A MOLECULAR AND EVOLUTIONARY STUDY OF THE β-GLOBIN GENE FAMILY OF *SMINTHOPSIS CRASSICAUDATA*

By

Steven J.B. Cooper B.Sc.(Hons.)

A thesis submitted in total fulfilment of the

requirements for the degree of

Doctor of Philosophy

Department of Genetics

University of Adelaide

January 1991

To my parents,
for their support, encouragement
and belief in the value of
education

TABLE OF CONTENTS

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

LIST OF FIGURES

List of figures (continued)

List of Figures (continued)

LIST OF TABLES

SUMMARY

This thesis contains the results of a molecular and evolutionary study of the β-globin gene family of the Australian dasyurid marsupial *Sminthopsis crassicaudata*. The principle aim of the project was to examine the question of how many progenitors of the five ancestral eutherian β-globin loci: ε, γ, η, δ and β; exist in marsupials, given that the times of their individual duplications coincided approximately with the time that the eutherian and marsupial lineages diverged from a common ancestor. It was also an aim of this project to investigate whether marsupials have β-like globin genes expressed specifically during embryonic and pouch young development.

A *S.crassicaudata* genomic DNA library in the bacteriophage EMBL3 was screened using a *Dasyurus viverrinus* cDNA β-globin clone (pDG-5.2) as a probe. A single phage clone, that cross-hybridized to the *D.viverrinus* β-globin probe, was isolated and partially characterized by restriction mapping. Hybridization to pDG-5.2 was localized by Southern analysis to a 3.75 kb BamHI / SalI fragment, which was sub-cloned into pBR322. A detailed restriction map and Southern analysis of this subclone (pSG-2) further localized the hybridization to a 460 bp HindIII fragment. Known restriction sites in the pSG-2 insert were used to generate fragments of 300 to 400 bp containing regions of the 460 bp HindIII fragment and flanking sequences. These fragments were cloned into the appropriate restriction site of the M13 bacteriophage vectors Mp18 and Mp19 and were sequenced using the "di-deoxy" chain termination procedure.

Sequence comparisons with mouse adult and embryonic β-globin genes revealed that the subclone pSG-2 contained the 5' region and 1st and 2nd exons of a β-like globin gene. The 3rd exon and 3' flanking sequences of the gene were located in a 2nd subclone (pSG-12) which is contiguous with pSG-2. The entire gene was sequenced on both strands and an extra 2.5 kb of flanking sequence 5' to the CAP site and 3' to the end of the 3rd exon of the gene was also determined from a single strand.

Comparison of the DNA sequence of the gene with eutherian and *Didelphis virginiana* adult and embryonic β-like globin genes revealed highly conserved promoter signals, intron/exon splice sites and 3' poly-adenylation signals. The three exons of the gene are open reading frames and conceptually translate into a polypeptide consisting of 146 amino acid residues. The length of the 2nd intron (1438 bp) of S.c-β is much longer than the 2nd intron of all known eutherian β-globin genes, but is similar in length to the 1465 bp 2nd intron of the *D. virginiana* adult β-globin gene (Koop and Goodman, 1988). cf Wu and Li (1985)

Non-synonymous and synonymous divergence values between the *S. crassicaudata* β-like globin gene (S.c-β) and eutherian and *D. virginiana* β-like globin genes were calculated using the procedure of Li and Wu (1985) and Li 40, and reveal that S.c-β most closely resembles the *D. virginiana* and eutherian ϵ-globin genes. Its coding sequence differs at non-synonymous sites from *D. virginiana* ϵ, *D. virginiana* β, human ϵ, human β and *D. viverrinus* β (partial sequence) by 11%, 17%, 9%, 14% and 11% respectively. These data suggest that S.c-β is orthologous to the

D. virginiana and eutherian non-adult β-globin genes. Additional evidence for this comes from the presence of a phenylalanine residue at the 3rd position of the conceptually translated polypeptide chain of S.c-β. This is a synapomorph of eutherian non-adult β-globin polypeptides and is also present in the *D. virginiana* putative ε-globin polypeptide.

To investigate the possibility that S.c-β was expressed during embryonic development, RNA was isolated from an embryo, pouch young of between 1 and 10 days of age *post partum* and adult bone marrow and Northern analysed using a region of S.c-β (1st intron and 2nd exon) as a probe. The results showed that the probe hybridized more strongly to embryonic mRNA than to pouch young an adult bone marrow RNA, providing evidence that S.c-β is expressed during embryonic development. In addition, the results suggest that S.c-β is either switched off or expressed at low level within a day or two of birth.

To provide further evidence that S.c-β was not the major adult β-globin gene of *S. crassicaudata*, α- and β-globin polypeptides were purified by ion-exchange column chromatography and a partial amino-acid sequence, consisting of 60 amino-acids from the N-terminal, was determined for each polypeptide. The partial sequence of the adult β-globin polypeptide was compared to the 1st 60 amino-acids of the conceptually translated sequence of S.c-β. The two sequences differ at 20 amino-acid sites out of 60. This result provides evidence that S.c-β is not the major adult β-globin gene of *S. crassicaudata*.

This thesis also investigates the number and organization of β-globin genes in

S.crassicaudata using two approaches. The first approach, involved an attempt to chromosome walk from S.c-β and isolate the adult β-globin gene of *S.crassicaudata*. DNA sequences 5' to S.c-β and 12 kb 3' to S.c-β were used as probes to re-screen the genomic DNA library and isolate overlapping phage clones. The *D.viverrinus* β-globin clone (pDG-5.2) and a region of S.c-β, containing the 1st intron and 2nd exon, were also used as probes. Three phage clones were isolated and partially characterized by restriction mapping. Two of the phage clones contained identical inserts spanning a 15 kb region of DNA located 5' to S.c-β. The other phage clone extended from the 2nd intron of S.c-β a further 5.7 kb from the end of the λSG-3 insert in a direction 3' to S.c-β. Southern analysis of these phage clones failed to show the presence of any new β-like globin loci.

The second approach that was used, to investigate the number and organization of β-globin genes, involved an extensive Southern blot analysis of *S.crassicaudata* DNA. Three restriction fragment length variants at the S.c-β locus and two variants at another two β-like globin loci were detected, each of which cross-hybridized to S.c-β. One of these loci, which hybridized strongly to the *D.viverrinus* β-globin clone (pDG-5.2), was found by family studies to be closely linked to S.c-β. The results of the Southern analysis provide evidence that this locus is the putative adult β-globin gene of *S.crassicaudata*. The 3rd locus failed to hybridize to pDG-5.2 suggesting that it is likely to be a pseudogene derived by duplication of an ancestral ϵ-globin gene. Although family data on this locus were limited, they did indicate that it was not closely linked to either the putative adult β-globin gene or S.c-β.

Analysis of PstI restriction fragment length variants provided evidence that S.c-β and the putative adult β-globin gene are located on a 30 kb Pst I fragment. This indicates that the "chromosome walking" had proceeded to within 10 kb or less of the putative adult β-globin gene. Overall the results suggest that at least three β-like globin loci exist in *S.crassicaudata* with strong evidence for a 30 kb gene cluster consisting of one functional embryonic β-globin gene (S.c-β) and one functional adult β-globin gene. A 3rd β-like globin gene, possibly a pseudogene, is unlinked to the functional β-globin gene cluster. The results therefore lend support to the hypothesis proposed by Koop and Goodman (1988) that two progenitors of the five linked eutherian β-globin loci: ϵ, γ, η, δ and β; existed at the time the marsupial and eutherian lineages separated from a common ancestor and that "these 2 genes were already differentiated with respect to their promoter regions and developmental expression" (Koop and Goodman, 1988).

A comparison of the rates of nucleotide substitution of eutherian ϵ-globin genes is also presented in this thesis. "Relative rate tests" were carried out on 5 eutherian ϵ-globin genes using S.c-β as an outgroup sequence. The results showed that the mouse ϵ-globin gene is the fastest evolving gene and the goat ϵ-globin gene is the slowest evolving gene at both synonymous and non-synonymous sites. The human and rabbit ϵ-globin genes appear to have evolved at similar rates. These results suggest that rates of evolution have not been constant during the evolution of eutherian ϵ-globin genes, but do not provide definitive evidence to support the proposal that rates of gene evolution are influenced by generation time.

This thesis also examines the evolutionary relationships of monotremes, marsupials and eutherians using phylogenetic trees constructed for adult α- and β-globin amino acid sequence data. To provide further data for this analysis, the *S.crassicaudata* partial adult β-globin amino-acid sequence was combined with the *D.viverrinus* partial β-globin sequence, which was missing the 1st 60 amino-acid sites, giving a full length polypeptide. This sequence was assumed to be representative of β-globin sequences of the Order Dasyuromorphia. Phylogenetic trees were constructed for α-globin and β-globin polypeptides using the "Protpars" computer program of Phylip (version 3.2, made available by J. Felsenstein, Washington University). This program infers the most parsimonious unrooted phylogenetic trees directly from protein sequences.

Using β-globin data, from 5 marsupial, 2 monotreme and 4 eutherians, a single consensus tree was obtained which depicted marsupial sequence divergence as ancestral to the monotreme and eutherian sequences. This branching order is in contra-distinction to the taxonomic position of marsupials which places them closer to eutherians than to egg-laying monotremes (Stonehouse, 1977). Using α-globin data, from 4 marsupial, 5 eutherian and 2 monotremes, two consensus trees were obtained each of which depicted monotreme sequence divergence as ancestral to the marsupial and eutherian sequences. The implications of these discrepant results are further discussed in this thesis.

DECLARATION

This thesis contains no material published elsewhere, except where due reference is made in the text. No material from this thesis has been presented for any degree or diploma. No other person's work has been used without due acknowledgement.

Signed

Date:

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Rory Hope for his enthusiasm and encouragement throughout this project, and for his valuable assistance under difficult circumstances in the preparation of this thesis.

I am grateful to the following people who helped me in different aspects of the project: Prof. J.H. Bennett and Mrs D. Golding for provision of animals from the *S.crassicaudata* colony, Clive Chesson for his assistance with protein extraction and electrophoresis procedures, Sharon Hutchins for preparing *S.crassicaudata* genomic DNA samples, Prof. P.G. Martin and Julie Dowd for their assistance with protein sequencing procedures, Dr. A.V. Sivaprasad for his assistance with DNA sequencing procedures, Dr. J. Felsenstein for provision of the "Phylip" computer programs for analysis of protein data, Dr.'s W-H. Li, C-I. Wu and C-C. Luo for providing a computer program to calculate divergence values, and Dr. Matthew O'Connell for assistance with RNA procedures.

For their helpful advice and discussion I am also grateful, to Dr. J.A.M. Graves, Mr Clinton Hale, Dr. D.L. Hayman, Dr. J. Kelly, Dr. L. Selwood, Dr. J.N. Timmis, Dr. M. Westerman, Dr. A. Wilton, and Dr. P. Woolley.

For assistance in the preparation of this manuscript I would like to thank Denise O'Keefe and Mrs Helen Skene for typing sections of the manuscript, Jo Richter, Mrs Yvonne La Rose and Doug Pottrell for their assistance with photography, and Tania Bezzobs and Dr. Mike Westerman for proof-reading and corrections of the manuscript.

I would also like to thank my fellow post grad students (especially Jo) at Adelaide University, and the "BCG club" (Denise, Neil and Colin) at La Trobe University for providing an enjoyable working (and social) environment.

Finally, I would like to thank Tania Bezzobs for her friendship, support and encouragement, and for sharing my life over the last four years of my Ph.D.

List of Abbreviations

amp^R	ampicillin resistant
bp	base pair
BSA	Bovine Serum Albumin
cDNA	synthesized DNA copy of mRNA
Ci	Curie
cpm	counts per minute
DNA	Deoxyribonucleic Acid
ddNTP	dideoxynucleoside triphosphate
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylene diamine tetra acetic acid
ETBR	ethidium bromide
HPLC	high pressure liquid chromatography
kb	kilobase
LMT	low melting temperature agarose
my	million years
mya	million years ago
mRNA	messenger RNA
NaOAc	sodium acetate
OD_x	optical density at x nm
PEG	polyethylene glycol
pfu	plaque forming units
RNA	Ribonucleic acid
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate
tet_s	tetracycline sensitive
TLC	thin layer chromatography
Tris	Tris(hydroxymethyl amino ethane)
v/v	volume per volume
w/v	weight per volume
λ	lambda bacteriophage

CHAPTER 1

INTRODUCTION

"It is only through the application of a variety of scientific approaches to a variety of organisms that we can come to terms with the biological world, its evolution and its richness."

(J.A.M. Graves, R.M. Hope and D.W. Cooper, 1990)

In higher vertebrates, the oxygen carrying molecule haemoglobin is composed of two α- and two β-globin polypeptide chains, with each type of chain being encoded by separate α- and β-globin genes. During the evolution of higher vertebrates, a series of tandem gene duplications and adaptations of the ancestral α- and β-globin genes, resulted in the formation of α-like and β-like globin gene clusters each with at least two or more differentially expressed member genes. The advent of recombinant DNA technology in the mid 1970's enabled these genes to be studied in detail at the molecular level and also provided an opportunity to investigate their evolutionary history by comparative DNA studies.

β-globin gene clusters have been extensively studied, at the molecular level, in species from four orders of eutherian mammals: Lagomorpha, Artiodactyla, Primates and Rodentia. These studies have revealed a remarkable diversity in gene number and organization in each cluster, both within and between orders (see

Section 2.5). DNA sequence comparisons involving eutherian β-like globin genes and comparisons of the chromosomal arrangement and expression of these genes have provided evidence that an ancestral β-globin cluster existed prior to the eutherian radiation (Goodman *et al.*, 1984; Hardison, 1984). These studies further revealed that the progenitors of eutherian β-like globin genes arose by a series of tandem duplications, which probably occurred after the separation of the avian and mammalian lineages. The first duplication of a single ancestral β-globin locus is estimated to have occurred between 155 and 200 million years ago (mya) and resulted in the production of two β-like globin genes, one of which became developmentally delayed in its expression, and the other became expressed in embryonic tissues only. Goodman *et al.* (1984) suggested that prior to the eutherian radiation the embryonic gene duplicated a further two times and the adult gene duplicated once, producing a β-globin cluster consisting of three embryonic (ϵ, γ and η) and two adult expressed genes (δ and β) linked in the order: 5'-ϵ-γ-η-δ-β-3'. These later duplications are estimated to have occurred between 90 and 140 mya, a period which coincides with the estimated divergence time of the marsupial and eutherian lineages (100 - 150 mya, Air *et al.*, 1971).

It is therefore of considerable evolutionary interest and a principle objective of the research presented in this thesis, to study the β-globin gene family of a marsupial and determine how many of the progenitors of the five eutherian ancestral loci (ϵ, γ, η, δ and β) may have existed in a common ancestor of eutherians and marsupials. It was also an aim of this project to investigate whether

marsupials have β-like globin genes which are differentially expressed during development and adapted to the unique oxygen transport requirements of pouch young life. Sequence comparisons involving β-like globin genes of eutherians and marsupials may also aid in the identification of conserved sequences which may be important in the regulation of β-globin gene expression.

Prior to the start of this Ph.D project, there were no published reports of the number and organization of β-like globin genes in a marsupial. In 1988, however, Koop and Goodman published a paper reporting the isolation and characterization of two β-like globin genes from the marsupial *Didelphis virginiana* (Koop and Goodman, 1988: see Section 2.6). The coding sequence of one of these genes (β^m) conceptually translated into the known sequence of the adult β-globin polypeptide. The second gene (ϵ^m) was more similar in coding and promoter sequence to embryonic β-like globin genes of eutherians than to the adult expressed genes. No further β-like globin genes could be isolated from a *D.virginiana* genomic DNA library or detected by Southern analysis of genomic DNA from this species. Koop and Goodman therefore concluded that just "2 progenitors of the 5 linked genes (ϵ, γ, η, δ and β) existed by the time of the Eutherian (placental mammal) - Metatherian (marsupial mammal) split and that these 2 genes were already differentiated with respect to their promoter regions and developmental expression". Their study, however, was unable to confirm the expected 5'-ϵ-β-3' linkage arrangement of the two β-like globin genes and did not prove conclusively that the ϵ^m-globin gene was in fact expressed in the embryo. Thus Koop and

Goodman's proposal that ϵ^m was already differentiated with respect to its developmental expression is speculative. It is possible that ϵ^m is a minor adult β-globin gene and fortuitously or by a slow down in the rate of evolution of its nucleotide sequence, it is similar in DNA sequence to eutherian ϵ-globin genes.

The evolution of the β-globin cluster, particularly during the eutherian radiation, has been accompanied by major alterations in both gene number and expression. At least four evolutionary mechanisms have contributed to these changes:

i) Tandem duplications of one or more genes. This mechanism resulted in an increase in number of β-globin genes in artiodactyls from 4 to 9 in cattle, (Schimenti and Duncan, 1985), and 12 in goats (Townes *et al.*, 1984).

ii) Gene conversion by unequal crossing over resulting in gene loss or concerted evolution. This mechanism resulted in the 5' region of the human $^{A}\gamma$-globin gene on one chromosome (A) resembling more closely the corresponding region of the $^{G}\gamma$-globin gene on the same chromosome than its allele on a homologous chromosome (Scott *et al.*, 1984). Gene conversion in Lemuroidea, also resulted in the fusion of the 5' region of $\psi\eta$ with the 3' region of the δ-globin gene, giving a 4 gene cluster (Jeffreys *et al.*, 1982).

iii) Recruitment of genes into different developmental regimes. This mechanism resulted in the γ-globin genes of simian primates (Ceboidea, Cercopithecoidea and Hominoidea) becoming developmentally delayed in their expression, from embryonic to foetal life (Goodman *et al.*, 1987; Tagle *et al.*, 1988).

iv) Inactivation of genes by the accumulation of DNA mutations to produce pseudogenes. This mechanism resulted in the silencing of the η-globin gene of primates and the δ-globin genes of lagomorphs, rodents and artiodactyls (Hardies *et al.*, 1984).

Thus, during the evolution of eutherians there have been major changes in gene number and expression both within and between orders. It is therefore possible that more than two progenitors of the five ancestral eutherian β-like globin loci (ϵ, γ, η, δ and β) existed prior to the divergence of the marsupial and eutherian lineages, but by gene conversion and inactivation, only two genes remained functional in the lineage leading to *D.virginiana*. A recent re-classification of marsupials by Aplin and Archer (1987) suggests that marsupials can be grouped into at least seven extant orders and separate American and Australian cohorts. This classification reflects the level of diversity which exists in the marsupial infra-class and also highlights the obvious fact that a single American marsupial, *D.virginiana*, can not be considered a representative of all marsupials. It is therefore important to study a β-like globin gene family in an Australian marsupial to clarify the conclusions reached by Koop and Goodman (1988). Hence, the species chosen for study in this project was the Australian dasyurid marsupial *Sminthopsis crassicaudata*. The immediate aims of this project were to:

i) Isolate and characterize at the molecular level the β-globin gene family of *S.crassicaudata*.

ii) Verify that marsupials have β-like globin genes that are expressed during

embryonic development and investigate whether marsupials also have β-like globin genes specifically adapted to the oxygen transport requirements of pouch young.

iii) Determine whether the β-like globin genes of *S.crassicaudata* are clustered, ie closely linked.

The most widely used procedure for investigating the number and organization of β-globin gene families has involved the isolation of genomic clones of β-like globin genes from bacteriophage or cosmid DNA libraries. By constructing these libraries from partially digested genomic DNA fragments, the complete β-globin gene cluster can be isolated in a series of overlapping phage or cosmid clones. Such a library was constructed in the bacteriophage EMBL3 from partial MboI digests of *S.crassicaudata* DNA (Cooper, 1985). Genomic DNA clones containing regions of the β-globin gene cluster in *S.crassicaudata* were isolated by screening this library with a cDNA β-globin clone from the dasyurid marsupial *Dasyurus viverrinus* (Wainwright and Hope, 1985). The number and organization of β-like globin genes in *S.crassicaudata* were also investigated by a detailed Southern analysis of DNA samples from randomly chosen animals and family DNA samples. To aid in the identification of cloned β-like globin genes, amino acid sequence data were obtained from purified α- and β-globin chains isolated from adult *S.crassicaudata*. In addition, the timing of expression of a cloned β-like globin gene was investigated by Northern analysis of adult, pouch young and embryonic RNA. These studies were greatly facilitated by the availability of a large breeding colony of *S.crassicaudata* (Bennett *et al.*, 1990).

CHAPTER 2

LITERATURE REVIEW.

2.1 The genetics and evolution of marsupials.

Class Mammalia consists of three major extant groups or infraclasses: monotremes, marsupials and eutherians. The marsupial and eutherian infraclasses have been classified into a single subclass "Theria" and monotremes to a separate subclass "Prototheria" although there is conjecture as to whether this subdivision is justified (see Section 2.12). The marsupial infraclass comprises over 250 extant species arranged into 16 families, which includes 13 Australian and 3 American families (Kirsch and Calaby, 1977). Higher order classification of marsupials is still the subject of much debate, and where it has been necessary in this thesis to use a higher order classification, the scheme proposed by Aplin and Archer (1987) has been adopted. Hence Australian and American marsupials have been grouped into separate "Cohorts", Australdelphia and Ameridelphia with each consisting of six orders and five orders of marsupials respectively. Therefore the species studied in this thesis, *S. crassicaudata*, belongs to the family Dasyuridae of the Order Dasyuromorphia and Cohort Australidelphia.

Marsupials are most obviously distinguished from eutherians and monotremes by their mode of development. Eutherians nurture their young, within a uterus, to a

relatively mature stage of development while monotreme young develop within a large yolky egg. Marsupials are born after a relatively short gestation period and spend much of their early development attached to a teat, within a fold of skin called a pouch. The term "pouch young" is used in this thesis to refer to young marsupials at this period of development, while the term "embryo" refers to that period of development before birth.

Globin sequence comparisons suggest that marsupials and eutherians diverged between 100 and 150 mya during the mid or early Cretaceous period (Air *et al.*, 1971; Cooper *et al.*, 1977; Hope *et al.*, 1990). Thus approximately 200 to 300 my of independent evolution separate marsupials and eutherians, so it is perhaps not surprising that marsupials are consistently providing new variants of fundamental genetic processes which are found to be conserved in eutherians. For example, the process of gene dosage compensation by X-inactivation in eutherians is distinctly different from X-inactivation in marsupials. Inactivation of the X chromosome in eutherians is random, occurring equally as frequently on either the paternally or the maternally derived X chromosome. In marsupials, however, the paternally derived X chromosome is preferentially inactivated and inactivation of some loci on the chromosome is often incomplete (see review by Cooper *et al.*, 1990).

Another fundamental genetic process which is relatively conserved in eutherians but is variant in marsupials is the process of meiotic recombination. In eutherians, there are only small differences between the sexes in linkage values, and and where differences do occur, males generally show closer linkage between a given pair of

genes than females. In the marsupial *S.crassicaudata*, however, Bennett *et al.* (1986) reported significantly reduced autosomal recombination values in females, which was reflected by "extreme localization of chiasmata" on the autosomes during meiosis, a feature which was absent in males. Such extreme sexual disparity in chiasma location on an autosome has not previously been reported for any mammal. Cytological studies by Hayman *et al.* (1988) in the South American opossum *Monodelphis domestica* also revealed differences in chiasmata distribution between the sexes, similar to that found in *S.crassicaudata*. Therefore, differences in recombination values between the sexes may also be expected in this species. More recent studies, however, indicate that sexual disparity in chiasma location is not a general feature of all marsupials (D.L. Hayman, pers. comm.).

The presence of these and other variant genetic processes in marsupials, exemplifies the importance of studying evolutionarily distant species in order to gain an understanding of the organization, function and evolution of the mammalian genome.

2.2 The fat-tailed dunnart, *Sminthopsis crassicaudata.*

The availability of a pedigreed laboratory colony of the small insectivorous marsupial *S.crassicaudata* in the Department of Genetics, University of Adelaide, was the principle reason why this species was chosen for study. The colony was established from about 30 wild-caught animals and has been maintained as a self-

sustaining colony since 1975. Litters are produced uniformly throughout the year and their average offspring size, at weaning, is approximately 5.3 young, making this species particularly amenable to studies of Dasyurid reproductive biology and genetics (reviewed by Bennett *et al.*, 1990). There also appears to be an abundant source of genetic variation in the colony, with blood protein variation detected at 8 loci and restriction fragment length polymorphisms (RFLP's) recently detected at a further 3 loci (see Bennett *et al.*, 1986, 1990). With particular reference to this project, the colony provides a useful source of material for RNA expression studies of globin genes during development. However, the size of the animals (adult weight 14-16 g and young at birth about 16 mg) limits the use of this species for studies of haemoglobin ontogeny at the protein level.

2.3 The structure and function of haemoglobin.

The principle function of haemoglobin is to bind oxygen and transport it to the tissues of an organism where under the appropriate physiological conditions the oxygen is released. Higher vertebrates have evolved tetrameric haemoglobins to perform this function, which are composed of two β-globin and two α-globin polypeptide chains. Each chain is covalently bound to a single iron-porphyrin ring called haem. An oxygen molecule binds reversibly and non-covalently to an iron atom found in the centre of each haem. Vertebrates also contain a monomeric oxygen transport protein, myoglobin, which is found in muscle tissue, where it facilitates the

diffusion of oxygen from the vascular system to muscle mitochondria (Wittenberg, 1970).

Lower vertebrates such as the lamprey use haemoglobin-like molecules which are monomeric in structure, for oxygen transport (Love *et al.*, 1971). Significantly, the lamprey haemoglobin molecules aggregate when deoxygenized to form dimers and tetramers which dissociate when the polymer binds oxygen (Briehl, 1963). It is believed therefore that lamprey haemoglobin is an "intermediate step in the evolution from monomeric to tetrameric haemoglobin" (Li and Riggs, 1970). According to Monod *et al.*(1965) "The most decisive factor in the emergence and selective maintenance" of the tetrameric form of haemoglobin in higher vertebrates is the "inherent co-operativity of its structure". This co-operative behaviour of tetrameric haemoglobin is the ability of a single haemoglobin molecule to alter its quaternary structure and increase its affinity for oxygen as successive molecules of oxygen bind, and conversely to decrease its affinity for oxygen as successive molecules of oxygen are released. This is characterized by a sigmoidal oxygen saturation curve in contrast to the hyperbolic curve of myoglobin which is non co-operative in structure (Weissbluff, 1974). This behaviour of haemoglobin has important physiological consequences as it enables haemoglobin not only to bind oxygen but also to release oxygen readily in order that the tissues of an organism may be adequately supplied with oxygen.

The oxygen affinity of haemoglobin in higher vertebrates is regulated by H^+ and CO (the Bohr effect) and by the organic phosphate 2,3-diphosphoglycerate (DPG).

H^+ and CO reduce the oxygen affinity of haemoglobin and thereby enhance the release of oxygen in metabolically active tissue. DPG binds preferentially to deoxyhaemoglobin and also has the effect of reducing the oxygen affinity of haemoglobin. Embryonic haemoglobins of eutherian mammals bind less DPG than adult haemoglobins and therefore have a higher affinity for oxygen. This is physiologically important in the transfer of oxygen from mother to embryo *via* the placenta.

The three-dimensional structure of human and horse deoxy- and liganded haemoglobin have been determined by Perutz and co-workers using X-ray diffraction techniques, and have enabled amino acid residues with important functional roles to be identified (Perutz, 1970; Fermi, 1975; Ladner *et al.*, 1977). Five functional categories have been specified for individual amino acid residues: (1) residues involved in globin-haem interactions (these residues provide the necessary environment for oxygen uptake and dissociation) (2) residues in the $\alpha_1\beta_2$ contact (these residues are involved in co-operative alterations of the quaternary structure of haemoglobin), (3) residues in the $\alpha_1\beta_1$ contact, (4) residues of the β-globin chain that bind DPG, (5) residues involved in the Bohr effect (see Fermi (1975) for a list of the functional residues). The haem contact, $\alpha_1\beta_2$ contact and $\alpha_1\beta_1$ contact residues were defined by Perutz as being those amino acid sites containing at least one atom within 4 Angstrom units of the haem or adjacent subunit atoms (Perutz *et al.*, 1968). Studies of human haemoglobin variants provides further evidence for the functional roles of specific amino acid residues (Perutz and Lehmann, 1968). Mutations of α- and β-globin

chains with known abnormal properties generally occur at functional residues (Eaton, 1980). In addition, comparisons of α- and β-globin amino acid sequences amongst higher vertebrates show significant conservation of functional amino acid residues particularly at $\alpha_1\beta_2$ and haem contact residues and at residues involved in the Bohr effect (Goodman *et al.*, 1975; Hope *et al.*, 1990).

2.4 The structure and evolution of the haemoglobin gene.

Haemoglobin-like oxygen carrying molecules are found in a wide range of organisms including vertebrates, invertebrates, plants and bacteria. The DNA sequences of over 100 globin genes have been determined and the amino acid sequences are now known for several hundred globin chains. Amino acid and nucleotide sequence comparisons provide strong evidence that haemoglobin genes existed in an ancestor common to modern plants and animals, and may perhaps have evolved before the divergence of prokaryotes and eukaryotes (Landsmann *et al.* 1986, Wakabayashi *et al.* 1986, Goodman *et al.* 1987, 1988).

All functional vertebrate globin genes studied to date share a 3 exon / 2 intron structure with the 2 introns being located at identical positions in each gene. Although DNA sequence data from invertebrate globin genes are limited, a *Lumbricus terrestris* globin gene is also found to have the 3 exon / 2 intron structure characteristic of vertebrates, while surprisingly the insect *Chironomus thumni* globin gene contains no introns (Jhiang *et al.* 1988, Antoine and Niessing, 1984). Globin genes in legume

plants and non-legumes such as *Trema tomentosa* and *Parasponia andersonii* have an alternative gene structure consisting of 4 exons and 3 introns (Bogusz *et al.* 1988, Jensen *et al.*, 1981, Lee *et al.* 1983). The 1st and 3rd introns of the gene, however, are located "at positions identical to the introns in vertebrate globins" (Landsmann *et al.* 1986). Gō (1981) identified 4 structural units (F1, F2, F3 and F4) in vertebrate α- and β-globin chains, of which F1 and F4 were encoded by the 1st and 3rd exons respectively, while F2 and F3 were encoded by the 2nd exon of the globin gene. Gō postulated that an ancestral globin gene may have contained an intron between the exon "blocks" encoding F2 and F3, which was subsequently lost prior to the evolution of vertebrates. The central intron of plant globin genes is almost precisely located at the position predicted by Gō, providing support for the proposal by Gilbert (1978) and Blake (1978) that genes could evolve by the shuffling together and reassortment of exons encoding functional units or domains. However, there is still conjecture that the haemoglobin gene evolved in this manner, as alternate approaches for detecting functional domains have failed to reveal exon related domains in haemoglobin (Rashin, 1981).

The conservation of intron position in plant and animal haemoglobin genes lends support to the idea of a mono-phyletic origin of all plant and animal haemoglobin genes, with an ancestral gene consisting of a 4 exon, 3 intron structure. An alternative explanation was proposed by Hyldig-Nielson *et al.*(1982) in which horizontal transfer of globin genes from animals to plants was responsible for the evolution of globin genes in plants. Recent data from Bogusz *et al.* (1988), which

indicate that globin genes may be widespread in the plant kingdom, suggest this alternative is unlikely. Runnegar (1984) proposed that a "primitive ancestral globin gene contained 3 exons, the first two of which were derived from a cytochrome-like DNA sequence". Further studies of the exon / intron organization of haemoglobin genes in plants and invertebrates, particularly lower invertebrates may help to shed some light on this issue.

It has been proposed that myoglobin, α- and β-globin genes of higher vertebrates arose by successive duplications of a primordial globin gene between 500 and 800 mya. The myoglobin/ haemoglobin duplication (600 - 800 mya) preceded the divergence of α- and β-globin genes (450 - 500 mya) although the timings of these duplications are still the subject of debate (Hunt *et al.*, 1978; Kimura, 1981; Czelusniak *et al.*, 1982). The myoglobin gene contains two introns at identical positions to α- and β-globin genes but the lengths of these introns are significantly greater in size (4.8 kb and 3.4 kb in seal myoglobin introns 1 and 2 respectively) (Blanchetot *et al.*, 1983; Weller *et al.*, 1984). Amongst the α- and β-globin genes of mammals, the 1st intron is generally smaller than the 2nd intron and shows little variation in length, ranging in size from 108 bp to 132 bp (Blanchetot *et al.*, 1983). Two exceptions are the mouse βh0-globin gene and the human embryonic α-globin gene 1st introns which are 336 bp and 886 bp in length respectively (Hill *et al.*, 1984; Proudfoot *et al.*, 1982). In contrast, the 2nd intron is much more variable in length, ranging in size from between 103 bp in an α-globin gene of the goat (*Capra hircus*) to up to 1857 bp in the ϵ-globin gene of the marsupial *Didelphis virginiana* (Schon *et*

al., 1981; Blanchetot *et al.*, 1983; Koop and Goodman, 1988). It has not yet been determined whether this apparent conservation of length in the 1st intron has any functional significance, or whether it just reflects a low frequency of deletion / insertion events which have had insufficient time to substantially alter the lengths of these regions.

2.5 The evolution of α- and β-globin gene clusters.

During early vertebrate evolution, the tandemly duplicated proto α- and proto β-globin genes mutually adapted to enable tetramers to form from two α-like and two β-like globin polypeptides. Goodman *et al.* (1975, 1987) have proposed that during this period "the vertebrates were evolving into larger more active animals and thus needed an allosteric hemoglobin that could efficiently deliver oxygen to respiring tissues" and further that the rates of globin evolution "were much faster in the early jawed vertebrates and tetrapods than in the amniotes descending to mammals and birds". They deduced that this increase in the rate of globin evolution was the result of positive natural selection acting at prospective $\alpha_1\beta_1$ and $\alpha_1\beta_2$ contact sites as well as sites involved in the Bohr effect. These proposals, however, are still the subject of much debate. Kimura (1981) was particularly critical of both the phylogenetic procedures used by Goodman and his colleagues and of their reliance on dubious fossil derived divergence times, for example, the time of divergence of vertebrate myoglobin and haemoglobin.

During higher vertebrate evolution the ancestral α- and β-globin genes underwent further tandem duplications and adaptations to produce families of closely linked α-like and β-like globin genes. While the level of expression of α- and β-like globin genes was coordinated, each producing equal quantities of α- and β-globin chains, the individual members of each family showed differential expression during development.

For example, the amphibian *Xenopus laevis* globin gene family is comprised of at least 12 genes, of which 6 encode α-globin and 6 encode β-globin polypeptides (Hosbach *et al.*, 1983; Jeffreys *et al.*, 1980). The twelve genes are arranged in two almost identical clusters, designated I and II, with six genes in each cluster. Each cluster contains two adult expressed genes (α^A and β^A) and four genes expressed in larval tissue only (α_a^L, α_b^L, β_a^L and β_b^L) arranged in the order: 5'-α_a^L-α_b^L-α^A-β^A-β_a^L-β_b^L-3' (Jeffreys *et al.*, 1980; Hosbach *et al.*, 1983; Patient *et al.*, 1980). The amphibian *Xenopus tropicalis* which last shared a common ancestor with *X.laevis* approximately 110 - 120 mya, contains half the DNA content and chromosome number of *X.laevis* and only one α-β cluster (Knochel *et al.*, 1986; Jeffreys *et al.*, 1980). Jeffreys *et al.* (1980) proposed that the extra globin gene cluster in *X.laevis* is the result of chromosome duplication by tetraploidization of a *X.tropicalis*-like ancestor, which is estimated to have occurred between 40 and 60 mya (Knochel *et al.*, 1986). The presence of linked α- and β-globin genes in amphibians provides strong evidence that "vertebrate α- and β-globin genes evolved by tandem duplication of a single primordial globin gene" (Jeffreys *et al.*, 1980).

While the α-like and β-like gene families remained linked in the amphibians, they are located on separate chromosomes in avians and marsupial and eutherian mammals (Deisseroth *et al.*, 1977,1978; Hughes *et al.*, 1979; Wainwright and Hope, 1985).

The α-globin and β-globin gene families of humans (*Homo sapians*), for example, are located on chromosomes 16 and 11 respectively (Deisseroth *et al.*, 1977, 1978). The α-globin gene family of humans consists of four functional genes (ζ, $\alpha 1$, $\alpha 2$ and θ) and three non-functional pseudogenes ($\psi\zeta$, $\psi\alpha 2$ and $\psi\alpha 1$) which are tightly clustered together and arranged in the order: 5'-ζ-$\psi\zeta$-$\psi\alpha 2$-$\psi\alpha 1$-$\alpha 2$-$\alpha 1$-θ-3' (Orkin, 1978; Lauer *et al.*, 1980; Proudfoot and Maniatis, 1980). The ζ-globin gene is expressed in nucleated red cells formed in the placental yolk sac but is inactive following the onset of erythropoiesis in the liver. In contrast, the $\alpha 1$- and $\alpha 2$-globin genes are only partially active in the yolk sac and become fully expressed in the liver and bone marrow (Weatherall and Clegg, 1979; Lauer *et al.*, 1980). The recently identified $\theta 1$-globin gene of humans is thought to be functional, based on the observation that its overall gene structure and nucleotide sequence has been conserved during primate evolution, and on the observation that it is transcribed in cells of erythroid origin (Marks *et al.*, 1986; Shaw *et al.*, 1987; Leung *et al.*, 1987; Hsu *et al.*, 1988). However, the physiological role of this gene in these cells is unknown.

The β-globin gene family of humans consists of five functional genes (ϵ, $^{A}\gamma$, $^{G}\gamma$, δ and β) and one pseudogene ($\psi\eta$) arranged in the order: 5'-ϵ-$^{A}\gamma$-$^{G}\gamma$-$\psi\eta$-δ-β-3' over a region spanning approximately 60 kb of DNA (Fritsch *et al.*, 1980; Efstratiadis *et*

al., 1980). The 5'-3' order of the genes is analagous to the order of their expression during development. The embryonic ϵ-globin gene is expressed in nucleated red cells formed in the placental yolk sac and is fully active at 5 weeks after conception. This gene, however, becomes progressively suppressed at 6-7 weeks while the $^{A}\gamma$ and $^{G}\gamma$-globin genes become active. Following the onset of erythropoiesis in the liver, foetal macrocytes are formed and γ-globin becomes the predominant β-like globin. By 7 - 8 weeks, the ϵ-globin gene is inactive, the γ-globin genes are fully active and the β-globin gene becomes partially expressed. There is a steady increase in β-chain synthesis and a corresponding steady decrease in γ-chain synthesis throughout gestation. The major switch from γ to β-chain synthesis occurs approximately at the time of birth, 36 weeks after conception. Synthesis of the minor adult globin δ is also activated at this stage (for a review see Weatherall and Clegg, 1979).

The chicken globin gene family comprises four β-like globin genes (ϵ, ρ, β^{H} and β) and three α-like globin genes (π^{1}, α^{A} and α^{D}) with each family also arranged in separate clusters. The genes ϵ, ρ and π^{1} are each expressed in embryonic or primitive red blood cells of 2 - 5 day old chickens, β^{H} and β are expressed in "definitive cells" which appear at 6 - 7 days and α^{A} and α^{D} are expressed in both cell lines (Bruns and Ingram, 1973). The β-globin genes are arranged in the order: 5'-ρ-β^{H}-β-ϵ-3' which is unique amongst higher vertebrates in having the embryonic ϵ-globin gene located on the 3' side of the cluster (Dolan *et al.*, 1981). Roninson and Ingram (1982) compared the DNA sequences of the ρ- and ϵ-globin genes and deduced that each arose by independent duplication of the adult β-globin gene, and not by the tandem

duplication of an embryonic-like globin gene. This could therefore explain why these genes are located at opposite ends of the cluster. The chicken α-globin genes are arranged in the order: 5'-π^1-α^A-α^D-3' which is similar to the arrangement of embryonic and adult α-globin genes in humans (Engel and Dodgson, 1980; Dodgson *et al.*, 1981). Proudfoot *et al.* (1982) calculated the amount of pairwise sequence divergence between the chicken π^1, chicken α^A, human ζ and human α1-globin genes, and found that the chicken π^1 / chicken α^A, human ζ / human α1, chicken α^A / human ζ, human α1 / chicken π^1 divergences were almost identical (approximately 38%). In addition they found that the chicken π^1 / human ζ and chicken α^A / human α1 divergences were also identical. These results suggested that the chicken and human α-globin genes were monophyletic in origin. Independent studies by Czelusniak *et al.* (1982) using amino acid sequence data and a geneological reconstruction method based on the maximum parsimony procedure, came to a similar conclusion. Both studies showed that the α-globin gene tandemly duplicated approximately 400 mya producing two gene lines, one of which became the progenitor of adult α-globin genes and the other the progenitor of embryonic α-globin genes of avians and mammals.

In contrast it was shown that independent β-like duplications and convergent evolution were responsible for the embryonic ρ locus of avians and the embryonic ϵ loci of mammals (Czelusniak *et al.*, 1982; Proudfoot *et al.*, 1982). The progenitors of the embryonic β-like globin genes and the adult β-globin genes of eutherians, thus arose by the tandem duplication of an ancestral β-like globin gene between 155 and 200 mya, in a period after the separation of the avian and mammalian lineages

(Efstratiadis *et al.*, 1980, Czelusniak *et al.*, 1982, Goodman, 1981; Proudfoot *et al.*, 1982). Similarly the ρ and ϵ-globin loci of the chicken arose by independent tandem duplications of the ancestral β-globin gene in the avian lineage, after the separation of avians and mammals from their common ancestor (Roninson and Ingram, 1982; Czelusniak *et al.*, 1982).

