. Asterisks indicate the two restriction endonuclease sites which cannot be present in the α_A globin mRNA sequence due to essential nucleotide differences between the codons for the amino acids at positions 110 and 120 (see Figure VI,4).



-100
-200
-300
-400
-500
Beta
AUG
UAA
poly A
?
26
35
37
60
72
94
110
177
182
225
284
294
342
395
399
411
489
501
505
512



Alpha$_s$
AUG
UAA
poly A
?
38
53
55
74
230
245
269
279
286
304
327
355
360
389
444
450

it seems likely that a cDNA probe from α_S globin mRNA would cross-react with α_A globin gene sequences. This factor will have to be taken into account in the mapping of chromosomal gene sequences and analysis of genomal DNA clones, as has already been reported (Engel and Dodgson, 1978 and Hughes *et al.*, 1979).

The α_A globin mRNA restriction digest map should differ from that of α_S globin mRNA (Figure VI,7), however, as at least two, possibly four, of the required nucleotide differences form part of restriction endonuclease recognition sites in the α_S globin mRNA sequence. This difference, and the assignment of all major cDNA restriction fragments to either α_S or β globin mRNA sequences (Table VI,1), makes the presence of α_A globin cDNA sequences in total cDNA restriction digests unlikely, except possibly as a minor species. Results presented in this chapter are consistent with expression of the α_S globin gene replacing that of the α_A globin gene in phenylhydrazine induced anaemic chickens.

CHAPTER VII

FINAL DISCUSSION

CHAPTER VII - FINAL DISCUSSION

1. Introduction

The main aim of this chapter is to summarize the major conclusions of the work described in this thesis and relate them to current relevant research.

2. Hemoglobin Switching in Chickens

Adult chicken hemoglobin is composed of two types, A and D, which differ in their α chain component. Consequently, mRNA derived from adult cells active in globin synthesis would be expected to contain sequences of coding potential for α_A, α_D and β globin. To increase the level of cells actively synthesising hemoglobin, hemolytic anaemia was induced in 12 week old chickens by the injection of phenylhydrazine. Hemolytic anaemia in chickens causes an increase in erythropoiesis and release into the bloodstream of immature erythroid cells, mainly mid and late polychromatic erythrocytes (Attardi *et al.*, 1966; Kabat and Attardi, 1967). In mRNA derived from these cells, sequences coding for normal adult β globin chains were found but not for either α_A or α_D chains.

The induction of hemolytic anaemia in chickens by phenylhydrazine appears to have induced the expression of an unusual α globin, α_S. This phenomenon may be related to induced "hemoglobin switching" seen in sheep and goats. Control of globin gene expression is not a simply defined phenomenon, and switching of synthesis from one type to another can involve molecular and cellular events (see section I,2,D,iii).

Peptide or nucleotide sequence analysis of chicken

embryonic globin chains or mRNA, respectively, would provide definite evidence of whether or not α_S is an embryonic globin chain, although peptide maps of the embryonic α-like chains (Brown and Ingram, 1974) suggest that it is not.

Switching of hemoglobin types is characterized by changes in peptide chain components. In humans this mainly involves the β-like chains whose gene locus is transcriptionally ordered in the same sense as temporal expression (Weatherall and Clegg, 1979 and Dahl and Flavell, pers. comm.). In chickens, switching involves mainly α-like chain components of which there are at least four (possibly six, Brown and Ingram, 1974 and this thesis). The location and arrangement of these genes in a locus similar to that of the β globin genes in man would reinforce the hypothesis that differentiation in erythropoiesis involves first the "switching on" of regions of the genome (including α and β globin loci) and then the selective and temporal expression of chain types from each locus. The finding that chicken red blood cells released prematurely into the bloodstream synthesize an atypical adult hemoglobin would be consistent with this hypothesis if the α_S gene lies upstream from the α_A gene.

