GLOBIN GENE MAPPING IN THE MARSUPIAL,

DASYURUS VIVERRINUS.

by

Brandon John Wainwright, B.Sc.(Hons.)

Department of Genetics,
University of Adelaide.

A thesis submitted to the University of Adelaide, South Australia,

for the degree of Doctor of Philosophy in March 1984.

THE NATIVE CAT, *DASYURUS VIVERRINUS*.

To my parents -

for their support, tolerance, and belief in education.

SUMMARY

This thesis contains the results of a gene mapping study carried out on an Australian marsupial, the native cat, *Dasyurus viverrinus*. The investigations presented in this thesis were aimed at determining the chromosomal location of the globin genes in the native cat, and these studies fell broadly into two categories; (1) construction and characterisation of nucleic acid probes encoding the major α- and β-globin polypeptides of the native cat, and (2) the use of such probes to map the globin sequences by *in situ* hybridisation to fixed metaphase chromosomes, and by hybridisation to DNA extracted from metaphase chromosomes purified by fluorescence-activated cell sorting.

Total RNA from native cat reticulocytes was fractionated by affinity chromatography and sucrose gradient centrifugation. A 10S RNA fraction was found to direct the synthesis of native cat globin polypeptides in a wheat germ cell free system. A cDNA library was constructed from total reticulocyte mRNA and putative globin cDNA clones were identified on the basis of hybridisation studies using a 10S cDNA probe. Recombinant plasmids were subjected to restriction enzyme analysis and clones chosen for further study. Hybrid-arrest translation experiments identified putative α- and β-globin clones. Nucleotide sequencing identified two recombinants, pDG 73 and pDG 5 as encoding the native cat α- and β-globin polypeptides, respectively. Plasmid pDG 73 was found to contain the entire α-globin coding and 3' non-coding regions, as well as 10 bases of 5' non-coding, whilst pDG 5 contained the entire β-globin 3' non-coding region, but only 60% of the coding region. Plasmids pDG 73 and pDG 5 contain the first

marsupial coding sequences to be cloned and characterised.

The amino acid sequence of the entire native cat α-globin polypeptide, as well as the partial amino acid sequence of the β-globin polypeptide were deduced from the nucleotide sequence data. Using the native cat globin amino acid sequences and the previously published amino acid sequences of eutherian and grey kangaroo globins, the date for the marsupial/eutherian divergence was calculated to be 155 million years ago, based upon a simple "molecular clock" model. This figure differs from a previously published divergence date estimate of 130 million years ago calculated by Air *et al.* (1971) based upon data from globin polypeptides and using an identical "molecular clock" model to that used in this study. The estimate by Air *et al.* (1971), however, was based on data from a smaller number of species than was used in this study, and data from only one marsupial species, the grey kangaroo, was included.

This thesis also examines the evolutionary relationship between macropod, dasyurid and didelphid marsupials using data from globin polypeptide amino acid sequences and contrasts this to relationships derived from immunological data.

The per cent corrected nucleotide sequence divergence between native cat α- and β-globin coding regions and the globin coding regions of eutherian globin genes was calculated for replacement site and silent site substitutions. The inclusion of the native cat data, the first from a marsupial, as well as the inclusion of the recently published data from the amphibian *X. laevis*, enabled a more accurate assessment of the evolutionary rate of globin gene replacement site substitution than previously published by Perler *et al.* (1980). The addition of new data from the native cat and *X. laevis* served to

highlights the variability in the nature of the relationship between silent site substitutions and evolutionary time, and confirmed the possibility that there are two rates of silent site substitution, dependent on the amojnt of divergence time between two species.

In situ hybridisation experiments using tritiated native cat globin probes localised the α- and β-globin sequences to separate chromosomes; α-globin to chromosome number 2 and β-globin to chromosome number 4. This result is consistent with the asyntenic chromosomal arrangement of the α- and β-globin genes in all other mammals studied so far. This is the second report of the specific localisation of any marsupial autosomal gene, other than those coding for rRNA.

The chromosomes of the native cat (2n=14) were fractionated into six groups of high purity by fluorescence-activated cell sorting. Only chromosomes numbers 2 and 3 could not be resolved on the basis of absolute DNA contents. The fluorescence distribution of the native cat chromosomes was shown to provide an accurate estimate of the relative DNA content and frequency of each chromosome type. DNA was extracted from purified metaphase chromosomes and probed with native cat α- and β-globin sequences in "Southern blot" experiments. The localisation of all β-globin sequences to chromosome number 4 was confirmed, and the α-globin probe was found to hybridise to the sorted fraction containing chromosomes numbers 2 and 3, a result which did not contradict the results obtained by *in situ* hybridisation.

DECLARATION

Except where due reference is given, this thesis contains no material which has been accepted for the award of any degree or diploma in any University and, to the best of my knowledge and belief, contains no material published or written by another person. I consent to this thesis being made available for photocopying and loan.

B.J. Wainwright

March 1984.

ACKNOWLEDGEMENTS

I am grateful to my supervisor Dr. R.M. Hope for his interest,enthusiasm and friendship throughout the course of this project. I am indebted to Prof. G.E. Rogers of the Department of Biochemistry and all the members of "Keratin Korna" for their provision of facilities and advice in the initial stages of the cloning experiments.Thanks are also due to many other people who provided help in various ways; to Dr. D.L. Hayman for his interest and valuable discussion; to Mrs. L. Billett for her tuition in cell culture proceedures and for her assistance in many aspects of the preparation of this thesis; to Messrs. C. Chesson, R. Goodwins, and D. Pottrell for their assistance; to Dr.G. Munro for her advice and assistance with computing proceedures; and to Mrs. C. Leach for statistical advice.

My task was made much easier by the happy atmosphere, and sometimes mania, created by my fellow postgraduate students,Cathy Shanahan, David Jones, Peter Sharp,and Gail Skidmore. I am grateful for the assistance given me by Ms. G. Psaltis, who also skilfully typed this thesis.

Finally , I should like to thank Therese ,whose idea of a night out did not really include reading a sequencing gel, and whose constant support was a source of great comfort.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AMV	Avian myeoblastosis virus
bp	base pair
BPB	Bromophenol Blue
BCIG	5-Bromo-4-chloro-3'-indolyl-β-galactoside
cDNA	synthetic DNA complementary to messenger RNA
Ci	Curie
cpm	counts per minute
CsCl	Cesium chloride
d	dalton
DNA	Deoxyribonucleic acid
ds cDNA	double-stranded cDNA
dNTP	Deoxynucleoside triphosphate
dpm	disintegrations per minute
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
Hepes	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
oligo-dT	oligodeoxythymidylic acid
PEG	Polyethylene glycol
PIPES	Piperazine-N-N'-bis[2-ethanesulfonic acid
pfu	plaque forming units
polyU	polyuridylic acid
POPOP	1,4,-Bis(5-phenyl-2-oxazolyl)-benzene
PPO	2,5,-Diphenyloxazole
RNA	Ribonucleic acid

S	Svedberg units
SDS	Sodium dodecyl sulphate
Temed	N,N,N',N'-Tetramethylethylenediamine
TCA	Trichloroacetic acid
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA

CHAPTER 1. INTRODUCTION

The number of genes which have been assigned to specific mammalian chromosomes has increased rapidly over the last decade, largely as a result of somatic cell genetic techniques. Not only has the number of gene assignments in humans increased from approximately 210 in 1975 to approximately 450 in 1981 (HGM-3, 1975; HGM-6, 1981*), but there has been interest in the mapping of species other than man, in particular mouse (550 genes assigned as of HGM-6, 1981), rat (44), rabbit (40), and in the non-human primates (75). Smaller numbers of genes have been mapped to the chromosomes of other eutherian mammals such as sheep, cattle, and pigs (HGM-6, 1981).

There are a number of reasons for interest in the chromosomal arrangement of genes in species other than man. Information obtained from other species may be useful for the investigation of certain human medical and genetical problems by providing a suitable animal model, and a knowledge of gene organisation on the chromosomes may aid in an understanding of the way in which genes are regulated.

Perhaps the primary interest in comparative gene mapping to date has focused on the evolutionary conservation of gene arrangement. Extensive homologies have been demonstrated between the chromosome maps of man and mouse (Lalley *et al.*, 1978), as well as between groups of loci in a large number of mammalian species (HGM-6, 1981). The apparent conservation of the chromosomal arrangement of some genes between mammalian species has raised questions regarding the restrictions imposed by gene arrangement on karyotypic evolution. Does the conservation of gene arrangement between species result from selective differences due to the arrangement of certain genes, or does

---

* HGM-6 - Proceedings of the 6th Annual Human Gene Mapping Workshop.

it largely reflect a lack of time for chromosomal evolution to break up linkage groups?

Most gene mapping studies have concentrated on eutherian mammals, and in particular, on man and mouse. A few genes have been mapped in metatherian mammals using somatic cell hybrids (Graves *et al.*, 1979; Donald and Hope, 1981; Donald and Adams, 1981) and pedigree analysis (see Cooper *et al.*, 1971). Marsupials are thought to have diverged from eutherian mammals about 1.3×10^8 years b.p. (Air *et al.*, 1971). The opportunity for changes in gene arrangement through karyotypic evolution has been much greater between marsupial and placental mammals than it has been between more recently diverged species, such as man and mouse. Because of the greater opportunity for karyotypic evolution to alter gene arrangement between marsupial and placental mammals compared to gene arrangement within eutherians, it is of interest to examine the chromosomal arrangement of genes in marsupials, especially those genes which show conservation of arrangement in all eutherians studied.

A large proportion of genes that have been assigned to mammalian chromosomes code for enzyme markers (HGM-6, 1981). The chromosomal assignment of a gene coding for an enzyme using somatic cell hybrids requires that the gene product be expressed in the hybrids, that a suitable assay for that enzyme exists, and that the two parental forms of the enzyme can be distinguished. Using recombinant DNA techniques, the chromosomal localisation of any sequence for which a suitable probe exists can be determined, without the need for gene expression. This can be accomplished using cell hybrids (d'Eustachio and Ruddle, 1983), *in situ* hybridisation to fixed metaphase chromosomes (Henderson, 1982), or molecular analysis of purified metaphase chromosomes (Padgett *et al.*, 1977; Lebo *et al.*, 1982).

Globin genes constitute perhaps the most intensively studied eukaryotic multigene family. Vertebrate haemoglobin molecules are composed of an association of four globin polypeptides, two α-like and two β-like chains. The globins are of interest because they represent an example of a co-ordinately expressed, developmentally regulated multigene family (see Dayhoff, 1975; Weatherall and Clegg, 1979). Recent molecular analysis of vertebrate globin genes has revealed many details of their organisation (see Jeffreys, 1981). The common evolutionary origin of vertebrate globins is demonstrated by their amino acid homology, and by the similar structure of their genes (see Jeffreys, 1981). Homologies between the α- and β-globin genes indicate that they arose by duplication of a primordial globin gene about 5.7×10^8 years ago. The α- and β-globin genes have been found to be closely syntenic in the amphibian *Xenopus laevis* (Jeffreys *et al.*, 1980), but on separate chromosomes in the chicken and in all eutherian mammals studied to date (HGM-6, 1981).

The apparent conservation of the syntenic relationship of the α- and β-globin genes in eutherian mammals raised the question as to whether a similar arrangement exists in marsupials.

Marsupials of the family Dasyuridae display a large range of size and morphology (Rofe, 1979). They have, however, a highly conserved 2n=14 karyotype (Young *et al.*, 1982), the chromosomes of all species having an identical G-banding pattern, or being related by simple inversions or reciprocal interchanges (Rofe, 1979). No gene loci have been assigned to specific chromosomes in dasyurid marsupials.

The main objectives of the research presented in this thesis were:-

1) To construct and characterise marsupial α- and β-globin cDNA clones.

2) To use such clones to examine the evolution of marsupial globin

protein and globin mRNA.

3) To determine the chromosomal location of the α- and β-globin genes in dasyurid marsupials.

The karyotypic conservation between members of the Dasyuridae makes them of particular interest for gene mapping studies in marsupials. An examination of the chromosomal arrangement of genes, or groups of genes, between species, would enable a relationship between G-banding pattern and gene arrangement to be determined. Dasyurids also possess the advantages of low diploid chromosome number, and large variation in chromosome size and/or morphology within each species, which makes them particularly suitable for cytological studies and the application of chromosome fractionation techniques. The species *Dasyurus viverrinus*, (native cat, see frontispiece), was chosen for this study because such animals represented a good compromise between availability and blood volume. The cloning of polyA$^+$ reticulocyte RNA requires a volume of blood larger than that which could be obtained from the available small species, such as *Sminthopsis crassicaudata*. The larger tiger cats (*Dasyurus maculatus*) and Tasmanian devils (*Sarcophilus harrisii*) could not be obtained.

Dasyurid x rodent somatic cell hybrids have been constructed, but none of these contained identifiable marsupial chromosomes, thus preventing their use in gene mapping studies (Wainwright *et al.*, in preparation). In this project, the following approaches were used to map the dasyurid α- and β-globin sequences. Cloned native cat α- and β-globin cDNA sequences were used as probes for *in situ* hybridisation to fixed metaphase chromosomes. To confirm the results obtained by the *in situ* experiments, the metaphase chromosomes of the native cat were purified using a fluorescence-activated cell sorter

(FACS) and the DNA from the purified chromosomes subsequently probed with the native cat α- and β-globin probes. The data derived from these approaches were used to determine the chromosomal assignment of native cat α- and β-globin gene sequences.

CHAPTER 2. LITERATURE REVIEW

2.1 GENE MAPPING

A gene map shows the positions of genes, relative to one another, or to cytologically identifiable features of the chromosome, such as centromeres or G-bands. When discussing gene mapping, it is important to distinguish between genetic linkage (as displayed on a linkage map) and physical linkage (as displayed on a chromosome map). The detection and estimation of genetic linkage requires the phenotypic scoring of parents and offspring from informative matings, in families where segregation is occurring at two or more loci. Two genes are said to be linked if the recombination fraction between them is significantly less than 0.5. Genes which are on the same chromosome are said to be syntenic (Renwick, 1971). A linkage map and a chromosome map are presenting different types of information about the relationships between gene loci, and it should be realised that the terms "linkage map" and "chromosome map" cannot be used interchangeably.

The generation of chromosome maps can be accomplished using interspecific somatic cell hybrids. The first cell hybrids were made by the fusion of mouse and rat cells by Ephrussi and Weiss (1965). It was subsequently shown that genes from both parental species were being expressed in the hybrid cells and that the hybrid karyotype was less stable than the karyotype of either parental cell. Weiss and Ephrussi (1966) then showed that in rat x mouse hybrid cells there was non-random segregation of chromosomes, with the rat chromosomes being preferentially eliminated. It was this phenomenon of the preferential elimination of the chromosomes of one species in an interspecific hybrid that opened up an entirely new approach to gene mapping in mammals.

By the use of an appropriate selective system (see Chu and Powell, 1976) hybrid clones can be isolated which have differentially eliminated chromosomes of one of the parental species, but retain the chromosome bearing the selective marker or markers, in addition to an assortment of non-selected chromosomes. The analysis of such hybrid clones enables the presence/absence of a specific gene product to be correlated with the presence/absence of a specific chromosome, thus facilitating a chromosomal assignment for that gene. The most efficient assignment of genes can be achieved by the use of a "clone panel" of hybrids, in which each chromosome of the species under investigation has a unique pattern of presence and absence amongst the member clones of the panel (Creagan and Ruddle, 1975). Some genes are difficult to map using somatic cell hybrids. When using somatic cell hybrids for gene mapping it is assumed that no undetectable chromosome rearrangement occurs within the hybrid, that the homologous gene products of both parental species can be distinguished, that all of the genes in the species one wishes to map are expressed constitutively, and that the chromosomes of both parental species are cytologically distinct. Since these assumptions do not always hold (Ruddle, 1970), the data generated by the use of interspecies cell hybrids must be subjected to careful scrutiny. By the use of recombinant DNA techniques somatic cell hybrids can now be analysed for the presence of specific gene sequences without the need for the expression of those sequences in the hybrid cells (Owerbach *et al.*, 1980). In addition to this, it is unlikely that the restriction maps shown by the parental gene sequences will be indistinguishable using a number of different restriction endonucleases with different base specificities (Jeffreys *et al.*, 1979). Despite the advantages of a molecular approach to the analysis of somatic cell hybrids, the possible cytogenetic instability

of such hybrids will still require that any gene assignment be confirmed, ideally by the use of an independent method of gene mapping.

A more direct approach to gene mapping involves the use of *in situ* hybridisation of radioactively labelled sequences to fixed metaphase chromosomes. The technique of *in situ* hybridisation has the advantages that localisations can be performed on diploid cells, and that gene expression is unnecessary for chromosomal localisation. The main disadvantage of gene mapping by *in situ* hybridisation is the problem of cross-hybridisation of probe with other sequences in the genome, and the subsequent inability to distinguish the signal expected from a specific site from that of a cross-reacting site, without first performing *in vitro* DNA hybridisation kinetics studies (Henderson, 1982). The recent technical advances associated with *in situ* hybridisation studies are reviewed in section 7.1.

In several cases, gene assignments have been achieved using molecular hybridisation to purified metaphase chromosomes (Hughes *et al.*, 1979; Lebo *et al.*, 1979, 1982; Collard *et al.*, 1983). This approach to gene mapping has the advantage that a restriction endonuclease analysis of DNA purified from specific metaphase chromosomes can be performed, and hence is not subject to the ambiguities of hybridisation specificity of *in situ* hybridisation nor, to a lesser extent, the cytological uncertainty of somatic cell hybrids. The two major disadvantages associated with chromosome fractionation methods are, 1) the lack of resolution of most fractionation methods for many karyotypes; and 2) the large amount of time and the costs involved when using high resolution methods such as flow cytometry. Flow cytometry and sorting, and gene mapping by hybridisation to purified metaphase chromosomes are further reviewed in section 8.1.

2.2 COMPARATIVE GENE MAPPING

2.2.1 The Mammalian X Chromosome

The mammalian X chromosome represents a special case of the conservation of gene content. Based upon the apparent conservation of size of the mammalian X chromosome, and a small amount of data on the X-chromosome map of several species, Ohno (1967) proposed that any gene which was X-linked in one mammalian species would be X-linked in all others. The number of known mammalian X-linked loci has grown considerably since, and no known exceptions to "Ohno's rule" have beem confirmed, with the possible exception of the gene for steroid sulphatase (Pearson and Roderick, 1978; HGM-7, 1983).

The apparent evolutionary conservation of the mammalian X chromosome is thought to be a consequence of the evolution of a dosage compensation system for X-linked genes (Ohno, 1973). Ohno (1967) proposed that the mammalian X and Y chromosomes evolved from a pair of autosomes and the differentiation of one of those homologues into the Y chromosome was accompanied by a loss of genetic material from that chromosome. Subsequently, a mechanism for the inactivation of one of the X chromosomes in the homogametic sex evolved to compensate for the disparity in dosage between it and the heterogametic sex for X-linked genes. Regardless of the origin of X-chromosome inactivation, any fragmentation of the X chromosome would disturb this dosage compensation mechanism and would be subject to strong natural selection. This hypothesis led Ohno (1973) to speculate that any genes present on the X chromosome at the time of the evolution of dosage compensation would since have remained on the X chromosome as the result of a "frozen accident.

Whilst Ohno's hypothesis advocates the conservation of the genetic content of the mammalian X chromosome, it does not imply that the order

of the genes on the X is conserved. The gene loci *GLA*, *GPD*, *HPT*, and *PGK* have been shown to be X-linked in many species, but are not always found in the same order along that chromosome (HGM-7, 1983).

The various models of X-inactivation and aspects of X chromosome conservation have been discussed in detail in a number of reviews (see Lyon, 1972, 1974; Cattenach, 1975; Gartler and Andina, 1976; Graves, 1983).

2.2.2 Comparative Mapping of Autosomes

As gene mapping data have accumulated, it has become increasingly apparent that the arrangement of certain autosomal genes has also been conserved during mammalian evolution. Most gene mapping work has been conducted on man and mouse, both being relatively straightforward to map using somatic cell hybrids, and hence conservation of syntenic groups is more obvious between these species than between any others. For example, a large region of mouse chromosome 4 and the short arm of human chromosome 1 appear to be highly conserved (Lalley *et al.*, 1978a). These chromosomes show an identical arrangement of the genes *ENO-1*, *PGD*, *PGM-2* and *AK-2*. This gene arrangement appears to have been conserved over a period of 80 million years, the estimated time of divergence of man and mouse (Lalley *et al.*, 1978b). Over 60 homologous genes have been mapped in both man and mouse, and appreciable synteny homology is observed, with over half of those loci showing arrangements common to both species (Pearson *et al.*, 1982).

Nash and O'Brien (1982) have shown that regions corresponding to 20% by length of the human genome show gene content and possible G-banding homology to the domestic cat, based upon a chromosome map of 33 loci of the domestic cat. It is possible that a more extensive map of the cat genome will result in the demonstration of even greater homology.

Work by Motizot *et al.* (HGM-7, 1983) has shown that the synteny of *PGM-2* and *MPI* has been conserved from fishes through to mammals. Syntenic autosomal homologies which appear to be conserved across all mammalian species studied to date include the loci *PGP* and *ENO*; *PGM-3* and *MOD1*; *GAPD* and *TPI*; *TK* and *GALK*; and *SOD1* and *HRC* (HGM-7, 1983).

That quite striking conservation of autosomal gene arrangement has occurred over millions of years of evolution is now beyond doubt. There is, however, considerable debate over the significance of such conservation. Fisher (1930) was the first to propose that there may be a selective advantage to certain combinations of alleles in a double heterozygote, and hence that selection could act to reduce or increase recombination between interacting genes. The idea that there is a selective advantage associated with certain gene arrangements was supported by Bodmer and Parsons (1962) and Bodmer (1975). It was argued by Bodmer (1975) that chromosomal rearrangements are fixed sufficiently often during karyotypic evolution for any extensive synteny homologies between widely divergent species to be disrupted, unless they result in the conferral of a selective advantage. Bodmer's argument, however, is based only upon data derived from the rate of chromosomal evolution through pericentric inversions in primates. Bush *et al.* (1977) and Bengtsson (1980) have found that the rate of chromosomal evolution in primates is approximately between 2 and 2,000 times faster than the karyotypic change observed in all other groups of mammals apart from horses (0.5X) and lagomorphs (1.3X). Clearly, an argument based upon rates of mammalian karyotypic evolution should not consider primates alone.

Whilst functionally related genes in bacteria tend to be closely associated on the chromosome, gene regulation in higher organisms appears much more complex. Location of functionally related genes

in close proximity to one another in vertebrates appears not to be necessary for their co-ordinate regulation (Hughes *et al.*, 1979), although instances have been found of close linkage between genes involved in a common metabolic pathway (Spandidos and Siminovitch, 1977). Indeed, Harris (1978) has noted several examples of genes subject to co-ordinate regulation such as β_2 microglobulin and the HLA region, LDH-A and LDH-B, and α and β globin, which are on separate chromosomes in all eutherian mammals so far studied, and argues that there may be a selective advantage associated with such separate gene location, which leads to its preservation.

Ohno (1973) takes a different point of view to that of Bodmer and Harris, and argues that the main types of chromosomal change associated with vertebrate karyotypic evolution, Robertsonian fissions and fusions, do not alter linkage arrangements within a chromosome arm, leading to a conservation of synteny for large blocks of autosomal genes.

There are many examples of autosomal genes which do not show conservation of their synteny relationship and this is obviously compatible with the hypotheses of both Bodmer (1975) and Ohno (1973). For the merits of these two views to be assessed, a considerable body of data will have to be assembled on gene arrangement homologies between species. The earlier the evolutionary divergence of species exhibiting conservation of gene arrangement, the less likely is the conservation due to effects associated with a lack of opportunity for chromosomal evolution, as suggested by Ohno (1973).

2.3 THE GLOBIN GENE SYSTEM

2.3.1 Haemoglobin - Structure and Function

Vertebrates have evolved two mechanisms for supplying their

cells with adequate oxygen to support aerobic life:- a circulatory system to deliver oxygen to the cells of an organism and oxygen carrying molecules which overcome the low solubility of oxygen in water.

Lower vertebrates, such as the lamprey (Dayhoff, 1975) use primitive, monomeric haemoglobin-like oxygen carriers. Oxygen binds reversibly to iron atoms found at the centre of prosthetic haem groups which form an integral part of these molecules while their protein component provides a suitable environment for the haem group to reversibly carry oxygen (Kendrew *et al.*, 1961). During evolution, duplication of a primitive globin gene and variation in the two new genes resulted in the production of two different chains, α and β. The mutual adaptation of these two chains permitted tetramers to form from two α chains and two β chains. These four polypeptide chains are held together by non-covalent interactions and each chain contains one haem group.

The tetrameric form is normal for all mammalian haemoglobins and it has three important advantages over monomeric oxygen carrying molecules. Firstly, the oxygen-binding curve of tetrameric haemoglobin is sigmoidal in nature, which means that the binding of oxygen to one haem group facilitates the binding of oxygen to other haems on the same molecule. Secondly, H^+ and CO_2 promote the release of O_2 from haemoglobin (the Bohr effect), and this is physiologically important in enhancing the release of O_2 in metabolically active tissues. Conversely, O_2 promotes the release of H^+ and CO_2 in the lungs. Thirdly, the affinity of haemoglobin for O_2 is further regulated by 2,3-diphosphoglycerate (DPG), a molecule which lowers the oxygen affinity of haemoglobin. Foetal haemoglobins have a higher O_2 affinity than do adult haemoglobins because they bind less DPG (see Vander *et al.*,

1976).

A comparison of the amino acid sequences of globin polypeptides from many species shows that certain amino acids are invariant (Dayhoff, 1975). This group of conserved amino acids can be assigned to various functional roles within the haemoglobin molecule including haem binding, peptide chain contact, Bohr effect and DPG binding (Perutz and Ten Eyck, 1972; Goodman *et al.*, 1975).

2.3.2 Arrangement of Human Globin Genes

The α-like and β-like subunits of human haemoglobin are encoded by a small family of genes whose individual members are expressed at different times during development. Structural characterisation of globin polypeptides and mRNAs and extensive clinical investigations of inherited disorders in haemoglobin expression make the human globin genes a model system for studying the molecular basis of human genetic disease, and the molecular genetics of eukaryotic gene regulation.

Humans have five different β-like globins, δ, ε, $^{G}\gamma$, $^{A}\gamma$, and β and two α-like globins, ζ and α (Weatherall and Clegg, 1979). The earliest embryonic haemoglobin, Hb Gower 1, consists of ε and ζ polypeptide chains. At approximately 8 weeks of gestation, the embryonic chains are gradually replaced by the adult α globin chain and two different β like chains $^{G}\gamma$ and $^{A}\gamma$. The γ chains differ only from one another in the presence of glycine or alanine at position 136. During the transition period between embryonic and foetal development, Hb Gower 2 ($\alpha_2\varepsilon_2$) and Hb Portland ($\zeta_2\gamma_2$) are detected. HbF ($\alpha_2\gamma_2$) becomes the predominant species throughout the remainder of foetal life. Beginning just before birth, the γ chains are gradually replaced by the adult β and δ globin polypeptide chains. At 6 months after birth, 97-98% of the haemoglobin in HbA ($\alpha_2\beta_2$) while HbA_2 ($\alpha_2\zeta_2$) account for

about 2%. Small amounts of foetal haemoglobin can be found in adult peripheral blood (Bunn *et al.*, 1977).

Human globins are coded for by two unlinked clusters of genes, with the α-like globins being situated on human chromosome 16 (Deisseroth *et al.*, 1977) and the β-like cluster situated on chromosome 11 (Deisseroth *et al.*, 1978).

The arrangement and structure of the globin genes within each cluster has been determined by genomic blotting experiments and analysis of recombinant bacteriophage containing globin genomal genes (see Maniatis, 1980). The α-globin gene cluster consists of three functional genes, an embryonic globin gene (ζ) and two adult/foetal α-globin genes ($\alpha_1\alpha_2$) arranged in the order 5'-ζ-α_2-α_1-3' (Orkin, 1978; Lauer *et al.*, 1980; Proudfoot *et al.*, 1982). All genes are orientated in the same direction of transcription and are arranged along the chromosome in the order in which they are expressed. The β-globin gene cluster consists of five functional genes, one embryonic (ε), two foetal ($^{A}\gamma$ $^{G}\gamma$), a minor adult gene (δ) and a major adult gene (β), arranged along the chromosome in the direction of 5'-ε-$^{G}\gamma$-$^{A}\gamma$-δ-β-3', with the arrangement reflecting the order in which they are expressed during human ontogeny (Fritsch *et al.*, 1980; Little *et al.*, 1979; Baralle and Proudfoot, 1982).

2.3.3 General Features of Globin Gene Structure

(a) Gene duplication and concerted evolution

A common feature of all globin gene clusters is the occurrence of pairs of genes which appear to have resulted from some form of duplication event; these include the α_1-α_2, $^{G}\gamma$-$^{A}\gamma$, and δ-β gene pairs, as well as the pseudogenes and their corresponding "functional" genes.

The mechanism of unequal crossing-over between homologous regions is thought to be responsible for the duplication of gene sequences. Early studies of the Hb Lepore β globin chain showed that its likely origin was an unequal crossover event between β and δ globin genes. Subsequent studies at the DNA level by Flavell *et al.* (1978) demonstrated that an unequal crossover event had indeed taken place to form the β-δ fusion product. An entire gene can be duplicated if regions of homology are found outside the actual gene itself. An unequal crossover event could then result in a duplication of the gene in question on one strand, with the reciprocal product of that gene being deleted.

Evidence that unequal crossing-over can occur within the human α-globin cluster and results in a duplication or deletion of genes is provided by the frequent occurrence of chromosomes containing one (Orkin *et al.*, 1979), and three (Gossens *et al.*, 1980) α globin genes in some human populations. The chromosome containing one α globin gene is associated with a common form of α thalassaemia (see Lauer *et al.*, 1980). Comparison of the end points of the deletion associated with this disorder with the location of blocks of homologous sequence within the α_1-α_2 gene pair strongly suggests that the deletion results from unequal crossing-over between homologous sequences (Lauer *et al.*, 1980; Liebhaber *et al.*, 1981). Deletions which are indistinguishable from those found in α thalassaemia 2 have occurred spontaneously *in vitro* in the cloned α-globin gene cluster during propagation in *E. coli* (Lauer *et al.*, 1980).

Structural analyses within the human α globin cluster have revealed that the sequences within and flanking the α globin genes show extensive sequence homology (Lauer *et al.*, 1980). With the exception of some small regions of non-homology, each adult α-globin gene is contained

within a duplication unit of about 4 kb (Proudfoot and Maniatis, 1980). The α-globin proteins encoded at these loci have remained identical even though the duplication of the α-globin gene probably occurred before the radiation of the mammals (Zimmer *et al.*, 1980). Since gradual accumulation of mutations within these two genes would have caused divergence in the primary structure of the proteins over this time period, it has been suggested that the two genes may have evolved in parallel. The parallel or coincidental evolution of two genes has been termed "concerted evolution" (Zimmer *et al.*, 1980). Comparisons of the structure of the two human γ globin loci, $^{A}\gamma$ and $^{G}\gamma$, also support the concept of concerted evolution of genes (Little *et al.*, 1979; Slightom *et al.*, 1980). Two different mechanisms for "correction" of closely associated genes have been proposed. Slightom *et al.*, (1980) have postulated an intrachromosomal gene conversion model to explain the pattern of sequence identity between the human $^{A}\gamma$- and $^{G}\gamma$-globin genes. Alternatively, Zimmer *et al.* (1980) have proposed a model for α-globin gene correction which involves interchromosomal unequal crossing-over events. Both models involve some form of recombination.

Liebhaber *et al.* (1981) have found evidence of frequent exchanges of DNA within the human α-globin gene cluster, which is consistent with the idea that mechanisms exist for the suppression of allelic polymorphism, resulting in the phenomenon of concerted evolution. It is unclear, however, whether any selective advantage is gained by the maintenance of homology within a duplicated locus.

(b) Intervening sequences

Two intervening sequences have been found in all functional globin genes thus far studied (see Abelson, 1979; Cleary *et al.*,

1980; Efstratiadis *et al.*, 1980; Maniatis *et al.*, 1980; Dodgson and Engel, 1983; Patient *et al.*, 1980; Hosbach *et al.*, 1983). In the β-like globin genes, the first intron is located between codons 30 and 31, while the second intron is between codons 104 and 105 (see Hosbach *et al.*, 1983). In α-like globin genes, the introns are in homologous positions to those in the β-like globin genes (see Dodgson and Engel, 1983). The size of the introns appears to be remarkably conserved between species, except in the amphibian *Xenopus laevis* which shows a large variation in intron size in some of the adult and larval genes (Hosbach *et al.*, 1983), and in a human (Proudfoot *et al.*, 1982) and chicken (Engel *et al.*, 1983) embryonic α-globin gene, where the introns are much larger than in other globin genes. The significance of the relatively larger introns in these genes is unknown, but they could reflect gene regulation by insertion of DNA elements into the introns (see Hosbach *et al.*, 1983; Roninson and Ingram, 1982).

The presence of introns in both the amphibian α- and β-globin genes in precisely homologous positions within each gene, suggests that introns were present in the ancestral gene which gave rise to the α- and β-globin genes in *X. laevis*, more than 500 million years ago (Jeffreys *et al.*, 1980; Patient *et al.*, 1980; Hosbach *et al.*, 1983).

The nucleotide sequence of the introns within the globin genes appears to have diverged considerably between species (Efstratiadis *et al.*, 1980; Lauer *et al.*, 1980). The apparent conservation of position and to a large extent size, but not base sequence, implies that it may not necessarily be the composition of the intron, but its position within the gene that is important. The function of introns is largely unknown. Introns appear to be associated with the regulation of gene expression (Hamer and Leder, 1979; Leder *et al.*, 1980; Felber *et al.*, 1982). Introns are known in some instances to

divide protein coding regions into functional "domains" (Blake, 1979; Eaton, 1980), and as such may have implications for the evolution of structural genes (Gilbert, 1978; Reanny, 1979).

(c) Pseudogenes

In addition to the genes which code for globin polypeptides, other DNA sequences which hybridise to α- and β-globin gene probes have been found at the globin gene loci. Globin pseudogenes have been detected in many species such as man (Fritsch *et al.*, 1980; Smithies *et al.*, 1982), rabbit (Hardison *et al.*, 1979), mouse (Jahn *et al.*, 1980; Nishioka *et al.*, 1980; Vanin *et al.*, 1980), goat (Cleary *et al.*, 1980; Schon *et al.*, 1981) and chicken (Engel and Dodgson, 1980).

The globin pseudogenes analysed so far exhibit greater than 75% sequence homology to their corresponding "normal" genes, but none of them can code for functional polypeptides because of the presence of small deletions or insertions that result in inactivation of the gene. Some pseudogenes have altered intron/exon junctions, and one mouse α-globin pseudogene has had both of its introns precisely removed to have an uninterrupted coding sequence (Breathnach *et al.*, 1978; Lerner *et al.*, 1980; Vanin *et al.*, 1980; Nishioka *et al.*, 1980; Engel and Dodgson, 1980; Lacy and Maniatis, 1980).

All pseudogenes which have been analysed to date appear to have diverged from their functional counterparts relatively recently. (See Lacy and Maniatis, 1980; Efstratiadis *et al.*, 1980; Proudfoot and Maniatis, 1980; Proudfoot *et al.*, 1982.) Each pseudogene must therefore have arisen independently in the different gene clusters since their divergence times (55-60 million years ago) postdate the divergence of α- and β-globin genes (570 million years) and the

radiation of the mammalian species involved (85 million years).

With the exception of an α-globin pseudogene in the mouse, all globin pseudogenes analysed so far have been syntenic with the functional globin genes. Popp *et al* (1981) and Leder *et al.* (1981) have reported asynteny of the two α-globin pseudogenes and the mouse α-globin gene cluster, with pseudogenes α3 and α4 being on chromosomes 15 and 17 respectively, whilst the functional α-globin genes are closely linked on chromosome 11. The significance of this observation is not known.

Whether pseudogenes have any function is unknown. Vanin *et al.* (1980) have postulated that pseudogenes may function in the regulation of gene expression by providing the transcriptional process with a "runoff" pathway. Alternatively, they suggest that pseudogenes may encode RNA molecules which have some function in the transcriptional process. However, the most likely explanation of the function of pseudogenes is that they represent a stage in the evolutionary process whereby genetic diversity is generated by gene duplication and subsequent divergence of one of the duplicated genes (to create a pseudogene).

2.3.4 Evolution of Globin Genes

Data from protein sequence studies indicate that vertebrate α- and β-globin genes have a common origin (Dayhoff, 1972). Nucleotide sequence data have since confirmed that the α- and β-globin genes are indeed related (reviewed in Jeffreys, 1981).

Globin genes in the amphibian *Xenopus laevis* have been studied at the molecular level. The major adult haemoglobin in *Xenopus* has been found to be coded for by two genes, designated α_1^A and β_1^A, with those genes being closely linked (Jeffreys *et al.*, 1980; Patient *et al.*, 1980). Homologies between the α_1^A and β_1^A globin genes and their

chromosomal arrangement suggests that they have arisen by tandem duplication of a primordial globin gene, approximately 570 million years ago (Jeffreys *et al.*, 1980). An unlinked *Xenopus* globin gene cluster coding for minor haemoglobin species has also been found (Jeffreys *et al.*, 1980). This cluster is thought to have resulted from a tetraploidisation event in an ancestor of *X. laevis*. An extant representative of such an ancestor, *X. tropicalis*, has been shown to have a single globin gene cluster.

The organisation of the *X. laevis* adult and larval globin genes has been determined, in addition to the nucleotide sequences of some of the genes (Jeffreys *et al.*, 1980; Patient *et al.*, 1982; Kay *et al.*, 1983; Knochel *et al.*, 1983; Hosbach *et al.*, 1983). The major globin gene cluster consists of adult (α_1^A,β_1^A) and larval (α_{1a}^L,α_{1b}^L,β_{1a}^L,β_{1b}^L) genes arranged 5'-α_{1a}^L-α_{1b}^L-α_1^A-β_1^A-β_{1a}^L-β_{1b}^L-3'. The minor globin gene cluster also consists of two adult genes (α_2^A,β_2^A) and four larval genes (α_{2a}^L,α_{2b}^L,β_{2a}^L,β_{2b}^L) arranged in a similar fashion to the major cluster, although distances separating the genes vary from each cluster, which may reflect a recent divergence of the two clusters (Patient *et al.*, 1982; Hosbach *et al.*, 1983).

From calculations on the human and chicken β-globin gene cluster it has been deduced that the embryonic, foetal and adult gene β-globin arrangement in those species has resulted from relatively recent gene duplications, having occurred within the last 200 million years of evolution (Efstradiadis *et al.*, 1980; Roninson and Ingram, 1982). Amphibians diverged from the mammalian and avian lines of evolution about 350 million years ago (Goodman *et al.*, 1975). It is obvious, therefore, that the larval and adult globin genes of *X. laevis* have arisen by gene duplication events independent of those which formed the globin gene clusters of higher vertebrates. The duplications

by which the larval and adult genes of α- or β-globin type arose, must have occurred early in evolution since Widmer (1981) found no evidence of cross-hybridisation between larval and adult genes within the same cluster and hence must have diverged considerably. The presence of a similar arrangement of larval genes in both the major and minor gene clusters suggests that the α^L and β^L genes underwent gene duplications before the genome duplication which gave rise to the minor globin gene cluster. The tandemly duplicated larval genes within each of the globin gene clusters (major and minor) are found to cross-hybridise more strongly to each other than they do to their orthologous counterparts, implying that extensive homologies between the linked larvel genes are probably maintained by "concerted evolution" (Hosbach *et al.*, 1983).

In contrast to the situation in *X. laevis*, the α- and β-globin gene clusters have been found to be asyntenic in the chicken (Hughes *et al.*, 1979). It is possible that such an arrangement may have arisen by a polyploidisation in the reptilian ancestors of birds, perhaps 300 million years ago, with subsequent divergence of each cluster (Ohno, 1973). An event such as polyploidisation would create multiple copies of an αβ globin cluster (two per tetraploidisation) in the genome. This could leave at least one copy of each gene free from functional constraint and hence able to diverge. If there was not a selective advantage associated with close linkage of the α- and β-globin genes, then it is quite possible that polyploidisation and subsequent divergence would result in asynteny between globin genes. Apart from *X. laevis*, the α- and β-globin genes have been found to be asyntenic in all vertebrates examined. The chromosomal arrangement of the globin genes has not been examined in any vertebrate which predates the amphibians, i.e. the fishes.

(a) Evolution of the β globin gene cluster in birds and mammals

Chickens have four different β-like globin chains ε, ρ, β^H and β, and three different α-like globin chains, Π, α^D and α^A. There are three different embryonic globins, HbE ($\alpha_2{}^A\varepsilon_A$), HbP ($\Pi_2\rho_2$), HbM ($\alpha_2{}^D\varepsilon_2$), with adult globins HbA ($\alpha_2{}^A\beta_2$) and HbD ($\alpha_2{}^A\beta_2$) (Moss and Thompson, 1969). The β-like genes have been shown to be arranged in the order 5'-ρ-β^H-β-ε-3' (Dolan *et al.*, 1981). The gene order in the β globin is unique amongst the vertebrates studied to date, with an embryonic, ε, being present on the 3' side of the cluster. Roninson and Ingram (1982) have compared the sequence of the ε and ρ genes to determine their relationship. It was concluded that each embryonic gene probably arose by independent duplication of the adult β globin gene, and was not derived by duplication of an ancestral embryonic globin gene that had already diverged from a precursor adult globin gene. This would explain why the ρ- and ε-globin genes are not located next to each other like other globin gene pairs that are expressed co-ordinately in development. Chickens and mammals diverged from a common ancestor approximately 270 million years ago (Goodman *et al.*, 1975) and during that time it is likely that the β-globin clusters evolved by independent duplication events since, the appearance and divergence of the embryonic and minor globin genes are well characterised in the mammals (Lacy *et al.*, 1979; Barrie *et al.*, 1981).

The β-globin cluster in the BALB/c mouse has been partially characterised by Jahn *et al.* (1980). Seven β-like sequences have been found, and are arranged as follows: 5'-εy3-βh0-βh1-βh2-βh2-β^{maj}-β^{min}-3'. The genes which code for the major adult β chain (β^{maj}) and the minor adult β chain (β^{min}) appear to have resulted from a single duplication event, and are analogous

to the human β and δ globin genes. In man, the most 5' of the adult β globin genes codes for the minor haemoglobin chain, whereas in the BALB/c mouse, the reverse is true (Jahn *et al.*, 1980). It appears that the structure of the BALB/c β^{maj},β^{min} gene pair has evolved in a different manner to that of the human cluster, probably by independent gene duplications in an ancestor common to both man and mouse, a situation similar to that which is thought to have occurred in the chicken. The general organisation of the mouse β globin cluster is similar to that of humans, the order of genes being the order in which they are developmentally expressed, with the βh2 gene being highly diverged, and most likely a pseudogene (Konkel *et al.*, 1979). It appears then that the overall structure of the β globin clusters in man and mouse have developed by a series of independent but similar gene duplications.

The arrangement of primate β-related globin genes has been determined for a number of species (Barrie *et al.*, 1981; Martin *et al.*, 1983). A prosimian, the brown lemur, has been found to have a single β-globin gene cluster arranged 5'-ε-γ-β-3', with the γ and β genes being separated by a gene containing the 3' end of a β-globin gene, preceded by sequences related to the 5' end of the ε-globin gene (Barrie *et al.*, 1981). A similar arrangement of β-globin genes has been found in the rabbit (Lacy *et al.*, 1979) and hence this arrangement was probably present in a common ancestor of the prosimian and the rabbit some 85 million years ago, before the radiation of mammals (Romero-Herrera *et al.*, 1973).

The absence of a γ-globin gene in the prosimian gene cluster implies that the event which gave rise to the γ-globin gene in primates occurred 40-70 million years ago - the time of divergence of the prosimian and the simian lines (Romero-Herrera *et al.*, 1973).

Barrie *et al.* (1981) have found that the organisation of β-globin cluster in the gorilla and yellow monkey is indistinguishable from man, but a new world monkey, the owl monkey, has a single γ-globin gene. This suggests that the $^{G}\gamma$-$^{A}\gamma$ duplication occurred about 20-40 million years ago, the time of divergence of the new world monkeys from the great apes and old world monkeys. The δ globin polypeptide is not expressed in old world monkeys, although the δ globin gene is present in the β-globin cluster (Martin *et al.*, 1980). Martin *et al.* (1983) have found that the nucleotide sequence of the rhesus and colobus monkeys (old world monkeys) δ-globin gene was 94% homologous to the functional human δ globin gene. It is thought that the silencing of the δ-globin gene in old world monkeys was probably a result of three mutations present in the DNA 5' to the putative δ-globin gene transcription initiation site (Martin *et al.*, 1983).

(b) Evolution of the α-globin gene cluster in birds and mammals.

The complex and variable structure of the vertebrate β-globin locus has allowed many valuable evolutionary comparisons to be made. Until recently the α-globin locus had been subject to comparatively fewer evolutionary studies than the β-locus. The evolutionary history of the vertebrate α-globin locus is now emerging as being of particular interest in that it appears to have a different evolutionary history to that of the β-globin locus. The α-globin gene cluster in the chicken consists of three functional genes, an embryonic gene (Π') and two adult genes (α^{A},α^{D}) arranged 5'-Π'-α^{A}-α^{D}-3' along the chromosome (Dodgson *et al.*, 1981; Dodgson and Engel, 1983). The absence of duplicated adult α-globin loci in amphibians (Hosbach *et al.* 1983) implies that the α-globin duplication event may have occurred some

350 million years ago, the time of divergence of the amphibian and avian lines (see Goodman, 1975). Proudfoot *et al.* (1982) found that the amount of sequence divergence between the chicken α^A and Π' genes was nearly identical to the amount of sequence divergence between the human $\alpha 1$ and ζ genes. In addition, the chicken Π'/human ζ and chicken α/human α divergences were also identical. This implies that the embryonic and adult α globin genes began to diverge from each other long before the birds and mammals began to diverge. That is, the chicken Π' and human ζ genes most likely diverged from a common ancestral α-like globin gene. Similar evidence comes from a calculation of the relationship between the human ζ/α genes which gives a divergence time of approximately 400 million genes, implying that distinct embryonic and adult α-globin genes appeared prior to, or about the time of the appearance of amphibians. This is consistent with the fact that the human α/*Xenopus* α divergence is approximately the same as the human ζ/human α divergence (Proudfoot *et al.*, 1982). Since the α and β globins are closely linked in amphibians but unlinked in other vertebrates, it is likely that the ζ/α divergence began prior to the time the clusters were separated.

In summary, it appears likely that all embryonic and adult α-like globin genes may have evolved from a common ancestral gene and have since diverged from each other at a constant rate (see Proudfoot *et al.*, 1982). Proudfoot *et al.* (1982) have calculated that the human β-like embryonic gene (ε) is the product of a much more recent evolutionary event than the divergence of the α-like embryonic gene (ζ). Hence, even though the temporal and quantitative expression of ζ and ε globin genes must be co-ordinately controlled during development, their evolutionary histories are clearly different.

It is clear that the evolutionary history of the vertebrate α-globin gene cluster is distinct from that of the β-globin cluster. In the β-globin gene cluster, many of the genes appear to have arisen through a series of independent gene duplications in each evolutionary line (see previous section), whereas the α-globin genes appear to have diverged in a constant manner leading to the conclusion that they may all have evolved from an α-like gene in a common ancestor of amphibians and birds, more than 500 million years ago.

2.4 GENETIC STUDIES IN MARSUPIALS

The order Marsupialia comp rises approximately 250 species in 8 superfamilies (Kirsch and Calaby, 1977). The marsupials are estimated to have diverged from eutherian mammals approximately 1.3×10^8 years ago (Air *et al.*, 1971).

The structural genes for the enzymes G6PD and PGK-A have been shown to be X-linked in kangaroos using family studies (Richardson *et al.*, 1971; Cooper *et al.*, 1971). From these studies and others (Johnston *et al.*, 1975; VandeBerg *et al.*, 1977; Cooper *et al.*, 1977) it was shown that the process of X inactivation in marsupials appeared to differ fundamentally from that of eutherian mammals in that it appeared to be non-random, with the paternally derived X-chromosome being preferentially inactivated in females. In some tissues and in cultured fibroblasts, inactivation of paternal genes is either incomplete or absent, and the degree to which this occurs differs for G6PD and PGK-A (VandeBerg *et al.*, 1977; Cooper *et al.*, 1977a; Johnston *et al.*, 1978; Raphael and Cooper, 1978). These observations led Cooper *et al.* (1977b) to postulate that dosage compensation

in marsupials does not operate at the level of the whole chromosome, but rather that different segments of the X chromosome may be under independent control.

The chromosomes of a wide range of marsupials have been studied by conventional cytogenetic techniques and DNA measurements (Sharman, 1973, 1974; Hayman and Martin, 1974; Sharp, 1982; Hayman and Sharp, 1981). Rofe (1979) examined the G-banding pattern of chromosomes from a large number of macropod marsupials. From these studies a phylogenetic pathway of the Macropodidae was constructed, with most species differing by simple inversions and Robertsonian events. The chromosomes of the superfamily Dasyuridae have also been analysed by the G-banding technique (Rofe, 1979; Young *et al.*, 1982) and were found to show a large amount of karyotypic conservation, with any differences between species being due to variation in the amount and position of heterochromatin, small inversions, or, in one case, a reciprocal translocation.

The determination of gene arrangement in marsupials is of interest not only for establishing a clearer picture of marsupial evolution, but also for assessing the forces governing linkage relationships between genes over the evolutionary distance which separates marsupial and placental mammals.

Apart from family studies involving the genes for G6PD and PGK-A, very few genes have been mapped in marsupials. Marsupials have many advantages for genetic and somatic cell genetic analysis (see Cooper, 1974; Graves and Hope, 1977a). These advantages include low numbers of large, easily distinguished chromosomes and the potential availability of a large number of isozyme differences from eutherian cells. Many difficulties were encountered in attempts to hybridise marsupial and eutherian cells, but after a series of studies on the

co-cultivation of rodent and marsupial cells (Graves and Hope, 1977), fusion (Graves and Hope, 1977b), heterokaryon formation and activity (Graves *et al.*, 1977; Graves and Hope, 1978) and the development of appropriate selective systems (Hope and Graves, 1978a), successful production of rodent x marsupial hybrids was achieved. Marsupial chromosomes were found to be preferentially eliminated from rodent x marsupial cell hybrids.

However, problems have been encountered mapping marsupial genes using these hybrids due to almost complete segregation of the chromosomes of the marsupial parent. Whilst all hybrids express at least one marsupial isozyme (the selected marker), very few contain readily identifiable marsupial chromosomes or chromosome fragments (Hope, pers. comm.; Graves, pers. comm.). Hope and Graves (1978b) observed extensive fragmentation of the marsupial genome soon after fusion with mouse cells. It is possibly as a result of such fragmentation that difficulties have been encountered in obtaining complete and/or identifiable marsupial chromosomes in the hybrids.

Despite these problems some hybrids containing a number of marsupial chromosomes have been obtained for the species *Macropus rufus*, the red kangaroo (Donald and Hope, 1981), *Macropus rufogriseus*, the red-necked wallaby (Sykes and Hope, in preparation), and *Pseudocheirus peregrinus*, the ring-tailed possum (Hope, pers. comm.). Using these hybrids, the genes for HPRT, G6PD and PGK-A have been assigned to the red kangaroo X chromosome (Donald and Hope, 1981), HPRT and PGK-A to the *M. rufogriseus* X chromosome (Sykes and Hope, in preparation) and to the *P. peregrinus* X chromosome (Hope, pers. comm.). Only two autosomal genes have been assigned to marsupial chromosomes; the gene for LDH-A to the *M. rufus* chromosome 5 in *M. rufus* x rodent cell hybrids (Donald and Adams, 1981), and a gene

coding for the marsupial antigen GA-1 to the long arm of the *M. rufogriseus* chromosome number 3, also using cell hybrids (Sykes and Hope, in preparation).

Sixteen somatic cell hybrids between the dasyurid marsupial *Sminthopsis crassicaudata* and rodents have been produced (Wainwright and Hope, unpublished results). Although these hybrids express marsupial isozymes they do not contain any cytologically identifiable marsupial material (Wainwright *et al.*, in preparation). Some linkage groups are known for *S. crassicaudata*, but as yet there are no gene assignments in *S. crassicaudata* (Bennett, pers. comm.). Clearly, if any genes are to be mapped in *S. crassicaudata*, and probably in other dasyurid marsupials, an alternative method to the use of somatic cell hybrids is preferable.