In summary, the evolutionary histories of the α and β-globin gene families have been quite distinct, yet remarkably, the temporal and quantitative expression of α- and β-globin genes have remained coordinately controlled during development (Proudfoot *et al.*, 1982).

2.6 The evolution of eutherian β-globin gene clusters.

Eutherian β-globin gene families generally consist of four or more member genes clustered together in a region of DNA spanning 25 to 120 kb (Fig. 2.1). The order and arrangement of β-globin gene clusters from representatives of four orders of eutherian mammals are shown in Fig. 2.1.

The rabbit (*Oryctolagus cuniculus*) β-globin gene cluster, consists of four loci of which two are expressed in embryonic erythrocytes and one is expressed in foetal and adult tissues (Fig. 2.1, Lacy *et al.*, 1979, Hardison *et al.*,1979, Rohrbaugh and Hardison, 1983). The fourth locus $\psi\beta 2$ is a pseudogene and is incapable of producing a functional globin (Lacy *et al.*, 1979).

The β-globin gene family of the mouse *Mus domesticus* has also been extensively

studied at the molecular level. Three haplotypes have been described at the β-globin locus of the laboratory mouse : Hbb^d, Hbb^s and Hbb^p. The Hbb^d mouse synthesizes two types of adult β-globin (β_{major} and β_{minor}) in differing amounts, the Hbb^s mouse synthesizes one type of adult β-globin and the Hbb^p mouse is similar to Hbb^d but synthesizes a variant form of β_{minor}. DNA and protein studies of these haplotypes reveal that all three possess two adult β-globin genes in a cluster of seven genes (Fig. 2.1; Gilman, 1976; Jahn *et al.*, 1980; Weaver *et al.*, 1981;). The functional role of βh0 (Fig. 2.1) has not been determined, although its general sequence organization suggests that it is unlikely to be a pseudogene (Hill *et al.*, 1984; Padgett *et al.*, 1987). The βh2 and βh3 genes (Fig. 2.1), however, have evolved rapidly in comparison to the other loci, and βh3 has accumulated mutations which would render its product inactive, suggesting that both are pseudogenes (Edgell *et al.*, 1983; Brown *et al.*, 1982). The two adult genes of the Hbb^s haplotype diverge in sequence by just 0.4% over a region bordered by the translational start and stop codons, compared with 16% divergence between Hbb^d β_{major} and β_{minor} (Erhart *et al.*, 1985; Perler *et al.*, 1980). This suggests that gene conversion event(s) have maintained the similarity of the Hbb^s adult β-globin genes (Erhart *et al.*, 1985).

The largest known β-globin gene cluster studied to date occurs in the goat (*Capra hircus*) and consists of twelve genes spanning 120 kb (Fig. 2.1). The ϵ genes are expressed in the embryo while β^F, β^C and β^A are expressed in foetal, pre-adult (after birth for approximately 3 months) and adult tissues respectively. Comparisons of the nucleotide sequences of the twelve genes suggest that the goat β-globin gene

cluster evolved by the triplication of a four gene ancestral cluster (5'-ϵ-ϵ-ψβ-β-3'), and also reveal that β^F and β^C were descended from an adult β-globin gene and then recruited into non-adult developmental regimes (Townes *et al.*, 1984). Additional evidence for these conclusions comes from a study of the cow (*Bos taurus*) β-globin gene cluster which consists of a duplication of the four gene ancestral cluster and includes a recently duplicated pseudogene (Schimenti and Duncan,1985). The β-globin gene cluster of a B haplotype sheep (*Ovis aries*) also consists of a duplicated four gene ancestral cluster (Garner and Lingrel, 1988).

Hardison (1984) compared the entire rabbit and human β-globin gene clusters, and assigned orthologous and paralagous relationships amongst the genes. His results suggested an ancestral β-globin gene cluster consisting of the four gene set 5'-ϵ-γ-δ-β-3' existed prior to the radiation of lagomorphs and primates. His conclusion, however, that this cluster "predated the mammalian radiation" was highly speculative, as the divergence time he estimated for ϵ- and γ-globin genes was between 110 and 130 mya, while it has been estimated that marsupials and monotremes diverged from eutherians between 110 and 150 mya (Air *et al.*, 1971). It is perhaps conceivable that Hardison (1984) did not include marsupials and monotremes in his definition of "mammal". Support for a four gene ancestral cluster was also provided in an independent study of the mouse β-globin gene cluster which was published in a series of two papers (Hardies *et al.*, 1984; Hill *et al.*, 1984).

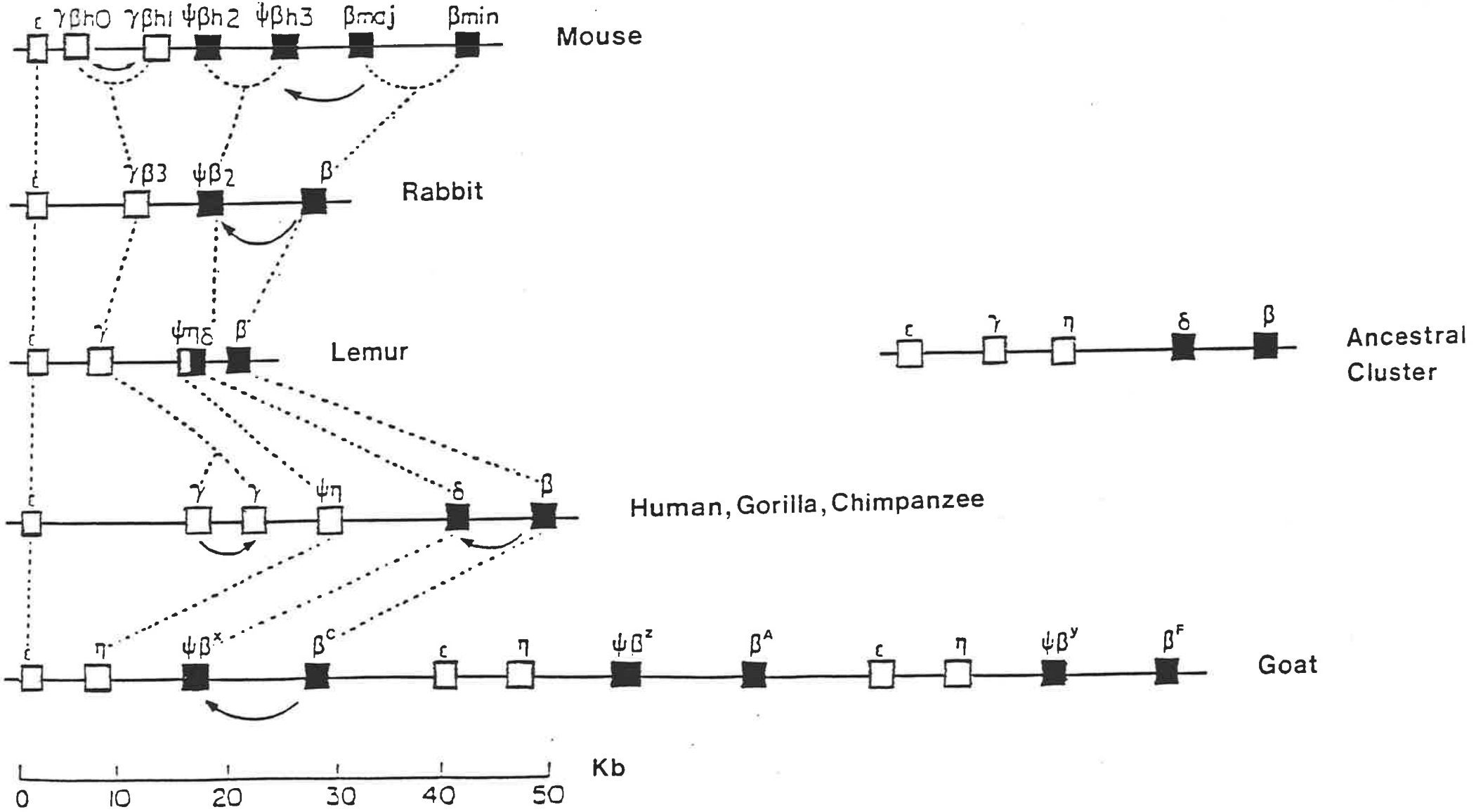
Phylogenetic reconstructions using the parsimony procedure by Goodman *et al.* (1984) on β-like globin gene families of the four eutherian orders Primates,

Lagomorpha, Artiodactyla and Rodentia, and independent studies by Harris *et al.* (1984), provided evidence that a five gene cluster (5'-ϵ-γ-η-δ-β-3') predated the eutherian radiation (Fig. 2.1). The three genes ϵ, γ and η were embryonically or foetally expressed genes and δ and β were expressed in adults. Thus Goodman *et al.* (1984) proposed that a series of four tandem duplications of β-globin genes occurred before the divergence of these four eutherian orders. The first of these duplications is thought to have occurred between 155 and 200 mya producing two gene loci, one of which was the progenitor of ϵ, γ and η-globin genes and the other the progenitor of δ and β-globin genes (Goodman *et al.*, 1984; Efstratiadis *et al.*, 1980; Hardison, 1984; Harris *et al.*,1984). The three duplications leading to the progenitors of ϵ, γ, η, δ and β-globin loci were estimated to have occurred between 90 and 140 mya (Goodman, 1981; Czelusniak *et al.*, 1982). Goodman *et al.* (1984) further proposed that during or just prior to the evolution of the Lagomorpha and Rodentia lineages, the η-globin locus was deleted leading to a four gene cluster for each of these 2 orders (Fig. 2.1). During the evolution of the Artiodactyla β-globin gene cluster the η-globin gene was retained and the γ-globin gene was deleted while Primates retained all five ancestral β-globin loci. The η-globin gene in Primates however became a non-functional pseudogene (Fig. 2.1; Jeffreys *et al.*, 1982; Harris *et al.*, 1984; Koop *et al.*, 1986, 1989).

Fig. 2.1

Organizations of loci in ß-globin gene clusters of eutherians.

This diagram was copied from Goodman *et al.* (1984) with minor alterations. The mouse ß-globin cluster is from Edgell *et al.* (1983); the rabbit from Hardison (1984); lemur from Barrie *et al.* (1981); human, chimpanzee and gorilla from Barrie *et al.* (1981) and the goat cluster from Shapiro *et al.* (1983). The ancestral eutherian cluster is that proposed by Goodman *et al.*(1984). Dotted lines connect orthologous genes in each cluster. Open boxes designate genes expressed during embryonic development, or as in the case of $\psi\eta$ of humans, descended from such a gene. Filled boxes designate loci expressed during adult life. Curved lines with arrowheads denote that gene conversions occurred between the two loci indicated.


ε γβh0 γβh1 ψβh2 ψβh3 βmaj βmin
Mouse
ε γβ3 ψβ2 β
Rabbit
ε γ ψη δ β
Lemur
ε γ γ ψη δ β
Human, Gorilla, Chimpanzee
ε γ η δ β
Ancestral Cluster
ε η ψβx βc ε η ψβz βA ε η ψβy βF
Goat
Kb
0 10 20 30 40 50

2.7 The evolution of β-globin gene clusters in marsupials.

Excluding the results of the study reported in this thesis, only two studies at the DNA level of a marsupial β-globin gene family have been described. Wainwright and Hope (1985) isolated and sequenced α- and β-globin cDNA clones from the dasyurid *Dasyurus viverrinus*. In addition, they used these cDNA clones to determine the chromosomal location of the α- and β-globin genes in *D.viverrinus*, by *in situ* hybridization to fixed metaphase chromosomes and by Southern analysis of DNA extracted from flow sorted chromosomes. Their results indicated that the α- and β-globin genes of *D.viverrinus* were asyntenic, thus confirming that asynteny of α- and β-globin genes predates the divergence of the marsupial and eutherian lineages.

Koop and Goodman (1988) isolated and sequenced two β-like globin genes from *Didelphis virginiana* (North American opossum). One of the genes was similar in its coding and promoter sequences to the eutherian non-adult β-like globin genes, while the other gene conceptually translated into the known sequence of the *D.virginiana* adult β-globin chain. Southern blot analysis of *D.virginiana* DNA provided evidence that just two functional β-like globin genes exist in this species. In addition, a "geneological reconstruction" based on the maximum parsimony procedure "joined opossum β^m to the root of a branch leading to eutherian β-genes and joins opossum ϵ^m to the root of a branch leading to eutherian ε, γ and η genes" (Koop and Goodman, 1988). Koop and Goodman therefore concluded that just "two progenitors of the five

linked genes" (5'-ϵ-γ-η-δ-β-3') "existed by the time of the eutherian (placental mammal) - metatherian (marsupial mammal) split and these two genes were already differentiated with respect to their promoter regions and developmental expression". It was also noted that the embryonically expressed ϵ-globin gene in *D.virginiana* and in eutherian mammals, in general, have evolved at a consistently slower rate than the adult β-globin gene. Koop and Goodman suggested that "less room may exist for variation of internal conditions in the embryonic stage of life than in later stages". The study by Koop and Goodman, however, was unable to confirm the expected 5'-ϵ-β-3' linkage arrangement of the 2 β-like globin genes and did not prove conclusively that the ϵ-globin gene was in fact expressed in the embryo.

2.8 Globin pseudogenes.

Pseudogenes are a common feature of multi-gene families and are frequently found within mammalian α- and β-globin gene families. In general, pseudogenes fall into one of two categories designated as either non-processed or processed.

Non-processed pseudogenes share the same DNA organization as a related functional gene, but contain multiple sequence defects which effectively silence the gene at either the transcriptional or the translational level. This type of pseudogene is thought to arise either from the tandem duplication and silencing of a functional gene or by the duplication of a pre-existing non-processed pseudogene (see Lacey and Maniatis, 1980). Non-processed pseudogenes are therefore usually found closely

linked to their functional counterparts, although an example of a dispersed non-processed pseudogene has been described in the α-globin gene family of the BALB/c mouse *Mus domesticus* (Leder *et al.*, 1981).

Processed pseudogenes resemble mRNA transcripts of functional genes in that they completely lack intervening sequences (introns), possess the remnants of poly(A) tails at their 3' ends and are often flanked by direct repeats characteristic of transposable elements (Hollis *et al.*, 1982). They are believed to arise by the reverse transcription of a spliced RNA transcript followed by integration of the cDNA copy into the genome at sites distinct from the location of the parent gene. Processed pseudogenes lack the promoter signals necessary for their expression and are therefore immediately non-functional following their integration into the genome. An exception to this is the human *Alu* family of repetitive DNA sequence elements which closely resemble reverse transcribed copies of the small cytoplasmic 7S and 4.5S RNA's, and are capable of transcription by RNA polymerase III (Schmid and Jelinek, 1982; Young *et al.*, 1982).

One example of a processed pseudogene in the globin gene family has been described in the BALB/c mouse *Mus domesticus*. The α-globin gene family of the mouse consists of at least two adult α-globin genes and one embryonic gene, as well as two closely related pseudogenes which are located on different chromosomes (Leder *et al.*, 1981). One of these pseudogenes (ψα3) has completely lost both introns in accordance with the GT / AT rule of RNA splicing (Vanin *et al.*, 1980; Nishioka *et al.*, 1980). This pseudogene, while showing up to 80% homology to the mouse adult

α-globin genes, has accumulated sufficient base changes to become functionally silenced. It does, however, contain sequences homologous to the adult α-globin gene promoter region, suggesting that an aberrant transcript of the gene acted as the intermediary for its reverse transcription and and integration into the genome. Recently, a new processed globin pseudogene has been described in the human α-globin gene family which is related to the θ1-globin gene and maps to the long arm of chromosome 22 (Shen *et al.*, 1989).

Non-processed globin pseudogenes have been described in four orders of eutherians: Lagomorpha, Rodentia, Primate and Artiodactyla, and are found closely linked to their functional counterparts (Edgell *et al.*, 1983; Hardison and Marogot, 1984; Barrie *et al.*, 1981; Shapiro *et al.*, 1983). DNA sequence comparisons of the β-globin pseudogenes of the rabbit (*Oryctolagus cuniculus*), mouse (*Mus domesticus*) and goat (*Capra hircus*) indicate that they are orthologous to the Primate δ-globin genes (Hardison, 1984; Hardies *et al.*, 1984). The δ-globin gene orthologues thus provide evidence that the progenitors of non-processed pseudogenes may be initially active, and that gene duplication precedes the silencing of the gene. After silencing, pseudogenes escape further selective pressures which act to preserve their sequence and evolve rapidly at rates greater than or equal to the synonymous codon substitution rate in functional genes or the minimal neutral rate of 5×10^{-9} nucleotide substitutions / site /year (Li *et al.*, 1981; Miyata and Hayashida, 1981).

Any function which pseudogenes may have is unknown. Their prevalence and conservation of position in eutherian β-globin gene clusters suggest they may play a

role (active or passive) in the maintenance of a linked set of regulated genes. Alternatively they may just act as "generators of diversity that could ultimately appear within active genes" through mechanisms such as gene conversion (Jeffreys, 1982). Hardies *et al.* (1984) accounted for their accumulation in the middle of β-globin gene clusters in the following model: "The cluster is surrounded on both sides by control elements that set up domains suitable for gene expression..... Duplications on either side of the cluster would leave genes in the middle in a region that falls too far away from its control elements. These genes would be destined to a future of poor expression and eventual inactivation". Recent studies which indicate that tissue-specific β-globin control elements do in fact lie 3' to the adult β-globin gene and 5' to the ε-globin gene (see next section) provide some support for this model.

2.9 The control of β-globin gene expression.

The differential expression of β-globin genes during ontogeny provides a model system in which to examine the mechanisms responsible for the tissue specific and developmental control of gene expression. In the β-globin gene family these control mechanisms must ensure firstly that individual β-globin genes are transcribed by RNA polymerase II, secondly that β-globin genes are expressed exclusively in cells of erythroid origin, thirdly that individual β-globin genes are expressed at precise developmental times and fourthly that β-globin genes remain coordinately expressed, in conjunction with the expression of α-globin genes. Several research strategies have

been employed to deduce these control mechanisms, and have revealed that in general this control operates at the transcriptional level and involves CIS-acting regulatory sequences which interact with specific TRANS-acting DNA binding proteins.

The β-globin promoter.

One strategy which was used involved sequence comparisons of non-coding regions of β-globin genes, both within and between species, with the aim of identifying conserved sequences with possible functional roles. Two conserved sequences located 5' to the β-globin gene were identified, which had structural similarities to prokaryotic promoter elements. The first element identified was an AT rich sequence (the "Goldberg-Hogness", "TATA" or "ATA" box) located 20-30 bp upstream of the RNA initiation site (CAP site) which was strikingly similar to the "Pribnow box" of prokaryotic promoters (5'-TATAAT-3') (Rosenberg and Court, 1979). The second sequence identified was the "CCAAT" box located 70 - 90 bp upstream of the CAP site which was similar to several *Escherchia coli* promoters (Efstratiadis *et al.*, 1980; Benoist *et al.*, 1980).

DNA sequence comparisons are not sufficient to prove a functional role for conserved elements, and so techniques were developed to alter their sequence *in vitro* and examine the effect this had on the expression of an adjacent β-globin gene *in vivo*. Dierks *et al.* (1983) introduced deletions into the 5' flanking region of a cloned rabbit β-globin gene and transfected the modified gene into mouse 3T6 cells (a mouse erythroid cell line). β-globin transcripts were assayed by S1 nuclease mapping and

revealed three regions upstream of the CAP site that were required for "maximal transcription" of the gene *in vivo* :- the "ATA box", the "CCAAT box" and a region between nucleotides -83 and -111 bp from the CAP site containing 2 copies of the sequence "CACCC". These results were verified by Myers *et al.* (1986) using a chemical based procedure for generating large numbers of single base substitutions in the 5' region of a mouse β-globin gene and a HeLa cell "transient expression assay" to determine the effects of these substitutions on gene transcription.

A third strategy that has been used to deduce the sequences involved in the control of β-globin gene expression, has been to determine the DNA mutations responsible for the naturally occuring β-globin disorders, such as the β-thalassemia's. Three point mutations were identified within the "ATA box" and two within the "CACCC box" of thalassemic individuals, which resulted in the reduction of normal β-globin gene expression by up to 80% (Orkin and Kazazian, 1984). These results confirm that the "ATA box" and the "CACCC box" are important for efficient β-globin gene expression *in vivo*.

The adult β-globin gene therefore consists of at least 4 promoter elements, including one "ATA box" ($\sim$30 bp upstream of the CAP site), one "CCAAT box ($\sim$76 bp upstream), and two "CACCC boxes" ($\sim$90 - 106 bp upstream). Similar sequences have also been identified in the promoter region of ϵ- and γ-globin genes (Hardison, 1983). The embryonic β-like globin genes however have only one "CACCC box". The human γ-globin gene has two "CCAAT boxes" although only one is required for efficient promoter function (Anagnou *et al.*, 1986). A recent study

by Antoniou *et al.* (1988) suggests a fifth promoter element may be present in the human adult β-globin gene, located 160 bp upstream of the CAP site, although the precise sequence and function of this element have not been determined.

Tissue specific and developmental control of β-globin gene expression.

The identification of DNA sequences which influence the tissue and developmental control of β-globin gene expression was originally made possible by the development of two transcription assay systems. One system involved the use of mouse erythroleukemia (MEL) cells, which are virus-transformed cells arrested at the pro-erythroblast stage of erythroid development (Chao *et al.*, 1983; Charnay *et al.*, 1984; Wright *et al.*, 1984). Several chemical inducers (eg. dimethyl sulfoxide) lead to induction of erythroid maturation and an approximately 20 fold increase in exogenous adult β-globin expression. Expression of exogenous ε-, $^{A}\gamma$- and $^{G}\gamma$-globin genes, however, is not induced by these agents. The DNA sequences required for induction of β-globin expression can be determined by transforming MEL cells with specific cloned β-globin gene constructs and assaying for transcription by S1 nuclease mapping. The second system has involved the introduction of cloned globin genes into the mouse germ-line and an analysis of their expression in the resulting transgenic animals (Chada *et al.*, 1985; Magram *et al.*, 1985).

Studies using these two approaches revealed that regulatory sequences closely linked to the β-globin gene were required for tissue specific regulation of the gene

(Chao *et al.*, 1983; Charnay *et al.*, 1984; Wright *et al.*, 1984; Chada *et al.*, 1985; Magram *et al.*, 1985). In particular it was shown that sequences located within the 3rd exon of the β-globin gene and a sequence lying 555 - 800 bp 3' to the gene could individually activate a cloned globin gene (Behringer *et al.*, 1987; Kollias *et al.*, 1987; Antoniou *et al.*, 1988). These sequences could induce expression when located either 5', 3' or in an opposite orientation from the promoter. Such properties are characteristics of enhancer-like elements. An enhancer has also been identified 3' to the chicken adult β-globin gene using HD3 cell (inducible avian erythroid precursor cell line) and embryonic erythrocyte transcription assay systems (Hesse *et al.*, 1986; Emerson *et al.*, 1987; Choi and Engel, 1988).

Sequence elements involved in the regulation of human γ-globin genes have also been studied intensively. These studies have made use of naturally occuring point mutations located 5' to the γ-globin gene , which are associated with the disease "hereditary persistence of foetal haemoglobin" (HPFH), a condition in which the γ-globin gene is expressed into adult life (Giglioni *et al.*, 1984; Stamatoyannopoulos and Nienhuis, 1987). An enhancer element located 3' to the gene was identified using a human K562 erythroleukemia cell transcription assay, a cell line normally expressing foetal but not adult β-globin genes (Bodine and Ley, 1987). In addition, 2 regions which influenced γ-globin gene expression were identified at positions which coincided with the location of mutations in HPFH individuals approximately 200 bp upstream of the CAP site (Lloyd *et al.*, 1989; Martin *et al.*, 1989). One particular mutation (at -175, T -> C) was found to increase expression of the γ-globin gene in erythroid

cells (Martin *et al.*, 1989).

Sequences which influence the expression of the human ϵ-globin gene have also been described. Cao *et al.* (1989) found an element located between -177 and -392 bp relative to the CAP site which acts as a transcriptional silencer and therefore negatively regulates the expression of the ϵ-globin gene. Interestingly, a transcriptional "silencer" has also been identified 5' to the ζ-globin gene (embryonic α-globin gene), a gene which is asynchronously regulated with the ϵ-globin gene in humans. It has been postulated that both silencers may interact with common TRANS-acting regulatory factors (Peschle *et al.*, 1985; Lamb *et al.*, 1989). Two silencers located 5' to the adult β-globin gene, which bind a common repressor protein, have also been described (Berg *et al.*, 1989).

It has been observed that cloned globin gene constructs introduced into cell lines or the mouse genome are consistently expressed at lower levels than their endogenous counterparts (Chada *et al.*, 1985). However, the introduction into the mouse genome of a β-globin gene construct containing flanking sequences 50 kb 5' and 20 kb 3' of the gene resulted in transcription levels which were as high as the endogenous mouse β-globin gene and "at a level directly related to its copy number yet independent of its position of integration into the genome" (Grosveld *et al.*, 1987). A series of major erythroid-specific DNase I hypersensitive sites were mapped at positions approximately 6-20 kb 5' of the CAP site of the ϵ-globin gene and 20 kb 3' to the CAP site of the adult β-globin gene (Tuan *et al.*, 1985).

Independent studies by Tuan *et al.* (1989) and Talbot *et al.* (1989) revealed the

existence of a "dominant control region" or "locus activation region" (LAR) encompassing the DNase I hypersensitive sites 10.2 - 11 kb 5' of the ϵ-globin gene, which is capable of conferring high level globin gene expression in erythroid cells. When LAR is linked to both a foetal and adult β-globin gene there is a normal switch from foetal to adult β-globin gene expression in transgenic mice. This switch in expression is not observed when LAR is linked individually to these genes leading to the suggestion that the foetal to adult β-globin switch in humans "is controlled through a mutually exclusive interaction between LAR and either the γ- or β-globin genes resulting in the expression of only one gene at any given moment" (Enver *et al.*, 1990).

In summary, the control of expression of β-like globin genes is influenced by a wide range of CIS-acting regulatory sequences including promoters, silencers, enhancers and a locus activating region. Each of these regulatory sequences interacts with specific TRANS-acting factors (see Evans *et al.*, 1988; Mantovani *et al.*, 1988, 1989; Berg *et al.*, 1989; Evans and Felsenfeld, 1989). The precise mechanisms of these interactions remain unknown.

2.10 Globin macromolecular sequences in marsupials: evidence for different molecular forms of haemoglobin.

There have been just two scientific publications reporting globin DNA sequences in marsupials (see Section 2.7), while in contrast globin protein sequences

in marsupials have been well documented, particularly from the macropodid marsupials, (see Thompson and Air, 1971; Thompson *et al.*, 1969; see review by Hope *et al.*, 1990).

Varying molecular forms of adult haemoglobin were detected in a number of macropodid species. Two major forms of β-globin were detected in the grey kangaroo *Macropus giganteus* and were subsequently shown to involve polymorphic variation at a single β-locus (Thompson *et al.*, 1969; Air and Thompson, 1969). A polymorphism involving the same amino acid substitution (2 Glu-His) was also detected in the wallaroo *Macropus robustus*, yet at least one fixed amino acid difference (56 Ala-Gly) distinguishes *M.giganteus* and *M.robustus*. Thompson and Air (1971) suggested that "either the grey kangaroo / wallaroo common ancestor had the polymorphism at β2 and the glycine -> alanine change became incorporated into both alleles by independent mutations or by crossing over with selection of only one of the products, or the β2 polymorphism has evolved independently in both species". Two variant β-chains were also detected in the agile wallaby *Macropus aglis* and partial amino-acid sequence data from a small peptide revealed that the chains differed at two non-adjacent sites (Thompson and Air, 1971). In addition, α- and β-chain variants have been reported for the potoroo *Potorous tridactylus* although the molecular and genetic basis for this variation is unknown (Thompson and Air, 1971). Much of the variation in β-chain sequence detected by Thompson and his colleagues can be attributed to polymorphic variation at a single genetic locus. The possibility remains however, that two adult β-globin loci may exist in *M.agilis* and *P.tridactylus* and

perhaps other macropodid and marsupial species.

Molecular forms of haemoglobin restricted in the timing of their expression to pouch young development have been reported by a number of authors, providing evidence that developmentally regulated globin genes exist in marsupials. Hope (1970), using starch gel electrophoresis, detected 2 electrophoretic forms of haemoglobin in brush tail possum (*Trichosurus vulpecula*) pouch young of less than 10 days old (*post partum*), which were not found in older animals. Richardson and Russell (1969) also detected a non-adult form of haemoglobin in *M.rufus, M.giganteus* and *M.robustus* pouch young less than 70 days old. The switch from embryonic to adult haemoglobin production appeared to vary from animal to animal, occurring as early as 7 or as late as 40 days *post partum*. More recently, Holland *et al.* (1988), using the procedure of isoelectric focussing, detected four types of non-adult haemoglobin in *Macropus eugenii* pouch young. In this species it was observed that replacement of the embryonic haemoglobin with adult haemoglobin was complete by day 13 (*post partum*) unlike the situation that occurs in the larger macropods *M.rufus, M.giganteus* and *M.robustus*. It was also noted by Richardson and Russell (1969) and Holland *et al.* (1988), that there is a rapid turnover of erythrocytes during the first 7 to 15 days of pouch young development, as erythrocytes change from a large nucleated megaloblastic form, to a normoblastic and non nucleated form. Richardson and Russell proposed that "if the change in haemoglobin type occurs before the change in the red blood cell population from foetal to adult, then the HbF disappears quite quickly from circulation". This would appear to be the case in *M.eugenii*, where there

is a rapid switch from embryonic to adult haemoglobin in pouch young between 3 and 7 days old *post partum* and also in *T.vulpecula* where no embryonic forms of haemoglobin were detected after 10 days of pouch young development (Holland *et al.*, 1988; Hope, 1970).

Studies of the oxygen-haemoglobin binding characteristics of blood in marsupial pouch young have been reported by Baudinette *et al.* (1988), Holland *et al.* (1988) and Murphy *et al.* (1977). Murphy *et al.* (1977) showed that *D.virginiana* pouch young haemoglobin had a higher affinity for oxygen than adult haemoglobin. If the Hb-O_2 affinity in the neonate reflects that which exists *in utero*, the *D.virginiana* embryonic haemoglobin has similar physiological properties to eutherian embryonic haemoglobins. In contrast, studies by Baudinette *et al.* (1988) and Holland *et al.* (1988) on the Hb-O_2 binding characteristics of pouch young haemoglobin in *M.eugenii* report lower Hb-O_2 affinities in pouch young than in adult blood. Therefore, *M.eugenii* appears to employ an Hb-O_2 affinity strategy favouring "tissue oxygen delivery to a greater degree" which may be advantageous to a small rapidly growing pouch young (Baudinette *et al.*, 1988). Similar studies have not been performed in *S.crassicaudata* or any other marsupial to date.

It is clear from the above studies that haemoglobin proteins, restricted to embryonic or pouch young development occur widely in marsupials. By implication, embryonic β-like globin genes are therefore expected to exist in marsupials.

2.11 The molecular clock hypothesis: relative rate tests of gene evolution.

The molecular clock hypothesis, which states that rates of amino-acid substitutions in proteins are constant over time, has been a controversial issue since it was first proposed by Zuckerkandl and Pauling in 1965 (see Wilson *et al.*, 1977; Kimura, 1983). It has been supported by protein sequence comparisons and immunological distance measurements (Wilson *et al.*, 1977; Kimura, 1983) but refuted in studies by Morris Goodman and his colleagues (Goodman *et al.*, 1975; Czelusniak *et al.*, 1982). The latter workers contend that accelerated evolution often occurred after the duplication of a gene and during adaptive radiation. In particular, it has been emphasized that rates of nucleotide substitution have slowed down during the evolution of higher Primates and during Hominoid descent (Goodman *et al.*, 1971; Wu and Li, 1985; Britten, 1986; Koop *et al.*, 1986; Li and Tanimura, 1987), although a number of studies indicate this conclusion is still contentious (see Easteal, 1988; Shaw *et al.*, 1989). The reason for the Hominoid rate slow down is unclear, although it has been suggested that differences in the number of germ-line DNA replications per year (generation time differences) or changes in DNA repair mechanisms may be responsible (Li and Tanimura, 1987; Britten, 1986).

There have been several attempts to re-evaluate the molecular clock hypothesis (see Wu and Li, 1985; Li *et al.*, 1985; Easteal, 1988), using a procedure called the "relative rate test" (Sarich and Wilson, 1967). This test involves a comparison of the evolutionary distance from an "ingroup" taxon (1) to an "outgroup" taxon (3) with that

from a second "ingroup" taxon (2) to the same "outgroup". If such distances are approximately equal then one concludes that rates of evolution have been similar in the two "ingroups". It should be noted, however, that the relative rate test can not detect changes in the rate of evolution of genes in two lineages if these changes occur proportionally (Fitch, 1976).

Using this test Wu and Li (1985) compared the coding regions of 11 genes from rodents (mouse or rat) and man, using genes from either the goat or rabbit as an outgroup. They concluded that there was a significantly faster rate of nucleotide substitution in the rodent lineage than in the lineage leading to humans and that the differences in substitution rate could be accounted for by the shorter generation times and thus higher germ-line mutation rates in rodents. In particular, it was found that the synonymous rate of substitution in rodents was 4 - 10 times higher than in higher primates (Li and Tanimura, 1987) and that "non-synonymous nucleotide substitutions proceed roughly in a clock-like manner as long as there is no large change in generation time." (Li *et al*, 1985).

Easteal (1988) investigated the molecular clock hypothesis using relative rate tests of β-like globin genes in species from four eutherian orders; Artiodactyla, Rodentia, Lagomorpha and Primate. The relative rate tests he used were based on branching orders established by cladistic analyses of nucleotide sequences which have no reliance on the fossil record. He concluded that the "globin genes are evolving at the same rate in the different mammalian orders" and further that "the rate of molecular evolution has not slowed down in the primate lineage leading to humans".

Easteal suggested that previous studies, which indicated a rate slow down in the Hominoid lineage, were "based on an incorrect interpretation of the poor fossil record" and proposed that "either the mammalian orders diverged earlier than has been supposed or humans diverged from African apes later than the currently accepted 5 to 8 mya". His results also showed, in contrast to those of Wu and Li (1985), that the molecular evolutionary rate of globin genes is not correlated with generation time.

A requirement of the relative rate test is that the relative divergence order of the compared species or genes is known, and that a reference species or gene forms an outgroup. In the relative rate tests reported by Wu and Li (1985) and Easteal (1988), the outgroup taxa were chosen on the basis of questionable divergence orders of eutherian mammals. Wu and Li (1985) used fossil derived divergence orders, and claimed that artiodactyls, carnivores, and lagomorphs "branched off before or around the time of the primate - rodent split". Easteal (1985, 1987) was particularly critical of these divergence orders, arguing that there are too many inconsistencies in the fossil record of mammals, for divergence orders to be inferred using the available fossil data. He therefore used cladistic analyses of β-like globin nucleotide sequences to establish the divergence orders of the species or genes in his tests. Problems in this approach occur, however, if there is non-uniformity of evolutionary rates in individual branches of the phylogenetic tree. For example, if there were significantly faster rates of gene evolution in the rodent lineage, then the tree proposed by Easteal may be incorrect, and the reference sequences used in his tests may not be outgroups. The use of incorrect outgroups in these studies would be fatal to the conclusions drawn.

2.12 The phylogeny of mammals: inter-relationships of marsupials, monotremes and eutherians.

Traditionally, eutherians and marsupials have been assigned to the Subclass Theria and monotremes to a separate Subclass (Prototheria) of Class Mammalia. These assignments were based on morphological studies of the anatomy of the mammalian skull (Hopson, 1970; Kermack and Kielan-Jaworowska, 1971). Kemp (1983) re-analysed these data by application of cladistic analysis and concluded that the subdivision of monotremes into a separate subclass was unjustified. Collaborative evidence for this conclusion comes from new fossil finds (Archer *et al.*, 1985; Kielan-Jaworowska *et al.*, 1987) which suggest that monotremes may be more closely related to therian mammals than was previously assumed.

The inter-relationships of the three mammalian infra-classes have also been investigated using amino acid sequence comparisons (Whittaker *et al.* 1980) and "geneological reconstructions" (Goodman *et al.*, 1985; 1987) of α-globin, β-globin and myoglobin sequences. It was Whittaker *et al.* (1972) who first noticed some anomalies in the globin data, with respect to the branching order of monotremes, marsupials and eutherians. They observed that "the sequence of the echidna β-chain would not support any palaeontological estimate that gave a date for marsupial divergence (from eutherians) more recent than for monotremes". Whittaker and Thompson (1975) reached a similar conclusion with respect to the platypus β-globin. In their final paper on monotreme globins Whittaker *et al.* (1980) concluded that

calculations of the date of monotreme divergence from eutherians and marsupials, using myoglobin differences, indicate "a divergence off the marsupial line while α- and β-globin calculations indicate a closer affinity with the placental mammals".

Such incongruences have also been reported by Morris Goodman and his colleagues in cladistic analyses of α-globin, β-globin and myoglobin amino acid sequences, involving geneological reconstructions based on the maximum parsimony procedure (Goodman *et al.*, 1985: 1987). In these studies it was shown that geneological reconstructions of myoglobin and β-globin sequences depict monotremes as an "ancient branch of Eutheria". In contrast, geneological reconstructions of α-globin sequences depict "Prototheria (monotremes) as the sister group of Theria". More recently, studies by Czelusniak *et al.* (1990) using geneological reconstructions of combined sequences of the three protein chains, α-globin, β-globin and myoglobin, showed that "the consensus of most parsimonious trees has Eutheria, Marsupialia and Monotremata emerge at the same point from the stem of Mammalia". To account for these results Czelusniak *et al.* (1990) suggested that the anomalous monotreme-eutherian grouping "is due to fortuitous homoplasy between monotreme and eutherian globin sequences" and further that "the period of time between the two ancestral branch points separating the three infra-classes of living mammals was much shorter than the periods of independent evolution in the three lineages".

Westerman and Edwards (1991) using the technique of DNA-DNA hybridization, obtained a date of 110 - 137 mya for the divergence of the monotreme and marsupial lineages (based on a date for the divergence of didelphis and kangaroos

of between 103 and 128 mya). The upper estimate is close to the figure of 143 mya (Whittaker *et al.* 1980) for this divergence, and 149 mya (Clemens *et al.*, 1989) for the divergence of the monotreme lineage from the combined marsupial/eutherian lineage, both dates being estimated from α-globin, β-globin and myoglobin data assuming a constant rate of amino acid substitution (the latter being based on a date of 135 mya for the divergence of eutherians and marsupials). Clemens *et al.* (1989) concluded that "the estimated date of divergence of the monotreme lineage from the combined metatherian/eutherian lineage is probably not significantly greater than that of the metatherian (marsupial) lineage from the eutherian lineage".

Molecular data therefore lend support to the idea that monotremes should be included within the Class Theria as a sister group of eutherians and marsupials, but do not provide definitive evidence as to whether the divergence of the monotreme lineage from an ancestral mammalian stock occurred prior to the divergence of the marsupial and eutherian lineages. To resolve this incongruence, sequence data are required from a wider range of proteins and from a wider range of taxa. For example, the inclusion in globin phylogenies of α-globin, β-globin and myoglobin sequences from a dasyurid and representatives of other marsupial families should reduce errors in the derived phylogenetic tree caused by non-uniformity of evolutionary rates in separate lineages.

CHAPTER 3

MATERIALS AND METHODS

3.1 Materials

Chemicals

All chemicals were of analytical reagent grade and purchased from BDH except for the following:

Acrylamide, bis-acrylamide were "Electran" grade and purchased from BDH.

CM-Sephadex C-50 (capacity: 4.5 +/- 0.5 meq/g; particle size: 40 - 120 μ) was purchased from Pharmacia.

Ammonium per sulfate (Ultra Pure) and mixed bed resin (AG 501-lx8, 20 - 50 mesh) was purchased from Biorad.

Low melting temperature (LMT) agarose "Ultra Pure" was purchased from BRL and "Seaplaque" LMT agarose purchased from FMC.

Nitrocellulose membranes were purchased from Schleicher and Schuell.

Cellogel 500 was purchased from Chemetron, Milan.

Isotopes

α^{32}P-dCTP was purchased from Bresatec.

Enzymes

Bovine Pancreas DNaseI, RNaseA, Klenow fragment of DNA polymerase I (sequencing grade), restriction enzymes and bovine serum albumin (BSA; DNase free) were purchased from Boehringer-Mannheim.

T4 DNA ligase and *E.coli* DNA polymerase I were purchased from Bresatec.

Bal31 was purchased from Promega.

DNA

Nucleoside triphosphates and dideoxy nucleoside triphosphates were purchased from Boehringer Mannheim.

λ DNA digested with HindIII, SPP-1 DNA digested with EcoRI and synthetic random hexamer primers were purchased from Bresatec.

High molecular weight DNA markers were purchased from BRL.

Salmon sperm DNA was purchased from Sigma.

pBR322 DNA was a gift from Neil Dear.

M13 DNA was a gift from A. Sivaprasad, Biochemistry Dept., University of Adelaide.