Hybridization data (Dodgson *et al.*, 1979) suggest that the genes in the chicken β-like globin locus are not transcriptionally and developmentally coincident, and therefore this general hypothesis may be incorrect. Similar studies have found related "pseudo-gene" sequences in the vicinity of *Xenopus laevis* 5S ribosomal genes (Jacq *et al.*, 1977) and the chicken ovalbumin gene (Royal *et al.*, 1979), therefore sequence analysis of chicken globin mRNAs and genes is

necessary to unequivocally determine the structure of the gene loci.

The analysis of embryonic globin mRNA through cDNA cloning will indicate the temporal expression of genes in either the α or β locus during development and thereby complement the cellular studies of chicken erythropoiesis (Chapman and Tobin, 1979; Keane *et al.*, 1979) as well as establish whether the α_S sequence described here represents a normal embryonic α-like globin chain.

3. Chicken Globin mRNA

From analysis of *in vitro* constructed recombinant DNA containing sequences derived from chicken globin mRNA, the nucleotide sequences of the major gene transcripts expressed in the red blood cells of chickens with induced hemolytic anaemia were derived.

To ensure that the globin mRNA sequences investigated were representative of the major polysomal mRNA species present in anaemic chickens, a comparison of restriction endonuclease patterns of double-stranded cDNA and of individual α_S and β globin coding cDNA clones was made. This verified that these two sequences were the major mRNA species present.

Comparative analyses were undertaken with the analogous sequences of different species in an attempt to assign functional significance (in terms of control of gene expression) of certain sequences by their selective conservation between the species.

A. 5' Untranslated region

By the nature of their construction, the chicken globin

cDNA clones and the sequences derived from them do not represent a complete copy of the mRNA sequence. Extensive sequences of the β globin mRNA 5' untranslated region (51 bases) were derived from pCG_{β}-3 and compared to the analogous sequences of rabbit and human β globin. No more than four contigous bases were found to be homologous in this region and therefore if a functional site does exist (e.g., for ribosome binding) then there is little selective pressure to maintain the specificity of this sequence.

B. Coding region

Both chicken globin mRNA sequences have high GC content (α_S 59%, β 57%) compared with the genome (41%, Sinclair and Brown, 1971). This selection is most evident in the redundant bases within the coding region (α_S 54% NNC, 27% NNG and β 49% NNC, 30% NNG). In contrast to the high GC content there is an infrequent occurrence of the C-G doublet (α_S 23 C-G c.f. 44 G-C, β 19 C-G c.f. 43 G-C).

Although these features are not seen in all eukaryote mRNA sequences (e.g., chicken ovalbumin mRNA has a relatively high AU content, McReynolds *et al.*, 1978) they are common to most sequences studied to date.

C. 3' Untranslated region

Only limited homology was found on comparison of the 3' untranslated regions of analogous globin mRNA sequences of mammals with those of chicken (e.g., chicken and rabbit β globin, 54%, chicken α_S and rabbit α globin 69%). Comparison of chicken α_S and β globin mRNA sequences in this region revealed a degree of homology (66% see Figure VI,6) not previously seen in sequences from different chain types.

Retention of these sequences in chicken may reflect the evolutionary relationship established for globin genes by amino acid sequence analysis (Goodman *et al.*, 1975).

Further support to the assignment of an ancestral position for the chicken genes is seen in the degree of homology between chicken β globin and human γ globin 3' untranslated regions (Poon *et al.*, 1978) of about 60%, compared with 45% maximum homology between human γ globin and human β globin sequences in this region (data not shown).

More striking than the overall homology between the α_S and β sequences in this region is the conservation of a 22 base palindrome in both mRNA types in the same relative position in the sequence, despite base changes (see Figure VI,6). The role of such a sequence can only be speculative, but may involve transcriptional controls or binding of specific proteins concerned with mRNA stability (Weatherall and Clegg, 1979).

4. Further Studies

While the results reported in this thesis are of interest in their own right, significant exploitation of the products of this research is envisaged.

As a continuation of research into the relationship between gene sequences and their selective expression, three major approaches are evident.