Very few genetic studies at the nucleic acid level have been carried out in marsupials. Satellite DNA's have been studied in the macropod marsupials *Macropus rufogriseus* (Dunsmuir, 1976) and *Macropus robustus robustus* (Venolia and Peacock, 1981). The *M. robustus robustus* satellite was found to cross hybridise to the long arm of the *M. rufogriseus* X chromosome, indicating that this sequence has been conserved in the two species. Apart from the α- and β-globin mRNA in a dasyurid marsupial, no marsupial gene sequences corresponding to structural genes have been cloned (Wainwright, in preparation).

CHAPTER 3. ISOLATION AND PARTIAL CHARACTERISATION OF NATIVE CAT GLOBIN mRNA.

3.1 INTRODUCTION

It has been shown that vertebrate globin mRNA can be isolated from reticulocytes (see Taylor, 1979). The major protein product of the reticulocyte is globin, which constitutes greater than 50 per cent of all protein synthesised in the cell (see Harrison, 1974) Purified 10S mRNA from reticulocytes yields greater than 95% globin protein when translated in cell free systems (Paterson and Roberts, 1977), inferring that globin mRNA constitutes by far the largest proportion of mRNA present in the 10S fraction.

Many eukaryotic mRNA molecules have a polyadenylic acid tract at their 3' terminus (Lee *et al.*, 1971; Edmonds *et al.*, 1971; Darnell *et al.*, 1971). It has been shown that globin mRNA sequences from a number of vertebrates including man (Nudel *et al.*, 1977), mouse (Lingrel *et al.*, 1972), and chicken (Engel and Dodgson, 1978) will bind oligo-dT, due to the presence of poly-A tracts on the 3' ends of the RNA. A number of techniques have been developed utilising the poly-A segment as a specific method for the selection of mRNA. Materials used include millipore filters (Lee *et al.*, 1971); unmodified cellulose (Schutz *et al.*, 1972); poly-U bound to filters, to cellulose or to sepharose (Kates, 1970; Kates *et al.*, 1973; Lundberg and Pearson, 1972); and oligo-dT bound to cellulose (Aviv and Leder, 1972). The more efficient methods involve the formation of stable base-pairing between the poly-A tail of the mRNA, and the complementary nucleotides, either poly-U or oligo-dT attached to an inert support.

To determine the nature of any mRNA species present in an RNA

fraction, a sensitive and efficient assay system is required. Cell free systems derived from wheat germ (Roberts and Paterson, 1973) or from rabbit reticulocytes (Pelham and Jackson, 1975) have been shown to translate RNA molecules from a variety of sources (see Taylor, 1979). The translation of adult human and adult rabbit globin mRNA in a wheat germ cell free systems leads to the production of full length globin polypeptides, with low background (Roberts and Paterson, 1973; Paterson *et al.*, 1975).

This chapter describes the isolation and partial chracterisation of the adult globin polypeptides of the native cat and their comparison to the *in vitro* translation products of native cat reticulocyte poly-A^+ mRNA. The purification of a fraction of native cat mRNA encoding globin proteins is also described.

3.2 RESULTS

3.2.1 Purification of Globin Protein from the Native Cat

Whole blood from a male native cat was obtained and globin protein extracted by standard methods as described in section 10.2.1. The globin protein extracted was analysed on an SDS-acrylamide gel. Figure 3.1 shows that the "globin protein" consisted primarily of a polypeptide(s) with a molecular weight of approximately 14,000d. In addition, a smaller amount of protein of molecular weight 25,000d can be observed. Chromatography of the "globin protein" on a carboxymethylcellulose (CMC) column resulted in the elution of two distinct protein peaks (Fig. 3.2). Analyses of the eluted peaks by SDS gel electrophoresis (Fig. 3.1) shows that both peaks consist of a protein or proteins of a molecular weight of 14,000d. Each peak from the CMC column was analysed by electrophoresis on cellulose

Figure 3.1

Electrophoresis of native cat globin protein on a 10% polyacrylamide gel. Approximately 30μg of protein was loaded into each sample well.

1. fraction A from a CMC column (see Fig. 3.2)
2. unfractionated globin protein
3. fraction B from a CMC column (see Fig. 3.2).

Molecular weights were determined by co-electrophoresis with

ovalbumin	43 kd
α-chymotrypsinogen	25.7 kd
egg white lysozyme	14.3 kd
β-trypsin inhibitor	6.2 kd

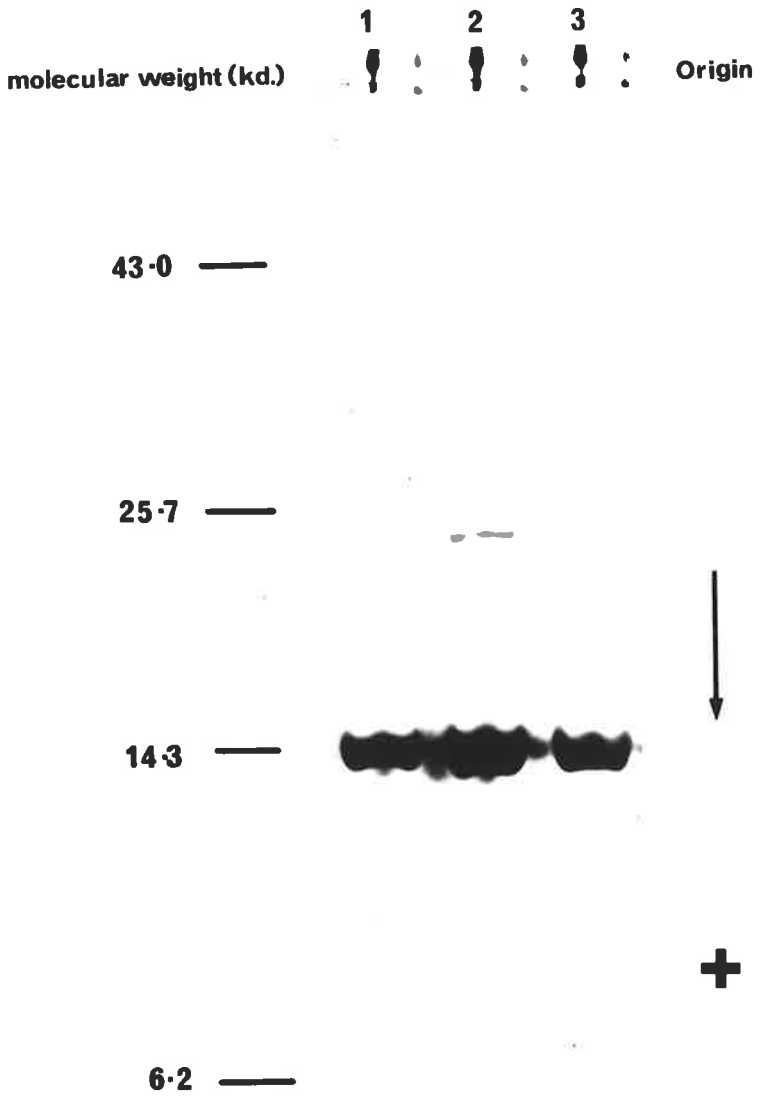
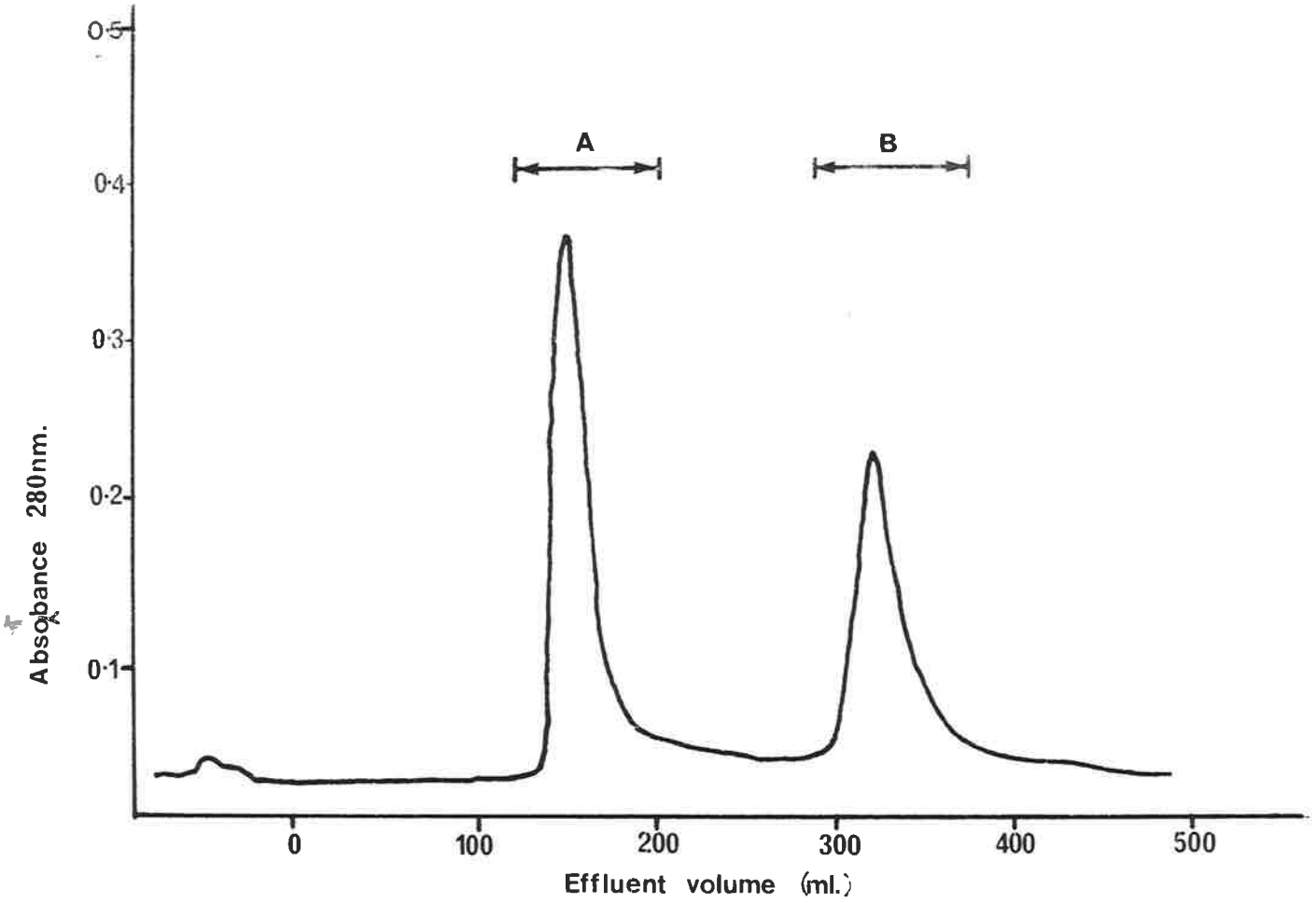

1
2
3
molecular weight (kd.)
Origin
43·0
25·7
14·3
6·2
+

Figure 3.2

Carboxymethyl cellulose chromatography of globin protein. 30mg of globin protein was loaded on to a 1.0 x 12cm CMC column and eluted using a buffer gradient from 0.02M Na^+ pH 6.7 to 0.07M Na^+ pH 6.7. Each peak was collected as indicated and analysed by SDS-polyacrylamide electrophoresis. (see Fig. 3.1)


0·5
0·4
0·3
0·2
0·1
Absorbance 280nm.
A
B
0
100
200
300
400
500
Effluent volume (ml.)

acetate. The results in Figure 3.3 show that each peak resolves into a separate band under the cellulose acetate electrophoresis conditions, which separate proteins based largely upon their net charge.

3.2.2 Preparation of Poly-A^+ RNA from the Native Cat

Reticulocytes were obtained from the blood of two native cats made anaemic by daily injections of phenylhydrazine hydrochloride. Total RNA was extracted from the reticulocytes and passaged through a poly-U sepharose column as described in section 10.2.3. A typical column elution profile as shown in Figure 3.4 and shows the presence of a large peak of unbound RNA, and a smaller peak that was initially bound to the column and subsequently eluted in the presence of 90% formamide. The bound and unbound fractions were then ethanol precipitated and the relative amounts of RNA measured. It was found that approximately 5% of total cellular RNA bound to the poly-U sepharose column.

3.2.3 Translation of Poly-A^+ RNA

Native cat reticulocyte RNA fractions which had not bound (presumed poly-A^-) and bound (presumed poly-A^+) to a poly-U sepharose column were translated in a wheat germ cell free system. The unbound RNA fraction was found not to stimulate the incorporation of tritiated leucine significantly over background, whilst the addition of 1 µg of poly-A^+ RNA resulted in a 50-fold increase in TCA precipitable material over background. The translation products of the poly-A^+ RNA were analysed by gel electrophoresis (Fig. 3.5) and were found to be polypeptides of molecular weight 14 kd and 25 kd. Proteins of molecular weights of 14 kd and 25 kd have both previously been shown to be present in protein extracts from native cat reticulocytes

Figure 3.3

Cellulose acetate electrophoresis of globin protein. Samples were dissolved at a concentration of 2mg/ml and electrophoresed for 5h at 225V. Protein was visualised by staining with 0.5% amido black.

Samples are:

1. 5µg of protein from CMC column peak "a" (see Fig. 3.2)
2. 9µg of total globin protein
3. 5µg of protein from CMC column peak "B" (see Fig. 3.2).

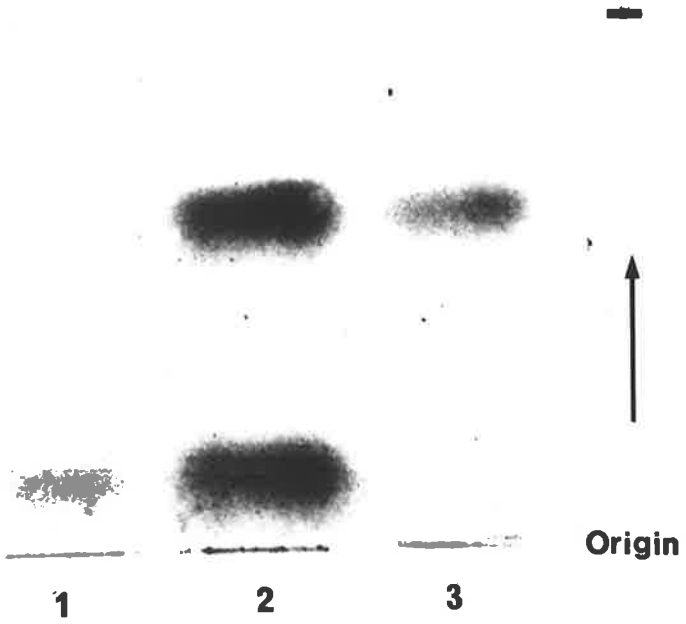

Origin
1
2
3

Figure 3.4

Poly-U Sepharose chromatography of native cat reticulocyte RNA. Approximately 4mg of total RNA was loaded on to a 5ml poly-U Sepharose column and unbound RNA was eluted in the presence of 25% formamide. Bound RNA was eluted by the addition of buffer containing 70% formamide. The bound material was approximately 7% of the total RNA applied to the column.

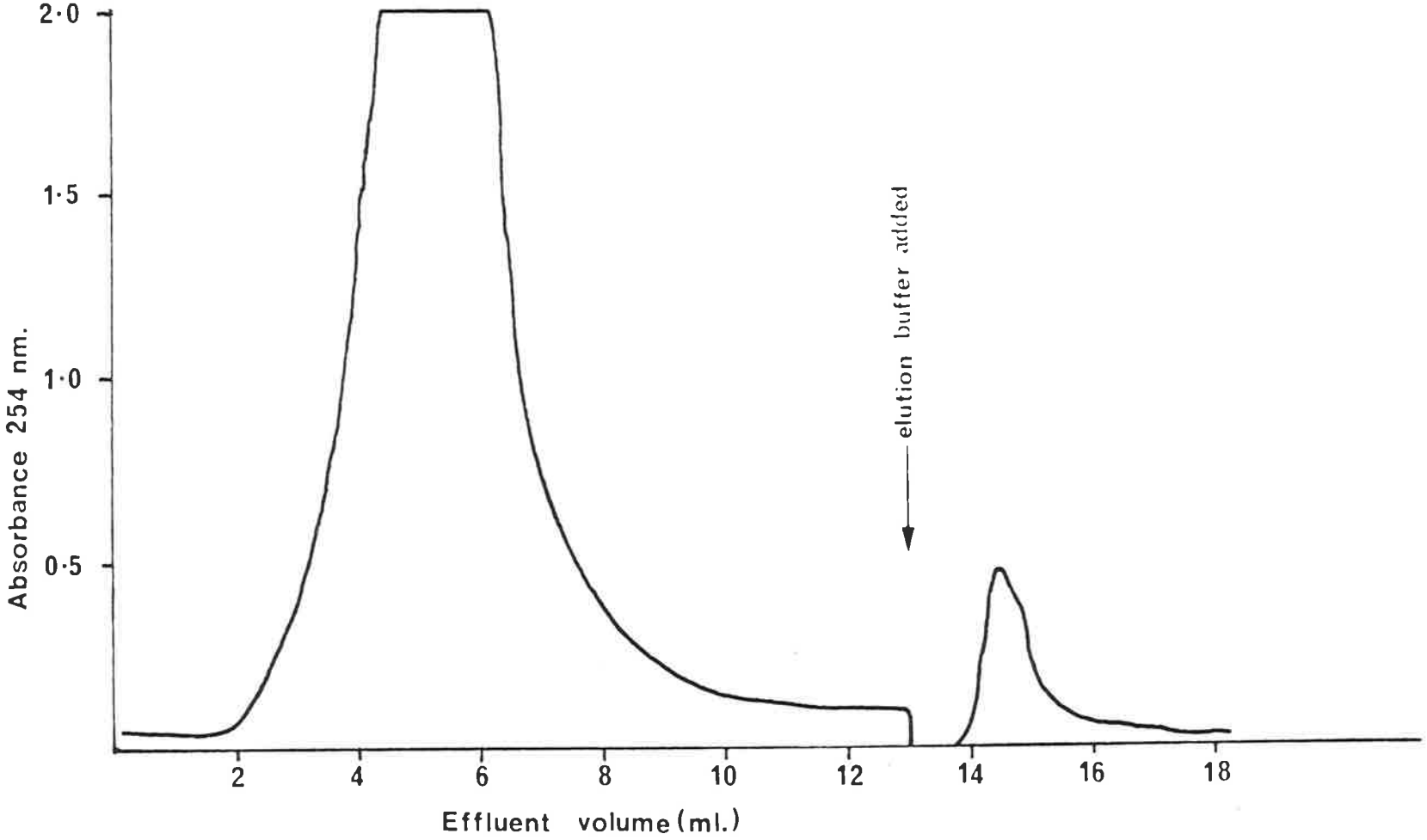

2·0
1·5
1·0
0·5
Absorbance 254 nm.
elution buffer added
2
4
6
8
10
12
14
16
18
Effluent volume (ml.)

Figure 3.5

SDS-polyacrylamide analysis of translation products.

Total and fractionated reticulocyte RNA was translated in a wheat germ cell-free system and analysed on a 10% SDS-polyacrylamide gel. The gel was subjected to fluorography and exposed for 3 days (samples A and B) or 6 days (samples C and D). Molecular weights were calculated from molecular weight markers:- ovalbumin (43.0 kd), α-chymotrypsinogen (25.7 kd), lysozyme (14.3 kd), cytochrome (12.3 kd)

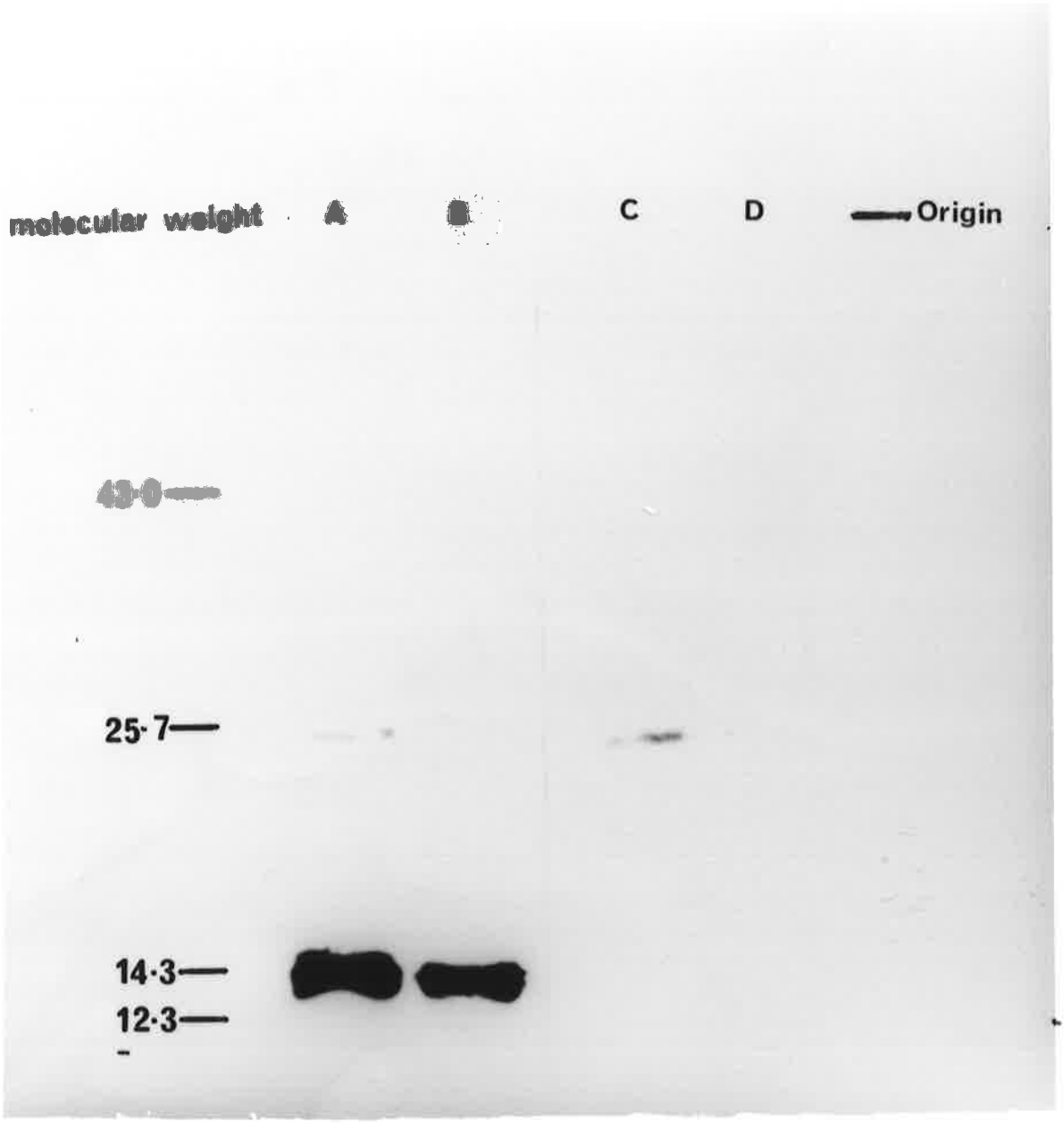
Samples are:

A. poly-A^+ RNA

B. 10S RNA

C. 14-24S RNA

D. >24S RNA


molecular weight
A
B
C
D
Origin
43·0
25·7
14·3
12·3

(section 3.2.1).

3.2.4 Isolation of 10S and mRNA

In an attempt to isolate a relatively pure fraction of globin mRNA, the poly-A^+ RNA was fractionated on sucrose gradients. An absorbance profile representative of such a fractionation is shown in Figure 3.6.

As well as a peak corresponding to a size of 10S, significant amounts of 18S and 28S RNA were also detected. The presence of 18S and 28S RNA's, presumed to consist predominantly of ribosomal RNA, suggests that there was a degree of non-specific binding of RNA to the poly-U sepharose column. It is interesting to note that 5S RNA does not appear to have bound to the poly-U column. Fractions from the gradient were pooled as indicated in Figure 3.6. The result of the translation of these fractions is shown in Figure 3.5. It was found that greater than 90% of the translational activity, as measured by the incorporation of ^3H leucine, was contained in the 10S RNA fraction. As shown in Figure 3.5, the 10S translation products have a molecular weight of approximately 14 kd. Translational activity resulting in the production of the 25 kd polypeptide(s) was detected in an RNA fraction with a size range of approximately 14-24S.

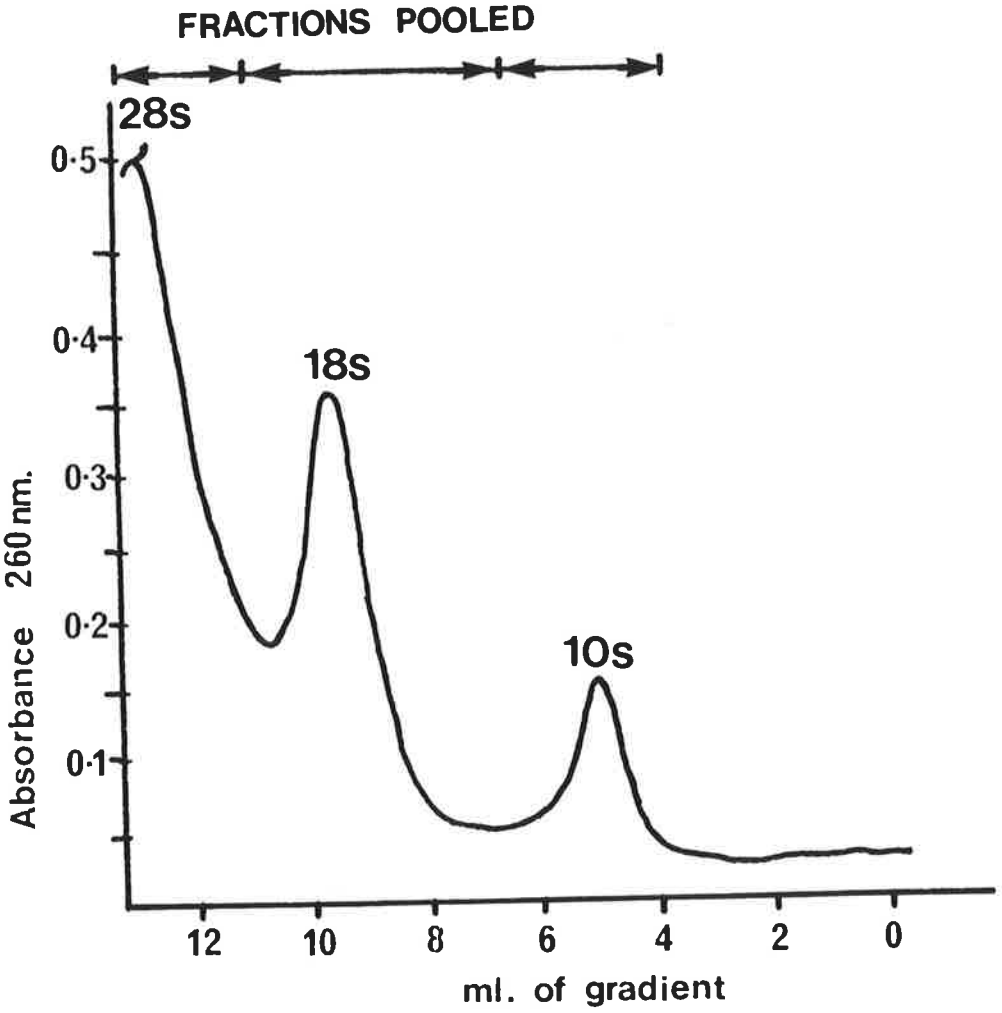
Assuming that the greater proportion of the 18S and 28S peaks were non-specifically bound RNA, then it appears that approximately 25-30% of the RNA fraction bound to the poly-U column was polyadenylated RNA capable of stimulating protein synthesis in the wheat germ cell free system.

3.3 DISCUSSION

The adult globin proteins from a number of marsupials have

Figure 3.6

Sucrose gradient fractionation of affinity purified RNA. 40μg of affinity purified reticulocyte RNA was loaded on to a 13 ml linear 10-40% sucrose gradient and centrifuged at 36,000 rpm for 17 hours. The gradient was fractionated as indicated. Recovery of 10S RNA was approximately 25% of the total RNA. Approximate RNA molecular weights were calculated from identical gradients run in parallel containing chicken embryo 28S, 18S and 5S RNA - a kind gift from Dr. J. Brooker.


FRACTIONS POOLED
28S
18S
10S
0·5
0·4
0·3
0·2
0·1
Absorbance 260 nm.
12
10
8
6
4
2
0
ml. of gradient

been extensively characterised by Thompson and co-workers (Thompson *et al.*, 1969; Air and Thompson, 1969; Beard and Thompson, 1970). In all cases the adult α- and β-globin chains have displayed the amino acid sequence conservation and electrophoretic properties characteristic of vertebrate globin proteins. Apart from a report of a polymorphism involving the adult β-globin chain of the grey kangaroo (Thompson *et al.*, 1969), and the observation of the appearance of foetal haemoglobin in pouch young of the possum (Hope, 1969) and grey kangaroo (Richardson and Russell, 1969), no minor or embryonic/foetal globins have been described in marsupials. The results presented in this chapter show that a globin protein extract derived from the red blood cells of an adult native cat consisted largely of a polypeptide or polypeptides of a molecular weight of 14 kd. the molecular weight expected for globin protein. In addition, an unidentified protein or proteins of a molecular weight of 25 kd was also present at low levels. It is not known whether the 25 kd protein is a globin protein, but such an interpretation is highly unlikely, based upon its apparent molecular weight. It is also unlikely that the 25 kd protein band results from an aggregation of globin chains since itw molecular weight is lower than that expected from such an aggregate. Fractionation of the native cat globin protein extract by carboxymethyl cellulose chromatography resulted in the separation of two classes of polypeptide chain distinguishable by electrophoresis. This result was consistent with the behaviour of kangaroo (Thompson *et al.*, 1969), chicken (Moss and Thompson, 1969), and human (Clegg *et al.*, 1965; Nudel *et al.*, 1975) globins under similar conditions. In all cases the more cathodal band has been identified as α-globin and the other band as β-globin. The Results to be presented in Chapter 5 demonstrate that the two protein bands shown in Figure 3.3 correspond to native cat

globin, the more cathodal band being α-globin and the other, β-globin. It is interesting to note that the 25 kd protein is not present in protein fractions bound to the CMC column.

Column chromatography of total reticulocyte RNA on poly-U sepharose resulted in the isolation of an RNA fraction with the ability to direct protein synthesis in a cell free system. The major products directed by the affinity purified RNA gave molecular weights of approximately 25 kd and 14 kd. Whilst the 14 kd polypeptides have not been directly identified as globins, all indirect evidence available is consistent with that conclusion. The identity of the 25 kd protein is unknown. It is difficult to estimate the relative amount of each RNA coding for the 25 kd and 14 kd proteins present in the poly-A^+ RNA sample based upon *in vitro* translation products. Even mRNA molecules that are normally expressed in the same cell can have different *in vitro* translational activities, for example, rabbit α-globin mRNA translates more efficiently than rabbit β-globin mRNA in a wheat germ cell free system (Roberts and Paterson, 1975) and ovalbumin always translates more efficiently than conalbumin in the reticulocyte lysate system (Palmiter, 1974). In addition, ^3H leucine was the only labelled amino acid used in the translation system and depending on the relative proportions of leucine in each polypeptide, the relative amount of each mRNA species may be over- or underestimated. Previously, there has been no report of the isolation of mRNA from any marsupial species. Information available from other mammals suggested that an enrichment for globin mRNA might be obtained by the purification of a 10S fraction from reticulocyte poly-A^+ RNA. Sucrose gradient fractionation of the affinity purified RNA showed significant amounts of 18S and 28S RNA as well as a distinct peak at 10S, with most of the translational activity of the RNA being in the 10S fraction. Although the proportion

of the 18S and 28S peaks may consist of mRNA encoding other proteins and aggregates of 10S, and other RNA molecules, it is likely that the majority of the 18S and 28S RNA was non-specifically bound ribosomal RNA. Perhaps a lesser degree of non-specific binding of RNA to the poly-U sepharose column may have been obtained if the RNA was heat disaggregated before passaging, or the RNA was passaged several times through the column. Due to the limited amount of RNA available for these experiments, any procedures which increased the chance of sample contamination by RNAase through increased handling were avoided wherever possible.

Translation of the 10S RNA and subsequent analyses of the translation products showed that the reticulocyte 10S RNA encoded the 14 kd polypeptide(s). The 25 kd protein was not detected in the translation products of the 10S RNA.

In summary, a fraction of native cat reticulocyte RNA has been isolated which directs the synthesis of 14 kd polypeptides in a wheat germ cell free system. Indirect evidence supports the identification of the 14 kd polypeptides as globins. On the basis of these findings a cDNA library constructed from native cat poly-A^+ reticulocyte RNA should contain adult α- and β-globin cDNA clones.

CHAPTER 4. MOLECULAR CLONING OF RETICULOCYTE POLY-A$^+$ mRNA SEQUENCES

4.1 INTRODUCTION

The isolation of pure mRNA species, which can subsequently be used as in the analysis of mRNA structure and as probes for analysis of gene arrangement, can be achieved by the use of recombinant DNA techniques. This procedure was first used in the purification of rabbit β-globin mRNA sequences (Rougeon and Mach, 1975; Maniatis *et al.*, 1975; Rabbits, 1976). The basic approach to the molecular cloning of mRNA sequences involves the synthesis of a double stranded complementary DNA molecule (ds cDNA) from an mRNA template. Each ds cDNA molecule is subsequently joined to a vector DNA molecule, usually a plasmid, that has the ability to replicate autonomously inside a host bacterial cell. The recombinant molecules are subsequently used to transform *E. coli* cells which are then plated out for single colonies. Since each colony is derived from a single cell, and each cell normally only acquires a single recombinant molecule, each colony carries copies of a single mRNA sequence. If the procedures involved in the cloning protocol do not tend to select for or against certain sequences, then the final population of recombinants will reflect the starting population of RNA molecules.

The first step in the production of ds cDNA suitable for cloning is the synthesis of cDNA on the RNA template using avian myeloblastosis virus (AMV) reverse transcriptase. The most commonly used primer for this reaction is oligo dT_{12-16} which binds to the 3' poly-A tract present in most eukaryotic mRNA molecules.

As all known globin mRNA species contain 3' poly-A tracts, and many of them have been cloned using oligo dT_{12-16} priming, this method was used in attempts to clone native cat globin mRNA sequences.

AMV reverse transcriptase leaves a short hairpin loop at the 5' end of the newly synthesised cDNA and this segment of DNA is able to act as a self-primer to copy back along the cDNA molecule (Efstratiadis *et al.*, 1975). This synthesis can be carried out using either *E. coli* DNA polymerase I (Efstratiadis *et al.*, 1975), or AMV reverse transcriptase, which can also act as a DNA dependent DNA polymerase (Rougeon and Mach, 1976). Many investigators have utilised AMV reverse transcriptase to synthesise the second cDNA strand. Although there is one report that second-strand synthesis of immunoglobin light chain cDNA could not be achieved with AMV reverse transcriptase (Rougeon and Mach, 1976), the success of a large number of experiments indicates that this is not a general problem.

The final step in the preparation of cDNA for cloning is the cleavage of the covalently joined primer loop at the 5' end of the molecule, and the removal of any single stranded regions, using the single strand specific nuclease S_1 (Vogt, 1973).

The cDNA molecules may be joined to vector DNA molecules in several ways. The first of these is the addition of homopolymeric oligodeoxynucleotides to the 3' end of the DNA using terminal deoxynucleotidyl transferase (Jackson *et al.*, 1972; Lobban and Kaiser, 1973). When complementary base-pairing nucleotides are added to the ds cDNA and vector DNA respectively, the sequences will anneal to form a circular molecule capable of transforming competent *E. coli* cells. The "A-T tailing" procedure was used by Maniatis *et al.* (1976) and Rabbits (1976) for the cloning of rabbit β-globin mRNA sequences and G-C "tailing" has been used for the cloning of *Xenopus laevis* globin mRNA sequences (Kay *et al.*, 1980).

Other methods for generating recombinant DNA molecules utilise the blunt end ligation properties of T_4-DNA ligase to add a synthetic

oligonucleotide "linker", containing the recognition sequence for a specific restriction endonuclease, to the ends of the ds cDNA. Cleavage with the restriction enzyme yields complementary sequences which may be annealed to similar sequences in the vector DNA molecules (see Deacon *et al.*, 1980).

More recently, methods have been developed which obviate the need for S_1 nuclease treatment of the ds cDNA molecule and therefore tend to result in the production of a greater proportion of "full length" ds cDNA molecules. Zain *et al.* (1979) have transformed *E. coli* with mRNA:cDNA hybrids that have been joined to plasmid vectors. The bacterial host removes the mRNA and replaces it with DNA. In addition to the absence of S_1 treatment, this method requires the synthesis of only one cDNA strand. The major disadvantage of this method is that an RNA:DNA hybrid has a 10X lower transformation efficiency than DNA:DNA hybrids (Zain *et al.*, 1979).

In practice, the tailing cDNA approach involves far less handling of the cDNA than "linker" methods and is more efficient than linker ligation in the generation of recombinant molecules. In addition, Bolivar *et al.* (1977) have shown that if a 3' poly-dG sequence is added to PstI cut vector DNA and this is joined to poly-dG "tailed" ds cDNA, it is possible to regenerate the PstI cleavage site, allowing excision of the inserted ds cDNA sequences.

The plasmid pBR 322 (Bolivar *et al.*, 1977; Sutcliffe, 1978) was chosen as the vector molecule in the following experiments because it is small, well characterised, contains a single PstI recognition site, and carries two antibiotic resistance genes which aid in the selection of recombinant molecules.

This chapter describes the preparation of ds cDNA molecules from native cat poly-A^+ reticulocyte mRNA, the molecular cloning of these

sequences and the selection of clones containing sequences which hybridise to native cat 10S RNA.

4.2 RESULTS

4.2.1 Preparation of ds cDNA from Poly-A^+ RNA

A poly-A^+ RNA fraction isolated as described in Chapter 3 was used as the starting material for ds cDNA synthesis. Synthesis of cDNA from this template was carried out using AMV reverse transcriptase and oligo dT_{12-16} priming (see section 10.2.9). For preparative synthesis 6 µg of affinity purified RNA was used, which corresponds to approximately 2 µg of 10S RNA (see section 3.2.1). Under the conditions detailed in section 10.2.9, 14% copy was achieved, yielding an estimated 0.3 µg of cDNA. AMV reverse transcriptase was used in the second strand synthesis, giving 40% back copy with a final yield of approximately 0.14 µg of ds cDNA. The ds cDNA was diluted into digestion buffer and cleaved with S_1 nuclease.

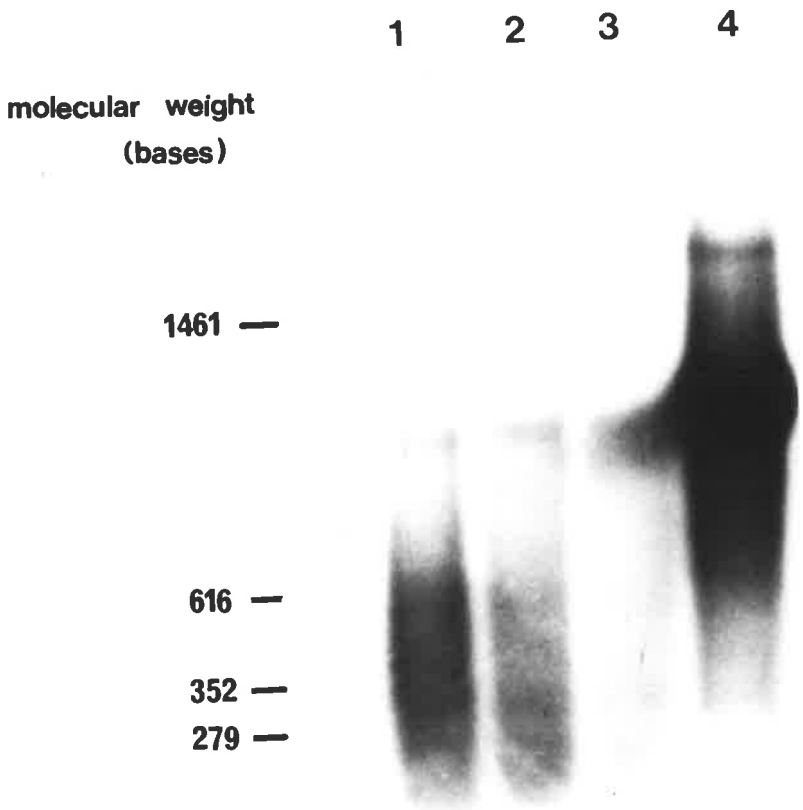
The cDNA first strand synthesis, ds cDNA synthesis, and S_1 cleavage were all analysed by alkaline agarose gel electrophoresis. Figure 4.1 shows that the first strand synthesis resulted in the production of a heterogeneous population of cDNA molecules, with a significant number being greater than 500 bases in length. After second strand synthesis it can be seen that a large proportion of the single stranded cDNA molecules have been converted into double stranded form, resulting in a major band at a molecular weight of approximately 550-600 b.p. (approximately 1100-1200 bases on an alkaline agarose gel). A small proportion of double stranded molecules can be seen to have an apparent molecular weight of approximately 900 b.p. After treatment with S_1 nuclease the ds cDNA molecules revert to a distribution which is

Figure 4.1

Analysis of cDNA and ds cDNA on a 1.2% alkaline agarose gel.

1. cDNA
2. ds cDNA after S_1 nuclease treatment
3. cDNA
4. ds cDNA

Molecular weight markers were generated by end labelling a Sau 96 digest of pBR 322 with α-^{32}P dNTPs.


1
2
3
4
molecular weight
(bases)
1461
616
352
279

similar to that shown by the single stranded cDNA. This indicates that the S_1 nuclease has successfully cleaved the 5' hairpin loop of a large proportion of the double stranded molecules.

The ds cDNA was fractionated by Sephacryl 1000 chromatography. While most of the ds cDNA was greater than 300 base pairs in length, it was important to remove as much of the low molecular weight material as possible as it may have been in large molar excess over the high molecular weight material. The presence of a large number of short ds cDNA molecules results in the cloning of these sequences in preference to the longer molecules (Maniatis *et al.*, 1982).

The Sephacryl 1000 elution profile is shown in Figure 4.2. The highest molecular weight fraction was selected as indicated and then collected by precipitation. The final yield of this high molecular weight fraction of ds cDNA was approximately 2 ng.

4.2.2 Tailing of ds cDNA

Tailing of the ds cDNA was performed using calf thymus terminal deoxynucleolidyl transferase (TdT) and dCTP as dubstrate. Roychoudhury *et al.* (1976) have shown in the presence of Co^{2+} ions TdT will preferentially add nucleotide triphosphates to the 3' end of a double stranded DNA molecule without prior treatment with λ exonuclease. The reaction was monitored as described in section 10.2.9. Figure 4.3a shows a time course for the addition of dCTP to the ds cDNA. The reaction was stopped by the addition of the chelating agent EDTA when it was calculated that an average of 20 residues had been added to each end of the molecule.

4.2.3 Tailing of the Plasmid Vector DNA

Plasmid pBR 322 DNA was linearised by two separate digestions

Figure 4.2

Sephacryl 1000 chromatography of ds cDNA. The ds cDNA fractions used for subsequent cloning experiments are indicated. Arrows indicate the approximate elution volume of some DNA fragments resulting from a Hpa II digestion of pBR 322. A pBR 322 digest was end-labelled with α-^{32}P dCTP, passaged on the sephacryl 1000 column and the fractions analysed by gel electrophoreses.

DNA fragment sizes are:

A. 622 bp

B. 527 bp

C. 404 bp

D. 309 bp

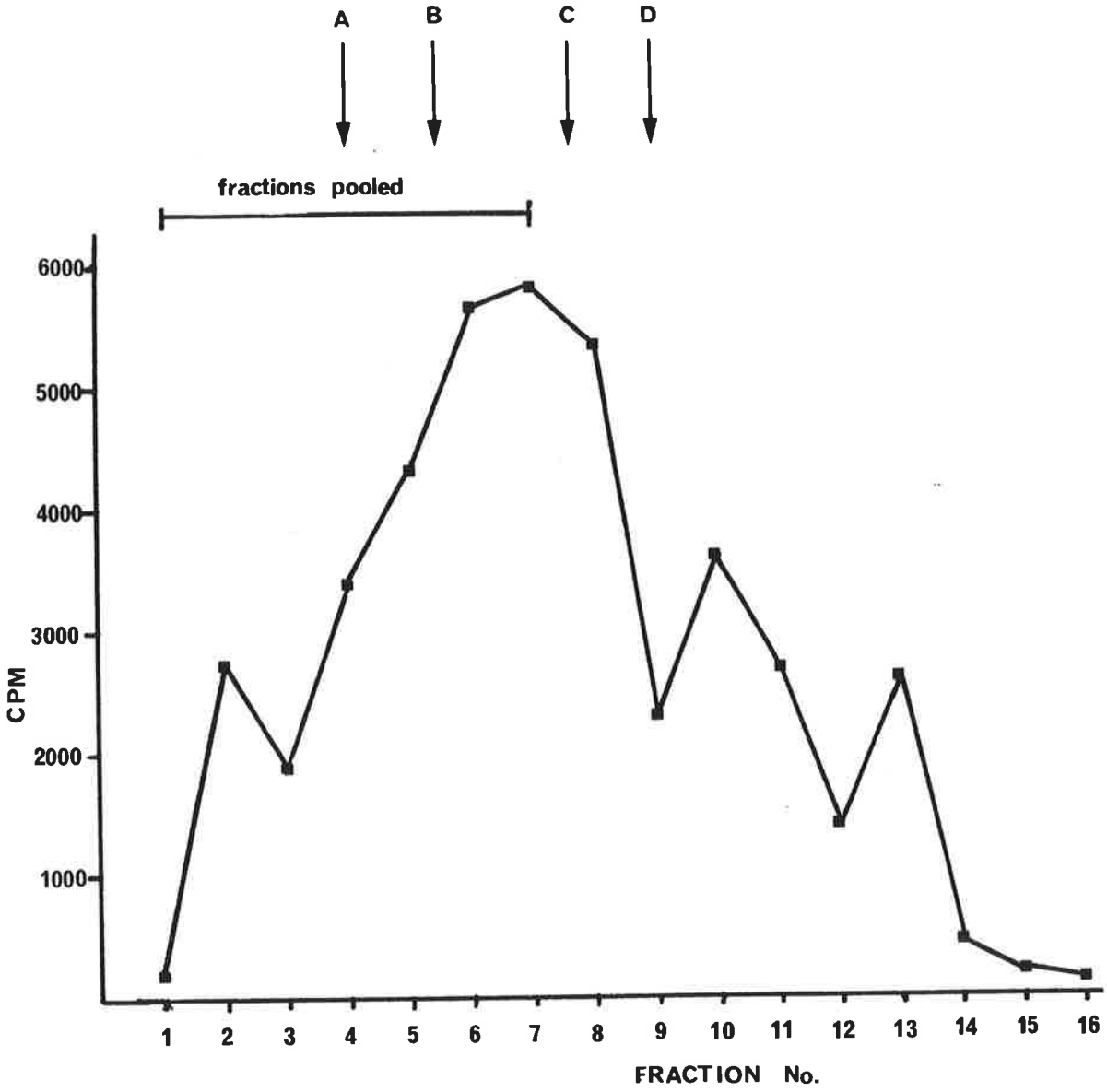
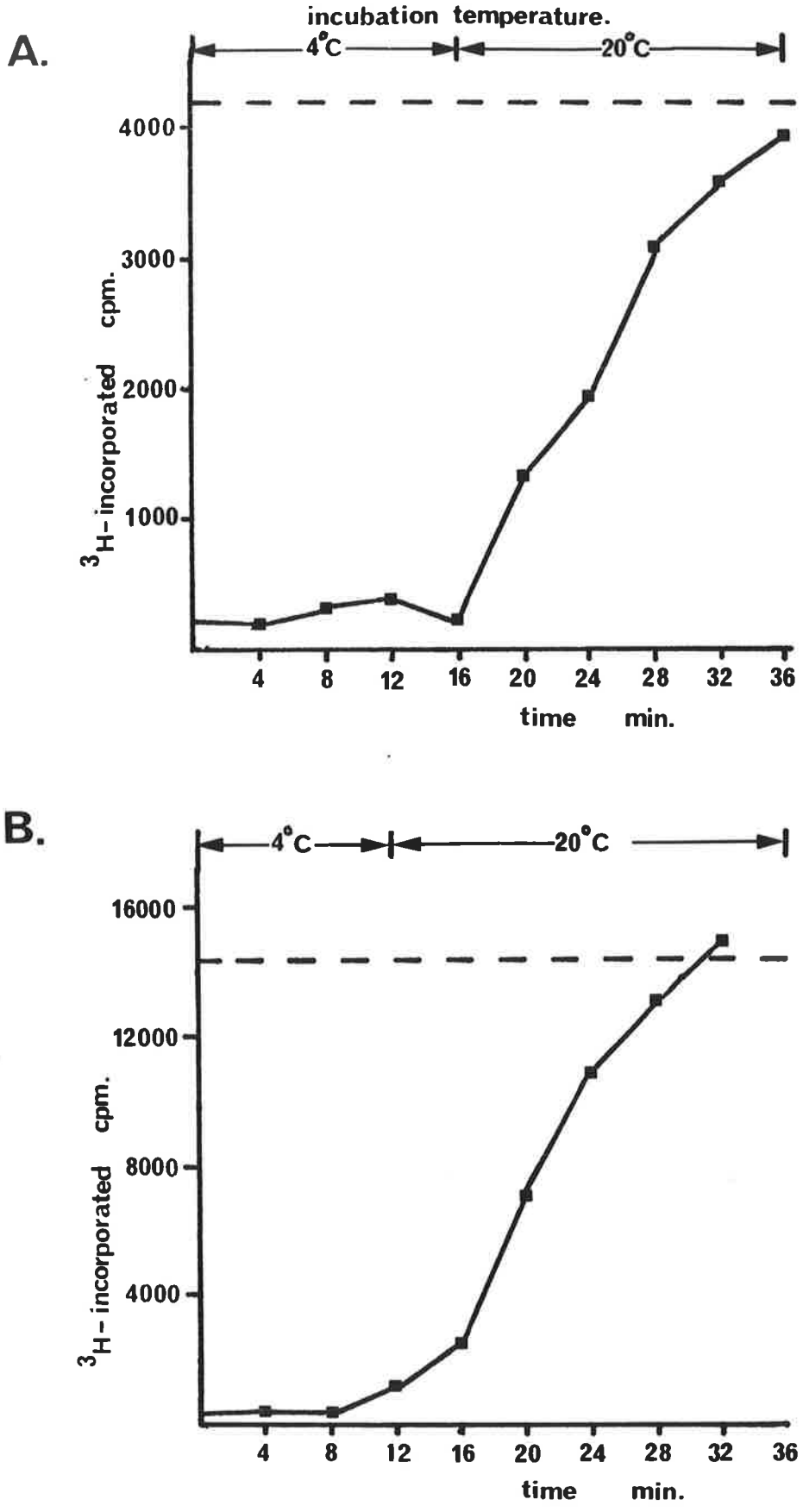

A
B
C
D
fractions pooled
6000
5000
4000
3000
2000
1000
CPM
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
FRACTION No.

Figure 4.3

Homopolymer tailing of DNA. The DNA, either reticulocyte ds cDNA or Pst I cleaved pBR 322 was incubated with terminal deoxynucleotidyl transferase and ^{3}H-dCTP or ^{3}H-dGTP. 1µl samples of the reaction mix were taken and the amount of acid insoluble radioactively determined. The broken line indicates the number of counts incorporated equivalent to the addition of about 20 residues per 3' end of the molecules.

A. poly-dC tailing of ds cDNA

B. poly-dG tailing of Pst I cleaved pBR 322.


A.
incubation temperature.
4°C
20°C
4000
3000
2000
1000
^{3}H- incorporated cpm.
4
8
12
16
20
24
28
32
36
time min.
B.
4°C
20°C
16000
12000
8000
4000
^{3}H- incorporated cpm.
4
8
12
16
20
24
28
32
36
time min.

with PstI to ensure complete cleavage. Deoxyguanosine tails were added and the reaction was stopped when it was calculated that an average of 20 residues per end had been added (Fig. 4.3b).

4.2.4 Annealing and Transformation

Approximately equimolar amounts of dG-tailed pBR 322 and dC-tailed ds cDNA were annealed using a high salt hybridisation solution to stabilise G-C base pairing. The resultant circular molecules were used to transform competent *E. coli* cells.