E.coli strains

NM538 (ED8654 cured of sup E) was a gift from Dr J.N. Timmis.

MC1061 was a gift from Neil Dear.

JM101 was a gift from Dr R.M. Hope, Genetics Dept., University of Adelaide.

Experimental animals

S.crassicaudata were obtained from colonies maintained by the Genetics Department, R.A. Fisher Laboratories, University of Adelaide, and were made available by Prof. J.H. Bennett.

Films

X-ray film was purchased from Kodak.

Solutions and Media

All solutions and media were made as specified by Maniatis *et. al.* (1982) except for the following:

PBS: 0.8% w/v NaCl
0.02% w/v KCl
0.115% w/v Na_2HPO_4
0.02% w/v KH_2PO_4
pH 7.4

1 x TM: 10 mM Tris-HCl pH 8.0
10 mM $MgCl_2$

YT-broth: 1% w/v bacto yeast extract
1.6% w/v bacto-tryptone
0.5% w/v NaCl

M13 minimal + glucose medium:

1.05% w/v K_2HPO_4

0.45% w/v KH_2PO_4

0.1% w/v $(NH_4)_2SO_4$

0.05% w/v sodium-citrate

0.02% w/v $MgSO_4$

0.2% w/v glucose

0.0005% w/v Thiamine HCl

25 x BXE: xylene cyanol- 25 mg

bromo-phenol blue- 25 mg

dissolved in 1 ml 0.25 mM EDTA (pH 8.0)

10 x TBE (pH 8.3):

Tris 7-9- 108.0 g

Na_2EDTA- 9.3 g

boric acid- 55 g

dissolved in 1 litre of H_2O

3.2 Methods

3.2.1 Isolation of recombinant clones from a bacteriophage genomic DNA library.

1. Preparation of NM538 plating cells.

The *E. coli* host NM538 was cultured to saturation overnight at 37°C in L-broth supplemented with 20mM $MgSO_4$ and 0.2% w/v maltose. Cells were pelleted by centrifugation at 5000g for 5 minutes, resuspended in half the original culture volume of 10mM $MgSO_4$ and stored at 4°C for up to 3 days.

2. Titration of bacteriophage stocks.

Ten-fold serial dilutions of the phage stock (100μl) were mixed with 100μl of plating cells and incubated at 37°C for 15 minutes. Top-layer agar or agarose (3μl for each 8.5cm diameter petri dish), pre-warmed to 45°C, was added to each solution and the mixture was immediately plated onto 1.5% w/v L-agar plates. When the top layer had set, the plates were incubated upside down at 37°C overnight.

3. Plaque-lifting of bacteriophage DNA to nitrocellulose.

(Benton and Davis, 1977)

Plates of bacteriophage were kept at 4°C for one hour to harden the top-layer agarose. Nitrocellulose filters, cut to the dimensions of the plate, were carefully layered onto the top-layer agarose surface. Using a 26 gauge hyperdermic needle dipped in Indian ink, ink marks were placed at random positions by piercing through the nitrocellulose filter into the bottom L-agar. After a one minute transfer the nitrocellulose filter was removed, dried briefly and placed DNA side down onto Whatman paper (3MM) saturated with 0.5M NaOH, 1.0M NaCl for one minute. The filter was then dried for one minute and placed DNA side down onto Whatman paper saturated with 0.5M Tris-HCl (pH 7.5), 1.5M NaCl for one minute, followed by transfer to Whatman paper saturated with 2 x SSC for one minute. The filter was air dried and baked at 80°C *in vacuo* for two hours.

4. Hybridization of plaque filters to radio-labelled nucleic acid probes.

Plaque filters were washed in 10 x Denhardts, 2 x SSPE at 65°C for a minimum of 2 hours, and then pre-hybridized for 5 hours at 42°C in a buffer containing 50% v/v formamide, 5 x SSPE, 5 x Denhardts, 0.1% w/v SDS, 100μg/ml sonicated and single stranded salmon sperm DNA, 50μg/ml *E. coli*

chromosomal DNA. DNA probes were radio-labelled (see sections 3.2.3.5 and 3.2.3.6), denatured, and added directly to the pre-hybridization buffer. Incubation was continued for 18-24 hours at 42°C. Filters were washed 3 times in 2 x SSC, 0.1% w/v SDS at room temperature, followed by 2 washes in 1 x SSC, 0.1% w/v SDS at 65°C. Filters were wrapped in clear plastic, taped to X-ray films and autoradiographed for one to seven days in the presence of an intensifying screen at -70°C. Before developing the autoradiograph, ink marks were located on the plaque filters and using a needle to stab through the ink marks, corresponding marks were made on the X-ray film. Plaques showing hybridization to the probe were located by aligning the ink marks on the agar plate with the corresponding marks on the autoradiograph.

5. Preparation of bacteriophage stocks from a single plaque.

Single plaques were removed from plates in "agar plugs" by stabbing through the bottom agar with the end of a pasteur pipette. The "agar plug" was resuspended in 1ml SM buffer with a drop of chloroform, and left at 4°C overnight. An aliquot (100μl) of the phage suspension was mixed with 100μl of plating cells, incubated at 37°C for 15 minutes and plated onto an 8.5cm petri dish containing L-agar as described in section 3.2.2. After overnight incubation at 37°C, SM buffer (5ml/plate) was added to the plate and left at room temperature for several hours with moderate shaking. The phage suspension was removed and

debris was pelleted by centrifugation at 5000g for 5 minutes. The phage stock was titred and the above procedure repeated until a phage particle concentration of 10^9 - 10^{10} pfu's /ml was obtained. Phage stocks were stored in SM at 4°C with a few drops of chloroform.

6. Rapid small scale isolation of bacteriophage DNA.

Small scale preparations of bacteriophage DNA were prepared using the plate-lysate procedure described by Maniatis *et al.*, (1982, pp 371-372).

7. Large-scale preparation of recombinant bacteriophage particles.

Liquid cultures of bacteriophage (recombinant EMBL3 clones) were prepared using the procedure described by Maniatis *et al.*, (1982, pp 77-78) with minor modifications. The strain of *E. coli* used was NM538, a non-selective host for EMBL3. Cells were grown in L-broth supplemented with 20mM $MgCl_2$. A bacteriophage particle : cell ratio of 5 x 10^8 pfu's : 10^{10} cells was used in 500mls of pre-warmed L-broth + 20mM $MgCl_2$ to obtain confluent lysis of cells after 16 hours of incubation at 37°C.

8. Purification of EMBL3 bacteriophage particles.

Bacteriophage particles were purified using the procedure of Yamamoto *et al.* (1970) as described in Maniatis *et al.* (1982, p 80) with minor modifications. After PEG precipitation the bacteriophage suspension was extracted twice with chloroform and stored at 4°C. The bacteriophage particles were not further purified on caesium chloride gradients.

9. Large scale preparation of bacteriophage DNA.

Aliquots (500μl) of the PEG purified phage stocks (section 3.2.5) were mixed with 5μl 10% w/v SDS and 5μl 0.5M EDTA (pH 8.0), and incubated at 68°C for 15 minutes. The mixture was extracted two times with phenol/chloroform (1:1) and once with chloroform. To the final aqueous layer, an equal volume of isopropanol was added to precipitate the bacteriophage DNA. The DNA was removed by spooling, washed in 70% v/v ethanol, dried briefly and dissolved in 200μl TE (pH 8.0). The DNA was reprecipitated by adding 20μl of 3M NaOAc (pH 5.2) and 440μl of ethanol. The DNA was pelleted by centrifugation at 12,000g for 5 minutes, washed in 70% w/v ethanol, dried briefly *in vacuo* and resuspended in 200μl TE (pH 8.0). If necessary, for subcloning experiments, the bacteriophage DNA (up to 20μg) was further purified on a Nensorb column (Du Pont) using the procedure specified by the manufacturer.

3.2.2 Preparation of subclones in the bacterial plasmid pBR322.

1. Ligation of bacteriophage DNA into pBR322.

pBR322 (2μg) and bacteriophage DNA (5μg) were each cleaved with Bam HI and Sal I (section 3.2.3.1). The digested DNA's were precipitated by the addition of 2/3 volume of 5M ammonium-acetate and 2 volumes of ethanol at -20°C, and pelleted by centrifugation for 15 minutes in a microfuge at 4°C. DNA pellets were washed once in 70% v/v ethanol, dried *in vacuo* and resuspended in H_2O. Ligations were performed by mixing 5ng of the digested pBR322 DNA in a total reaction volume of 20μl 1 x ligase buffer (0.05M Tris-HCl (pH 7.4), 10mM $MgCl_2$, 10mM DTT, 1mM spermidine, 1mM ATP, 100μg/ml BSA: DNase free); T4 DNA ligase (1 unit) was added and the mixture was incubated in a 5 litre beaker of H_2O (starting temperature approximately 25°C), which was placed at 4°C overnight (final temperature of H_2O, after 18 hours, approximately 10°C).

2. Transformation of *E. coli* with plasmid DNA.

A single colony inoculate of MC1061 in 5ml L-broth was grown overnight at 37°C. The culture was diluted 1/50 v/v with L-broth and incubated at 37°C until an $O.D._{600}$ of 0.6 was reached. The cells were chilled on ice for 2 hours and pelleted by centrifugation at 8,000g for 10 seconds at 4°C. The supernatant was

discarded and the cells were resuspended in a ½ volume (25ml) of ice-cold 0.1M $MgCl_2$. The cells were pelleted by centrifugation at 8,000g for 30 seconds, resuspended in a 1/20 volume (2.5ml) of ice-cold $CaCl_2$ and left on ice for 2 - 16 hours. Competent cells (0.2ml) were mixed with 10ng of the transforming DNA and held on ice for 30 minutes with occasional mixing. The mixture was incubated at 42°C for 2 minutes, held on ice for 30 minutes and warmed to room temperature over 20 minutes. L-broth (0.5ml) was added to the transformed cells and the mixture was incubated at 37°C for 30 minutes. The transformed cells were then mixed with 3ml of 0.8% w/v L-agar, and plated onto a 1.5% w/v L-agar plate supplemented with the appropriate antibiotic (12.5μg/ml tetracycline). The plate was incubated at 37°C overnight.

3. Detection of recombinant plasmid DNA.

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

4. Rapid small scale preparation of recombinant plasmid DNA.

Plasmid DNA from small bacterial cultures (< 5ml) was prepared using the procedure of Maniatis *et al.*, (1982, pp 368-9). Transformed cells were cultured in L-broth supplemented with 50μg/ml ampicillin.

5. Large scale preparation of recombinant plasmid DNA.

(Birnboim and Doly, 1980)

A single colony of bacteria (transformed with a recombinant plasmid) was used to inoculate 5ml of L-broth containing the appropriate antibiotic (50μg/ml ampicillin) and the culture was grown overnight at 37°C with shaking. The culture was diluted into 500mls of L-broth (containing 50μg/ml ampicillin) and incubated at 37°C with shaking until an $O.D._{600}$ of 0.6 was reached. Chloramphenicol (85mg dissolved in 2.5 ml of ethanol) was added and the culture was further incubated overnight at 37°C with shaking.

Cells were pelleted by centrifugation at 5,000g for 10 minutes, and resuspended in 3.6ml of buffer A (25mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0) and 50mM glucose). Lysozyme (8mg dissolved in 400μl of buffer A), was added and the mixture kept on ice for 40 minutes. A total of 8ml of buffer B (0.2M NaOH, 1.0% w/v SDS) was added, mixed by inversion and kept on ice an additional 10 minutes. 5ml of 3.0M NaOAc (pH 4.6) was added, mixed by

inversion and after a 40 minute incubation on ice, cellular debris and chromosomal DNA was pelleted by centrifugation at 40,000g for 15 minutes at 4°C. The supernatant was decanted and was treated with RNase A (DNase free) at a final concentration of 10μg/ml for 2 hours 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 nucleic acids were precipitated by the addition of 2 volumes of ice cold ethanol. After a minimum of 2 hours at -20°C the nucleic acids were pelleted by centrifugation at 15,000g for 30 minutes and the pellet was washed once in 70% v/v ethanol and dried under vacuum. The pellet was then resuspended in 1.6ml of H_2O to which was added 400μl of 4M NaCl and 2ml of 13% w/v PEG 6000. The mixture was left on ice for at least 2 hours and the plasmid DNA was pelleted by centrifugation at 15,000g for 15 minutes at 4°C. The supernatant was discarded and last traces were removed with a pasteur pipette. The pellet was washed in 70% v/v ethanol, dried under vacuum and resuspended in 500μl TE (pH 8.0).

3.2.3 Characterization of recombinant plasmid or bacteriophage clones.

1. Restriction digest conditions.

Restriction digests of DNA were carried out under the conditions specified by the supplier.

2. Partial restriction analysis of DNA.

A reaction mixture was prepared with 5μg of DNA in 150μl of 1 x restriction buffer (manufacturers' recommended reaction buffer). 30μl of the reaction mixture was dispensed into a 1.5 ml polypropylene tube (tube 1) and 15μl was dispensed into eight other tubes, on ice. Four units of the restriction enzyme was added to tube 1, mixed well and 15μl of this mixture was transferred to tube 2. The two-fold serial dilution was continued through to tube 8. No enzyme was added to tube 9. The tubes were incubated at 37°C for 1 hour and the reaction was stopped by chilling to 0°C and adding EDTA to a final concentration of 20mM. Each of the digests was then analysed by electrophoresis on an agarose gel.

3. Agarose gel electrophoresis.

Electrophoresis was carried out on horizontal slab gels containing agarose and 1 x TAE buffer. DNA was visualized by staining with 5μg/ml ethidium bromide in 1 x TAE buffer for 20 minutes and examination under UV (254nm) light.

4. Poly-acrylamide gel electrophoresis.

Electrophoresis was carried out on vertical gels (16cm wide, 18cm long, 0.15cm thick), containing 6% w/v poly-acrylamide, at 20V/cm for 2 hours. The electrophoresis buffer used was 1 x TBE (90mM Tris-borate (pH 8.0) and 2.5mM EDTA). Gels were stained in 5μg/ml ethidium bromide in 1 x TBE buffer for 20 minutes and DNA was visualized under UV (254nm) light.

5. Purification of DNA from LMT agarose gels.

(Weislander, 1979)

DNA samples were electrophoresed on horizontal slab gels containing LMT agarose (BRL or Seaplaque), at 7.5V/cm for 2 - 3 hours at 4°C. The gels were stained with ETBR (1μg/ml) for 15 minutes and viewed briefly under UV (254nm) light to determine the location of DNA fragments. Excised fragments were mixed with an equal volume of 0.1M Tris-acetate (pH 7.5), 5mM EDTA, 0.5M NaCl and heated to 65°C until the agarose was melted. The mixture was extracted with an equal volume of phenol (at 37°C), followed by re-extraction of the aqueous layer once with phenol/chloroform (1:1) and once with chloroform. DNA was precipitated by the addition of 2 volumes of ethanol at -20°C and recovered by centrifugation at 12,000g for 30 minutes at 4°C. The pellet was washed once in 70% v/v ethanol, dried *in vacuo* and resuspended in 5 - 10μl of TE (pH 8.0).

6. *In vitro* synthesis of radio-labelled DNA.

(i) Oligo-labelling of DNA restriction fragments.

(Hodgson *et al.*, 1987)

Oligo-labelling reactions were carried out in a total volume of 20μl. Random hexamer primers (60ng) were combined with 200ng of template DNA and H_2O to a final volume of 12μl. After immersion in a boiling water bath for 3 minutes, the mixture was placed on ice and 2μl of 10 x oligo-labelling buffer (10 x buffer: 0.5M Tris-HCl (pH 6.9), 0.1M $MgSO_4$, 1mM DTT, 0.6mM each of dATP, dGTP, and dTTP) and 50μCi of α^{32}P-dCTP (> 3,000 Ci/mM) were added. The solution was mixed and centrifuged briefly in a microfuge. Klenow fragment of DNA polymerase I (5 units) was added, mixed briefly and the reaction mixture was incubated for 2 hours at room temperature. Unincorporated nucleotides were then removed as described in Section 3.2.3.7.

(ii) Nick Translation of double-stranded DNA

Nick translation of double stranded DNA was carried out in a 20μl reaction volume, using a nick translation kit (Bresatec) and the procedure specified by the manufacturer.

7. Spun column separation of unincorporated nucleotides from radio-labelled DNA.

Biogel P60 (Biorad) columns were prepared in 0.5ml polypropylene centrifuge tubes. A 27 gauge needle was used to make one hole in the bottom of the tube and four holes in the lid. Columns were prepared with 100μl of 50 - 100 mesh Biogel P60 in the lower fraction and 400μl of 100 - 200 mesh in the upper fraction. The column was compacted by centrifugation at 500g for 2 minutes and washed once with 100μl of stop buffer (10mM Tris-HCl (pH 7.5), 1mM EDTA, 0.2% w/v SDS). The radio-labelled DNA reaction mix was diluted to 50μl with stop buffer and loaded onto the top of the column. After centrifugation at 500g for 2 minutes, 50μl of stop buffer was added and the column was centrifuged again at 500g for 2 minutes. A 1μl aliquot of the 100μl effluent was mixed with 1ml of scintillation fluid and the number of cpm/μg of DNA was determined using a Beckman scintillation counter.

8. Southern transfer of bacteriophage or plasmid DNA to nitrocellulose.

DNA restriction fragments were fractionated on agarose slab gels and transferred to nitrocellulose membranes using the method of Southern (1975), as modified by Wahl *et al.*, (1979). In this procedure DNA in the gel was partially hydrolysed by treatment with 0.25M HCl for 15 minutes.

9. Hybridization of Southern filters with radio-labelled DNA probes.

Southern membranes were sealed in plastic bags and pre-hybridized for 5 hours in a buffer containing 5 x SSPE, 5 x Denhardts, 0.1% w/v SDS, 100μg/ml single stranded salmon sperm DNA at 65°C with shaking. Radio-labelled DNA probes were denatured in a boiling water bath for 5 minutes, mixed with 1ml of the pre-hybridization buffer and added directly to the pre-hybridization solution in the sealed plastic bag. Incubation was continued for 18 - 24 hours at 65°C in a shaking water bath. Filters were then washed 3 times in 2 x SSC, 0.1% w/v SDS for 10 minutes at room temperature followed by two higher stringency washes (conditions specified in Chapter 4). Filters were covered, while still wet, in plastic wrap and exposed to X-ray film in the presence of an intensifying screen at -70°C for 4 - 24 hours.

3.2.4 DNA sequencing of M13 subclones.

Sequencing procedures were obtained from A. Sivaprasad, Biochemistry Dept., University of Adelaide.

1. Ligation of DNA fragments in M13.

Ligations were carried out as specified in Section 3.2.2.1 with minor modifications. 20ng of replicative form (RF) DNA was ligated to a two-fold molar

excess of passenger DNA in a 20μl reaction volume. The ligation buffer used was 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5% w/v PEG 8000, 1mM ATP and 1mM DTT.

2. Preparation of competent JM101 bacterial cells.

M13 minimal medium + glucose (5ml) was inoculated with JM101 and grown overnight, with shaking at 37°C. The culture was diluted 1/100 v/v into 10ml of 2 x YT medium and grown with shaking at 37°C to an $O.D._{600}$ of about 0.6 to 0.8. The culture was chilled on ice for 20 minutes and then cells were pelleted at 5,000g for 5 minutes at 4°C and resuspended in 2ml of ice-cold 0.1M $CaCl_2$. The cells were kept on ice for at least one hour and then transfected with recombinant M13 DNA.

3. Transfection of JM101 with recombinant M13 DNA.

Competent cells (200μl) were mixed on ice with 1 to 10μl of the ligated DNA and left on ice for at least 40 minutes. The cells were then heat-shocked at 45°C for 2 minutes, mixed with 3ml of 0.8% w/v YT-agar, 20μl 24mg/ml IPTG (isopropylthio-β-D-galactoside; dissolved in H_2O), 20μl 20mg/ml BCIG (5-bromo-4-chloro-3-indolyl-β-D-galactoside; dissolved in di-methyl formamide) and 100μl of log-phase JM101 (OD_{600}: 0.6 - 0.8). The mixture was plated on M13 minimal +

glucose plates and incubated overnight at 37°C.

4. Single strand template preparation.

M13 minimal + glucose medium (5ml) was inoculated with JM101 (from a glycerol stock) and cultured overnight at 37°C with shaking. The fresh overnight culture was diluted 1/100 v/v with 2 x YT broth and 2ml aliquots were dispensed into sterile screw-capped polypropylene tubes of 10ml capacity. Colourless plaques from a fresh plate were tooth picked into each tube (or 20μl of a liquid stock of phage was used to inoculate the cells). Controls with 2 x YT medium alone, and cells with no phage were also set up. The tubes were incubated on a rotating shaker for 5 to 6 hours at 37°C. Cells were pelleted by centrifugation at 5,000g for 5 minutes. The supernatant was poured into 1.5 ml polypropylene tubes and further centrifuged in a microfuge for 10 minutes. 1ml of supernatant was removed from each tube and mixed with 270μl of a PEG solution (20% w/v polyethylene glycol 8000, 2.5M NaCl). The remainder of the supernatant was stored at -20°C as a stock of the clone. The tubes were left at room temperature for 15 minutes and then centrifuged in a microfuge for 5 minutes. The supernatant was discarded and the tubes were further centrifuged for 5 seconds. The last traces of supernatant were removed using a drawn-out glass pipette and each pellet was resuspended in 200μl TE. Phenol (100μl) was added to the suspension and vortexed gently. The mixture was left at room temperature for 5 minutes and

centrifuged for 2 minutes in an eppendorf microfuge. 150μl of the supernatant was mixed with 6μl of 3M NaOAc (pH 4.6) and 400μl of ice-cold ethanol and stored overnight at -20°C. The DNA was pelleted by centrifugation for 30 minutes at 4°C, washed in ethanol and dried *in vacuo*. The DNA was then dissolved in 25μl of TE and stored at -20°C.

5. Preparation of Bal31 digested DNA subclones in M13.

pSG-12 DNA (10 μg) was digested to completion with BamHI and precipitated by adding 2/3 volume of 5M ammonium acetate and 2 volumes of ethanol. The DNA was pelleted, washed once in 70% v/v ethanol, dried *in vacuo* and resuspended in 20 μl H_2O. The digested DNA was exposed to Bal31 in a total reaction volume of 250μl of 1 x Bal31 digestion buffer (0.6M NaCl, 12mM $CaCl_2$, 12mM $MgCl_2$, 20mM Tris-HCl (pH 8.0), 1mM EDTA). The mixture was heated to 30°C. Aliquots (50μl) were removed every 2 minutes and immediately mixed with 50μl phenol containing 25mM EDTA. The mixtures were centrifuged for 2 minutes in an eppendorf microfuge and the aqueous phases were removed and re-extracted with phenol/chloroform (1:1) and chloroform. DNA was precipitated by the addition of 2 volumes of ethanol to the final aqueous layer and overnight incubation at -20°C. DNA was pelleted by centrifugation for 30 minutes in a microfuge, washed once in 70% v/v ethanol and dried *in vacuo*. The DNA was dissolved in 10μl of H_2O and fragment ends were repaired by treatment with 2.5

units of the Klenow fragment of DNA polymerase I in 1 x oligo-labelling buffer (50mM Tris-HCl (pH 8.0), 10mM $MgSO_4$, 0.1mM DTT and 0.06mM each of dTTP, dATP, dCTP and dGTP), at 37°C for 40 minutes. The Klenow was inactivated by heating the reaction mixture at 70°C for 5 minutes. Bal31 digested DNA from each time interval was digested with EcoRI and deletions of the 3.5 kb EcoRI fragment of pSG-12 were purified from a LMT gel (Section 3.2.3.5) and ligated into the EcoRI/SmaI restriction sites of Mp18.

6. Sequencing of single stranded M13 DNA templates-protocol for 3 clones.

(i) Annealing.

In 3 1.5 ml polypropylene tubes, 6μl of single stranded M13 DNA was mixed with 1μl 10 x TM, 1μl of M13 17-mer primer and 2μl of H_2O. Each mixture was incubated at 60°C for 15 minutes in a heating block, cooled slowly to room temperature over 30 minutes and centrifuged briefly.

(ii) Label.

In 3 separate eppendorf tubes 10μCi of α-^{32}P-dCTP (Bresatec, 1800μCi/mM) was mixed with 1μl of label supplement (31.5μM dCTP in H_2O).

(iii) Reagents.

Zero's for labelled dCTP (quantities shown are in μl):

	A°	C°	G°	T°
0.5mM dATP	2	40	40	40
0.5mM dGTP	40	40	3	40
0.5mM dTTP	40	40	40	3
10 x TE	10	10	10	10

Zero mixes were stored at -20°C for several months.

dideoxynucleotides:

ddA - 0.5mM

ddC - 0.1mM

ddG - 0.3mM

ddT - 0.8mM

(iv) Reaction mix.

In 4 separate tubes marked (A, C, G, T) 4μl of the N° mix was mixed with

4μl of the appropriate ddN.

(v) Polymerization Reaction.

Three rows, each with 4 tubes labelled T, C, G, A were arranged on a rack and 2μl of the reaction mix was dispensed to the side of the appropriate nucleotide tubes. The annealed DNA was mixed with the label and 2μl was dispensed to the side of each of the 4 tubes of the three respective clones. The tubes were centrifuged briefly in a microfuge at 37°C. Klenow (2μl of Boehringer Klenow sequencing grade, 5 units/μl) was diluted into 28μl 1 x TM and 2μl of the enzyme mixture was dispensed to the side of each tube. The reactions were started by brief centrifugation and were then incubated at 37°C for 15 minutes. At t=13 minutes a chase solution was prepared (28μl of 0.25mM dNTP's mixed with 1μl Klenow (5 units/μl)) and 2μl was dispensed to the side of each tube. At t=15 minutes the tubes were centrifuged for 5 seconds and incubation continued for a further 15 minutes. Reactions were stopped by the addition of 4μl of a freshly prepared formamide loading buffer (made by mixing 95μl formamide, 1μl 1M NaOH and 4μl 25 x BXE). The DNA was denatured by heating to 100°C for 2.5 minutes before being examined by gel electrophoresis.

(vi) Gel Electrophoresis.

The DNA samples (2μl of the above reaction mix) were electrophoresed on vertical 6% w/v polyacrylamide gels containing 50% w/v urea in 1 x TBE (pH 8.3; see Section 3.1). Dimensions of gels used were 40cm long x 20cm wide x 0.25mm thick or 40cm x 40cm x 0.25mm. The gels were pre-electrophoresed for 15 minutes at 1500V before being loaded with the denatured DNA samples. Electrophoresis was continued at 1200V for 2 hours in a "short" run and for up to 6 hours in "long" runs. After electrophoresis, DNA was fixed in the gel by soaking the gel in 10% v/v acetic acid, 20% v/v methanol (or ethanol) for 15 - 20 minutes and dried in an oven at 80°C for 2 hours. Gels were then covered with plastic wrap and exposed to X-ray film for 18 - 24 hours at room temperature.

3.2.5 Northern analysis of RNA isolated from embryos, pouch young and adult bone marrow.

1. Isolation of adult bone marrow from *S. crassicaudata*.

Tibia and femur bones from adult animals were stripped of skin and muscle tissue and broken at each end. A 27 gauge needle and a syringe containing 200μl of PBS was used to flush out the marrow from each bone. RNA was then immediately extracted from the bone marrow suspension.

2. Extraction of RNA from pouch young, embryos and adult bone marrow.

(Chomczynski and Sacchi, 1987)

A whole animal in the case of pouch young and embryos, or bone marrow from adult animals was homogenized in 1ml solution D (4M guanidinium thiocyanate (Fluka 50990), 25 mM Sodium citrate (pH 7.0), 0.5% w/v sarcosyl, 0.1 M 2-mercaptoethanol: added just before use) in a 10ml polypropylene tube using a motor driven teflon homogenizer.

Sequentially 0.1ml of 2M sodium acetate (pH 4.0), 1ml of phenol (TE saturated) and 0.2ml of chloroform - isoamyl alcohol mixture (49:1) were added to the homogenate, with thorough mixing after each reagent was added. The mixture was left on ice for 15 minutes and centrifuged at 10,000g for 20 minutes at 4°C. 700μl of the aqueous phase was transferred to a 1.5ml polypropylene tube and RNA was precipitated by the addition of 700μl of isopropanol at -20°C for at least 1 hour. RNA was pelleted by centrifugation for 20 minutes in an eppendorf microfuge at 4°C. The pellet was dissolved in 0.3ml of solution D and RNA precipitated with 1 volume isopropanol at -20°C for 1 hour. After centrifugation for 15 minutes at 4°C in an eppendorf microfuge, the RNA pellet was vortexed in 75% ethanol, pelleted, dried *in vacuo* for 5 minutes and dissolved in 50μl H_2O at 65°C for 10 minutes. The RNA was stored at -70°C until use.

3. Formaldehyde agarose gel electrophoresis of RNA.

RNA samples were electrophoresed on formaldehyde-agarose gels prepared as follows: for 75ml of gel, 15ml of formaldehyde and 1.5ml of 0.5M Na-phosphate (pH 7.0) was added to 58.5ml H_2O containing 1.125g agarose, at 65°C. The gel "running buffer" was 10mM Na-phosphate (pH 7.0) which was recycled between anode and cathode tanks every 20 - 30 minutes. Before being loaded onto the gels RNA (5μl) was mixed with 2 volumes of denaturing solution (50% v/v formamide, 30% v/v formaldehyde, 10mM Na-phosphate (pH 7.0)), heated to 65°C for 10 minutes and stored on ice.

4. Northern transfer of RNA to Hybond N^+ nylon membranes.

RNA was transferred from formaldehyde agarose gels to Hybond N^+ membrane (Amersham) using a 0.05M NaOH transfer buffer and the protocol specified by the manufacturer.

5. Hybridization of Northern filters with radio-labelled DNA probes.

Pre-hybridization and hybridization conditions were identical to those used in Section 3.2.3.9. Filters were washed 2 times in 2 x SSC, 0.1% w/v SDS at room temperature and then washed under the conditions specified in Chapter 4. Filters

were covered in plastic wrap while still wet and exposed to X-ray film at -70°C for 2 - 7 days using an intensifying screen.

3.2.6 Amino-acid sequencing of purified α- and ß-globin polypeptides.

1. Isolation of *S. crassicaudata* globin protein from red blood cells.

(Thompson *et al.*, 1969).

Blood (300 - 400μl) was obtained from the the sub-orbital synus of adult animals. The cells were pelleted at 12,000g for 30 seconds in a microfuge, washed 3 times in 0.15M NaCl, and lysed by the addition of 2 volumes of H_2O and 0.5 volumes of CCl_4. Samples from 4 animals were pooled and centrifuged at 40,000g to remove cell debris. After centrifugation the supernatant was brought to 0.1M NaH_2PO_4 (pH 6.0) and haemoglobin precipitated by the addition of 10 volumes of acid acetone at -20°C (1.5ml of 10N HCl per 100ml acetone). The precipitated globin was centrifuged at 100g for 5 minutes and washed 4 times in cold (-20°C) acetone. The pellet was dried *in vacuo* and dissolved in 3ml of H_2O. After dialysis two times against a 1000-fold volume of H_2O the globin was freeze-dried and stored at -20°C.

2. Cellulose Acetate electrophoresis of globin.

(Ueda and Schneider, 1969).

Strips of cellulose acetate (Cellogel 500) were equilibrated with running buffer (4.66M Urea, 45mM Tris-HCl (pH 9.0), 0.5% v/v 2-mercaptoethanol) by slow shaking for 1 hour. Sample wells (0.8cm) were marked with a pencil near the middle of the gel and 3μl samples of globin protein dissolved in running buffer, were loaded. Electrophoresis was carried out at 250 V for 3 hours at 4°C. Gels were then stained with an Amido-black solution (0.5% w/v Amido black, 45% v/v methanol, 10% v/v acetic acid) for 5 minutes, followed by destaining in 47.5% methanol, 10% acetic acid for 30 minutes. Protein was fixed by soaking the cellogel for 30 minutes in 2% v/v glacial acetic acid, 20% v/v glycerol.

3. Ion exchange column chromatography of globin.

(Clegg *et al.*, 1966)

Globin samples were chromatographed on CM-sephadex C-50 columns in 8M urea buffers. Urea solutions were deionized by the addition of mixed bed resin (Biorad), then sodium phosphates (Na_2HPO_4 and NaH_2PO_4) were added to give the required Na^+ concentration of the start and limit buffers at pH 6.7. 2-mercaptoethanol (3.5ml/litre) was added just before running the column as the

absorbance of urea buffers containing 2-mercaptoethanol increases over time. The freeze-dried globin (20mg) was dissolved in 2ml of start buffer (0.04M Na-phosphate (pH 7.5), 8M Urea, 0.35% v/v 2-mercaptoethanol) and left on ice for several hours. The sample was loaded, without pH adjustment, on a 1.5cm x 10cm long column, containing CM-sephadex pre-equilibrated with start buffer, and was washed into the ion exchange medium with 2ml of start buffer. The column was exposed to a linear gradient using a peristaltic pump, with 150ml of start buffer and 150ml of a limit buffer (0.08M Na-phosphate (pH 7.5), 8M Urea, 0.35% v/v 2-mercaptoethanol). The effluent was monitored at a wavelength of 280nm and 3ml fractions were collected. Fractions corresponding to peak OD_{280} values were pooled and the globin chains carboxymethylated for 15 minutes using 0.94g iodoacetic acid (freshly dissolved in 5ml of 3M Tris) per 100ml of buffer; the mixture being adjusted to pH 8.4. The polypeptide chains were separated from reaction products and urea by dialysis against H_2O followed by ultra filtration using an "Amicon" ultrafiltration Cell Model 52, with a filter PM30 which has a molecular weight cut off of 30,000. Residual urea was removed by further dialysis against H_2O and the chains were recovered by freeze drying and stored at -20°C.

4. Amino-acid sequence determination of purified α- and ß-globin polypeptides from the N-terminal.

Amino acid sequencing was performed using the procedures described by Martin and Jennings (1983). Sequencing was done on a Beckman automatic sequencer 890c fitted with a cold trap. The protein (5mg) was dissolved in 0.5ml of formic acid (98%), loaded into a spinning cup, partially dried and then 3mg of polybrene in 0.5ml of H_2O was loaded and dried. The sequence program used was essentially the same as that of Hunkapiller and Hood (1978). This allows the processing of 14 residues per day. The anilino-thiazolinone derivatives taken from the sequencer were converted to phenylthiohydantoin derivatives (called PTH-amino acids) by heating under nitrogen at 65°C for 10 minutes in methanolic-HCl, chilling and drying in a stream of nitrogen. PTH-amino acids were identified with high-pressure liquid chromotography (HPLC) using a Waters instrument fitted with a Model 660 solvent programmer, a radial compression nodule, a Waters Radial-Pak (Resolve C_{18} 10μ) column, and monitored at 254nm and 313nm (the latter to detect PTH-dehydrothreonine). PTH-amino acids were dissolved in 100 μl of a mixture of equal parts of methanol and ethyl acetate and about 10 μl was injected into the Waters instrument. Solution A was 0.1 M Na-acetate at pH 6.0 and solution B was acetronitrile. For a standard run at 5ml per minute, the initial conditions were 17% v/v A changing over 9 minutes to 35% v/v using curve 5 on the programmer, followed by 6 minutes of isocratic elution. HPLC separated all

except 2 pairs of amino acids (PTH-valine and PTH-methionine, and PTH-phenylalanine and PTH-isoleucine). These were identified by thin layer chromatography (TLC) using solvent system V of Jeppsson and Sjoquist (1967).

3.2.7 Southern analysis of *S. crassicaudata* genomic DNA.

1. Agarose gel electrophoresis.

S. crassicaudata genomic DNA (10μg) was digested in 30μl reaction volumes (Section 3.2.3), mixed with 3μl of 10x loading buffer and directly loaded onto horizontal slab gels (15cm wide x 20cm long x 0.6cm thick) containing 0.8% w/v agarose in 1x TAE. A molecular weight standard (λ digested with HindIII; SPP-1 digested with EcoRI; BRL high molecular marker) was co-electrophoresed on each gel. Electrophoresis was carried out at 25 Volts for 24 hours and DNA visualized by staining the gel in ethidium bromide and examination under UV (254nm) light.

2. Southern transfer of genomic DNA to hybond N^+.

DNA was transferred to hybond N^+ using the alkali transfer procedure as specified by the manufacturer; Amersham. The DNA was first hydrolysed by treatment with 0.25M HCl for 15 minutes and transferred to Hybond N^+ using a

0.4M NaOH transfer solution. Transfer was carried out for 3 - 4 hours and filters were rinsed briefly in 2 x SSC and stored at -20°C before use.

3. Hybridization of genomic Southern membranes with radio-labelled DNA probes.

Nucleic acid probes were radio-labelled using the "oligo-labelling procedure" (Section 3.2.3.6), denatured and hybridized with Hybond N^+ filters at 65°C using the conditions described in Section 3.2.3.9. To avoid background problems it was ensured that the probe concentration did not exceed 20 ng/ml. After hybridization, filters were washed two times in 2 x SSC, 0.1% w/v SDS at room temperature for 10 minutes, followed by two washes each for 30 minutes in 1 x SSC, 0.1% w/v SDS at 65°C. Wet filters were covered with plastic wrap and exposed to X-ray film in the presence of an intensifying screen at -70°C for 2 to 14 days. The sizes of fragments showing hybridization with nucleic acid probes were estimated by comparing their mobility with known standards. Filters were stripped of nucleic acid probes by incubating the filter in a solution containing 0.5% w/v SDS at 100°C for 10 minutes.

3.2.8 Relative rate tests of marsupial and eutherian β-like globin genes.

Rates of non-synonymous substitutions per site at non-degenerate and two-fold degenerate sites, and synonymous substitutions per site at two-fold and four-fold degenerate sites, were estimated using the procedure of Li *et al.* (1985). Relative rate tests were carried out using the procedure of Wu and Li (1985).

3.2.9 Phylogenetic tree constructions.

Phylogenetic trees were constructed using the "Protpars" program of Phylip (version 3.2) available from Joseph Felsenstein, University of Washington: USA. Full details of this program are described in the documentation of Phylip. The program infers the most parsimonious unrooted trees directly from protein sequences by counting the number of nucleotide substitutions required to evolve the protein in each phylogeny. These numbers are estimated as follows:-

i) any changes of amino acid are consistent with the genetic code;

ii) changes between two amino acids via a third are allowed and counted as two changes;

iii) synonymous changes are not counted.

The procedure is based on five assumptions common to parsimony methods (taken directly from the documentation of Protpars (version 3.2)):

i) change in different sites is independent.

ii) change in different lineages is independent.

iii) the ancestral amino acid state at any time is not known.

iv) the probability of a base substitution that changes the amino acid sequence is small over the length of time involved in a branch of the phylogeny.

v) the rates of change in different branches of the phylogeny do not vary by so much that two changes in a high-rate branch are more probable than one change in a low rate branch.

The output of the program consists of a list of the most parsimonious unrooted trees, a table of the number of changes of state or amino acid substitutions required at each site of the protein and a table for each branch giving "the states which are inferred to have been at the top end of the branch". If the inferred state was unknown (designated: "?") there were "multiple equally-parsimonious assignments of states".

CHAPTER 4

RESULTS

4.1 Isolation and characterization of a β-like globin gene from a *S.crassicaudata* genomic DNA library.

4.1.1 Cloning strategy.

In order to isolate genomic DNA clones of β-like globin genes from *S.crassicaudata* the following cloning strategy was chosen:

1. A bacteriophage genomic DNA library was constructed from partially digested restriction fragments of *S.crassicaudata* DNA. This was achieved prior to the commencement of the research presented in this thesis (Cooper, 1985).

2. A cDNA plasmid clone (pDG-5), isolated from the dasyurid marsupial *D.viverrinus*, was characterized by restriction mapping and DNA sequencing, and shown to contain a partial cDNA copy of adult β-globin RNA. This clone was used as a probe to screen the above genomic DNA library using the "plaque lifting" procedure of Benton and Davis (1977).

4.1.2 Construction of a *S.crassicaudata* genomic DNA library in the bacteriophage EMBL3.

To facilitate the isolation of genomic DNA clones of β-like globin genes from *S.crassicaudata*, a genomic DNA library was constructed in the bacteriophage EMBL3 (Frischauf *et al.*, 1983). EMBL3 is a lambda replacement vector, with a nominal cloning capacity of 9 to 23 kb, derived by modification of λ1059 (Karn *et al.*, 1980). Two BamHI restriction sites located either side of the phage stuffer region in λ1059 were replaced by poly-linker sequences containing SalI, BamHI, and EcoRI restriction sites. Digestion of EMBL3 with BamHI and EcoRI and removal of the poly-linker sequence, leaves the phage stuffer region incapable of re-ligating to the bacteriophage arms (Frischauf *et al.*, 1983). *S.crassicaudata* genomic DNA was partially digested with MboI, to obtain fragments in the size range 15 - 20 kb. This DNA was then dephosphorylated by treatment with calf intestinal alkaline phosphatase and directly cloned into the BamHI site of EMBL3. Recombinant bacteriophage were packaged *in vitro* and amplified on a P2 lysogenic *E.coli* host NM539. This step selects against re-ligated Spi^+ phage and therefore selects directly for phage with a genomic DNA insert. The insert can be removed from the bacteriophage by digestion with the endonuclease SalI. The library of recombinant bacteriophage particles was amplified and stored at 4 °C. The titre of the library was approximately 400,000 pfu/ml in August, 1985.