(1) The molecular cloning of regions of the chicken genome encompassing the α and β globin gene loci for direct analysis of the genes and adjacent sequences. The use of pure cloned cDNA probes and the information derived from them, such as restriction endonuclease maps, described in this thesis

should facilitate such an analysis.

(2) The molecular cloning and sequence analysis of embryonic globin cDNA. Information derived from this analysis should identify the temporal expression of genes within the α and β globin gene loci and determine whether the α_S globin chain is normally expressed during the development of the chicken.

(3) The use of cloned α_S and β globin sequences as an affinity column (DNA bound to cellulose, Aviv and Leder, 1972) for the purification of non-globin chicken erythroid mRNA sequences (affinity column run-through fraction) which should include those coding for histone H_5 and high mobility group protein, HMG-E. The comparative analysis of the mRNA and gene sequences of these tissue specific proteins with those of the globins would be invaluable in the assessment of convergent evolution of control sequences concerned with the selective expression of these genes.

REFERENCES

Adams, J.M. and Cory, S. (1975) *Nature 255*, 28-33.

Appels, R. and Wells, J.R.E. (1972) *J. Mol. Biol. 70*, 425-434.

Attardi, G., Parnas, H., Hwang, M-I.H. and Attardi, B. (1966) *J. Mol. Biol. 20*, 145-151.

Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA 69*, 1408-1412.

Beck, E., Sommer, R., Auerswald, E.A., Kurz, C., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1978) *Nucleic Acids Res. 5*, 4495-4503.

Baralle, F. (1977a) *Cell, 10*, 549-558.

Baralle, F. (1977b) *Cell, 12*, 1085-1095.

Baralle, F. and Brownlee, G.G. (1978) *Nature 274*, 84-87.

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

Borun, T.W., Ajiro, K., Zweidler, A., Dolby, T.W. and Stephens, R.E. (1977) *J. Biol. Chem. 252*, 173-180.

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

Brawerman, G. (1976) in *Progress in Nucleic Acid Research and Molecular Biology, volume 17*, 117-148.

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

Brown, D.D. and Blackler, A.W. (1972) *J. Mol. Biol. 63*, 75-83.

Brown, J.L. and Ingram, V.M. (1974) *J. Biol. Chem. 249*, 3960-3972.

Bruns, G.A.P. and Ingram, V.M. (1973) *Phil. Trans. Roy. Soc. B. 266*, 225-305.

Busslinger, M., Portmann, R. and Birnsteil, M. (1979) *Nucleic Acids Res. 6*, 2997-3008.

Calos, M.P. (1978) *Nature 274*, 762-765.

Caplan, A.I. and Ordahl, C.P. (1978) *Science 201*, 120-130.

Chapman, B.S. and Tobin, A.J. (1979) *Develop. Biol. 69*, 375-387.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry 18*, 5294-5299.

Clewell, D.B. (1972) *J. Bacteriol. 110*, 667-676.

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) *Nature 282*, 567-574.

Collins, J. and Bruning, H.J. (1978) *Gene 4*, 85-107.

Cummings, I.W., Lin, A.Y. and Salser, A.W. (1978) *Nature 276*, 418-419.

Dawkins, R. *The Selfish Gene* (Oxford University Press, 1976).

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

Dick, C. and Johns, E.W. (1969) *Biochim. Biophys. Acta 175*, 414-418.

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

Donelson, J.E. and Wu, R. (1972) *J. Biol. Chem. 247*, 4654-4660.

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

Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. (1976) *Cell, 7*, 279-288.

Efstratiadis, A., Kafatos, F.C. and Maniatis, T. (1977) *Cell, 10*, 571-585.

Engel, J.D. and Dodgson, J.B. (1978) *J. Biol. Chem. 253*, 8239-8246.

Forget, B.G., Housman, D., Benz, E.J. and McCaffrey, R.P. (1975) *Proc. Natl. Acad. Sci. USA 72*, 984-988.