Control experiments with pBR 322 gave a transformation efficiency of 2×10^7 colonies per microgram. Under the same conditions, the PstI digested, dG-tailed pBR 322 transformed with an efficiency of 2×10^5 colonies per microgram after self-ligation. When 2 ng of dG-tailed pBR 322 was annealed to dC-tailed ds cDNA approximately 1,000 *E. coli* transformants resulted. This was three times the level of background transformation, and equivalent to approximately 4×10^5 transformants per microgram of ds cDNA.

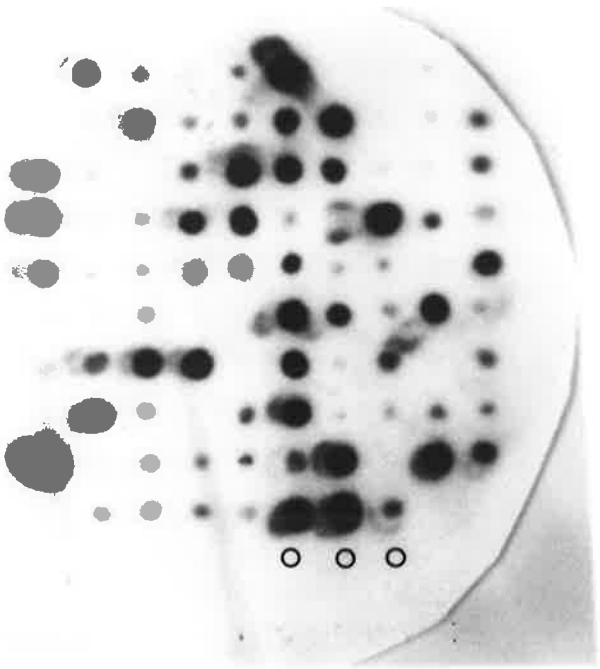
4.2.5 Screening of Transformants for Recombinant Plasmids

a) Antibiotic resistance

The plasmid pBR 322 carries genes for resistance to the antibiotics ampicillin and tetracycline (see Sutcliffe, 1978). The single PstI site of pBR 322 is located within the gene for ampicillin resistance (Ap^R). Insertion of DNA into the PstI site of pBR 322 results in the inactivation of the ampicillin gene. The transformants described in the previous section were isolated on agar plates containing tetracycline, and therefore all bacterial colonies were tetracycline resistant (Tet^R) and contained plasmid. Five hundred colonies were toothpicked in duplicate to tetracycline and ampicillin agar plates and approximately 200 Ap^S,Tet^R colonies were identified.

Figure 4.4

Detection of 10S cDNA sequences amongst reticulocyte recombinants obtained using the colony screening procedure. Colonies which were Tet^R, Ap^S were denatured *in situ*, immobilised on nitrocellulose paper, annealed to ^{32}P labelled 10S cDNA, washed in 1 x SSC / 0.5% SDS at 65°C, dried and autoradiographed. The three colonies circled were transformed with wild type pBR 322.



b) Identification of recombinants by hybridisation to 10S cDNA

The two hundred colonies identified as Ap^S,Tet^R were grown on nitrocellulose filters as described by Grunstein and Hogness (1975). The 10S poly-A^+ RNA described in Chapter 3 was used as template for the production of radioactively labelled cDNA. The cDNA was then used to screen the Ap^S,Tet^R colonies. A typical autoradiograph (Fig. 4.4) demonstrated hybridisation of 10S cDNA to greater than 70% of all colonies at levels significantly above the pBR 322 control colonies.

4.3 DISCUSSION

The production of ds cDNA from native cat poly-A RNA gave rise to molecules of two main size classes when analysed by alkaline agarose electrophoresis. One band, representing a large percentage of the total ds cDNA molecules, corresponds to a molecular weight of approximately 600 b.p., which is within the expected size range for globin ds cDNA molecules. Another, less intense, band could be discerned corresponding to a molecular weight of approximately 900 b.p. It is possible that the 900 b.p. band represents ds cDNA synthesised from mRNA coding for the 25 kd "unknown" protein, which has been shown to be present in the poly-A^+ RNA fraction (see section 3.2.2). Although the aim of these experiments was to isolate native cat globin cDNA recombinants, the poly-A^+ RNA fraction was chosen as the starting material rather than 10S RNA because the cloning of reticulocyte RNA species other than those encoding globin may have been of interest for further study. Treatment with S_1 nuclease was shown to cleave to 5' hairpin loop of the ds cDNA molecules, although the overall size of the cDNA molecules after treatment appeared to be smaller than was

initially observed with the first strand cDNA synthesis (see Fig. 4.1). Due to the limited amount of material available for these experiments, an investigation of the optimal amount of S_1 nuclease required for successful cleavage and minimal degradation of native cat ds cDNA molecules was not undertaken.

Considerable difficulty was encountered in the "tailing" procedure. The system was initially optimised using pBR 322 DNA digested with the restriction endonuclease Hha1, which cuts pBR 322 21 times thus providing 62 "ends" per molecule available for tailing. It was found that only one batch of TdT from one of two commercial suppliers was able to catalyse the addition of dG tails on to Hha1 digested pBR 322. It was also found that the batch of cacodylate used in the enzyme buffer affected the efficiency of tailing. Wu *et al.* (1981) suggest that it is necessary to add the $CoCl_2$ and the TdT simultaneously to the reaction mixture. When the $CoCl_2$ and TdT was not added under the conditions specified by Wu *et al.* (1981) the tailing of Hha1 digested pBR 322 had been established for the addition of homopolymer tails to double stranded DNA, ds cDNA molecules were successfully tailed using tritiated dCTP as substrate. The number of dG:dC residues required for optimal annealing and transformation efficiency have been determined by Peacock *et al.* (1981). In general, it was found that the number of residues on the plasmid and inserted DNA should be equal, with approximately 20 residues being added to each 3' end of the DNA for maximum efficiency. Accordingly, the tailing reactions detailed in Figure 4.3 were stopped when an estimated 20 residues were added to each 3' end.

The PstI digested DNA used for the tailing reaction was digested twice and an aliquot analysed by agarose gel electrophoresis to ensure that digestion had proceeded to completion. A high background of

non-recombinant transformants was observed, presumably due to a small amount of undigested pBR 322 and the recircularisation of plasmids which did not undergo tailing.

Of the ampicillin sensitive colonies recovered after transformation, approximately 30% failed to hybridise to 10S cDNA. It is likely that some of the non-hybridising colonies represent clones of mRNA species not present in 10S RNA, including mRNA encoding the 25 kd "unknown" protein. The annealing of tailed vector DNA molecules in the absence of ds cDNA molecules resulted in a proportion of non-recombinant transformants, of which approximately 15% were Ap^S, presumably due to small deletions in the ampicillin gene resulting from nuclease contamination. The nature of the non-hybridising clones has not been directly investigated.

The results presented in this chapter demonstrated that clones containing sequences complementary to 10S RNA have been obtained. Fifty such clones were selected for further study as detailed in Chapter 5.

CHAPTER 5. IDENTIFICATION AND CHARACTERISATION OF NATIVE CAT GLOBIN cDNA CLONES

5.1 INTRODUCTION

When a protocol similar to that described in the previous chapters was used to isolate human α- and β-globin cDNA clones (Wilson *et al.*, 1977) and *X. laevis* α- and β-globin cDNA clones (Kay *et al.*, 1980), most of the recombinants were found to fall into one of two classes of restriction endonuclease pattern, indicating the presence of two distinct species of cDNA clone. However, evidence from restriction enzyme analysis alone cannot confirm the identity of a cloned insert. Hybrid arrest translation (HART) has been used to identify the nature of cloned DNA sequences (Paterson *et al.*, 1977). Using the HART procedure, plasmid DNA containing a cDNA or DNA insert is hybridised to a limiting amount of mRNA. If the protein coding sequence contained in the cDNA insert is represented in the mRNA, stable/cDNA hybrids will form. When the mRNA/cDNA mixture is subsequently used to direct protein synthesis in a cell free translation system, only unbound mRNA will be translated. In this way, the absence of a particular translation product can be correlated with a particular cDNA clone, hence identifying that cDNA clone as encoding the absent translation product. The HART procedure has also been used to identify restriction fragments of viral DNA containing protein coding sequences (Paterson *et al.*, 1977). A modification of the HART procedure is the hybrid release translation in which the plasmid DNA is bound to a solid support and any hybridising mRNA is subsequently eluted and translated. The latter system has a distinct advantage over the HART procedure when the initial RNA being assayed consists of a mixture of many mRNA species,

leading to a complex pattern of translation products. The hybrid release procedure has been used either on its own or in combination with various antibody techniques to detect cDNA clones coding for rare (<0.5% of total mRNA) mRNA species (Parnes *et al.*, 1981).

The only direct evidence on the composition of a cDNA insert is the determination of its nucleotide sequence. The "dideoxy" chain termination DNA sequencing method of Sanger *et al.* (1977) provides a rapid and accurate method for DNA sequencing. It makes use of the ability of DNA polymerase to faithfully synthesise a complementary radioactive copy of a single stranded DNA template using a short DNA fragment as primer, and to incorporate at random a "dideoxy" analogue of each deoxynucleoside triphosphate into the newly synthesised DNA chain. Once the analogue is incorporated, the absence of a 3' hydroxyl group prevents the formation of a phosphodiester linkage, and the growing DNA chain is terminated.

The advantages of this method over other rapid DNA sequencing techniques lies in the high degree of base specificity obtained, the easily controllable enzyme reaction employed and the significantly lower amounts of radioactivity required. In addition, fewer manipulations are needed and there are less hazardous chemicals to deal with in the protocols.

Two disadvantages have limited the wide acceptance of this procedure; the requirement that the DNA be available in a single-stranded form and that for every 200-300 bases of sequence a specific primer complementary to the 3' end must be prepared. Recent genetic manipulation of bacteriophage M13 (Messing *et al.*, 1981) and the construction of a "universal" primer (Anderson *et al.*, 1980) have overcome these problems.

M13 is an *E. coli* male specific single-stranded filamentous DNA phage. Through genetic engineering the DNA has been specifically altered to permit cloning of fragments (Messing *et al.*, 1981).

The resultant phage has several characteristics essential for a good cloning vehicle. The intracellular double-stranded replicative form (RF) has:-

1) a small size (just over 7 kb);
2) a high copy number (about 200 molecules per cell);
3) multiple clustered cloning sites including EcoRI, BamHI (Sau3A), SalI, PstI, HindIII, AccI, and the blunt ended sites SmaI and HincII;
4) insertional activation of the β-galactosidase gene as indicated by a change from blue plaque to colourless plaque formation.

Apart from the double-stranded RF, the phage itself contains a single stranded DNA form which is packaged into a protein coat and extruded out of the cell without lysis. If foreign DNA is cloned directly into the RF, then this biological process serves to separate the complementary DNA strands. About 5-10 μg of pure single-stranded cloned DNA can be obtained per ml of culture.

Thus, DNA to be sequenced is cloned into one of the unique clustered cloning sites of the M13 RF DNA. After infection, the viral DNA containing the insert is readily isolated, thereby satisfying the requirement for a single stranded template. Furthermore, cloning into the same specific region obviates the need for isolation of many different primers, since a single primer can be used for all inserts.

Having determined the nucleotide sequence of an mRNA molecule it is then possible to examine the primary structure of that molecule,

and compare it to other eukaryotic mRNAs. Through such comparisons it is possible to observe the conservation of certain primary structures within the molecule which may play some role in its synthesis and/or expression.

A functionally active eukaryotic mRNA molecule is the product of a series of steps that include the transcription of an mRNA precursor from DNA, the processing of primary transcripts within the nucleus, and the subsequent transport of mature mRNA molecules into the cytoplasm and their association with ribosomes to initiate peptide synthesis (see Baralle, 1983; Williamson, 1980). The mRNA molecule can be divided into three domains which are, from 5' to 3', the leader, coding and trailer sequences. The leader and trailer regions contain sequences copied from the genomic DNA and features added post-transcriptionally.

Whereas prokaryote mRNA sequences contain a purine rich ribosome binding site adjacent to the AUG intitiation colon (Shine and Dalgarno, 1975), there is no obvious counterpart in eukaryotes (Baralle and Brownlee, 1978). The 5'-terminus of eukaryote mRNA is post-transcriptionally modified by the addition of a 7-methyl-guanosine "cap" structure (Furucchi *et al.*, 1975; Adams and Cory, 1975). This cap appears to be all that is essential for ribosome binding and efficient translation (Rosenberg and Peterson, 1979). The length of the 5'-untranslated region has an effect on the efficient translation of eukaryotic mRNAs (Rosenberg and Peterson, 1979). The general picture that emerges comparing cellular and viral 5' nontranslated regions is that of a wide variation. The length varies from just a few nucleotides, such as 3 for the immunoglobulin kappa chain (Hamlyn *et al.*, 1981) to more than 250 for mouse liver α-amylase (Hagenbuchle *et al.*, 1981). No particular sequence appears to be

universally conserved in the 5' non-coding region and the only recognisable signal features are the cap, the initiation codon AUG, and perhaps a limited complementarity to region near the 3' terminus of 18S rRNA (Tsujimoto and Suzuki, 1979). Baralle (1983) has examined the 5' non-coding regions of some vertebrate β-globin genes and found considerable homology between species, with both length and nucleotide sequence showing significant conservation (Fig. 5.1). The nucleotide sequence CTTPyTG which is evident in the 5' non-coding of vertebrate β-globin genes several nucleotides from the capping site has also found to be common to the 5' regions of several non-globin, eukaryotic mRNA molecules (Baralle and Brownlee, 1978).

The 3' non-coding region extends from the termination codon to the poly-A addition site. The length of the 3' non-coding region varies widely in eukaryotic cellular mRNA molecules. In chicken ovalbumin a 637 nucleotide-long region follows the coded messenger (McReynolds *et al.*, 1978), while the 3' non-coding region of mouse α-amylase mRNAs is only 30 nucleotides long, being considerably shorter than its 5' counterpart (Hagenbuchle *et al.*, 1981). The nucleotide sequence of the 3' non-coding regions of different eukaryotic mRNAs shows very little homology with the exception of short segments in the region near the poly-A addition sites. The hexanucleotide AAUAAA, first noted by Proudfoot and Brownlee (1976), precedes by 11 to 30 bases the poly-A addition site in most eukaryotic mRNA molecules (see Baralle, 1983) and is most likely to function in the processing and/or polyadenylation of the mRNA. Exceptions to this rule have been found in some viruses (Porter *et al.*, 1978). Variants of the AAUAAA motif have been found in rat amylase (McDonald *et al.*, 1979), anglerfish somatostatin (Hobart *et al.*, 1980), and some leukocyte interferon mRNA molecules (Goeddel *et al.*, 1981). As is the case for 5' non-coding

regions, it appears that some homology in 3' non-coding regions can be demonstrated between related genes, but apart from a small region involving the putative polyadenylation signal there does not appear to be any other sequences common to most eukaryotic 3' non-coding regions (Efstratiadis *et al.*, 1977; Baralle, 1983).

The function of the 3' non-coding region of eukaryotic mRNAs has not been clearly defined. Sequence comparisons have suggested that the particular sequences which comprise the 3' non-coding regions are not essential for mRNA function. Kronenberg *et al.* (1979) were unable to observe any effects of *in vitro* translation of rabbit β-globin mRNA when the 3' non-coding region was removed. The apparent redundancy of portions of the 3' region is demonstrated by the existence of multiple mRNAs encoding mouse dihydrofolate reductase (DHFR) which differ only in the length of their 3' non-coding regions (Nunberg *et al.*, 1980). Setzer *et al.* (1980) were unable to find any functional difference in the four polysomal DHFR mRNAs. They appear to produce identical proteins and their relative proportions are constant at various points in cellular growth, and they are of equal stability. It seems likely that redundant polyadenylation and/or transcription termination signals exist at the 3' end of the DHFR gene and may be recognised with varying degrees of efficiency, giving rise to multiple transcription products (Setzer *et al.*, 1980). Hence most of the 3' non-coding region in the longer DHFR mRNAs does not significantly affect their translational function. It has been suggested that the 3' non-coding region and the poly-A tail may have a role in the nuclear processing of mRNA precursor and/or transport between the nucleus and the cytoplasm. It appears certain that the poly-A tail plays some role in promoting mRNA stability (Gurdon, 1976; Huez *et al.*, 1978), but the role of putative polyadenylation signals and their relationship with the

surrounding sequences is yet to be defined.

This chapter describes the restriction analysis of a number of ds cDNA clones which hybridise to native cat reticulocyte 10S RNA, and the subsequent selection of two clones pDG 73 and pDG 5 for further analysis in hybrid arrest translation experiments. The complete nucleotide sequence of the pDG 73 and pDG 5 inserts is presented, confirming their identity as native cat α- and β-globin cDNA clones, respectively. The partial sequence of another β-globin cDNA clone, pDG 77, is also described. The structure of the native cat α- and β-globin mRNAs is compared to that of other eukaryotic, and in particular, globin mRNAs.

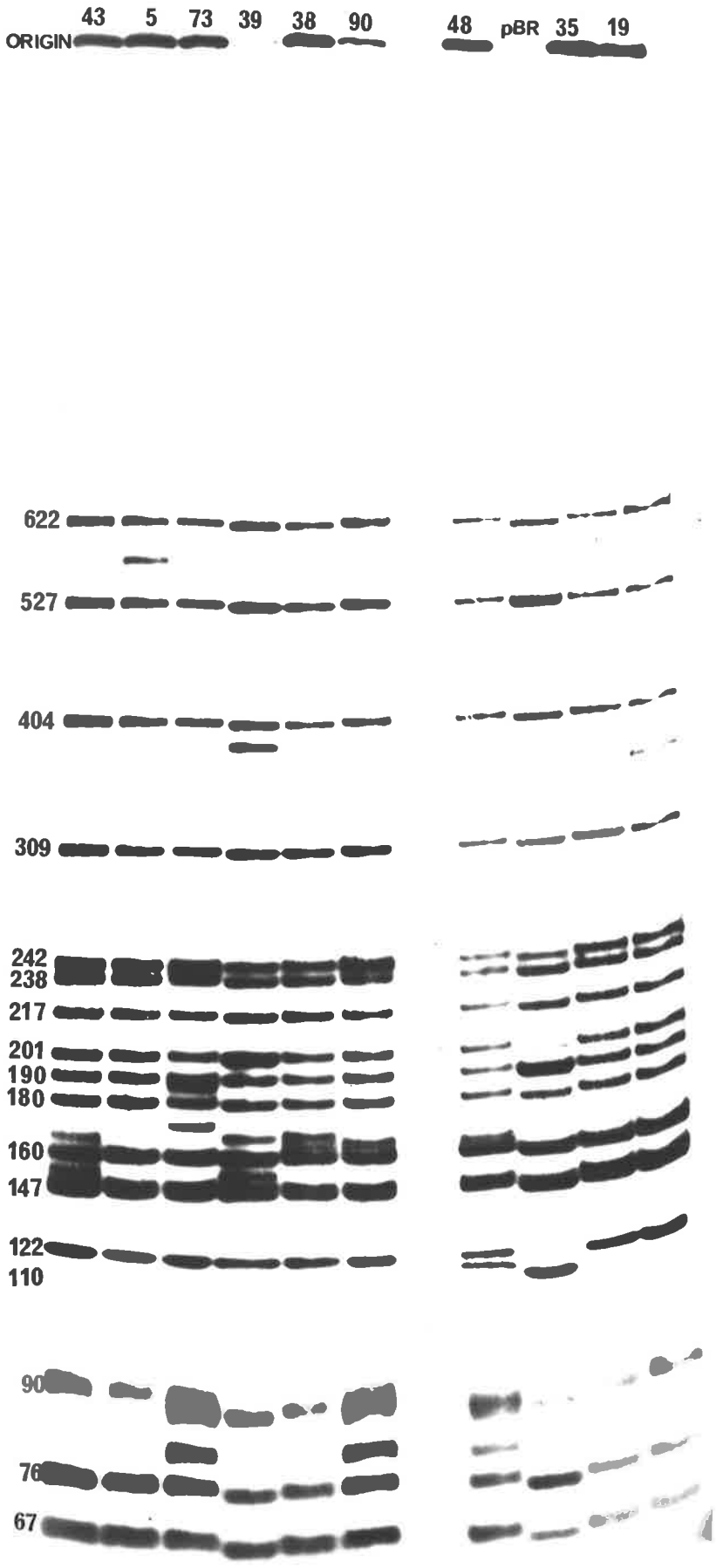
5.2 RESULTS

5.2.1 Restriction Analysis of 10S ds cDNA Clones

Fifty of the most strongly hybridising clones were selected as detailed in section 4.2.5. The plasmid DNA from 1 ml of saturated culture was isolated, digested with the restriction endonuclease HpaII, end labelled using the Klenow fragment of *E. coli* DNA polymerase I ("Klenow"), and analysed on 6% aqueous acrylamide gels. An example of such an analysis is shown in Figure 5.1 where it can be seen that the cDNA clones appear to fall into one of two classes based upon their HpaII restriction pattern. One group of clones, typified by pDG 5, have no HpaII recognition sites within the inserted ds cDNA. This leads to the production of one "extra" band when compared to the HpaII digest of pBR 322. The other group of clones, typified by pDG 73 have several HpaII sites contained in the inserted cDNA. All clones of this type have at least two or three bands of molecular weights 210 b.p., 85 b.p. and 80 b.p. in common. It was found that all clones lacked the 110 b.p. HpaII fragment of pBR 322. This is expected

Figure 5.1

Digestion of recombinants with Hpa II. Approximately 0.5µg of plasmid DNA from 1ml overnight cultures was digested with Hpa II, end labelled with α-^{32}P dCTP and fractionated on a 6% polyacrylamide gel. Hpa II digested pBR 322 was also electrophoresed to show which of the fragments were derived from the vector molecule. Plasmid designations are indicated.


ORIGIN
43
5
73
39
38
90
48
pBR
35
19
622
527
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67

since the 110 b.p. HpaII fragment contains the pBR 322 PstI recognition site, and insertion of DNA into this site will alter the size of that fragment.

One of each "class" of clone was selected for further analysis; pDG 5, with an insert length of approximately 500 b.p., and pDG 73 with an insert length of approximately 650 b.p. It was found that all clones contained inserts with sizes between 200 and 650 b.p.

5.2.2 Hybrid Arrest Translation of Selected Recombinants

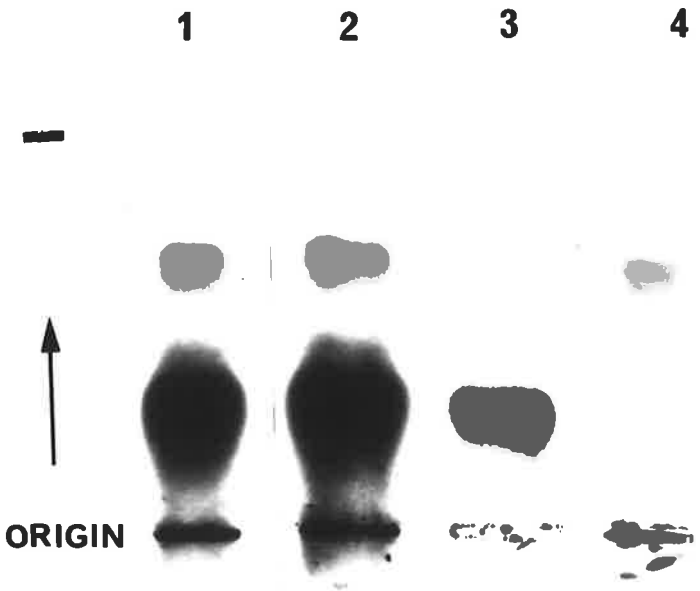
Based upon the findings of Paterson *et al.* (1977) it was decided to hybridise 10 µg of cloned plasmid DNA with 200 ng of 10S RNA. This was approximately a 10-fold molar excess of insert DNA over total RNA. After the hybridisation of plasmid DNA to RNA, the hybridisation mixtures were translated in a wheat germ cell free system using ^{35}S methionine as the labelled amino acid. The translation products were analysed by electrophoresis on cellulose acetate ("Cellogel") as described in section 10.2.6, Cellogel electrophoresis results in the separation of the native cat α- and β-globin chains. Figure 5.2 shows that hybridisation of plasmid pDG 5 to 10S RNA resulted in the translational inhibition of a specific polypeptide, as did hybridisation with plasmid pDG 73. Globin protein had previously been added to the translation products as carrier and was visualised by staining with Coomassie blue R250. A comparison of the stained gel with the autoradiograph of that gel showed that the polypeptide inhibited by pDG 5 corresponded to putative native cat β-globin (see Chapter 2). Similarly, pDG 73 was found to inhibit the translation of native cat α-globin mRNA. An area of the stained cellogel corresponding to the native cat globin polypeptides was excised and its radioactivity determined by liquid scintillation counting. The

Figure 5.2

Hybrid arrest translations of plasmids, pBR 322, pDG 5, and pDG 73 with reticulocyte RNA. Plasmid DNA was hybridised in excess to reticulocyte RNA and the DNA/RNA mixture translated in a wheat germ cell-free system in the presence of ^{35}S-methionine. ^{35}S-labelled translation products were visualised by autoradiography after electrophoresis on cellulose acetate.

The translation products are derived from:-

1. RNA with no added plasmid DNA.
2. pBR 322-hybridised RNA.
3. pDG 73-hybridised RNA.
4. pDG 5-hybridised RNA.


1
2
3
4
ORIGIN

results presented in Table 5.1 show clear inhibition of the synthesis of specific globin proteins with pDG 5 and pDG 73 but not with pBR 322. Due to the small amount of RNA available for these experiments it was decided not to split each hybridisation mixture and subsequently dissociate the DNA:RNA hybrids of part of the hybridisation mixture as an added control experiment. Such a procedure effectively reduces the amount of RNA available for translation and hence decreases the total amount of translated protein available for electrophoresis and subsequent autoradiography. Even in the absence of this control experiment it appears that the inhibition of mRNA translation observed in these experiments was specific to particular recombinant plasmids, and not an artefact of the hybridisation protocol.

Table 5.1

Quantitation of polypeptide synthesis in a cell free system after hybrid arrest translation.

Hybridising DNA	α-globin (cpm)	β-globin (cpm)	β/α ratio
none	2,795	7,808	2.79
pBR 322	1,077	3,134	2.91
pDG 73	410	2,760	6.73
pDG 5	1,482	248	0.17

5.2.3 Preparation of Bacteriophage M13 Vectors

About 1 milligram of M13 mp 93 replicative factor (RF) DNA was prepared from 1 litre of culture using a modified alkali lysis procedure, followed by CsCl centrifugation, as described in section

10.2.14. Cloning vehicles were prepared from the purified RF DNA.

3 µg of M13 mp 93 RF DNA was digested with a 5-fold excess of the appropriate restriction endonuclease. Three vectors were prepared by digestion with BamHI, AccI and SmaI, respectively. After digestion the DNA was diphosphorylated with calf intestinal phophatase, phenol extracted and precipitated with ethanol. The digested DNA was run on a 0.7% low gelling temperature agarose gel and the linearised RF extracted. These DNA preparations typically had a transformation efficiency of 2×10^2 transformants per microgram, which was 5×10^3 times lower than the transformation efficiency of RF. These DNAs were used as vectors in subsequent cloning experiments.

5.2.4 Preparation of DNA fragments for Cloning

Plasmid pDG 5 and pDG 73 DNAs were prepared by a modified alkali lysis procedure described in section 10.2.12. Each plasmid was digested with a range of restriction endonucleases including BamHI, HaeIII, Sau3A, RsaI, EcoRI and HpaII. The aim of this experiment was to generate a range of overlapping, independently derived, DNA fragments containing the ds cDNA inserted into the PstI site of the plasmids. Digestion products were electrophoresed on aqueous acrylamide gels with appropriate digests of the parental plasmid pBR 322 (Figs 5.3 and 5.4). Bands containing insert sequences were excised from the gel and the DNA eluted.

Only restriction fragments which contained at least one recognition site within the ds cDNA insert were eluted. Restriction endonucleases which do not have a recognition site within the ds cDNA insert result in the generation of insert DNA fragments bounded on both ends by dG:dC tails. Such fragments are unsuitable for dideoxy nucleotide sequencing.

Figure 5.3

Digestion of plasmid pDG 73 with Sau 3A, Hpa II, Hae III, and Bam HI. Sau 3A, Hpa II, and Hae III plasmid digests were electrophoresed on 6% polyacrylamide gels and visualised by staining with ethidium bromide. Bam HI digests were electrophoresed on 0.8% agarose gels.

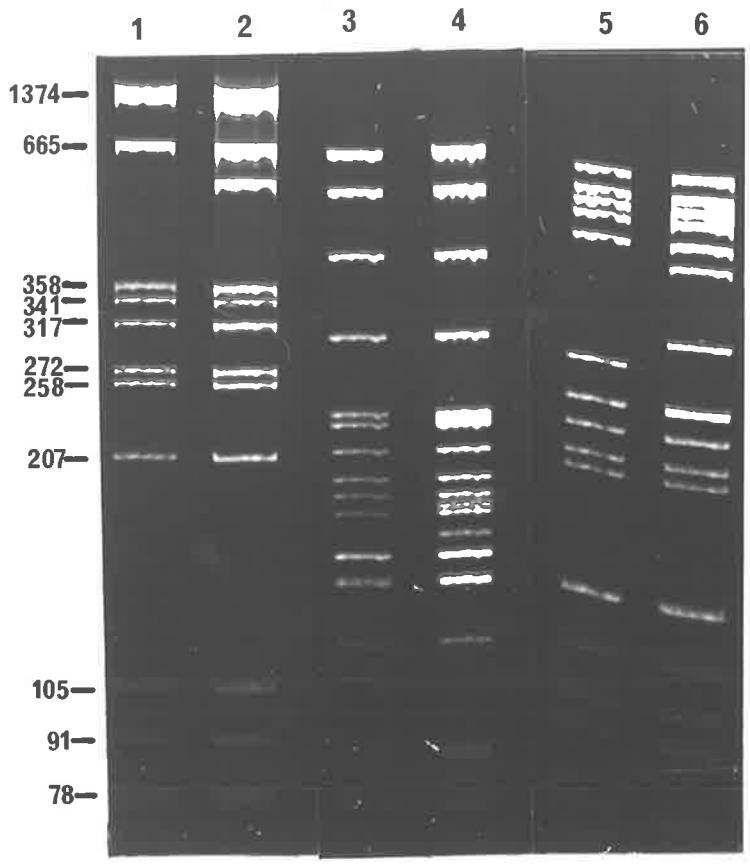
A) 1. Sau 3A digest of pBR 322
2. Sau 3A digest of pDG 73
3. Hpa II digest of pBR 322
4. Hpa II digest of pDG 73
5. Hae III digest of pBR 322
6. Hae III digest of pDG 73

molecular weights are in base pairs.

b) 1. Bam HI digest of pBR 322
2. Bam HI digest of pDG 73

molecular weights are in kilobase pairs.

A.



B.

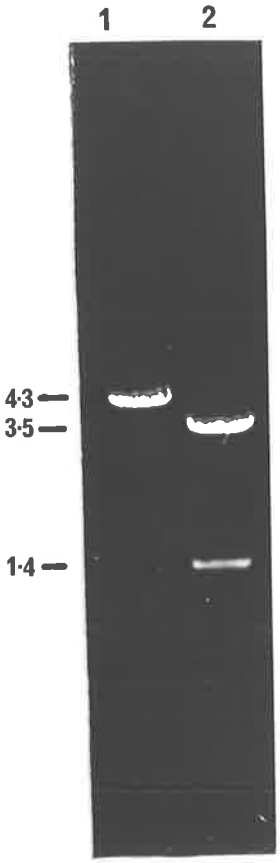


Figure 5.4

Digestion of plasmids pDG 5 and pDG 77 with Rsa I and Sau 3A. Digests were electrophoresed on 6% polyacrylamide gels and visualised by staining with ethidium bromide.

A) 1. Rsa I digest of pDG 77

2. Rsa I digest of pDG 5

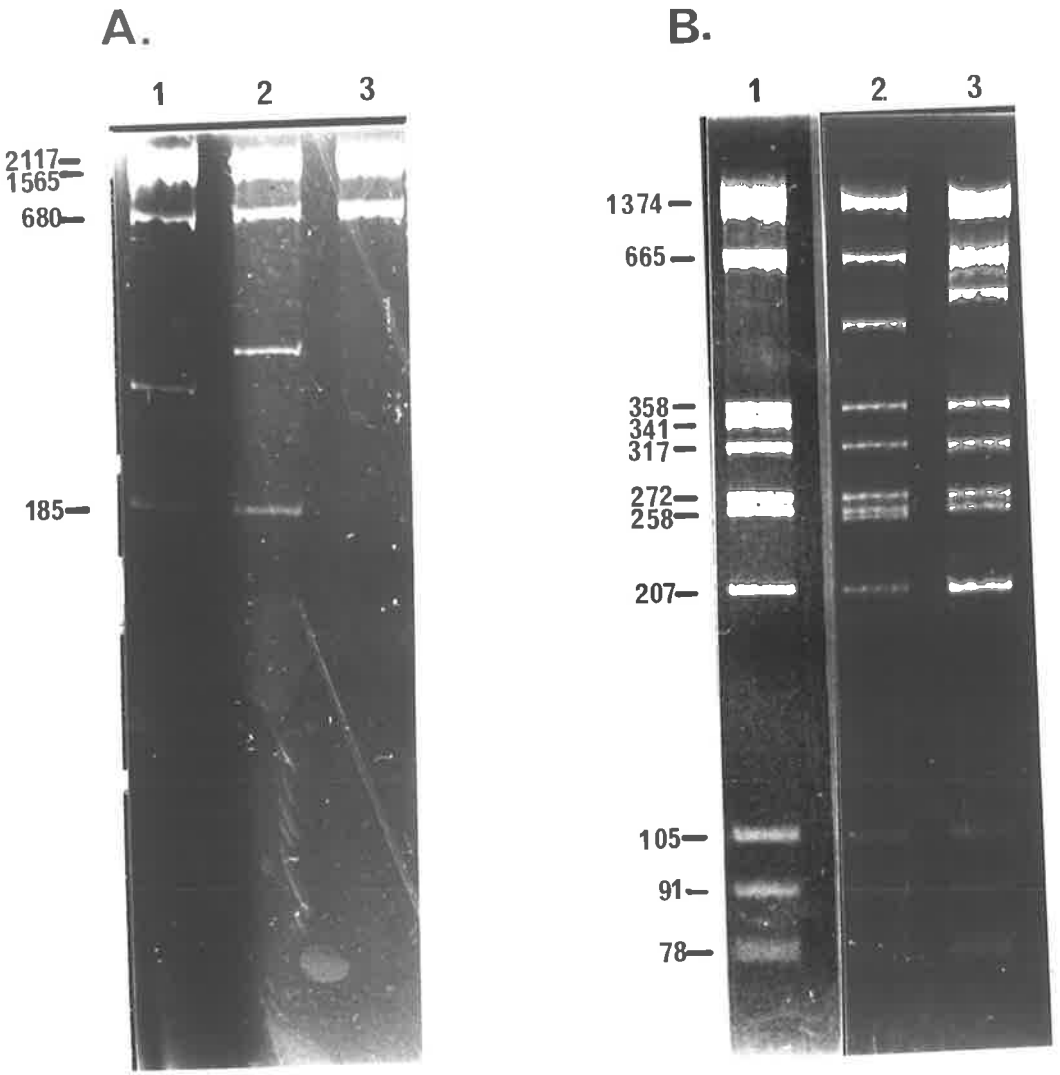
3. Rsa I digest of pBR 322.

B) 1. Sau 3A digest of pBR 322

2. Sau 3A digest of pDG 5

3. Sau 3A digest of pDG 77.

molecular weights are in base pairs.


A.
1
2
3
2117
1565
680
185
B.
1
2
3
1374
665
358
341
317
272
258
207
105
91
78

5.2.5 Ligation and Transformation

Equimolar amounts of the appropriate M13 vector and each restriction fragment were covalently joined using T_4 DNA ligase. The resultant recombinant molecules were used to transform competent *E. coli*.

When four nanograms of ligated DNA was used to transform host *E. coli* cells, between 1 and 50 clear (recombinant) plaques and from 1-5 blue (non-recombinant) plaques resulted. A 2-fold lower production of recombinants was observed when DNA fragments with blunt ends (e.g. HaeIII, RsaI digests) were used, compared to "sticky" ends such as those generated by BamHI and HpaII digests.

5.2.6 Preparation of Templates for DNA Sequencing

As large amounts of single stranded M13 DNA can be made from small cultures, 10-20 templates were prepared at a time.

Recombinant plaques were toothpicked into 1 ml of culture broth and grown for 6 hours. After removal of bacteria, M13 phage were precipitated from the supernatant with polyethylene glycol and the DNA purified as described in section 10.2.16.

DNA Sequencing Reactions

Up to six templates were sequenced at one time. Single stranded M13 template and primer were denatured by boiling and annealed at room temperature. Samples of annealed template-primer DNA were separated into four aliquots and DNA synthesis was carried out in the presence of the four deoxynucleoside triphosphates one of which was ^{32}P-labelled and in turn with each of the four dideoxynuceloside triphophates in four separate reactions. Each reaction gives a population of partially synthesised DNA molecules, each radioactive

and each having a common 5'-end, but each varying in length to a base specific 3' end. Parallel size fractionation of the four DNA synthesis reaction products by denaturing gel electrophoresis and autoradiography allows a sequence to be deduced. Even using very small amounts of ^{32}P-label (1 microcurie/reaction) at low specific activity (400-600 ci/mmole) enough incorporation was achieved to allow a sequence to be determined after 12 hours autoradiography, without an intensifying screen.

Figure 5.5 shows an autoradiograph of a 6% denaturing 40 cm polyacrylamide gel on which the bromophenol blue tracker dye has been electrophoresed to within 2 cm of the bottom of the gel. In this gel system the bromophenol blue migrates at the same rate as a 35 base fragment of single stranded DNA. The BamHI site in the vector is 29 bases from the 5' end of the primer molecule. Therefore the first few bases read from the autoradiograph correspond to vector DNA sequence and include the vector/insert DNA junction point.

5.2.7 Determination of the Nucleotide Sequence of pDG 73

Fragments generated from BamHI and HpaII digests of pDG 73 DNA were subcloned into mp 93 and their nucleotide sequence determined. Considerable difficulty was encountered in the determination of the nucleotide sequence of subclones which contained the dG:dC tails (Fig. 5.6). The fidelity of the polymerase reaction was severely affected by the dG:dC tails present at each end of the insert molecule. Others (Robins, 1983) have noticed the poor fidelity of the sequencing reactions when a run of cytosine residues is present on the template. When a piece of DNA is cloned into an M13 vector there are two possible orientations of that piece of DNA with respect to the cloning site. When a DNA fragment consists of pBR 322 DNA; a dG:dC tail; and then

Figure 5.5

DNA sequence determination. M13 recombinants containing the Bam HI fragments of pDG 73 (see Fig. 5.3) were used as templates in the enzymic dideoxy sequencing reactions. After polymerisation the samples were denatured and loaded on to a 6% polyacrylamide DNA sequencing gel. The M13 vector/insert DNA junction is indicated.

A. M13 subclone containing the 1.4kb Bam HI fragment of pDG 73.

B. M13 subclone containing the 3.5kb Bam HI fragment of pDG 73.

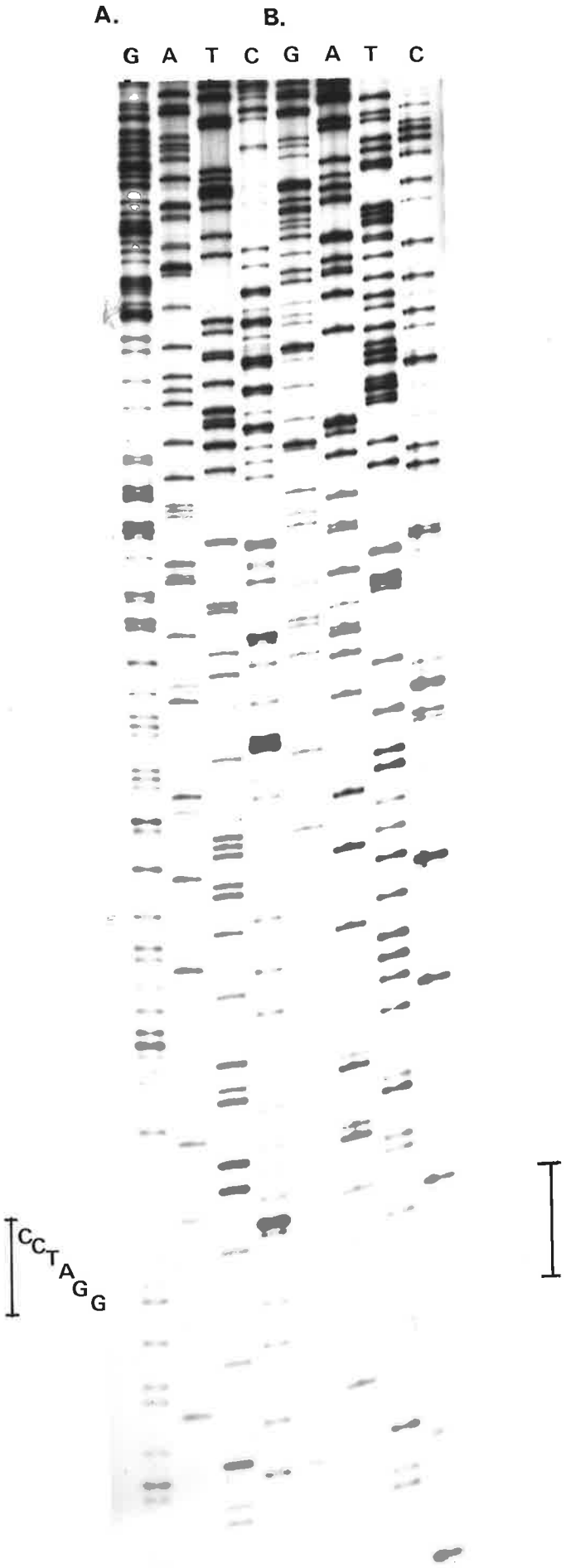

A.
B.
G A T C G A T C
CCTAGG

Figure 5.6

DNA sequence determination of the 350 bp Rsa I DNA fragment of pDG 5. The clone is orientated such that the pBR 322 DNA is adjacent to the priming site. To sequence the pDG 5 insert DNA, the polymerisation reactions must extend through the dG:dC tail. Loss of polymerase fidelity is observed after passage through the dC tail (dG residues read from the gel).
Only the pBR 322 nucleotide sequence is able to be determined.

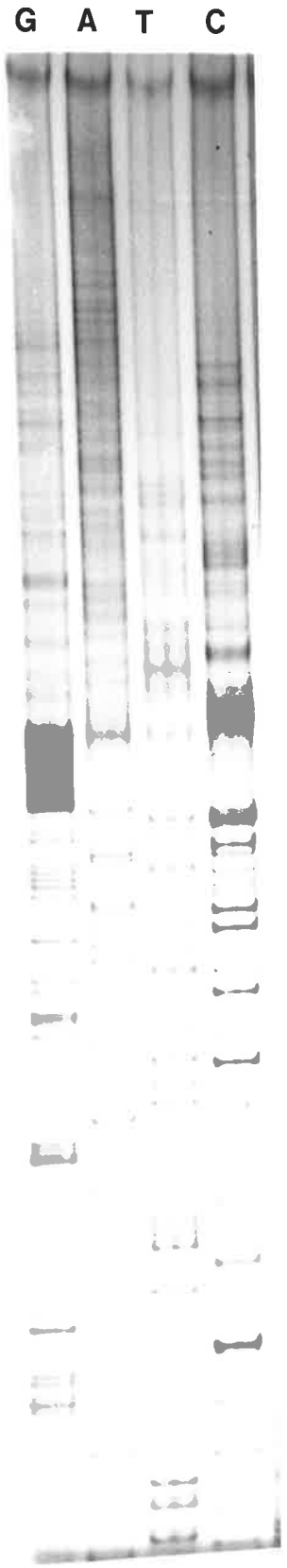

G
A
T
C

Figure 5.7

A. Production of single stranded M13 phage containing opposite strands of an inserted DNA fragment.

B. Clone orientation test.

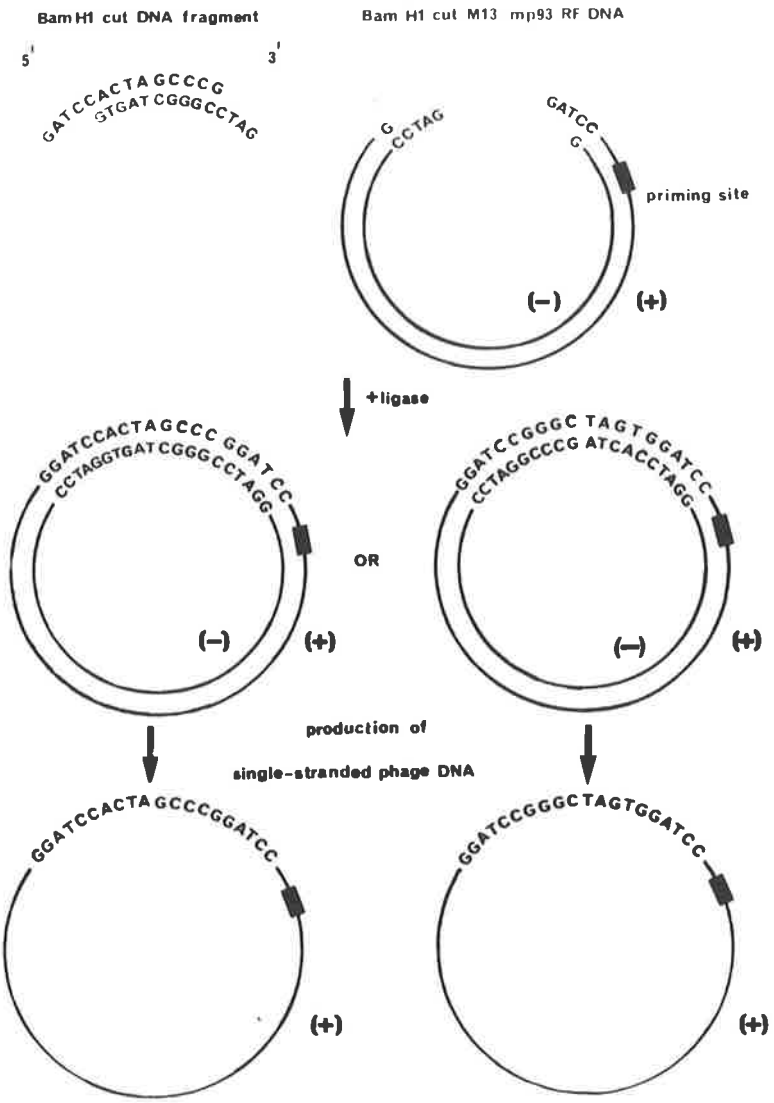
Single strand M13 phage DNA containing the pDG 73 170 bp. Hpa II fragment (see Fig. 5.1) in a known orientation was hybridised to the single stranded DNA of other M13 Hpa II subclones of unknown orientation and electrophoresed on a 0.8% agarose gel.

Samples 2-5 and 8-13 indicate M13 subclones containing the 170 bp Hpa II pDG 73 fragment in the same orientation as the reference clone.

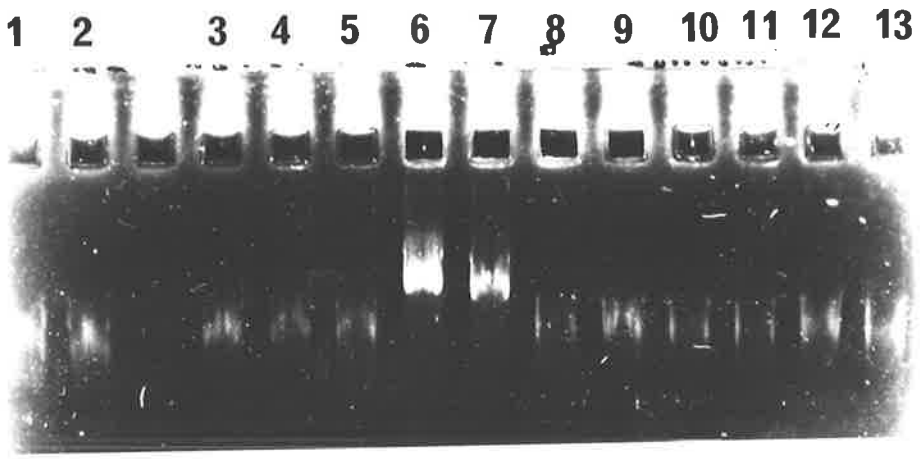
Samples 6 and 7 show hybridisation of the unknown clones to the reference clone and are therefore of an opposite orientation to the reference clone.

Sample 1 is the reference clone alone.

A.


BamH1 cut DNA fragment
5'
3'
GATCCACTAGCCCG
GTGATCGGGCCTAG
Bam H1 cut M13 mp93 RF DNA
G
CCTAG
GATCC
G
priming site
(–)
(+)
+ligase
GGATCCACTAGCCC GGATCC
CCTAGGTGATCGGGCCTAGG
(–)
(+)
OR
GGATCCGGGC TAGTGGATCC
CCTAGGCCCG ATCACCTAGG
(–)
(+)
production of
single-stranded phage DNA
GGATCCACTAGCCCGGATCC
(+)
GGATCCGGGCTAGTGGATCC
(+)

B.


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

the inserted DNA, the nucleotide sequence of the insert can only be determined reliably by dideoxysequencing when the insert DNA is adjacent to the priming site. If the fragment occurs in an orientation such that the pBR 322 DNA is adjacent to the priming site, then the sequencing reaction must proceed through a dC tail before reaching the insert DNA, and passage through the dC tail results in a loss of the sequential nature of the reaction, as seen in Figure 5.6.

Of eight M13 subclones containing a 170 b.p. HpaII fragment of pDG 73, all were found to be in an orientation not conducive to the determination of the nucleotide sequence of the inserted ds cDNA due to the presence of dC tails. A further 20 subclones of this fragment were screened by the "clone orientation test" procedure. In this procedure uncharacterised M13 subclones are hybridised to a previously characterised M13 "reference" subclone. The M13 phage single stranded DNA consists of the positive viral strand and the inserted DNA. If the reference clone and the "test" clone contain insert DNA in opposite orientations, then hybridisation will take place, giving a concatemer with a molecular weight double that of the reference clone or any of the test clones. This procedure is illustrated in Figure 5.7. Of the twenty 170 b.p. HpaII fragment recombinants tested, three were found to be of the opposite orientation to the "reference" clone and were successfully sequenced.

A total of seven different restriction fragments consisting either entirely or partially of inserted ds cDNA were sequenced, with all fragments, other than those containing dG:dC tails, being sequenced at least once in both directions.

Within the pDG 73 nucleotide sequence there is an open reading frame of 423 bases which gives a deduced amino acid sequence which corresponds closely to that of the α-globin chain of the grey kangaroo,

Figure 5.8

The complete nucleotide sequence of the pDG 73 insert DNA. The inferred amino acid sequence is shown directly below.