The advantage of using partially digested genomic DNA in the construction of

this library, is that a set of overlapping phage clones can be isolated extending many kilobases from the gene of interest. This makes it possible to study the organization of closely linked genes as well as nearby DNA sequences which may be important in gene regulation.

4.1.3 Characterization of the *D.viverrinus* cDNA plasmid clone, pDG-5.

Before a β-globin gene can be isolated from a genomic DNA library, it is necessary to have a radioactively labelled nucleic acid probe, sufficiently homologous to the gene, to enable hybridization and detection by autoradiography.

Wainwright (1984) synthesized cDNA copies of reticulocyte poly(A) mRNA from the dasyurid marsupial *D.viverrinus*. The cDNAs were cloned by dG-dC tailing and joining into the Pst I cleavage site of the bacterial plasmid pBR322. Restriction enzyme analysis, hybrid arrest translation experiments and DNA sequence analysis were used to identify two clones, pDG-5 and pDG-73, containing cDNA copies of β-globin and α-globin mRNA respectively. The pDG-5 clone was shown to contain a cDNA copy of a portion of the 2nd exon and complete 3rd exon, the 3' non-coding region and the poly (A) tail of adult β-globin mRNA.

Prior to the commencement of the research reported in this thesis, pDG-5 was the only available marsupial β-globin clone. It was therefore chosen as a probe to screen the *S.crassicaudata* genomic DNA library. Before using this clone as a probe, however, it was necessary to verify the results of Wainwright (1984).

pDG-5 was amplified and purified using a large-scale plasmid preparation procedure (Section 3.2.2). The PstI cDNA insert of pDG-5 was isolated from a low melting temperature (LMT) agarose gel and digested singly with the restriction enzymes Sau3A, RsaI, EcoRI, HpaII, SacI and HinfI, and double digested with RsaI/Sau3A, SacI/Sau3A, and SacI/HinfI. Restriction fragments were separated by electrophoresis on 1.5% w/v agarose and 6% w/v poly-acrylamide gels and their sizes were estimated by comparing their mobility to known standards (gels not shown). A restriction map of the pDG-5 PstI insert was deduced from these fragment sizes (Fig. 4.1.1.a). A restriction map was also deduced from the DNA sequence of the pDG-5 insert reported by Wainwright (1984) (Fig. 4.1.1.b). The above two restriction maps each contained single SacI, EcoRI, Sau3A and HinfI sites and two RsaI sites located at equivalent positions. However, the size of the PstI fragment of map a) was at least 50 bp shorter than the PstI fragment of map b) (Fig. 4.1.1a). This indicated that the pDG-5 clone, characterized in this study (henceforth referred to as pDG-5.2), contained a cDNA copy of β-globin mRNA, but appeared to be missing a portion of the 3' non-coding region and the poly (A) tail of the mRNA.

To provide further evidence that pDG-5.2 contained a cDNA copy of adult β-globin mRNA a partial DNA sequence of the clone was determined. The PstI insert of pDG-5.2 was digested with Sau3A to produce two DNA fragments of size 210 bp and 190 bp, each with a PstI and Sau3A cohesive end. The fragments were cloned into the PstI / BamHI site of the M13 bacteriophage MP19. Single stranded M13 DNA was prepared from seven colourless plaques and sequenced using the "dideoxy chain

termination" procedure (see section 3.2.4 for full details of the above procedures). DNA sequence data from each of the seven M13 clones indicated they all contained an identical Sau3A / PstI insert. A partial DNA sequence of pDG-5.2 was determined and was found to be identical in each of the seven M13 clones.

This sequence was compared to the DNA sequence of pDG-5 reported by Wainwright (1984) (Fig. 4.1.1b). The two sequences differed at eight nucleotide sites out of a total of eighty nucleotide sites determined. Four of the eight differences were at synonymous sites and therefore did not alter the conceptual amino acid sequence of the *D. viverrinus* β-globin polypeptide deduced by Wainwright (1984). The remaining four nucleotide site differences led to two amino acid replacements (Thr -> Trp and Thr -> Leu, see Fig. 4.1.1b).

The restriction map and DNA sequence comparisons described above indicate that pDG-5.2 contains a cDNA copy of β-globin mRNA, but is not identical to the cDNA β-globin clone (pDG-5) isolated and sequenced by Wainwright (1984). pDG-5.2 was derived from a β-globin mRNA species present in adult *D. viverrinus* 10S RNA, which suggests that this mRNA was transcribed from an adult β-globin gene. Either two adult β-globin genes exist in *D. viverrinus* or the nucleotide sequence differences of the two clones (pDG-5.2 and pDG-5) can be attributed to polymorphic variation at a single adult β-globin locus. Alternatively, there have been experimental errors in the determination of the DNA sequence of pDG-5 reported by Wainwright (1984).

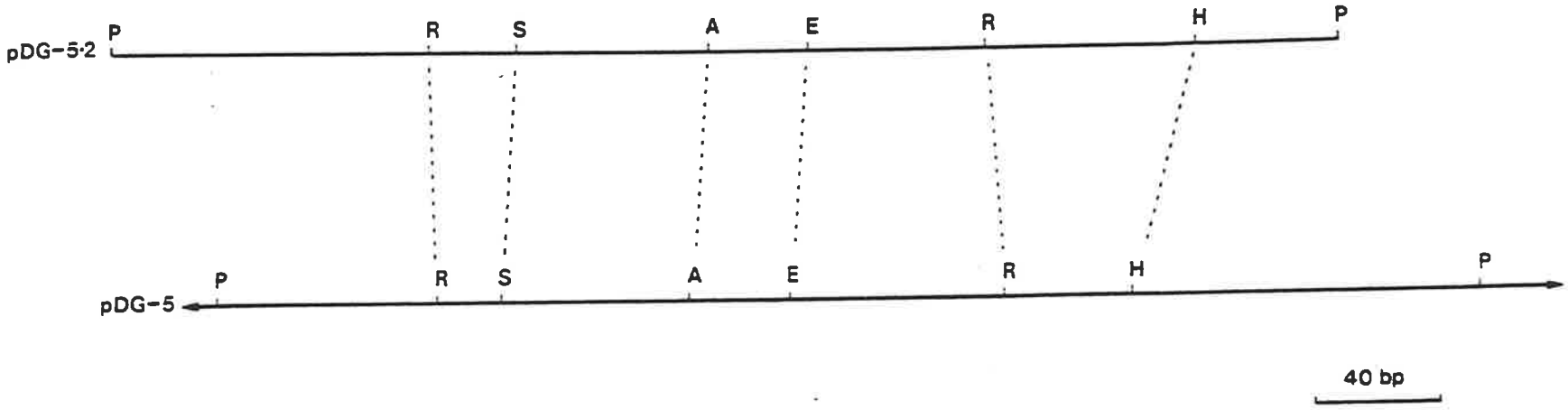
Fig. 4.1.1

A Comparison of the restriction map of pDG-5.2 with the restriction map of pDG-5 inferred from nucleotide sequence data (Wainwright, 1984).

Enzymes mapped are P - *PstI*; R - *RsaI*; S - *SacI*; A - *Sau3A*; H - *HinfI*. Dotted lines join sites which approximately coincide in the two maps.

B Comparison of the partial DNA sequence of pDG-5.2 with the DNA sequence of pDG-5 reported by Wainwright (1984). The two sequences conceptually translate into identical amino acid sequences, except at the two sites shown. The nucleotide sequence of pDG-5 is indicated only where it differs from that in pDG-5.2.

A



B

```
                                                                                                  trp
pDG-5.2     ATC TGC CTT GCT GAG CAC TTT GGC AAA GAA TTC ACC CCT GAG GTT CAG GCT GCC TGG CAG AAA
pDG-5                  G           T                                                  ACC       G
                                                                                      thr

            leu
pDG-5.2     CTT GTG GCT GGT GTG GCC
pDG-5       ACT
            thr
```

4.1.4 Isolation of the phage clone λSG-3 from the *S.crassicaudata* genomic DNA library.

To avoid background hybridization problems from dG-dC tails and pBR322 vector sequences in pDG-5.2, an 185 bp RsaI fragment, containing DNA sequences homologous to the 2nd and 3rd exons of a *D.viverrinus* adult β-globin gene, was used to probe the *S.crassicaudata* genomic library. Approximately 200,000 plaque forming units (pfu's) from the library were plated onto 30 cm x 30 cm square petri dishes (100,000 pfu's / plate) and grown for approximately 16 hours until the plaques were almost touching. The plaques were transferred to nitrocellulose filters and the filters were then baked for 2 hours at 80 °C, treated with a Denhardt's solution at 65 °C before pre-hybridization at 42 °C in a solution containing 50% v/v formamide. Full details of these procedures are described in section 3.2.1.

The 185 bp RsaI fragment of pDG-5.2 (500 ng) was purified from a LMT agarose gel, radio-labelled by nick translation to a specific activity of 6 x 10^7 cpm/μg, and hybridized with the above filters at a concentration of 2.4 ng/ml. Hybridization was carried out for 24 hours at 42 °C, and the filters were washed to a stringency of 2 x SSC, 0.1% w/v SDS at 65 °C and autoradiographed for 2 days.

A total of seven plaques showed possible hybridization to the pDG-5.2 probe and each were isolated from the original phage plates in plaque stabs containing up to 50 plaques. The plaque stabs were resuspended in SM buffer and the seven bacteriophage suspensions were re-plated at a low density (approximately 500

plaques/8.5 cm petri dish). Bacteriophage DNA was transferred to nitrocellulose and probed with the PstI cDNA insert of pDG-5.2 using the same procedures and conditions described above. One filter showed some positively hybridizing plaques (Fig. 4.1.2.a). The remaining six nitrocellulose filters failed to show any positive hybridization to the pDG-5.2 probe. This suggested that six out of the seven plaque stabs were most probably isolated from regions of the bacteriophage plates showing background hybridization to the probe, due to the low stringency conditions. Single plaques showing hybridization to pDG-5.2 were isolated, re-plated at low density, plaque lifted and re-probed with pDG-5.2, until 100% of the plaques on the plate showed positive hybridization to pDG-5.2 (Fig. 4.1.2b).

4.1.5 Partial characterization of the phage clone λSG-3.

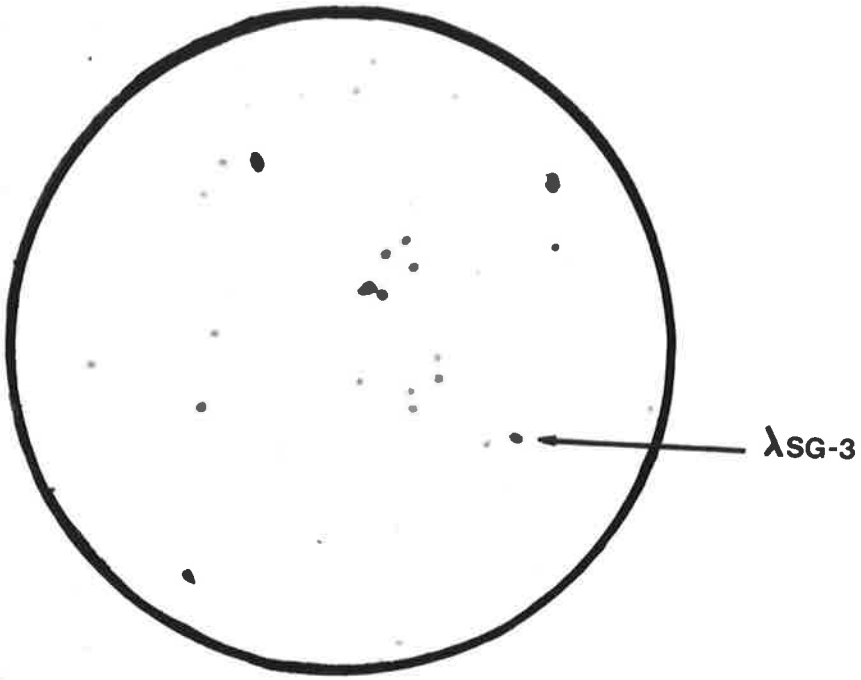
Bacteriophage DNA was purified from a single plaque, showing positive hybridization to pDG-5.2, using the plate lysate procedure described in section 3.2.1. The bacteriophage DNA (henceforth referred to as λSG-3) was digested with the enzymes SacI, PstI, HindIII, BamHI, EcoRI and SalI, and DNA fragments were separated by electrophoresis on 0.4% w/v and 1.5% w/v agarose gels using standard procedures. DNA fragment sizes were estimated by comparison with known standards and a partial restriction map of λSG-3 was determined (Fig. 4.1.3). The position of one HindIII and two EcoRI sites were not determined at this time.

Fig. 4.1.2

A: Autoradiograph of a plaque lift showing positive hybridization with the PstI insert of pDG-5.2. A single plaque (λSG-3) was isolated, replated at low density, plaque lifted and re-probed with pDG-5.2.

B: Autoradiograph of a plaque lift after two rounds of the above purification procedure, showing positive hybridization of the pDG-5.2 probe with 100% of all plaques.

A



B

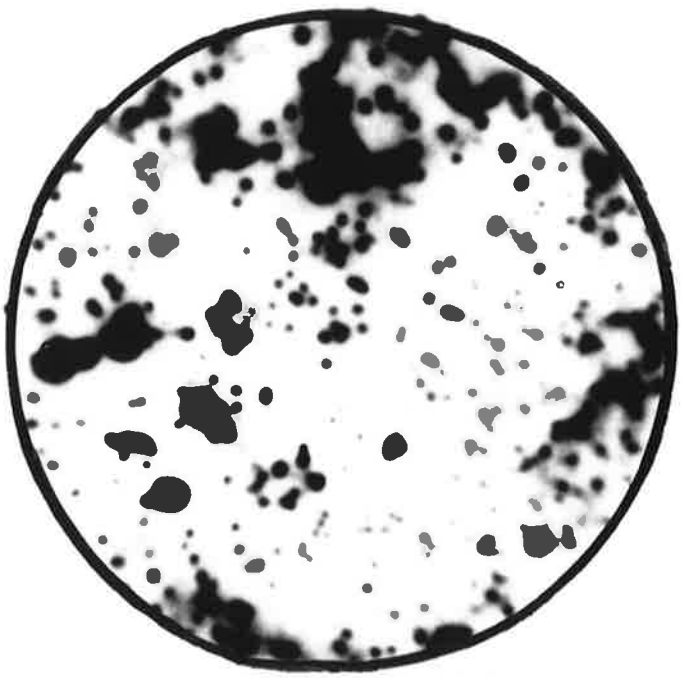


Fig. 4.1.3

Partial restriction map of λSG-3 showing the regions of DNA which were subcloned into pBR322 (pSG-2 and pSG-12). Enzymes mapped are: S:- *SalI*; H:- *HindIII*; B:- *BamHI*; E:- *EcoRI*; C:- *SacI*. The thick lines designate the EMBL3 left and right arms which are not drawn to scale. The locations of at least two *EcoRI* sites were not determined for this map, but are shown in a more detailed restriction map of the plasmid subclone pSG-12.

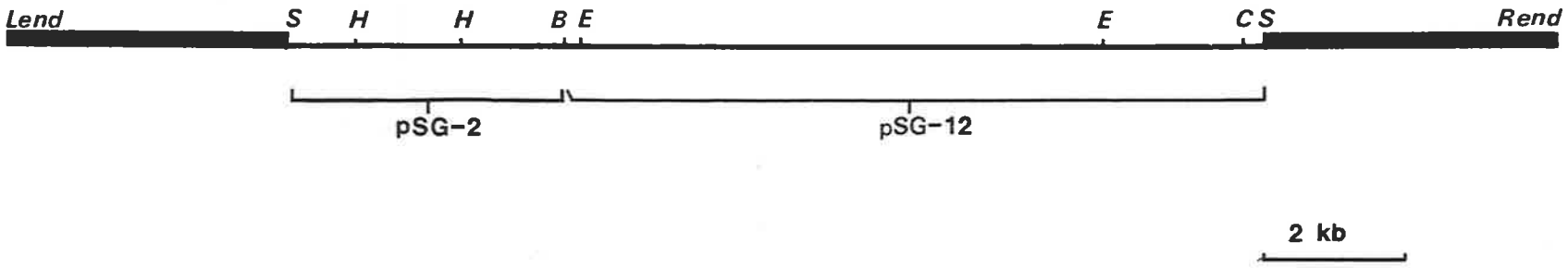

Lend
S
H
H
B
E
E
C
S
Rend
pSG-2
pSG-12
2 kb

Fig. 4.1.4

A: Restriction analysis of λSG-3.

Restriction fragments were separated on a 0.4% w/v agarose gel (12.5cm long) by electrophoresis at 15V for 20 hours.

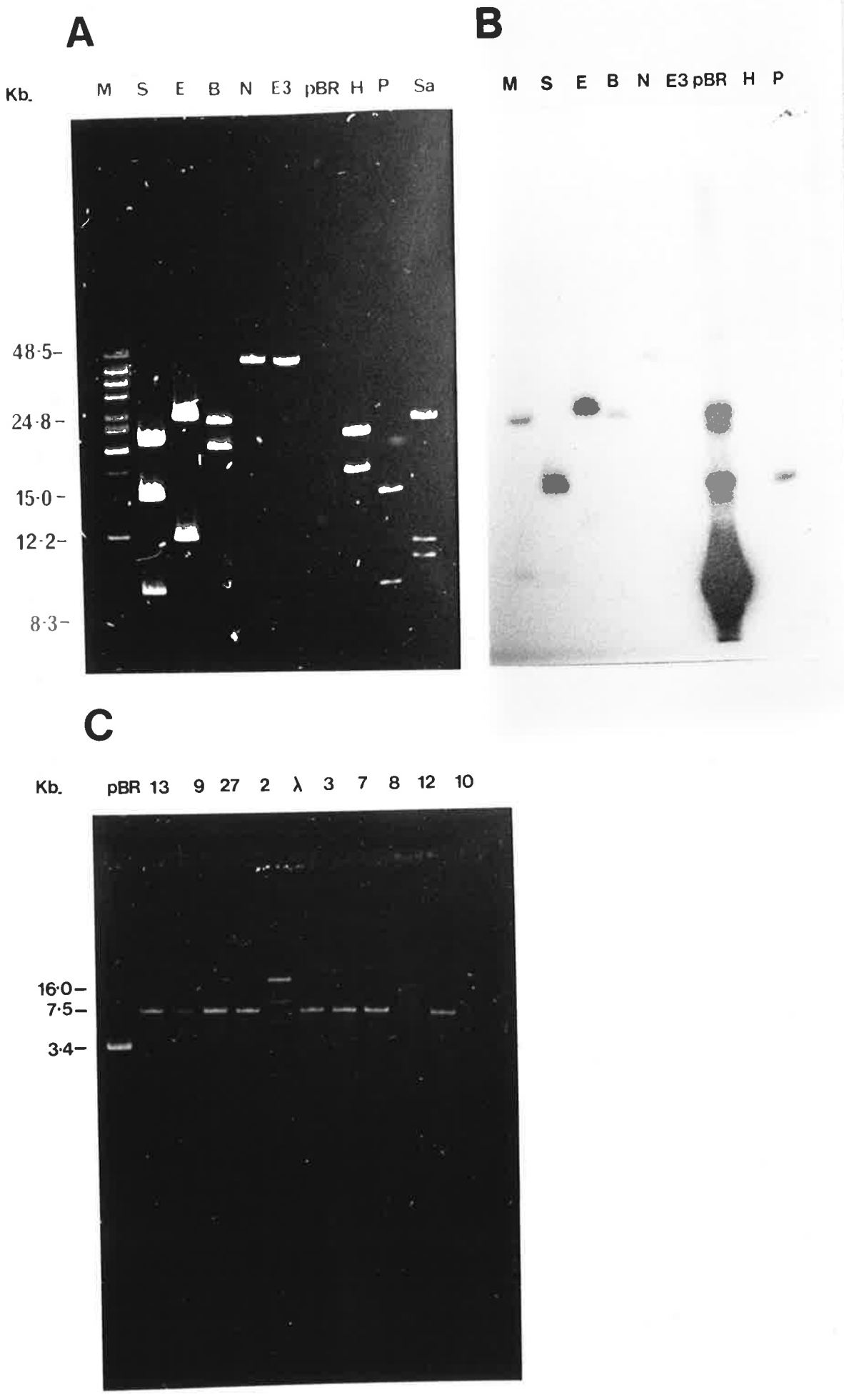
Lanes are: M - high molecular weight marker (BRL)
S - λSG-3 digested with *SalI*
E - λSG-3 digested with *EcoRI*
B - λSG-3 digested with *BamHI*
N - λSG-3 undigested
E3 - EMBL3 DNA undigested
pBR - pBR322 DNA undigested
H - λSG-3 digested with *HindIII*
P - λSG-3 digested with *PstI*
Sa - λSG-3 digested with *SacI*

B: Southern analysis of λSG-3.

DNA fragments in the above gel were Southern transferred to nitrocellulose, hybridized with radio-labelled pDG-5.2 (whole plasmid) and autoradiographed overnight. The lanes are as specified above. Note strong hybridization of the probe with the 15kb *SalI* genomic insert of λSG-3, and an approximately 24kb *BamHI* and *EcoRI* fragment.

C: Restriction analysis of plasmid subclones.

Plasmid DNA was purified using the mini-plasmid preparation procedure (Section 3.2.2.4) and 10μl aliquots of the DNA were digested with *BamHI*. Fragments were separated by electrophoresis on a 1% w/v agarose gel (12.5cm long). Numbers refer to individual subclones: note pSG-2 which is approximately 7.5kb and pSG-12 which is approximately 16kb in size. These clones were selected for further analysis. Molecular weight markers are pBR:- pBR322 (linear) and λ:- λ DNA digested with *HindIII*.


A
Kb.
M S E B N E3 pBR H P Sa
48·5–
24·8–
15·0–
12·2–
8·3–
B
M S E B N E3 pBR H P
C
Kb.
pBR 13 9 27 2 λ 3 7 8 12 10
16·0–
7·5–
3·4–

DNA fragments from the 0.4% w/v gel (Fig. 4.1.4a) were Southern transferred to nitrocellulose, and probed with pDG-5.2 using standard hybridization conditions and washings at a stringency of 2 x SSC, 0.1% w/v SDS (Section 3.2.3). Hybridization of pDG-5.2 was localized to a 24 kb BamHI and a 24 kb EcoRI fragment (Fig. 4.1.4b). The EMBL3 left arm encompasses 19.4 kb in each of the 2 fragments, indicating that the region hybridizing pDG-5.2 was located on an approximately 4.5 kb SalI/ BamHI or SalI / EcoRI fragment adjacent to the EMBL3 left arm.

4.1.6 Generation of the subclones pSG-2 and pSG-12.

To determine a more detailed restriction map of λSG-3 and further characterize the region of this clone which hybridized to pDG-5.2, it was necessary to make plasmid subclones of the SalI insert of λSG-3. The bacterial plasmid pBR322 was chosen as the vector molecule for the subcloning step.

Subcloning strategy.

When λSG-3 was digested with BamHI and SalI, a total of four fragments resulted: a 19.4 kb fragment containing the EMBL3 left arm, a 9.2 kb fragment containing the EMBL3 right arm, a 4 kb SalI / BamHI fragment and a 10 kb SalI / BamHI fragment. Of these fragments, only the 4 kb and 10 kb BamHI/ SalI fragments could be ligated into the BamHI and SalI sites of pBR322. The 4 kb

fragment, which hybridized to pDG-5.2, would be expected to preferentially ligate into pBR322 because of its size. A sub-cloning strategy was therefore chosen, involving the digestion of λSG-3 with BamHI and SalI and the direct cloning of the resultant fragments, without purification, into the single BamHI and SalI sites of pBR322. The BamHI and SalI sites of pBR322 are located in the tetracycline resistance gene and therefore foreign DNA cloned into these sites would confer tetracycline sensitivity and ampicillin resistance to a host bacterial cell.

Subcloning results.

λSG-3 phage particles were amplified using a liquid lysate procedure (Section 3.2.1). A ratio of 10^{10} NM538 cells to 5 x 10^8 pfu's in 500 ml pre-warmed L-broth supplemented with 20 mM MgSO4, gave high titre confluent phage cultures. The phage particles were concentrated using PEG 6000 and phage DNA was purified using an adaption of the plate lysate procedure (Section 3.2.1). In this technique, the phage DNA was spooled after the isopropanol precipitation step, dissolved in TE (pH 8.0) and precipitated a second time with ethanol. The combination of this DNA isolation procedure with the liquid lysate procedure for amplifying EMBL3 phage particles, gave high yields and DNA of sufficient purity for restriction enzyme analysis. The use of a "Nensorb" column (Du Pont) to further purify the phage DNA, gave DNA of sufficient purity for cloning. In contrast, the plate lysate amplification procedure gave low yields and poor quality EMBL3 DNA; in most cases such DNA was incapable of digestion with restriction enzymes.

Native pBR322 DNA (10 ng), double digested with BamHI and SalI, was ligated to 100 ng of "Nensorb" purified λSG-3 DNA double digested with BamHI and SalI, in a total volume of 20 μl of 1 x ligation buffer. As a control, 10 ng of pBR322 double digested with BamHI and SalI was ligated to itself. Half of each ligation mix was used to transform competent MC1061 bacteria. Included in the transformation experiment was a control in which no DNA was added to the competent cells. The transformed cells were plated onto L-agar containing ampicillin and grown overnight at 37 °C. Approximately 1000 bacterial colonies were observed on the plate containing MC1061 transformed with the recombinant pBR322. Only 20 colonies were observed on one control plate (MC1061 transformed with the re-annealed pBR322), indicating a low background of colonies with native pBR322, and no colonies were observed on the second control plate (un-transformed MC1061).

One hundred and forty three colonies were picked and replica plated onto duplicate tetracycline and ampicillin L-agar plates, marked with a grid. After an overnight incubation, results indicated that 140 of the 143 colonies were both tetracycline sensitive (tet^S) and ampicillin resistant (amp^R). Nine amp^R tet^S colonies were streaked out onto ampicillin plates to obtain single colonies. Isolated colonies were picked and used to prepare plasmid DNA, from each of the nine clones, using the "mini-plasmid" preparation procedure. The plasmid DNA's were digested with BamHI and electrophoresed on a 1% w/v agarose gel (Fig. 4.1.4c). All nine clones were larger than linear pBR322 (4362 bp), indicating that they contained inserts. Eight of the plasmid clones were of an identical size (approximately 7.5 kb), which

suggested they possibly contained the 4 kb BamHI / SalI fragment of λSG-3. One of these plasmids, pSG-2, was chosen for further studies. One out of the nine plasmid clones, pSG-12, was approximately 14 kb in size which suggested that it possibly contained the 10 kb BamHI/ SalI fragment of λSG-3.

4.1.7 Characterization of the subclone pSG-2.

Plasmid DNA from the clone pSG-2 was prepared using a large scale plasmid purification procedure, and was digested with a wide range of restriction endonucleases, including HindIII, BamHI, RsaI, Sau 3A, HpaII, PstI, SalI, EcoRI, SacI, KpnI, PvuII, ClaI, XhoI, and BglII. DNA fragments were separated by electrophoresis on agarose and poly-acrylamide gels (Fig. 4.1.5). Fragment sizes were estimated by comparison with known standards, and a restriction map of pSG-2 was derived (Fig. 4.1.6). Double and triple digestions were performed in some cases to determine the location of restriction sites. To localize more frequently occurring restriction sites such as Sau3A and RsaI, the four HindIII fragments (of length: 4.8 kb, 1.6 kb, 0.94 kb and 0.46 kb) of pSG-2 were purified from a 1.5% w/v LMT agarose gel, and both singly and double digested with RsaI and Sau3A. Fragments were separated by electrophoresis on a 6% w/v acrylamide gel and a 2% w/v agarose gel, using pBR322 digested with HpaII as a known standard. The approximate locations of Sau3A and RsaI sites were determined for three of the four HindIII fragments, but were not precisely determined for the 4.8 kb HindIII fragment. No

restriction sites for EcoRI, ClaI and XhoI were detected in the insert (Fig. 4.1.6).

The purified HindIII fragments of pSG-2 were digested with RsaI and Sau3A, electrophoresed on a 1% w/v agarose gel and Southern transferred to nitrocellulose (Section 3.2.3). The filter was probed with the PstI cDNA insert of pDG-5.2 using standard hybridization conditions at 42 °C, and washed to a stringency of 1 x SSC, 0.1% w/v SDS. The results are shown in Fig. 4.1.7. The 460 bp HindIII fragment of pSG-2 hybridized strongly to the pDG-5.2 probe. Faint hybridization of the probe was also observed on the 4.8 kb HindIII fragment of pSG-2, while no hybridization was observed on the 940 bp HindIII fragment which lies between the 460 bp and 4.8 kb HindIII fragments (see restriction map of pSG-2 in Fig. 4.1.6).

Given that eutherian β-globin genes generally have 2nd introns in the size range 500 - 900 bp, the Southern results suggested that the 940 bp HindIII fragment may lie in the 2nd intron of a β-globin gene, with the 460 bp fragment containing the 1st and 2nd exons and the 1.2 kb SalI / HindIII fragment containing the 3rd exon of the gene. Alternatively, the 460 bp HindIII fragment could contain the 3rd exon, with the 1st and 2nd exons being located in the 1.2 kb SalI / HindIII fragment. Only a DNA sequence analysis of the subclone pSG-2 could unequivocally distinguish between these possibilities.

Fig. 4.1.5

Restriction analysis of the plasmid subclone pSG-2.

A: Poly-acrylamide gel electrophoresis of pSG-2 restriction fragments (Section 3.2.3.4).

Lanes are:
1. pSG-2 double digested with *HindIII* and *PstI*
2. pSG-2 digested with *HpaII*
3. pBR322 digested with *HpaII*
4. pSG-2 double digested with *HpaII* and *RsaI*
5. pSG-2 digested with *Sau3A*
6. pSG-2 double digested with *Sau3A* and *RsaI*
7. pSG-2 digested with *RsaI*

B: Poly-acrylamide gel electrophoresis of pSG-2 restriction fragments.

pSG-2 DNA was digested with *HindIII* and the 4 resultant restriction fragments of length: 4.84kb, 1.6kb, 0.94kb and 0.46kb; were purified from a 1.5% w/v LMT gel (Section 3.2.3.5). Each of the above *HindIII* fragments was digested with *RsaI* and *Sau3A* and fragments were separated by electrophoresis on a 6% w/v poly-acrylamide gel (Section 3.2.3.4). H/R, H/S and H designate *HindIII/RsaI*; *HindIII/Sau3A* and *HindIII* digests respectively. Numbers are 1:- 4.84kb *HindIII* fragment; 2:- 1.6kb *HindIII* fragment; 3:- 0.94kb *HindIII* fragment; 4:- 0.46kb *HindIII* fragment; 5:- total pSG-2 DNA. P and K designate *PvuII* and *KpnI* digests of total pSG-2 DNA.

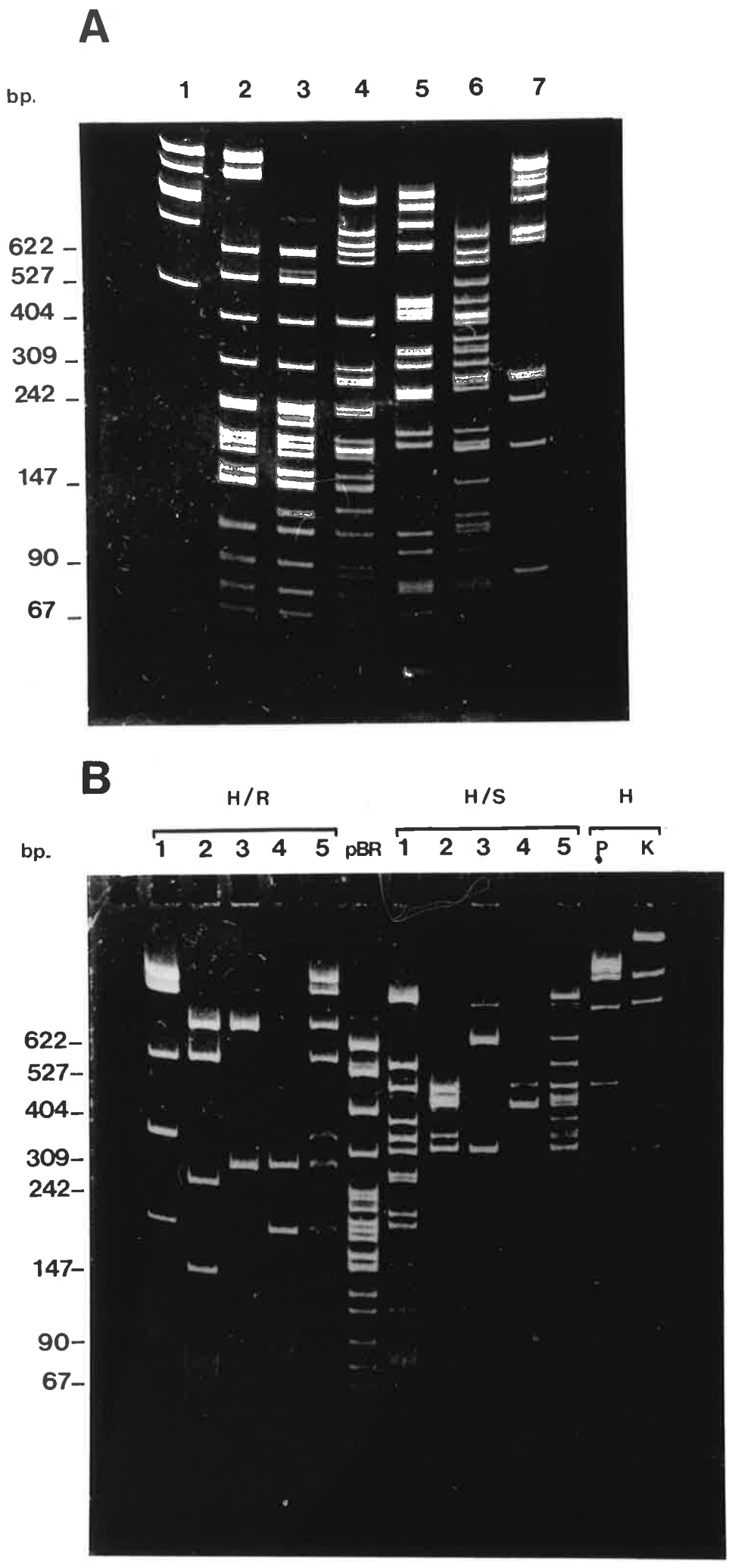

A
bp.
1 2 3 4 5 6 7
622
527
404
309
242
147
90
67
B
H/R
H/S
H
bp.
1 2 3 4 5 pBR 1 2 3 4 5 P K
622
527
404
309
242
147
90
67

Fig. 4.1.6

Restriction map of the genomic insert of the plasmid pSG-2.

Enzymes mapped are: S:- *SalI*; P:- *PstI*; A:- *Sau3A*; R:- *RsaI*; H:- *HindIII*; Bg:- *BglII*; Pv:- *PvuII*; K:- *KpnI*; C:- *SacI*; Hp:- *HpaII*; B:- *BamHI*. No *EcoRI*, *ClaI*, *XhoI* and *XbaI* sites were found in the genomic insert of pSG-2. Fragment sizes are in kb.

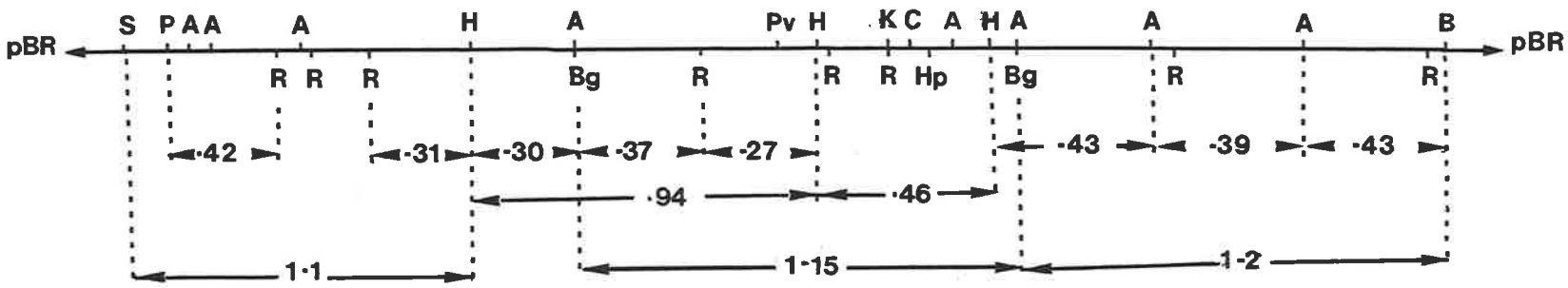

pBR
S
P
A
A
A
H
A
Pv
H
K
C
A
H
A
A
A
B
pBR
R
R
R
Bg
R
R
R
Hp
Bg
R
R
·42
·31
·30
·37
·27
·43
·39
·43
·94
·46
1·1
1·15
1·2

Fig. 4.1.7

A: Restriction analysis of pSG-2.

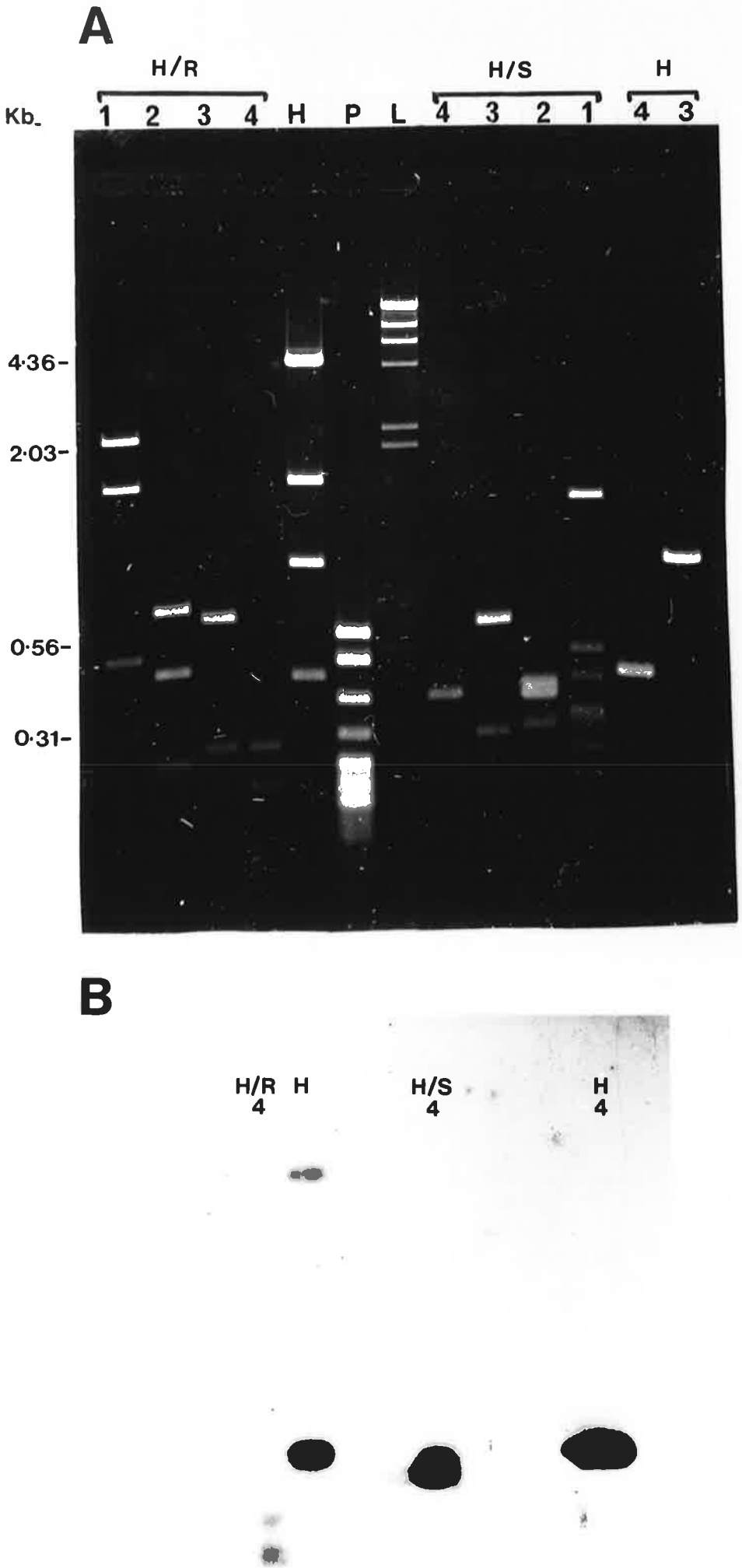
Restriction fragments were separated on a 1.5% w.v agarose gel (12.5cm long) by electrophoresis at 75V for 3 hours.

H/R and H/S designate *RsaI* and *Sau3A* digests respectively of LMT purified *HindIII* restriction fragments: 1:- 4.84kb *HindIII* fragment; 2:- 1.6kb *HindIII* fragment; 3:- 0.94kb *HindIII* fragment and 4:- 0.46kb *HindIII* fragment.