Furuichi, Y., Morgan, M., Shatkin, A.J. Jelinek, W., Salditt-Georgieff, M. and Darnell, J.E. (1975) *Proc. Natl. Acad. Sci. USA 72*, 1904-1908.

Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) *Nature 278*, 428-434.

Garel, A. and Axel, R. (1976) *Proc. Natl. Acad. Sci. USA 73*, 3966-3970.

Glynn, I.M. and Chappell, J.B. (1964) *Biochem. J. 90*, 147-149.

Goetz, G., Esmailzadeh, A.K. and Huang, P.C. (1978) *Biochim. Biophys. Acta 517*, 236-245.

Goodman, M., Moore, W.G. and Matsuda, G. (1975) *Nature 253*, 603-608.

Graf, T., Ade, N. and Beug, H. (1978) *Nature 275*, 496-501.

Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) *Eur. J. Biochem. 36*, 32-38.

Guerry, P., Le Blanc, D.J. and Falkow, S. (1973) *J. Bacterial. 116*, 1064-1066.

Gurdon, J.B. and Woodland, H.R. (1970) *Current Topics in Developmental Biology 5*, 39-70.

Horiuchi, K. and Zinder, N.D. (1975) *Proc. Natl. Acad. Sci. USA 72*, 2555-2558.

Howk, R.S., Anisowicz, A., Silverman, A.Y., Parks, W.P. and Scolnick, E.M. (1975) *Cell 4*, 321-327.

Hozumi, N. and Tonegawa, S. (1976) *Proc. Natl. Acad. Sci. USA 73*, 3628-3632.

Hughes, S.H., Stubblefield, E., Payvar, F., Engel, J.D., Dodgson, J.B., Spector, D., Cordell, B., Schimke, R.T. and Varmus, H.E. (1979) *Proc. Natl. Acad. Sci. USA 76*, 1348-1352.

Hunt, L.T. and Dayhoff, M.O. (1976) *Atlas of Protein Sequence and Structure* (M.O. Dayhoff, Ed.) *5(2)*, 191-223.

Ingram, V.M. *Hemoglobin in Genetics & Evolution* (Columbia University Press, 1963).

Itzhaki, R.F., Hell, A. and Birnie, G.D. (1978) *Nucleic Acids Res. 5*, 739-750.

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

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

Kabat, D. and Attardi, G. (1967) *Biochim. Biophys. Acta. 138*, 382-399.

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

Keane, R.W., Lindblad, P.C., Pierik, L.T. and Ingram, V.M. (1979) *Cell 17*, 801-811.

Keller, E.B. and Calvo, J.M. (1979) *Proc. Natl. Acad. Sci. USA 76*, 6186-6190.

Kendrew, J.C., Watson, H.C., Stradberg, B.E., Dickerson, R.E., Phillips, D.C. and Shore, V.C. (1961) *Nature 190*, 666-670.

Knochel, W., Lange, D. and Hendrick, D. (1976) *Mol. Biol. Reports 3*, 143-150.

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

Kornerg, R.D. (1974) *Science 184*, 868-871.

Kreig, P. and Wells, J.R.E. (1976) *Biochemistry 15*, 4549-4558.

Kronenberg, H.M., Roberts, B.E. and Efstratiadis, A. (1979) *Nucleic Acids Res. 6*, 153-166.

Lehninger, A.L. In *Biochemistry* (Worth Publishers, Inc., 1970) pp. 558-559.

Leibovitch, S.A. and Harel, J. (1978) *Nucleic Acids Res. 5*, 777-787.

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

Levy, S., Sures, I. and Kedes, L.H. (1979) *Nature 279*, 737-739.

Lewin, B. (1975) *Cell 4*, 11-20 and 77-93.

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

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) *Proc. Natl. Acad. Sci. USA 72*, 1184-1188.

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

Marotta, C.A., Wilson, J.T., Forget, B.G. and Weissman, S.M. (1977) *J. Biol. Chem. 252*, 5040-5053.