K. positions where the native cat amino acid sequence differs from that of the grey kangaroo α-globin amino acid sequence;

H. position where the native cat amino acid sequence differs from that of the human α-globin amino acid sequence.

5'

AGCCACCAUG

```
        GUG CUC UCG GAU GCU GAC AAG ACU CAC GUG AAA GCC AUC UGG GGU AAG GUG GGA GGC
        val leu ser asp ala asp lys thr his val lys ala ile trp gly lys val gly gly
K               ala             gly
H               pro                 asn             ala                         ala

        CAC GCC GGU GCC UAC GCA GCU GAA GCU CUU GCC AGA ACC UUC CUC UCC UUC CCC ACU
        his ala gly ala tyr ala ala glu ala leu ala arg thr phe leu ser phe pro thr
K       glu                         gly     glu
H       glu     gly                         glu     met

        ACC AAA ACU UAC UUC CCC CAC UUC GAC CUG UCC CCC GGC UCC GCC CAG AUC CAG GGU
        thr lys thr tyr phe pro his phe asp leu ser pro gly ser ala gln ile gln gly
K                                               his                         ala
H                                               his                 lys

        CAU GGU AAG AAG GUA GCC GAU GCC CUU UCC CAG GCU GUU GCC CAC CUG GAC GAC CUG
        his gly lys lys val ala asp ala leu ser gln ala val ala his leu asp asp leu
K                                   ile gly             gly     ile
H                                       thr asn                 val         met

        CCC GGA ACC CUG UCC AAA CUA AGC GAC CUG CAC GCC CAC AAG CUG AGA GUG GAU CCC
        pro gly thr leu ser lys leu ser asp leu his ala his lys leu arg val asp pro
K
H       asn ala         ala

        GUG AAC UUC AAG CUC CUC UCU CAC UGC CUG AUC GUG ACU CUG GCC GCC CAU CUG AGC
        val asn phe lys leu leu ser his cys leu ile val thr leu ala ala his leu ser
                                            leu         phe                 gly
K                                           leu                             pro
H

        AAG GAU UUG ACU CCC GAA GUG CAC GCC UCC AUG GAC AAG UUC UUU GCC UCU GUG GCU
        lys asp leu thr pro glu val his ala ser met asp lys phe phe ala ser val ala
K       ala                                 leu                 ala
H       glu phe                             leu             leu

        ACC GUG CUG ACC UCG AAG UAC CGU UAAGUUGUGUCGGAAGCCAGGGAUUUCCACUGAGAUUCGGACC
        thr val leu thr ser lys tyr arg
K
H

        GCGAAUCAUCCGGGCCUGCGGUUCCUAGUGGAAUUCAAUCCUCAUCGAUGGAGAUGGAGUUUCGAAUAAAGGUUC

        AAGUUGpolyA(7)
```

3'

Macropus giganteus (Air *et al.*, 1971). Figure 5.8 shows the complete nucleotide sequence of the pDG 73 insert and a comparison of the deduced native cat α-globin amino acid sequence to that of grey kangaroo and human α-globin.

5.2.8 Determination of the Nucleotide Sequence of pDG 5

Plasmid DNA pDG 5 was digested with the restriction endonucleases Sau3A and RsaI (Fig. 5.4) and DNA fragments containing ds cDNA insert isolated, and inserted into the appropriate M13 cloning vector. Similar difficulties in the sequencing of the pDG 5 subclones to those detailed in the previous section for pDG 73 were encountered for particular DNA fragments and overcome in the same manner. Except for DNA fragments containing dG:dC tails, all subclones were sequenced at least once in both directions. The complete nucleotide sequence of pDG 5 is shown in Figure 5.9 and displays an open reading frame of 258 bases which corresponds well to amino acids numbers 60 to 146 of the adult grey kangaroo and human β-globin polypeptides.

5.2.9 Partial Nucleotide Sequence of pDG 77

The native cat reticulocyte cDNA library was screened with the pDG 5 M13 subclone, Rsa1.3, which had previously been shown to contain an internal restriction fragment encoding part of the native cat β-globin polypeptide. Approximately half of the clones screened hybridised to Rsa3.1 probe (Fig. 5.10). One of those clones, pDG 77, was selected for further study. After digestion with restriction endonucleases RsaI and Sau3A (Fig. 5.4) DNA fragments were subcloned into M13 vectors. The partial nucleotide sequence of this clone has been determined from subclone M13 Sau1.7 and is shown in Figure 5.11. It can be seen that the known nucleotide sequence of the pDG 77 insert DNA corresponds exactly to that of a region of the pDG 5 insert except at one position.

Figure 5.9

The complete sequence of the pDG 5 insert DNA. The inferred amino acid sequence is shown directly below.

K. positions where the native cat amino acid sequence differs from that of the grey kangaroo β-globin amino acid sequence;

H. positions where the native cat amino acid sequence differs from that of the human β-globin amino acid sequence.

5'

GUC AGA GCC CAU GGC GCU AAG GUG CUG GUC UCC UUU GGU GAU GCU GUC AAG AAC CUG
val arg ala his gly ala lys val leu val ser phe gly asp ala val lys asn leu

K *leu* *ala* *ile*
H *lys* *lys* *gly ala* *ser* *gly leu ala his*

GAC AAC CUG AAG GGU ACC UUU GCC AAA CUG AGU GAG CUC CAC UGU GAC AAG CUG CAC
asp asn leu lys gly thr phe ala lys leu ser glu leu his cys asp lys leu his

K
H *thr*

GAG GAC CCU GAG AAC UUC AAG CUC CUG GGC AAC AUC CUG GUG AUC UGC CUG GCU GAG
glu asp pro glu asn phe lys leu leu gly asn ile leu val ile cys leu ala glu

K *val* *ile*
H *val* *cys val* *his*

CAU UUU GGC AAA GAA UUC ACC CCU GAG GUU CAG GCU GCC ACC CAG AAG ACU GUG GCU
his phe gly lys glu phe thr pro glu val gln ala ala thr gln lys thr val ala

K *ile asp ser* *val* *leu*
H *pro* *tyr* *val*

GGU GUG GCC AAC GCU CUG GCC CAC AAG UAC CAC UAAACUCCUGCCUCCUCUGGGUCUUCAACCU
gly val ala asn ala leu ala his lys tyr his

K
H

GUUGAAUCCCCUGUUCUCCAUGUUGUCCAUCUUUUGCCAUGGGUGAAUGGGCCCUAUGGCCAUAGCCUUGCCUAG

AAUAAAGGUUCAUUUAUUCCAAUUCpolyA(20)

3'

Figure 5.10

Detection of β-globin-like sequences amongst reticulocyte recombinants. 34 of the most strongly hybridising colonies (see Fig. 4.4) were denatured *in situ*, immobilised on nitrocellulose paper, annealed to a ^{32}P-labelled M13 subclone containing the 185 bp Rsa I fragment of pDG 5. Colonies were plated in duplicate. Clone pDG 77 (arrowed) was selected as containing sequences homologous to the pDG 5 Rsa I fragment M13 DNA subclone.

pBR 322 control colonies are indicated.

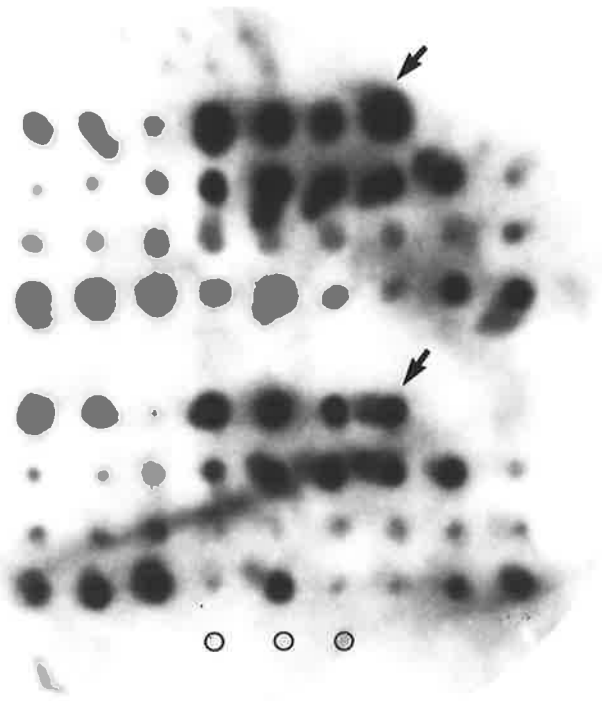


Figure 5.11

Nucleotide sequence of pDG 77 M13 subclone containing the 570 bp Sau 3A fragment (see Fig. 5.4). This clone was detected using the pDG 5 185 bp Rsa I fragment.

The dots indicate regions of the M13 subclone for which the nucleotide sequence was not determined.

Sequence differences detected between pDG 77 and pDG 5 are indicated by asterisks.

```
pDG 5  GAUCUGCCUGGCUGAGCAUUUUGGCAAAGAAUUCACCCCUGAGGUUCAGGCUGCCACC
pDG 77 GAUCUGCCUGGCUGAGCAUUUUGGCAAAGAAUUCACCCCUGAGGUUCAGGCUGCCACC

pDG 5  CAGAAGACUGUGGCUGGUGUGGCCAACGCUCUGGCCCACAAGUACCACUAAACUCCUG
pDG 77 CAGAAGACUGUGGCUGGUGUGGCCAACGCUCUGGCCCACAAGUACCACUAAACUCCUG

                                                *
pDG 5  CCUCCUCUGGGUCUUCAACCUGUUGAAUCCCCUGUUCUCCAUGUUGUCCAUCUUUUGCC
pDG 77 CCUCCUCUGGGUCUUCAACCUGUUGAAUCCCGUGUUCUCCAUGUUGUCCAUCUUUUGCC

pDG 5  AUGGGUGAAUGGGCCCUAUGGCCAUAGCCUUGCCUAGAAUAAAGGUUCAUUUAUUCCAA
pDG 77 AU............complete sequence not determined.............

pDG 5  UUCCpolyA(20)
pDG 77 .........
```

5.3 DISCUSSION

Restriction analyses of native cat 10S cDNA clones was accomplished by digestion with the enzyme HpaII and subsequent end-labelling. It was decided to adopt this approach rather than visualisation of EtBr stained DNA for the following reasons:-

1) The use of HpaII enables the Klenow fragment of DNA polymerase I to fill in 3' ends with α-^{32}P dCTP. This permits the analysis of very small amounts of DNA, typically 100 ng, with quantitative detection of small DNA fragments after autoradiography for several hours.
2) The enzyme HpaII cuts at a 4-base recognition sequence and is more likely to cut the ds cDNA insert than an enzyme with a 6-base recognition sequence, leading to the generation of a relatively large number of small DNA fragments which can be analysed on an acrylamide gel with great accuracy.
3) DNA fragments generated by digestion with HpaII can be cloned directly into the AccI site of M13 phage mp 93.

The observation that the clones appeared to fall into one of two categories of restriction pattern was consistent with the library being composed of native cat α- and β-globin cDNA clones.

Hybrid arrest translation experiments demonstrated that plasmids pDG 73 and pDG 5 significantly inhibited the translation of specific mRNA species in a wheat germ cell free system, whilst pBR 322 had no effect on the specificity of mRNA translation (Fig. 5.2). When bands corresponding to the native cat major adult α- and β-globin proteins were excised from the gel, significant and specific inhibition of globin synthesis was confirmed (Table 5.1). It is most likely that the observed 1:3 ratio of radioactivity incorporated into the α- and

β-globin protein using 10S poly-A RNA reflects a quantitative difference in the number of methionine residues present in each globin polypeptide.

The recombinant plasmids pDG 73 and pDG 5 have been shown by nucleotide sequencing to encode the native cat α- and β-globin polypeptides, respectively. The insert in pDG 73 contains the complete coding and 3' non-coding region of the α-globin mRNA and in addition, seven bases of the 5' non-coding region. The pDG 5 clone, however, was found not to contain the entire β-globin coding region, although it extends to the poly-A tail in the 3' direction. Screening of the 10S cDNA library with a portion of the pDG 5 insert failed to find any clones which were likely to extend further into the β-globin coding region than pDG 5, based upon the restriction pattern and DNA insert length of the hybridising clones.

On the basis of their restriction pattern (section 5.2.1), approximately 20% of the α-globin cDNA clones in the 10S cDNA library are likely to contain at least the entire coding region. This contrasts with the situation for the β-globin clones. The reasons for the apparent disparity in size between the two classes of clones are not clear. It is reasonable to assume that the RNA used to prepare the cDNA library was not degraded to a significant extent, since full length translation products appear after *in vitro* translation (see Chapter 2). Also, if it were degraded it is unlikely that the preferential degradation of the β-globin mRNA would occur. It is possible that some secondary structure present in the β-globin mRNA prevented complete cDNA synthesis by the AMV reverse transcriptase, without inhibiting the ability of the single stranded cDNA to self-prime and undergo second strand synthesis. Analysis of the products of cDNA synthesis were presented in Chapter 3. In that analysis it was shown that the single- and double-stranded cDNA products

corresponded to the size expected of 10S RNA, but there was a proportion of transcripts which were not of a size corresponding complete copies of the RNA. It is possible that this reflects the premature termination of β-globin cDNA synthesis, although this is unlikely to be sompletely responsible for the smaller transcripts. The treatment of ds cDNA molecules with S_1 nuclease to cleave the 5' "hairpin" loop sometimes results in the inadvertant removal of varying amounts of the 5' ends of the molecules, and in some cases where there is an incomplete second strand synthesis, the 3' ends. This phenomenon, although most likely responsible for the low number of full length α-globin cDNA clones is unlikely to account for the relatively much shorter β-globin cDNA clones, unless there was some differential sensitivity to S_1 nuclease displayed by the β-globin molecules.

Very little is known of the number and types of globin chains present in the red blood cells of an adult native cat. The α-globin amino acid sequence derived from the nucleotide sequence of the pDG 73 insert is homologous to the amino acid sequence of the adult α-globin polypeptide of the grey kangaroo and of man. It is likely that the polypeptide encoded in the pDG 73 nucleotide sequence is the major adult α-globin because restriction analysis of the cDNA library has shown that the distinctive HpaII restriction pattern displayed by pDG 73 is common to a large percentage of all clones. The β-globin cDNA clones are all characterised by the absence of a HpaII recognition site within the cDNA insert. Due to the small size of most of the β-globin inserts, the lack of a restriction site is not necessarily indicative of the identity of that insert. Consequently, another recombinant, pDG 77, was selected using a probe consisting entirely of pDG 5 β-globin coding sequence, and its partial nucleotide determined. The complete correspondence of the nucleotide sequence

of pDG 77 with pDG 5 in both the coding and 3' non-translated regions examined strongly suggests that both clones are derived from an mRNA species which is relatively abundant in adult native cat 10S RNA, and is therefore most likely to represent the major adult β-globin mRNA sequence.

The nucleotide sequence of pDG 73 includes seven bases of the native cat α-globin mRNA 5' non-coding region, excluding the initiation codon. Alignment of those seven nucleotides with the corresponding 5' sequences of other α-globin mRNAs demonstrates clear homology over that region with the human, rabbit, chicken and toad molecules (Fig. 5.12). The motif ACC is found to precede initiation codon in the 5' α-globin mRNAs examined, although this does not appear to be a general feature of other eukaryotic mRNA molecules. The native cat α-globin 3' non-coding region extends 123 bases from the termination codon to the poly-A tail, compared to 109 for human, 86 for rabbit, 97 for chicken, and 110 for the toad. Of most interest within the native cat 3' non-coding region is the presence of the hexanucleotide AAUAAA approximately seven bases from the poly-A tail. It is difficult to estimate the exact length of any base from the poly-A tail when examining a cDNA sequence since it is not possible to determine the exact location of the site of poly-A addition. The presence of the putative polyadenylation signal has now been shown to be present in the mRNA 3' non-coding region in species representing every vertebrate infraclass.

The 3' non-coding region of the native cat β-globin mRNA is 133 nucleotides in length, excluding the translation termination codon. The putative polyadenylation signal, AAUAAA, is found approximately 19 bases from the site of polyadenylation. Figure 5.13 shows that the region from the putative polyadenylation signal to the 3' end of the molecule of native cat β-globin mRNA shows sequence homology with other

Figure 5.12

Alignment of the 5' non-coding regions of α-globin mRNA.

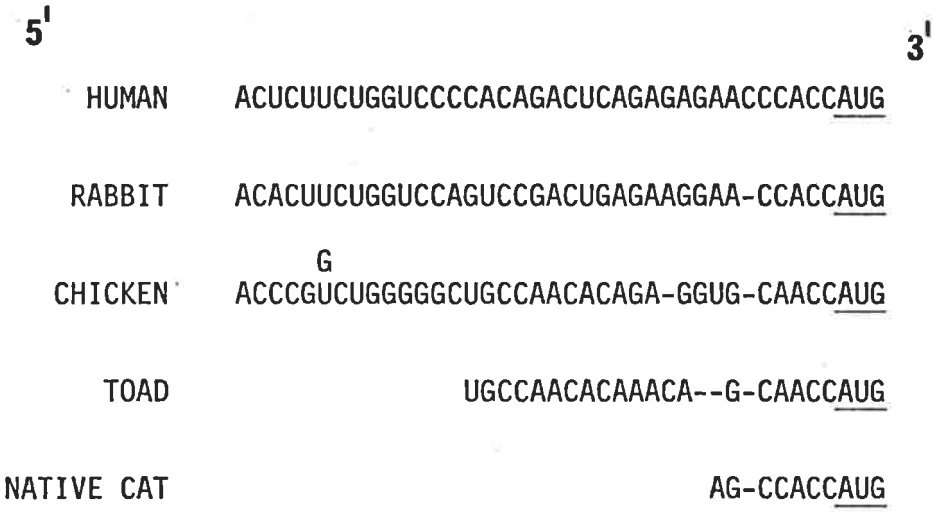

5'
3'
HUMAN ACUCUUCUGGUCCCCACAGACUCAGAGAGAACCCACCAUG
RABBIT ACACUUCUGGUCCAGUCCGACUGAGAAGGAA-CCACCAUG
G
CHICKEN ACCCGUCUGGGGGCUGCCAACACAGA-GGUG-CAACCAUG
TOAD UGCCAACACAAACA--G-CAACCAUG
NATIVE CAT AG-CCACCAUG

Figure 5.13

Alignment of β-globin mRNA molecules from the putative polyadenylation signal to the poly-A addition site.

5' 3'

HUMAN	AAUAAAAAACAUUUAUUUUCAUUGCpolyA
RABBIT	AAUAAAGGAAAUUUAUUUUCAUUGCpolyA
CHICKEN	AAUAAAGUUCAUUCAGUGACACTCpolyA
TOAD	AAUAAAAAGAAAAGUUUCUUCACAUUCUpolyA
NATIVE CAT	AAUAAAGGUUCAUUUAUUCCAAUUCpolyA

β-globin mRNAs when aligned at the AAUAAA motif. Maximum homology is observed with the rabbit where 10 out of 19 nucleotides correspond directly. The region from the AAUAAA sequence to the termination codon shows minor homologies to those sequences present in human, mouse, rabbit, chicken and toad β-globin mRNAs. Williams *et al.* (1980) have noted a region of dyad symmetry 7 or 8 nucleotides 5' to the poly-adenylation signal of many vertebrate β-globin mRNAS, however, no such symmetry is found in the native cat β-globin mRNA.

In summary, restriction analysis and hybrid arrest translation experiments indicated that two cDNA clones pDG 73 and pDG 5 contained nucleotide sequences corresponding to native cat α- and β-globin mRNAs, respectively. Nucleotide sequencing of pDG 73 and pDG 5 confirmed the clones as encoding native cat globin mRNA sequences. In addition, another clone pDG 77, was identified by partial sequence analysis as containing sequences identical to those found in the pDG 6 insert DNA. The structure and composition of the native cat α- and β-globin mRNAs elucidated in these experiments highlights the highly related nature of all vertebrate globin mRNA molecules.

CHAPTER 6. ASPECTS OF GLOBIN MOLECULAR EVOLUTION

6.1 INTRODUCTION

A comparison of the amino acid sequences of orthologous proteins has revealed that molecular evolution appears to occur at a constant rate with time, the rate varying between different classes of proteins (e.g. histones, globins, fibrinopeptides etc.) (see Wilson *et al.*, 1977a). This phenomenon has been referred to as the "molecular clock". Whilst the molecular clock has proved useful for evolutionary studies, there is much conjecture as to its theretical basis. There is argument about why the rate of protein evolution should appear to be constant with respect to solar time rather than generation time. Also, it is possible that the observed constancy of rate could result from averaging saltatory changes over a long evolutionary period (see Kimura, 1983; Wilson *et al.*, 1977a).

Two competing theories on the nature of the driving force behind molecular evolution have been proposed. Kimura (1969) has suggested that most of the mutations that are fixed during evolution are selectively neutral. This has become known as the "neutral theory". The "selection theory" as represented by Clarke (1970) predicts that most mutations that are fixed are advantageous to the organism as a whole. Proponents of both theories accept that many mutations are deleterious and are eliminated by natural selection and, accordingly, that advantageous mutations have a role in protein evolution. Kimura (1969) first proposed the neutral theory to account for the discrepancy between the apparent rate at which amino acids have been substituted in protein evolution, and the rate allowable according to the cost of such natural selection as calculated by the procedure of Haldane (1957).

He argued that this discrepancy would disappear if there were essentially no advantage associated with an allele during the period when it was going to fixation.

The neutral theory was proposed in part to explain the molecular clock phenomenon and, as such, the nature of the driving force behind the molecular clock has become a central issue in selectionist/neutralist debate. For neutral alleles, the rate of fixation by drift can be readily shown to be equal to the rate at which such alleles occur by mutation. Fluctuations in population size and environment are not expected to affect the rate of substitution of neutral alleles (see Kimura, 1983). Consequently, for a given protein, the rate of substitution should have a constant value dependent only on the functional constraints on the protein. King and Jukes (1969) argue that if evolutionary change was due primarily to selection under the influence of changing environment, adaptive radiation and changes in population size, then one would expect evolution to be subject to a burst of rapid change in some species, and relative stability in others. This would be expected to lead to a non-constant rate of evolution of orthologous proteins. The "selectionist" school of molecular evolution argues that the molecular clock may be the response to a constantly changing environment through the fixation of selectively advantageous mutants (Maynard-Smith, 1969). The theoretical grounds of the neutralist argument have been vigorously contested by, amongst others, Clarke (1970), Maynard-Smith (1968), and Richmond (1970), whilst proponents of the neutral theory have been notably Kimura (1969), King and Jukes (1969), and Corben and Uzzel (1970).

The ability to purify specific pieces of DNA and to determine their nucleotide sequence using recombinant DNA technology has resulted

in the accumulation of a considerable body of data which has contributed to the selectionist/neutralist debate. Studies of apparently "functionless" DNA such as pseudogenes and introns and of synonymous base changes in functional protein-coding regions, have led to the conclusion that, in general, evolution is more rapid where changes do not alter the functional efficiency of proteins, but whether some of these changes are truly without selective value is subject to conjecture.

In this chapter, the protein and nucleotide sequence data generated from the study of the native cat α- and β-globin mRNAs are added to the globin data from other species, and various aspects of molecular evolution are re-examined.

6.2 RESULTS

6.2.1 Evolutionary Studies Based Upon Amino-acid Sequence Differences

Air *et al.* (1971) estimated the date for the marsupial/eutherian divergence from evolutionary studies of the α- and β-globin, and myoglobin polypeptides of the grey kangaroo, *Macropus giganteus*. Many more α- and β-globin protein sequences have since been determined for eutherian species, as well as the β-globin amino acid sequence of another macropod marsupial, the potoroo *Potorous tridactylus* (Thompson and Air, 1971), and the amino acid sequences for the native cat globins.

Air *et al.* (1971) calculated evolutionary rate, Y, expressed as years per amino acid substitution per 100 residues in the following manner. If d is the number of amino acid differences between two polypeptide chains, and n is the total number of amino acids in each polypeptide, the number of amino acid substitutions, D, in each of them since divergence, can be expressed in terms of each 100 residues as,

$$D = \frac{d}{2} \times \frac{100}{n} \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

If the time (in years) since divergence of two polypeptides is t, then,

$$Y = \frac{t}{D} = \frac{2nt}{100d} \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

To obtain a date for the marsupial/eutherian divergence, equation (2) is employed using a value for Y calculated from non-marsupial data only, to avoid using the data in a circular manner.

The matrices of Tables 6.1 and 6.2 show the values of d for the α- and β-globin polypeptides, respectively. These values were calculated in this study, using the amino acid sequences compiled by Dayhoff (1975), except for the amino acid sequences of the toad (*X. laevis*) α- and β-globin polypeptides which were inferred from nucleotide sequence data presented by Kay *et al.* (1983) for α-globin mRNA, and Williams *et al.* (1980) for the β-globin mRNA. Since the available amino acid sequence data for the native cat β-globin polypeptide are incomplete (see Chapter 5), Table 6.2 also includes a calculation of the number of amino acid sequence differences between β-globins which considers only differences in the region from amino acid number 60 to the carboxyl terminus.

Table 6.3 gives the Y values calculated from the data in Tables 6.1 and 6.2, and includes two Y value calculations for β-globin polypeptides, one being based on differences considering the entire molecule, and another being based upon a partial sequence, as previously described.

Table 6.1

Number of amino acid differences between α-globins.‡

	human	monkey	rabbit	mouse	horse	goat	sheep	chicken	toad	carp	native cat	kangaroo	opossum
human	0												
monkey	4	0											
rabbit	25	25	0										
mouse	15	15	27	0									
horse	18	16	25	17	0								
goat	21	17	28	22	17	0							
sheep	21	18	29	22	17	3	0						
chicken	35	35	44	39	40	42	42	0					
toad	52	52	59	58	54	56	56	51	0				
carp	71	71	74	70	70	68	69	73	71	0			
native cat	24	24	36	34	32	31	31	37	51	74	0		
grey kangaroo	24	23	34	32	26	34	34	37	52	69	17	0	
opossum	40	40	50	43	42	41	41	60	70	77	39	42	0

‡ References are given in the text.

Table 6.2

* Number of amino acid sequence differences between β-globins from amino acid numbers 60-146.

Values in parentheses are differences from complete sequences.‡

	human	monkey	rabbit	mouse	horse	goat	sheep	chicken	toad	native cat	kangaroo	potoroo
human	0											
monkey	3(7)	0										
rabbit	7(15)	7(16)	0									
mouse	15(26)	13(26)	13(28)	0								
horse	17(25)	17(28)	15(25)	21(37)	0							
goat	18(28)	14(26)	16(29)	21(34)	19(34)	0						
sheep	21(32)	16(26)	17(28)	21(34)	19(32)	4(4)	0					
chicken	26(38)	26(38)	26(42)	27(44)	22(39)	20(39)	22(38)	0				
toad	49(67)	47(69)	48(64)	50(64)	45(60)	41(60)	41(63)	45(58)	0			
native cat	17	18	19	23	17	24	24	24	44	0		
grey kangaroo	21(26)	21(25)	20(25)	20(31)	18(24)	23(29)	25(31)	26(31)	43(51)	10	0	
potoroo	19(28)	19(27)	19(25)	19(31)	18(29)	23(29)	26(32)	24(33)	44(53)	12	2(5)	0

‡ References are given in the text.

Table 6.3

Average rate of evolution of globin chains.

	Hb α chain	Hb β chain	
		whole molecule	amino acids 60→146
Y value ± s.d.	13.4 ± 5.06	12.7 ± 5.0	12.8 ± 4.8
n*	45	36	36
± s.e.m.	0.72	0.82	0.81

* n is the number of comparisons

6.2.1.1 Calculation of the marsupial/eutherian divergence date

Using the Y values shown in Table 6.3, and the amino acid sequences of the native cat and grey kangaroo α- and β-globin polypeptides and potoroo β-globin, divergence estimates were calculated according to equation (2) (see Table 6.4). The divergence estimates of 148 and 165 million years using the α- and β-globin polypeptide sequences respectively, may be compared with values of 139 and 137 million years obtained by Air *et al.* (1971).

6.2.1.2 Other divergence dates

Using the amino acid sequence differences shown in Tables 6.1 and 6.2 for the native cat, kangaroo and potoroo globin polypeptides, a date of approximately 81 million years was calculated for the dasyurid/macropod divergence (see Table 6.5).

Stenzel (1974) has determined the amino acid sequence of the α-globin polypeptide from a South American marsupial, the opossum, *Didelphis marsupialis*. Table 6.6 shows that estimates of 186, 190 and 199 million years were obtained for the divergence of the opossum α-globin from dasyurid, macropod, and eutherian α-globins, respectively. Clearly, such results are incompatible with the marsupial/eutherian divergence dates calculated in the previous section, as well as with the calculations of Air *et al.* (1971).

6.2.2 Nucleotide Sequence Divergence Estimates

Nucleotide sequence divergence estimates were calculated as described by Perler *et al.* (1980). Changes in the coding regions of homologous genes either result in amino acid replacements (replacement

Table 6.4

Estimated number of years elapsed since marsupial/eutherian divergence.

	Protein	
	α-globin	β-globin
Y value	13.4 ± 0.74	12.8 ± 0.81
divergence estimate (million years)	148	165
95% confidence limits	132-164	145-185

Table 6.5

Estimated number of years elapsed since dasyurid/macropod divergence.

	Protein	
	α-globin	β-globin
Y value	13.4 ± 0.74	12.8 ± 0.81
divergence estimate (million years)	81	82
95% confidence limits	89-73	92-72

Table 6.6

Estimated number of years elapsed since didelphid/non-didelphid divergence.

	Protein α-globin
Y value	13.4 ± 0.74
divergence from dasyurids (m.y.)	190
divergence from macropods (m.y.)	194
divergence from eutherians (m.y.)	199

site substitutions) or lead to the appearance of synonymous codons (silent site substitutions).

Using nucleotide sequence data for human (Wilson *et al.*, 1980), rabbit (Heindell *et al.*, 1978), mouse (Nishioka and Leder, 1979), chicken (Dodgson and Engel, 1983), toad (Kay *et al.*, 1983), native cat (this study) α-globin coding regions, and human (Wilson *et al.*, 1977b), rabbit (van Ooyen *et al.*, 1979), mouse (Konkel *et al.*, 1979), chicken (Richards *et al.*, 1979), toad (Williams *et al.*, 1980), native cat (this study) β-globin coding regions, the per cent replacement and per cent silent site substitutions for pairwise comparisons between the aforementioned species were calculated (see Table 6.7). Since only partial nucleotide sequence of the native cat β-globin is known (see Chapter 5), the nucleotide sequence divergences were also calculated for pairwise comparisons involving β-globin coding regions from codon number 60 up to, but not including, the termination codon. These are presented in Table 6.7. The divergence values for the partial β-globin sequences were found to correspond closely to those calculated for the entire coding region; only four out of thirty comparisons gave divergence values differing by 10% or more.

A linear relationship between the estimated time of divergence based upon palaeontological data (Romer, 1966) and per cent nucleotide substitution can be demonstrated for replacement site substitution in the α- (Figure 6.1) and β- (Figure 6.2) globin coding regions.

An estimation of the "unit evolutionary period" (UEP), defined by Wilson *et al.* (1977a) as the time (in million years) required to fix a 1% nucleotide sequence divergence between species, gives values of 11.3 and 11.0 for the α- and β-globins, respectively.

The relationship between the time of divergence and per cent

Table 6.7

Per cent nucleotide substitution for globin genes.

	Replacement sites			Silent sites		
			GLOBIN	GENES		
	α	β	β178-438	α	β	β178-438
human/rabbit	11.4	6.3	5.8	32.1	42.4	44.9
human/mouse	8.8	13.1	12.7	83.0	49.4	45.8
human/native cat	16.7	-	18.4	84.6	-	60.1
human/chicken	21.9	25.4	27.2	75.6	70.4	72.5
human/toad	42.9	51.2	48.7	205.0	147.1	164.3*
rabbit/mouse	11.9	12.3	12.9	82.7	64.0	61.7
rabbit/native cat	15.9	-	18.6	77.8	-	55.1
rabbit/chicken	23.0	23.9	22.9	64.9	81.2	92.6*
rabbit/toad	36.1	48.4	42.4*	172.9	118.2	114.7
mouse/native cat	13.5	-	14.1	70.1	-	66.2
mouse/chicken	20.8	27.2	27.8	87.9	79.8	91.3*
mouse/toad	41.4	-	43.8	152.3	132.6	124.8
native cat/chicken	27.5	-	23.7	112.8	-	89.4
native cat/toad	40.6	-	42.7	170.9	-	128.6
chicken/toad	33.2	45.8	42.4	121.7	144.4	154.09

* Difference >10% whole molecule value

Figure 6.1

Relationship between per cent nucleotide substitution and time since divergence for vertebrate α-globin genes.

The solid line represents the relationship for replacement site substitutions and the dotted lines represent two alternative interpretations of the relationship between per cent silent site substitution and time since divergence.

Squares represent the silent site data points.

Triangles represent the replacement site data points.

All data are presented in Table 6.7.

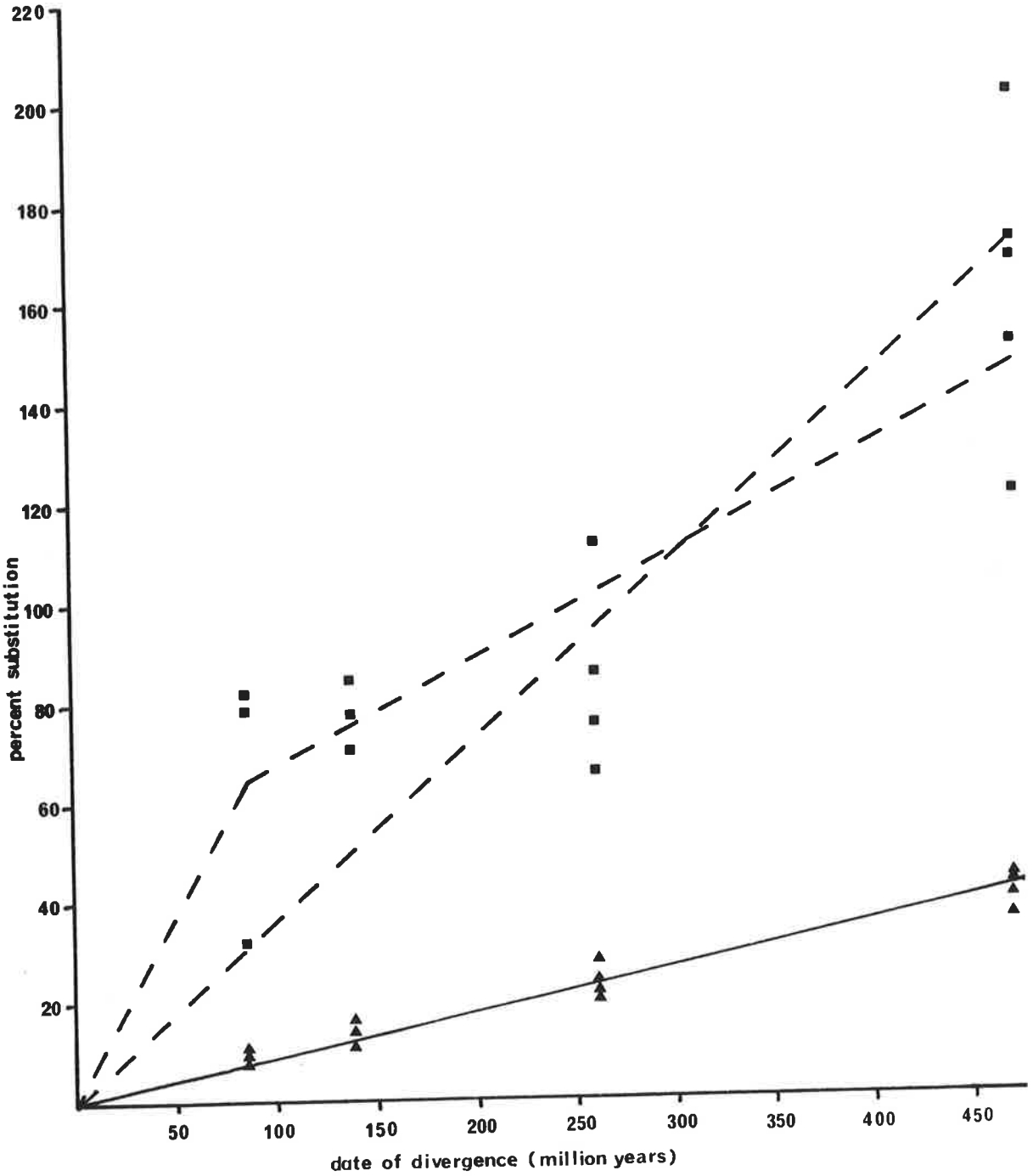

220
200
180
160
140
120
100
80
60
40
20
percent substitution
50
100
150
200
250
300
350
400
450
date of divergence (million years)

Figure 6.2

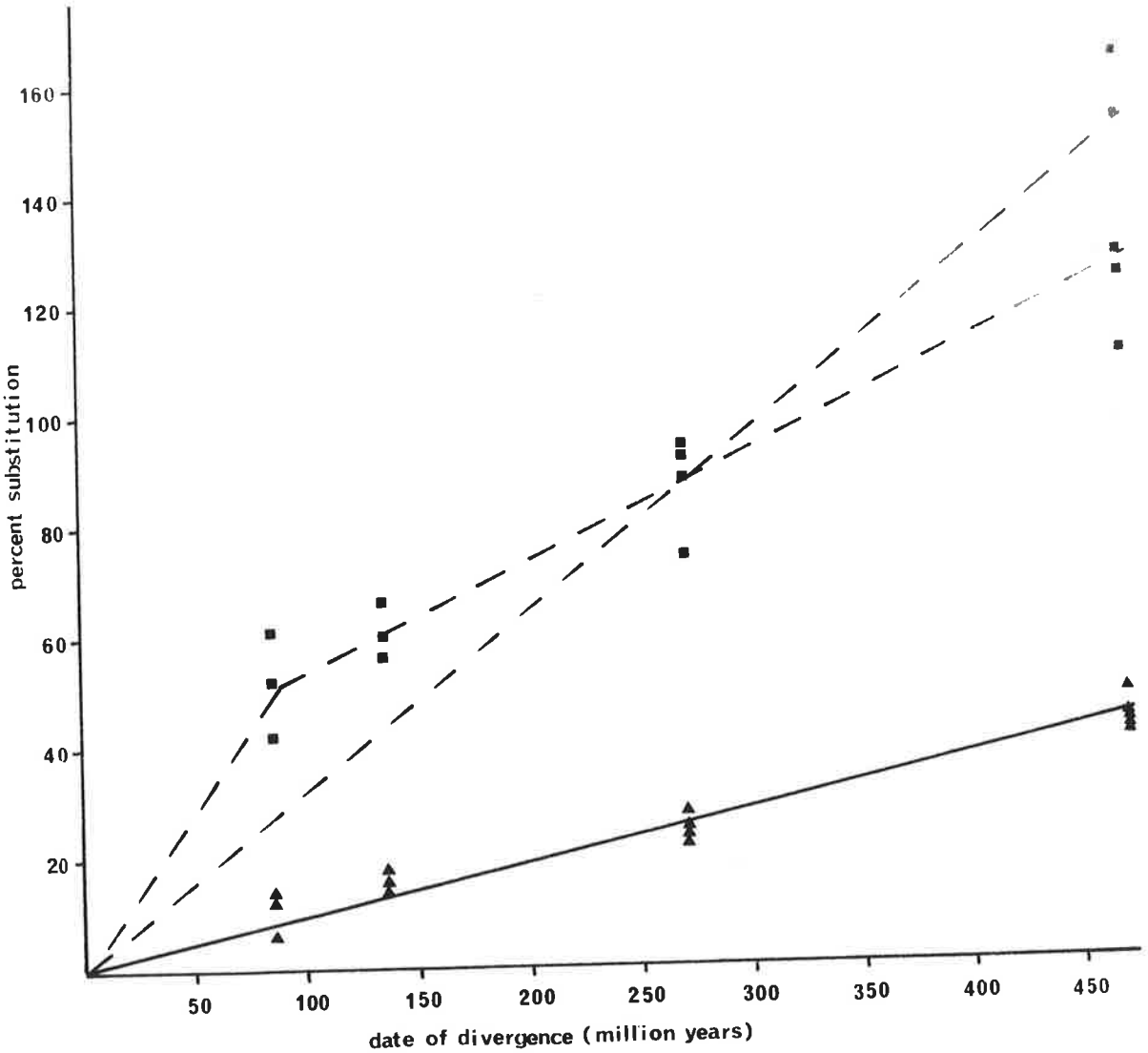
Relationship between per cent nucleotide substitution and time since divergence for vertebrate β-globin genes.

The solid line represents the relationship for replacement site substitutions and the dotted lines represent two alternative interpretations of the relationship between per cent silent site substitution and time since divergence.

Squares represent the silent site data points.

Triangles represent the replacement site data points.

All data are presented in Table 6.7.


160
140
120
100
80
60
40
20
percent substitution
50
100
150
200
250
300
350
400
450
date of divergence (million years)

silent site substitution is not clear. Two possible interpretations of that relationship are presented for the α- and β-globin genes in Figures 6.1 and 6.2, respectively, one involving a single linear relationship, and the other a biphasic linear relationship.

6.3 DISCUSSION

6.3.1 Divergence Date Estimates Based Upon Amino-acid Differences

Air *et al.* (1971) described a linear relationship between the date of divergence of pairs of species and the number of amino acid differences between homologous polypeptides in those species. The polypeptides considered were α-globin, β-globin, and myoglobin. Using such a relationship they derived an estimate of 130 million years ago for the date of the marsupial/eutherian divergence. Since their study, globin amino acid sequences have been determined for many more species, including the potoroo (Thompson and Air, 1971) and the native cat (this study). Consequently it has been possible now to expand the total number of comparisons from 38 (Air *et al.*, 1971) to 83 for the calculation of Y. The availability of three further marsupial globin sequences allow a total of 35 marsupial/eutherian pairwise comparison to be made, compared with 10 comparisons used by Air *et al.* (1971) in their calculation.

The Y values in this study for the β-globin polypeptides were calculated from amino acids 60→146 only, except for *X. laevis* which has a total of 145 amino acids (Williams *et al.*, 1980), and was aligned with other β-globins from amino acids 59→145. This was done so that the partial amino acid sequence of the native cat β-globin, derived from the nucleic acid studies detailed in previous chapters, could be used in these calculations. The evolutionary rate estimated

for the partial β-globin polypeptide was found not to differ from that for the complete molecule. This was not unexpected since the functional sites defined by Perutz (1979) are distributed proportionately in the two β-globin polypeptide fragments (amino acids 1→59; and amino acids 60→146).

Using the Y values shown in Table 6.3, an estimate of 156 million years (being an average of 148 and 165 million years) was obtained for the time since marsupials and eutherians diverged. This may be compared with the estimate of 138 million years obtained by Air *et al.* (1971) using data from α- and β-globin polypeptides. The divergence estimate derived in this study based on α-globin polypeptide data (148 million years) was within the 95% confidence limits of the divergence estimate proposed by Air *et al.* (1971) using α-globin polypeptide data (123-154 million years). Similarly, the estimate obtained in this study using β-globin data (165 million years) is also within the 95% confidence limits calculated by Air *et al.* (1971) from β-globin data (104-170 million years). The estimates obtained here, while not differing significantly from those of Air *et al.* (1971), have smaller 95% confidence limits, suggesting that Air *et al.* (1971) may have underestimated the time since the marsupial/eutherian divergence.

The divergence time estimate of Air *et al.* (1971) (approximately 130 million years) resulted from averaging estimates derived from α-globin, β-globin, and myoglobin amino acid sequence data. An estimate of 138 million years was obtained when only α- and β-globin data was used by Air *et al.* (1971). The Y value calculated from the non-marsupial myoglobin data was based upon only six comparisons, all of which were between eutherians. It would be of interest to recalculate the estimate of the time since the marsupial/eutherian divergence using a larger sample of myoglobin amino acid sequence

data than used by Air *et al.* (1971).

It should be noted that the divergence date based upon α-globin amino acid sequences calculated by Air *et al.* (1971), 139 million years, assumes an average marsupial/eutherian amino acid difference of 28.8, although the value calculated from data published by Air *et al.* (1971) gives an average difference of only 26.6. An average marsupial/eutherian α-globin polypeptide difference of 26.6 gives an estimated divergence date of 128 million years, not 139 million years. Despite this discrepancy, the figure of 139, quoted from their own data by Air *et al.* (1971), has been used in this discussion.

Air *et al.* (1971) have cited palaeontological evidence suggesting that marsupials diverged from eutherian mammals early in the Cretaceous period, about 130 million years ago (see Slaughter, 1968). A divergence date of around 155 million years ago, such as calculated here, would place the marsupial/eutherian divergence in the late Jurassic period. Several different theories have been advanced as to the early dispersive history of marsupials. Lillegraven (1969) suggested that marsupials "arose" in North America and from there dispersed to South America by the late Cretaceous period. In North America, the only continent from which there is a good Cretaceous fossil record, the marsupials had diverged into three families and at least five genera and thirteen species by the end of that period (Lillegraven, 1974). Tedford (1974) argued that marsupials arose in the South American part of Gondwanaland and dispersed northwards to North America. Many arguments have been advanced for the radiation of marsupials from South America to Australia via Antarctica, rather than from Australia to the Americas (see Clemens, 1977; Keast, 1977). Although marsupials appear to have been well established on the American continent in the mid to late Cretaceous

period, deficiencies in the fossil record of South America limit the conclusions about the date of marsupial/eutherian divergence. A divergence date in the late Jurassic, such as suggested by the comparison of amino acid sequences presented here, is entirely plausible.

A comparison of dasyurid and macropod globin polypeptides has given a divergence estimate of 80 million years, in the mid to late Cretaceous period. The Australian fossil record, which does not extend beyond the lower Miocene or upper Oligocene periods (about 27 million years g.p.) (Tedford, 1974), shows the presence of nearly all major marsupial genera and families, indicating a long previous history of evolution and differentiation. It has been suggested that the South American-Antarctica junction persisted only up until the early to mid Cretaceous period (Dalziel *et al.*, 1973). If this were the case, then dasyurid/macropod radiation would probably have occurred after the separation of Antarctica/Australia from South America, and before the subsequent separation of Australia from Antarctica, about 50 million years b.p. (Hayes and Ringis, 1973). This may account for the absence of members of the Dasyuridae or Macropodidae on the American continent. An alternative explanation for the apparent lack of radiation of metatherians is that they were largely eliminated by competition with eutherian species on the American continent, whilst on the Australian continent they had few competitors (see Hayman, 1977).

Using immunological data derived from studies on albumins, Maxson *et al.* (1975) have estimated the dasyurid/macropod divergence date as 55 million years, an estimate outside the 95% confidence limits of the calculations presented in this study (see Table 6.5). No estimate was given of statistical error, so it is difficult to compare their

value with the 80 million years estimated from globin polypeptide data. Whilst a molecular clock based upon immunological studies of albumin has proved useful in evolutionary studies (Sarich, 1973; Lowenstein *et al.*, 1981), the rate of albumin evolution as measured by immunological methods varies considerably along certain lineages such as some groups of marsupials, and, as such, the albumin-based "clock" is prone to errors (see Maxson *et al.*, 1975).

Comparisons of the α-globin amino acid sequence of the opossum, *Didelphis marsupialis*, a South American marsupial, with other α-globin amino acid sequences, give divergence date estimates of 186, 190 and 199 million years ago for dasyurids, macropods and eutherians, respectively (see Table 6.6). The opossum is known as a "living fossil", having evolved morphologically very little from its Cretaceous counterpart (Stenzel, 1974). The fossil record and chromosomal studies have shown didelphids to be metatherians (see Hayman and Martin, 1974; Sharman, 1962) and thus it appears that the divergence estimate derived from α-globin polypeptide studies is inconsistent with all other evidence, including all estimates of the date of marsupial/eutherian divergence. The opossum α-globin polypeptide obviously displays an evolutionary rate exceeding that of other vertebrate α-globin molecules. Based upon the selectionist assumption that the rate of evolution is limited by the cost of natural selection, Stenzel (1974) has suggested that in organisms which appear to have undergone very little morphological change over long periods of evolutionary time (such as "living fossils"), an unusually high rate of molecular evolution may be expected at some loci. His argument appears to be based upon the assumption that the slow rate of morphological change in "living fossils" reflects a reduced rate of substitution at certain loci, and that this reduction

allows a more rapid rate of substitution at other loci, such as the α-globin locus.

The opossum α-globin is the only amino acid sequence data to be published from a didelphid, and has not yet been confirmed. It is obviously important to obtain more molecular data from the opossum before general trends in its molecular evolution can be recognised. Vertebrate α-globin molecules show amino acid conservation at a total of 46 positions, excluding the opossum data, with those positions all having functional properties defined by Perutz (1976). With the addition of the data from the opossum α-globin polypeptide chain, the number of residues in common to all α-globins is reduced to 20. This observation highlights the extraneous nature of the opossum α-globin data published by Stenzel (1974). It would be interesting to investigate this problem by determining whether the rapid evolution of certain loci in the didelphids is a real phenomenon, and if so, what the factors are which determine the loci that evolve at an enhanced rate.

Using a molecular clock based upon immunological studies of albumin, Lowenstein *et al.* (1981) derived an estimate of 70 million years for the didelphid/non-didelphid marsupial divergence. As discussed earlier, the "albumin clock" may not be entirely accurate, but the estimate of 70 million years derived by Lowenstein *et al.* (1981) is certainly more compatible with palaeontological data than the estimate derived in this study based upon α-globin polypeptides.

It should be noted when studying evolution using molecular techniques that the interpretation of palaeontological data is highly subjective, especially in the absence of a complete fossil record, and that the accuracy attempted using molecular data often exceeds the possible accuracy of the fossil record.

6.3.2 Nucleic Acid Sequence Divergence Studies

Since the study by Perler *et al.* (1980) on mammalian and avian globin and preproinsulin genes, many more globin gene sequences have become available for evolutionary comparison. In particular, the elucidation of the coding sequences of marsupial (this study) and amphibian (Williams *et al.*, 1980; Kay *et al.*, 1983) α- and β-globin genes has enabled the estimation of nucleic acid sequence divergence for vertebarte classes not previously represented in the study of Perler *et al.* (1980).

A linear relationship was described between per cent replacement site substitution and evolutionary divergence for the α- and β-globin coding regions (see Figures 6.1 and 6.2). The α- and β-globin gene divergence estimate data were presented separately so that any small differences between the evolutionary divergence rates of the two genes might be recognised. No such differences were found. An average UEP value of 10 for replacement site substitutions calculated by Perler *et al.* (1980) by pooling data from the α- and β-globin genes, may be compared with the value of 11.1 for α-globin genes and 11.0 for β-globin genes obtained here.

The linear relationship between per cent replacement site substitution and divergence time was not unexpected in view of the "molecular clock" phenomenon observed when comparing the amino acid sequence divergences of orthologous proteins. As pointed out by Jeffreys (1981), many of the divergence dates used in nucleotide sequence divergence studies are often derived largely from protein sequence molecular clock expectations, leading to an analysis which is to some extent circular in nature. The divergence date estimates used in this study were compiled from palaeontological data by Goodman

et al. (1975). It should be noted, however, that divergence date estimates based on the fossil record, are sometimes assessed in the knowledge of divergence estimates based upon molecular evidence, and as such, divergence estimates based upon palaeontological data may not always be totally independent of molecular data (see Wilson *et al.*, 1977a).

The relationship between per cent silent site substitution and divergence time showed greater variability than the comparison involving replacement sites (see Figures 6.1 and 6.2). It is clear that the overall rate of silent site substitutions in α- and β-globin genes is much higher than the rate of replacement site substitutions.

Perler *et al.* (1980) have suggested that globin gene silent site substitutions show an UEP of approximately 1 for a period of up to 100 million years of divergence, after which the rate decreases to an UEP of approximately 9. A similar interpretation of the data presented in this study is possible. Figures 6.1 and 6.2 show UEP values of 1.1 and 1.8 for a period up to 90 million years difference, followed by UEP values of 4.4 and 5.3 for the α- and β-globin genes, respectively. The addition of the globin gene data from the amphibian *X. laevis* has contributed to the discrepancy between the UEP estimates of this study and those of Perler and her co-workers for silent site substitution at periods greater than 100 million years. It must be emphasised that the relationships presented in Figures 6.1 and 6.2 are just two interpretations of the data, and result from attempts to fit *linear* relationships to the data points. Figures 6.1 and 6.2 depict an attempt to fit a single linear distribution to the α- and β-globin silent site divergence data points. In both cases a single linear function appears inappropriate to describe the relationship between

per cent silent site substitution and evolutionary divergence.

Jeffreys (1981) has pointed to the problems of correcting for multiple substitution events at high levels of substitution when using a "random substitution" model of nucleotide sequence divergence, such as the model presented by Perler *et al.* (1980). Accordingly, it is difficult to define a relationship between silent site substitutions and divergence time when high levels of substitution are evident, such as those displayed by *X. laevis* and chicken. Whilst the large variability in the data precludes the use of silent site divergence as a molecular clock over long periods of time, it may be useful for comparing sequences separated by less than 100 million years divergence.

Perler *et al.* (1980) have suggested that there may be two classes of silent site substitution; the initial substitution rate may reflect the accumulation of neutral changes which saturate at a divergence time of about 100 million years, the second class of silent substitution, which is largely observed after divergences exceeding 100 million years, achieve fixation at a slower rate than the first due to selective forces associated with them. The results presented in this study show that the date from marsupial and amphibian globin genes tend to reinforce the observation by Perler *et al.* (1980) of an initial, more rapid, rate of accumulation of silent changes, followed by a more gradual rate of fixation of silent changes after a period of 100 million years divergence.