Other lanes are as follows: H:- pSG-2 total DNA digested with *HindIII*
P:- pBR322 digested with *HpaII*
L:- λ digested with *HindIII*
H4:- 0.46kb *HindIII* fragment of pSG-2
H3:- 0.94kb *HindIII* fragment of pSG-2

B: Southern analysis of pSG-2.

DNA fragments from the above gel were Southern transferred to nitrocellulose, hybridized with the *PstI* insert of pDG-5.2 (radio-labelled by nick translation) and autoradiographed overnight. Note strong hybridization of the probe to the 0.46kb and 4.84kb *HindIII* fragments of pSG-2.


A
H/R
H/S
H
Kb.
1 2 3 4 H P L 4 3 2 1 4 3
4·36-
2·03-
0·56-
0·31-
B
H/R H
4
H/S
4
H
4

4.1.8 DNA sequencing strategy.

A sequencing strategy was chosen, in which available restriction sites in pSG-2 were used to generate subclones in the bacteriophage M13, and these subclones were sequenced using the "dideoxy chain termination" procedure (Sanger *et al.*, 1980). As restriction sites in pSG-2 were located at most every 500 bp apart, the entire genomic insert of pSG-2 could be sequenced on both strands using this strategy.

The M13 vectors MP18 and MP19 were used as cloning vectors for the generation of single stranded templates. Both vectors are 7250 bp in length and contain a multiple cloning region composed of a wide range of restriction sites including EcoRI, BamHI, HindIII, SalI and SmaI (Messing, 1983). These restriction sites are located in the β-galactosidase gene of MP18 and MP19, and therefore the insertion of foreign DNA fragments into these sites leads to a colour change of recombinant M13 plaques from blue to colourless, when growth medium contains the compounds X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, a chromogenic substrate) and IPTG (isopropylthio-β-D-galactoside, a gratuitous inducer of β-galactosidase) (Sambrook *et al.*, 1989).

4.1.9 Preparation of pSG-2 subclones in the M13 vectors MP18 and MP19.

pSG-2 DNA was double digested with HindIII and SalI, and fragments were isolated and purified from a 1.5% w/v LMT agarose gel. Each fragment was further

digested with a range of enzymes chosen to produce fragment sizes in the range 200 - 400 bp, and capable of ligation into the multiple cloning region of MP18 and MP19.

Each of MP18 and MP19 replicative form (RF) DNA, was digested with the appropriate restriction enzyme to produce compatable 3' or 5' overhangs for ligation to the above restriction fragments. Where Sau3A was used to generate pSG-2 inserts, BamHI was used to digest MP18 or MP19, in order to produce compatible cloning sites. Fragments generated by digestion with RsaI, which produces a "blunt" ended DNA molecule, were ligated into the SmaI site of MP18 or MP19.

A 3:1 molar excess of the restriction fragment to be cloned and 10 ng of the appropriate M13 vector (Mp18 or Mp19) were ligated, and then used to transform *E.coli* JM101. Transformed bacteria were plated onto petri dishes containing minimal medium supplemented with glucose, BCIG and IPTG. Single stranded M13 templates were prepared from colourless plaques and the DNA sequences of the pSG-2 inserts were determined (for full details of these procedures see Section 3.2.4).

4.1.10 The DNA sequence of a β-like globin gene in pSG-2.

DNA sequence determined from the 460 bp HindIII fragment of pSG-2 was compared to the mouse adult (βmaj) and embryonic (βh1) β-like globin genes and showed strong homology to the first and second exons of these genes (Fig. 4.1.8). The presence of the highly conserved globin intron 5' splice sequence "GTAAGT" and 3' sequence "(C/T)nGCTATAG" enabled the direct identification of intron / exon

boundaries. The exon regions of the 460 bp HindIII fragment show 81 % and 79 % sequence homology respectively to the corresponding regions of the mouse adult and embryonic β-like globin genes.

The remainder of exon 1 and the upstream regions of the *S.crassicaudata* β-like globin gene (henceforth referred to as S.c-β), were located in the contiguous 940 bp HindIII fragment of pSG-2 (Fig. 4.1.9). Three distinct regions of homology, the "ATA box", the "CCAAT box", and the "CACA box" promoter elements, were also observed in the upstream regions of S.c-β.

The orientation of S.c-β in the 940 bp and 460 bp HindIII fragments indicated that the 2nd intron and possibly the 3rd exon of S.c-β may be present in the contiguous 1.2 kb BamHI / HindIII fragment of pSG-2 (Fig. 4.1.6). Comparisons of the DNA sequence of the 1.2 kb BamHI / HindIII fragment of pSG-2 with the mouse adult and embryonic β-like globin gene 3rd exons, failed to show any sequence homology. This confirmed the results of a Southern analysis of pSG-2, where the plasmid clone, pDG-5.2, which contains sequences homologous to the 3rd exon of a *D.viverrinus* adult β-globin gene, failed to hybridize to this 1.2 kb fragment (Fig. 4.1.7). This result suggests that the 3rd exon of S.c-β lies in the contiguous 10 kb BamHI / SalI fragment of λSG-3 (Fig. 4.1.3). Additional evidence suggests this fragment is present in the plasmid subclone pSG-12 (Section 4.1.6).

The M13 subclones generated and sequenced in the plasmid pSG-2 are shown in Fig.4.1.10. The 1st intron, 1st and 2nd exons, and a portion of the 2nd intron of S.c-β were sequenced on both DNA strands. In addition, 575 bp of upstream flanking

DNA (Fig.4.1.10) was sequenced on both strands and a further 1.1 kb of DNA upstream of the gene was sequenced on one strand only (see Appendix).

One small compression in the DNA sequence of S.c-β was observed in a region towards the 5' end of the 2nd intron. This region consists of the sequence :- 5'...TTTAAAA[GGGCC]A[GGGCC]AAAATAAA...3' with the compression occurring in the bracketed regions. When determining the sequence of the non-transcribed strand (complementary to the sequence shown above), the compression occurred in the 2nd bracketed region, while sequencing of the transcribed strand gave a compression in the 1st bracketed region. On combining the two single strand sequences, the above DNA sequence was assumed to be correct for this region of the gene.

The Southern results reported in section 4.1.7 showed that the 4.8 kb HindIII fragment hybridized to the PstI insert of pDG-5.2 (Fig. 4.1.7). This suggested that the 1.2 kb SalI / HindIII fragment of pSG-2 may have contained the 3rd exon of a β-like globin gene. The DNA sequence data reported above, suggest that this fragment does not contain the 3rd exon or the 1st and 2nd exons of S.c-β. In addition, DNA sequence data from the 1.2 kb SalI / HindIII fragment, failed to show any significant homology to a mouse adult β-globin gene (comparison not shown), suggesting that this fragment does not contain DNA sequences from a second β-like globin gene. The most likely explanation for the hybridization of the PstI insert of pDG-5.2 to the 4.8 kb HindIII fragment of pSG-2, is that small quantities of pBR322 vector sequences were co-purified with the PstI insert during its isolation from a LMT agarose gel.

Fig. 4.1.8

DNA sequence comparisons involving the 460bp *HindIII* fragment of pSG-2 and mouse ϵ- (ßh1) and ß- (ßmaj) globin genes. Shown are nucleotide sites at which S.c-ß differs from the mouse ϵ- and ß-globin sequences. Dotted lines indicate regions of the mouse ϵ- and ß-globin sequences which were not compared with the S.c-ß sequence due to problems with aligning the three sequences.

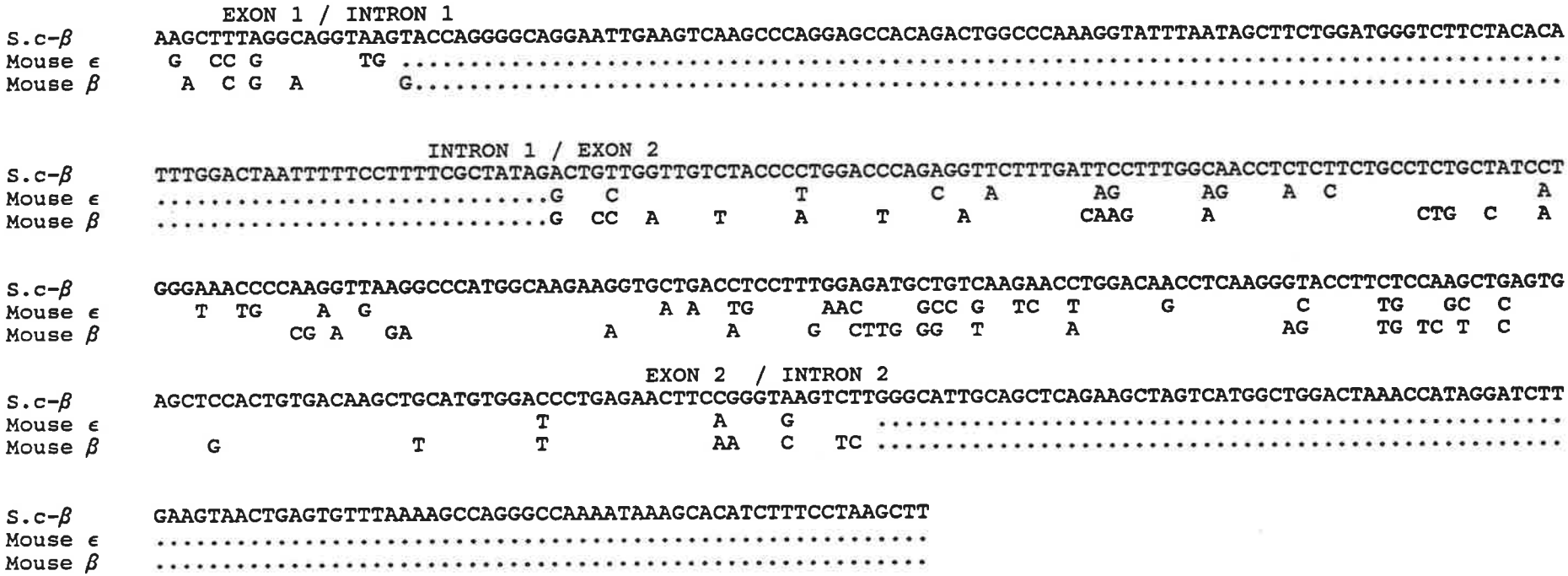

EXON 1 / INTRON 1
S.c-β AAGCTTTAGGCAGGTAAGTACCAGGGGCAGGAATTGAAGTCAAGCCCAGGAGCCACAGACTGGCCCAAAGGTATTTAATAGCTTCTGGATGGGTCTTCTACACA
Mouse ε G CC G TG
Mouse β A C G A G
INTRON 1 / EXON 2
S.c-β TTTGGACTAATTTTTCCTTTTCGCTATAGACTGTTGGTTGTCTACCCCTGGACCCAGAGGTTCTTTGATTCCTTTGGCAACCTCTCTTCTGCCTCTGCTATCCT
Mouse ε G C T C A AG AG A C A
Mouse β G CC A T A T A CAAG A CTG C A
S.c-β GGGAAACCCCAAGGTTAAGGCCCATGGCAAGAAGGTGCTGACCTCCTTTGGAGATGCTGTCAAGAACCTGGACAACCTCAAGGGTACCTTCTCCAAGCTGAGTG
Mouse ε T TG A G A A TG AAC GCC G TC T G C TG GC C
Mouse β CG A GA A A G CTTG GG T A AG TG TC T C
EXON 2 / INTRON 2
S.c-β AGCTCCACTGTGACAAGCTGCATGTGGACCCTGAGAACTTCCGGGTAAGTCTTGGGCATTGCAGCTCAGAAGCTAGTCATGGCTGGACTAAACCATAGGATCTT
Mouse ε T A G
Mouse β G T T AA C TC
S.c-β GAAGTAACTGAGTGTTTAAAAGCCAGGGCCAAAATAAAGCACATCTTTCCTAAGCTT
Mouse ε
Mouse β

Fig. 4.1.9

DNA sequence comparisons involving regions of the 940bp *HindIII* fragment of pSG-2 and mouse ϵ- (ßh1) and ß- (ßmaj) globin genes. Shown are nucleotide sites at which S.c-ß differs from the mouse ϵ- and ß-globin sequences. Dotted lines indicate regions of the mouse ϵ- and ß-globin sequences which were not compared with the S.c-ß sequence. The promoter sequences "CACCC", "CAAT" and "ATA", and the initiation codon (ATG) are underlined. The *HindIII* site (AAGCTT) within the 1st exon of the S.c-ß is also underlined.

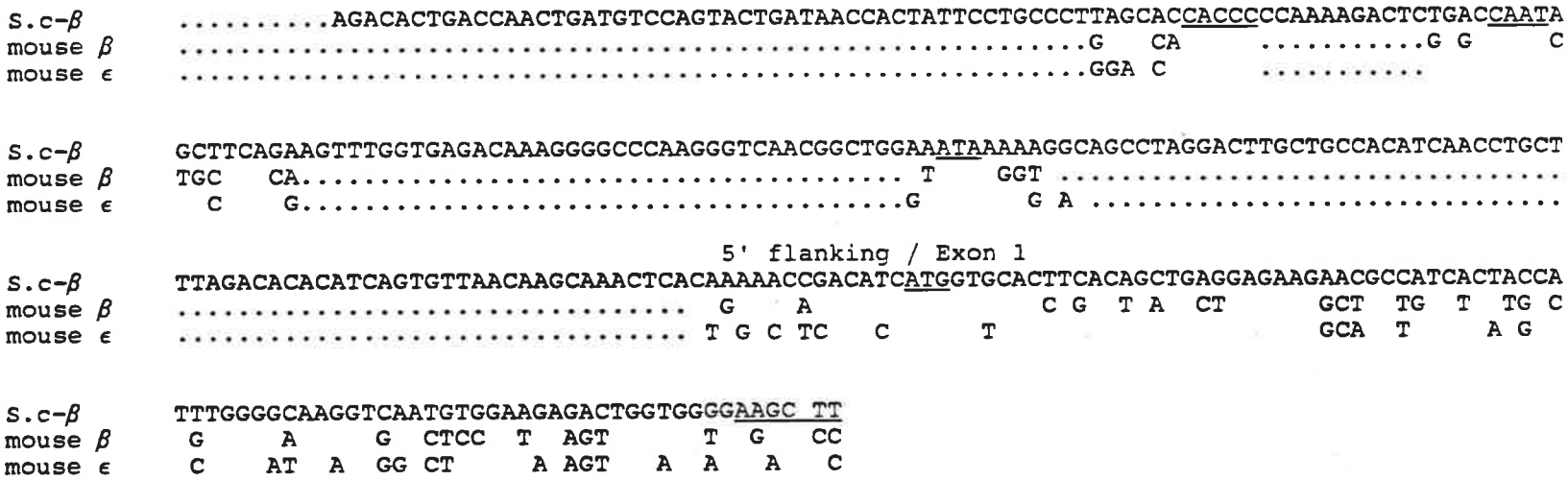

S.c-β ..........AGACACTGACCAACTGATGTCCAGTACTGATAACCACTATTCCTGCCCTTAGCACCACCCCCAAAAGACTCTGACCAATA
mouse β .....G CA .....G G C
mouse ϵ .....GGA C .....
S.c-β GCTTCAGAAGTTTGGTGAGACAAAGGGGCCCAAGGGTCAACGGCTGGAAATAAAAAGGCAGCCTAGGACTTGCTGCCACATCAACCTGCT
mouse β TGC CA..... T GGT .....
mouse ϵ C G.....G G A .....
5' flanking / Exon 1
S.c-β TTAGACACACATCAGTGTTAACAAGCAAACTCACAAAAACCGACATCATGGTGCACTTCACAGCTGAGGAGAAGAACGCCATCACTACCA
mouse β ..... G A C G T A CT GCT TG T TG C
mouse ϵ ..... T G C TC C T GCA T A G
S.c-β TTTGGGGCAAGGTCAATGTGGAAGAGACTGGTGGGGAAGC TT
mouse β G A G CTCC T AGT T G CC
mouse ϵ C AT A GG CT A AGT A A A C

Fig. 4.1.10

A: Regions of pSG-2 which were subcloned into M13 and sequenced using the "dideoxy chain termination" procedure.

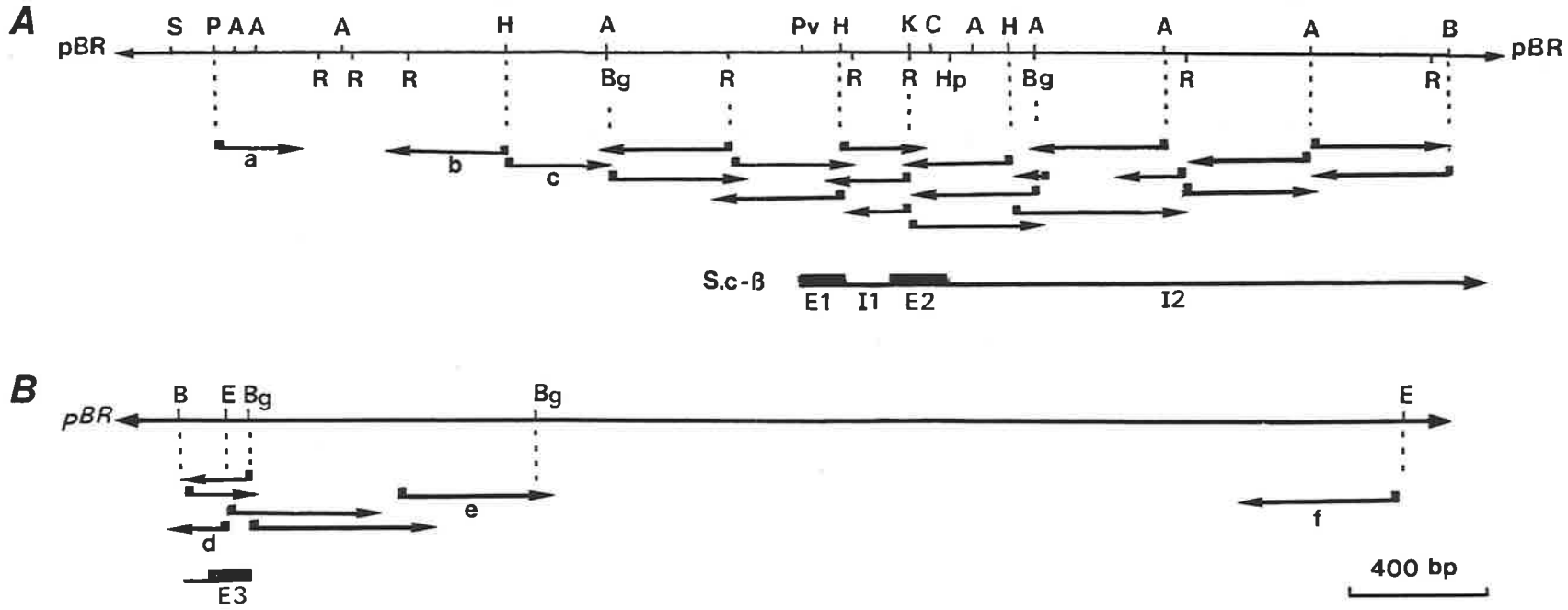
B: Regions of pSG-12 which were subcloned and sequenced.

Restriction fragments (as indicated between broken lines, and spanning one or more broken lines) were subcloned into MP18 and MP19. Lines with an arrowhead below the restriction map signify regions of each subclone which were sequenced; from the closed box to the point of the arrow. Each restriction site within the gene (shown below restriction map) was sequenced over at least once, to verify that restriction fragments were contiguous.

Enzyme sites on each map are: A:- Sau3A; B:- BamHI; Bg:- BglII; C:- SacI; E:- EcoRI; H:- HindIII; Hp:- HpaII; K:- KpnI; P:- PstI; Pv:- PvuII; R:- RsaI.

In the diagram of S.c-ß, E1, E2 and E3 (thick lines) refer to the 1st, 2nd and 3rd exons of the gene respectively; I1 and I2 (thin lines) refer to the 1st and 2nd introns of the gene respectively.

Sequences a,b,c,e and f are shown in the appendix. d- designates the subclone of λSG-3 which was used to sequence over the shared BamHI site of pSG-2 and pSG-12.


A
pBR
S P A A A H A Pv H K C A H A A A B
pBR
R R R Bg R R R R Hp Bg R R
a
b
c
S.c-ß
E1 I1 E2 I2
B
pBR
B E Bg Bg E
e
f
d
E3
400 bp

These contaminating sequences subsequently hybridized to vector sequences contained within the 4.8 kb HindIII fragment.

4.1.11 Characterization of the plasmid subclone pSG-12.

In order to isolate and sequence the remainder of S.c-β, the plasmid subclone pSG-12 was characterized to determine whether it contained the 10 kb BamHI / SalI fragment of λSG-3 which is contiguous with the 4 kb BamHI / SalI fragment of pSG-2. Plasmid DNA from the subclone pSG-12 was prepared using the large-scale plasmid purification procedure and digested with BglII, EcoRI, BamHI, SalI, SacI and KpnI. DNA samples from single and double digestions were electophoresed on 1% w/v agarose gels. Fragment sizes were estimated by comparison with known standards and were used to construct a restriction map of pSG-12 (Fig. 4.1.11). EcoRI sites were located on the map by digesting pSG-12 DNA with 0.5 x serial dilutions of EcoRI from 2 units/μg to 0.25 units/μg, and estimating partially digested EcoRI fragment sizes after electrophoresis on a 1% w/v agarose gel (Fig. 4.1.12a).

Southern analysis of pSG-12 DNA, using the pDG-5.2 PstI insert as a probe, showed that the BamHI / SalI insert of pSG-12 and a 5.7 kb BglII and a 3.4 kb EcoRI fragment hybridized strongly to the probe (Fig. 4.1.12b). When the 5.7 kb BglII fragment was digested with BamHI, the amount of hybridization of the probe was significantly reduced. This suggested that the 3rd exon of S.c-β may be located on a 200 bp BamHI / BglII fragment of pSG-12.

Fig. 4.1.11

Restriction map of pSG-12.

Enzymes mapped are: B:- BamHI; E:- EcoRI; Bg:- BglII; C:- SacI; S:- SalI. Enlarged is the region containing the 2nd intron (I2), 3rd exon (E3) and 3' flanking sequence of S.c-ß. Sizes of fragments are in kb. pBR designates pBR322 DNA adjacent to the SalI / BamHI insert of pSG-12.

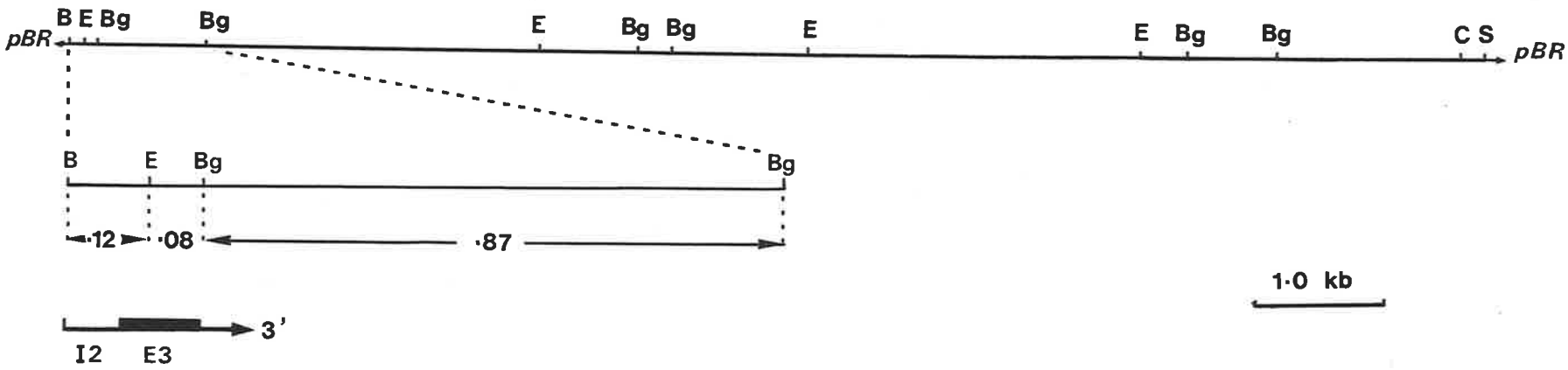

pBR
B E Bg
Bg
E
Bg Bg
E
E Bg
Bg
C S
pBR
B
E
Bg
Bg
·12
·08
·87
1·0 kb
3'
I2
E3

Fig. 4.1.12

A: Partial restriction analysis of pSG-12.

pSG-12 DNA was digested with 0.5 x serial dilutions of EcoRI: 2:- 2 units; 3:- 1 unit; 4:- 0.5 unit; 5:- 0.25 unit. Fragments were separated by electrophoresis on a 1% w/v agarose gel (shown in figure).

Molecular weight standard used are: 1:- SPP-1 digested with EcoRI; 6:- λ digested with HindIII

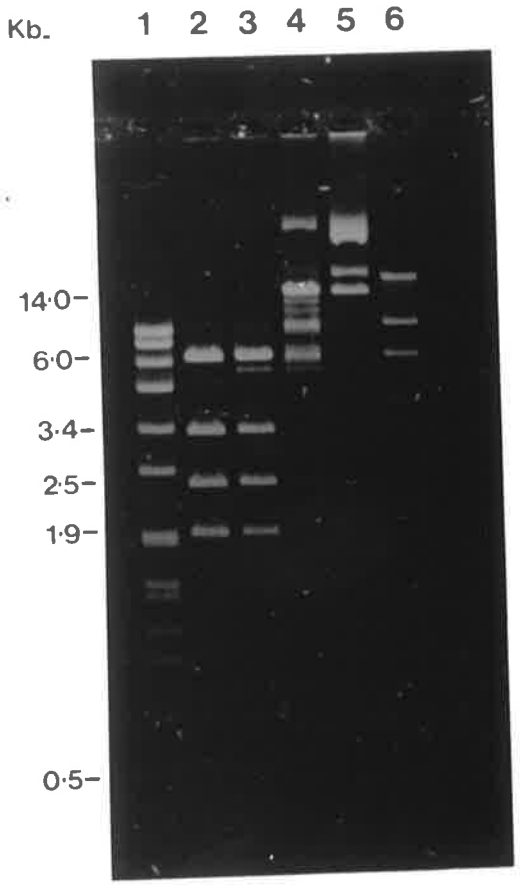
B: Restriction analysis of pSG-12.

pSG-12 was digested with: 1:- BglII; 2:- BamHI/BglII; 3:- BamHI/SalI; 4:- BamHI; 5:- EcoRI/SalI; 6:- EcoRI. Fragments were separated by electrophoresis on a 1% w/v agarose gel (shown in figure). λ digested with HindIII (not shown) was used as a molecular weight standard.

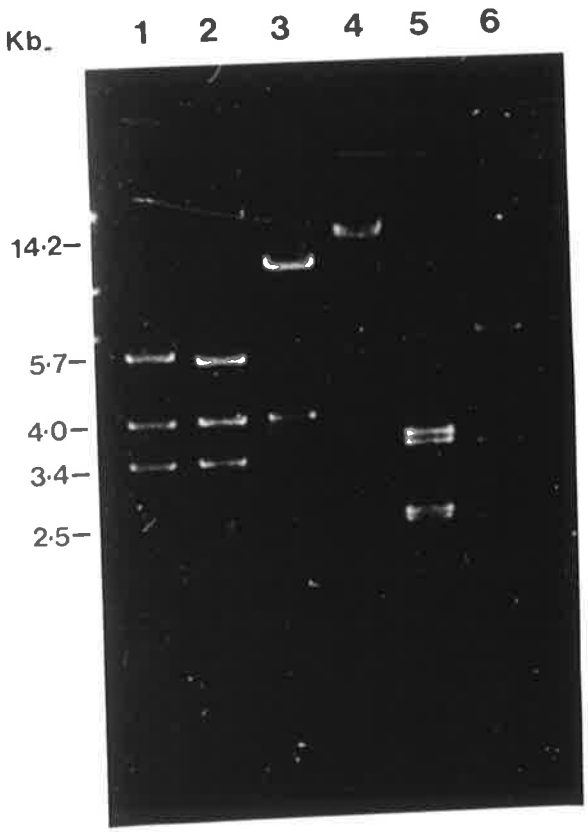
C: Southern analysis of pSG-12.

DNA fragments from the above gel (Fig. B) were Southern transferred to nitrocellulose and probed with the PstI insert of pDG-5.2 (radio-labelled using the oligo-labelling procedure). Shown is an autoradiograph after 4 hours exposure with no intensifying screen.

A



B



C

The faint hybridization of the probe to DNA fragments containing pBR322 vector sequences indicated that the probe may have been slightly contaminated with pBR322 DNA (Fig. 4.1.12b). This, however, would have no bearing on the above conclusion.

The restriction mapping and Southern results described in this section suggest firstly, that pSG-12 contains the 10 kb BamHI / SalI fragment of λSG-3, and secondly that this fragment contains the 3rd exon of S.c-β, in a region of DNA immediately adjacent to the BamHI site.

4.1.12 The generation of pSG-12 subclones in M13.

Two strategies were used to generate pSG-12 subclones in M13. Firstly, known restriction sites adjacent to the BamHI site of pSG-12 (Fig. 4.1.12), were used to generate M13 subclones containing the 200 bp BamHI / BglII fragment, the contiguous 870 bp BglII fragment, and the 3.4 kb EcoRI fragment of pSG-12.

The second strategy involved using the enzyme Bal31 to generate deletions into the 3.4 kb EcoRI fragment of pSG-12 (Fig. 4.1.14). pSG-12 was digested with BamHI and treated with Bal31 (Section 3.2.4). Aliquots from the digestion were taken at 2 minute intervals, treated with the large fragment of DNA polymerase (Klenow) to blunt end repair the DNA molecules, and digested with EcoRI. The aliquots were analysed by electrophoresis on a 1.0% w/v agarose gel to determine the rate of digestion of DNA by the Bal31 enzyme (Fig. 4.1.13a). It was found that treatment with Bal31 resulted in approximately 150 bp of double stranded DNA being

removed every 2 minutes in each direction from the BamHI site of pSG-12. The deletion products of the 3.4 kb EcoRI fragment of pSG-12, were isolated and purified from a 1.2% w/v LMT agarose gel and ligated into the EcoRI / SmaI site of MP18.

Single stranded M13 templates were prepared from colourless plaques and analysed by electrophoresis on agarose mini-gels. It was found that of 35 single stranded M13 clones analysed, 24 contained deletions of M13, and only 11 contained inserts. Comparisons of the mobility of these single stranded M13 clones, with recombinant M13 clones containing 500 bp and 3.4 kb inserts, revealed that where pSG12 DNA had inserted into MP18, it was generally of a short length (Fig. 4.1.13b). This was despite the fact that the insert DNA was originally at least 2.7 kb in size before ligation. This failure to ligate Bal31 treated DNA fragments into M13 was possibly caused by the presence of contaminants in the ligation mix from a bad batch of LMT agarose. M13 single stranded templates containing pSG-12 inserts, including several Bal31 deletion inserts, were sequenced using the "dideoxy chain termination" procedure (Section 3.2.4).

4.1.13 DNA sequence from the plasmid subclone pSG-12.

DNA sequence determined from the 200 bp BamHI / Bgl II fragment of pSG-12 was found to be closely homologous to the 3rd exon of the mouse adult and embryonic β-globin genes and the 3rd exon cDNA sequence of a *D. viverrinus* adult β-globin mRNA (Fig. 4.1.14a).

Fig. 4.1.13

A: Analysis of Bal31 deletion products.

pSG-12 was digested with BamHI and then treated with Bal31 for varying time intervals (Section 3.2.4): 0:- 0 minutes; 2:- 2 minutes; 4:- 4 minutes; 6:- 6 minutes; 8:- 8 minutes; 10:- 10 minutes. Bal31 digested DNA from each time interval was further digested with EcoRI and analysed by electrophoresis on a 1% w/v agarose gel (shown in figure). Note that approximately 150bp of DNA is removed from the 3.4kb EcoRI fragment every 2 minutes.

B: Analysis of M13 subclones containing Bal31 deletion fragments.

Single stranded M13 templates were analysed by electrophoresis on a 0.8% w/v agarose gel. Lanes show MP18 clones after ligation with the following DNA fragments:

1 - Deletion products of the 3.4kb EcoRI fragment after a 6 minute treatment with Bal31
2 - 500bp EcoRI fragment of pSG-12
3 - no DNA, ie. native MP18
4,5,6 - Deletion products of the 3.4kb EcoRI fragment of after a four minute treatment with Bal31
7 - 3.4kb EcoRI fragment of pSG-12
8 - Deletion products of the 3.4kb EcoRI fragment after a 2 minute treatment with Bal31.

Fig. 4.1.14

A: Comparisons involving DNA sequence from the BamHI site of pSG-12 and mouse ϵ-(ßh1) and ß-(ßmaj) globin genes: showing the remainder of the 2nd intron, complete 3rd exon and 3' non-coding region of S.c-ß. For the mouse ϵ- and β-globin genes, only differences from the S.c-ß sequence are shown. Dotted lines indicate regions which were not compared between the three sequences. Underlined are the BamHI site (GGATCC) of pSG-12, the EcoRI site (GAATTC) the BglII site (AGATCT) and the 3' poly-adenylation signal (AATAAA).

B: DNA sequence of a tandem repeat located 312bp from the end of the 3rd exon of S.c-ß.

A

```
                                                                                                  Intron 2 / Exon 3
S.c-β          GGATCCTGACTAGAGCTTTTTCCTTATAAGGTGCCTGAGTTTTTCTTTTTCTCTCTCTCTCTTACAGCTCCTTGGGAATGTTCTGGTTATTGTTATGGCTGCT
Mouse ε        ..........................................................................      G  C   A GA C  G     GC   GCCAC
Mouse β        ..........................................................................      G  C  CA GT   G     CC TT  A

                                                                                                                  Exon 3 / 3' Non coding
S.c-β          CATTTCAACAAGGAATTCACTCCTGAAGTGCAGGCTGCTTTTCAGAAGCTGGTGACTGGTGTGGCCAATGCTTTGGCTCACAAGTACCACTAAATGTAGATCT
Mouse ε          CC TGG      T     C  C CT CA       C  C     G    G   A       C  C
Mouse β             TGC           C  A  G          C GG           T   A           C  T C          T

S.c-β          ATCTGTGGTGGGGGTCTATGGGCAAATGGAAGCCCAACCATTCATACTGCCTTGTAGTCCTAGCCCTTAACTAATAAAGTGCTTTTGGCCTTGCAGTTCATTG
Mouse ε        ..........................................................................     ..........................
Mouse β        ..........................................................................     ..........................

S.c-β          TCATTGTGTT
```

B

```
CATTATGGAGAGGCAAAAG ACAGTCCCTGCTCTGGTGGAGCT ACAGTCCCTGCTCTAGTGGAGCT TACAATATAATTGTGTGTGT
```

The 870 bp Bgl II fragment of pSG-12 was sequenced on one strand only and was shown to contain the 3' non-coding region of S.c-β (Fig. 4.1.14a). Attempts to obtain DNA sequence of this fragment on the complementary strand proved unsuccessful. Sixteen clones of the 870 bp BglII fragment were analysed on a sequencing gel by sequencing the T-track only. All sixteen clones contained the 870 bp BglII fragment in the same orientation. Attempts to subclone smaller regions of this fragment also proved unsuccessful. It is possible that regulatory elements may be present in the 870 bp BglII fragment, which somehow interfere with the replication of M13 when inserted in a specific orientation.

DNA sequence ($\sim$450 bp) from this fragment was found to "overlap" with DNA sequence from an M13 clone generated by the Bal31 procedure which contained an approximately 3 kb insert. A notable feature of this overlapping DNA sequence was the presence of a tandem repeat (Fig. 4.1.14). It was therefore possible that the two M13 clones shared a repeat sequence and were not necessarily contiguous. To provide evidence that the two clones were contiguous, 200 ng of single-strand template DNA from the Bal31 M13 clone was labelled by primer extension and used to probe a Southern blot of pSG-12 DNA (Section 3.2.3). The autoradiograph showed strong hybridization of the probe to the 3.4 kb EcoRI fragment of pSG-12 (Fig. 4.1.15b). This provided evidence that the Bal31 M13 clone contained a deletion of the 3.4 kb EcoRI fragment and DNA sequence from this fragment would therefore be contiguous with sequence determined from the 870 bp BglII fragment of pSG-12. Several other M13 subclones produced by the Bal31 procedure were also sequenced, but were not

found to overlap with DNA sequence from the 3' flanking region of S.c-β (see Appendix for sequence data).

The M13 clones of pSG-12 that were sequenced are shown in Fig. 4.1.10. The remainder of the 2nd intron and the complete 3rd exon of S.c-β which were present in the plasmid subclone pSG-12 were sequenced on both strands. An extra 800 bp of 3' flanking sequence was sequenced on one strand only.

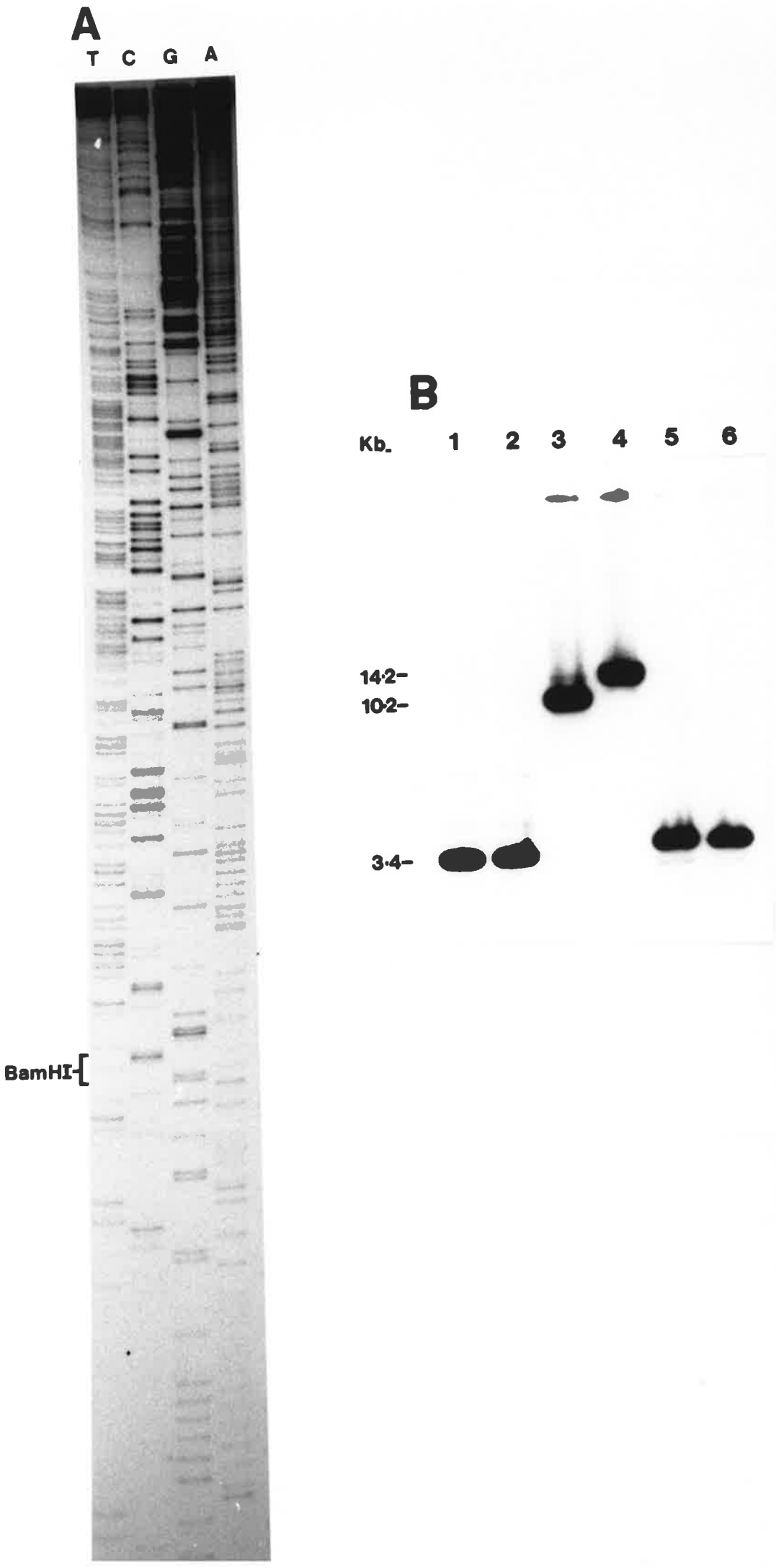
4.1.14 Sequencing over the BamHI site of λSG-3.

Although the entire DNA sequence of S.c-β had been determined from both strands in each of the plasmid subclones pSG-2 and pSG-12, it was also necessary to sequence over the shared BamHI site of the two clones, to verify that they were contiguous. This could only be achieved by cloning a fragment containing this restriction site from the phage clone λSG-3. λSG-3 DNA was digested with HindIII and EcoRI, and the 1.3 kb HindIII / EcoRI fragment, containing the BamHI site of λSG-3, was isolated and purified from a LMT agarose gel and ligated into the HindIII / EcoRI site of MP19. The DNA sequence of this subclone (Fig. 4.1.15a) verified the presence of only a single BamHI site in λSG-3, and confirmed that the sequence of the β-like globin gene from pSG-2 was contiguous with the sequence determined from pSG-12.