Martial, J.A., Hallewell, R.A., Baxter, J.D. and Goodman, H.M. (1979) *Science 205*, 602-606.

Matsuda, G., Takei, H., Wu, K.C. and Shiozawa, T. (1971) *Int. J. Protein Res. 3*, 173-174.

Matsuda, G., Maita, T., Mizuno, K. and Ota, H. (1973) *Nature 244*, 244.

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

Maxam, A. and Gilbert, W. (1979) *Methods in Enzymology*, in press.

McKay, R. (1979) *Nature 282*, 556-557.

McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Divol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) *Nature 273*, 723-728.

Molgaard, H.V., Perucho, M. and Ruiz-Carrillo, A. (1980) *Nature 283*, 502-504.

Nakanashi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N. and Numa, S. (1979) *Nature 278*, 423-427.

Neelin, J.N. and Butler, G.C. (1961) *Can. J. Biochem. Physiol. 39*, 485-491.

Neinhuis, A.W. and Stamatoyannopoulos, G. (1978) *Cell 15*, 307-315.

Nishioka, Y. and Leder, P. (1979) *Cell 18*, 875-882.

O'Farrell, P.Z., Cordell, B., Valenzuela, P., Rutter, W.J. and Goodman, H.M. (1978) *Nature 274*, 438-445.

Ogata, R.T. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA 75*, 5851-5854.

Olins, A.L. and Olins, D.E. (1974) *Science 183*, 330-334.

O'Malley, B.W. and Means, A.R. (1974) *Science 183*, 610-620.

Paterson, B.M. and Rosenburg, M. (1979) *Nature 279*, 692-696.

Peacock, A.C. and Dingham, C.W. (1967) *Biochemistry 6*, 1818-1827.

Pemberton, R.E., Housman, D., Lodish, H.F. and Baglioni, C. (1972) *Nature 235*, 99-102.

Perutz, M.F. and Ten Eyck, L.F. (1972) *Cold Spring Harb. Symp. quant. Biol. 36*, 295-310.

Philipp, M., Ballinger, D. and Seliger, H. (1978) *Naturwissenschaften 65*, 388-389.

Poon, R., Kan, Y.W. and Boyer, H.W. (1978) *Nucleic Acids Res. 5*, 4625-4630.

Preobrazhensky, A.A. and Spirin, A.A. (1978) in *Progress in Nucleic Acids Research and Molecular Biology, volume 21*, 1-38.

Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA 72*, 784-788.

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

Proudfoot, N.J. (1977) *Cell 10*, 559-570.

Proudfoot, N.J., Gillam, S., Smith, M. and Longley, J.I. (1977) *Cell 11*, 807-818.

Rodriguez, R.L., Tait, R., Shine, J., Bolivar, F., Heynecker, H., Betlach, M. and Boyer, H.W. (1977) *Molecular Cloning of Recombinant DNA*, 73-84, Academic Press, Inc., New York.

Rosenburg, M. and Paterson, B.M. (1979) *Nature 279*, 696-701.

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

Russell, G.J., Walker, P.M.B., Elton, R.A. and Subak-Sharpe, J.H. (1976) *J. Mol. Biol. 108*, 1-15.

Scott, A.C. (1975) *Ph.D. Thesis*, University of Adelaide.

Scott, A.C. and Wells, J.R.E. (1976) *Nature 259*, 635-638.

Seeburg, P.H., Shine, J., Martial, J.A., Ullrich, A., Baxter, J.D. and Goodman, H.M. (1977a) *Cell 12*, 157-165.

Seeburg, P.H., Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977b) *Nature 270*, 486-494.

Schroeder, H.W., Liarakos, C.D., Gupta, R.C., Randerath, K. and O'Malley, B.W. (1979) *Biochemistry 18*, 5798-5808.

Seidman, M.M. and Cole, R.D. (1977) *J. Biol. Chem. 252*, 2630-2639.

Shine, J. and Dalgarno, L. (1975) *Nature 254*, 34-38.