It is difficult to obtain an accurate estimate of the initial silent site UEP. The initial silent site UEP estimates presented in Figures 6.1 and 6.2 could be significantly decreased by assuming that saturation of "neutral" sites occurs at less than 85-100 million years divergence. The lack of data in the divergence period up to 85 million

years is a serious deficiency in this analysis. By comparing the nucleotide sequences of recently duplicated gene pairs, such as the human β-δ globin genes, a better estimate of the initial rate of silent site divergence might be obtained. Unfortunately, the calculation of the divergence times of recently diverged gene pairs is often largely based upon replacement site or amino acid comparisons, and the use of such divergence dates, albeit for silent site studies, may result in a somewhat circular analysis. From their study, Perler *et al.* (1980) proposed that the observed amount of replacement changes in globin genes could not be accounted for by a "neutral mutation rate" defined by a UEP of 0.7 (being the initial rate of silent site substitution for α- and β-globin genes). They therefore concluded that the "replacement clock" is driven primarily by selection, not neutral mutation. This conclusion assumes that the initial rate of silent site substitution is an accurate estimate of the "neutral" mutation rate. It is unlikely that a large proportion of the initial silent site substitutions are strictly neutral, but even if so, the results from this study suggest an initial rate of silent site substitution for globin genes with a UEP of approximately 1.5. A neutral mutation rate defined by a UEP of 1.8 - 2.0 could easily account for the amount of replacement changes observed for globin genes when using the calculations of Perler *et al.* (1980).

In conclusion, a re-examination of the relationship between per cent nucleotide sequence substitution and time of divergence for globin genes has confirmed the linear nature of the relationship for replacement site substitutions found by Perler *et al.* (1980). The addition of new data to that of Perler *et al.* (1980) for silent site substitutions has highlighted the variability in the relationship between silent site

substitution and time of divergence. This study has confirmed the possibility that there are two rates of silent site substitution, dependent on the amount of time separating the divergence of species. Problems in the estimation of the initial rate of silent site substitution and whether this is a realistic representation of the "neutral mutation rate" were found to severely limit conclusions as to the basis for the molecular clock.

CHAPTER 7. GLOBIN GENE MAPPING BY *IN SITU* HYBRIDISATION

7.1 INTRODUCTION

The techniques of *in situ* hybridisation to fixed metaphase chromosomes was developed in several laboratories (Gall and Pardue, 1969; John *et al.*, 1969; Buongiorno-Nadelli and Amaldi, 1970). The opportunity to visualise nucleic acid:chromosomal DNA hybridisation has resulted in a large number of studies which have determined the chromosomal assignment of specific genes or sequences (see Henderson, 1982).

Initially, *in situ* hybridisation was used to localise the chromosomal sites of repeated sequences on both polytene and diploid chromosomes of many species (reviewed by Miklos and John, 1979; Henderson, 1982). Early claims for the localisation of unique genes on diploid chromosomes by *in situ* hybridisation were subject to much controversy. Price *et al.* (1972) reported the localisation of globin sequences to human chromosomes 2, and r or 6, using *in situ* hybridisation of *in vivo* labelled rabbit reticulocyte mRNA. Based upon calculations involving the specific activity of the probe, the autoradiographic efficiency of the emulsion, and the exposure time, Bishop and Jones (1972) and Prensky and Holmquist (1973) argued that the number of silver grains deposited over chromosomes 2, and 5 or 6 in the experiment of Price *et al.* (1972) could not have been due to low multiplicity hybridisation events. Subsequent research has shown that the α-globin gene sequences are located on human chromosome 16 (Deisseroth *et al.*, 1977a) and the β-globin sequences are located on chromosome 11 (Deisseroth *et al.*, 1977b; Lebo *et al.*, 1979).

A major factor limiting the application of *in situ* hybrisisation for the localisation of unique gene sequences is the need for high purity nucleic acid probes. In the case of relatively abundant nucleic

acids such as ribosomal RNA and satellite DNA sequences, sufficient quantities of highly purified probes can be isolated by standard physical methods. However, the relatively low cellular concentrations of most mRNA molecules, and the difficulty in purifying them, preclude a physical isolation approach. Small amounts of contaminating RNA species can produce a detectable signal over their gene sites that can be stronger than the signal generated from the major RNA species of the probe, especially if the contaminant is encoded by a repeated sequence such as ribosomal RNA. In the experiments of Price *et al.* (1972) it is likely that a false localisation occurred through contamination of their probe with some sequence of high multiplicity. The advent of recombinant DNA technology has essentially eliminated the problem of contaminating nucleic acid species. Pure probes for the analysis of any gene or genomic sequence can now be prepared.

Probe purity has been one of the two major drawbacks which have, up until recently, prevented the chromosomal assignment of low multiplicity sequences, the other being the strength of the signal generated by a hybridisation event. The detection limits of *in situ* hybridisation were initially determined by quantitative studies in which 5S RNA and ribosomal RNA were hybridised to *Drosophila* polytene chromosomes (Szabo *et al.*, 1977) or human diploid chromosomes (Cote *et al.*, 1980). In these studies it was calculated that about 0.12 - 0.6 silver grains per site of hybridisation would occur for a unique gene of about 10^3 base pairs after a 60 day autoradiographic exposure period, using a probe of specific activity of 1-5 x 10^8 dpm/µg. Probes of a specific activity of the order of 10^8 dpm/µg can be generated by direct iodination using carrier free ^{125}I. Tritium labelled probes can also be produced with specific activities of 0.5 - 1.5 x 10^8 dpm by the *in vitro* synthesis of complementary RNA cRNA (Malcolm *et al.*,

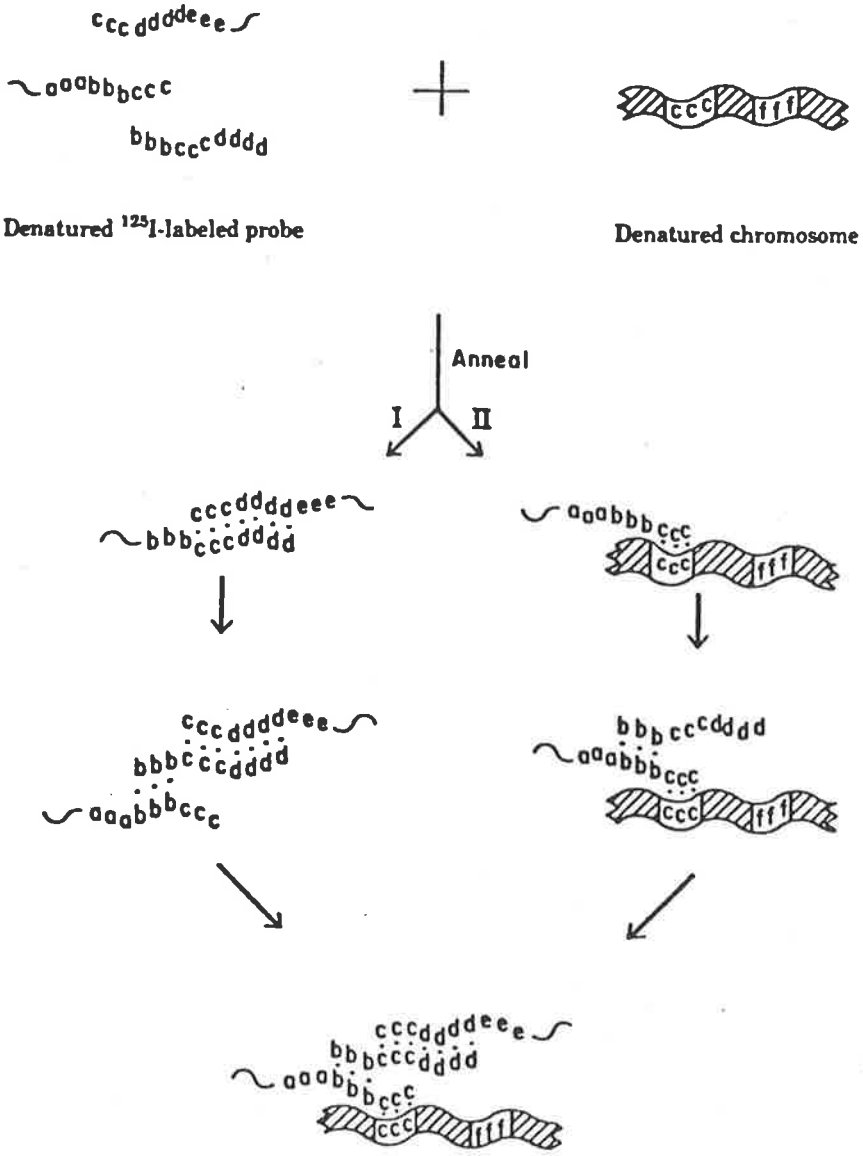
1982) or by nick translation (Trent *et al.*, 1982). Although the specific activities of tritium labelled probes are lower, they have the advantage of a lower energy β emission which improves autoradiographic resolution. The signal produced by probes with such specific activities is sufficient for localisation of low multiplicity gene sequences if the signal to noise ratio is very high and a large number of metaphases are analysed. Malcolm *et al.* (1981) and Henderson *et al.* (1978) have successfully used ^{3}H-cRNA for the localisation of globin genes in man and mouse, respectively. Studies such as these made it clear that for convenient and routine localisation of unique genes, some enhancement of the signal produced by the hybridisation of a single probe molecule to its complementary sequence was required.

Signals can be enhanced most easily by increasing the specific activities of the probe molecules. Using carrier free ^{125}I-deoxycytidine triphosphate, specific activities of greater than 10^{9} dpm/μg have been obtained by nick translation (Gerhard *et al.*, 1981). The signal from a given hybrid can also be increased by indirectly increasing the probe size. When nick translated probes are used, the signal from a given hybrid can be amplified through the formation of probe networks. Nick translation of a recombinant plasmid results in the generation of a random collection of labelled fragments consisting of both insert and plasmid sequences. With the presence of 10% dextran sulphate in the hybridisation buffer, the plasmid and other non-hybridised sequences will form a network using the original hybridised molecule as the base molecule (Fig. 7.1). Dextran sulphate has been shown to increase the rate of DNA:DNA hybridisation in solution, on filters, and *in situ* (Alwine and Kemp, 1977). Alwine *et al.* (1979) have shown that in the presence of 10% dextran sulphate, the signal obtained using double stranded probes on DNA blots is up to 100-fold greater than that

Figure 7.1

Probe network formation in the presence of 10% dextran sulphate.

from Gerhard *et al.*, 1981.


Denatured ^{125}I-labeled probe
Denatured chromosome
Anneal
I
II

obtained with a single stranded probe of comparable specific activity. The increase in signal using dextran sulphate has enabled a number of effectively unique genes to be mapped to chromosomal loci using autoradiographic exposure times of between 5 and 22 days. These include the human α-globin gene (Gerhard *et al.*, 1981), the human insulin gene (Harper *et al.*, 1981), the human placental lactogen growth hormone cluster (Harper *et al.*, 1982), the sequences coding for human leukocyte, fibroblast and immune interferon (Trent *et al.*, 1982), the human actin genes (Soriano *et al.*, 1982), and the Syrian hamster CAD genes (Wahl *et al.*, 1982). These genes were all localised by nick translating plasmid probes with ^{125}I or ^{3}H-labelled nucleotides.

Alwine *et al.* (1979) have determined the optimal probe size for network formation to be 500-1000 base pairs, and that probe size is more important for signal detection than the use of very high specific activity probes. Soriano *et al.* (1982) found that the signal at a single actin gene site yielded an average of two silver grains per day of exposure, compared to an expected signal of 0.01-0.02 silver grains per day in the absence of probe networks. At that level of enhancement any gene for which a hybrid plasmid exists should be easily mapped on diploid chromosomes.

Another technical improvement which has made a significant contribution to chromosomal gene localisation is the development of techniques for combining chromosome banding techniques with *in situ* hybridisation. Many *in situ* hybridisation experiments require statistical analyses to unequivocally determine the significance of a specific hybridisation event, compared to background. Essential for such analyses, and indeed for all gene assignments by *in situ* hybridisation, is the need for positive chromosome identification. Some banding methods involve prephotographing Giemsa banded chromosomes

following treatment with trypsin (Robins *et al.*, 1981) or 1% lipsol detergent (Malcolm *et al.*, 1982), or staining chromosomes with Quinacrine mustard to obtain Q bands after hybridisation (Lawrie and Gosden, 1980). Harper *et al.* (1981) have used a staining/destaining procedure with Wright's stain to obtain post-hybridisation G-banding.

In spite of the use of radioactively labelled probes for *in situ* hybridisation, the track of the decay particle and the thickness of the autoradiographic emulsion, even with low energy tritium probes, limits the precision with which target sequences can be localised within chromosomes. As an alternative to radiolabelled probes, a number of researchers have used immunological or affinity techniques for the chromosomal localisation of gene sequences. Bauman *et al.* (1981) have chemically coupled fluorescein or rhodamine to the 3' terminus of RNA molecules and used these as probes for *in situ* hybridisation to *Drosophila* chromosomes. Apart from the speed of localisation, another major advantage in using a fluorescent label is the marked increase in resolution. For green-light emission, the resolving power of a fluorescent microscope is 0.25 μm, whereas the autoradiograph resolution ^{3}H or ^{125}I is at best 1 μm and 4 μm, respectively (Rudkin and Stollar, 1977; Bauman *et al.*, 1981).

Polynucleotide probes have also been detected without radioisotopic labels by exploiting the specific interaction between biotin, a water soluble vitamin, and avidin, an egg white glycoprotein. Nucleotide analogues of dTTP and dATP that contain a biotin molecule covalently attached to the C-5 position of the pyrimidine ring through 'linker arms' have been synthesised by Langer *et al.* (1981). These nucleotides can be used as substrates for the various polymerising reactions and result in DNA or RNA products which can effectively bind to complementary DNA or RNA sequences (Langer *et al.*, 1981). The sites of hybridisation

can be detected immunologically using anti-biotin antibodies in consort with a second antibody tagged with a fluorescent, enzymatic or electron dense reagent, or by affinity labelling using preformed complexes of avidin and biotinylated derivatives of horseradish peroxidase or intestinal alkaline phosphate. Langer-Safer *et al.* (1981) have used biotinylated probes for the localisation in *Drosophila* of five separate cloned DNA sequences which are specifically expressed in the fat bodies of third-instar larvae, a cloned tRNA gene, and a cloned transposable element. The sites of hybridisation were detected fluorometrically or cytochemically using second antibodies (rabbit antibiotin antibodies) conjugated with fluoroscein or horseradish peroxidase. Manuelides *et al.* (1982) have also used biotinylated probes to localise mouse satellite DNA to fixed metaphase chromosomes derived from a mouse cell line. Second antibodies tagged with colloidal gold in concert with biotinylated probes have been used to map satellite sequences in *Drosophila* (Wu *et al.*, 1982) and mouse (Hutchinson *et al.*, 1982). The colloidal gold spheres are electron dense compared to chromatin and can be visualised with high resolution under the electron microscope.

The major drawback with all non-isotopic detection procedures is their lack of sensitivity. At present there has not been a report of the successful localisation of a unique sequence to chromosomes from a diploid cell. Autoradiographic procedures have been refined to such an extent that unique sequences can be localised. Such procedures are, at present, preferred for the localisation of sequences of low multiplicity to chromosomes.

The previous chapters have described the construction, isolation and characterisation of native cat α- and β-globin cDNA clones. Such cDNA clones can be used to determine the chromosomal localisation of α- and β-globin-like sequences by using the technique of *in situ*

hybridisation. The *in situ* hybridisation method of Trent *et al.* (1981) has been used to assign the chromosomal position of various interferon genes to human metaphase chromosomes, and this procedure was used in the assignation of native cat α- and β-globin sequences.

The *in situ* hybridisation procedure of Trent *et al.* (1982) makes use of ^{3}H-labelled probes. This was thought to be preferable to the procedure of Gerhard *et al.* (1981) in which the probes are labelled with ^{125}I. ^{3}H-labelled probes give lower backgrounds and higher resolution than iodinated probes, are cheaper to produce, decay much less rapidly, and do not present as great a personal radiation safety hazard. Tritiated probes suffer from the disadvantage of being of lower specific activity than their iodinated counterparts, leading to a correspondingly longer exposure time.

7.2 RESULTS

7.2.1 Chromosome Preparations

Metaphase chromosomes were prepared from a diploid cell line, dv♂, established from a pinna explant from a male native cat. Dividing cells were arrested at metaphase by the addition of colcemid for 1 hour and fixed metaphase preparations were prepared as described in section 10.2.18. Those preparations which were thought of sufficient quality for further study were spotted on to glass microscope slides and stored dessicated at room temperature. Slide preparations were always used within the first week of storage. The karyotype of the native cat derived from dv♂ is shown in Figure 7.2.

7.2.2 Probe Preparation

Plasmids pDG 73 (α-globin), pDG 5 (β-globin), and pBR 322 were linearised by digestion with the restriction endonuclease EcoRI

Figure 7.2

Metaphase chromosome of the native cat, *Dasyurus viverrinus*.


1
2
3
4
5
6
XY

and subsequently nick translated in the presence of tritiated dTTP and dGTP as described in section 10.2.7. A specific activity of 5-8 x 10^7 dpm/µg was routinely obtained.

7.2.3 Hybridisation of pDG 73 DNA to Native Cat Metaphase Chromosomes

In situ hybridisation of the pDG 73 probe was carried out as detailed in section 10.3.19. Four probe concentrations were used, 0.05, 0.1, 0.2, and 0.4 µg/ml. After 21 days a slide which had been treated with each probe concentration was analysed by examination of photographs of metaphase spreads which contained the complete diploid chromosome complement. Twenty metaphases were examined at each concentration to determine the chromosome:cytoplasm ratio of silver grains. For each metaphase a circle was drawn which encompassed all of the chromosomes, then the number of labelled chromosomal sites was determined and compared to the number of grains in the "cytoplasm". This was done to obtain some measure of the autoradiographic background. De la Chapelle (pers. comm.) has used a similar procedure in the determination of optimum probe concentration for the mapping of human immunoglobulin light chain genes. Table 7.1 shows that the chromosomal:cytoplasm silver grain ratio varied from 1:1.14 for a probe concentration of 0.2 µg/ml to 1:5.14 for a probe concentration of 0.05 µg/ml. On the basis of these results, the slides which were hybridised at a probe concentration of 0.2 µg/ml were chosen for further analysis.

A total of 140 metaphase spreads were analysed for the presence of silver grains after an exposure time of 22 days (Fig. 7.3a). The metaphases analysed in this experiment were from four separate slides, all of which were prepared, probed and analysed in a single experiment. Metaphases were found to have an average of 1.3 labelled chromosomal

Table 7.1

Signal to noise ratio estimation for pDG 73 (α-globin).

Probe concentration µg/ml	Grains after 21 day exposure /20 metaphases		Chromosome:cytoplasm ratio
	chromosome	cytoplasm	
0.05	7	36	1 : 5.14
0.1	21	39	1 : 1.85
0.2	25	41	1 : 1.64
0.4	29	60	1 : 2.14

Table 7.2

Signal to noise ratio estimation for pDG 5 (β-globin).

Probe concentration µg/ml	Grains after 19 day exposure /20 metaphases		Chromosome:cytoplasm ratio
	chromosome	cytoplasm	
0.05	12	42	1 : 3.50
0.1	23	41	1 : 1.78
0.2	36	67	1 : 1.86
0.4	45	106	1 : 2.35

Figure 7.3

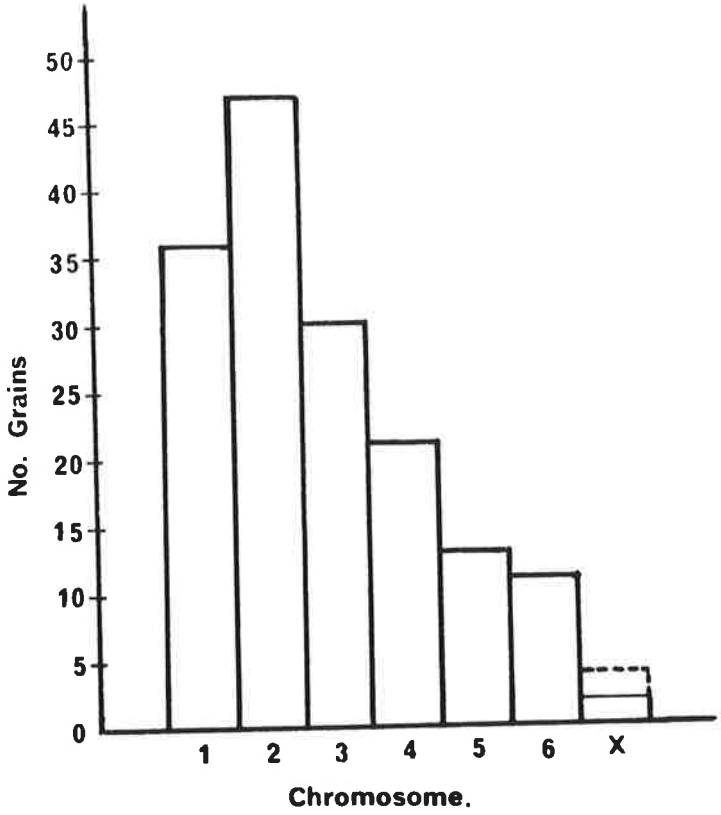
Results of *in situ* hybridisation of tritium-labelled pDG 73 to native cat metaphase chromosome.

A. Absolute number of grains per chromosome.

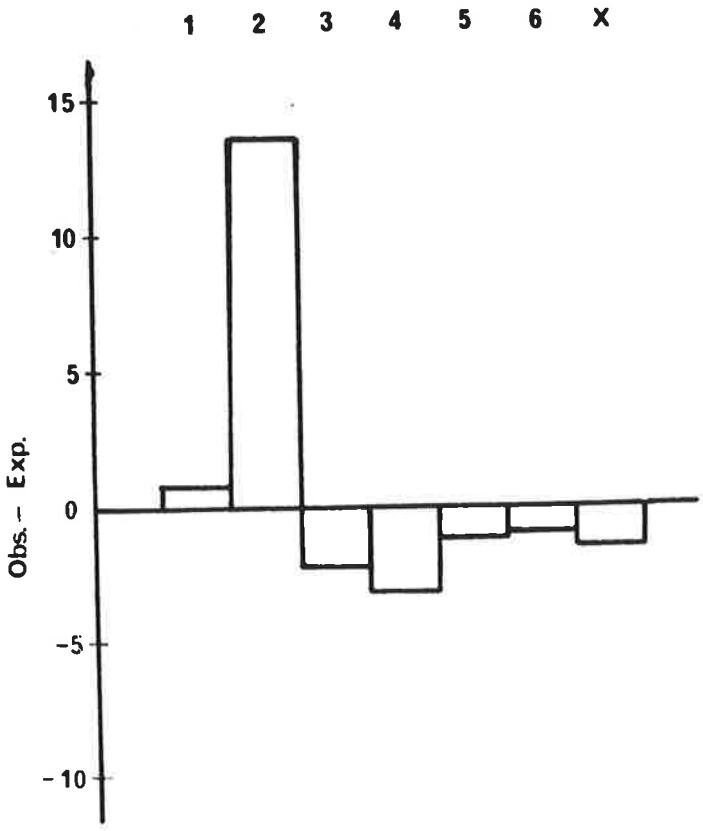
B. Number of grains adjusted for length of the chromosomes.

Dotted line indicates the adjustment made for two X chromosomes.

A


50
45
40
35
30
25
20
15
10
5
0
No. Grains
1
2
3
4
5
6
X
Chromosome.

B


1
2
3
4
5
6
X
15
10
5
0
-5
-10
Obs. – Exp.

sites per cell. As can be seen from Figure 7.3a, there is no obvious difference between the labelling of each chromosome when the absolute number of labelled chromosomal sites is considered. This result can be tested for significance in the following manner. If one assumes that there had been no significant hybridisation of pDG 73 DNA to the chromosomes of the native cat then one would expect the silver grains to be distributed at random over all chromosomes. The expected number of grains per chromosome was therefore calculated by dividing the total number of grains present over chromosomes by the relative proportion of the genome represented by each chromosome (see section 9.2.1). Figure 7.3b shows the result of subtracting the expected values from those observed using a probe concentration of 0.2 µg/ml. It can be seen that chromosome number 2 appears to be more heavily labelled than any of the other chromosomes. The mean displacement of the observed from expected was 2.9 grains per unit chromosome length, with a standard deviation of 4.94. The displacement from expected for chromosome number 2 gave a t_6 value of 2.16, corresponding to a probability of between 0.05 and 0.025. The displacement from expected for chromosome number 2 was therefore significant at the 5% level.

The distribution of silver grains observed along the length of chromosome number 2 is shown in Figure 7.4. Chromosome number 2 divided into thirds based upon length measurements. Chromosome number 2 is a metacentric (Fig. 7.2), and in the absence of chromosome banding, it was not possible to unambiguously orientate each chromosome. For this reason the data were pooled into two classes, the "central" third and the "telomeric" thirds. The results presented in Figure 7.4a suggest that the "central" third is significantly labelled, with 65% of the silver grains being present over one-third of the chromosome.

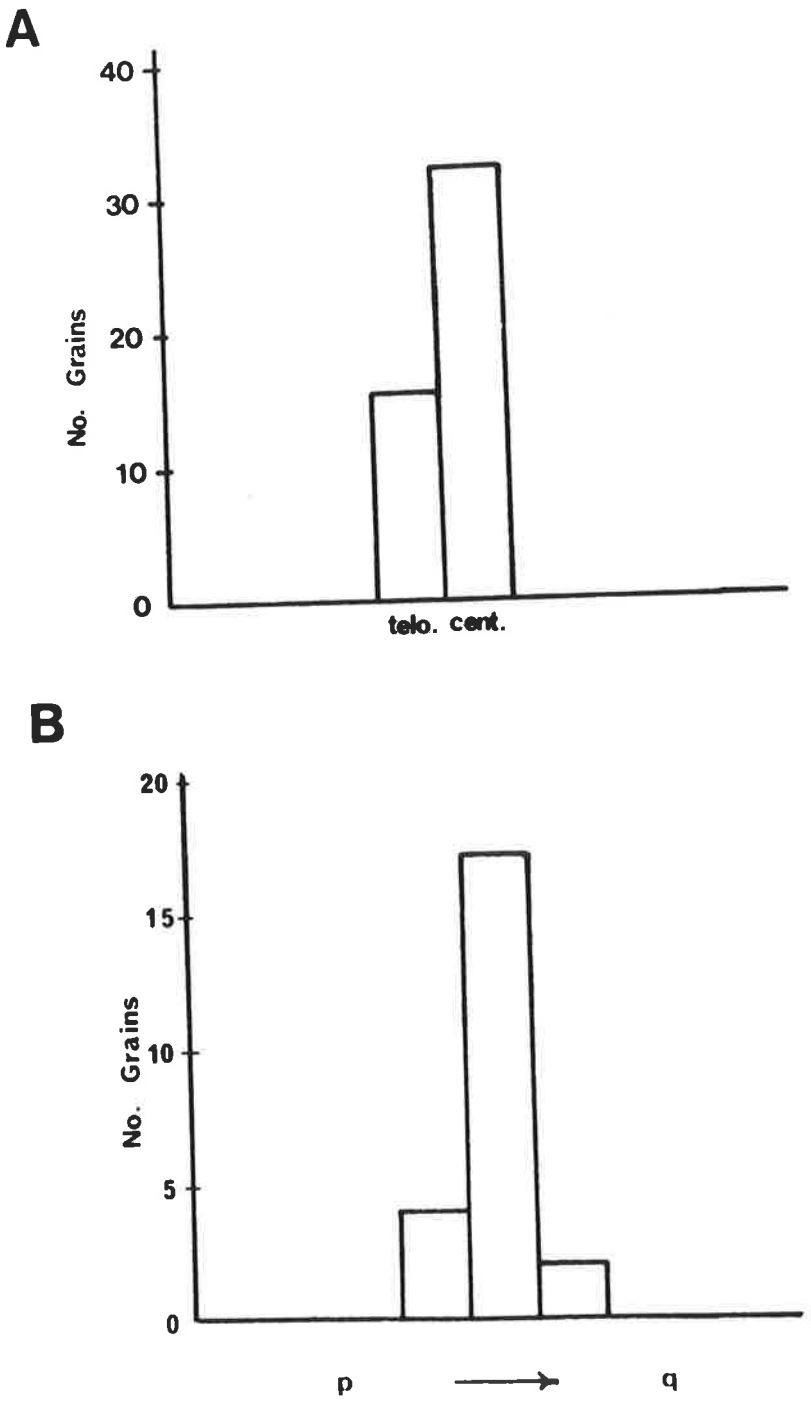
A representative metaphase chromosome spread from these experiments

Figure 7.4

Distribution of silver grains along the length of chromosomes.

A. Distribution of silver grains along the length of chromosome 2 after probing with ^{3}H pDG 73 DNA. The chromosome was divided into three regions, the "central" third and the "telomeric" thirds. The data from the telomeric thirds were pooled and compared to the control third as indicated.

B. Distribution of the silver grains along the length of chromosome 4 after probing with ^{3}H pDG 73 DNA. The chromosome was divided into three regions. The data from the short arm (p) to the long arm (q) are presented.


A
40
30
20
10
0
No. Grains
telo. cent.
B
20
15
10
5
0
No. Grains
p
q

is shown in Figure 7.5.

7.2.4 Hybridisation of pDG 5 DNA to Native Cat Metaphase Chromosomes

In situ hybridisation of the pDG 5 probe was carried out as described in section 10.3.19. Four probe concentrations were examined, 0.05, 0.1, 0.2 and 0.4 µg/ml. After 19 days an example of each probe concentration was examined. Twenty metaphases were examined at each concentration to determine the chromosome:cytoplasm ratio of silver grains, as described in section 7.2.3. The results are presented in Table 7.2. The lowest chromosome:cytoplasm ratio (1:1.78) was obtained when pDG 5 DNA was used at a concentration of 0.1 µg/ml and it was therefore decided to further analyse metaphase spreads using this probe concentration. A total of 80 metaphase spreads were analysed for the presence of silver grains after an exposure period of 20 days. An average of 1.4 labelled chromosomal sites/metaphase was observed.

Figure 7.6a shows the chromosomal distribution of silver grain after probing with pDG 5 DNA. It appears from Figure 7.6a that chromosome number 4 may be significantly labelled. The data was subjected to an analysis identical to that described for the pDG 73 hybridisation (section 7.2.3). The results are presented in Figure 7.6b. The mean displacement of observed values from expected values was 2.8 grains/chromosome unit length, with a standard deviation of 4.80. The displacement value observed for chromosome number 4 gave a t_6 value of 2.27, giving a result significant at the 5% level ($0.05>P>0.025$).

Chromosome number 4 was divided into thirds based upon chromosome length, and the distribution of silver grains along that chromosome determined (Fig. 7.4b). Of all labelled sites, 76% were found to

Figure 7.5

Metaphase spread showing hybridisation of tritium labelled pDG 73 DNA to native cat chromosomes.

The silver grain present on chromosome 2 is arrowed.

Figure 7.6

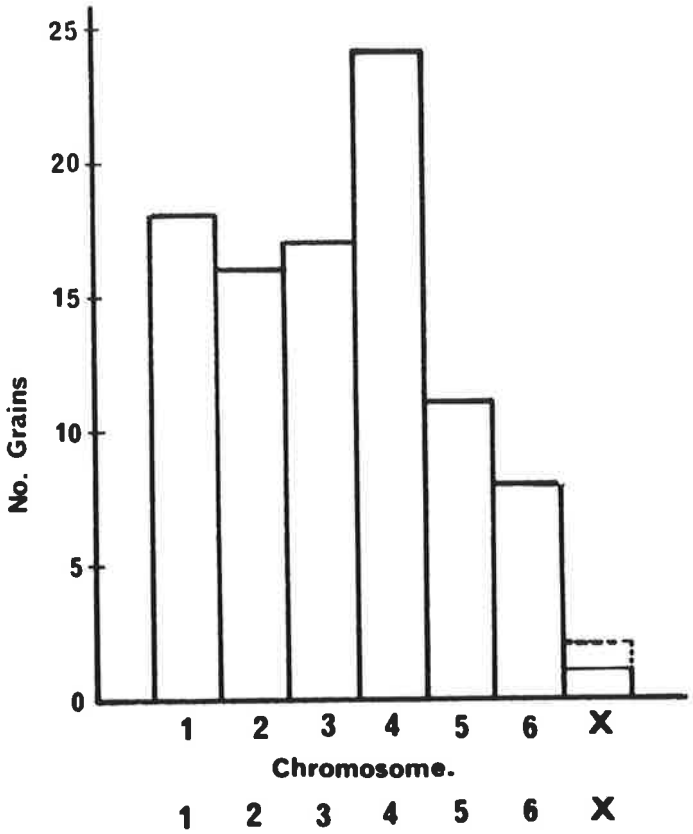
Results of *in situ* hybridisation of tritium-labelled pDG 5 to native cat metaphase chromosomes.

A. Absolute number of grains per chromosome.

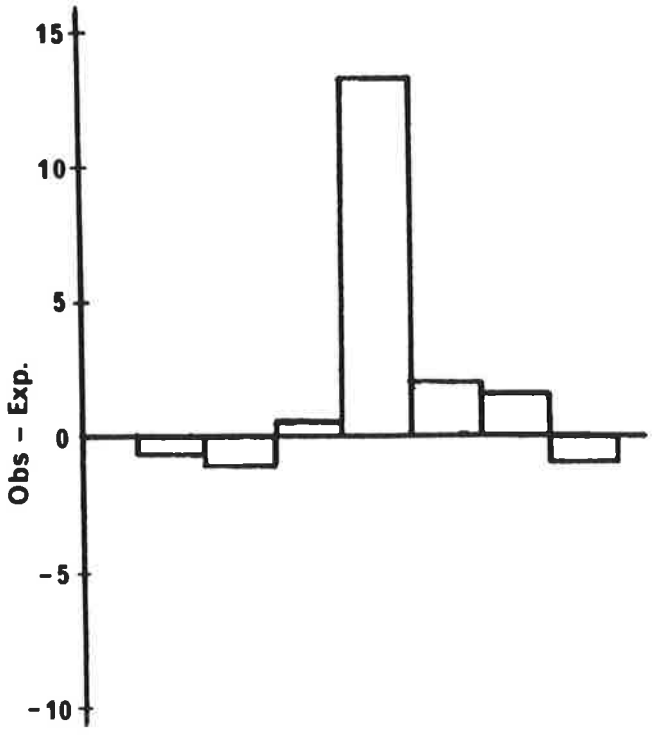
B. Number of grains adjusted for length of the chromosomes.

Dotted line indicates the adjustment made for two X chromosomes.

A


25
20
15
10
5
0
No. Grains
1 2 3 4 5 6 X
Chromosome.

B


15
10
5
0
-5
-10
Obs - Exp.
1 2 3 4 5 6 X

reside within the central third of chromosome number 4.

A representative metaphase spread from this experiment is shown in Figure 7.7.

7.2.5 Control Hybridisation with pBR 322

When pBR 322 DNA is nick translated in the presence of ^{32}P labelled deoxynucleoside triphosphates and probed to total native cat DNA in "Southern" blot analyses, no specific hybridisation is observed, indicating a lack of homology between pBR 322 and native cat DNA. Even so, the *in situ* hybridisation experiments presented in this chapter included pBR 322 controls.

Tritium labelled pBR 322 was hybridised to native cat metaphase spreads at probe concentrations of 0.1 and 0.2 µg/ml. Sixty metaphase spreads from each probe concentration were analysed after an exposure time of 22 days. The results for the two probe concentrations are presented in Figure 7.8a. There was no apparent hybridisation of pBR 322 DNA to the native cat chromosomes when total grain counts per chromosome were tabulated. No significant hybridisation was apparent when the data were adjusted for chromosome length (Fig. 7.8b).

7.2.6 Other Hybridisation

Although only the analyses for the 0.2 µg/ml pDG 73 probe concentration and the 0.1 µg/ml pDG 5 probe concentration are presented in detail, the results from the other probe concentrations are summarised in Table 7.3 (pDG 73) and Table 7.4 (pDG 5). For both probes the 0.05 µg/ml concentration showed non-significant hybridisation to the chromosome which has previously been shown to bind probe, i.e. pDG 73 and chromosome number 2, and pDG 5 and chromosome number 4. The higher concentrations of both probes failed to show significant hybridisation at the 5% level.

Table 7.3

pDG 73 (α-globin)

Effect of pDG 73 probe concentration on *in situ* hybridisation.

Probe concentration (μg/ml)	No. metaphases	Total no. chromosomal silver grains	Chromosome 2 sig. labelled?		σ from mean displacement
			5%	10%	
0.05	40	16	No	No	0.54
0.1	40	48	Yes	Yes	2.12
0.4	40	61	No	Yes	1.78

Table 7.4

pDG 5 (β-globin)

Effect of pDG 5 probe concentration on *in situ* hybridisation.

Probe concentration (μg/ml)	No. metaphases	Total no. chromosomal silver grains	Chromosome 2 sig. labelled?		σ from mean displacement
			5%	10%	
0.05	40	22	No	No	0.71
0.2	40	68	Yes	Yes	2.29
0.4	40	87	No	Yes	1.75

Figure 7.7

Metaphase spread showing hybridisation of tritium-labelled pDG 5 DNA to native cat chromosomes.

The silver grain present on chromosome 4 is arrowed.

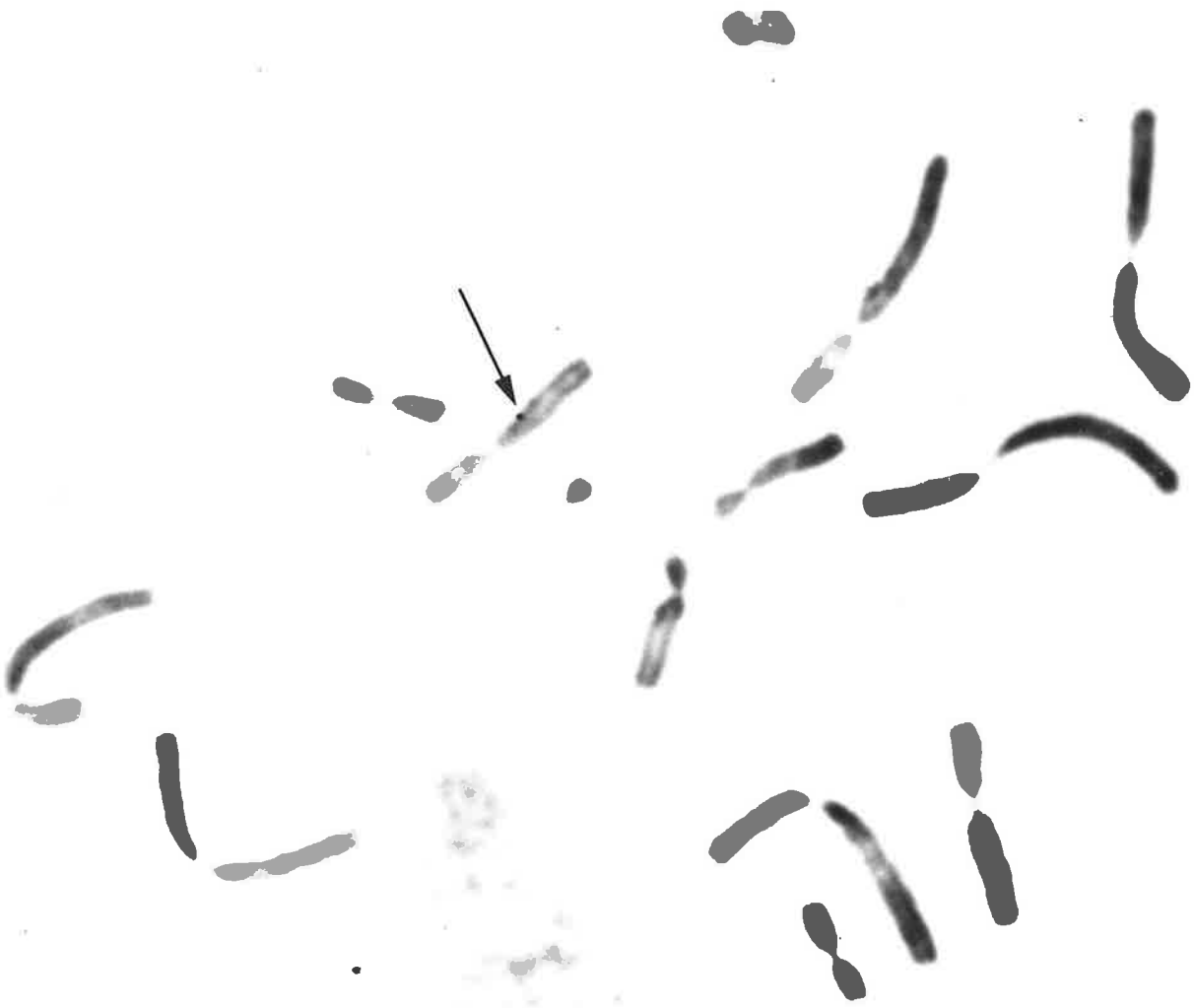


Figure 7.8

In situ hybridisation of pBR 322 DNA to native cat chromosomes.

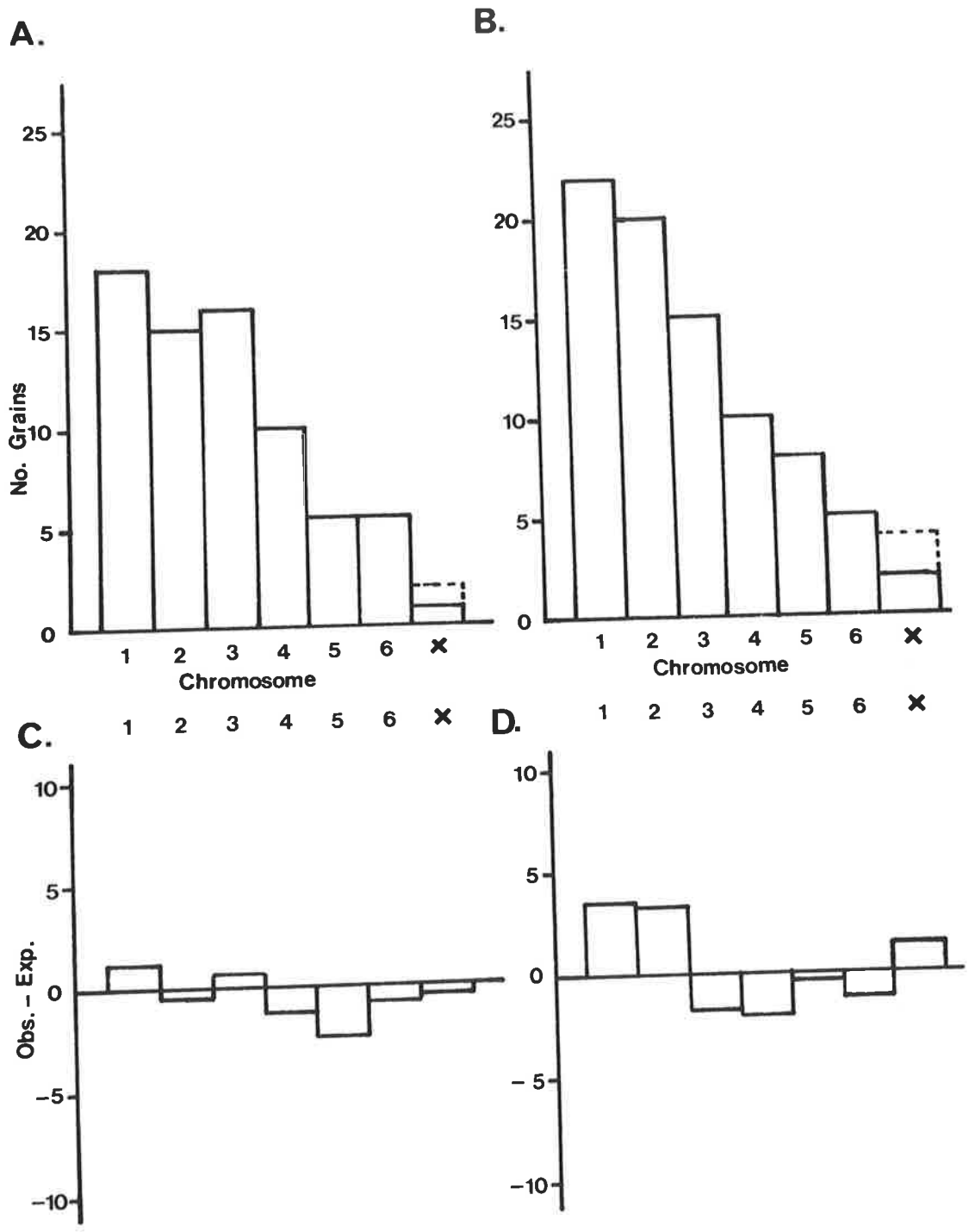
A. Absolute number of grains scored per chromosome using a probe concentration of 0.1µg/ml.

B. Absolute number of grains scored per chromosome using a probe concentration of 0.2µg/ml.

C. Number of grains per chromosome adjusted for chromosome length. Probe concentration was 0.1µg/ml.

D. Number of grains per chromosome adjusted for chromosome length. Probe concentration was 0.2µg/ml.

Dotted lines indicate the adjustment made for two X chromosomes.


A.
B.
C.
D.
No. Grains
Obs. – Exp.
Chromosome
1 2 3 4 5 6 ×
0
5
10
15
20
25
–5
–10

7.2.7 Effects of Increased Exposure

The *in situ* hybridisation procedure contains a large number of variable factors which either along or in combination may affect the efficiency of detection of a hybridisation event. One of these variables is autoradiographic exposure time. In the experiments detailed in the previous sections, exposure time was held to either 20 or 21 days, based upon the findings of Trent *et al.* (1982) and Harper *et al.* (1982). Using pDG 73 DNA at 0.1 µg/ml as a model probe, the effect of exposure time was investigated. The ratio of chromosomal silver grains to "cytoplasmic" grains was calculated for a variety of exposure times (Table 7.5). Also included is the incidence of double labelling. Double labelling is defined as the occurrence of adjacent silver grains over a chromosomal site, e.g. two silver grains over a single site on chromosome number 2 when using the pDG 73 probe. The data show a decrease in the apparent signal to noise ratio due to a progressive increase in "cytoplasmic" silver grains with increased exposure time.

7.3 DISCUSSION

As described in section 7.2.3, a sample of metaphases was examined and the chromosome:"cytoplasm" ratio of silver grains determined for each probe concentration. This was done in an attempt to quantify autoradiographic background. A low signal to noise ratio may at best require that a large amount of data be collected before a statistically significant result emerges, or, at worst, prohibit the detection of specific probe hybridisation to the chromosomes.

High autoradiographic backgrounds can be generated in a number of ways. The autoradiographic emulsion may have a high background due to gradual exposure during storage. Accordingly, the emulsion used

Table 7.5

Effect of increased exposure time on *in situ* localisation using pDG 73 at 0.1 µg/ml.

Exposure time (days)	Grains/20 metaphases		Chromosome:cytoplasm ratio	Specific double labelling per 20 metaphases
	Chromosome	cytoplasm		
21	21	39	1 : 1.85	0
25	22	45	1 : 2.05	0
28	22	45	1 : 2.05	0
35	25	56	1 : 2.24	0
42	27	70	1 : 2.59	1

in these experiments was checked by coating a clean glass microscope slide and examined immediately after drying. Probe concentration has been shown to play a major role in the level of autoradiographic background (Malcolm *et al.*, 1981). In an *in situ* experiment where the signal will be low, such as the mapping of low multiplicity sequences, it is important to use a range of probe concentrations such that appropriate hybridisation kinetics are established with the presence of low amounts of non-specific hybridisation. Table 7.1 showed that a low concentration (0.05 µg/ml) and a high concentration (0.4 µg/ml) of pDG 73 DNA, the chromosome to cytoplasm silver grain ratio was lower than the two intermediate values. At a low probe concentration the chromosomal grain count (7) was approximately one-third that of the 0.2 µg/ml probe, whilst the cytoplasmic grain counts were comparable. With the highest probe concentration the chromosomal grain count was comparable to that of the 0.2 µg/ml probe, yet the cytoplasmic grain count was higher. This data clearly illustrates the effects of pDG 73 probe concentration on signal to noise ratio.

When the hybridisations involving pDG 5 were analysed in the same manner, similar results were obtained, except the data indicated that the optimal probe concentration was 0.1 µg/ml, not 0.2 µg/ml, as was the case for the pDG 73 probe. Other workers have reported successful *in situ* localisations at probe concentrations ranging from 0.1 µg/ml to 0.5 µg/ml (Harper *et al.*, 1982; Trent *et al.*, 1982).

It appears that the optimal probe concentration varies with factors such as the nature and size of the probe, specific activity and hybridisation conditions.

After analysis of the data, it was found that only chromosome number 2 was significantly labelled when probed with pDG 73 DNA. Control experiments have shown that pBR 322 DNA does not hybridise

to genomic DNA isolated from the native cat cell dv♂ or to DNA isolated from native cat liver. This leads to the conclusion that α-globin-like sequences are located on chromosome number 2 of the native cat. The number and nature of native cat embryonic and foetal α-globin genes is unknown, as is their degree of cross-hybridisation to an adult α-globin probe under the conditions of *in situ* hybridisation. It is difficult to determine whether the localisation of α-globin-like sequences to chromosome number 2 represents the location of only the adult genes or represents the adult genes plus related sequences such as embryonic, foetal or pseudogenes. Whilst chromosome number 2 was the only chromosome which displayed significant labelling, it could be argued that the resolution necessary to detect unique sequences had not been achieved in these experiments and that, for example, three copies of the adult gene sequence are present on chromosome number 2 in tandem, but a single copy of an adult α-globin gene is present on another chromosome, but was not detected because the *in situ* hybridisation procedure had not been optimised to a degree where hybridisation to a truly unique sequence would be detected. It is possible to level this criticism at all *in situ* hybridisation experiments. A calculation involving the total expected signal with network formation over a unique site and autoradiographic emulsion efficiency, leading to an expectation of the number of silver grains deposited per day would not be of sufficient accuracy to determine whether one or two copies of a gene are being detected in an experiment. These weaknesses in the *in situ* procedure clearly illustrate the necessity for the use of at least two independent methods of gene assignment before a localisation is confirmed. The sensitivity of the *in situ* technique also may not be such that slightly diverged α-globin sequences, which will not hybridise as efficiently to the pDG 73 probe as adult sequences, will

be detected in these experiments.

When chromosome number 2 was divided into thirds, as described in section 7.2.3, the central third of the chromosome was significantly labelled, adding further weight to the assignment.

Labelling was found to be slightly less significant ($P>0.9$) probing at a concentration of 0.4 µg/ml and non-significant ($P<0.9$) for the 0.05 µg/ml probe. The most likely explanation for this is that at the low probe concentration (0.05 µg/ml) any hybridisation was obscured by the low signal, and at the high concentration (0.4 µg/ml) the high background made it necessary to analyse more than 40 metaphases before a result significant at the 5% level was obtained. These results vindicate the use of the "chromosome:cytoplasm" silver grain ratio as a rough indication of signal to noise ratio in *in situ* hybridisation experiments.

When pDG 5 DNA was used as probe, chromosome number 4 was found to be significantly labelled over all other chromosomes, and division of the chromosome into thirds showed significant labelling over the central third. As in the case of the α-globin probe, it is not possible to unequivocally state that all sequences which hybridise to pDG 5 are located on chromosome number 4. It is clear, however, that adult β-globin sequences are present on chromosome number 4.

An examination of the effects of pDG 5 probe concentration showed an identical pattern to that of the pDG 73 probe (Table 7.5). A non-significant result was obtained at a probe concentration of 0.05 µg/ml after the analysis of 40 metaphases. A non-significant result at the 5% level was also obtained for the 0.4 µg/ml probe concentration, but this is once again thought to reflect the effects of relatively high autoradiographic background on small sample sizes.

Although the *in situ* experiments detailed in this chapter have

achieved the stated aim of localisation of α- and β-globin sequences to native cat chromosomes, not all variables affecting *in situ* hybridisation efficiency were investigated. The effects of probe concentration have been discussed. Another variable to be investigated was the effect of exposure time on signal to noise ratio. All data discussed so far are from autoradiographs which were developed after an exposure time of approximately three weeks. This exposure time was chosen on the basis of the exposure times used by several authors using tritiated probes of similar specific activity and concentration (Harper *et al.*, 1981; Harper *et al.*, 1982; Trent *et al.*, 1982). Table 7.6 shows the effect of increased exposure time on the chromosome:cytoplasm silver grain ratio using pDG 73 DNA at 0.1 µg/ml. The chromosome: cytoplasm ratio, which appears to be a reasonable measure of the relative signal to noise ratio, decreased (i.e. signal relatively lower) as the exposure time was increased. Increased exposure time resulted in an increase in both specific (central third of chromosome number 2) and non-specific labelled chromosomal sites. However, increased exposure time resulted in a decreased signal to noise ratio, presumably due to saturation of specifically labelled sites, whilst the non-specific sites due to background radiation and non-specifically bound probe continued to increase. It appears then, that no added advantage would be gained through an increase of exposure time beyond three weeks, under the conditions used in these experiments.