Fig. 4.1.15

A: Autoradiograph of a sequencing gel showing that a single BamHI site is present in an M13 clone containing the 1.3kb HindIII/EcoRI fragment of λSG-3 (clone d, Fig 4.1.10).

B: Southern analysis of pSG-12 DNA digested with: 1:- BglII; 2:- BamHI/BglII; 3:- BamHI/SalI; 4:- BamHI; 5:- EcoRI/SalI; 6:- EcoRI; using the Bal31 M13 clone (e, Fig. 4.1.10) as a probe. The Southern was hybridized under standard conditions at 65°C (Section 3.2.7) with the radio-labelled probe (prepared by primer extension from the "universal" M13 primer using the Oligo-labelling procedure; Section 3.2.3.6). The filter was washed to a stringency of 0.1 x SSC, 0.1% w/v SDS at 65°C and autoradiographed for 2 hours at room temperature with no intensifying screen. The autoradiograph shows strong hybridization of the probe with the 3.4kb BglII and 3.4kb EcoRI fragments of pSG-12.


A
T C G A
BamHI
B
Kb.
1 2 3 4 5 6
14·2–
10·2–
3·4–

4.1.15 The complete DNA sequence of a β-like globin gene from *S.crassicaudata.*

The complete DNA sequence of S.c-β is shown in Fig. 4.1.16. The location of the putative CAP site and the 5' promoter sequences, including the "ATA box", the "CAAT box", and the "CACCC box", were determined by comparing the gene with mouse adult and embryonic β-like globin genes. More detailed comparisons of this region are described in Section 4.2. In addition, the 3' poly-adenylation signal (AATAAA) and intron / exon splicing signals, were also identified by their homology with corresponding regions of the mouse β-like globin genes.

There are no suggestions from the overall structure of the gene, that it would be non-functional. The introns split the gene in positions that are identical to those in all known mammalian globin genes. The exons are open reading frames, and conceptually translate to give a polypeptide consisting of 146 amino acid residues, analagous to all known mammalian β-globin genes. One notable feature of this gene is the size of the second intron (1438 bp), which is much larger than the 2nd intron of any eutherian β-like globin studied to date (see Section 4.2).

Fig. 4.1.16

The complete DNA and inferred amino acid sequence of the *S. crassicaudata* ß-like globin gene. The gene is divided into three exons and two introns, with each of the introns flanked by conserved donor and acceptor splicing signals. The 5' flanking region of the gene contains "CACCC" and "CAAT" promoter elements, an RNA polymerase II binding site ("ATA") and a putative CAP site. The conceptual amino acid sequence is 146bp in length and is given in a the standard three letter code (see appendix). The 3' flanking region contains the termination codon TTA (underlined) and a poly-adenylation signal (AATAAA). The length of the 1st and 2nd introns of the gene are 120 bp and 1438 bp respectively. The entire DNA sequence of the gene was determined on both DNA strands with the exception of the last 113 bp of 3' flanking sequence.

```
GATCTGACAGGTAAACTATTCCTTATTCTCCTAATTTGCTATCACCCTTTGTGTCTAAATCATGGATTCATTTTTACCTT
ATTTTGATATATGGTGTGAAATGTCTATCTATAGCTAGTTTTCTTTCTATACTATTTTTCAGTTTTTCCAAAAGTTTTTT
TTTTTTTTTTTGTCAAATAGTGAGTTTTTATCCTAGTGGCTAGTCTGGATTTATAAATAATAAATTTATAAATAATAGAT
TATTATATTCATGACTGTGTTTTGTGTTTCTAATCTATTCCATTTATATAAGGGAAACTATTATTTGGATTAGAGAATTA
GCATGCAAGGACAGGAGACCCTACAGGTCTCCAAATACAGAGACACTGACCAACTGATGTCCAGTACTGATAACCACTAT
                caccc              ccaat
TCCTGCCCTTAGCACCACCCCCAAAAGACTCTGACCAATAGCTTCAGAAGTTTGGTGAGACAAAGGGCCAAGGGTCAACG
      ata                          cap
GCTGGAAATAAAACAGCAGCCTAGGACTTGCTGCCACATCAACCTGCTTTAGACACACATCAGTGTTAACAAGCAAACTC

                MetValHisPheThrAlaGluGluLysAsnAlaIleThrThrIleTrpGlyLysValAsnValGl
ACAAAAACCGACATCATGGTGCACTTCACAGCTGAGGAGAAGAACGCCATCACTACCATTTGGGGCAAGGTCAATGTGGA

uGluThrGlyGlyGluAlaLeuGlyAr
AGAGACTGGTGGGGAAGCTTTAGGCAGGTAAGTACCAGGGGCAGGAATTGAAGTCAAGCCCAGGAGCCACAGACTGGCCC

                                                              gLeuLeuValVal
AAAGGTATTTAATAGCTTCTGGATGGGTCTTCTACACATTTGGACTAATTTTTCCTTTTCGCTATAGACTGTTGGTTGTC

TyrProTrpThrGlnArgPhePheAspSerPheGlyAsnLeuSerSerAlaSerAlaIleLeuGlyAsnProLysValLy
TACCCCTGGACCCAGAGGTTCTTTGATTCCTTTGGCAACCTCTCTTCTGCCTCTGCTATCCTGGGAAACCCCAAGGTTAA

sAlaHisGlyLysLysValLeuThrSerPheGlyAspAlaValLysAsnLeuAspAsnLeuLysGlyThrPheSerLysL
GGCCCATGGCAAGAAGGTGCTGACCTCCTTTGGAGATGCTGTCAAGAACCTGGACAACCTCAAGGGTACCTTCTCCAAGC

euSerGluLeuHisCysAspLysLeuHisValAspProGluAsnPheArg
TGAGTGAGCTCCACTGTGACAAGCTGCATGTGGACCCTGAGAACTTCCGGGTAAGTCTTGGGCATTGCAGCTCAGAAGCT
AGTCATGGCTGGACTAAACCATAGGATCTTGAAGTAACTGAGTGTTTAAAAGCCAGGGCCAAAATAAAGCACATCTTTCC
TAAGCTTGGCATGAATGACTTCAGATATCAAAGGAGATCTGTTTCAAACATATGCATATACTCCTCTTCTTTCTATCTCC
TGATAGCACTTTCCATCTCTCTTTCATCATACACTCCTTTTTTTGGCCTTCATTGCTTCATTTCTAGTTCGTTTCTTTCT
CTTCTTTTTGTCTACTTTTTCTAATATTTCTCCTATCATCCTTCCCAATTTTTTTCTTTAGTGATTATTTTTTTTCAATT
TTTCTCCAACTGTCCCATTTTCTCTTTCCCTCTAATCATTTTGCTTGGACACATTCATGACACATTCATATTCTCCTCAT
GTTTCTCATCTCTTTTCACTGGATACATAATTCTCAAGTAACCTTTCTGCTTTTGAAAAGTGGGATGATTTGGTGTTCTG
CTTTAATTTGCTTTTCATCTAATTTGTGCTTTTTGTGGATCTTTGCTTCTCTGTCATCATCTTCTACCTCTACTTGATAC
CCCAATTCTGATTAGTACTCATGGAAGTCTCATGCTTAATCTTGCAAAGACAGTGCTACATCTCCATTTATTATCTTTTT
AAATCGGGAATCATATTGTTTGCACTAAATACTGAGACTTGATTTTGGAGTGAGGAATATTGGGTTCAAATGAAGATTTT
TCTTTATTAGTTCTGTGATTTTATGAAAGTCACTCTATACTTGAGTCTATTTCTTCAGTTACAAAATAAGGAGTTGGACT
AGACTTCTACATCCCTTCTATGTGAATATCATTTTTCTTCCTTTTTTCTTCATTCTCATTTTTTTCACTCTTTCCTCTTC
TCTCTTTCCTTATTTTTTTCTTCCAATTGATCAGTTTTTCTTTTTCCTTTCCTTTCCAACCTCTTTAACATTCCAATTAG
TTCTTTTCAAGTTGGTAAATTAGGCCTGTCAAGTCCTTTATCTTAGTGTTATAAACTTTGGCCTATGAGAAGCAGATTTA
GGGTTGAGGAAGTTTTGGGAGGCAAGGAGGAAAGGGTTTTTAAGTAAGAAGAAGAAGGCAAGATAAACCCAAGAAAATAG
ATTAGGACAGAAATGGAACTGAGTATCTTATTTGGGTCTAAGGACAGGAGGAGCTGGAGAGGAAGAGACAGGGTCCTTTT
GAATAAAGAACTAAGGACAACAGGAACAGATGATATCTTATTCTGTAATAGGTGACCTGAGAATAGTTTTGGACATGGGT
AGGAGAACATGGATTCTATAGTAGTGGTACTATATTAAGAGAACATGGGTAGACTTCCCAGGGATCCTGACTAGAGCTTT

                                                        LeuLeuGlyAsnValLeuValIleValMetAl
TTCCTTATAAGGTGCCTGAGTTTTTCTTTTTCTCTCTCTCTCTTACAGCTCCTTGGGAATGTTCTGGTTATTGTTATGGC

aAlaHisPheAsnLysGluPheThrProGluValGlnAlaAlaPheGlnLysLeuValThrGlyValAlaAsnAlaLeuA
TGCTCATTTCAACAAGGAATTCACTCCTGAAGTGCAGGCTGCTTTTCAGAAGCTGGTGACTGGTGTGGCCAATGCTTTGG

laHisLysTyrHis
CTCACAAGTACCACTAAATGTAGATCTATCTGTGGTGGGGGTCTATGGGCAAATGGAAGCCCAACCATTCATACTGCCTT
                  ataaa
GTAGTCCTAGCCCTTAACTAATAAAGTGCTTTTGGCCTTGCAGTTCATTGTCATTGTGTT
```

4.2 Sequence comparisons involving the *S.crassicaudata* β-like globin gene, S.c-β.

4.2.1 DNA sequence comparisons involving the 5' promoter region.

In the following section the 5' promoter region of S.c-β is compared to corresponding regions of *D.virginiana* and eutherian β-like globin genes, to determine whether S.c-β is orthologous (homologous sequences that diverged through speciation; Fitch, 1970) to the *D.virginiana* and eutherian embryonic (ε, γ, η) or adult (δ, β) β-like globin genes.

Rationale.

In the 5' promoter regions of mammalian β-like globin genes there are a number of presumptive shared derived features (synapomorphs) of both the embryonic and adult set of genes, which can provide evidence of orthology. For example, the "ATA" box eutherian adult β-globin genes is embedded in the sequence: GGGCATAAAG and the "CAAT" box is generally of the form AGCCAAT. In eutherian embryonic β-like globin genes, however, the "ATA box" is embedded in the sequence : AAGAATAAAAG (Hardison, 1983), and "CAAT" boxes are of the form GACCAAT (Harris et al., 1986; Baralle et al., 1980; Hill et al., 1984). In addition, the distance between the "ATA" and "CAAT" boxes is approximately 9 bp longer in eutherian embryonic β-like globin genes compared with the adult genes. There is also

evidence that adult β-globin genes, in general, have 2 "CACCC" boxes while embryonic β-like globin genes have a single "CACCC" box. Little is known of the functional significance of these promoter "synapomorphs" and therefore their presence or absence in the 5' promoter region of S.c-β may not provide conclusive evidence of orthologous and paralogous (homologous sequences that diverged after gene duplication; Fitch, 1970) relationships between marsupial and eutherian β-like globin genes.

Results.

The S.c-β "ATA" box, located 30 bp 5' to the putative CAP site, was found to most closely resemble the *D.virginiana* ε-globin "ATA" box and is of the form AATAAAA found in all known eutherian embryonic β-like globin genes (Fig. 4.2.1). The S.c-β "CAAT" box, located 83 bp 5' to the putative CAP site, has precisely the same sequence as the "CAAT" box of the *D.virginiana* ε-globin gene, and differs by only one nucleotide from the mouse ε-globin "CAAT" box region (Fig. 4.2.1). Unlike the *D.virginiana* ε-globin gene, however, S.c-β has only a single "CAAT" box. The S.c-β "CAAT" box is of the form: GACCAAT and is located 48 bp from the "ATA" box, compared with 41 bp and 52 bp respectively for the *D.virginiana* adult and embryonic genes. It therefore most closely resembles the "CAAT" boxes of the *D.virginiana* and eutherian embryonic β-globin genes in both its sequence and location.

Fig. 4.2.1

Comparisons of "ATA", "CAAT" and "CACA" promoter elements in marsupial and eutherian ß-like globin genes.
Lower case letters signify that the nucleotide is identical to that in the S.c-ß sequence. The *D.virginiana* ϵ and ß sequences were taken from Koop and Goodman (1988); mouse ϵ(ßhl) from Hill *et al.* (1984), mouse ßmaj from Konkel *et al.* (1978), human ϵ from Baralle *et al.* (1980) and human ß from Lawn *et al.* (1980).

Comparisons of "ATA" boxes in mammalian β-like globin genes.

S.c-β	.....GGAAATAAAACA.....
D.virg ε	.....gCaaataaaaGG.....
D.virg β	.....AgGGataaaaGG.....
mouse ε (βh1)	.....AAGaataaaaGa.....
mouse βmaj	.....gCaTataaGGTG.....
human ε	.....CAGaataaaaGG.....
human β	.....ggGCataaaaGT.....

Comparisons of "CAAT" boxes in mammalian β-like globin genes.

S.c-β	.....TGACCAATAGCTTCAGA....
D.virg ε	.....tgaccaatagcttcaga....
D.virg β	.....tAGccaatagACATCCa....
mouse ε (βh1)	.....tgaccaatagcCtcaga....
mouse βmaj	.....GgGccaatCTGCtcaga....
human ε	.....tgaccaatGActtTTAa....
human β	.....tgGccaatCTCCCAGga....

Comparisons of "CACA" boxes in mammalian β-like globin genes.

S.c-β	.....AGCACCACCCCCAA.....
D.virg ε	.....GAcTccaccccACa.....
D.virg β	.....agcCAcaccccAaC.....
mouse ε (βh1)	.....GAcCccaccccAAC.....
mouse βmaj	.....agcCAcacccTGCA.....
human ε	.....GAcTccaccccTGa.....
human β	.....agcCAcacccTAGG.....

The "CACCC" box of S.c-β is located 14 bp upstream of the "CAAT" box. S.c-β has a single "CACCC" box, a characteristic of all known eutherian embryonic β-globin genes and the *D.virginiana* ϵ-globin gene (Koop and Goodman, 1988). The S.c-β "CACCC" box is embedded in the sequence:- CACCACCCCC. It differs by only 1 nucleotide from the *D.virginiana*, mouse and human ϵ-globin gene "CACCC" boxes and by at least 2 nucleotides from the *D.virginiana*, mouse and human adult β-globin gene "CACCC" boxes (Fig. 4.2.1).

The above DNA sequence comparisons involving the 5' promoter region of S.c-β provide evidence that the gene is orthologous to the *D.virginiana* and eutherian embryonic β-globin genes.

4.2.2 Comparisons involving the 1st and 2nd introns of S.c-β.

The locations of the introns (intervening sequences) in S.c-β were determined by homology with the mouse adult and embryonic β-like globin genes (Section 4.1.10) and are analogous to the location of introns in all known vertebrate globin genes. Comparisons involving the intron / exon junctions of S.c-β and *D.virginiana* and eutherian β-like globin genes, revealed the presence in S.c-β of conserved "donor" and "acceptor" splicing sites (Fig. 4.2.2). These sites are required for accurate and efficient splicing of introns from "pre-mRNA" (see Mount, 1982). The "donor" and "acceptor" consensus sequences reported by Mount (1982) match closely with the "donor" and "acceptor" sequences of both introns in S.c-β (Fig. 4.2.2).

The 5' "donor" splice sequence in intron 1 of S.c-β most closely resembles the corresponding sequence of the *D. virginiana* adult β-globin gene (Fig 4.2.2). In intron 2, however, the "donor" splice sequence of S.c-β most closely resembles the corresponding sequence of the *D. virginiana* ϵ-globin gene. No obvious shared derived sequence features of the intron/ exon splice junctions of embryonic or adult β-like globin genes were apparent in these comparisons.

The size of the 1st intron of S.c-β (120 bp) is similar to that found in most eutherian β-like globin genes, which range from 116 bp to 132 bp in length (Table 4.2.1). The *D. virginiana* adult and embryonic β-globin genes have slightly smaller 1st introns of size 113 bp and 109 bp respectively (Table 4.2.1). The one exception to this narrow range of 1st intron sizes occurs in the mouse β_{h0}-globin gene which is 336 bp in length (Hill *et al.*, 1984).

The 2nd intron of vertebrate β-globin genes is larger in all known cases than the 1st intron and much more variable in size (573 bp - 1857 bp, see Table 4.2.1). The S.c-β 2nd intron is 1438 bp in size, almost 400 bp larger than the largest known eutherian β-globin gene 2nd intron (1040 bp in the goat ϵ^{II}-globin gene, Shapiro *et al.*, 1983), and approximately similar in size to the 2nd intron of the *D. virginiana* adult β-globin gene (1465 bp). The 2nd intron of the *D. virginiana* ϵ-globin gene is almost 400 bp larger again (1857 bp).

Fig. 4.2.2

Comparisons of intron/exon junctions in intron 1 and intron 2 of marsupial and eutherian ß-like globin genes.

Lower case letters signify that the nucleotide is identical to that in the S.c-ß sequence. Upper case letters signify that the nucleotide is different from that in the S.c-ß sequence. The source of each sequence is given in Fig. 4.2.1.

Comparisons of intron / exon junctions in intron 1.

	5'-DONOR	ACCEPTOR-3'
S.c-β	GTAAGTACC	$(C/T)_{12}$GCTATAG
D.virg ϵ	gtaagtaTA	$(c/t)_{12}$CctaCag
D.virg β	gtaagtacc	$(c/t)_{12}$CcCaAag
mouse ϵ (βh1)	gtaagGaAT	gTCTtag
mouse βmaj	gtTGgtaTc	TTtTtag
human ϵ	gtaagCaTT	TTCatag
human β	gtTGgtaTc	CcCTtag
consensus*	gt(a/g)agtt	$(c/t)_n$n(c/t)ag

(n > 10)

Comparisons of intron / exon junctions in intron 2.

	5'-DONOR	ACCEPTOR-3'
S.c-β	GTAAGTCTT	$(C/T)_{21}$CTTACAG
D.virg ϵ	gtaagtctt	$(c/t)_{20}$cCtacag
D.virg β	gtaagtTCt	$(c/t)_{12}$TCtacag
mouse ϵ (βh1)	gtCagtTCt	Gttacag
mouse βmaj	gtGagtctG	cCCacag
human ϵ	gtGagtTCA	ctAacag
human β	gtGagtctA	cCCacag
consensus*	gt(a/g)agtt	$(c/t)_n$n(c/t)ag

(n > 10)

* consensus intron /exon junctions from Mount (1982).

TABLE 4.2.1

Comparisons of intron lengths amongst vertebrate globin genes[1].

Species	Gene	Intron 1	Intron 2
α-globin related			
human	α_2	117	140
mouse	α	122	135
goat	$^{I}\alpha$	108	103
Xenopus laevis	α^1	171	338
β-globin related			
human	β	130	850
	δ	128	889
	$^{A}\gamma$	122	866
	ϵ	122	850
mouse	β^{maj}	116	653
	β^{min}	116	628
	βh0	336	841
	βh1	116	807
rabbit	β1	126	573
goat	β^{A}	128	906
	β^{C}	132	847
	γ	129	827
D.virginiana	β^{m}	113	1465
	ϵ^{m}	109	1857
S.crassicaudata	β-like	120	1438
chicken	β	92	810
X.laevis	β^1	192	841
Myoglobin			
seal	Mb	~4800	~3400
human	Mb	~5800	~3600

1. Intron lengths were taken from Blanchetot *et al.* (1983) except for the following: *D.virginiana* β^{m} and ϵ^{m} from Koop and Goodman (1988); human myoglobin (Mb) form Weller *et al.* (1984); chicken β from Dolan *et al.* (1983). Lengths are in base pairs.

Sequence alignment of the 2nd introns of S.c-β and the *D.virginiana* ϵ- and β-globin genes, revealed no significant stretches of sequence homology (results not shown). It was noted, however, that long tracks of dT's interspersed occasionally with dC's are a common feature of the 2nd intron of these genes, particularly in the central regions of the intron. In S.c-β, for example, 42% of all nucleotides are dT's in a region spanning 908 bp, 170 bp 3' of the "donor" splice site. This feature was not observed to the same degree in human, goat, mouse, rabbit or avian β-like globin genes (Konkel *et al.*, 1984; Shapiro *et al.*, 1983; Dolan *et al.*, 1983; Baralle *et al.*, 1980; Lawn et al., 1980).

4.2.3 Protein sequence comparisons.

Koop and Goodman (1988) determined the "most parsimonious geneological arrangement for > 80 β-globin genes" in mammals and concluded that six non-synonymous substitutions occurred on the "nascent adult haemoglobin β-chain gene line" and two non-synonymous substitutions occurred on the "nascent embryonic gene line". Comparisons of the conceptually translated amino acid sequences of S.c-β and the *D.virginiana* ϵ and β-globin genes, indicated that the S.c-β chain contains the 2 amino acid substitutions : leucine -> phenylalanine and alanine -> serine at positions 3 and 52 respectively, that are postulated to have occurred on the nascent ϵ-globin gene line (Fig. 4.2.3). Of the 6 substitutions proposed for the nascent β-globin gene line, 2 occurred in S.c-β: methionine -> leucine and isoleucine -> valine at positions

78 and 111 respectively. In particular, the amino acid substitution at position 3 of the S.c-β chain (leu -> phe) was suggested by Koop and Goodman (1988) to be important for embryonic haemoglobin function.

Comparison of the amino acid sequence of S.c-β with the *D. virginiana* adult β-globin gene at putative "contact sites" reveals a number of differences (Table 4.2.2). The two genes differ at seven out of forty five contact sites, with one difference occurring at a haem contact site (residue 85), one at an $\alpha_1\beta_2$ contact site (residue 39) and five at $\alpha_1\beta_1$ contact sites (residues 51, 55, 116, 119 and 128, Table 4.2.2).

In contrast, S.c-β and the *D. virginiana* ϵ-globin gene differ at only 3 contact sites (residues 55, 116 and 119; all of which are $\alpha_1\beta_1$ contact sites). At functional sites, S.c-β is clearly more closely related to the *D. virginiana* ϵ-globin gene.

4.2.4 Coding nucleotide sequence comparisons involving S.c-β: estimation of synonymous and non-synonymous divergence values.

Rationale.

The most informative region of S.c-β, that can be used to determine whether the gene is orthologous to either the embryonic or adult set of β-like globin genes, is the coding region of the gene. In comparing the coding regions of genes, it is necessary to distinguish between non-synonymous (amino acid replacement) substitutions and synonymous or silent substitutions, which do not alter the encoded

Fig. 4.2.3

Comparisons involving the conceptually translated sequences of S.c-ß and the *D.virginiana* ϵ^M- and $ß^M$- globin genes (Koop and Goodman, 1988).

Abbreviations: D.v-ϵ:- *D.virginiana* ϵ^M-globin; D.v-ß:- *D.virginiana* $ß^M$-globin; amino acids are abbreviated by the standard one letter code (see appendix); indicated only if a substitution with respect to S.c-ß has been detected in the *D.virginiana* ϵ^M- and $ß^M$-globin sequences.

	10	20	30	40	50	60
S.c-β chain:-	VHFTAEEKNA	ITTIWGKVNV	EETGGEALGR	LLVVYPWTQR	FFDSFGNLSS	ASAILGNPK
β-globin chain:-	VHLSAEEKGV	INGIWSKVSV	DQTGAEALGR	LLIVYPWTSR	FFDHFGDLSS	AKGVMGNAK

TABLE 4.2.2

Comparisons of S.c-β and *D.virginiana* ϵ- and β-globin conceptual amino acid sequences at putative contact sites[1].

* - DPG binding site
\+ - α_1-β_2 contact site
x - α_1-β_1 contact site
h - haem contact site

contact site	1	2	30	33	34	35	36	37	39	40	41	42
type of site	*	*	x	x	x	x	+	+	+	+	h	h
S.c-β	V	H	R	V	V	Y	P	W	Q	R	F	F
D.virg. ϵ-												
D.virg. β-									T			

contact site	51	55	63	66	67	70	82	85	88	91	92	96
type of site	x	x	h	h	h	h	*	h	h	h	h	h
S.c-β	A	L	H	K	V	S	K	F	L	L	H	L
D.virg. ϵ-	M											
D.virg. β-	M							Y				

contact site	97	98	99	101	102	103	106	108	112	115	116	119
type of site	+	+,h	+	+	+,h	h	h	x	x	x	x	x
S.c-β	H	V	D	E	N	F	L	N	I	A	A	N
D.virg. ϵ-											S	G
D.virg. β-											E	G

contact site	122	123	124	125	127	128	131	141	143
type of site	x	x	x	x	x	x	x	h	*
S.c-β	F	T	P	E	Q	A	Q	L	H
D.virg. ϵ-									
D.virg. β-						V			

1. Contact sites are those identified by Perutz *et al.* (1968) and Perutz (1976). DPG binding sites are those identified by Kilmartin (1976) Abbreviations: amino acids by the standard one letter code (see appendix); indicated only if a substitution with respect to S.c-β has been detected in the *D.virginiana* ϵ- and β-globin genes (Koop and Goodman, 1988).

amino acid sequence. This is necessary because the rate of synonymous substitution is known to be much higher than that of non-synonymous substitutions. In comparing closely related sequences, where no more than one nucleotide substitution occurs in each pair of homologous codons, the number of synonymous and non-synonymous substitutions can be determined by directly counting the number of amino acid changing and silent nucleotide substitutions. In more distantly related sequences, where more than one nucleotide substitution may occur in each pair of codons, several evolutionary pathways may exist between the codons and it is more difficult to estimate the number of synonymous and non-synonymous substitutions. Perler *et al.* (1980) developed a statistical procedure for estimating these numbers by assuming equal weights for each possible pathway and averaging the total number of synonymous and non-synonymous substitutions. More elaborate methods were developed by Miyata and Yasunaga (1980) and Li *et al.* (1985) which did not assume equal weights for different evolutionary pathways between pairs of codons. The method of Li *et al.* (1985), which was used in the evolutionary comparisons presented in this section, takes into account different transitional and transversional rates of nucleotide substitution and also considers the relative frequency of codon changes, using weights obtained from comparisons between a wide range of mammalian genes. Each site is classified as either non-degenerate (all possible substitutions are non-synonymous), 2-fold degenerate (substitutions may be silent or non-synonymous) and 4-fold degenerate (all possible substitutions are silent). The non-synonymous and synonymous rates of substitution are computed for each of the above three

classifications. According to Li *et al.* (1985) this method is superior to other methods, particularly when the number of synonymous and non-synonymous substitutions per site approaches one. Computer simulation studies by Nei and Gojobori (1986) indicated that all the current available methods tended to under-estimate the number of non-synonymous substitutions when the number of substitutions per site was very large (> 1). This should not pose any serious problems in the present study as comparisons between the *D. virginiana* and eutherian β-like globin genes indicated that non-synonymous divergence values between these genes were generally low (< 0.3) (Koop and Goodman, 1988).

Results.

The rates of synonymous and non-synonymous substitutions between the coding regions of S.c-β and goat, human, rabbit, chicken, mouse, *D. viverrinus* and *D. virginiana* β-like globin genes, were estimated using the procedure of Li *et al.* (1985). S.c-β differs at non-synonymous sites from the *D. virginiana* ϵ^m and β^m-globin genes by 11% and 17% respectively and from the partial *D. viverrinus* adult β-globin mRNA sequence by 11% (Table 4.2.3a). Over the same partial coding sequence, spanned by the *D. viverrinus* β-globin mRNA, S.c-β differs from the *D. virginiana* ϵ and β-globin genes by 9% and 14% respectively (Table 4.2.4). These data suggest that S.c-β is not orthologous to marsupial adult β-globin genes. Additional evidence for such a suggestion comes from comparisons of S.c-β with eutherian adult and non-

adult β-globin genes (Table 4.2.3a). At non-synonymous sites, S.c-β differs from human ε, mouse ε, human β and mouse β-globin genes by 9%, 13% 14% and 17% respectively (Table 4.2.3a). That S.c-β is more similar to the human ε-globin gene than it is to the adult β-globin mRNA sequence of *D. viverrinus* suggests strongly that S.c-β is orthologous to the *D. virginiana* and eutherian non-adult β-globin genes.

At synonymous sites, S.c-β differs from the *D. virginiana* ε, human ε, *D. virginiana* β and human β-globin genes by 33%, 65%, 88% and 94% respectively (Table 4.2.3b). These results again provide evidence that S.c-β is orthologous to the *D. virginiana* and eutherian non-adult β-globin genes.

At non-synonymous sites the human ε-globin gene is more similar to S.c-β (9% divergence) than it is to the human γ-globin gene (10% divergence). Similarly, the mouse ε-globin gene is more similar to S.c-β (13% divergence) than it is to the human γ-globin gene (14% divergence). On the assumption of constant rates of evolution in each gene lineage, these data provide evidence that the ε and γ-globin gene lineages separated before the divergence of the eutherian and marsupial lineages. In contrast, at synonymous sites, the human γ-globin gene is more similar to the human ε-globin gene (56% divergence) than it is to either the *D. virginiana* ε-globin gene or S.c-β (66% and 65% divergence respectively). These synonymous divergence values support the hypothesis, proposed by Koop and Goodman (1988), that the ε and γ-globin gene lineages separated after the divergence of marsupials and eutherians. The implications of the above results are discussed in Chapter 5.

Another interesting feature of the non-synonymous divergence values (Table

4.2.3a), is the apparent highly conserved nature of the ϵ-globin genes. The maximum non-synonymous divergence value between any pair of ϵ-globin genes is 14%. In contrast, there is generally at least 20% divergence and up to 24% divergence between eutherian and marsupial adult β-globin genes (Table 4.2.3a). S.c-β also appears highly conserved, differing from the chicken adult β-globin gene by just 15% at non-synonymous sites.

The non-synonymous divergence values for individual exons of marsupial and eutherian β-like globin genes are shown in Table 4.2.5. The 2nd exon clearly is the most highly conserved region of the genes. The *D. virginiana* ϵ-globin gene and S.c-β differ by just 4% at non-synonymous sites, compared with 26% and 15% divergence in the 1st and 3rd exons respectively (Table 4.2.5). The non-synonymous divergence value in the 1st exon of these genes (26%) is much higher than the divergence values between S.c-β and the human ϵ, mouse ϵ and *D. virginiana* β-globin genes (13%, 16% and 17% respectively) (Table 4.2.5). It was not possible to determine the statistical significance of these observations due to the generally low (<20) number of substitutions in each individual exon for each class of substitution. It does appear, however, that the *D. virginiana* ϵ-globin gene has evolved faster in the 1st exon than either S.c-β or the *D. virginiana* β, human ϵ and mouse ϵ-globin genes. Further data from other marsupial ϵ-globin genes would be useful to test this hypothesis.

TABLE 4.2.3a

Non-synonymous divergence values (percent) over pairs of globin gene coding regions[1].

S.c-β	-									
D.virg ϵ	11	-								
D.virg β	17	20	-							
D.viv β^*	11	15	10	-						
mouse βh1	13	14	23	17	-					
mouse βmaj	17	20	24	20	21	-				
human β	14	16	21	15	21	13	-			
human ϵ	9	12	19	14	13	16	16	-		
human γ	12	13	23	17	14	18	18	10	-	
chick β	15	20	23	18	20	26	22	18	22	-

TABLE 4.2.3b

Synonymous divergence values (percent) over pairs of globin gene coding regions.

S.c-β	-									
D.virg ϵ	33	-								
D.virg β	88	110	-							
D.viv β^*	62	69	66	-						
mouse βh1	96	107	165	64	-					
mouse βmaj	86	86	91	74	90	-				
human β	94	87	102	70	108	49	-			
human ϵ	65	66	95	66	72	66	62	-		
human γ	80	79	133	68	71	67	73	56	-	
chick β	78	101	128	97	143	87	75	80	88	-

1. Non-synonymous and synonymous divergence values were calculated using the procedure of Li *et al.* (1985): see appendix for DNA sequence data and references of the above genes. * - Partial sequence only.

TABLE 4.2.4

Non-synonymous divergence values (percent) over pairs of globin gene coding regions (partial sequence only)[1].

S.c-β	-									
D.virg ϵ	9	-								
D.virg β	14	16	-							
D.viv β	11	15	10	-						
mouse βh1	13	12	19	17	-					
mouse βmaj	16	18	19	20	22	-				
human β	13	15	20	15	20	10	-			
human ϵ	10	9	15	14	13	19	17	-		
human γ	13	13	19	17	14	21	17	11	-	
chick β	13	16	16	18	17	24	24	17	20	-

1. Non-synonymous and synonymous divergence values were calculated using the procedure of Li *et al.* (1985), using only the region of each sequence spanned by the partial sequence of the *D.viverrinus* β-globin mRNA (Wainwright, 1984): see appendix for sequence data and references of the above genes.

TABLE 4.2.5

Non-synonymous divergence values (percent) over pairs of globin gene coding regions in individual exons[1].

Exon 1

S.c-β	-				
D.virg ϵ	25.9	-			
D.virg β	16.2	34.0	-		
human ϵ	13.5	27.9	30.0	-	
mouse βh1	16.5	25.8	30.1	16.9	-

Exon 2

S.c-β	-				
D.virg ϵ	3.9	-			
D.virg β	7.3	7.0	-		
D.viv β^*	8.6	10.3	6.2	-	
human ϵ	7.2	6.9	10.1	11.5	-
mouse βh1	14.2	9.3	14.0	16.3	13.9	-

Exon 3

S.c-β	-				
D.virg ϵ	15.0	-			
D.virg β	21.0	24.8	-		
D.viv β	14.1	20.6	13.3	-	
human ϵ	12.4	11.6	18.8	16.4	-
mouse βh1	11.9	15.5	22.4	17.3	11.2	-

1. See appendix for sequence data and references for the above genes.

4.3 Studies of the expression of S.c-β: Northern analysis of adult, pouch young and embryonic RNA.

The results given in Section 4.2, while providing evidence that S.c-β is orthologous to the *D.virginiana* and eutherian embryonic β-like globin genes, do not prove that S.c-β is in fact expressed in the embryo. An expression study of the *D.virginiana* ϵ^m-globin gene has not been reported, nor has an embryonic β-chain polypeptide been identified in this species. It is possible that S.c-β may encode a minor adult β-globin chain which has not been detected at the protein level in *S.crassicaudata*. For these reasons, a study of S.c-β expression was undertaken to determine the developmental stage (embryo, pouch young or adult) at which this gene is expressed.

The expression of S.c-β can be examined at either the protein or the RNA level. To investigate whether S.c-β is expressed in embryonic or pouch young tissues at the protein level would require the determination of at least a partial amino acid sequence from an embryonic β-chain polypeptide and comparison of this sequence with the conceptually translated sequence of S.c-β. However, the size of *S.crassicaudata* pouch young (approximately 20 mg at birth) and embryos and the quantity of blood which could be isolated at these developmental stages would severely limit the use of this approach. The expression of S.c-β was therefore examined at the RNA level, using Northern analysis of total RNA isolated from whole embryos, pouch young, and adult bone marrow. To reduce the chance of RNA degradation from the possibly high

levels of RNases in whole animal tissues, RNA was isolated using a guanidinium thiocyanate extraction procedure (Section 3.2.5).

4.3.1 Isolation of RNA from an embryo, pouch young and adult bone marrow.

Individual *S.crassicaudata* pouch young were taken from a single litter over a seven day period. The female parent of the litter was examined just prior to birth and every 24 hours until offspring were detected in the pouch (Doreen Golding, pers. comm.). At this time, the first pouch young was removed from the pouch and RNA was immediately extracted. This animal was therefore less than 24 hours old *post partum* and is referred to here as the "one day old" pouch young. Single pouch young were then taken from the litter every 24 hours for three days (2 day, 3 day and 4 day old pouch young) and a further two pouch young were taken from the same litter after six days (7 day old pouch young). In addition, two pouch young (10 days old *post partum*) were obtained from a second litter. RNA was extracted from each pouch young (whole animal) within five minutes of their removal from the pouch. RNA was also extracted from a single whole embryo and adult bone marrow (Section 3.2.5). The precise age of the embryo was unknown, although the presence of well developed fore and hind limbs, and an enlarged allantois suggest the embryo was within a day or two of birth (Fig. 4.3.1, Dr. Lynne Selwood, pers. comm.). As it was only possible to isolate small quantities of RNA from animals less than four days old (*post partum*) and adult bone marrow, accurate quantification of the RNA using a

spectrophotometer was not possible. Aliquots (5 μl) of each RNA sample were co-electrophoresed on a 1.5% formaldehyde agarose gel together with RNA of a known quantity. Comparisons of the intensity of ethidium bromide fluorescence of RNA in each lane were used to estimate the approximate concentration of RNA in each sample.

4.3.2 Northern analysis of embryo, pouch young and adult bone marrow RNA.

Total RNA (approximately 10 - 20 μg) from each animal was electrophoresed on a 1.5% w/v formaldehyde agarose gel and Northern transferred to Hybond N^+. The filter was probed with the 460 bp HindIII fragment of pSG-2, containing the 1st intron and 2nd exon of S.c-β, under standard hybridization conditions at 65 °C and high stringency washing conditions (0.1 x SSC, 0.1% w/v SDS at 70 °C). These conditions were chosen to minimize cross-hybridization of the S.c-β probe to RNA transcribed from non-allelic β-like globin genes.

To provide comparative information on the relative amounts of RNA in each lane the same Northern filter was stripped of the S.c-β probe and re-probed with an *Aspergillus niger* rRNA gene (kindly provided by Dr. Matthew O'Connell). Hybridization of the Northern filter with the rRNA probe was carried out using standard conditions and washing at low stringency (1 x SSC, 0.1% w/v SDS at 65 °C; see Section 3.2.5.5).

Results.

When the Northern filter was probed with the 460 bp HindIII fragment of S.c-β, a single band was detected in each lane by autoradiography (Fig. 4.3.2a). The hybridization signal in the lane containing embryonic RNA was significantly greater than the signal in all other lanes. The second strongest signal was observed in the lane containing RNA from one day old pouch young. In addition, it can be seen that there was a significant reduction in the signal after day 4 of pouch young development (Fig. 4.3.2a). These results can be explained by either of the following two proposals:

i) S.c-β is expressed in both embryos, pouch young and adult animals and differences in hybridization signals detected were the result of differences in the quantity of mRNA or β-like globin mRNA in each lane of the Northern. Thus the stronger hybridization signal detected in the lane containing embryonic RNA was the result of a significantly greater quantity of mRNA in this lane compared to the quantity of mRNA in all other lanes.

ii) S.c-β is directly homologous to a β-like globin RNA present almost exclusively in embryos and in reduced amounts in pouch young less than 4 days old, and is not directly homologous to β-globin RNA found in pouch young greater than seven days old and adult bone marrow. This latter proposal is compatible with the hypothesis that S.c-β is expressed in embryonic tissues and is either not expressed at all during adult development or is only expressed at low levels.

Fig. 4.3.1

S.crassicaudata embryo within a day or two of birth.

A - amniotic cavity
B - mouth
C - allantois
D - fore limb
E - hind limb
F - portion of yolk sac

The scale of the photo is unknown.