Shine, J., Seeburg, P.H., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) *Nature 270*, 494-499.

Shooter, K.V., Goodwin, G.H., and Johns, E.W. (1974) *Eur. J. Biochem. 47*, 263-270.

Sinclair, J.H. and Brown, D.D. (1971) *Biochemistry 10*, 2761-2769.

Sippel, A. (1973) *Eur. J. Biochem. 37*, 31-40.

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

Stark, B.C., Kole, R., Bowman, E.J. and Altman, S. (1978) *Proc. Natl. Acad. Sci. USA 75*, 3717-3721.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977) *Gene 1*, 255-280.

Sterner, R., Boffa, L.C. and Vidali, G. (1978) *J. Biol. Chem. 253*, 3830-3836.

Sutcliffe, G. (1978) *Nucleic Acids Res. 5*, 2721-2728.

Tahara, S.M., Traugh, J.A., Sharp, S.B., Lundak, T.S., Safer, B. and Merrick, W.C. (1978) *Proc. Natl. Acad. Sci. USA 75*, 789-793.

Takei, H., Ota, Y., Wu, K.C., Kiyohara, T. and Matsuda, G. (1975) *J. Biochem. 77*, 1345-1347.

Taylor, J.M., Illemensee, R. and Summers, S. (1976) *Biochim. Biophys. Acta 442*, 324-330.

Therwath, A. and Scherrer, K. (1978) *Proc. Natl. Acad. Sci. USA 75*, 3776-3780.

Tilghman, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P. and Weissman, C. (1978) *Proc. Natl. Acad. Sci. USA 75*, 1309-1313.

Tsai, S.Y., Tsai, M-J., Lin, C-T. and O'Malley, B.W. (1979) *Biochemistry 18*, 5726-5731.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977) *Science 196*, 1313-1319.

Verma, I.M., Temple, G.F., Fan, H. and Baltimore, D. (1972) *Nature New Biol. 235*, 163-167.

Vidali, G., Boffa, L.C. and Allfrey, V.G. (1977) *Cell 12*, 409-415.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA 75*, 3727-3731.

Watson, D.C., Peters, E.H. and Dixon, G.H. (1977) *Eur. J. Biochem. 74*, 53-60.

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

Wellauer, P.K. and Dawid, I.B. (1974) *J. Mol. Biol. 89*, 379-395.

Weintraub, H. and Groudine, M. (1976) *Science 193*, 848-856.

Williams, A.F. (1970) *Ph.D. Thesis*, University of Adelaide.

Wilson, J.T., de Riel, J.K., Forget, B.G., Marotta, C.A. and Weissman, S.M. (1977) *Nucleic Acids Res. 4*, 2353-2367.

Young, N.S., Benz, E.J., Kantor, J.A., Kretschmer, P. and Neinhuis, A.W. (1978) *Proc. Natl. Acad. Sci. USA 75*, 5884-5888.

APPENDIX - PUBLICATIONS

Work described in this thesis has been presented in the following papers:

1. <u>Papers for Publication</u>

Molecular Cloning and Sequence Analysis of Adult Chicken β Globin cDNA (with J. Shine, A. Ullrich, J.R.E. Wells and H.M. Goodman) (1979) *Nucleic Acids Res.* *7*, 1137-1146.

Chicken Globin Genes: Nucleotide Sequence of cDNA Clones Coding for the Alpha Globin Expressed During Hemolytic Anaemia (with J.R.E. Wells) submitted to *J. Biol. Chem.*

2. <u>Papers Presented at Meetings</u>

Molecular Cloning and Sequence Analysis of cDNA Transcribed from Chicken Globin mRNA (with A. Ullrich, J. Shine, W.J. Rutter, H.M. Goodman and J.R.E. Wells) (1978) *Proc. Aust. Biochem. Soc.* *11*, 81.

Sequence Analysis of Chicken Globin cDNA Clones (1980) *Proc. Aust. Biochem. Soc.* *13*, 85.