Although *in situ* experiments involving both the pDG 73 and pDG 5 probes have resulted in the localisation of globin sequences to chromosomal regions of native cat chromosomes, it is desirable to be able to specifiy a localisation in terms of a specific chromosome band. Methods of chromosome banding after denaturation and subsequent *in situ* hybridisation have been described by several workers. Chandler and

Yunis (1978) have described a G-banding method for prometaphase chromosomes following hybridisation using a staining/destaining procedure with Wright's stain. Lawrie and Gosden (1980) have described a methodology for the Q-banding of human chromosomes after hybridisation. Methods for post-hybridisation banding are notoriously unreliable (A. de la Chapelle, pers. comm.). Chromosome denaturants not only denature chromosomal DNA, but also chromosomal protein, thus, stringent denaturation cannot be followed by G-banding. An alternative to post-hybridisation banding is to band chromosomes before denaturation and hybridisation and subsequently photograph and record large numbers of metaphases for comparison after autoradiography. The main advantage of the prebanding procedure is that metaphase chromosomes can be positively identified even if severe denaturation procedures make subsequent identification difficult. The major disadvantage of the prebanding/prephotographing procedure is that it is extremely time consuming.

Attempts were made to induce post-hybridisation banding of native cat metaphase chromosomes using the method of Chandler and Yunis (1978). A number of variables such as stain concentration, stain/destain time and frequency, and staining temperature were investigated. Significant G-banding was not induced under any of the conditions examined. The lack of banding could have been due to over-denaturation of the chromosomes or the failure to use prometaphase preparations, although banding was not observed when undenatured metaphases were tested, some of which were prometaphase spreads. Attempts to band the chromosomes using fluorochromes such as quinacrine dihydrochloride or Hoechst 33258 were unsuccessful. An extensive investigation of the response of dasyurid marsupial chromosomes to fluorochromes has not resulted in the reliable induction of banding, even under conditions which result in

banding for other marsupial and placental species (P. Sharp, pers. comm.). The reason for this phenomenon is not yet understood. Dasyurid chromosomes can be reliably G-banded using a standard trypsin/Giemsa approach, but it was reasoned that the treatment with trypsin and subsequent exposure to xylenes and alcohols during destaining might interfere with hybridisation efficiencies. Higher backgrounds and variable hybridisation efficiencies have been observed to occur after exposure to xylenes by Warburton *et al.* (1976).

The various native cat chromosomes are easily identified on the basis of their size and morphology. In the absence of any pre- or post-hybridisation banding, it was decided to divide the appropriate chromosomes according to length. A problem with chromosome number 2 is that it is a metacentric and, in the absence of banding, impossible to orientate from one metaphase to another. This did not, however, preclude localisation from a chromosome region but, rather, reduced the number of regions from three to two, i.e. central region or telomeric region.

In summary, *in situ* hybridisation techniques have been used to localise α- and β-globin sequences to chromosomes numbers 2 and 4, respectively. Each sequence was further localised to the central third of the respective chromosome.

CHAPTER 8

GENE MAPPING BY FLOW CYTOMETRY AND FLOW SORTING OF MARSUPIAL CHROMOSOMES

8.1 INTRODUCTION

The quantitative classification and purification of isolated metaphase chromosomes is facilitated by the application of flow cytometry and sorting (Gray *et al.*, 1975; Stubblefield *et al.*, 1975; Carrano *et al.*, 1976). In this approach, individual chromosomes are released from metaphase cells into an isolation buffer to which is added a DNA specific fluorochrome. The stained chromosomes are then forced to flow one-by-one through the illumination beam of a flow cytometer (Van Dilla *et al.*, 1974); the resulting fluorescence is measured and used to quantitatively classify each chromosome. Thousands of chromosomes can be measured each second so that a statistically precise fluorescence frequency distribution, representative of the total chromosome population, can be accumulated in a few Typically, these fluorescence distributions have several peaks superimposed on an exponential background continuum. Each peak represents a distinct group of chromosomes (ideally one peak for each chromosome type): its mean is proportional to the chromosomal group fluorescence and its area is proportional to the frequency of occurrence of chromosomes within that group. The background continuum is produced by chromosomal debris, clumps and other fluorescent cellular debris.

The information contained in a fluorescence distribution, namely chromosome group mean and frequency of occurrence, is similar to that obtained by conventional karyotyping so that the fluorescence distributions can, in a sense, be used as karyotypes (Gray *et al.*, 1975b; Carrano *et al.*, 1978b). The flow karyotypes are particularly

sensitive to aberrations which occur homogeneously in all cells, and which cause changes in chromosomal fluorescence (seen as a change in peak mean) or in chromosome loss or gain (seen as a change in peak area). Random aberrations which occur only in one or a few cells contribute to the background continuum in the fluorescence distribution. Carrano *et al.* (1978a) have used the height of the background continuum as a semiquantitative estimate of the frequency of heterogeneous chromosome aberrations induced by mutagens.

Flow sorting (Gray *et al.*, 1975; Carrano *et al.*, 1976) allows chromosomes within each fluorescence group to be collected. Flow sorting enables the determination of the identity of the chromosomes which produce the fluorescence peaks (Carrano *et al.*, 1979; Carrano *et al.*, 1982). Flow cytometry followed by flow sorting enables the purification of specific metaphase chromosomes to a degree of homogeneity not possible with other chromosome fractionation methods such as electrofocusing (Handel *et al.*, 1972), sucrose gradient centrifugation (Hughes *et al.*, 1979), zonal centrifugation (Burki *et al.*, 1973), or separation at low gravity forces (1-15g) through discontinuous sucrose gradients (Collard *et al.*, 1979, 1982). The main disadvantages of flow sorting compared to other chromosome fractionation methods are the amount of time required to accumulate large numbers of purified chromosomes, and the cost of the procedure.

Purified metaphase chromosomes can be used for a variety of biochemical investigations. Several genomic DNA libraries have been constructed from purified chromosomes (Davies *et al.*, 1981; Disteche *et al.*, 1981) enabling the gene sequences from specific chromosomes to be isolated. Disteche *et al.* (1982) have purified the "Cattanach translocation" chromosome using flow sorting to enable a study of the biochemical nature of an inactive mammalian X chromosome. Sawin *et al.*

(1979) have examined the *in vitro* transcripts generated from purified Chinese hamster metaphase chromosomes. Several laboratories have used purified metaphase chromosomes for gene mapping (Lebo *et al.*, 1979, 1982; Hughes *et al.*, 1979; Collard *et al.*, 1982). Gene mapping using purified chromosomes usually involves the extraction of DNA from specific chromosomes and subsequent filter hybridisation studies of that DNA using nucleic acid probes. This approach to gene mapping is useful where somatic cell hybrids are not available for the species one wishes to map and where low gene sequence copy number makes mapping *in situ* difficult. The main disadvantages to this approach are the limited chromosomal resolution for some species, the requirement for large numbers of metaphase chromosomes as the starting material, and the inability to map any gene for which there is no appropriate nucleic acid probe.

This chapter describes the characterisation of the metaphase chromosomes of the native cat by flow cytometry and flow sorting, and the use of purified native cat metaphase chromosomes to map α- and β-globin genes.

8.2 RESULTS

8.2.1 Chromosome Preparation

The native cat cell line dh♂, described in section 6.2, was cultured and cells at metaphase harvested as detailed in section 10.4 Metaphase chromosomes were released from the cells by gentle vortexing in the presence of digitonin (Sillar and Young, 1981). Approximately 4×10^7 metaphase cells were routinely obtained from six $95cm^3$ culture vessels. After staining at a final concentration of 50μg/ml ethidium bromide, the chromosomes were held at 4°C for up to 2 weeks before

flow cytometry without any discernible change in flow karyotype over that time.

8.2.2 Flow Cytometry and Flow Sorting

The operational parameters of the FACS IV have been described in section 10.2.21 and only parameters directly affecting the resolution of native cat chromosomes will be described here.

Much of the background fluorescence in a flow analysis of chromosomes comes from autofluorescent pieces of cellular debris, such as membrane fragments and other cellular debris. Such pieces of debris, although fluorescing with an intensity comparable to that of some of the chromosomes, produce a larger signal based upon forward light scatter because of the contracted nature of the chromosomes. Thus by programming the FACS to ignore or "gate" certain data channels, based upon forward light scatter, the measured background fluorescence can be reduced. This is shown in Table 8.1, where the percentage of signals due to chromosomes (and is thus a measure of background) has been calculated as described in section 10.2.21 for the fluorescence distribution.

Table 8.1

Effect of forward light scatter gating on measured fluorescent background.

Gate on scatter channels	Percantage of signals due to chromosomes
1 - 10	51.4
1 - 15	53.0
1 - 20	66.1
1 - 25	68.4

Using a flow rate of 500 chromosomes per second, the fluorescence distribution in Figure 8.1 was obtained for the dh♂ cell line. The dots represent the actual data points and the solid line represents the results of fitting a least squares Gaussian distribution program to the experimental data. The computer program used for fitting of normal and exponential distributions to the experimental data was modified from Bevington (1969) as described by Moore II (1975), and is presented in Appendix I. The use of such a program enables the determination of the mean and area of each fluorescence peak, taking into account random noise and the underlying exponential continuum.

The identification of the chromosome types contained in each peak was accomplished by sorting approximately 10,000 chromosomes from each fluorescence distribution peak on to glass microscope slides which were then fixed as described in section 10.2.21. The chromosomal constitution was determined by the examination of at least one hundred Giemsa stained chromosomes. The areas from which representative sorts were taken are indicated in Figure 8.1. Table 8.2 shows the chromosome type responsible for each fluorescence peak and its corresponding purity. Examples of each sort are shown in Figure 8.2.

By applying the least squares fitting program to the data, the relative areas of each peak were determined (relative frequency of each chromosome type), along with the peak means (relative DNA content of each chromosome). It can be seen in Table 8.3 that the flow karyotype corresponds well to the data generated by conventional karyotyping procedures and relative DNA contents based upon length measurements.

Figure 8.1

Fluorescence distribution of native cat metaphase chromosomes measured by flow cytometry. Chromosomes were stained with ethidium bromide and the fluorescence distribution was determined by measurement on a FACS IV. The dots represent the actual data points and the solid line is the result of fitting normal and exponential distributions to the experimental data.

Bars indicate the areas of the distribution which were sorted on to glass microscope slides and cytologically examined.

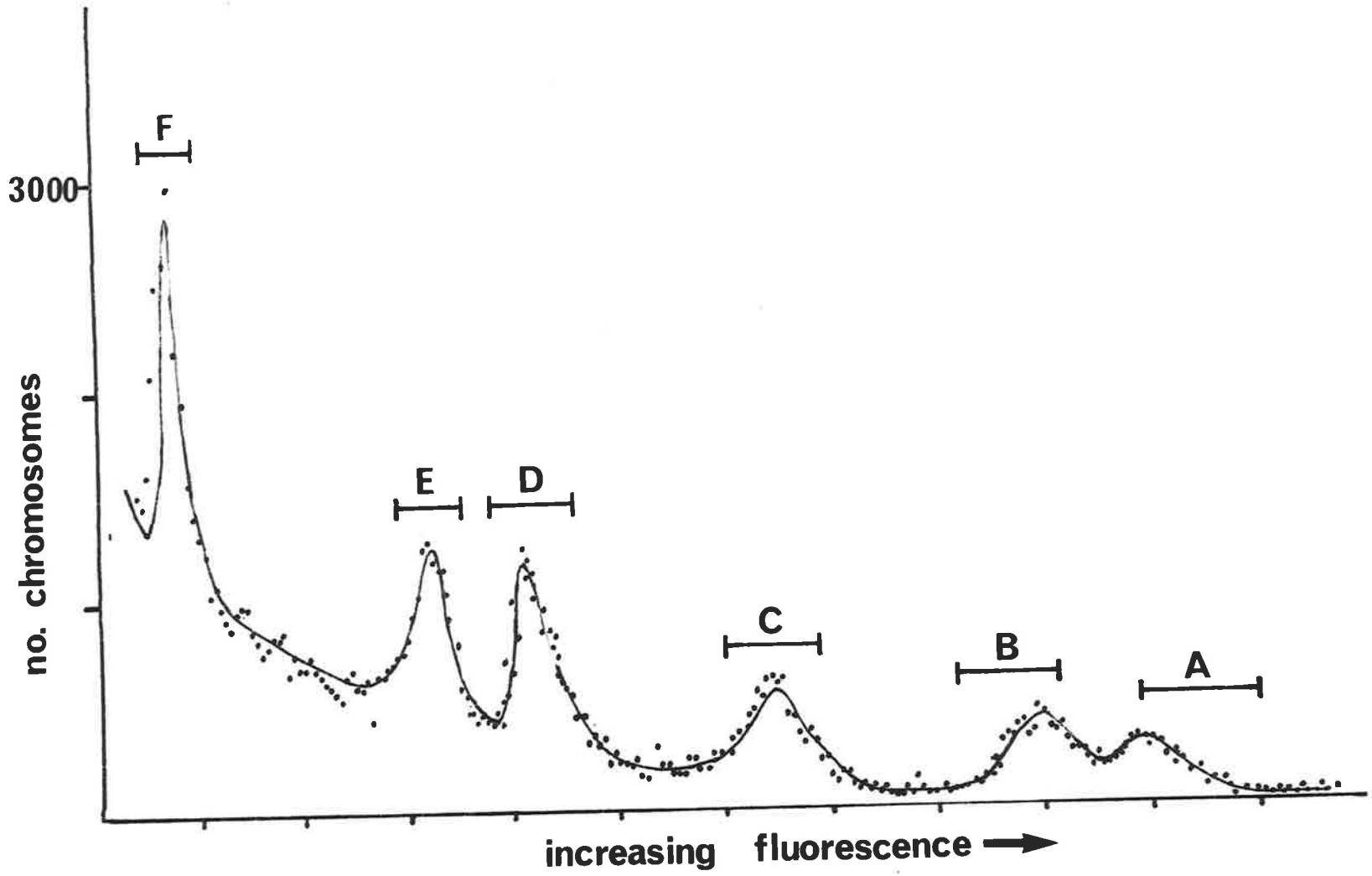

F
3000
E
D
C
B
A
no. chromosomes
increasing fluorescence

Table 8.2

Chromosome purity after flow sorting.

Peak designation	Chromosome designation	Sort purity %
A	1	72
B	2 3	37 39
C	4	89
D	5	86
E	6	84
F	X	91

Table 8.3

Chromosome measurement by flow cytometry.

Peak designation	Chromosome designation	Relative DNA content		Relative chromosome frequency	
		Flow cytometry	Chromosome length	Flow cytometry	Expected
A	1	1.00	1.00	2.09	2
B	2,3	0.92	0.95 0.87	3.89	4
C	4	0.67	0.66	1.95	2
D	5	0.44	0.48	2.06	2
E	6	0.34	0.35	2.05	2
F	X	0.14	0.15	0.96	1

Figure 8.2

Examples of sorted chromosomes. Chromosomes were sorted from areas of the fluorescence distribution as indicated in Fig. 8.1, fixed, stained with Giemsa, and examined. Letters correspond to the letters indicated in Fig. 8.1.

A. Chromosome 1

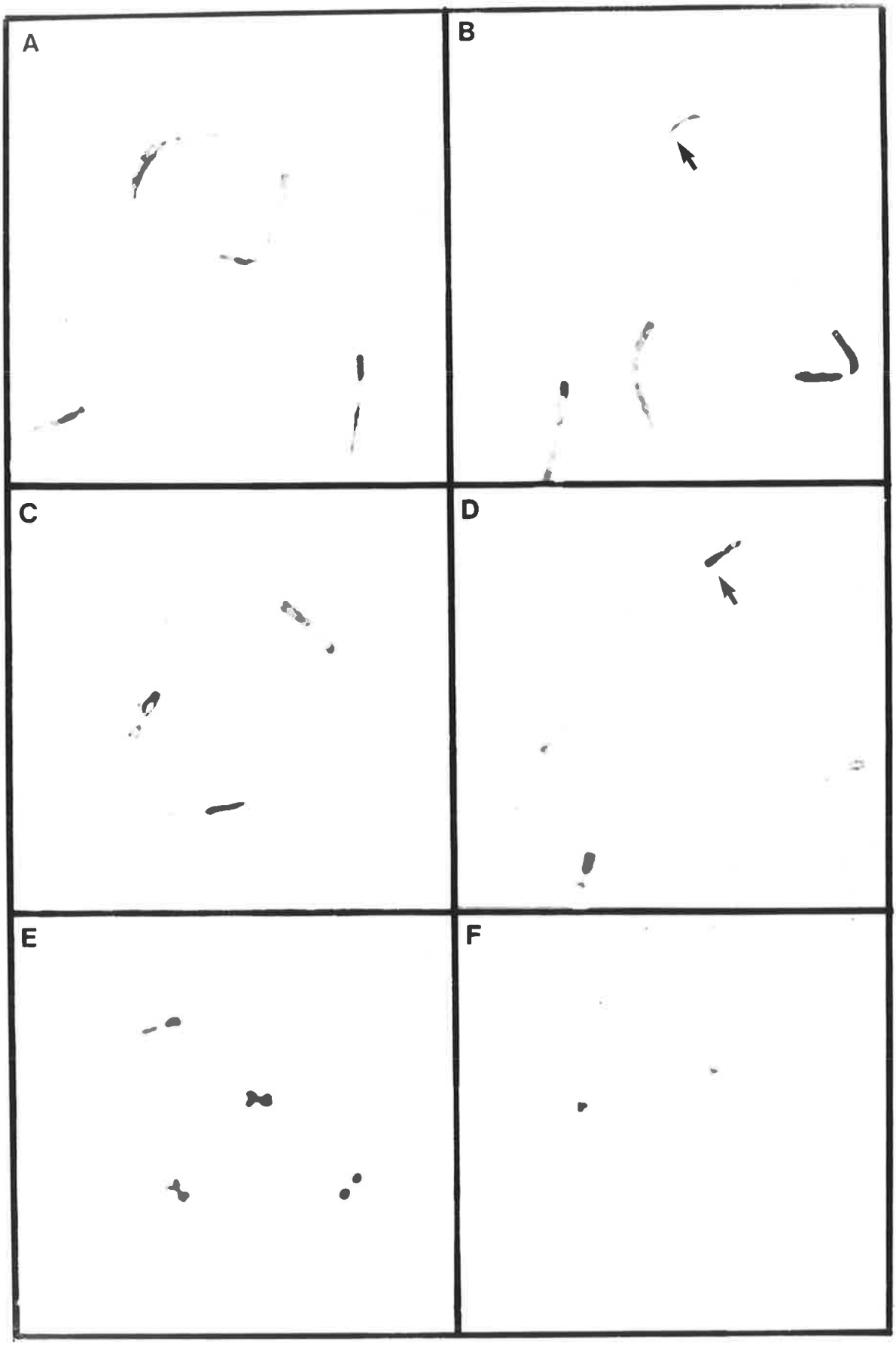
B. Chromosomes 2 and 3 (chromosome 3 arrowed)

C. Chromosome 4

D. Chromosome 5 (unidentified chromosome arm arrowed)

E. Chromosome 6

F. X chromosome


A
B
C
D
E
F

8.2.3 Mapping of Globin Genes Using Sorted Chromosomes

Due to the expensive and time consuming nature of flow sorting, it was decided that the most efficient sorting protocol (requiring the least number of sorting runs) would involve an initial fractionation into groups, rather than purify each chromosome individually. Having determined which subset of chromosomes contained the native cat globin sequences, that group was further purified as indicated in the following sections.

8.2.3.1 α-globin genes

An initial sorting run was performed which separated the native cat chromosomes into two groups, one containing chromosomes 1, 2 and 3, and the other containing the rest of the genome. At least 3×10^6 of each chromosome type was present in each sort. Typically 10μg of the dh♂ DNA was sufficient to enable the detection of native cat α- and β-globin-like sequences (Figure 8.3). This amount of DNA represents approximately 100,000 copies of the native cat genome, and hence 3×10^6 of each chromosome type represents a 30X excess over the amount of DNA required to detect native cat α- and β-globin sequences. The purity of each sort was monitored and each was found to consist predominantly of the chromosome types indicated (see Table 8.2). DNA was extracted from each sort as described in section 10.2.21. Using *in vivo* labelled tritiated DNA derived from human tissue culture cells (generously supplied by Dr. R. Harvey), the DNA recovery after undergoing the extraction procedures described in section 10.2.21 was estimated to be approximately 50%.

Figure 8.4 shows that after the sort containing chromosomes 1, 2 and 3 was digested with restriction endonuclease EcoR1, specific hybridisation of the pDG 73 (α-globin) probe to that sort could be

Figure 8.3

Southern blot analysis of native cat RNA. DNA extracted from the native cat cell line dv ♂ was digested with Eco R1, electrophoresed on a 0.8% agarose gel, transferred to nitrocellulose paper, probed, and washed to a stringency of 1 x SSC, 0.5% SDS at 65°C. Filters were exposed for 4 days in the presence of intensifying screens at -80°C.

A. Eco R1 cut DNA probed with ^{32}P-labelled pDG 5 DNA.

B. Eco R1 cut DNA probed with ^{32}P-labelled pDG 73 DNA.

Molecular weights were determined from a Hind III digest of λ phage, visualised by staining with EtBr.

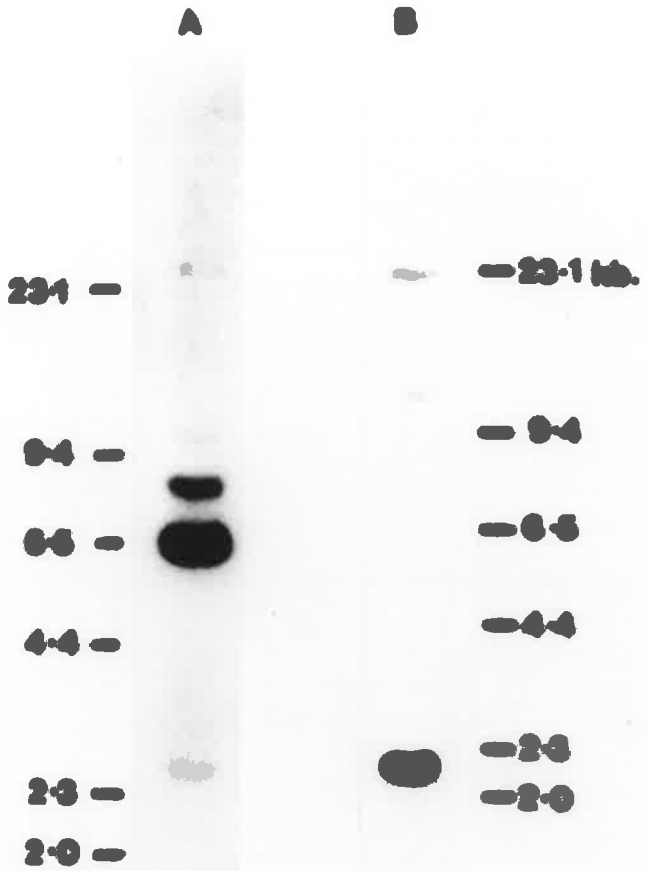
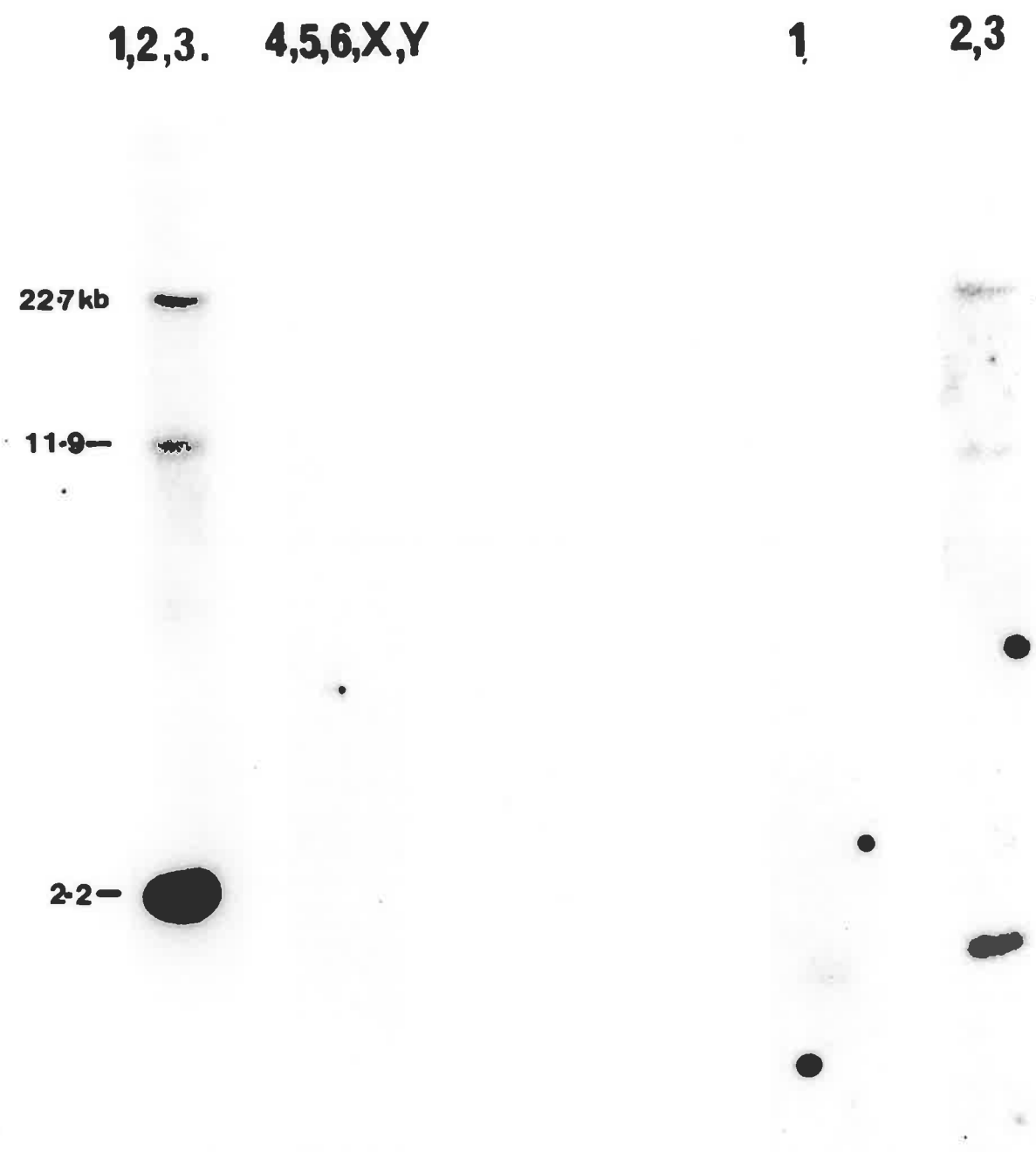

A
B
23·1
9·4
6·6
4·4
2·3
2·0
23·1 kb.
9·4
6·6
4·4
2·3
2·0

Figure 8.4

Filter hybridisation of ^{32}P-labelled pDG 73 DNA to DNA extracted from flow sorted chromosomes. DNA was digested with Eco R1, fractionated on a 0.8% agarose gel, transferred to nitrocellulose paper, probed, and washed to a stringency of 1 x SSC, 0.5% SDS at 65°C. Filters were exposed for 4 days in the presence of intensifying screens at -80°C.

Molecular weights of the hybridising DNA fragments were determined using molecular weight markers generated by a Hind III digest of λ phage, which was visualised by staining with EtBr.


1,2,3.
4,5,6,X,Y
1
2,3
22·7kb
11·9
2·2

detected, whereas no hybridisation was detected to an EcoR1 digest of chromosomes 4, 5, 6, X and Y.

When chromosomes 1, 2 and 3 were sorted into two groups, one containing chromosome 1 and the other chromosomes 2 and 3, hybridisation was observed to the sort containing chromosomes 2 and 3 but not to the group containing chromosome 1 (Fig. 8.4). There was, however, a very weak hybridisation detected to the chromosome 1 sort, presumably reflecting the quantitative detection of chromosomes 2 and 3 contamination of the chromosome 1 sort. The contamination of the chromosome 1 sort with chromosomes 2 and 3 was approximately 10% of the total number of chromosomes in that sort.

The DNA extracted from each pair of chromosome sorts was electrophoresed on the same gel, transferred to the same nitrocellulose filter and probed in the same hybridisation mixture so that any differential hybridisation due to experimental conditions was unlikely.

8.2.3.2 β-globin genes

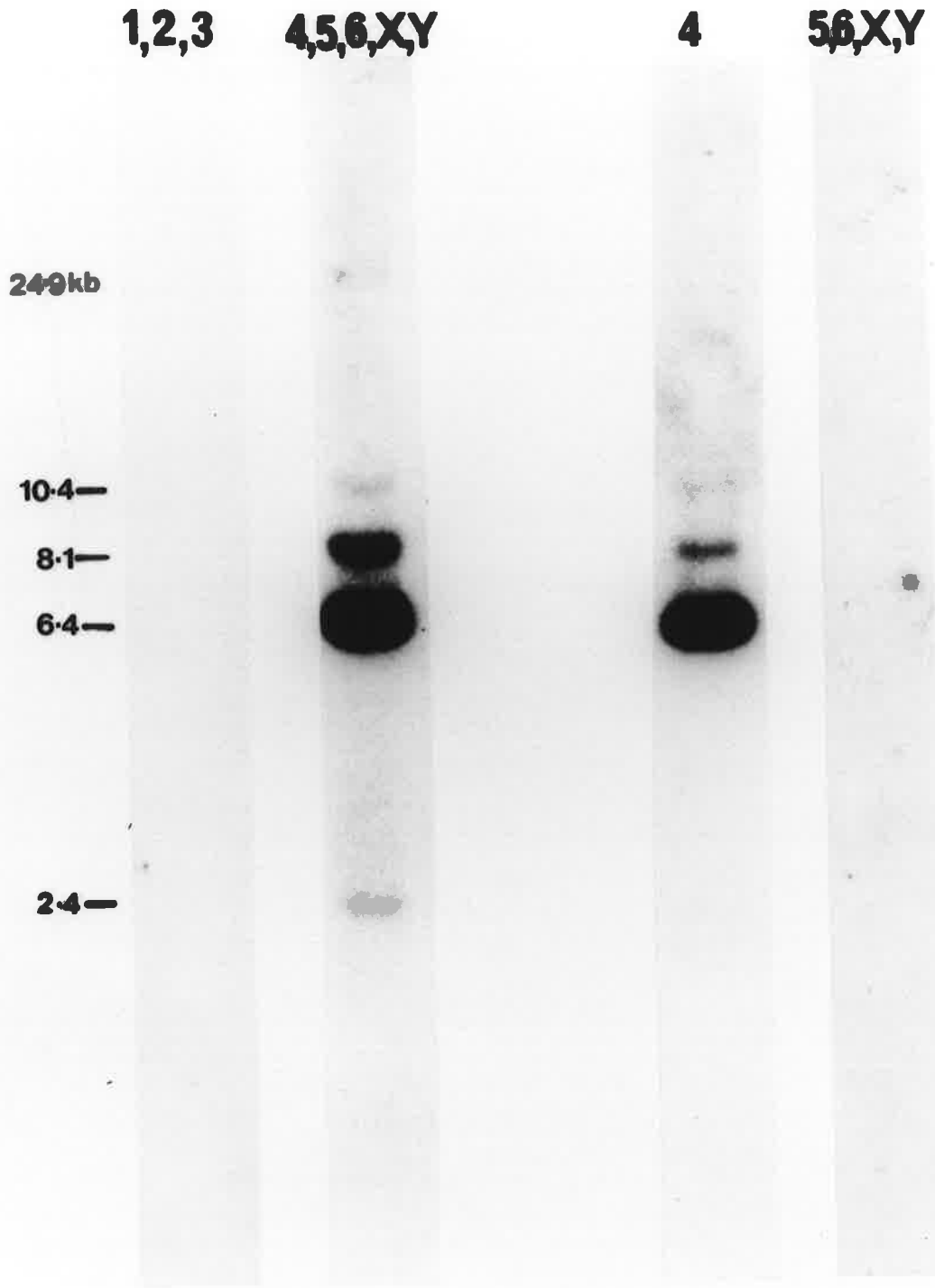
Native cat chromosomes were sorted into two groups, one group containing chromosomes 1, 2 and 3, and the other containing chromosomes 4, 5, 6, X and Y. The DNA was extracted from each chromosome group, digested with EcoR1, electrophoresed and transferred to nitrocellulose.

Hybridisation of plasmid pDG 5 (β-globin) to each sort resulted in the detection of β-globin related sequences in the sort containing chromosomes 4, 5, 6, X and Y (Fig. 8.5). The next sort analysed was that of chromosome 4 compared to chromosomes 5, 6, X and Y. The sort containing chromosome 4 was found to contain all of the DNA sequences hybridising to the pDG 5 probe.

Figure 8.5

Filter hybridisation of ^{32}P-labelled pDG 5 DNA to DNA extracted from flow sorted chromosomes. DNA was digested with Eco R1, fractionated on a 0.8% agarose gel, transferred to nitrocellulose paper, probed, and washed to a stringency of 1 x SSC, 0.5% SDS at 65°C. Filters were exposed for 4 days in the presence of intensifying screens at -80°C. The sample track containing chromosomes 5, 6, 7, X, was exposed for 6 days.

Molecular weights of the hybridising DNA fragments were calculated using molecular weight markers generated by a Hind III digest of λ phage which was visualised by staining with EtBr.


1,2,3
4,5,6,X,Y
4
5,6,X,Y
24·9kb
10·4
8·1
6·4
2·4

8.3 DISCUSSION

The FACS machine used in these studies is routinely used for cell studies and was not "set up" for chromosome work. The modifications of the FACS machine and its accompanying data processing made by some users (Dean and Pinkel, 1977; Carrano *et al.*, 1978; Young *et al.*, 1981) specifically for high resolution chromosome sorting were not available, nor was it possible to introduce them. Consequently only a limited number of variables could be investigated for the optimisation of conditions for flow cytometry and sorting of chromosomes. The effects of particle flow rate have been investigated elsewhere (Davies *et al.*, 1981: Cremer *et al.*, 1982). A flow rate of approximately 500 chromosomes per second was chosen on the basis of the work by Davies *et al.* (1981), as the minimum flow rate practical for preparative purposes. The native cat chromosomes were routinely stained with ethidium bromide before flow cytometry. Whilst the use of a general DNA fluorochrome such as ethidium bromide enables an accurate estimate of the relative DNA content of each chromosome, base-specific fluorochromes such as Hoechst 33258 (Ho) and chromomycin A3 (CA3) give lower coefficients of variation than EtBr and thus lead to better resolution (Langlois *et al.*, 1980; Carrano *et al.*, 1979). In these experiments it was not possible to use either Ho or CA3-stained chromosomes due to the high cost of operating the excitation laser in the wavelength range necessary for these fluorochromosomes.

Despite the lack of optimisation of the FACS IV for chromosome cytometry and sorting, the native cat chromosomes could be resolved into six distinct distributions, with each distribution, except for one, consisting predominantly of a single chromosome type. Using ethidium bromide staining, chromosomes 2 and 3 could not be resolved.

Although their relative DNA contents are almost identical (Hayman *et al.*, 1982), it is possible that some cytochemical differentiation could have been achieved using base specific fluorochromes. The fluorescence of native cat Y chromosome was not significantly different to that of background debris, and although it does not appear as a separate distribution, it was included in chromosome sorts, along with background "debris".

The correspondence of the native cat flow karyotype to that of the conventional karyotype in both relative chromosome frequency and relative DNA contents is consistent with the results obtained from other species (Gray *et al.*, 1975; Carrano *et al.*, 1976). It is interesting to note that the relative DNA content of 0.13 for the native cat X chromosome represents an actual value of approximately 3% of the total DNA of the genome, being the smallest mammalian X chromosome so far described. This result has been confirmed for another dasyurid marsupial, *Sminthopsis crassicaudata*, by Hayman *et al.* (1982) and Wainwright (unpublished results).

The determination of sort purity was achieved by the examination of general stained chromosomes. Gray *et al.* (1975) and Carrano *et al.* (1976) have used G-banding and Q-banding of sorted chromosomes to determine sort purity. Banding of the sorted native cat chromosomes was attempted using a variety of staining methods, but were all unsuccessful in the induction of banding. The metaphase chromosome suspensions were prepared in the presence of polyamines, which resulted in a greater degree of chromosome contraction than those methods employed by Carrano and co-workers using hexylene glycol-based homogenisation media (Sillar and Young, 1981). It is possible that the contracted state of the chromosomes inhibits the banding. The polyamine-based chromosome preparation method allows a more gentle

lysis of metaphase cells, and results in better differentiation between chromosomes and debris than hexylene glycol-based methods due to the contracted nature of the chromosomes, and thus results in better resolution with a smaller background continuum of debris and chromosome fragments (Young *et al.*, 1981; Sillar and Young, 1981). The native cat chromosomes are of distinctive size and morphology (see Chapter 6), and the lack of banding on sorted chromosomes was not considered a serious problem, although due to the highly contracted state of some of its chromosomes, there was a degree of uncertainty of identification. It is likely that the sort purity data presented in Table 8.2, although repeatable between experiments, would differ slightly from sort purity estimates based upon banded chromosomes.

A genomic mapping study of the native cat globin genes has not been undertaken. In the absence of such a study no conclusions can be reached as to which gene sequences are represented by the bands displayed in Figures 8.4 and 8.5. It is presumed that all of the hybridising restriction fragments contain globin-like sequences, being major or minor, adult or foetal, or functional or pseudo-genes. No hybridisation of pBR 322 sequences to native cat DNA has been detected (see Chapter 7).

Hybridisation of the native cat α-globin probe to EcoR1 digests of chromosomes 2 and 3 was observed. As has been previously discussed, chromosomes 2 and 3 could not be resolved by flow cytometry and sorting, so a positive assignment of the α-globin gene sequences to chromosome 2 and/or 3 was not possible. *In situ* hybridisation experiments presented in the previous chapter assigned α-globin sequences to chromosome 2. The results from flow sorting, whilst not directly confirming the *in situ* results, do not contradict them.

Hybridisation of the native cat β-globin probe to an EcoR1 digest

of chromosome 4 was observed. This result confirms the result obtained with the *in situ* hybridisation experiments, detailed in Chapter 7. Although the decision to perform a sort separating chromosome 4 from 5, 6, X and Y was based upon results from the *in situ* hybridisation experiments, this in no way detracts from the independent and unbiased nature of the gene assignment, since all chromosomes other than chromosome 4 clearly do not contain any β-globin-like sequences (Fig. 8.5). It is interesting to note that all of the native cat β-globin related sequences detected in these experiments appear to be present on a single chromosome, a result which has been observed for all species for which mapping data are available.

In summary, gene mapping experiments using metaphase chromosomes purified by flow cytometry have confirmed the asyntenic arrangement of the native cat α- and β-globin genes

CHAPTER 9. CONCLUDING DISCUSSION

The research presented in this thesis was, to a large extent, preliminary in nature. As a first step towards the elucidation of the organisation of the globin genes in a marsupial, more specifically, a dasyurid marsupial, it was decided to attempt to define the chromosomal location of the globin genes in the native cat. The unique evolutionary position of marsupials within the mammals provides the opportunity for valuable evolutionary comparisons between two infraclasses. Somatic cell genetic techniques have been used for gene mapping with only limited success in marsupials (see section 2.5) and in particular, dasyurid x rodent cell hybrids have not been isolated which contain cytologically identifiable marsupial chromosomal material. It was therefore decided to map the globin genes by constructing recombinant DNA probes and using such probes for *in situ* hybridisation, and by hybridisation to DNA extracted from purified metaphase chromosomes.

As in many other species, reticulocytes were found to provide a ready source of globin mRNA and globin polypeptides in the native cat. Globin polypeptides from native cat blood, and those resulting from *in vitro* translation of native cat mRNA, were identified on the basis of indirect evidence. However, subsequent nucleotide sequencing of the cDNA clones obtained from reticulocyte RNA showed that the original identification of native cat globin protein was correct.

No evidence was found in these studies of the existence of minor globin species present in the blood of adult native cats. The presence of minor globin species in human blood can be detected by isoelectric focusing or by extensive fractionation on CMC columns. Native cat blood proteins were not analysed by isoelectric focusing

nor were any of the globin peaks observed after fractionation on CMC columns further analysed by extensive CMC chromatography. If either of these procedures were employed, then one might reasonably expect to detect the presence of globin species other than the major adult α- and β-globin.

As far as could be determined, the native cat α- and β-globin mRNAs were not remarkably different from other vertebrate α- and β-globin mRNAs. The absence of a complete copy of the β-globin mRNA molecule amongst the reticulocyte library was disappointing. Assuming the RNA used to construct the library was not significantly degraded, it should be possible to determine the sequence of the entire native cat β-globin mRNA molecule by direct RNA sequencing, using the M13 DNA subclones as primers for the "dideoxy" reactions. Alternatively, full length clones might be obtained by adopting a different cloning strategy.

Evolutionary studies using marsupial globin gene sequences are of great interest because the date of marsupial/eutherian divergence (approximately 150 million years ago) is between the date of radiation of eutherian mammals (85 million years) and the divergence of the avian/non-avian lines (270 million years). Thus trends in molecular evolution can be more readily be defined with the addition of data from the marsupials. The reassessment of the date of the marsupial/eutherian divergence using globin polypeptide data gave an estimate approximately 25 million years later than the previously accepted estimate. The new estimate obtained in this study was within the upper range of the 95% confidence limit of the previous estimate of Air *et al.* (1971). The use of a significantly larger amount of both marsupial and eutherian data in this study than in previously published

studies was expected to provide a much more accurate estimate of the marsupial divergence date. This was vindicated in the significantly narrower 95% confidence limits obtained compared to other studies using globin polypeptide data.

A most interesting observation to result from the globin amino acid comparisons was the apparently anomalous rate of evolution of the α-globin polypeptide of a South American marsupial, the opossum, *Didelphis marsupialis*. The determination of the nucleotide sequence of the opossum α-globin gene, its flanking sequences, and the sequence of any other α-globin genes, would provide some very interesting data. Whether the mechanism of "concerted evolution" is evident in the opossum α-globin cluster is just one many interesting questions which could be asked of the nature of the opossum globin gene system.

The estimates of the amount of silent site substitution observed between representatives of various vertebrate classes reinforced the findings of Perler *et al.* (1980). The addition of the nucleotide sequence data from a marsupial and from an amphibian appeared to confirm that there might be two rates of silent site substitution, dependent upon the amount of evolutionary divergence between the two vertebrates being examined. At periods of up to about 100 million years divergence, the rate of silent site substitution in globin genes appears to proceed at a rate approximately 3-5X that observed at periods greater than 100 years divergence. The absence of data for divergences of less than 85 million years was found to seriously limit the accuracy of any estimate of the initial rate of silent site substitution.

In situ hybridisation experiments and hybridisation to DNA extracted from purified metaphase chromosomes showed that the α- and

β-globin gene sequences in the native cat are asyntenic, with the α-globin-like sequences being on chromosome 2 and the β-globin-like sequences being on chromosome 4. Blotting experiments to genomic DNA have not been performed to determine the approximate organisation of the native cat globin sequences. Consequently, it was not possible to identify DNA fragments of specific lengths as containing specific globin gene sequences. It appeared that all α-globin-like sequences (e.g. major adult, pseudo-, embryonic genes etc.) were most likely syntenic on chromosome 2, as were all of the β-globin-like gene sequences on chromosome 4. Clones containing sequences hybridising to the pDG 5 and pDG 73 probes have been obtained from a lambda phage genomic library of the wallaroo, *M. robustus* (A. van Daal, pers. comm.). Analysis of these clones will provide details of the organisation of the globin genes in a marsupial.

The asyntenic nature of the α- and β-globin genes has now been confirmed for a representative of the Metatheria. Dasyurid marsupials have a low diploid chromosome number (2n=14) compared to the majority of the mammals. If there were no significant selective advantage associated with an asyntenic arrangement of the α- and β-globin genes, then one might expect a higher probability of synteny between these two types of genes as the diploid chromosome number of the animal drops. For this reason the determination of the chromosomal location of the globin gene sequences in an animal such as the Indian muntjac (2n=6♀; 2n=7♂) could provide some interesting data.

The cytological relationships between the various members of the Dasyuridae are well defined. Using the globin gene probes developed in this study, the relationship between G-banding pattern and gene organisation within a mammalian family could be examined. Several

isozyme linkage groups have been defined for a member of the Dasyuridae, *Sminthopsis crassicaudata* (J.H. Bennett, pers. comm.). A study looking for globin polypeptide polymorphism is currently under examination in this species (R. Nancarrow, pers. comm.) in the hope that a chromosomal localisation of one or more of the linkage groups might be facilitated by the demonstration of the linkage of the globin gene loci to the established linkage groups, followed by localisation of the globin genes by *in situ* hybridisation.

Using somatic cell hybrids the gene for LDH-A has been assigned to red kangaroo (*M. rufus*) chromosome 5 (Donald and Adams, 1981). Jeffreys *et al.* (1979) have shown that the gene for LDH-A and the β-globin locus are syntenic on the short arm of chromosome 11 of man. It would be of interest to analyse the *M. rufus* x rodent somatic cell hybrids using the marsupial globin probes to determine whether such a relationship exists in the red kangaroo.

Globin genes have proved to be a useful system for the study of multigene families in a relatively large number of species. Results presented in this thesis have shown that the globin gene system also shows great potential for molecular studies in marsupials from the standpoint of gene organisation and evolution.

CHAPTER 10. MATERIALS AND METHODS

10.1 MATERIALS

Chemicals

All chemicals were of analytical reagent grade or of the highest available purity.

Isotopes

α-^{32}P-dCTP, α-^{32}P-dATP (1500-2000 Ci/mmole) was a kind gift from Dr. R.H. Symons.

^{35}S-methionine (400 Ci/mmole): Amersham.

^{3}H-ATP (113 Ci/mmole) and ^{3}H-dCTP (79 Ci/mmole): New England Nuclear.

^{3}H-GTP (33 Ci/mmole).

Enzymes

RNA-dependent DNA polymerase, calf thymus terminal deoxynucleotidyl transferase, *E. coli* DNA polymerase I, Klenow fragment *E. coli* DNA polymerase, S_1 nuclease, proteinase K and ribonuclease A were all purchased from Boehringer-Mannheim.

Restriction Enzymes were obtained from New England Biolabs. except for:

PstI which was obtained from P.L. Biochemicals.

Calf thymus DNA, yeast tRNA, nucleoside triphosphates were obtained from Sigma.

Poly-U Sepharose: Pharmacia.

Carboxymethyl cellulose: Bio-Rad.

Experimental Animals

Native cats were obtained through the offices of the Tasmanian National Parks and Wildlife Service.

Cell Culture Media

RPMI 1640: Flow Laboratories.

Foetal calf serum: GIBCO, C.S.L.

Penicillin, streptomycin and glutamine: Sigma.

10.2 METHODS

10.2.1 Preparation of Native Cat Erythroid Cells

The procedures followed were essentially those of Nienhuis *et al.* (1974). Anaemia was induced in three native cats by five daily injections of 2.5% phenylhydrazine hydrochloride in 48% ethanol, pH 7.0. On the seventh day after the injections had begun, approximately 50 ml of blood per animal was obtained by cardiac puncture. Blood was delivered into NKM (0.14M NaCl, 5mM KCl, 2mM $MgCl_2$) containing 0.5mg/ml of heparin. The red cells were washed three times in NKM by centrifugation for 10' at 5,000g, with the buffy coat being removed after each wash by aspiration.

10.2.2 Isolation of RNA (Battaglia and Melli, 1977)

All procedures involving RNA employed glassware and solutions sterilised by autoclaving and/or treatment with diethyl pyrocarbonate and KOH (Taylor, 1979). Washed red blood cells were suspended in TKM (0.05M Tris-HCl pH 7.5, 0.03M KCl, and 5mM $MgCl_2$) at a ratio of 1 vol. of packed cells to 2 vol. of buffer and were lysed by addition of Nonidet Np-40 to 0.5%. Lysate was centrifuged at 15,000g for 20 minutes and the supernatant collected. Total cytoplasmic

RNA was purified by three phenol-chloroform extractions followed by two chloroform extractions. The RNA was precipitated in 0.2M NaCl with 2.5 volumes of ethanol and was collected by centrifugation at 15,000g for 20 minutes. The RNA pellet was resuspended in water at a concentration of approximately 4mg/ml and stored frozen at -80°C.

10.2.3 Poly-U Sepharose Chromatography of RNA

(Shapiro and Schimke, 1975)

Up to 8mg of RNA was loaded in a single application to a 5ml column. Samples were brought to a final concentration of 30mM EDTA, 1% lauryl-sarcosine and diluted 4-fold with starting buffer (25% formamide, 0.7M NaCl, 50mM Tris-HCl pH 7.5, 10mM EDTA). After loading the column was washed with ten bed volumes of starting buffer. Bound RNA was eluted by the addition of a solution containing 90% formamide, 10mM $KHPO_4$, 10mM EDTA, 0.2% lauryl-sarcosine, pH 7.5. RNA was collected into 0.4ml fractions, precipitated, and resuspended in water at a final concentration of approximately 1mg/ml.

10.2.4 Sucrose Gradient Fractionation of RNA

(Crawford *et al.*, 1977)

Linear gradients of 10 to 40% sucrose (in 10mM Tris-HCl pH 7.4, 20mM NaCl, and 1mM EDTA) were prepared in SW41 cellulose nitrate tubes. Samples containing from 2 to 0.2mg of RNA in 0.4ml of 10mM Tris-HCl pH 7.4, 1mM EDTA were loaded on each tube and centrifuged at 4°C in the Spinco SW41 rotor for 16 hours at 36,000 r.p.m. (16,000g). The absorbance of 260nm from collected gradients was monitored and the fractions were pooled as noted in section 3.2, precipitated in ethanol, and stored at -20°C.

10.2.5 Preparation and Optimisation of a Wheat Germ Cell Free Translation System (Roberts and Paterson, 1973)

The wheat germ translation extract was prepared from commercial wheat germ (Canberra Mill), using the method of Marcu and Dudock (1974). Cell free translations were carried out in a 50µl reaction mix containing 20mM Hepes-KOH pH 7.5, 80mM KCl, 2.5mM Mg-acetate, 2mM DH, 1mM ATP, 20µM GTP, 9mM creatine phosphate, 4µg/ml creatine phosphokinase, 10µCi of ^{3}H-leucine or ^{35}S-methionine, 25µM of the other unlabelled amino acids and 25µl of the wheat germ extract. Approximately 100-500ng of RNA was added to each translation mix and incubations were carried out at 30°C for 1 hour, followed by an incubation for 15 minutes in the presence of 25µM leucine or methionine, depending on the labelled amino acid used.

For determination of incorporated radioactivity, aliquots of the translation mix were spotted on to Whatman GF/A filters, boiled for 5 minutes in 10% TCA, rinsed in two changes of cold 10% TCA, then rinsed with ether. After drying, toluene scintillant was added to the filters and then the radioactivity determined in a Packard scintillation spectrometer.

The translation system was optimised for K^{+} and Mg^{2+} concentration by assaying the incorporation of ^{3}H leucine into TCA precipitable counts directed by 10ng of native cat reticulocyte RNA. At a K^{+} concentration of 100mM optimal translational efficiency was sharply defined at 2.5mM Mg^{2+}. Using a Mg^{2+} of 2.5mM a broad range of K^{+} in concentration was found to be acceptable, with maximum incorporation achieved at 80mM K^{+}.

Analysis of translation products was accomplished by loading an aliquot of the translation mix directly on to an appropriate gel electrophoresis system (see 10.2.6).

10.2.6 Analysis of Native Cat Globin Protein

Extraction of globin from red blood cells

(Thompson *et al.*, 1968)

Native cat blood obtained by cardiac puncture was centrifuged at 4,000g for 10 minutes at 4°C. The cells were washed three times in 0.15M NaCl and lysed by the addition of 2 volumes of water and 0.5 volumes of CCl_4. Cell debris was removed by centrifugation at 40,000g. After centrifugation the supernatant was brought to 0.1M NaH_2PO_4, pH 6.0.

Globin was prepared by the addition of 20 volumes of acid acetone (1.5ml of 10N HCl per 100ml of acetone). The precipitated globin was centrifuged at 100g for 5 minutes and washed four times in cold (-20°C) acetone, dried under vacuum, and dissolved in water at a concentration of 10mg/ml. After dialysis against distilled water, the globin was freeze-dried and stored at -20°C.