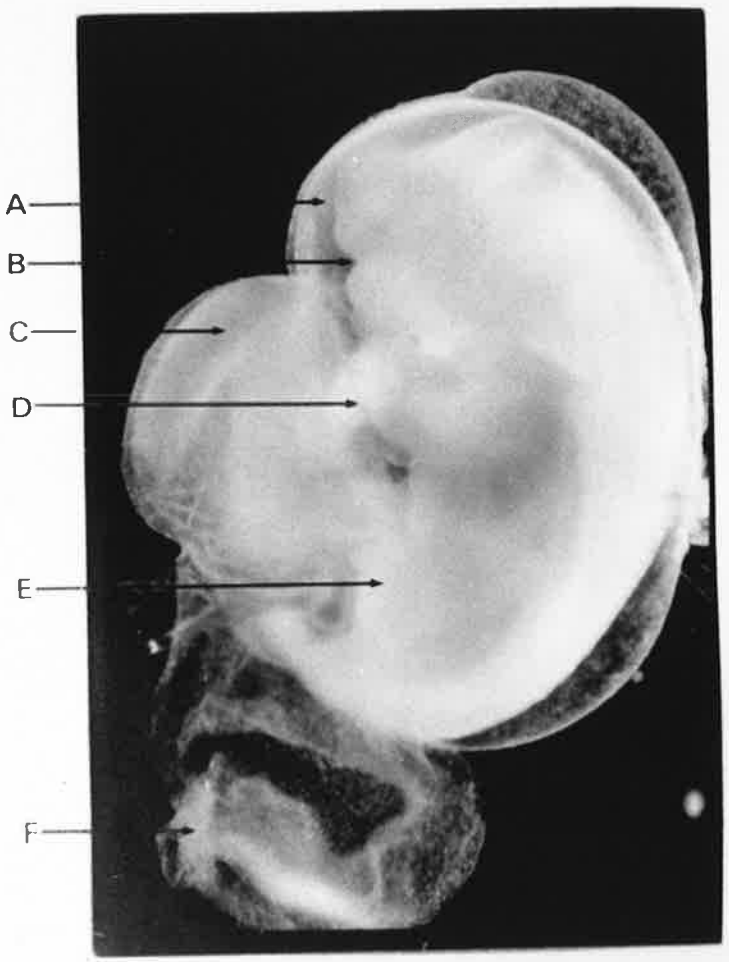

A
B
C
D
E
F

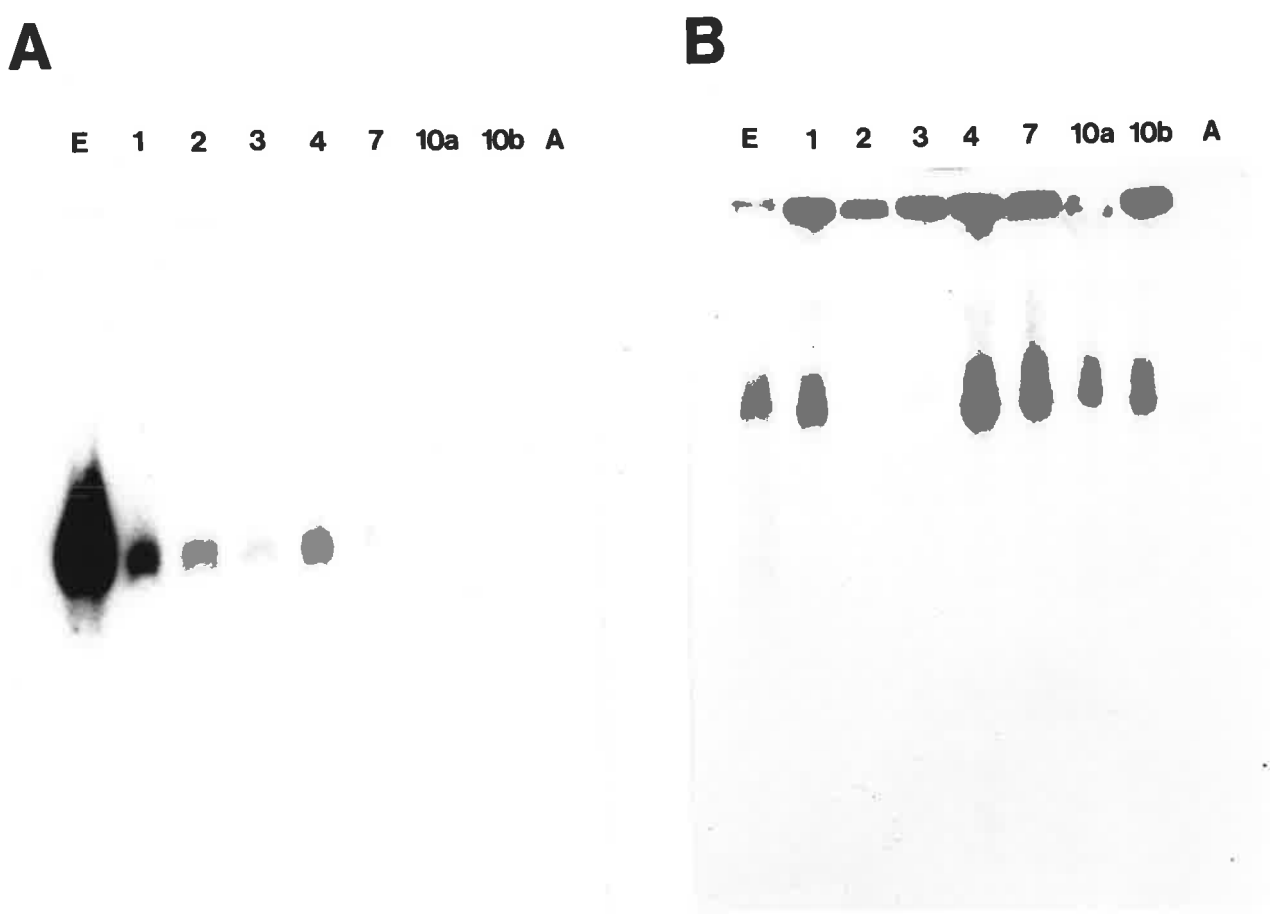
Fig. 4.3.2

Northern analysis of mRNA isolated from an embryo, pouch young (between 1 and 10 days old *post partum*) and adult bone marrow.

Total RNA (approximately 10 - 20μg) from each animal was electrophoresed on a 1.5% w/v formaldehyde agarose gel and Northern transferred to Hybond N^+. Lanes are: E:- embryo RNA; 1:- 1 day old pouch young RNA; 2:- 2 day old pouch young RNA; 3:- 3 day old pouch young RNA; 4:- 4 day old pouch young RNA; 7:- 7 day old pouch young RNA; 10a and 10b:- 10 day old pouch young RNA respectively, each isolated from different animals in the same litter; A:- adult bone marrow RNA (unknown quantity)

A: Autoradiograph of the above Northern probed with the 460bp HindIII fragment of pSG-2 (conditions given in text). The fragment was radio-labelled using the "oligo-labelling" procedure.

B: Autoradiograph of the above Northern probed with an *A.niger* rRNA gene (cloned into a pUC plasmid by Matthew O'Connell, University of Adelaide, 1986). The probe was radio-labelled by nick translation.


A
E 1 2 3 4 7 10a 10b A
B
E 1 2 3 4 7 10a 10b A

When the Northern was probed with the *A.niger* rRNA clone, a single band was again detected in each lane by autoradiography (Fig. 4.3.2b). The hybridization signal varied considerably between lanes, indicating that there were differing amounts of rRNA in each lane (Fig. 4.3.2b). It can be seen that the embryonic RNA hybridized less strongly to the rRNA probe than RNA from 1 day old, 4 day old, 7 day old and 10 day old pouch young. On the assumption that the ratio of rRNA : mRNA does not vary significantly between the embryo (within a day or two of birth) and pouch young less than 10 days old, these results imply that there were lower amounts of embryonic mRNA available for hybridization to the S.c-β probe than mRNA from the above pouch young. Thus the results provide evidence against the first proposal and in favour of the second, ie that S.c-β is expressed in embryonic tissues. However, the possibility that S.c-β is expressed at low levels during adult development can not be definitely ruled out from this study.

The rRNA probe hybridized to 4 day old pouch young RNA at a significantly greater level than RNA from 2 day old and 3 day old pouch young (Fig. 4.3.2b), suggesting that there was a significantly greater amount of 4 day old pouch young mRNA available for hybridization to the S.c-β probe than mRNA from 2 and 3 day old pouch young and therefore a stronger hybridization signal resulted (Fig. 4.3.2a). A less likely explanation for the increased hybridization of 4 day old pouch young RNA to the S.c-β probe is that there was a sudden increase in expression of S.c-β on day 4 followed by a marked decline in expression before day 7 of pouch young development.

Taken overall, the results presented here lend support to the proposition that S.c-β is orthologous to eutherian embryonic β-globin genes and is expressed during embryonic development in marsupials. There is also preliminary evidence to suggest that a switch in β-globin gene expression from S.c-β to an adult or pouch young β-globin gene occurs on or around the time of birth.

4.4 The partial amino acid sequences of the α- and β-globin chains of haemoglobin from adult *S.crassicaudata*.

Results presented in Section 4.3 provided evidence that S.c-β is not the major adult β-globin gene of *S.crassicaudata*. To provide further evidence for this proposal, a partial amino-acid sequence was determined from the β-globin chain of adult haemoglobin and compared to the conceptually translated sequence of S.c-β.

4.4.1 Purification of globin protein from *S.crassicaudata*.

Red cell haemolysates from three adult *S.crassicaudata* were pooled and globin protein was extracted (Section 3.2.6.1). The protein was analysed by electrophoresis on cellulose acetate under denaturing conditions, and it was found to consist predominantly of two major polypeptides (Fig. 4.4.1b). This result was consistent with the electrophoretic patterns of total globin protein from *D.viverrinus* (Wainwright, 1984) and from *Macropus giganteus* (Thompson *et al.*, 1969), with the

"slow" migrating polypeptide being the β-globin chain and the "fast" migrating polypeptide being the α-globin chain (Fig. 4.4.1b).

Chromatography of the globin protein on a CM-sephadex (C50) column resulted in the elution of two major protein peaks (Fig. 4.4.1a). Each peak was analysed by electrophoresis on cellulose acetate. The first peak to be eluted (A) was found to consist almost entirely of the "slow" migrating polypeptide, while the second peak consisted largely of the "fast" polypeptide but was also contaminated with the "slow" polypeptide (Fig.4.4.1).

4.4.2 Determination of partial amino acid sequences from the N-terminal of the purified globin polypeptides.

Amino acid sequencing was done on a Beckman automatic sequencer using 5 mg of purified polypeptide from each peak. The anilino-thiozolinone derivatives taken from the sequencer were converted to phenylthiohydantoin derivatives (called PTH-amino acids). PTH-amino acids were identified with high pressure liquid chromatography (HPLC) which separated all except two pairs of amino acids, PTH-valine and PTH-methionine, and PTH-phenylalanine and PTH-isoleucine. These were identified by thin layer chromatography (TLC) (full details of the above procedures are described in Section 3.2.6). There was some difficulty in distinguishing PTH-histidine by HPLC and in such cases the presence of a histidine residue was inferred from the absence of peaks for the other amino acids.

Fig. 4.4.1

a: Elution of globin from a CM-sephadex column. Globin samples (20mg) were chromatographed on CM-sephadex columns in 8M urea buffers using a linear gradient from 0.04M Na^+ to 0.08M Na^+. The effluent was monitored at 280nm and 3ml fractions were collected. Two major peaks resulted:- labelled A and B. After approximately 85ml of buffer had run through the column, an 8M urea buffer containing 2M Na^+ was pumped through the column to flush out globin protein still bound to the sephadex: hence the higher peak (B) resulted.

b: Cellulose acetate electrophoresis of globin protein.

Freeze dried globin protein was dissolved in electrophoresis buffer (4.66M urea, 45mM Tris-HCl (Ph 9.0), 0.5% v/v 2-mercaptoethanol) and electrophoresed on strips of cellulose acetate (Cellogel 500) at 250V for 3 hours at 4°C. Gels were stained with Amido-black, de-stained and fixed in a solution containing 2% v/v glacial acetic acid, 20% v/v glycerol.

Abbreviations are: N:- neat haemoglobin; A:- protein purified from peak A fractions; B:- protein purified from peak B fractions; O:- origin at which sample was loaded; F:- fast migrating polypeptide; S:- slow migrating polypeptide. The direction of migration of the polypeptides is towards the anode.

Fig. 4.4.2

Amino acid sequences of peak A and peak B proteins and comparison of these sequences by direct alignment from the N-terminal of the conceptual α-chain of *D. viverrinus* (*D.viv.* α; Wainwright, 1984) and the ß-chain of *M. giganteus* (*M.gig.* ß; Air and Thompson, 1969). Amino acids are abbreviated using the standard one letter code (appendix); indicated in the *D.viverrinus* and *M. giganteus* sequences, only if a substitution with respect to the peak A or peak B protein has been detected. Amino acids in parentheses could not be clearly distinguished from other amino acids by HPLC or thin layer chromatography and are therefore uncertain. Note that the sequence of the peak A protein most closely resembles the *M. giganteus* ß-globin chain and the sequence of the peak B protein most closely resembles the *D. viverrinus* conceptual α-globin chain.

```
                      10        20        30        40        50        60
peak A      V(H)LSAEEKGV INGIWSKVSV DQTGAEALG(R) LLIVYPWTS(R) FFD(H)FGDLSS AKGV(M)GNAKV
M.gig. β         T    NA  TSL G  AI E   G                              N   A    A P
D.viv. α       L SD DKTH  KAIWGKVGGH AGAY AEAL A  RTFLSFP T K  TYF P HF   P GSAQ I QGHGK
```

```
                    10        20        30        40        50        60
peak B      VLSGADKANV KAIWGKVGGN AGAYAGEALA (R)TFLSFPTTK TYFP(H)FDL(S)P G(S)ADVKG(H)G(H)
M.gig. β     HLT EEK A ITSLWGKVAI EQTGGEALGR  L LIVYPW SR FFDH F G     N A K  VMAN P K V
D.viv. α       D   TH          H      A                                 QIQ      K
```

Fig. 4.4.3

Comparison of the partial amino acid sequence of the adult ß-chain of *S.crassicaudata* with the conceptually translated sequence of S.c-ß (from the N-terminal). Amino acids are abbreviated using the standard one letter code (appendix); indicated only if a substitution with respect to the S.c-ß chain has been detected in the adult ß-chain. Note that the two chains differ at 20 amino acid sites out of 60.

	10	20	30	40	50	60
S.c-β chain:-	VHFTAEEKNA	ITTIWGKVNV	EETGGEALGR	LLVVYPWTQR	FFDSFGNLSS	ASAILGNPK
β-globin chain:-	VHLSAEEKGV	INGIWSKVSV	DQTGAEALGR	LLIVYPWTSR	FFDHFGDLSS	AKGVMGNAK

A partial amino acid sequence of the polypeptide purified from peak A, consisting of a total of 60 amino acids from the N-terminal, was determined and it was found to share 75% and 13% sequence homology when aligned from the N-terminal of the *M.giganteus* adult β-globin and *D.viverrinus* α-globin chains respectively (Fig. 4.4.2). These results suggest that the polypeptide purified from peak A was the adult β-globin chain of *S.crassicaudata*. A partial amino acid sequence from the N-terminal of the polypeptide purified from peak B was also determined. Contamination of this polypeptide with the β-globin chain from the first peak did not lead to any background problems during the HPLC identification of PTH-amino acids. The sequence from this polypeptide was found to share 85% and 12% sequence homology when aligned from the N-terminal of the *D.viverrinus* α-globin and the *M.giganteus* β-globin chains respectively, providing evidence that it was the adult α-globin chain of *S.crassicaudata* (Fig 4.4.2).

4.4.3 Comparison of the adult β-globin chain of *S.crassicaudata* with the conceptually translated sequence of S.c-β.

Alignment of the partial amino acid sequence of the adult β-globin chain of *S.crassicaudata* and the conceptually translated sequence of S.c-β revealed that there were 20 amino acid differences between the two sequences over 60 amino acid residues (Fig. 4.4.3). These results provide evidence that S.c-β is not the major adult β-globin gene of *S.crassicaudata* and lend support to the proposal that S.c-β is

expressed during embryonic development and is either not expressed or is expressed at low levels during adult life.

4.5 Chromosome walking in the β-globin cluster of *S.crassicaudata*: isolation of phage clones λSG-5, λSG-8 and λSG-1.

4.5.1 Rationale.

Koop and Goodman (1988) presented evidence that just 2 β-like globin genes exist in *D.virginiana* but were unable to confirm the expected 5'-ε-β-3' linkage arrangement of the 2 genes (Section 2.6). This linkage arrangement is expected because the *D.virginiana* ε-globin gene was shown to be orthologous to the embryonic ε-globin gene of eutherians which lies to the extreme 5' end of the β-globin cluster in all eutherian mammals studied to date.

Evidence presented in Sections 4.2 and 4.3 suggests that S.c-β is orthologous to the *D.virginiana* and eutherian ε-globin genes and is expressed in embryonic tissues. Comparison of a partial amino-acid sequence of the adult β-globin chain with the conceptually translated sequence of S.c-β indicates that at least one other functional β-globin gene exists in *S.crassicaudata* (Section 4.4). It is likely that this gene is orthologous to the *D.virginiana* and eutherian adult β-globin genes and is located to the 3' side of S.c-β. The possiblity also exists that there are additional functional and non-functional β-like globin genes in *S.crassicaudata*.

In order to investigate the above propositions, the *S.crassicaudata* genomic DNA library was re-screened with probes designed to detect β-like globin genes and a "chromosome walk" was carried out in both the 5' or 3' directions from S.c-β.

4.5.2 Isolation of phage clones λSG-5, λSG-8 and λSG-1.

Approximately 300,000 bacteriophage pfu's from the *S.crassicaudata* genomic DNA library were screened using a combination of probes derived from the phage clone λSG-3 and the *D.viverrinus* cDNA β-globin clone pDG-5.2. The 3.75 kb SalI / BamHI insert of pSG-5.2, a 2.5 kb EcoRI / SalI fragment at the extreme end of λSG-3 located 3' to S.c-β, and the PstI insert of pDG-5.2 were each purified from LMT agarose gels and labelled to specific activities of $> 10^8$ cpm/μg (Section 3.2.3.6). The probes were pooled and hybridized with plaque lifts from the genomic library under low stringency conditions (hybridization performed at 55 ^{0}C in standard buffer, membrane washing performed at 65 ^{0}C in 1 x SSC, 0.1% w/v SDS). Three plaques showing hybridization to these probes, λSG-8, λSG-5 and λSG-1 were isolated from separate plates and phage clones were purified after 2 rounds of re-plating and re-screening using the same pool of DNA fragments as probes.

DNA from each of the phage clones λSG-1, λSG-5 and λSG-8, was prepared from liquid lysates (Section 3.2.1) and each DNA sample was singly digested with the restriction enzymes EcoRI, BamHI and HindIII, and double digested with BamHI / SalI and EcoRI / SalI. The restriction fragments were separated by electrophoresis

on a 0.8% w/v agarose gel and fragment sizes were estimated by comparison with a known standard. The phage clones λSG-5 and λSG-8 had an identical pattern of restriction fragments (Fig. 4.5.1a) indicating that each contained the same *S.crassicaudata* genomic insert. Restriction maps of the phage clones λSG-5 and λSG-1 were determined for each of the enzymes HindIII, EcoRI, BamHI and SalI (Fig. 4.5.2). HindIII and EcoRI sites were mapped from partial digestions of λSG-1 and λSG-5 DNA (Section 3.2.3).

4.5.3 Characterization of the phage clones λSG-5 and λSG-1: determination of regions of overlap with λSG-3.

λSG-5 DNA was digested with the restriction enzymes EcoRI, SalI, BamHI and HindIII and electrophoresed on a 1% w/v agarose gel (Fig. 4.5.1b). Restriction fragments were Southern transferred to a nitrocellulose membrane and the filter was probed with the 1.1 kb SalI / HindIII fragment of pSG-2 located 5' to S.c-β (Fig 4.1.6). The autoradiograph revealed that the 2.3 kb EcoRI / SalI and 2.5 kb BamHI / SalI fragments, located at one end of the λSG-5 insert, hybridized to the probe (Fig 4.5.1c). The precise position of overlap is shown in Figure 4.5.3. The PstI restriction site of pSG-2 (see Fig. 4.1.6) was also expected to be located in the 2.3 kb EcoRI / SalI fragment of λSG-5, at an estimated distance of 1.3 kb from the SalI site. However, digestion of λSG-5 DNA with PstI and SalI failed to give this expected 1.3 kb fragment, indicating that this PstI restriction site was not present in λSG-5.

Fig. 4.5.1

A: Restriction analysis of the phage clones λSG-5 and λSG-8.

λSG-5 (1) and λSG-8 (2) DNA's were digested with the restriction enzymes: B:- BamHI; H:- HindIII; E:- EcoRI; E/S:- EcoRI/SalI; and fragments were separated by electrophoresis on a 1% w/v agarose gel. Note that λSG-5 and λSG-8 have an identical pattern of restriction fragments.

B: Restriction analysis of λSG-5.

λSG-5 DNA was digested with the restriction enzymes: B:- BamHI; B/S:- BamHI/SalI; S:- SalI; E:- EcoRI; E/S:- EcoRI/SalI; and fragments were separated by electrophoresis on a 1% w/v agarose gel. Molecular weight standard are: M1:- SPP-1 digested with EcoRI; M2:- λ digested with HindIII.

C: Southern analysis of λSG-5.

DNA fragments in the above agarose gel (B) were Southern transferred to Hybond N^+ and the filter was probed with the 1.1kb SalI/HindIII fragment of pSG-2 (radio-labelled using the oligo-labelling procedure). Shown is the autoradiograph after 4 hours exposure at -70°C, with an intensifying screen: note that the 2.5kb BamHI/SalI and 2.3kb EcoRI/SalI fragments hybridize strongly with the probe.

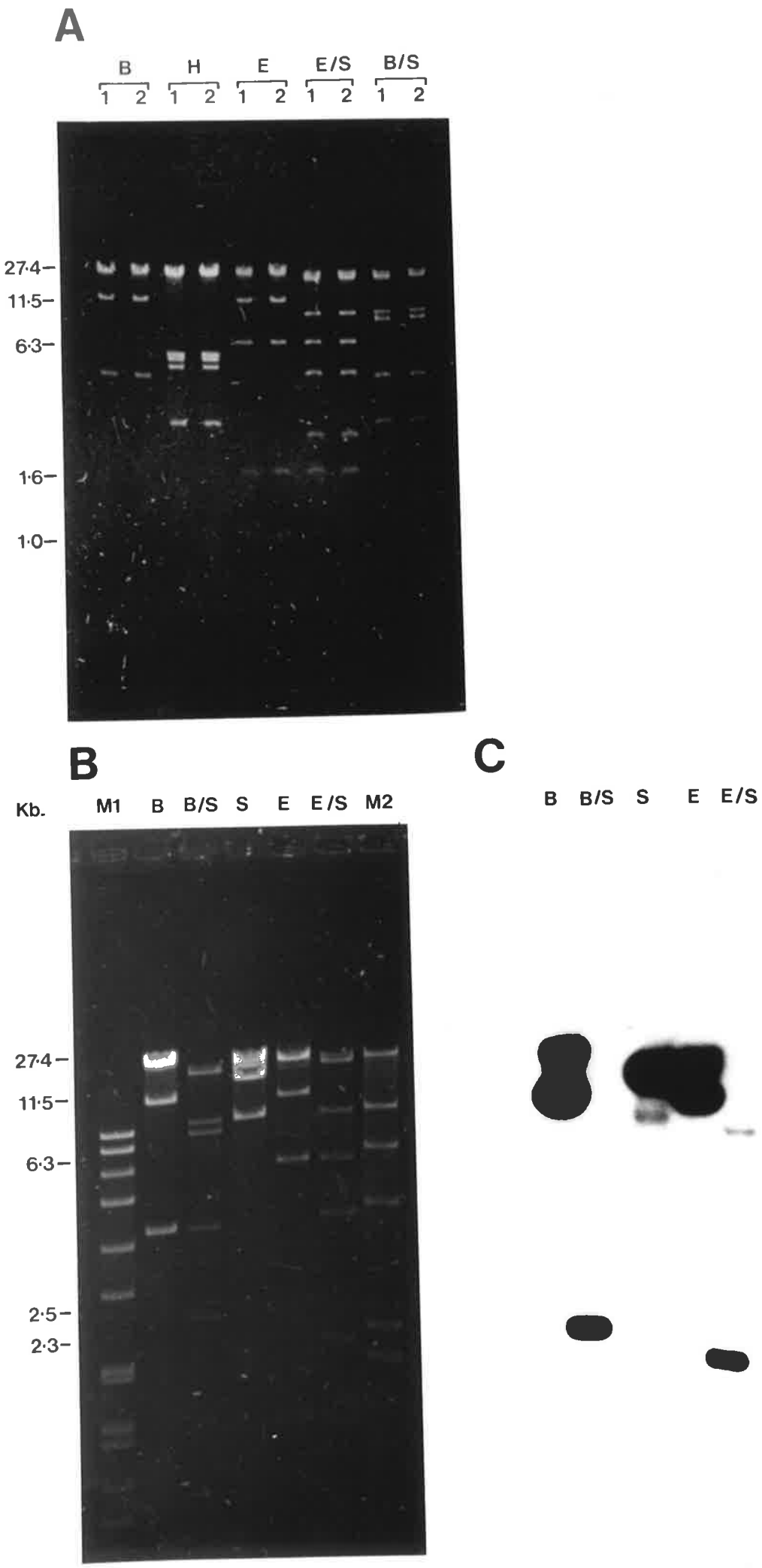
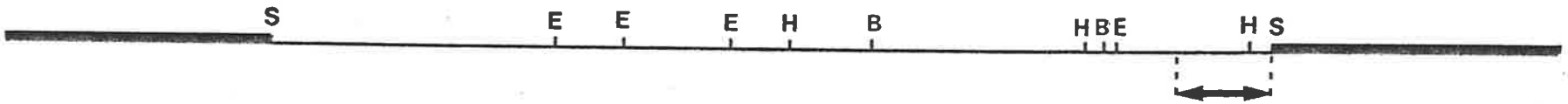

A
B
H
E
E/S
B/S
1 2 1 2 1 2 1 2 1 2
27·4–
11·5–
6·3–
1·6–
1·0–
B
Kb.
M1
B
B/S
S
E
E/S
M2
27·4–
11·5–
6·3–
2·5–
2·3–
C
B
B/S
S
E
E/S

Fig. 4.5.2

Restriction maps of λSG-5 and λSG-1.

Enzymes mapped are: E:- EcoRI; H:- HindIII; B:- BamHI; S:- SalI. No PstI site was found in the arrowed region of λSG-5. This region overlaps with the genomic insert of pSG-2. Thick lines denote the EMBL3 left and right arms which are not drawn to scale.

λSG-5

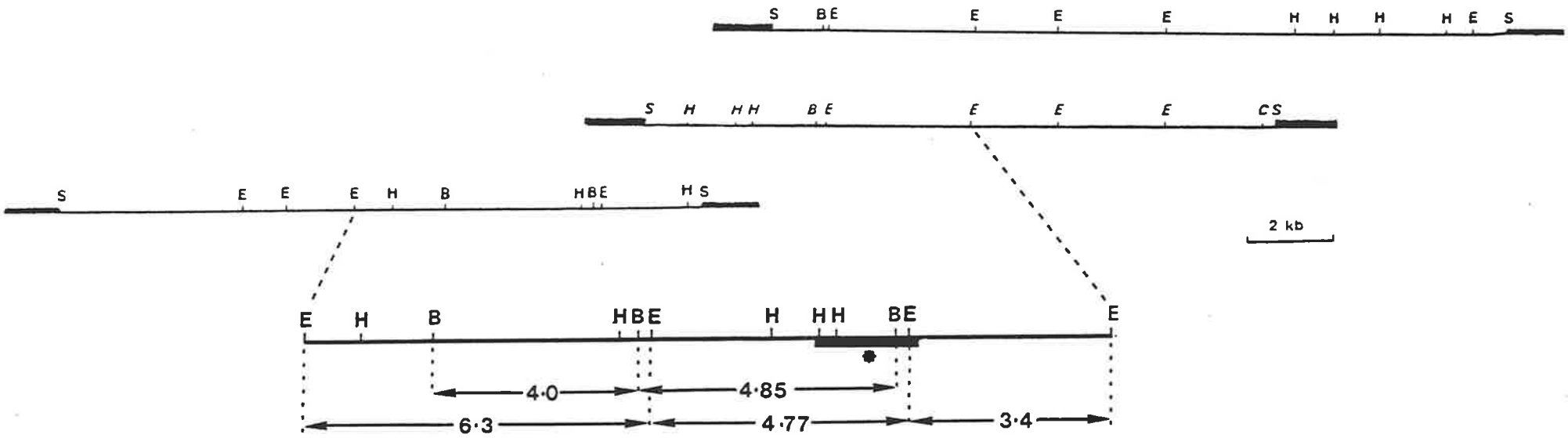


λSG-1

Fig. 4.5.3

Combined restriction maps of λSG-5, λSG-3 and λSG-1 showing the regions of overlap of the 3 clones.

Enzymes mapped are: E:- EcoRI; H:- HindIII; B:- BamHI; S:- SalI. Enlarged is a combined restriction map of EcoRI, HindIII and BamHI sites immediately adjacent to S.c-ß (*). Fragment lengths were inferred from the combined restriction maps of λSG-5 and λSG-3 and are given in kilobases.


S BE E E E H H H H E S
S H HH BE E E E CS
S E E E H B HBE H S
2 kb
E H B HBE H HH BE E
4·0
4·85
6·3
4·77
3·4

The region of overlap between λSG-1 and λSG-3 was deduced by comparing their restriction maps. Each clone contained 2.46 kb, 1.9 kb and 3.4 kb EcoRI restriction fragments. λSG-1 contained the 2nd intron and 3rd exon of S.c-β and extended an additional 6 kb from the end of λSG-3 in a direction 3' to S.c-β, giving a total of 16 kb of 3' flanking sequence (Fig. 4.5.3).

4.5.4 Southern analysis of the *S.crassicaudata* DNA used in the construction of the genomic DNA library.

In order to provide further evidence that the phage clones λSG-3 and λSG-5 contained DNA fragments which overlap, the sizes of genomic restriction fragments containing S.c-β were determined by Southern analysis of *S.crassicaudata* DNA and compared to the sizes of corresponding restriction fragments deduced from the overlapping restriction maps of λSG-3 and λSG-5. The Southern analysis was carried out on the same sample of DNA that was used in the construction of the genomic DNA library.

DNA from *S.crassicaudata* animal 407.1b was digested wth the enzymes BglII, HindIII, PstI, EcoRI and BamHI, electrophoresed on a 0.8% w/v agarose gel and Southern transferred to Hybond N^+ (Section 3.2.7). The filter was probed with the 460 bp HindIII fragment of pSG-2, which contains the first intron and entire second exon of S.c-β. Strong hybridization was detected between the probe and an approximately 4.5 kb BamHI fragment, a 4.4 kb EcoRI fragment, a 1.0 kb BglII

fragment and a 500 bp HindIII fragment (Fig. 4.5.4). These fragment sizes are in concordance with the sizes of restriction fragments, expected to contain S.c-β, which were deduced from the overlapping restriction map of λSG-3 and λSG-5 (Fig. 4.5.3). These results provide further evidence that λSG-3 and λSG-5 contained DNA fragments which overlap.

The autoradiograph also revealed strong hybridization of the probe to three PstI fragments, which were approximately 35 kb, 30 kb and 4.2 kb in size. Although the PstI restriction sites were not mapped in the phage clones λSG-5 and λSG-1, only a single PstI fragment was found in the λSG-3 insert at a position 5' to S.c-β (see Section 4.1.7). The PstI fragment containing S.c-β is therefore at least 14 kb in size. This suggests that S.c-β is located within one of the 30 kb PstI fragments.

4.5.5 Southern analysis of the phage clones λSG-3, λSG-5 and λSG-1 to determine whether β-like globin genes are present in these clones.

DNA samples from the phage clones λSG-5, λSG-3 and λSG-1, were digested with BamHI, EcoRI and HindIII and double digested with BamHI / SalI and EcoRI / SalI. Restriction fragments were separated by electrophoresis on a 1% w/v agarose gel (Fig. 4.5.5a), and Southern transferred to Hybond N^+. The filter was probed with the 460 bp HindIII fragment of pSG-2. The autoradiograph revealed that the probe hybridized strongly to the 4.0 kb BamHI / SalI and EcoRI / SalI fragments of λSG-3, which were shown previously to contain S.c-β (Fig. 4.5.5).

Fig. 4.5.4

Southern analysis of total genomic DNA from animal 407.1B.

407.1b DNA (10μg) was digested with the restriction enzymes: S:- SacI; G:- BglII; H:- HindIII; P:- PstI; E:- EcoRI; B:- BamHI. Fragments were separated by electrophoresis on a 0.8% w/v agarose gel (20cm long) and Southern transferred to Hybond N^+ (Section 3.2.7). The filter was probed with the 460bp HindIII fragment of pSG-2 (Section 3.2.7). Shown is the autoradiograph after exposure for seven days at -70°C with an intensifying screen: note strong hybridization of the probe to 12kb SacI, 1kb BglII, 0.5kb HindIII, 30kb PstI, 4.4kb EcoRI and 4.5kb BamHI fragments.


Kb.
S
G
H
P
E
B
30·0–
12·0–
4·5–
1·0–
0·5–

Fig. 4.5.5

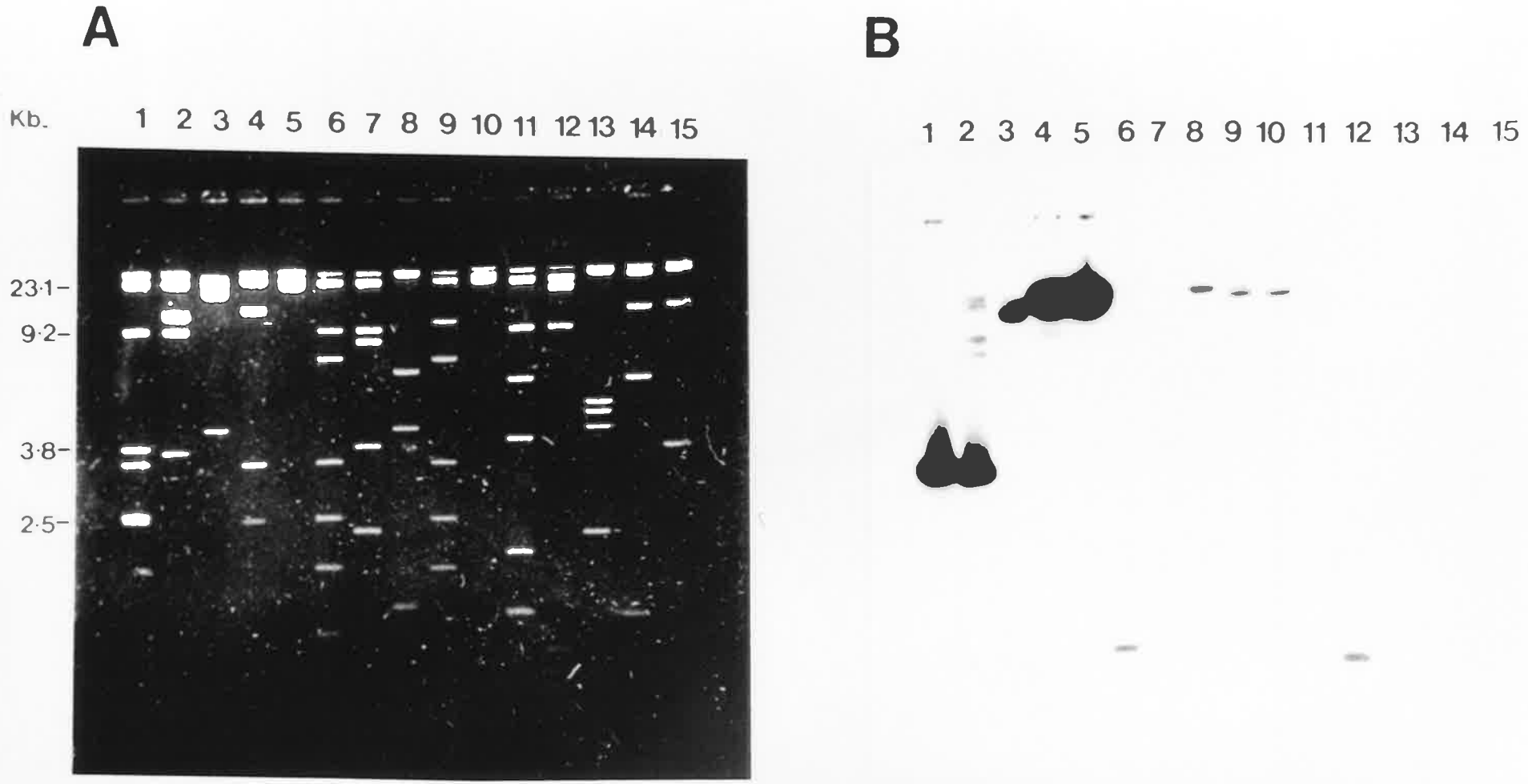
A: Restriction analysis of λSG-3, λSG-1 and λSG-5.

Phage DNA was digested with the enzymes referred to below. Restriction fragments were separated by electrophoresis on a 1% w/v agarose gel (20cm long) at 20V for 20 hours.

Lanes are: 1. λSG-3 - EcoRI/SalI; 2. λSG-3 - BamHI/SalI; 3. λSG-3 - HindIII; 4. λSG-3 - EcoRI; 5. λSG-3 - BamHI; 6. λSG-1 - EcoRI/SalI; 7. λSG-5 - BamHI/SalI; 8. λSG-1 - HindIII; 9. λSG-1 - EcoRI; 10. λSG-1 - BamHI; 11. λSG-5 - EcoRI/SalI; 12. λSG-1 - BamHI/SalI; 13. λSG-5 - HindIII; 14. λSG-5 - EcoRI; 15. λSG-5 - BamHI.

B: Southern analysis of λSG-3, λSG-1 and λSG-5.

DNA fragments from the above gel were Southern transferred to Hybond N^+ and probed with the 460bp HindIII fragment of pSG-2 (radio-labelled using the oligo-labelling procedure) as described in Section 3.2.7. Shown is the autoradiograph after 6 hours exposure at room temperature with no intensifying screen. Note strong hybridization of the probe with 3.8kb EcoRI/SalI and BamHI/SalI fragments of λSG-3.


A
Kb.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
23·1–
9·2–
3·8–
2·5–
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Weak hybridization was also detected with several restriction fragments of the phage clones λSG-1 and λSG-5. All of these fragments mapped to regions immediately adjacent to or within S.c-β. DNA sequence data from these regions, failed to show any significant homology to the 460 bp HindIII fragment of S.c-β (comparisons not shown). Therefore, it is possible that the probe shares sequence homology to a region of EMBL3, which is present in each of these fragments. Alternatively, the probe may have been contaminated with other HindIII fragments of pSG-2, during the isolation of the 460 bp HindIII fragment from a LMT agorose gel. No further fragments showed hybridization to the S.c-β probe, indicating that additional β-like globin genes were not present in the phage clones λSG-1, λSG-3 and λSG-5 (Fig. 4.5.5). The presence in these clones of highly divergent β-globin pseudogenes or genes derived from anciently separated β-like globin lineages can not be excluded.

In order to "walk" further in a 3' direction from S.c-β, a 0.86 kb EcoRI / SalI fragment of λSG-1 was used as a probe to screen approximately 100,000 pfu's of the *S.crassicaudata* genomic DNA library. No hybridizing clones were detected.

4.6 Southern analysis of *S.crassicaudata* genomic DNA: evidence for the presence of two closely linked β-like globin genes.

In the following section the number and arrangement of β-like globin genes in *S.crassicaudata* are investigated by Southern analysis of genomic DNA.

4.6.1 Rationale.

The results presented in Section 4.2, indicate that the 2nd exon of S.c-β, differs in sequence from the *D.virginiana* adult β-globin 2nd exon by 21% over 223 nucleotides and the partial 2nd exon of a *D.viverrinus* adult β-globin mRNA by just 12.6% over 135 nucleotides. Using the appropriate hybridization conditions, a probe containing the 2nd exon of S.c-β (for example the 460 bp HindIII fragment of pSG-2) would be expected to cross-hybridize to the adult β-globin gene of *S.crassicaudata* and to other functional β-like globin genes that were derived by duplication of either the adult or the embryonic β-globin locus. Similarly, a probe containing the 2nd exon region of pDG-5.2 (*D.viverrinus* cDNA adult β-globin clone) would be expected to hybridize to embryonic and adult β-globin genes.

An approximate estimate of the number of β-globin genes in *S.crassicaudata* can be obtained by Southern analysis of genomic DNA: using DNA digested with a wide range of restriction enzymes that recognize sequences of six bases and then directly counting the number of fragments so generated which hybridize to a short β-

globin probe (< 500 bp). Restriction sites are unlikely to fall within the small region to be probed, and usually, therefore, only a single band would be detected per gene copy. Of course if the fragment hybridizing the probe is sufficiently large to contain two β-globin genes, an under-estimate of the number of β-like globin genes may be obtained. In addition, the presence of restriction fragment length variants, which hybridize with the probe, could lead to an over-estimate of the number of β-like globin genes in those individuals which are heterozygous for alternative variants. This latter difficulty can be overcome by Southern analysis of DNA samples from families and by determining those DNA fragments which result from variation at a single β-like globin locus. In addition, if restriction fragment length variants are detected at more than one β-globin locus, it may be possible to use family data to obtain an estimate of the recombination fraction between the loci and therefore contribute to the development of a genetic map of the region.

4.6.2 Preliminary evidence for the existence of at least three β-like globin loci in *S.crassicaudata.*

DNA from animal 1650.1a was digested with the restriction enzymes HindIII, BamHI and EcoRI, Southern transferred to Hybond N^+ and probed with the 460 bp HindIII fragment of pSG-2 (Section 3.2.7). Four EcoRI, three BamHI and three HindIII fragments hybridized to the probe (Fig. 4.6.1a). Strongest hybridization was detected between the probe and a 4.5 kb BamHI, a 4.4 kb EcoRI and a 500 bp HindIII

fragment. These fragments were shown previously to contain regions of S.c-β (see Section 4.5.5).

When the filter referred to above was stripped of the S.c-β probe and re-probed with a 400 bp HindIII / SacI fragment of pDG-5.2 (a fragment containing the 2nd exon region of the *D. viverrinus* cDNA β-globin clone), the strongest hybridization was detected between the probe and the 4.0 kb and 3.2 kb EcoRI, 27 kb BamHI and 15 kb HindIII fragments, each of which had previously hybridized to the S.c-β probe (Fig. 4.6.1b). Given that pDG-5.2 contains a cDNA copy of an adult β-globin mRNA, these results suggest that the above fragments contain a putative adult β-globin gene. The presence of two EcoRI fragments (sizes 4.0 kb and 3.2 kb) that show a similar level of hybridization to the *D. viverrinus* β-globin probe can be explained in at least three ways:-

1. There may be two adult β-globin genes in *S. crassicaudata*.
2. An invariant EcoRI site is located within a single adult β-globin gene in a region of DNA encompassed by the probe.
3. The two fragments may be restriction fragment length variants at a single adult β-globin locus.