Chromatography of globin (Clegg *et al.*, 1965)

Globin samples were chromatographed on carboxymethyl cellulose columns in 8M urea buffers. Urea solutions were deionised by the addition of mixed bed resin, then sodium phosphates added to give the required Na^+ concentration of pH 6.7. Mercaptoethanol (final concentration 50mM) was added just before use since the absorbance of urea buffers containing mercaptoethanol increased with time. The acid-acetone precipitated globin was (100-200mg) was dissolved in 2ml of urea buffer, loaded without pH adjustment, and washed in with 1ml of starting buffer. For a column 1.0 x 12 cm a linear gradient was formed from 350ml each of starting buffer (0.02M Na^+ pH 6.7) and a limit buffer of 0.07M Na^+ concentration, pH 6.7.

The effluent was monitored at 280mμ and fractions collected. The fractions corresponding to peak values were pooled, and the globin chains carboxymethylated for 15 minutes using 0.94g iodoacetic acid per 100ml buffer in 5ml of 3M Tris pH 8.4.

The polypeptide chains were separated from urea and reaction products by dialysis against water, followed by 0.01N HCl. After freeze drying, the residual urea was removed by passage through a G10 Sephadex column (3.5 x 40 cm) equilibrated with 10% formic acid. The chains were recovered by freeze drying and stored at -20°C.

Cellulose acetate electrophoresis of globin

(Ueda and Schneider, 1969)

Cellulose acetate strips ("Cellogel 250") were equilibrated with running buffer (4.66M urea, 45mM Tris-HCl pH 9.0, 0.05% v/v β-mercaptoethanol) by slow shaking for 30 minutes. The origin in the middle of the strips was marked with a pencil and 3μl samples (5μg of globin chains) were applied in 0.8cm-long sample wells. Electrophoresis was carried out at a constant voltage of 225V for 5 hours at 4°C. The strips were stained with an Amido black solution (0.5% Amido black, 45% methanol, 10% acetic acid) for 10 minutes, destained in 47.5% methanol, 10% acetic acid for 30 minutes and then soaked for 1 minute in formalin and 3 minutes in 7% glycerol. Gel strips were then dried for 20 minutes on a glass plate at 80°C.

Analysis of translation products was performed in the same way as described, applying 3μl of the translation mix (with 5μg of globin as carrier) to the cellulose acetate. If it was needed to apply a greater proportion of the translation mix than 3μl, samples were precipitated by the addition of 1ml of acid-acetone in the presence of 20μg of globin protein as carrier. After standing at -20°C overnight, the

samples were centrifuged for 5 minutes at 10,000g at 4°C. The supernatant was removed and the pellet washed with cold acetone. The protein pellet was then dissolved in 10-15µl of electrophoresis buffer by solubilisation for 4h on ice. After electrophoresis the stain was fixed and the air dried gel was autoradiographed or the band corresponding to the carrier α- and β-globin chains cut from the gel and quantitated by liquid scintillation counting.

SDS-polyacrylamide electrophoresis of globin

(Weeds, 1976)

10% polyacrylamide, 0.25% bis-acrylamide slab gels were prepared in 10mM Tris, 10mM glycine, 0.1% SDS. Running buffer was 8mM Tris, 8mM glycine, 0.89% SDS, final pH 8.5. Globin protein was found not to dissolve in tris-glycine based loading buffer so a sample buffer consisting of 8M urea, 1% β-mercaptoethanol, 10mM Tris pH 8.0 was used. Samples were heated at 65°C for 10 minutes prior to loading. Samples were electrophoresed at 10-12 v/cm for 2½ hours at room temperature. Gels were stained in Coomassie blue R250 and destained by the method of Swank and Munkries (1971).

Gels used for the analysis of tritium labelled translation products were prepared and electrophoresed in the same manner. After fixation the labelled products were examined by fluorography. The gels were prepared for fluorography essentially as described by Bonner and Lasky (1974), except that 25% naphthalene, 1% PPO replaced PPO/POPOP as the fluorographic agent. Dried gels were placed into contact with X-ray film in the presence of an intensifying screen (Ilford fast tungstate) and exposed at -80°C.

10.2.7 *In vitro* Synthesis of Labelled DNA

Oligo-dT-primed reverse transcription

Oligo-dT-primed reverse transcription was carried out in a 20µl reaction mix containing up to 3g of RNA, 1mM each of dATP, dTTP, approximately 6M each of α-^{32}P-dCTP (400Ci/mmole) and α-^{32}P-dGTP (400Ci/mmole), 100mM Tris-HCl pH 8.3, 6mM $MgCl_2$, 30mM β-mercaptoethanol and 0.5µg digo-dT-dT_{12-14}. Reverse transcriptase was added and the solution incubated at 42°C for 60 minutes. The RNA template was removed boiling for 5 minutes. The mix was extracted with an equal volume of phenol:chloroform (1:1) and the aqueous phase loaded on to a 0.4 cm x 4 cm Sephadex G-50 column and eluted with 10mM Tris-HCl pH 7.4, 1mM EDTA, to remove unincorporated nucleotides.

Nick translation of double-stranded DNA

(Roop *et al.*, 1975)

Labelling double stranded DNA using *E. coli* DNA polymerase I was carried out in a 25µl reaction mix containing 5mM Tris-HCl pH 8.0, 5mM $MgCl_2$, 10mM β-mercaptoethanol, 50µg/ml bovine serum albumin, 4µM each of α-^{32}P-dATP and α-^{32}P-dCTP (2,000Ci/mmole), 1mM each of dGTP and dTTP, 10µM each of dCTP and dATP and up to 200ng of DNA. The DNA was nicked by the addition of 50pg of *E. coli* DNAase I and the reaction started by the addition of 2 units of *E. coli* polymerase I. The reaction mixture was incubated at 15°C for up to 3 hours, brought to a final concentration of 100mM NaCl in 10µl and then incubated at 37°C for 1 hour in the presence of 0.1mg/ml proteinase K. The solution was then made 0.1% SDS, phenol/chloroform extracted and the unincorporated nucleotides removed as described for oligo-dT-primed reverse transcription. If the labelled DNA was to be used as a hybridisation probe, the DNA strands were separated by boiling

the solution for 2 minutes and then snap-cooling in ice/ethanol.

Primed synthesis of M13 phage DNA probes

Approximately 400ng of the appropriate M13 DNA subclone was annealed in a 10µl reaction mixture (final 10mM Tris-HCl pH 7.4, 1mM $MgCl_2$, 50mM NaCl) to approxkmately 10pg of probe-primer (New England Biolabs.) by boiling for 30 seconds followed by gradual cooling. The solution was brought to a final concentration of 1mM each of dGTP and dTTP and 6µM each of α-^{32}P-dCTP (2,000Ci/mmole) and α-^{32}P-dATP (2,000Ci/mmole) in a total volume of 15µl. The synthesis reaction was initiated by the addition of 0.5µ of *E. coli* DNA polymerase I (Klenow fragment). The mixture was incubated at 14°C for 90 minutes. Labelled DNA was recovered by the addition of 50µg of yeast tRNA and two precipitations performed by the addition of $^1/_5$ volume of 10M ammonium acetate and 2 volumes of ethanol. M13 probes were not boiled before use.

10.2.8 Restriction Analysis of DNA

Digestion conditions

Restriction endonuclease digestion of DNA was carried out using the conditions for the appropriate enzyme detailed by the supplier. All reactions were stopped by the addition of EDTA to a final concentration of 5mM. The reaction mix was extracted with an equal volume of phenol of chloroform (1:1), and the aqueous phase ethanol precipitated.

Agarose gel electrophoresis

Electrophoresis of DNA for analytical purposes or for transfer to nitrocellulose was carried out on 14 cm x 16 cm x 0.3 cm slab gels containing 0.8-1% agarose. Electrophoresis buffer

consisted of 40mM Tris-acetate pH 8.2, 20mM Na-acetate, 1mM EDTA and electrophoresis was carried out at 10 v/cm for up to 3 hours. Genomic DNA was electrophoresed at 2 v/cm for 16 hours at 4°C. DNA was visualised by staining with 0.02% ethidium bromide solution (in electrophoresis buffer) for 15 minutes and examination under U.V. (254nm) light.

Polyacrylamide gel electrophoresis

Electrophoresis of DNA species less than about 1 kilobase in length was carried out on vertical 14 cm x 14 cm x 0.15 cm gels containing between 5% and 8% acrylamide. Electrophoresis buffer was 90mM Tris-borate pH 8.3, 2.5mM EDTA and electrophoresis was performed at 15 v/cm for 2 hours. DNA was visualised by ethidium bromide staining.

Transfer of DNA to nitrocellulose

Restricted DNA fractionated on 0.8% agarose slab gels was transferred to nitrocellulose filter paper using the method of Southern (1975), as modified by Wahl *et al.* (1979). In this procedure, the rapid transfer of DNA from the gel is facilitated by a partial hydrolysis of the DNA in the gel with 0.25N HCl.

The prehybridisation, hybridisation and washing conditions for nitrocellulose filters in Southern blot experiments were exactly as described by Wahl *et al.* (1979). The washed, dried nitrocellulose filter was placed in contact with X-ray film and exposed at -80°C in the presence of an intensifying screen.

10.2.9 Preparation and Tailing of Double-stranded cDNA

Synthesis of first strand

Oligo-dT-primed synthesis of the first strand on the

native cat RNA template was carried out as described in section 10.2.7, except that the concentration of the labelled nucleotides was increased to 500μM with cold nucleotides to ensure maximum cDNA synthesis. An aliquot of this reaction mixture was passaged through a Sephadex G-50 column to enable the determination of the efficiency of cDNA synthesis. The RNA template was removed by boiling the reaction mixture for 5 minutes.

Second strand synthesis

Second strand synthesis was accompoished by increasing the volume of the cDNA reaction mixture to 50μl, keeping the nucleotide and buffer concentration constant and the addition of 5 units of reverse transcriptase. After incubation at 37°C for 5 hours the reaction was terminated by bringing the solution to a final sodium sarkosyl concentration of 0.1% and holding on ice for 10 minutes. The volume of the reaction mix was increased to 100μl and the ds cDNA was recovered by passaging on a Sephadex G-50 column.

S_1 nuclease cleavage of ds cDNA

The second strand reaction mix was diluted into a 200μl final volume containing 30mM Na-acetate pH 4.6, 0.3M NaCl, 4.5mM $ZnCl_2$ and 150 units of S_1 nuclease. The S_1 digestion mix was incubated at 37°C for 30 minutes, neutralised by the addition of NaOH, and then extracted with an equal volume of phenol chloroform (1:1). The aqueous phase was passaged through Sephadex G-50 and the ds cDNA collected.

Gel electrophoresis of cDNA and products

The reaction products of cDNA and ds cDNA, synthesis and S_1 nuclease treatment of ds cDNA, were analysed by alkaline agarose gel electrophoresis. Approximately 30,000 cpm of each sample was

loaded per sample track on a 1% agarose gel containing 30mM NaOH, 2mM EDTA. The electrophoresis and sample buffers also contained 30mM NaOH, 2mM EDTA. After electrophoresis, gels were fixed by soaking in 10% TCA for 15 minutes, wrapped in plastic film, and exposed to X-ray film in the presence of an intensifying screen.

Size selection and tailing of ds cDNA

The ds cDNA eluate from the Sephadex G-50 column was loaded directly on to a 30 cm x 1 cm Sephacryl 1000 column equilibrated with 10mM Tris-HCl pH 7.4, 1mM EDTA. Fractions of 400µl were collected and appropriate samples were concentrated by freeze-drying and then precipitated in ethanol.

Poly-dC nucleotide tails were added to the ds cDNA using calf thymus terminal deoxynucleotidyl transferase. 400pmoles of ^{3}H-dCTP were dried down and resuspended in a 25µl volume containing 0.1M Na-Cacodylate pH 7.4, 1mM $CoCl_2$, 0.2mM DTT, and 0.1pmoles of ds cDNA. 1.0µl of terminal deoxynucleotidyl transferase was added and the reaction followed by the conversion of the ^{3}H-dCTP to a TCA insoluble form. When an average of 10-20 nucleotides per end had been added, the reaction was stopped by the addition of EDTA to a final concentration of 5mM.

10.2.10 Annealing and Transformation

Annealing to vector

0.1pmoles of pBR 322 DNA, cleaved with PstI and tailed with deoxyguanosine, as described above, was annealed to an equimolar amount of deoxycytosine tailed cDNA in 0.2M NaCl, 10mM Tris-HCl pH 8.2 by heating for 10 minutes at 65°C, incubating for 1 hour at 45°C and finally allowing the solution to cool slowly at 4°C. The annealed DNA

was stored at 4°C.

Transformation of *E. coli*

E. coli strain MC1061 was grown overnight at 37°C in Luria broth and then diluted $^{1}/_{50}$ into fresh Luria broth and grown to an A600 of 0.6. The cells were then chilled on ice for 2 hours, pelleted by centrifugation at 5,000g for 5 minutes and gently resuspended in $^{1}/_{20}$th of the original volume in ice-cold 0.1M $CaCl_2$. The cells were kept on ice for at least 1 hour and up to 24 hours. 0.2ml of competent cells was added to the transforming DNA, which had been diluted into 0.1ml of 0.1M Tris-HCl pH 7.4, and the mixture held on ice for 30 minutes, with occasional stirring. The cells were heat shocked at 42°C for 2 minutes, held on ice for a further 30 minutes, and then allowed to warm slowly to room temperature over a 20 minute period. 0.5ml of Luria broth was added to the transformed cells and incubated at 37°C for 30 minutes. The transformed cells were then mixed with 3ml of 0.7% L-agar and plated on to 1.5% L-agar plates containing 20µg/ml of tetracycline. These were incubated overnight at 37°C.

10.2.11 Detection and Examination of Recombinants

Screening for ampicillin sensitivity

Colonies from a transformation were transferred by toothpick to duplicate agar plates, one containing 20µg/ml tetracycline and the other containing 50µg/ml ampicillin. The colonies were grown overnight and tetracycline resistant, ampicillin sensitive transformants were selected for further study.

Colony screening by filter hybridisation

(Grunstein and Hogness, 1975)

Tetracycline-resistant, ampicillin sensitive colonies were transferred by toothpick to a master plate and to a sheet of nitrocellulose that had been boiled three times in distilled water and laid on to an L-agar plate containing 20µg/ml of tetracycline. The colonies were grown overnight on the nitrocellulose at 37°C, and the colonies lysed by transferring the nitrocellulose sequentially on to 3MM paper saturated with: 5µg/ml lysozyme in 50mM Tris-HCl pH 7.4, 10mM EDTA for 10 minutes; 0.5M NaOH, 0.2% Triton-X100 for 7 minutes, 1M Tris-HCl pH 7.4 for 2 minutes; 1M Tris-HCl pH 7.4 for 2 minutes and 1.5M NaCl, 0.5M Tris pH 7.0 for 4 minutes. The nitrocellulose filter was then washed in chloroform, left to air dry, then baked under vacuum for 2 hours. Hybridisation and washing conditions were as described for Southern blot experiments.

Miniscreen examination of plasmid recombinants

Isolation of plasmid DNA from small cultures was carried out as follows: 1ml cultures of each recombinant were grown overnight at 37°C in L-broth containing 20µl/ml tetracycline. Each culture was harvested by centrifugation for 1 minute in an Eppendorf microfuge. Each pellet was resuspended in 0.1ml of 50mM glucose, 25mM Tris-HCl pH 8.0, 10mM EDTA and left to stand at room temperature for 5 minutes. 0.2ml of 0.2N NaOH, 1% SDS was added to each sample and held on ice for 5 minutes, then 150µl of precooled 5N K-acetate pH 4.8 (3M K-acetate, 2N acetic acid) was added, mixed well and centrifuged for 1 minute. The supernatants were removed and nucleic acids precipitated by the addition of 2 volumes of room temperature ethanol. After 2 minutes plasmid DNA (and some RNA) was collected by centrifugation for 2 minutes, followed by a 70% ethanol wash, and

the pellets dissolved in 40µl Tris-HCl pH 7.4, 0.1mM EDTA.

10µl of each plasmid DNA (approx. 500ng of DNA) preparation was digested with the restriction endonuclease HpaII in a final volume of 25µl in the presence of 10µg/ml RNAase A. After digestion the DNA was end-labelled by the addition of 1µCi of α-^{32}P-dCTP and 0.2µ of *E. coli* DNA polymerase I (Klenow fragment) per DNA sample in the presence of 50mM NaCl, and a final reaction volume of 30µl. 10µl of each reaction was loaded on to a 10 cm x 40 cm x 0.05 cm 6% polyacrylamide gel and electrophoresed at 20 v/cm for 2½ hours. After electrophoresis the gel was fixed by washing with 12% acetic acid and was subsequently wrapped in plastic film and exposed to an X-ray plate at room temperature.

10.2.12 Large Scale Preparation of Recombination Plasmid DNA

(Birnboim and Doly, 1980)

Recombinant clones were grown overnight in L-broth supplemented with tetracycline (20µg/ml), diluted ^{1}/100 into 500 ml of fresh L-broth and grown to stationary phase. The cells were pelleted by centrifugation and resuspended in 4ml of ice-cold 25mM Tris-HCl pH 8.0, 10mM EDTA and 15% sucrose. After lysozyme was added to a final concentration of 3mg/ml, the solution was kept on ice for 40 minutes, 8ml of 200mM NaOH, 1% SDS was added and the mixture was left on ice for another 10 minutes. 5ml of 3.0M Na-acetate pH 5.5 was added, mixed gently and after a 40 minute incubation on ice, the cellular debris and chromosomal DNA was pelleted by centrifugation at 40,000g for 20 minutes at 4°C. The supernatant was decanted and was treated with 50µg of RNAase A for 1 hour at 37°C. The solution was then extracted once with an equal volume of phenol/chloroform (1:1),

once with an equal volume of chloroform and the nucleic acids were precipitated by the addition of 2 volumes of ethanol at -20°C for a minumum of 2 hours. The DNA was collected by centrifugation at 15,000g for 15 minutes. The pellet was resuspended in 1.6ml of H_2O, to which was added 400µl of 4M NaCl and 2ml of 13% PEG 6000. The mixture was left on ice for at least 1 hour and the DNA was pelleted by centrifugation at 15,000g for 15 minutes at 4°C, washed with cold 70% ethanol, briefly dried *in vacuo* and resuspended in 500µl of H_2O.

10.2.13 Hybrid Arrest Translation (Paterson and Roberts, 1977)

15µg of plasmid DNA was linearised with an appropriate restriction enzyme, phenol/chloroform extracted and precipitated as previously described. The plasmid DNA was resuspended in a total volume of 50µl and 10ng of RNA added. The DNA/RNA mixture was precipitated overnight at -20°C, pelleted by centrifugation of 15,000g for 10 minutes and the pellet resuspended in 2.5µl of 100mM PIPES pH 6.4, boiled for 30 seconds followed by rapid cooling in an ethanol dry ice bath. Formamide was added to a final concentration of 80% and NaCl to 0.4M in a final volume of 25µl. The hybridisation mixture was incubated at 48°C for 2 hours and the reaction was stopped by the addition of 200µl of ice cold H_2O and 25µg of yeast tRNA. The solution was precipitated overnight at -20°C and resuspended in a total volume of 20µl.

Hybridised samples were translated in the presence of 80µCi of ^{35}S-methionine (>400Ci/mmole) and analysed by electrophoresis as described in section 10.2.6.

10.2.14 Preparation of M13 Cloning Vectors

Preparation of M13 replicative form DNA

(Messing *et al.*, 1982)

To 3ml of 0.7% agar at 45°C was added 30µl of BCIG (20mg/ml in dimethylformamide), 20µl of IPTG (20mg/ml in H_2O), 0.2ml of exponential *E. coli* JM101 (lac, pro, supE, thi, F', trad D36, proAB, $lacI^q$, 2 delta m15), and 0.1ml of diluted M13 phage (ennough to give about 200 pfu). This mixture was poured on to a minimal (+ glucose) plate and incubated at 38°C for 9 hours.

A blue plaque was selected, toothpicked into 1ml of YT broth and grown with shaking for 6 hours. Meanwhile a 10ml culture of JM101 from a single colony on a minimal glucose plate was grown to an A600 of 0.5, and added to 1 litre of 2 x YT. When the A600 of this culture reached 0.5, the 1ml of phage solution was added and grown for 4 hours. Replicative form (RF) M13 DNA was prepared from pelleted cells by the alkali lysis method previously described in section 10.2.12.

Digestion and purification of RF DNA

The RF of M13 mp7, mp8 and mp9 was digested with the appropriate enzyme(s) to generate the desired termini. The linearised RF was fractionated on a 1% low gelling temperature agarose gel (see section 10.2.15) to remove any intact molecules, extracted, and resuspended at a final concentration of approximately 20ng/ml. The 5' terminal phosphate groups were removed by treatment with calf intestinal phosphate as described by Maxam and Gilbert (1980).

10.2.15 Subcloning of DNA Fragments into M13 Vectors

DNA fragment purification

DNA fragments of a length greater than 600 base pairs were electrophoresed on 1% low gelling temperature agarose gels as described in section 10.2.8. After visualisation by staining with ethidium bromide, bands containing DNA were excised from the gel in as small a volume as possible. The gel slice containing the DNA was subsequently melted by heating at 65°C for 5 minutes, 50µl of 10mM Tris-HCl, 1mM EDTA, pH 8.0 added, and the solution extracted with half volume of phenol which had previously been heated to 65°C. After recovery of the aqueous phase by centrifugation at 10,000g for 1 minute, the phenol extraction was repeated, and was followed by two extractions of the aqueous phase with diethyl ether. The DNA was then recovered by precipitation overnight at -20°C followed by centrifugation at 10,000g for 10 minutes at 4°C and the pellet washed twice with ice cold 70% ethanol and finally resuspended in 20µl of H_2O.

DNA fragments of a length less than 600 base pairs were electrophoresed on 6% polyacrylamide gels as described in section 10.2.8. After visualisation by ethidium bromide staining or, in some cases, by autoradiography, appropriate bands were excised and incubated at 37°C for 16 hours in 0.5ml of 0.5M NH_4-acetate, 10mM $MgCl_2$, 1mM EDTA, 0.1% SDS. After incubation the gel slice was removed and the DNA precipitated by the addition of 2 volumes of ethanol at -20°C for at least 12 hours. After precipitation the DNA was washed and resuspended as described for agarose gel extraction.

Ligation and transformation (Messing *et al.*, 1981)

The DNA fragment and appropriate M13 vector were combined in a ratio of approximately 3:1 in a 10µl reaction mix

containing 50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 1mM ATP and 10mM DTT. All ligations, regardless of the termini, were carried out using 0.5 units of T4 DNA ligase for at least 5 hours at 10°C.

E. coli strain JM101 was made competent essentially as described in section 10.2.14 except that cells were harvested after reaching an OD 600nm of 0.75 and finally resuspended in $^1/_{20}$th volume of ice cold 100mM $CaCl_2$. The cells were left on ice for at least one hour before use and the density of these cells was such that no additional bacteria were needed to act as a feeder lawn. The transformation was carried out as described in section 10.2.14.

10.2.16 Harvesting (+) Strand of Recombinant M13 Bacteriophage

White plaques were toothpicked individually into 1ml of a $^1/_{50}$ dilution of an overnight culture of JM101. The overnight culture was derived from a single JM101 colony and grown in minimal medium and was subsequently diluted into 2 x YT medium. Recombinant phage were multiplied at 37°C with vigorous shaking for 6 hours. Each 1ml culture was pelleted by centrifugation at 10,000g for 5 minutes and the supernatants removed, taking care to avoid the *E. coli* cell pellet. Phage were sedimented by the addition of a quarter volume of 2.5M NaCl, 13% PEG 6000, allowed to incubate at room temperature for 15 minutes, and collected by centrifugation for 5 minutes at 10,000g. The phage pellets were then resuspended in 100µl of 10mM Tris-HCl pH 8.0, 1mM EDTA and phenol extracted. This was followed by two extractions with 5 volumes of diethyl ether. The DNA was then precipitated by the addition of $^1/_{10}$ volume of 3M Na-acetate pH 5.5 and 2.5 volumes of ethanol at -20°C for 16 hours, collected centrifugation and resuspended in 30µl of H_2O and stored at 4°C.

10.2.17 Dideoxy Sequencing Reactions (Messing *et al.*, 1981)

Hybridisation

4µl of the single-stranded M13 template, 1µl of primer (1µ/µl of Biolabs. 17-mer), 1µl of 10X Hin buffer (60mM Tris-HCl pH 7.5, 60mM $MgCl_2$, 500mM NaCl) and 4µl of H_2O were combined in an Eppendorf tube which was placed in a small boiling water-bath and left to cool to room temperature (about 45 minutes). The tubes were centrifuged for 5 seconds to collect the condensate and left at 4°C until required.

Polymerisation

The chain termination reaction (Sanger *et al.*, 1977) was carried out as described by Messing *et al.* (1981) except that [α-^{32}P]-dCTP was the radiolabelled nucleotide and the incubations were carried out at 37°C. The final dideoxynucleotide concentrations used in this study were 0.3mM ddGTP, 0.6mM ddATP, 1.0mM ddTTP and 0.08mM ddCTP. It was convenient to combine the G° and ddGTP solutions, etc. of Messing *et al.* (1981) and to store these mastermixes at -80°C until required.

Sequencing gel electrophoresis

The samples were electrophoresed through a 6% polyacrylamide gel containing 50% urea. TBE pH 8.8 was used in these gels as the buffer provided greater resolution for long electrophoresis runs than the TBE pH 8.3 buffer used elsewhere in this study. In cases of prolonged electrophoresis of the samples through the gel, the buffer was changed every time the BPB tracker dye migrated 40cm. After electrophoresis, the gel was fixed for at least 10 minutes with 12% acetic acid, dried in an oven at 110°C for at least 30 minutes,

covered with plastic wrap and exposed to X-ray film for up to 12 hours at room temperature.

10.2.18 Cell Culture and Chromosome Preparation

Routine cell culture methods

Cells were grown as monolagers in plastic flasks or glass prescription bottles, maintained at 36°C in a human incubator with a 5% CO_2 atmosphere. Cells were subcultured by washing with phosphate-buffered saline (300mM NaCl, 50mM KCl, 16mM Na_2HPO_4, 3mM KH_2PO_4), dissociating with trypsin solution (300mM NaCl, 50mM KCl, 16mM Na_2HPO_4, 3mM KH_2PO_4, 0.5mM EDTA, 125g/l trypsin), and transferring an innoculum to a fresh vessel. The standard culture medium used was RPM1/640 supplemented with 10% foetal calf serum, 50µg/ml streptomycin, 60µg/ml pennicillin and 0.002mM glutamine.

Cells in foetal calf serum containing 5% dimethylsulphoxide were stored in heat-sealed ampoules in liquid nitrogen. They were frozen slowly using a liquid nitrogen programmed cooler.

Establishment of a cell line from the native cat

The ear of 1 year-old male native cat was shaved and sterilised by washing with 70% ethanol. After sterilisation a piece of tissue approximately 3 mm x 3 mm was removed and washed in three changes of standard cell culture medium. The tissue was then placed in a well of a 6-well cell culture plate and finely macerated using a scalpel. The macerated tissue was then divided equally amongst the six wells and a sterile 22 mm x 22 mm glass coverslip was placed on top of each piece of tissue. Culture medium was then added so that the coverslip was covered to a depth of approximately 5 mm. After 4-7 days dividing cells could be observed on the edges of the tissue,

and after two weeks the cells in most wells had grown to confluence. The cell line (dv♂) was then subcultured and stored as described above.

Chromosome preparation for *in situ* hybridisation

Chromosome preparations were obtained from logarithmically growing cultures, which were usually subcultured 12-16 hours before harvesting to achieve partial synchronisation and increase the yield of mitoses. They were treated for 1 hour with 0.2µg/ml colcemid before harvesting. After harvesting, cells were washed once with PBS, and suspended in hypotonic KCl (0.075M) at 37°C for 15 minutes. The cell suspensions were then centrifuged at 400g for 5 minutes and were then fixed in three changes of fixative (3 methanol : 1 glacial acetic acid), dropped on to clean microscope slides and air dried.

Slides were stored dessicated at -80°C for storage periods of greater than 1 week.

Chromosome suspensions for flow cytometry

Chromosome suspensions from the dv♂ cell line suitable for flow cytometry were prepared by a modification of the method of Sillar and Young (1981). Cell cultures were grown to approximately 50% confluence, then synchronised by the addition of thymidine to a final concentration of 5mM for 16 hours. After incubation in thymidine, the cell cultures were washed and fresh medium without thymidine added. After 8 hours colcemid was added to a final concentration of 0.2µg/ml for 12-16 hours. Mitotic populations were harvested by vigorous shaking of the culture flasks. Such preparations routinely had a mitotic index greater than 90%.

About 4×10^7 mitotic cells (typically from four 120cm^2 culture vessels) were resuspended in 0.075M KCl for 15 minutes at 37°C, pelleted by centrifugation, and resuspended in chromosome isolation

buffer (15mM Tris-HCl, 0.2mM spermine, 0.5mM spermidine, 2mM EDTA, 0.5mM EGTA, 80mM KCl, 20mM NaCl and 14mM β-mercaptoethanol, pH 7.2). The cells were washed twice in isolation buffer and the final pellet was resuspended in approximately 20 times its volume of ice cold isolation buffer plus 0.1% digitonin. This suspension (typically 2ml) was vortexed for 1 minute. This was usually sufficient to release the majority of the chromosomes from the metaphase cells. Occasionally chromosome suspensions were passaged several times through a 21g syringe needle to ensure the release of most of the chromosomes.

10.2.19 *In situ* Hybridisation (Trent *et al.*, 1982)

Slides were viewed by contrast phase microscopy and areas containing the metaphase cells located and marked by the use of a diamond pencil on the back of the slide. Each slide was treated with 100µg/ml RNAase A at 37°C for 1 hour and rinsed well in four changes of 2 x SSC (0.3M NaCl, 0.03M Na-citrate, pH 7.0) at room temperature. Slides were then dehydrated by passaging through an ethanol series. Chromosomal DNA was denatured by treatment with 70% formamide/2 x SSC, pH 7.0, at 70°C for 2 minutes. After denaturation slides were quickly rinsed in 70% ethanol for 1 minute, followed by dehydration through an ethanol series.

Probe DNA was labelled with tritiated nucleotides as described in section 10.2.7. Probe DNA was resuspended at final concentrations ranging from 0.05 to 0.4µg/ml in a 50% formamide solution containing 10% dextran sulphate and attaining final concentration of 2 x SSCP (0.24M NACl, 0.03M Na-citrate, 0.04M $NaPO_4$, pH 6.0). Probe also contained a final concentration of 250µg/ml carrier DNA. Before use the probe was denatured by heating to 70°C for 5 minutes followed by quick cooling in dry ice/ethanol.

Probe was placed on preselected areas of the slides (15µl for a 22 mm x 22 mm coverslip) and covered with glass coverslips. Slides were incubated in moist chambers at 37°C for 16 hours.

After hybridisation, unbound probe was removed by washing in four changes of 50% formamide/2 x SSC, pH 7.0, at 40°C, followed by six rinses in 2 x SSC at 40°C. Slides were then dehydrated through an ethanol series.

Slides were then dipped in a 50% solution of Ilford K2 emulsion, allowed to dry overnight, then exposed for 10-30 days at -80°C. Autoradiographs were visualised by developing in 50% Kodak D19 developer for 5 minutes followed by fixation for 2 minutes in Ilford "rapid fix". Slides were stained for 30 seconds in a 1% solution (in H_2O) of toluidine blue, and coverslips mounted.

10.2.20 Flow Cytometry and Flow Sorting

Chromosome samples were stained at a final concentration of 100µg/ml of ethidium bromide and analysed with a Becton Dickinson FACS IV Cell Sorter. A Spectra-Physics 164-05 laser producing an output of 0.8W at 488nm was used to excite the stained chromosomes. Scattered laser light was blocked with a Schott SP550 filter. Fluorescence signals detected by the photomultiplier tube were collected into one of 256 channels.

Sheath fluid (10mM Tris-HCl pH 7.6, 1mM EDTA, 50mM NaCl), flow rate and chromosome suspension concentration were adjusted to give a fluorescent particle flow rate of approximately 900-1200 particles per second.

For chromosomal analysis, photomultiplier tube voltages were advanced such that the fluorescence of the largest chromosome was approximately in channels 190-220, with fluorescence due to unbroken

nuclei and large aggregates of chromosomes etc. falling into the last (256) channel, and hence not interferring with the chromosomal distribution. Laser power and photomultiplier tube voltages were adjusted such that a linear photomultiplier response was obtained (the relationship between laser power and amplification being dependent on the measured fluorescence of the particles).

For analyses involving chromosomes a 70μm nozzle diameter was used and sheath fluid was passaged through a 0.45μm filter before entering the machine.

The raw data, in the form of channel values, was manually transferred to a Control Data Corporation Cyber 6400 computer. Data analysis was performed as described in Chapter 8 and Appendix I.

Peak composition analysis was performed by sorting approximately 10,000 chromosomes on to dry microscope slides. The chromosomes were then fixed by the drop-wise addition of 3:1 methanol:glacial acetic acid fixative. After fixation, chromosomes were stained with a 10% solution of Giemsa in phosphate buffer pH 6.8, for 2 minutes.

Chromosomes for DNA isolation were sorted into 1.5 ml Eppendorf tubes and frozen in liquid nitrogen. Chromosomes from two regions of interest were sorted simultaneously at rates of approximately 600-1000 particles per second. For the analysis of chromosomal DNA, approximately 10^6 chromosomes of each type could be sorted in 3-5 hours. Sort purity was monitored by withdrawing aliquots from each sort fraction and visualising the chromosomes as described above.

10.2.21 Isolation of DNA from Sorted Chromosomes

(Lebo *et al.*, 1979)

Each fraction was thawed and the chromosomes pelleted by centrifugation at 50,000g for 3 hours at 4°C. Each pellet was

resuspended in 1.0ml of digestion solution (200µg/ml proteinase K, 20µg/ml sonicated, heat denatured salmon sperm DNA, 50mM Tris-HCl, pH 7.4, 100mM NaCl, 1mM EDTA and 0.5% SDS). After incubation at 45°C for 5 hours the mixture was phenol extracted and precipitated by the addition of $^{1}/_{10}$ volume of 3M Na-acetate pH 5.5 and 2.5 volumes of ethanol. After precipitation at -20°C for 72 hours, the DNA was collected by centrifugation at 10,000g for 15 minutes at 4°C. The DNA pellet was washed twice with 70% ethanol and then dried under vacuum for 20 minutes. The DNA was then resuspended in a final volume of 30µl and digested with 50 units of EcoR1 for 3 hours.

10.2.22 Southern Blot Analysis of Sorted DNA

Restriction digest were loaded directly on to 0.8% horizontal agarose gels and subsequently transferred to nitrocellulose, exactly as described in section 10.2.8. The washed dried filters were probed with ^{32}P labelled native cat α- and β-globin sequences and were washed to a stringency of 1 x SSC.

10.2.23 Preparation of High Molecular Weight Tissue Culture DNA

Dh ♂ cells were grown in monolayer culture, removed by treatment with trypsin-versene and washed twice in PBS. Cells were then homogenised in 7ml of 50mM Tris pH 8.0, 10mM EDTA using a dounce homogeniser and 8g of $CsCl_2$ added. After the $CsCl_2$ added. After the $CsCl_2$ had dissolved, 1ml of 20% Sarkosyl was added and 400µl of 10mg/ml EtBr. Samples were centrifuged at 41,000 rpm for 40 hours. The high molecular weight DNA band was removed by side puncture. Ethidium bromide was removed by three ethanol precipitations at room temperature. The DNA was finally resuspended in 0.1mM EDTA pH 7.0.

10.2.24 M13 Clone Orientation Test

100µl each of a 1ml phage solution (see 10.2.16) were mixed and precipitated by the addition of $^{1}/_{10}$ volume of Na-acetate pH 5.5 and 2 volumes of ethanol, and held at -80°C for 15 minutes. The DNA was collected by centrifugation at 10,000g for 10 minutes. The pellet was washed in 70% ethanol and resuspended in 30µl of 0.1% SDS, 0.1% BPB, 0.2M NaCl, 10mM Tris-HCl pH 8.0. The mixture was then heated to 65°C for 10 minutes followed by incubation at 45°C for 1 hour. The samples were loaded directly on to a 0.8% agarose gel and the DNA visualised by staining with ethidium bromide.

APPENDIX I

```
PROGRAM BRANDON     73/173  OPT=1                                FTN 4.1+526

1                PROGRAM BRANDON(TAPE1,OUTPUT,TAPE3,TAPE2=OUTPUT)
                 DIMENSION X(100),A(5),SIGMAA(5),YFIT(100),XN(200),YN(200),
                1YFITN(200),Y(100),R(200),DIFF(200)
                 READ(1,1)NDISTS
5                FORMAT(I2)
               C          FOR EACH MAIN PEAKN READ IN DATA AND FIT CURVE
                 DO 90 K=1,NDISTS
                 READ(1,2) NPTS
                 FORMAT(I4)
0                DO 100 I=1,NPTS
                 READ(1,3) X(I),Y(I)
                 FORMAT(F3.0,1X,F4.0)
                 CONTINUE
                 NTERMS =5
5                READ(1,4) (A(J),J=1,NTERMS)
                 FORMAT(F8.2,3(1X,F7.2),1X,F6.4)
                 CHISQ1=0.0
                 L=1
                 FLAMBDA=0.001
0                CALL CURFIT(X,Y,NPTS,NTERMS,A,SIGMAA,FLAMBDA,YFIT,CHISQR)
                 WRITE(2,5) L,CHISQR
                 FORMAT(/////,10X,*ON ITERATION*,I2,5X,*CHISQR WAS*,F10.4)
                 N=0
                 WRITE(2,6)
5                FORMAT(/,7X,*X*,5X,*Y*,10X,*YFIT*)
                 DO 102 I=1,NPTS
                 N=N+1
                 WRITE(2,7) X(I),Y(I),YFIT(I)
                 FORMAT(5X,F4.0,2X,F4.0,6X,F6.1)
0                WRITE(3,7) X(I),Y(I),YFIT(I)
                 XN(N)=X(I)
                 YN(N)=Y(I)
                 YFITN(N)=YFIT(I)
                 CONTINUE
5                WRITE(2,8)
                 FORMAT(/,5X,*PARAMETERS*)
                 WRITE(2,9) (A(J),J=1,NTERMS)
                 FORMAT(5(2X,F10.4))
                 WRITE(2,9) (SIGMAA(J),J=1,NTERMS)
0                IF (ABS(CHISQR-CHISQ1).LE.0.01)90,11
                 CHISQ1=CHISQR
                 L=L+1
                 IF(L .GT.10)90,10
                 CONTINUE
5                RSM=0
               C          EVALUATE MSR
                 WRITE(2,20)
                 FORMAT(///,10X,*RESIDUALS*,/)
                 DO 105 K=1,N
0                DIFF(K) =(YN(K)-YFITN(K))
                 RSM =RSM + DIFF(K)*DIFF(K)
                 CONTINUE
                 RSM = RSM/(N -3*NTERMS - 2)
                 WRITE(2,19) RSM
5                FORMAT(///, * RESIDUAL  MEAN SQUARE *,F8.4) '
                 DO 106 K=1,N
                 R(K) = DIFF(K)/SQRT(RSM)
                 WRITE(2,18) XN(K),YN(K),R(K)
                 FORMAT(5X,F4.0,2X,F6.1,2X,F5.4)
0                CONTINUE
                 STOP
                 END
```

```
      SUBROUTINE CURFIT(X,Y,NPTS,NTERMS,A,SIGMAA,FLAMBDA,YFIT,CHI
      DOUBLE PRECISION ARRAY
      DIMENSION X(260),Y(260),A(5),SIGMAA(5),YFIT(260)
      DIMENSION WEIGHT(260),ALPHA(5,5),BETA(5),DERIV(5),ARRAY(5,5)
   11 NFREE=NPTS-NTERMS
      IF (NFREE) 13,13,20
   13 CHISQR=0
      GO TO 110
C           EVALUATE WEIGHTS
   20 DO 30 I=1,NPTS
      WEIGHT(I) = 1./Y(I)
   30 CONTINUE
C           EVALUATE ALPHA AND BETA MATRICES
   31 DO 34 J=1,NTERMS
      BETA(J) = 0
      DO 34 K=1,J
   34  ALPHA(J,K) = 0
   41  DO 50 I =1,NPTS
      CALL FDERIV(X,I,A,DERIV)
      DO 46 J=1,NTERMS
      BETA(J) = BETA(J) + WEIGHT(I) * (Y(I) - FUNCTN(X,I,A)) * DER
      DO 46 K=1,J
   46 ALPHA(J,K) = ALPHA(J,K) + WEIGHT(I)*DERIV(J)*DERIV(K)
   50 CONTINUE
   51 DO 53 J=1,NTERMS
      DO 53 K=1,J
   53 ALPHA(K,J) = ALPHA(J,K)
C           EVALUATE CHI SQUARE AT START
   61 DO 62 I=1,NPTS
   62 YFIT(I) = FUNCTN(X,I,A)
   63 CHISQ1 = FCHISQ(Y,NPTS,NFREE,YFIT)
C           INVERT MODIFIED CURVATURE MATRIX TO FIND NEW PARAMETRS
   71 DO 74 J=1,NTERMS
      DO 73 K=1,NTERMS
   73 ARRAY(J,K) = ALPHA(J,K)/SQRT(ALPHA(J,J)*ALPHA(K,K))
   74 ARRAY(J,J) = 1. + FLAMBDA
   80 CALL MATINV(ARRAY,NTERMS,DET)
   81 DO 84 J=1,NTERMS
      B(J) = A(J)
      DO 84 K=1,NTERMS
   84 B(J) = B(J) + BETA(K)*ARRAY(J,K)/SQRT(ALPHA(J,J)*ALPHA(K,K))
C           IF  CHISQ IS INCREASED ,INCREASE LAMBDA
C           AND TRY AGAIN
   91 DO 92 I=1,NPTS
   92 YFIT(I) = FUNCTN(X,I,B)
   93 CHISQR = FCHISQ(Y,NPTS,NFREE,YFIT)
      IF (CHISQ1 - CHISQR) 95,101,101
   95 FLAMBDA = 10.* FLAMBDA
      GO TO 71
C           EVALUATE PARAMETERS AND UNCERTAINTIES
  101 DO 103 J=1,NTERMS
      A(J) = B(J)
  103 SIGMAA(J) = DSQRT(ARRAY(J,J)/ALPHA(J,J))
      FLAMBDA = FLAMBDA/10.
  110 RETURN
      END
```

```
      SUBROUTINE FDERIV(X,I,A, DERIV)
      DIMENSION X(200),A(5),DERIV(5)
11    X1 = X(I)
      Z = (X1 - A(2))/A(3)
      ZZ = Z**2
      IF (ZZ - 50.) 21,15,15
15    DO 16 J=1,3
16    DERIV(J) = 0.
      GO TO 24
21    DERIV(1) = EXP(-ZZ/2.)
      DERIV(2) = A(1)*DERIV(1)*Z/A(3)
      DERIV(3) = DERIV(2)*Z
24    DERIV(4) = EXP(-A(5)*X1)
      DERIV(5) =-A(4)*X1*DERIV(4)
      RETURN
      END
```

```
      FUNCTION FUNCTN(X,I,A)
      DIMENSION X(200),A(5)
   11 X1 = X(I)
   12 FUNCTN = A(4)*EXP(-A(5)*X1)
   13 Z = (X1 - A(2))/A(3)
      ZZ = Z**2
      IF (ZZ - 50.) 15,20,20
   16 FUNCTN = FUNCTN + A(1)*EXP(-ZZ/2.)
   20 RETURN
      END
```

```
      FUNCTION FCHISQ(Y,NPTS,NFREE,YFIT)
      DOUBLE PRECISION CHISQ,WEIGHT
      DIMENSION Y(260),YFIT(260)
   11 CHISQ = 0
   12 IF (NFREE) 13,13,20
   13 FCHISQ = 0
      GO TO 40
C           ACCUMULATE CHI SQUARE
   20 DO 30 I=1,NPTS
      WEIGHT = 1./Y(I)
   30 CHISQ = CHISQ + WEIGHT*(Y(I) - YFIT(I))**2.
C           DIVIDE BY NO.DF
   31 FREE = NFREE
   32 FCHISQ = CHISQ/FREE
   40 RETURN
      END
```

```
      SUBROUTINE MATINV (ARRAY,NORDER,DET)
      DOUBLE PRECISION ARRAY, AMAX ,SAVE
      DIMENSION ARRAY(5,5), IK(5), JK(5)
   10 DET = 1
   11 DO 100 K=1, NORDER
      AMAX = 0
   21 DO 30 I=K, NORDER
      DO 30 J=K, NORDER
   23 IF (DABS(AMAX) - DABS(ARRAY(I,J))) 24, 24, 30
   24 AMAX = ARRAY(I,J)
      IK(K) = I
      JK(K) = J
   30 CONTINUE
   31 IF (AMAX) 41, 32, 41
   32 DET = 0
      GO TO 140
   41 I = IK(K)
      IF (I-K) 21, 51, 43
   43 DO 50 J=1, NORDER
      SAVE = ARRAY(K,J)
      ARRAY(K,J) = ARRAY(I,J)
   50 ARRAY(I,J) = -SAVE
   51 J = JK(K)
      IF (J-K) 21, 61, 53
   53 DO 60 I=1, NORDER
      SAVE = ARRAY(I,K)
      ARRAY(I,K) = ARRAY(I,J)
   60 ARRAY(I,J) = -SAVE
   61 DO 70 I=1, NORDER
      IF (I-K) 63, 70, 63
   63 ARRAY(I,K) = -ARRAY(I,K) / AMAX
   70 CONTINUE
   71 DO 80 I=1, NORDER
      DO 80 J=1, NORDER
      IF (I-K) 74, 80, 74
   74 IF (J-K) 75, 80, 75
   75 ARRAY(I,J) = ARRAY(I,J) + ARRAY(I,K)*ARRAY(K,J)
   80 CONTINUE
   81 DO 90 J=1, NORDER
      IF (J-K) 83, 90, 83
   83 ARRAY(K,J) = ARRAY(K,J) / AMAX
   90 CONTINUE
      ARRAY(K,K) = 1. / AMAX
  100 DET = DET * AMAX
  101 DO 130 L=1, NORDER
      K = NORDER - L + 1
      J = IK(K)
      IF (J-K) 111, 111, 105
  105 DO 110 I=1, NORDER
      SAVE = ARRAY(I,K)
      ARRAY(I,K) = -ARRAY(I,J)
  110 ARRAY(I,J) = SAVE
  111 I = JK(K)
      IF (I-K) 130, 130, 113
  113 DO 120 J=1, NORDER
      SAVE = ARRAY(K,J)
      ARRAY(K,J) = -ARRAY(I,J)
  120 ARRAY(I,J) = SAVE
  130 CONTINUE
  140 RETURN
      END
```

BIBLIOGRAPHY

ABELSON, J. (1979) RNA processing and the intervening sequence problem. *Ann. Rev. Biochem.*, 48: 1035-69.

ADAMS, J.M. and CORY, S. (1975) Modified nucleosides and bizarre 5'-terminal in mouse myeloma mRNA. *Nature*, 255: 28-33.

AIR, G.M. and THOMPSON, E.O.P. (1979) Studies on marsupial proteins. II. Amino acid sequence of the β-chain of haemoglobin from the grey kangaroo, *Macropus giganteus*. *Aust. J. Biol. Sci.*, 22: 1437-54.

AIR, G.M., THOMPSON, E.O.P., RICHARDSON, B.J. and SHARMAN, G.B. (1971) Amino acid sequences of kangaroo myoglobin and haemoglobin and the date of marsupial-eutherian divergence. *Nature*, 229: 391-4.

ALWINE, J.C., KEMP, D.J., PARKER, B.A., REISER, J., STARK, G.R. and WAHL, G.M. (1979) In 'Methods in Enzymology', 68: 220-42, ed. K. Moldave and L. Grossman. Academic Press: New York.

ALWINE, J.C., KEMP, D.J. and STARK, G.R. (1977) Method for the detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridisation with DNA probes. *Proc. Natl. Acad. Sci. USA*, 74: 5350-5.

ANDERSON, S., GAIT, M.J., MAYOL, L. and YOUNG, I.G. (1980) A short primer for sequencing DNA cloned into the single stranded phage vector, M13mp2. *Nuc. Acids res.*, 8 :1731

AVIV,H. and LEDER, P.(1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci.USA.*,69: 1408-12.

BARALLE, F.E. (1983) The functional significance of leader and trailer sequences in eukaryotic mRNAs. *Int. Rev. Cytol.*, 81: 71-106.

BARALLE, F.E. and BROWNLEE, G.G. (1978) AUG is the only recognisable signal sequence in the 5' non-coding regions of eukaryotic mRNA. *Nature*, 274: 84-7.

BARRIE, P.A., JEFFREYS, A.J. and SCOTT, A.F. (1981) Evolution of the β-globin gene cluster in man and the primates. *J. Mol. Biol.*, 149: 319-36.

BATTAGLIA, P. and MELLI, M. (1977) Isolation of globin messenger RNA of *Xenopus laevis*. *Develop. Biol.*, 58: 337-50.

BAUMAN, J.G.J., WIEGANT, J., VAN DUIJN, P., LUBSEN, H.M., SANDERMEIJER, P.J.A., MENNIG, W. and KUBLI, E. (1981) Rapid and high resolution detection of *in situ* hybridisation to polytene chromosomes using fluorochrome-labeled RNA. *Chromosoma*, 84: 1-18.

BEARD, J.M. and THOMPSON, E.O.P. (1970) Studies on marsupial proteins. III. N-bromosuccinimide cleavage of the α-chain of kangaroo haemoglobin and the amino acid sequence of the N-terminal fragment. *Aust. J. Biol. Sci.*, 23: 185-92.

BENGTSSON, B.O. (1980) Rates of karyotype evolution in placental mammals. *Hereditas*, 92: 37-7.

BEVINGTON, P.R. (1969) Data reduction and error analysis for the physical sciences. pp.204-46. McGraw-Hill: New York.

BIRNBOIM, H.C. and DOLY, J. (1980) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.*, 7: 1513-23.

BISHOP, J.O. and JONES, K.W. (1972) Chromosomal localisation of human haemoglobin structural genes. *Nature*, 240: 149-50.

BLAKE, C.C.F. (1979) Exons encode protein functional units. *Nature*, 277: 598.

BODMER, W.F. (1975) Analysis of linkage by somatic cell hybridisation and its conservation by evolution. In 'Chromosome Variations in Human Evolution', ed. A.J. Boyce, pp.53-61.

BODMER, W.F. and PARSONS, P.A. (1962) Linkage amd recombination in evolution. *Adv. in Genetics*, 11: 1-100.

BOLIVAR, F., RODRIGUEZ, R.L., GREENE, P.J., BETTACH, M.C., HEYMEKER, H.L. and BOYER, H.W. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2: 95-113.

BONNER, W.M. AND LASKEY, R.A. (1974) A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.*, 46, 83-8.

BREATHNACH, R., BENOIST, C., O'HARE, K., GANNON, F. and CHAMBON, P. (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA*, 75: 4853-7.

BUNN, H.F., FORGET, B.G. and RANNEY, H.M. (1977) 'Human Haemoglobins.' Philadelphia: Saunders.

BUONGIORNO-NADELLI, M. and AMALDI, F. (1970) Autoradiographic detection of molecular hybrids between rRNA and DNA in tissue sections. *Nature*, 225: 946-8.

BURKI, H.J., REGIMBAL, T.J. and MEL, H.C. (1973) Zonal fractionation of mammalian metaphase chromosomes and determination of their DNA content. *Preparative Biochem.*, 3: 157-82.

BUSH, G.L., CASE, S.M., WILSON, A.C. and PATTON, J.L. (1977) Rapid speciation and chromosomal evolution in mammals. *Proc. Natl. Acad. Sci.*, 74: 3942-6.

CARRANO, A.V., GRAY, J.W., LANGLOIS, R.G., BURKHART-SCHULTZ, K.J. and VAN DILLA, M.A. (1979) Measurement and purification of human chromosomes by flow cytometry and sorting. *Proc. Natl. Acad. Sci. USA*, 76: 1382-4.

CARRANO, A.V., GRAY, J.W., MOORE II, D.H., MINKLER, J.L., MAYALL, B.H. VAN DILLA, M.A. and MENDELSOHN, M. (1976) Purification of the chromosomes of the Indian muntjac by flow sorting. *J. Histochem. Cytochem.* 24: 345-56.

CARRANO, A.V., GRAY, J.W. and VAN DILLA, M.A. (1978) Flow cytogenetics: Progress towards chromosomal aberration detection. In 'Mutagen-induced Chromosome Change in Man', ed. H.J. Evans and D.L. Lloyd. Edinburgh University Press.

CATTANACH, B.M. (1975) Control of chromosome inactivation. *Ann. Rev. Genet.*, 9: 1-18.

CHANDLER, M.E. and YUNIS, J.J. (1978) A high resolution *in situ* hybridization technique for the direct visualization of labeled G-banded early metaphase and prophase chromosomes. *Cytogenet. Cell Genet.*, 22: 352-6.

CHU, E.H.Y. and POWELL, S.S. (1976) Selective systems in somatic cell genetics. In *Advances in Human Genetics*, 7 (H. Harris and K. Hirschhorn). Plenum Publishing: New York.

CLARKE, B. (1970) Selective constraints on amino-acid substitutions during the evolution of proteins. *Nature*, 228: 159.

CLEARY, M.L., HAYNES, J.R., SCHON, E.A. and LINGREL, J.B. (1980) Identification by nucleotide sequence analysis of a goat pseudo-globin gene. *Nucleic Acid Research*, 8: 4791-802.

CLEGG, J.B., NAUGHLOR, M.A. and WEATHERALL, D.J. (1965) An improved method for the characterisation of human haemoglobin mutants: identification of $\alpha_2\beta_2^{95Glu}$ haemoglobin N (Baltimore). *Nature*, 207: 945-7.

CLEMENS, W.A. (1977) Phylogeny of the marsupials. In 'The Biology of the Marsupials', ed. B. Stonehouse and D. Gilmore, p.51. Macmillan Press: London.

COLLARD, J.G., TULP, A., STEIGMAN, J., BOEZEMAN, J., BAUER, F.W., JONGKIND, J.F. and A. VERKERK (1980) Separation of large quantities of Chinese hamster chromosomes by velocity sedimentation at unit gravity followed by flow sorting (FACS II). *Exptl. Cell Res.*, 130: 217-27.

COLLARD, J.G., SCHIIVER, J., TULP, A. and MEULENBROCK, M. (1983) Localisation of genes on fractionated rat chromosomes by molecular hybridisation. *Exptl. Cell Res.*, 137: 463-9.

COOPER, D.W. (1974) Some aspects of the genetics of Australian marsupials with an appendix on the genetics of monotremes. *J. Aust. Mammal. Soc.*, 1: 155-74.

COOPER, D.W., EDWARDS, C., JAMES, E., SHARMAN, G.B., VANDEBERG, J.L. and GRAVES, J.A.M. (1977a) Studies on metatherian sex chromosomes. VI. A third state of an X-linked gene: partial activity for the paternally derived Pgk-A allele in cultured fibroblasts of *Macropus giganteus* and *M. parryi*. *Aust. J. Biol. Sci.*, 30: 431-43.

COOPER, D.W., JOHNSTON, P.G., SHARMAN, G.B. and VANDEBERG, J.L. (1977b) The control of gene activity on eutherian and metatherian X chromosomes: A comparison. In 'Reproduction and Evolution', ed. J.H. Calaby and C.H. Tyndale-Biscoe, pp.81-87. Australian Academy of Science: Canberra.

COOPER, D.W., VANDEBERG, J.L., SHARMAN, G.B. and POOLE, W.E. (1971) Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X-inactivation. *Nature New Biol.*, 230: 155-7.

CORBIN, K.W. and UZZEL, T. (1970) Natural selection and mutation rates in mammals. *Amer. Nat.*, 104: 37.

COTE, B.P., UHLENBECK, O.C. and STEFFENSEN, D.M. (1980) Quantitation of *in situ* hybridization of ribosomal ribonucleic acids to human diploid cells. *Chromosoma*, 80: 349-67.

CRAWFORD, R.J., SCOTT, A.C. and WELLS, J.R.E. (1977) Organisation of sequences in avian globin mRNA. *Eurp. J. Biochem.*, 72: 291-9.

CREAGAN, R.P. and RUDDLE, F.H. (1977) New approaches to human gene mapping by somatic cell genetics. In 'Molecular Structure of Human Chromosomes', ed. J.J. Yunis, pp.90-142. Academic Press: New York.

DALZIEL, J.W.D., KLINGFIELD, R., LOWRIE, W. and CEPDYKE, N.C. (1973) Palaeomagnetic data from the southernmost Andes and the Antarctandes. In 'Implications of Continental Drift to the Earth Sciences', Vol. I, ed. D.H. Tarling and S.K. Runeon, pp.87-101. Academic Press: London.

DARNELL, J.E., WALL, R. and TUSHINSKI, R.J. (1971) An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. *Proc. Natl. Acad. Sci. USA*, 68: 1321-5.

DAVIES, K.E., YOUNG, B.D., ELLES, R.G., HILL, M.E. and WILLIAMSON, R. (1981) Cloning of a representative genomic library of the human X chromosomes after sorting of flow cytometry. *Nature*, 293: 374-6.

DAYHOFF, M.O. (1975) 'Atlas of Protein Sequence and Structure.' National Biomedical Research Foundation, Washington D.C.

DEACON, N.J., SHINE, J. and NAORA, H. (1980) Complete nucleotide sequence of a cloned chicken α-globin cDNA. *Nuc. Acids Res.*, 8: 1186-99.

DEAN, P.N. and PINKEL, D. (1978) High resolution dual laser flow cytometry. *J. Histochem. Cytochem.*, 26: 622-7.

DEISSEROTH, A., NIENHUIS, A., LAWRENCE, J., GILES, R., TURNER, P. and RUDDLE, F.H. (1978) Chromosomal localisation of human β-globin gene on human chromosome 11 in somatic cell hybrids. *Proc. Natl. Acad. Sci. USA*, 75: 1456-60.

DEISSEROTH, A., NIENHUIS, A., TURNER, P., VELEZ, R., ANDERSON, W.F., RUDDLE, F., LAWRENCE, J., CREAGAN, R. and KUCHERLAPATI, R. (1977) Localisation of the human α-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridisation assay. *Cell*, 12: 205-18.

D'EUSTACHIO, P. and RUDDLE, F.H. (1983) Somatic cell genetics and gene families. *Science*, 220: 919-24.

DISTECHE, C.M., CARRANO, A.V., ASHWORTH, L.K., BURKHART-SCHULTZ, K. and LATT, S.A. (1981) Flow sorting of the mouse Cattanach X chromosome translocation in the active and inactive state. *Cytogenet. Cell Genet.*, 29: 189-97.

DISTECHE, C.M., CERKER, S.H., LOJEWSKI, A., KUNKEL, L.M., EISENHARD, M., SAHAR, E. and LATT, S.A. (1981) Construction of an X-enriched lambda phage library from sorted mouse chromosomes. *Cytometry*, 2: 95-6.

DODGSON, J.B. and ENGEL, J.D. (1983) The nucleotide sequence of the adult chicken α-globin genes. *J. Biol. Chem.*, 258: 4623-9.

DODGSON, J.B., McCUNE, K.C., ROSLING, D.J., KRUST, A. and ENGEL, J.D. (1981) Adult chicken α-globin genes α^A and α^D. No anaemic shock α globin exists in domestic chickens. *Proc. Natl. Acad. Sci.*, 78: 5998-6002.

DOLAN, M., SUGARMAN, B.J., DODGSON, J.B. and ENGEL, J.D. (1981) Chromosomal arrangement of the chicken β-type globin genes. *Cell*, 24: 669-77.

DONALD, J.A. and ADAMS, M.A. (1981) An autosomal gene assignment on a marsupial: the gene for LDH-A is on chromosome 5 of the red kangaroo, *Macropus rufus*. *Biochem. Genet.*, 19: 901-8.

DONALD, J.A. and HOPE, R.M. (1981) Mapping a marsupial X chromosome using kangaroo-mouse somatic cell hybrids. *Cytogenet. Cell Genet.*, 29: 127-37.

DUNSMUIR, P. (1976) Satellite DNA in the kangaroo *Macropus rufogriseus*. *Chromosoma*, 56: 111-25.

EATON, W.A. (1980) The relationship between coding sequences and function in haemoglobin. *Nature*, 284: 183-5.

EDMONDS, M., VAUGHAN, M.H. and NAKAZATO, H. (1971) Polyadenylic acid sequences in the heterogenous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: Possible evidence for a precursor relationship. *Proc. Natl. Acad. Sci. USA*, 68: 1336.

EFSTRATIADIS, A., KAFATOS, F.C. and MANIATIS, T. (1977) The primary structure of rabbit β-globin mRNA as determined from cloned DNA. *Cell*, 10: 571-85.

EFSTRATIADIS, A., MANIATIS, T., KAFATOS, F.C., JEFFREY, A. and VOURNAKIS, J.N. (1975) Full length and discrete partial reverse transcripts of globin and chorion mRNAs. *Cell*, 4: 367.

EFSTRATIADIS, A., POSAKONY, J.W., MANIATIS, T., LAWN, R.M., O'CONNEL, C.O., SPRITZ, R.A., DEKIEL, J.K., FORGET, B.G., WEISSMAN, S.M., SLIGHTOM, J.L., BLECHL, A.E., SMITHIES, O., BARALLE, F.E., SHOULDERS, C.C. and PROUDFOOT, N.J. (1980) The structure and evolution of the human β-globin gene family. *Cell*, 21:653-68.

ENGEL, J.D. and DODGSON, J.B. (1978) Analysis of the adult and embryonic chicken globin genes in chromosomal DNA. *J. Biol. Chem.*, 253: 8239-46.

ENGEL, J.D. and DODGSON, J.B. (1980) Analysis of the closely linked adult chicken α-globin genes in recombinant DNAs. *Proc. Natl. Acad. Sci. USA*, 77: 2596-600.

ENGEL, J.D., RUSLING, D.J., McCURE, K.C. and DODGSON, J.B. (1983) Unusual structure of the chicken embryonic alpha-globin gene, pi^1. *Proc. Natl. Acad. Sci. USA*, 80: 1392-6.

EPHRUSSI, B. and WEISS, M.C. (1965) Interspecific hybridisation of somatic cells. *Proc. Natl. Acad. Sci. USA*, 53: 1040-2.

FELBER, B.K., ORKIN, S.H. and HAMER, D.H. (1982) Abnormal RNA splicing causes one form of alpha thalassaemia. *Cell*, 29: 895-902.

FISHER, R.A. (1930) 'The Genetical Theory of Natural Selection.' Clarendon Press: Oxford.

FLAVELL, R.A., KOOTER, J.M., DeBEER, E., LITTLE, P.F.R. and WILLIAMSON, R. (1978) Analysis of the β- δ-globin gene loci in normal and Hb Lepore DNA: direct determination of gene linkage and intergene distance. *Cell*, 15: 25-41.

FRTISCH, E.F., LAWN, R.M. and MANIATIS, T. (1980) Molecular cloning and characterisation of the human β-like globin gene cluster. *Cell*, 19: 959-72.

FURUCCHI, Y., LA FIANDRA, A. and SHATKIN, A.J. (1977) 5'-terminal structure and mRNA stability. *Nature*, 266: 235-8.

GALL, J.G. and PARDUE, M.L. (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. USA*, 64: 600-4.

GARTLER, S.M. and ANDINA, R.J. (1976) Mammalian X-chromosome inactivation. *Adv. Hum. Genet.*, 7: 99-140.

GERHARD, D.S., KAWASAKI, E.S., BANCROFT, F.C. and SZABO, P. (1981) Localisation of a unique gene by direct hybridisation *in situ*. *Proc. Natl. Acad. Sci. USA*, 78: 3753-9.

GILBERT, W. (1978) Why genes in pieces? *Nature*, 271: 501.

GOEDDEL, D.V., LEUNG, D.W., DULL, T.J., GROSS, M., LAWN, R.M. McCANDLISS, R., SEEBURG, P., ULRICH, A., YELVOTONS, E. and GRAY, P.W. (1981) The structure of eight distinct cloned human leukocyte interferon cDNAs. *Nature*, 290: 20-6.

GOODMAN, M., MOORE, W.G. and MATSUDA, G. (1975) Darwinian evolution in the geneology of haemoglobin. *Nature*, 253: 603-8.

GOOSSENS, M., DOZY, A.M., EMBURG, S.H., ZACHARIADES, Z., HADJIMINAS, M.G., STAMATOYANNOPOULOS, G. and Y.W. KAN (1980) Triplicated α-globin loci in humans. *Proc. Natl. Acad. Sci. USA*, 77: 518-21.

GRAVES, J.A.M. (1983) Inactivation and reactivation of the mammalian X-chromosome - on the threshold of molecular biology. In 'Development in Mammals', Vol. 5, ed. M.H. Johnson. Elsevier Press.

GRAVES, J.A.M., CHEW, G.K., COOPER, D.W. and JOHNSTON, P.G. (1979) Marsupial-mouse cell hybrids containing fragments of the marsupial X-chromosome. *Somat. Cell Genet.*, 5: 481-9.

GRAVES, J.A.M. and HOPE, R.M. (1977a) Fusion and hybridisation of marsupial and eutherian cells. I. Growth of marsupial cells and co-cultivation with eutherian cells. *Aust. J. Biol. Sci.*, 30: 445-9.

GRAVES, J.A.M. and HOPE, R.M. (1977b) Fusion and hybridisation of marsupial and eutherian cells. II. Fusion of marsupial cells. *Aust. J. Biol. Sci.*, 30: 461-9.

GRAVES, J.A.M. and HOPE, R.M. (1978) Fusion and hybridisation of marsupial and eutherian cells. IV. Activity of heterokaryons. *Aust. J. Exptl. Biol. Med. Sci.*, 56: 341-50.

GRAVES, J.A.M., HOPE, R.M. and MacGREGOR, A. (1977) Fusion and hybridisation of marsupial and eutherian cells. III. Heterokaryon formation. *Aust. J. Exptl. Biol. Med. Sci.*, 55: 625-33.

GRAY, J.W., CARRANO, A.V., MOORE II, D.H., STEINMETZ, L.L., MINKLER, J., MAYALL, B.H., MENDELSOHN, M.L. and VAN DILLA, M.A. (1975b) High speed quantitative karyotyping by flow microfluorometry. *Clin. Chem.*, 21: 1258.

GRAY, J.W., CARRANO, A.V., STEINMETZ, L.L., VAN DILLA, M.A., MOORE II, D.H., MAYALL, B.H. and MENDELSOHN, M.L. (1975a) Chromosome measurement and sorting by flow systems. *Proc. Natl. Acad. Sci. USA*, 72: 1231-4.

GRUNSTEIN, M.L. and HOGNESS, D.S. (1975) Colony hybridisation: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA*, 72: 3961-5.

GURDON, J.B., De ROBERTS, E.M. and DARLINGTON, G. (1976) Injected nuclei in frog oocytes provided a living cell system for the study of transcriptional control. *Nature*, 260: 116-20.

GURDON, J.B., LINGREL, J.B. and MARBAIN, G. (1973) Message stability in injected frog oocytes: long life of mammalian α- and β-globin messages. *J. Mol. Biol.*, 80: 539-51.

HAGENBUCHLE, O., TOSI, M., SCHIBLER, U., BOVEY, R., WELLAUER, P.K. and YOUNG, R.A. (1981) Mouse liver and salivary gland α-amylase mRNAs differ only in 5' mon-translated sequences. *Nature*, 289: 643-6.

HALDANE, J.B.S. (1957) The cost of natural selection. *J. Genet.*, 55: 511-24.

HAMER, D.H. and LEDER, P. (1979) SV40 recombinants carrying a functional RNA splice junction and polyadenylation site from the chromosomal mouse β^{maj} globin gene. *Cell*, 17: 737-47.

HAMLYN, P.H., GAIT, M.J. and MILSTEIN, C. (1981) Complete sequence of an immunoglobin mRNA using specific priming and the dideoxy-nucleotide method of RNA sequencing. *Nuc. Acids Res.*, 9: 4485-94.

HANDEL, A.M., ALONI, Y., RAFTOY, M.A. and ATTARDI, G. (1972) Electro-focusing analysis of HeLa cell metaphase chromosomes. *Biochemistry*, 11: 1654-63.

HARDISON, R.C., BUTLER, E.T., LACY, E., MANIATIS, T., ROSENTHAL, N. and EFSTRATIADIS, A. (1979) The structure and transcription of four linked rabbit β-like globin genes. *Cell*, 18: 1285-97.

HARPER, M.E., BARRERA-SALDANA, M.A. and SAUNDERS, G.F. (1982) Chromosomal localisation of the human placental lactogen-growth hormone gene cluster to 17q22-24. *Am. J. Hum. Genet.*, 34: 227-34.

HARPER, M.E., ULLRICH, A. and SAUNDERS, G.F. (1981) Localisation of the human insulin gene to the distal end of the short arm of chromosome 11. *Proc. Natl. Acad. Sci. USA*, 78: 4458-60.

HARRIS, H. (1978) Multilocus enzymes in man. In 'Ciba Foundation Symposium No. 66, Human Genetics: Possibilities and Realities.'

HARRISON, P.R., CONTIE, D., AFFARA, N. and PAUL, J. (1974) *In situ* localization of globin messenger RNA formation. *J. Cell Biol.*, 63: 402-13.

HAYES, H.E. and RINGIS, J. (1973) Seafloor spreading in the Tasman Sea. *Nature*, 243: 454-58.

HAYMAN, D.L. (1977) Chromosome number - constancy and variation. In 'The Biology of the Marsupials', ed. B. Stonehouse and D. Gilmore, p.27-48. Macmillan Press: London.

HAYMAN, D.L., ASHWORTH, L.K. and CARRANO, A.V. (1982) The relative DNA contents of the eutherian and marsupial X chromosomes. *Cytogenet. Cell Genet.*, 34: 265-70.

HAYMAN, D.L. and MARTIN, P.G. (1974) Monotremata and Marsupialia. Mammalia I, Vol. 4: Chordata 4. In 'Animal Cytogenetics' ed. B. John, pp.107. Gebrüder Borntraeger: Berlin.

HAYMAN, D.L. and SHARP, P.J. (1981) Hoechst 33258 induced uncondensed sites in marsupial chromosomes. *Chromosoma*, 83: 249-62.

HEINDELL, H.C., Llu, A., PADDOCK, G.V., STRIDNICKA, G.M. and SALSER, W.A. (1978) The primary sequence of rabbit α-globin mRNA. *Cell*, 15: 43-54.

HENDERSON, A.S. (1982) Hybridisation to mammalian chromosomes. *Int. Rev. Cytol.*, 76: 1-43.

HOBART, P., CRAWFORD, R., SHEN, L.P., PICTET, R. and RUTTER, W.J. (1980) Cloning and sequence analysis of cDNAs encoding two distinct somatostatin precursors found in the endocrine pancreas of anglerfish. *Nature*, 288: 137-41.

HOPE, R.M. (1969) Genetic Variation in Marsupials. Ph.D. Thesis, University of Adelaide.

HOPE, R.M. and GRAVES, J.A.M. (1978a) Fusion and hybridisation of marsupial and eutherian cells. V. Development of selective systems. *Aust. J. Biol. Sci.*, 31: 293-301.

HOPE, R.M. and GRAVES, J.A.M. (1978b) Fusion and hybridisation of marsupial and eutherian cells. VI. Hybridization. *Aust. J. Biol. Sci.*, 31: 527-43.

HOSBACH, H.A., WYLER, T. and WEBER, R. (1983) The *Xenopus laevis* globin gene family: chromosomal arrangement and gene structure. *Cell*, 32: 45-53.

HUEZ, G., MARBAIN, G., GALLWITZ, D., WEINBERG, E., DEVOS, R., HUBERT, E. and CLENTO, Y. (1978) Functional stabilisation of HeLa cell histone messenger RNAs injected into *Xenopus* oocytes by 3'-OH polyadenylation. *Nature*, 271: 572-3.

HUGHES, S.H., STUBBLEFIELD, E., PAYVAR, F., ENGEL, J.D., DODGSON, J.B., SPECTOR, P., CORDELL, B., SCHIMKE, R.T. and VARMUS, H.E. (1979) Gene localisation by chromosome fractionation: globin genes are on at least two chromosomes and three oestrogen-inducible genes are on three chromosomes. *Proc. Natl. Acad. Sci. USA*, 76: 1348-52.

HUTCHINSON, N.J., LANGER-SAFER, P.R., WARD, D.C. and HAMKALO, B. (1982) *In situ* hybridization at the electron microscope level: hybrid detection by autoradiography and colloidal gold. *J. Cell Biol.*, 95: 609-18.

JACKSON, D.A., SYMONS, R.H. and BERG, P. (1972) Biochemical method for inserting new genetic information into DNA of simian virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 69: 2904-9.

JAHN, C.L., HUTCHINSON, III, A., PHILLIPS, S.J., WEAVER, S., HAIGWOOD, N.L., VOLIVA, C.F. and EDGELL, M.H. (1980) DNA sequence organisation of the β-globin complex in the BALB/c mouse. *Cell*, 21: 159-68.

JEFFREYS, A.J. (1981) Recent studies of gene evolution using recombinant DNA. In 'Genetic Engineering 2', ed. R. Williamson. Academic Press.

JEFFREYS, A.J., CRAIG, I.W. and FRANCKE, U. (1979) Localisation of $^{G}\gamma$ and $^{A}\gamma$, and β-globin genes on the short arm of human chromosome 11. *Nature*, 281: 606-8.

JEFFREYS, A.J., WILSON, V., WOOD, D. and SIMONS, J.P. (1980) Linkage of adult α- and β-globin genes in *X. laevis* and gene duplication of tetraploidisation. *Cell*, 21: 555-64.

JOHN, H.A., BIRNSTIEL, M. and JONES, K.W. (1969) RNA-DNA hybrids at the cytological level. *Nature*, 223: 582-7.

JOHNSTON, P.G., SHARMAN, G.B., JAMES, E.A. and COOPER, D.W. (1978) Studies on metatherian sex chromosomes. VII. Glucose-6-phosphate dehydrogenase expression in tissues and cultured fibroblasts of kangaroos. *Aust. J. Biol. Sci.*, 31: 415-24.

JOHNSTON, P.G., VANDEBERG, J.L. and SHARMAN, G.B. (1975) Inheritance of erythrocyte glucose-6-phosphate dehydrogenase in the red necked wallaby, *Macropus rufogriseus* (Desmarest), consistent with paternal X-inactivation. *Biochem. Genet.*, 13: 235-42.

KATES, J. (1970) Transcription of the vaccinia virus genome and the occurrence of polyriboadenylic acid sequences in messenger RNA. *Cold Spring Harbor Symp. Quant. Biol.*, 35: 743.

KATES, J., BYERS, M.J. and JOHNS, E.W. (1973) Detection and utilization of poly(A) sequences in messenger RNA. *Methods in Cell Biology*, 7: 53.

KAY, R.M., HARRIS, R., PATIENT, R.K. and WILLIAMS, J.G. (1980) Molecular cloning of cDNA sequences coding for the major α- and β-globin polypeptides of adult *Xenopus laevis*. *Nuc. Acids Res.*, 8: 2691-701.

KAY, R.M., HARRIS, R., PATIENT, R.K. and WILLIAMS, J.G. (1983) Complete nucleotide sequence of a cloned cDNA derived from the major adult α-globin mRNA of *X. laevis*. *Nuc. Acids Res.*, 11: 1537-43.

KEAST, A. (1977) Historical biogeography of the marsupials. In 'The Biology of the Marsupials', ed. B. Stonehouse and D. Gilmore, p.69. Macmillan Press: London.

KENDREW, J.C., WATSON, H.C., STRADBERG, B.E., DICKERSON, R.E., PHILLIPS, D.C. and SHORE, V.C. (1961) A partial determination by X-ray methods, and its correlation with chemical data. *Nature*, 190: 666-70.

KIMURA, M. (1969) Evolutionary rate at the molecular level. *Nature*, 217: 624.

KIMURA, M. (1983) 'The Neutral Theory of Molecular Evolution.' Cambridge University Press.

KING, J.L. and JUKES, T.H. (1969) Non-Darwinian evolution. *Science*, 164: 788.

KIRSCH, J.A.W. and CALABY, J.H. (1977) The species of living marsupials - an annotated list. In 'The Biology of Marsupials', ed. B. Stonehouse and D. Gilmore, pp.9-26. Macmillan Press: London.

KNOCHEL, W., MEYERHEF, W., HUMMEL, S. and GRUNDMANN, U. (1983) Molecular cloning and sequencing of mRNAs coding for minor adult globin polypeptides of *X. laevis*. *Nuc. Acids Res.*, 11: 1543-53.

KONKEL, D.A., MAIZEL, J.V. and LEDER, P. (1979) The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. *Cell*, 18: 865.

KRONENBERG, H.M., ROBERTS, B.E. and EFSTRATIADIS, A. (1979) The 3' noncoding region of β-globin mRNA is not essential for *in vitro* translation. *Nuc. Acids Res.*, 6: 153-66.

LACY, E., HARDISON, R.C., QUON, D. and MANIATIS, T. (1979) The linkage arrangement of four rabbit β-like globin genes. *Cell*, 18: 1273-83.

LACY, E. and MANIATIS, T. (1980) The nucleotide sequence of a rabbit β-globin pseudogene. *Cell*, 21: 545-53.

LALLEY, P.A., FRANCKE, U. and MINNA, J.D. (1978a) Homologous genes for enolase, phosphoglucomutase and adenylate kinase are syntenic on mouse chromosome 4 and human chromosome 1p. *Proc. Natl. Acad. Sci. USA*, 75: 2382-86.

LALLEY, P.A., MINNA, J.D. and FRANCKE, U. (1978b) Conservation of autosomal gene synteny groups in mouse and man. *Nature*, 274: 160-3.

LANGER, P.R., WALDROP, A.A. and WARD, D.C. (1982) Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. USA*, 78: 6633-7.

LANGER-SAFER. P.R., LEVINE, M. and WARD, D.C. (1982) Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci. USA*, 79: 4381-5.

LANGLOIS, R.G., CARRANO, A.V., GRAY, J.W. and VAN DILLA, M.A. (1980) Cytochemical studies of metaphase chromosomes by flow cytometry. *Chromosoma*, 77: 229-51.

LAUER, J., SHEN, C.-K. and MANIATIS, T. (1980) The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions. *Cell*, 20: 119-30.

LAURIE, S. and GOSDEN, J. (1980) The identification of human chromosomes by quinacrine fluorescence after hybridisation *in situ*. *Hum. Genet.*, 53: 371-6.

LEBO, R.V., CARRANO, A.V., BURKHART-SCHULTZ, K., DOZY, A.M., YU, L.-C. and KAN, U.W. (1979) Assignment of human β-, γ-, and δ-globin genes to the short arm of chromosome 11 by chromosome sorting and DNA restriction enzyme analysis. *Proc. Natl. Acad. Sci. USA*, 76: 5804-8.

LEBO, R.V., KAN, Y.W., CHEUNG, M.C., CARRANO, A.V., YU, L.-C., CHANG, J.C., CORDELL, B. and GOODMAN, H.M. (1982) Assigning the polymorphic human insulin gene to the short arm of chromosome 11 by chromosome sorting. *Hum. Genet.*, 60: 10-15.

LEDER, A., SWAN, D., RUDDLE, F., d'EUSTACHIO, P. and LEDER, P. (1981) Dispersion of α-like globin genes of the mouse to three different chromosomes. *Nature*, 293: 196-200.

LEDER, P., HANSEN, J.N., KONKEL, D., LEDER, A., NISHIOKA, Y. and TALKINGTON. C. (1980) Mouse globin system: a functional and evolutionary analysis. *Science*, 209: 1336-42.

LEE, S.Y., MENDECKI, J. and BRAWERMAN, G. (1971) A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells. *Proc. Natl. Acad. Sci. USA*, 68: 1331.

LIEBHABER, S.A., GOOSSENS, M. and KAU, Y.-W. (1981) Homology and concerted evolution at the α1 and α2 loci of human α-globin. *Nature*, 290: 26-9.

LILLEGRAVEN, J.A. (1969) Latest Cretaceous mammals of upper part of Edmonton formation of Alberta, Canada, and review of marsupial-placental dichotomy in mammalian evolution. *Palaeontol. Contrib. Univ. Kans., Art.* 50.

LILLEGRAVEN, J.A. (1974) Biogeographical considerations of the marsupial-placental dichotomy. *Ann. Rev. Ecol. System*, 5: 263-83.

LINGREL, J.B., MORRISON, M.D. and GOSHI. J. (1972) The separation of mouse reticulocyte 9S RNA into fractions of different biological activity by hybridisation to poly V-cellulose. *Biochem. Biophys. Res. Acta*, 49: 775-81.

LITTLE, P.F.R., FLAVELL, R.A., KOOTER, J.M., ANNINSON, G. and WILLIAMSON, R. (1979) Structure of the human foetal globin gene locus. *Nature*, 278: 227-31.

LOBBAN, P.E. and KAISER, A.D. (1973) Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.*, 78: 453-71.

LOWENSTEIN, J.M., SARICH, U.M. and RICHARDSON, B.J. (1981) Albumin systematics of the extinct mammoth and Tasmanian wolf. *Nature*, 291: 409-11.

LUNDBERG, U. and PERSSON, T. (1972) Isolation of mRNA from KB cells by affinity chromatography on polyuridylic acid covalently linked to sepharose. *Eurp. J. Biochem.*, 31: 246.

LYON, M.F. (1972) X-chromosome inactivation and developmental patterns in mammals. *Biol. Rev.*, 47: 1-35.

LYON, M.F. (1974) Evolution of X-chromosome inactivation in mammals. *Nature*, 250: 651-53.

MacDONALD, R.J., CREROV, M.M., SWAIN, W.F., PICTET, R.L., THOMAS, G. and RUTTER, W. (1980) Structure of a family of rat amylase genes. *Nature*, 287: 117-22.

MALCOLM, S., BARTON, P., MURPHY, C. and FERGUSON-SMITH, M.A. (1981) Chromosomal localisation of a single copy gene by *in situ* hybridization - human β-globin genes on the short arm of chromosome 11. *Ann. Hum. Genet.*, 45: 135-41.

MALCOLM, S., BARTON, P., MURPHY, C., FERGUSON-SMITH, M.A., BENTLEY, D.L. and RABBITTS, T.H. (1982) Localisation of human immunoglobulin K light chain variable region genes to the short arm of chromosome 2 by *in situ* hybridisation. *Proc. Natl. Acad. Sci. USA*, 79: 4957-61.

MANIATIS, G. (1980) Recombinant DNA procedures in the study of eukaryotic genes. In 'Cell Biology', Vol. 3, 563-608.

MANIATIS, T., FRITSCH, E.F., LAUER, J. and LAWN, R.M. (1980) The molecular genetics of human hemoglobins. *Ann. Rev. Genet.*, 14: 145-78.

MANIATIS, T., FRITSCH, E.F. and SAMBROOK. J. (1982) Molecular cloning. A laboratory manual. Cold Spring Harbour Press.

MANIATIS, T., JEFFREY, A. and KLEID, D.G. (1975) A nucleotide sequence of the rightward operator of phage λ. *Proc. Natl. Acad. Sci. USA*, 77: 1184-88.

MANIATIS, T., SIM, G.K., EFSTRATIADIS, A. and KAFATOS, F.C. (1976) Amplification and characterization of a β-globin gene synthesized *in vitro*. *Cell*, 8: 163-82.

MANUELIDIS, L., LANGER-SAFER, P.R. and WARD, D.C. (1982) High-resolution mapping of satellite DNA using biotin-labeled DNA probes. *J. Cell Biol.*, 95: 619-25.

MARTIN, S.L., VINCENT, K.A. and WILSON, A.C. (1983) Rise and fall of the delta globin gene. *J. Mol. Biol.*, 164: 513-38.

MARTIN, S.L., ZIMMER, E.A., KAN, Y.W. and WILSON, A.C. (1980) Silent δ-globin gene in Old World monkeys. *Proc. Natl. Acad. Sci. USA*, 77: 3563-66.

MAXAM, A. and GILBERT, W. (1980) In 'Methods in Enzymology' Vol. 65 Part 1, pp.599-560, ed. L. Grossman and K. Moldave. Academic Press: New York.

MAXSON, L.R., SARICH, V.M. and WILSON, A.C. (1975) Continental drift and the use of albumins as an evolutionary clock. *Nature*, 255: 397-400.

MAYNARD-SMITH, J. (1968) "Haldane's dilemma" and the rate of evolution. *Nature*, 219: 1114.

McREYNOLDS, L., O'MALLEY, B.W., NISBET, A.D., FOTHERGILL, J.E., GIVOL, D., FIELDS, S., ROBERTSON, M. and BROWNLEE, G.G. (1978) Sequence of chickens ovalbumin mRNA. *Nature*, 273: 723-8.

MESSING, J., CREA, R. and SEEBURG, P.H. (1981) A system for shotgun DNA sequencing. *Nuc. Acids Res.*, 9: 309-21.

MIKLOS, G.L. and JOHN, B. (1979) Heterochromatin and satellite DNA in man: properties and prospects. *Am. J. Hum. Genet.*, 31: 264-80.

MOORE II, D.H. (1975) Use of residuals in fitting normal (Gaussian) distributions. UCRL - 76507.

MOSS, B.A. and THOMPSON, E.O.P. (1969) Haemoglobins of the adult domestic fowl *Gallus domesticus*. *Aust. J. Biol. Sci.*, 22: 145.

NASH, W.G. and O'BRIEN, S.J. (1982) Conserved regions of homologous G-banded chromosomes between orders in mammalian evolution: Carnivores and primates. *Proc. Natl. Acad. Sci. USA*, 79: 6631-5.

NISHIOKA, J. and LEDER, P. (1979) The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. *Cell*, 18: 875-82.

NISHIOKA, Y., LEDER, A. and LEDER, P. (1980) Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. *Proc. Natl. Acad. Sci. USA*, 77: 2806-9.

NUDEL, U., RAMIREZ, F., MARKS, P.A. and BANK, A. (1977) Preparative polyacrylamide gel electrophoretic purification of human α- and β-globin messenger RNAs. *J. Biol. Chem.*, 252: 2182-6.

NUNBERG, J.H., KAUFMAN, R.J., CHANG, A.C.Y., COHEN, S.N. and SCHIMKE, R.T. (1980) Structure and genomic organization of the mouse dehydrofolate reductase gene. *Cell*, 19: 355-64.

OHNO, S. (1967) 'Sex Chromosomes and Sex-linked Genes.' Springer-Verlag: Berlin.

OHNO, S. (1973) Ancient linkage groups and frozen accidents. *Nature*, 244: 259-62.

ORKIN, S.H. (1978) The duplicated human α-globin genes lie close together in cellular DNA. *Proc. Natl. Acad. Sci. USA*, 75: 5950-4.

ORKIN, S.H., LAZARUS, H., ALTAY, C., GURGEY, A., WEATHERALL, D.J. and NATHAN, D. (1979) The molecular basis of α-thalassemias: frequent occurrence of dysfunctional α loci among non-asians with Hb H disease. *Cell*, 17: 33.

OWERBACH, D., RUTTER, W.J., MARTIAL, J.A., BUXTO, J.D. and SHOWS, T.B. (1980) Genes for growth hormone, chorionic somatomammotropin, and growth hormone-like gene on chromosome 17 in humans. *Science*, 209: 289-92.

PADGETT, T.G., STUBBLEFIELD, E. and VARMUS, H.E. (1977) Chicken macrochromosomes contain an endogenous provirus and micro-chromosomes contain sequences related to the transforming gene of ASU. *Cell*, 10: 649-57.

PATERSON, B.M., BRYAN, E. ROBERTS and EDWARD L. HULF. (1977) Structural gene identification and mapping by DNA-mRNA hybrids - arrested cell-free translation. *Proc. Natl. Acad. Sci. USA*, 74: 4370-4.

PATIENT, R.K., ELKINGTON, J.A., KAY, R.M. and WILLIAMS, J.G. (1980) Internal organisation of the major adult α- and β-globin genes of *Xenopus laevis*. *Cell*, 21: 565-73.

PARNES, A.R., VELAN, B., FELSENFELD, A., RAMANATHAN, L., FERRINI, U., APPELLA, E. and SEIDMAN, J.G. (1981) Mouse β_2-microglobulin cDNA clones: a screening procedure for cDNA clones corresponding to rare mRNAs. *Proc. Natl. Acad. Sci. USA*, 78: 2253-7.

PEACOCK, S.L., McIVER, C.M. and MONOBAN, J.J. (1981) Transformation of *E. coli* using homopolymer-linked plasmid chimeras. *Biochim. Biophys. Acta*, 655: 243.

PEARSON, P.L. and RODERICK, T.H. (1978) Report of the committee on comparative mapping. In *Cytogenet. Cell Genet.*, 22: 150-62. 4th International Workshop on Human Gene Mapping, Winnipeg, 1977.

PEARSON, P.L., RODERICK, T.H., DAVISSON, M.T., LALLEY, P.A. and OBRIEN, S.J. (1982) Report of the committee on comparative mapping. *Cytogenet. Cell Genet.*, 32: 208-20.

PELHAM, H.R.B. and JACKSON, R.J. (1976) An efficient mRNA-dependent translation system from reticulocyte lysates. *Eurp. J. Biochem.*, 67: 247-56.

PERLER, F., EFSTRATIADIS, A., LOMEDICO, P., GILBERT, W., KOLODNER, R. and DODGSON, J. (1980) The evolution of genes: the chicken preproinsulin gene. *Cell*, 20: 555-66.

PERUTZ, M.F. (1976) Structure and mechanism of haemoglobin. *Br. Med. Bull.*, 32: 195-216.

PERUTZ, M.F. and TEN EYCK, L.F. (1972) Stereochemistry of co-operative effects in hemoglobin. *Cold Spring Harbour Symp. Quant. Biol.*, 36: 295-310.

POPP, R.A., LALLEY, P.A., WHITNEY III, J.B. and ANDERSON, W.F. (1981) Mouse α-globin genes and α-globin-like pseudogenes are not syntenic. *Proc. Natl. Acad. Sci. USA*, 78: 6362-6.

PORTER, A.G., FELLNER, P., BLACK, D.N., ROWLANDS, D.J., HARRIS, T.J.R. and BROWN, F. (1978) 3'-terminal nucleotide sequences in the genome RNA of picornaviruses. *Nature*, 276: 298-301.

PRENSKY, W. and HOLMQUIST, G. (1973) Chromosomal localisation of human haemoglobin structural genes: techniques queried. *Nature*, 243: 44.

PRICE, P.M., CONOVER, J.H. and HIRSCHORN, K. (1977) Chromosomal localisation of human haemoglobin structural genes. *Nature*, 237: 340-2.

PROUDFOOT, N.J. and BROWNLEE, G.G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA. *Nature*, 263: 211-4.

PROUDFOOT, N.J., GIL, A. and MANIATIS, T. (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell*, 31: 553-63.

PROUDFOOT, N.J. and MANIATIS, T. (1980) The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. *Cell*, 21: 537-44.

RABBITS, T.H. (1976) Bacterial cloning of plasmids carrying copies of rabbit globin messenger RNA. *Nature*, 260: 221-5.

RAPHAEL, K.A. and COOPER, D.W. (1978) Studies on metatherian sex chromosomes. VIII. Evidence for an absence of dosage compensation at the glucose-6-phosphate dehydrogenase locus in cultured cells of *Macropus rufogriseus*. *Aust. J. Biol. Sci.*, 31: 425-31.

REANNEY, D. (1979) RNA splicing and polynucleotide evolution. *Nature*, 277: 598-600.

RENWICK, J.H. (1971) The mapping of human chromosomes. *Ann. Rev. Genet.*, 5: 1-81.

RICHARDS, R.J., SHINE, J., ULRICH, A., WELLS, J.R.E. and GOODMAN, H.M. (1979) Molecular cloning and sequence analysis of adult chicken β-globin cDNA. *Nuc. Acids Res.*, 7: 1137-46.

RICHARDSON, B.J. and RUSSEL, E.M. (1969) Changes with ages in the proportion of nucleated red blood cell types and in the type of haemoglobin in kangaroo pouch young. *Aust. J. Exptl. Biol. Med. Sci.*, 47: 573-80.

RICHMOND, R.C. (1970) Non-Darwinian evolution: a critique. *Nature*, 225: 1025.

ROBERTS, B.E. and PATERSON, B.M. (1973) Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. *Proc. Natl. Acad. Sci. USA*, 70: 2330-4.

ROBINS, A.J. (1983) Purification and Characterisation of Chicken Globin mRNA Sequences. Ph.D. Thesis, University of Adelaide.

ROBINS, D.M., RIPLEY, S., HENDERSON, A.S. and AXEL, R. (1981) Transforming DNA integrates into the host chromosome. *Cell*, 23: 29-39.

ROFE, R. (1979) G-banding Studies in Australian Marsupials. Ph.D. Thesis, University of Adelaide.

ROMERO-HERRERA, A.E., LEHMANN, A.E., JOYSEY, K.A. and FRIDAY, A.E. (1973) Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. *Nature*, 246: 389-95.

RONINSON, J.B. and INGRAM. V.M. (1982) Gene evolution in the chicken β-globin cluster. *Cell*, 28: 515-21.

ROOP, D., TSAI, M. and O'MALLEY, B. (1980) Definition of the 5' and 3' ends of transcripts of the ovalbumin gene. *Cell*, 19: 63-8.

ROSENBERG, M. and PATERSON, B.M. (1979) Efficient cap-dependent translation of polycistronic prokaryotic mRNAs is restricted to the first gene in the operon. *Nature*, 279: 696-701.

ROUGEN, F. and MACH, B. (1976) Stepwise biosynthesis *in vitro* of globin genes from globin mRNA by DNA polymerase of avian myeloblastosis virus. *Proc. Natl. Acad. Sci. USA*, 73: 3418-22.

ROYCHOUDHURY, R., JAY, E. and WU, R. (1976) Terminal labelling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nuc. Acids Res.*, 3: 101-16.

RUDDLE, F.H. (1970) Utilisation of somatic cells for genetic analysis: possibilities and problems. In 'Control Mechanisms in the Expression of Cellular Phenotypes', ed. H.A. Padykula, pp.233-264. Academic Press: New York.

RUDKIN, G. and STOLLER, B. (1977) High resolution detection of DNA-RNA hybrids *in situ* by direct immunofluorescence. *Nature*, 265: 472-3.

SANGER, F., NICKLEN, S. and COULSON, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74: 5463-7.

SARICH, V.M. (1973) The giant panda is a bear. *Nature*, 245: 218-20.

SAWIN, V.L., ROWLEY, J.D. and CARRANO, A.V. (1979) Transcription and hybridisation of ^{125}I-cRNA from flow sorted chromosomes. *Chromosoma*, 70: 293-304.

SCHON, E.A., CLEARY, M.L., HAYNES, J.R. and LINGREL, J.B. (1981) Structure and evolution of goat γ-, β^{C}-, and β^{A}-globin genes: three developmentally regulated genes contain inserted elements. *Cell*, 27: 359-69.

SCHUTZ, G., BEATO, M. and FEIGELSON, P. (1972) Isolation of eukaryotic messenger RNA on cellulose and its translation *in vitro*. *Biochem. Res. Comm.*, 49: 680-9.

SETZER, D.R., McGEOGAN, M., NUNBERG, J.H. and SCHIMKE, R.T. (1980) Size heterogeneity in the 3' end of dihydrofolate reductase messenger RNAs in mouse cells. *Cell*, 22: 361-70.

SHAPIRO, D.J. and SCHIMKE, R.T. (1975) Immunochemical isolation and characterisation of ovalbumin messenger ribonucleic acid. *J. Biol. Chem.*, 250: 1759-64.

SHARMAN, G.B. (1973) The chromosomes of non-eutherian mammals. In 'Cytotaxonomy and Vertebrate Evolution', ed. A.B. Chiarelli and E. Capanna, pp. 485-530. Academic Press: London.

SHARMAN, G.B. (1974) Marsupial taxonomy and phylogeny. *Australian Mammalogy*, 1: 137-54.

SHARP, P. (1982) Sex chromosome pairing during male meiosis in marsupials. *Chromosoma*, 86: 27-47.

SHINE, J. and DALGARNO, L. (1975) Determinant of cistron specificity in bacterial ribosomas. *Nature*, 264: 34-8.

SILLAR, R. and YOUNG, B.D. (1981) A new method for the preparation of metaphase chromosomes for flow analysis. *J. Histochem. Cytochem.*, 29: 74-8.

SLAUGHTER, B. (1968) Earliest known marsupials. *Science*, 162: 254.

SLIGHTON, J.L., BLECHL, A. and SMITHIES, O. (1980) Heiman fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell*, 21: 627-38.

SORIANO, P., SZABO, P. and BERNARDI, G. (1982) *EMBO Journal*, 1: 579-83.

SOUTHERN, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, 98: 503-517.

SPANDIDOS, D.A. and SIMINOVITCH, L. (1977) Linkage of markers controlling consecutive biochemical steps in CHO cells as demonstrated by chromosome transfer. *Cell*, 12: 235-42.

STENZEL, P.L. (1974) Opossum Hb chain sequence and neutral mutation theory. *Nature*, 252: 62-3.

STUBBLEFIELD, E., DEAVAN, L. and CRAM, L.S. (1975) Flow microfluorometric analysis of isolated Chinese hamster chromosomes. *Exp. Cell Res.*, 94: 464-8.

SUTCLIFFE, J.G. (1978) pBR 322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. *Nuc. Acids Res.*, 5: 2721-8.

SWANK, R.T. and MUNKRES, K.D. (1971) Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. *Analytical Biochem.*, 39: 462-77.

SZABO, P., ELDER, R., STEFFENSEN, D.M. and UHLENBECK, O.C. (1977) Quantitative *in situ* hybridization of ribosomal RNA species to polytene chromosomes of *Drosophila melanogaster*. *J. Mol. Biol.*, 115: 539-63.

TAYLOR, J.M. (1979) The isolation of eukaryotic messenger RNA. *Ann. Rev. Biochem.*, 48: 681-717.

TEDFORD, R.H. (1974) Marsupials and the new paleogeography. In 'Paleogeographic provinces and provinciality', ed. C.A. Ross. *Soc. Econ. Paleontol. Mineral Spec. Pub.*, No. 21: 109-26.

THOMPSON, E.O.P. and AIR, G.M. (1971) Studies on marsupial proteins. VI. Evolutionary changes in β-globins of the macropodidae and the amino acid sequence of β-globin from *Potorous tridactylus*. *Aust. J. Biol. Sci.*, 24: 1199.

THOMPSON, E.O.P., HOSKEN, R. and AIR, G.M. (1969) Studies on marsupial proteins. I. Polymorphism of haemoglobin of the grey kangaroo, *Macropus giganteus*. *Aust. J. Biol. Sci.*, 22: 449-62.

TRENT, J.M., OLSON, S. and LAWN, R.M. (1982) Chromosomal localisation of human leukocyte, fibroblast, and immune interferon genes by means of *in situ* hybridisation. *Proc. Natl. Acad. Sci. USA*, 79: 7809-13.

TSUJIMOTO, Y. and SUZUKI, Y. (1979) The DNA sequence of *Bombyx mori* fibroin gene including the 5' flanking, mRNA coding, entire intervening and fibroin protein coding regions. *Cell*, 18: 591-600.

UEDA, S. and SCHNEIDER, R.G. (1969) Rapid differentiation of polypeptide chains of hemoglobin by cellulose acetate electrophoresis of hemolysates. *Blood*, 34: 230-5.

VANDEBERG, J.L., COOPER, D.W., SHARMAN, G.B. and POOLE, W.E. (1977) Studies on metatherian sex chromosomes. IV. X-linkage of PGK-A with paternal X-inactivation confirmed in erythrocytes by pedigree analysis. *Aust. J. Biol. Sci.*, 30: 115-25.

VANDER, J.G., SHERMAN, T. and YOUNG, J. (1976) 'Physiology'. Wiley and Sons.

VAN DILLA, M., STEINMETZ, L., DAVIS, D., CALVERT, R. and GRAY, J. (1974) High speed cell analysis and sorting with flow systems: Biological applications and new approaches. *IEEE Trans. Nucl. Sci.*, 21: 714-720.

VANIN, E.F., GOLDBERG, G.S., TUCKER, P.W. and SMITHIES, O. (1980) A mouse α-globin-related pseudogene lacking intervening sequences. *Nature*, 286: 222-6.

VAN OOYEN, J., VAN DEN BERG, J., MANTEL, N. and WEISSMANN, C. (1979) Comparison of total sequence of a cloned rabbit β-globin gene and its flanking regions with a homologous mouse sequence. *Science*, 206: 337-44.

VENOLIA, L. and PEACOCK, W.J. (1981) A highly repeated DNA from the genome of the wallaroo (*Macropus robustus robustus*). *Aust. J. Biol. Sci.*, 34: 97-113.

VOGT, V.M. (1973) Purification and further properties of single-stranded-specific nuclease from *Aspergillus oryzae*. *Eurp. J. Biochem.*, 33: 192.

WAHL, G.M., STON, M. and STARK, G.R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci. USA*, 76: 3683-7.

WAHL, G.M., VITTO, L., PADGETT, R.A. and STARK, G.R. (1982) Single-copy and amplified CAD genes in Syrian hamster chromosomes localized by a highly-sensitive method for *in situ* hybridization. *Mol. Cell Biol.*, 2: 308-19.

WARBURTON, D., YU, M.T., ATWOOD, K.C. and HENDERSON, A.S. (1976) The location of RN5S genes on the chromosomes of primates. *Cytogenet. Cell Genet.*, 16: 440-2.

WEATHERALL, D.J. and CLEGG, J.B. (1979) Recent developments in the molecular genetics of human hemoglobin. *Cell*, 16: 467-479.

WEEDS, A.G. (1976) Light chains from slow-twitch muscle myosin. *Eurp. J. Biochem.*, 66: 157-73.

WEISS, M.C. and EPHRUSSI, B. (1966) Studies of interspecific (rat x mouse) hybrids: 1. Isolation, growth and evolution of the karyotype. *Genetics*, 54: 1095.

WIDMER, H.J., ANDRES, A.-C., NIESSING, J., HOSBACH, H.A. and WEBER, R. (1981) Comparative analysis of cloned larval and adult globin cDNA sequences of *Xenopus laevis*. *Dev. Biol.*, 88: 325-32.

WILLIAMS, J.G., KAY, R.M. and PATIENT, R.K. (1980) The nucleotide sequence of the major β-globin mRNA from *Xenopus laevis*. *Nuc. Acids Res.*, 8: 4247-59.

WILLIAMSON, R. (1980) The processing of mRNA and its relation to mRNA. In 'Cell Biology' Vol. 3. Academic Press: New York.

WILSON, A.C., CARLSON, S.S. and WHITE T.J. (1977a) Biochemical evolution. *Ann. Rev. Biochem.*, 46: 573-639.

WILSON, J.T., de RIEL, J.K., FORGET, B.G., MAROTTA, C.M. and WEISSMAN, S.M. (1977b). Nucleotide sequence of 3' untranslated portion of human alpha globin mRNA. *Nuc. Acids Res.*, 4: 2353.

WILSON, J.T., WILSON, L.B., REDDY, V.B., CAVALLESCO, C., GHOSH, P.K. de RIEL, J.K., FORGET, B.G. and WEISSMAN, S.M. (1980) Nucleotide sequence of the coding portion of human α-globin messenger RNA. *J. Biol. Chem.*, 255: 2807-15.

WU, M. and DAVIDSON, N. (1981) Transmission electron microscopic method for gene mapping on polytene chromosomes by *in situ* hybridization. *Proc. Natl. Acad. Sci. USA*, 78: 7059-63.

YOUNG, B.D., FERGUSON-SMITH, M.A., SILLAR, R. and BOYD, E. (1981) High resolution analysis of human peripheral lymphocyte chromosomes by flow cytometry. *Proc. Natl. Acad. Sci. USA*, 78: 7727-31.

YOUNG, G.J., MARSHALL GRAVES, J.A., BARBIERI, I., WOOLLEY, P.A., COOPER, D.W. and WESTERMAN, M. (1982) The chromosomes of Dasyurids (Marsupialia). In 'Carnivorous Marsupials', ed. M. Archer. Roy. Zool. Sco. N.S.W.: Sydney.

ZAIN, A., SAMBROOK, J., ROBERTS, R.J., KELLER, FRIED, M. and DUNN, A.R. (1979) Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA. *Cell*, 16: 851-61.

ZIMMER, E.A., MARTIN, S.L., BEVERLY, S.M., KAN, Y.W. and WILSON, A.C. (1980) Rapid duplication and loss of genes coding for the α chains of haemoglobin. *Proc. Natl. Acad. Sci. USA*, 77: 2158-62.