The pDG-5.2 probe did not hybridize to the 15 kb BamHI, 7.6 kb EcoRI and 20 kb HindIII fragments which were detected by the S.c-β probe (Fig. 4.6.1), which indicates that either the probe was not sufficiently homologous to these fragments to form a stable hybrid or the probe was missing that portion of the S.c-β probe which had previously hybridized to those fragments.

In order to provide evidence that the restriction fragments which hybridized to the S.c-β and pDG-5.2 probes did not contain an α-globin gene, the Southern filter was again stripped and re-probed with a *D.viverrinus* cDNA α-globin clone (pDG-73) (Wainwright, 1984). The autoradiograph revealed that for each of the restriction enzymes EcoRI, HindIII and BamHI there were two hybridizing fragments neither of which corresponded in size to the fragments which hybridized to the S.c-β and pDG-5.2 β-globin probes (Fig. 4.6.1c).

A second filter was prepared from EcoRI, BglII and SacI digestions of DNA from two animals (1542.1b and 1920.1b) and KpnI and PstI digestions of DNA from three animals (1542.1b, 1527.1c and 1920.1b, in which 1542.1b and 1527,1c were parents of 1920.1b) and was probed with the 460 bp HindIII fragment of pSG-2 (Fig. 4.6.2). This fragment has been shown by DNA sequence analysis to contain single SacI and KpnI sites and no EcoRI, BglII and PstI sites. The numbers of fragments which hybridized to the S.c-β probe are set out below:-

Numbers of fragments which hybridized to S.c-β.

	animal identification number		
Enzyme	1542.1b	1527.1c	1920.1b
EcoRI	4	5*	3
PstI	4	2	2
KpnI	4	4	4
SacI	4	-	4
BglII	3	-	3

* - determined from a separate Southern filter.

Fig. 4.6.1

Southern analysis of total genomic DNA from one animal (1650.1a).

1650.1a DNA (10μg) was digested with the restriction enzymes: E:- EcoRI; B:- BamHI; H:- HindIII. Fragments were separated by electrophoresis on a 0.8% w/v agarose gel (20cm long) at 25V for 24 hours, and Southern transferred to Hybond N^+.

A: Autoradiograph of above filter probed with the 460bp HindIII fragment of pSG-2 (Section 3.2.7), after 5 days exposure at -70°C with an intensifying screen.

B: Autoradiograph of above filter probed with the 400bp HindIII/SacI fragment of pDG-5.2 (a fragment containing the 2nd exon region of the *D. viverrinus* cDNA ß-globin clone); (Section 3.2.7). The filter was autoradiographed for 10 days at -70°C with an intensifying screen.

C: Autoradiograph of above filter probed with the *D.viverrinus* cDNA α-globin clone pDG-73 (Wainwright, 1984); radio-labelled by nick translation (Section 3.2.3.6). Hybridization and washing conditions were as specified in Section 3.2.7. The filter was autoradiographed for 14 days at -70°C with an intensifying screen.

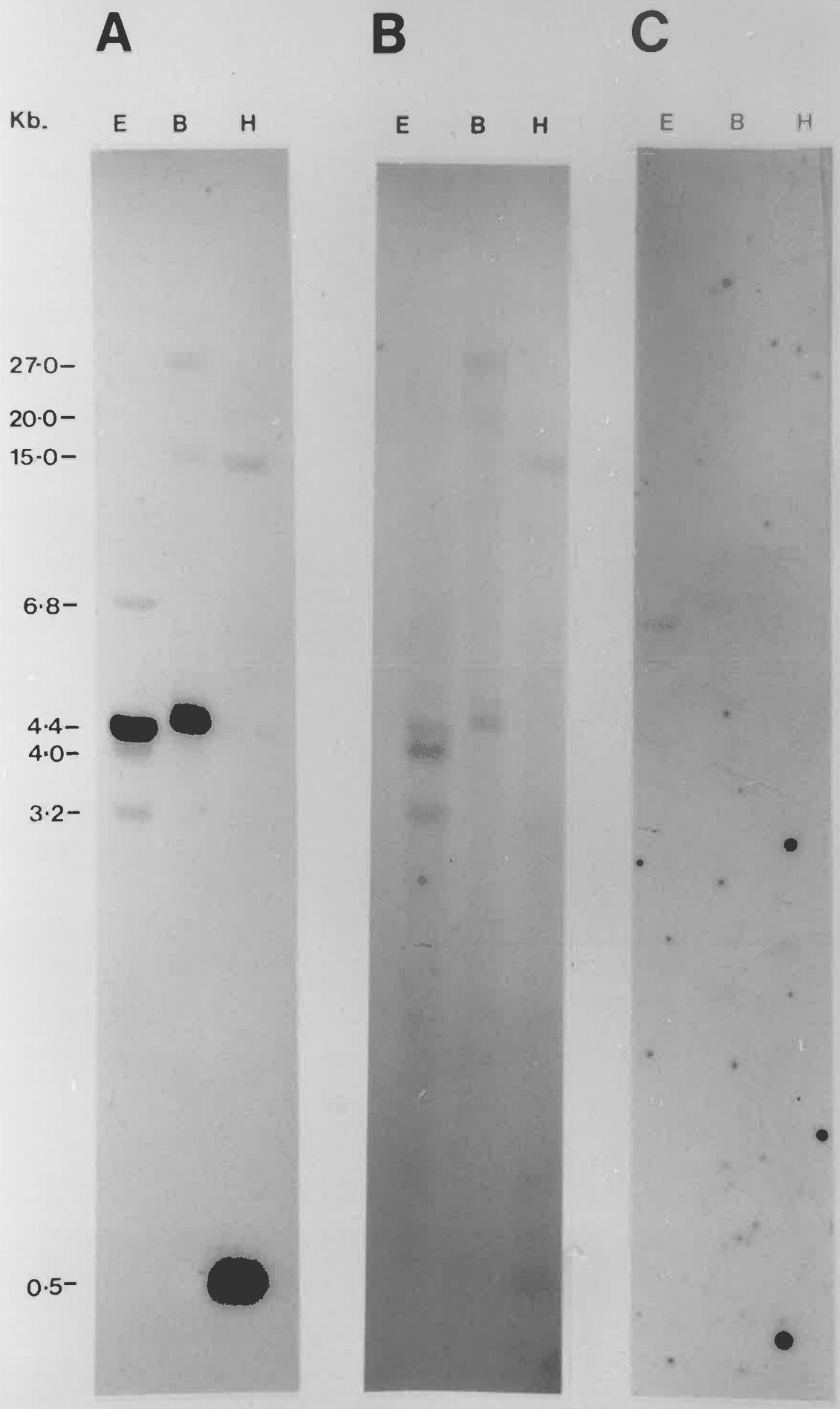
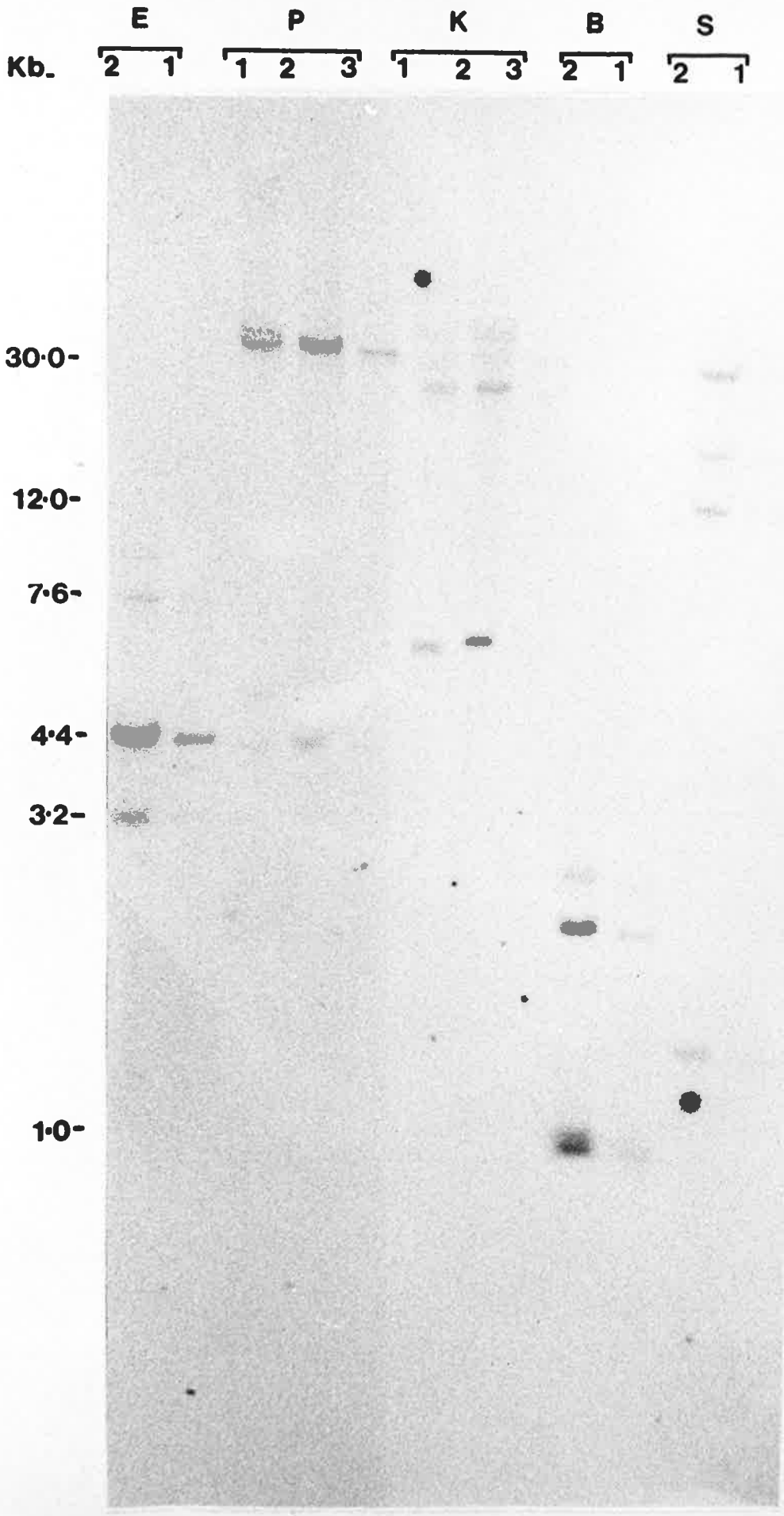

A
B
C
Kb.
E B H
E B H
E B H
27·0–
20·0–
15·0–
6·8–
4·4–
4·0–
3·2–
0·5–

Fig. 4.6.2

Southern analysis of total genomic DNA from three animals: 1542.1b (1); 1920.16 (2); 1527.1c (3).

DNA samples (10μg) were digested with the restriction enzymes: E:- EcoRI; P:- PstI; K:- KpnI; B:- BglII; S:- SacI. Fragments were separated by electrophoresis on a 0.8% w/v agarose gel (20cm long) at 25V for 24 hours, and Southern transferred to Hybond N^+. The filter was probed with the 460bp HindIII fragment of pSG-2 as described in Section 3.2.7. Shown is an autoradiograph after exposure for one week at -70°C with an intensifying screen.


E
P
K
B
S
Kb.
2 1
1 2 3
1 2 3
2 1
2 1
30·0-
12·0-
7·6-
4·4-
3·2-
1·0-

Two of the 4 KpnI and SacI fragments referred to above can be accounted for as containing regions of S.c-β. For the restriction enzymes EcoRI, BglII and PstI, only one fragment could be accounted for as containing S.c-β. These results also indicate that PstI and EcoRI restriction fragment length variants were present in the above individuals.

The results presented in this section provide preliminary evidence for the existence of at least three β-like globin genes (including S.c-β and at least one adult β-globin gene) in *S.crassicaudata*. This conclusion is based on the following observations:-

For six out of seven restriction enzymes used in the Southern analyses, at least three of the resultant fragments hybridized to a short DNA probe containing the 1st intron and 2nd exon of S.c-β. Of these three fragments only one could be accounted for as containing S.c-β. The two remaining fragments therefore contain either: i) separate β-like globin genes or ii) regions of a single adult β-globin gene. This latter possibility is unlikely as it would imply that for the six restriction enzymes referred to above, either the 223 bp 2nd exon of the adult β-globin gene contains the appropriate recognition sequence of the enzyme or the four animals studied above were heterozygous for restriction fragment variants at the adult β-globin locus.

In the following sections the possibility that three β-like globin genes exist in *S.crassicaudata* is further investigated by Southern analysis of DNA samples from families.

4.6.3 The identification of variant BamHI restriction sites linked to S.c-β.

DNA samples from family M1920 (2 parents and 6 offspring) and from family M1939 (2 parents and 4 offspring) were digested with BamHI, Southern transferred to Hybond N^+ and probed with the 460 bp HindIII fragment of pSG-2 (Section 3.2.7). Restriction fragment length variation involving three BamHI fragments of length 4.5 kb, 6.5 kb and 8.9 kb (B1-4.5, B1-6.5, B1-8.9) was detected in M1920 and variation involving two fragments of length 4.5 kb and 8.9 kb (B1-4.5, B1-8.9) was detected in M1939 (Fig. 4.6.3). The inheritance of these variants in each family was in accordance with Mendelian expectations for the inheritance of allelic variants at an autosomal locus. The observation that one of these fragments (4.5 kb) contained S.c-β (Sections 4.5.3 and 4.6.2) indicates that the BamHI variants, detected by the S.c-β probe, result from variation at restriction sites closely linked to S.c-β.

A map of BamHI restriction sites adjacent to S.c-β, deduced from the combined restriction maps of λSG-3 and λSG-5, indicates that each of the BamHI variants could be ascribed to the presence or absence of two BamHI sites located approximately 3 kb and 5 kb upstream of S.c-β (Fig. 4.6.4).

Two BamHI fragments of invariant size were also detected by the S.c-β probe (Fig. 4.6.3). One of these fragments (approximate size 27 kb) was shown to contain a putative adult β-globin gene of *S.crassicaudata* (Section 4.6.2). No variation in length of this fragment was observed in the family M1920, but a doublet of this

fragment appeared in offspring 4 of M1939. The slightly larger of the two fragments was not observed in either of the parents of this offspring, and it is likely that it resulted from partial digestion of genomic DNA.

4.6.4 The identification of variant EcoRI restriction sites linked to S.c-β, the putative adult β-globin locus and a 3rd β-like globin locus.

DNA samples from the family M1625 (2 parents and 8 offspring) were digested with EcoRI, Southern transferred to Hybond N^+ and probed with the 460 bp HindIII fragment of pSG-2 (Section 3.2.7). A complex array of between three and six DNA fragments hybridized to the probe in each individual (Fig. 4.6.5a). The inheritance of these fragment length variants can most simply be interpreted as follows:-

1. The 11.4 kb and 4.4 kb variants (E1-4.4, E1-11.4) contain regions of S.c-β and result from the presence or absence of an EcoRI restriction site located approximately 3 kb upstream of S.c-β (see Fig. 4.6.4). The evidence for this proposal comes from the observation that both fragments hybridized more strongly to the S.c-β probe than all other fragments and that the 4.4 kb EcoRI fragment was previously shown to contain S.c-β (Sections 4.5.4 and 4.6.2).

2. The 4.0 kb, 3.4 kb and 3.2 kb variants (E2-4.0, E2-3.4, E2-3.2) contain regions of an adult β-globin gene and result from the presence or absence of two EcoRI

restriction sites located adjacent to or within this gene. The evidence for this proposal comes from the observation that each of the fragments showed a similar level of hybridization to the S.c-β probe and the 4.0 kb and 3.2 kb EcoRI fragments hybridized more strongly to the pDG-5.2 probe than all other fragments (Section 4.6.2). These results, together with the results reported in Section 4.6.2, provide evidence that a single adult β-globin gene is present in *S.crassicaudata*.

3. The 7.6 kb and 6.8 kb variants (E3-7.6, E3-6.8) contain a third β-like globin sequence (henceforth referred to as the ϵ_2-globin locus) and result from the presence or absence of a single EcoRI restriction site. Each of the fragments hybridized to the S.c-β probe at a lower level than all other fragments and the 6.8 kb fragment failed to hybridize to a pDG-5.2 probe (Section 4.6.2). These results provide evidence that the ϵ_2-globin locus has diverged considerably from both S.c-β and the putative adult β-globin locus and is therefore likely to be a pseudogene (this proposal is further discussed in Chapter 5).

The inheritance of the restriction fragment length variants at each of the above globin loci was consistent with Mendelian expectations for the inheritance of allelic variants at an autosomal locus. Other possible interpretations were found to be either inconsistent with the results reported in Section 4.6.2 or inconsistent with Mendelian expectations.

Fig. 4.6.3

Southern analysis of total genomic DNA from the families M1920 and M1939: detection of BamHI restriction fragment length variants.

DNA samples (10μg) from family M1920 (P1:- 1542.1b; P2:- 1527.1c; 1:- 1920.1a; 2:- 1920.1b; 3:- 1920.1c; 4:- 1920.1d; 5:- 1920.1f; 6:- 1920.1g) and family M19239 (P1:- 1650.1a; P2:- 1571.1f; 1:- 1939.1a; 2:- 1939.1b; 3:- 1939.1d; 4:- 1939.1f) were digested with BamHI. Fragments were separated by electrophoresis on a 0.8% w/v agarose gel (20cm long) at 25V for 24 hours, and Southern transferred to Hybond N^+. The filter was probed with the 460bp HindIII fragment of pSG-2 (Section 3.2.7). Shown is an autoradiograph after exposure at -70 ^{0}C for one week with an intensifying screen.

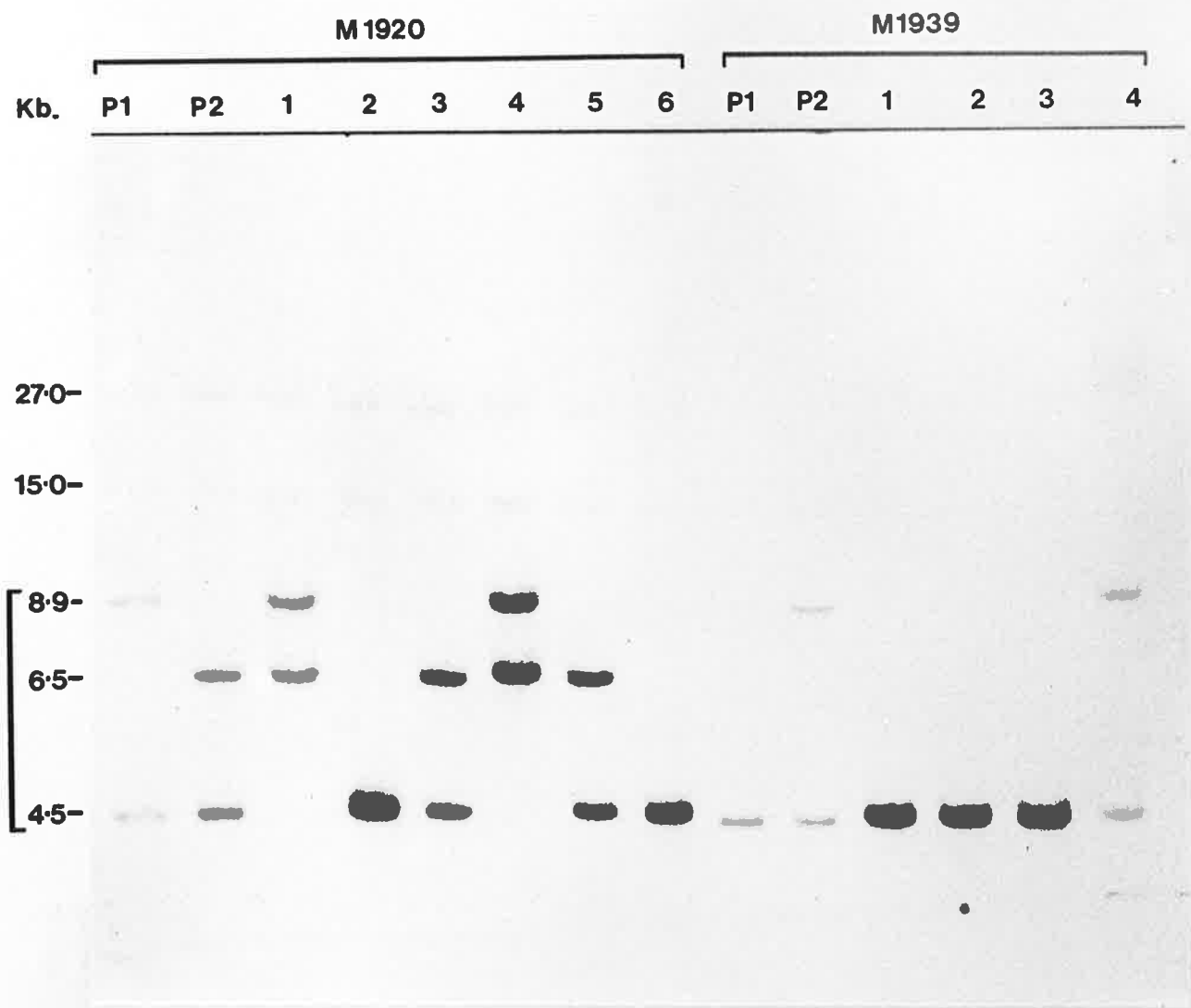

M1920
M1939
Kb.
P1
P2
1
2
3
4
5
6
P1
P2
1
2
3
4
27·0–
15·0–
8·9–
6·5–
4·5–

Fig. 4.6.4

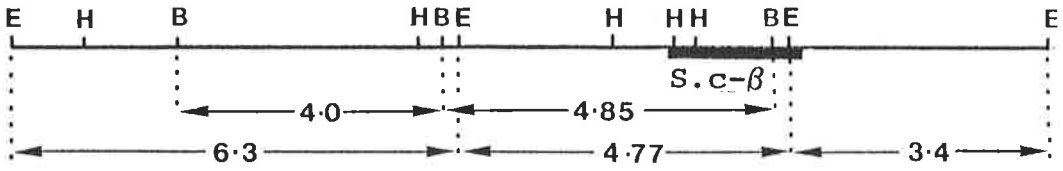
A: Combined restriction maps of λSG-5 and λSG-3 showing restriction sites immediately adjacent to S.c-ß (shown as a thick line).

Enzymes mapped are: E:- EcoRI; H:- HindIII; B:- BamHI. Fragment sizes are kilobases.

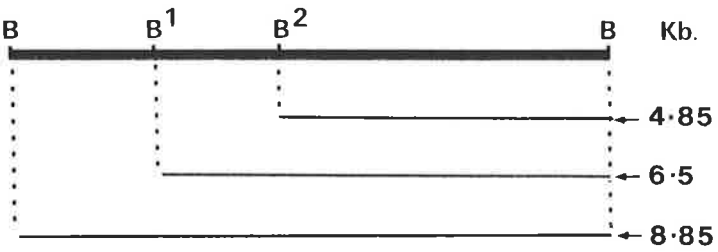
B: BamHI restriction fragment length variants inferred from the combined restriction maps of λSG-5 and λSG-3. The fragment lengths are generated by the presence or absence of the BamHI sites B^1 and B^2. B^2 is present in the phage clone λSG-5. The location of B^1 is predicted from the presence in the *S.crassicaudata* colony of a 6.5 kb BamHI variant, detected by Southern analysis of genomic DNA (Fig. 4.6.3). Note that the fragment lengths: 4.85 kb and 8.85 kb are similar to those detected by Southern analysis of total genomic DNA (4.5 kb and 8.9 kb respectively).

C: EcoRI restriction fragment length variants inferred from the combined restriction maps of λSG-5 and λSG-3. The fragment lengths are generated by the presence or absence of the EcoRI site E^1. Note that the fragment lengths 4.77 kb and 11.07 kb are similar to those detected by Southern analysis of total genomic DNA (4.4 kb and 11.4 kb respectively).

A


E
H
B
HBE
H
HH
BE
E
S.c-β
4·0
4·85
6·3
4·77
3·4

B


B
B1
B2
B
Kb.
4·85
6·5
8·85

C

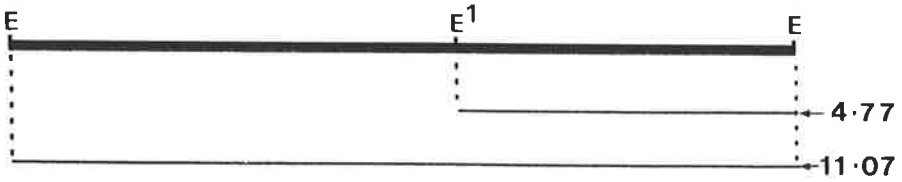

E
E1
E
4·77
11·07

Fig. 4.6.5

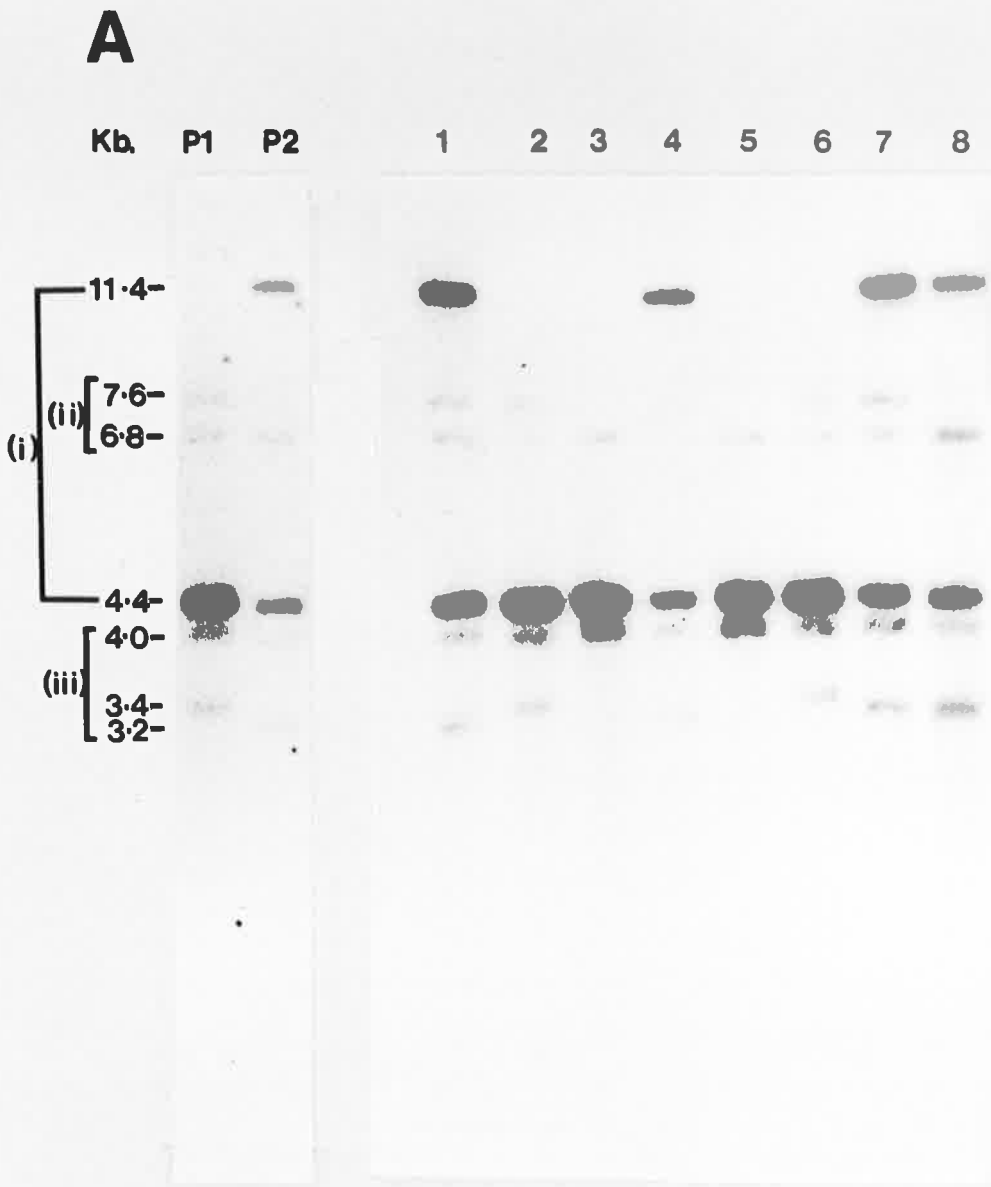
A: Southern analysis of total genomic DNA from the family M1625: detection of EcoRI fragment length variants.

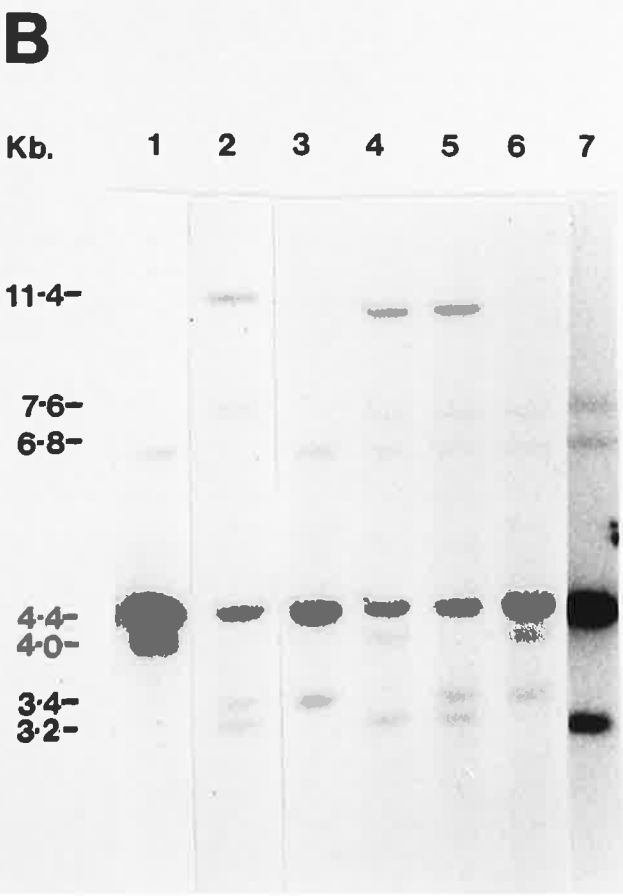
DNA samples (10μg) from the family M1625 (P1:- 1157.1a; P2:- 1211.1d; 1:- 1625.1h; 2:- 1625.1g; 3:- 1625.1f; 4:- 1625.1e; 5:- 1625.1d; 6:- 1625.1c; 7:- 1625.1b; 8:- 1625.1a) were digested with EcoRI. Fragments were separated by electrophoresis on a 0.8% w/v agarose gel (20cm long) at 25V for 24 hours, and Southern transferred to Hybond N^+. The filter was probed with the 460bp HindIII fragment of pSG-2 (Section 3.2.7). Shown is an autoradiograph after exposure at -70°C for one week with an intensifying screen. Numbers in parentheses are:

(i) EcoRI fragment length variants (E1-11.4; E1-4.4) at the S.c-ß locus.
(ii) EcoRI fragment length variants (E3-7.6; E3-6.8) at a third ß-like globin locus.
(iii) EcoRI fragment length variants (E2-4.0, E2-3.4; E2- 3.2) at the putative adult ß-globin locus.

B: Southern analysis of *S.crassicaudata* DNA: genotypes of EcoRI variants at each ß-like globin locus.

Southern filters were prepared and probed with the 460bp HindIII fragment of pSG-2 as described in Section 3.2.7. The autoradiograph shows EcoRI fragment length variation in animals: 1:- 1920.1b; 2:- 1940.2a; 3:- 1940.2d; 4:- 1940.1a; 5:- 1940.1c; 6:- 1940.1b; 7:- 1939.1a (see appendix for genotypes).


A
Kb.
P1
P2
1
2
3
4
5
6
7
8
11·4–
7·6–
6·8–
4·4–
4·0–
3·4–
3·2–
(i)
(ii)
(iii)


B
Kb.
1
2
3
4
5
6
7
11·4–
7·6–
6·8–
4·4–
4·0–
3·4–
3·2–

Southern analysis of DNA from further families revealed examples of six "genotypes" (variants at each locus on each homologous chromosome) of E2 variants at the adult β-globin locus, three "genotypes" of E3 variants at the ϵ_2-globin locus and two "genotypes" of E1 variants at the S.c-β locus (Fig. 4.6.5b). These results lend support to the above interpretation of the inheritance of these EcoRI variants.

4.6.5 Linkage analysis of the S.c-β, adult β- and ϵ_2-globin loci.

Parents of 14 litters (containing > 4 offspring) were typed for the BamHI variants at the S.c-β locus (B1) and EcoRI variants at the S.c-β locus (E1), the adult β-globin locus (E2) and the ϵ_2-globin locus (E3) (Table 4.6.1). These typings were used to determine, for each pair of globin loci, which families contained both one doubly heterozygous parent and one doubly homozygous parent. Assuming linkage between the two globin loci, a recombination fraction can be estimated from these families by determining the proportion of recombinant "genotypes" in the offspring. The recombination fraction can thus be estimated between the S.c-β and adult β-globin loci in the families M1355, M1940.1, M1940.2, M1382, M1361 and M1625, representing a total of 43 offspring; between the S.c-β and ϵ_2-globin loci in the families M1382, M1746, M1625, M1920 and M1939, representing a total of 30 offspring; and between the adult β- and ϵ_2-globin loci in the families M1625 and M1920, representing a total of 14 offspring.

TABLE 4.6.1. BamHI and EcoRI variants in parents of *S.crassicaudata* families.

Family	N	Animal	S.c-β locus		β-globin locus	ϵ_2-globin locus
			BamHI	EcoRI	EcoRI	EcoRI
M1355	5	1076.1a	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.2 E2-3.4	E3-6.8 E3-6.8
		1040.1f	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
M1940.1	8	1520.1b	B1-8.9 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-7.6 E3-6.8
M1940.2	9	1443.2e	B1-4.5 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
M1382	5	1220.1c	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1073.1f	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
M1230	5	1073.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-3.2	E3-6.8 E3-6.8
		960.1d	B1-8.9 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-6.8 E3-6.8
M1746	7	1494.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
		1471.1d	B1-8.9 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-7.6 E3-6.8
M1593	6	1316.1c	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-7.6*
		1137.1f	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
M1361	8	953.1c	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-6.8 E3-6.8
		1073.1g	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	-
M1625	8	1157.1a	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1211.1d	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
M1520.3	6	1199.1b	B1-4.5 B1-4.5	-	-	-
		1132.1c	B1-4.5 B1-4.5			
M1381		1181.1a	B1-4.5 B1-8.9	-	-	-
		1068.1d	B1-4.5 B1-6.5			
M1877	7	1492.1b	B1-8.9 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-7.6 E3-6.8
		1508.1e	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
M1920	6	1542.1b	B1-6.5 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-6.8 E3-6.8
		1527.1c	B1-4.5 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
M1939	4	1650.1a	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1571.1f	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8

S.crassicaudata from all the above informative families and the families M1593 and M1877, were typed for BamHI and EcoRI fragment length variants at each of the S.c-β, adult β- and ϵ_2-globin loci (Table A1, Appendix). For technical reasons, DNA samples from the families M1361 (excluding 1361.1f and 1361.1g) and M1746 could not be typed for the EcoRI variants at each locus. Individuals in these families, however, were typed for the BamHI variants at the S.c-β locus. The "genotypes" of each parent were derived for each pair of globin loci between which the recombination fraction could be estimated (Table A2, Appendix). Each of the offspring was then classified as having either a parental or recombinant "genotype" (assuming a given linkage phase in the doubly heterozygous parent) (Table A2). Lod scores were determined for recombination fractions between 0 and 0.45, in increments of 0.05, using tabulated lod score data for families with less than 8 offspring, and by direct calculation for families with 8 or more offspring using the procedure of Smith (1968) (Tables 4.6.2a, b and c). Total lod scores were determined for each recombination fraction (θ).

For linkage analysis of the S.c-β and adult β-globin loci, the maximum lod score ($z = 8.428$) was obtained at a recombination fraction of $\theta = 0.00$, providing strong evidence that the embryonic (S.c-β) and putative adult β-globin loci of *S.crassicaudata* are closely linked (Table 4.6.2a).

For linkage analysis of the S.c-β and ϵ_2-globin loci, "genotypes" were available for a total of 23 offspring. For a recombination fraction of $\theta = 0.1$ a total lod score of $z = -2.905$ was obtained and a maximum lod score of $z = -0.02$ was obtained at

$\theta = 0.45$ (Table 4.6.2b). These results provide evidence that the S.c-β and ϵ_2-globin loci are not closely linked.

For linkage analysis of the adult β- and ϵ_2-globin loci, "genotypes" were available from families M1625 and M1382 only, representing a total of 13 offspring. Analysis of these data provided evidence that the two loci are not closely linked ($z < -2$ at a recombination value of $\theta = 0.1$, see Table 4.6.2c).

Haplotypes for the S.c-β and adult β-globin loci were determined for each parent of the 10 families referred to above and M1230, and are shown below:-

Haplotypes	Frequency
B1-4.5 E1-4.4 E2-3.2	11
B1-6.5 E1-4.4 E2-3.4	6
B1-8.9 E1-11.4 E2-3.2	5
B1-8.9 E1-4.4 E2-3.4	8
B1-4.5 E1-4.4 E2-4.0	14

At the S.c-β locus three B1 variants and two E1 variants were detected and at the adult β-globin locus three E2 variants were detected. Therefore 18 haplotypes are theoretically possible, taking account of these variants. Of the 18 possible haplotypes only five were present in the 21 parents and since no recombination was detected between the S.c-β and adult β-globin loci, only five haplotypes were found in the offspring of these 20 parents. These results indicate that recombination between the adult β-globin and S.c-β loci is rare, providing further evidence that they are closely linked.

Table 4.6.2a. Data on linkage between ϵ and ϵ_2^1.

family	Z	Recombination Fraction (θ) 0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
M1382	Z(3:2)	$-\infty$	-1.44	-0.89	-0.59	-0.39	-0.25	-0.15	-0.08	-0.04	-0.01
M1625	Z(5:3)	$-\infty$	-1.91	-1.12	-0.70	-0.45	-0.15	-0.09	-0.04	-0.02	-0.00
M1920	Z(4:2)	$-\infty$	-1.19	-0.67	-0.41	-0.25	-0.03	-0.01	-0.00	-0.00	-0.00
M1939	Z(3:1)	$-\infty$	-0.46	-0.23	-0.12	-0.06	-0.28	-0.16	-0.09	-0.04	-0.01
TOTAL		$-\infty$	-5.00	-2.91	-1.82	-1.15	-0.71	-0.21	-0.21	-0.10	-0.02

Table 4.6.2b. Data on linkage between β and ϵ_2^1.

M1625	Z(5:3)	$-\infty$	-1.91	-1.12	-0.70	-0.45	-0.15	-0.09	-0.04	-0.02	-0.00
M1382	Z(3:2)	$-\infty$	-1.44	-0.89	-0.59	-0.39	-0.25	-0.15	-0.08	-0.04	-0.01
TOTAL		$-\infty$	-3.35	-2.01	-1.29	-0.84	-0.40	-0.24	-0.12	-0.06	-0.01

Table 4.6.2c. Data on linkage between ϵ and β^1.

M1940	Z(15:0)	3.91	3.88	3.53	3.16	2.76	2.34	1.89	1.41	0.89	0.34
M1382	Z(5:0)	1.20	1.09	0.98	0.85	0.72	0.58	0.44	0.29	0.15	0.04
M1355	Z(5:0)	1.20	1.09	0.98	0.85	0.72	0.58	0.44	0.29	0.15	0.04
M1625	Z(8:0)	2.11	1.93	1.74	1.54	1.33	1.11	0.87	0.61	0.35	0.11
TOTAL		8.42	7.99	7.23	6.40	5.53	4.61	3.63	2.60	1.54	0.53

1. Lod scores (z) were calculated using tabulated Lod score data for families with less than 8 offspring and by direct calculation in families with more than 8 offspring using the procedure of Smith (1968).

4.6.6 Detection of PstI fragment length variants in the *S.crassicaudata* family M1920.

DNA samples from family M1920 were digested with PstI, Southern transferred to Hybond N^+ and probed with the 460 bp HindIII fragment of S.c-β (Section 3.2.7). Four variant fragment lengths of sizes 35 kb, 30 kb, 5.0 kb and 4.2 kb were detected (Fig. 4.6.6). The 30 kb and 35 kb fragments segregated in accordance with Mendelian expectations for autosomal inheritance in this family, but did not segregate individually from either of the 5.0 kb and 4.2 kb fragments. The 5.0 kb and 4.2 kb PstI fragments also segregated in accordance with Mendelian expectations for autosomal inheritance. The inheritance of the PstI variants (P1-35, P1-30 and P2-5.0, P2-4.2) in M1920 provides evidence that the P1 and P2 loci are not closely linked ($z < -2$ at $\theta = 0.01$, see Fig. 4.6.7). Strongest hybridization of the S.c-β probe was detected with the 30 kb and 35 kb fragments suggesting that these fragments contain S.c-β. Therefore the third locus ϵ_2 is likely to be located in the 5.0 kb and 4.2 kb PstI fragments and the adult β-globin locus must also be located in the 31 kb and 35 kb PstI fragments.

Further evidence for these proposals was obtained by stripping the membrane of the S.c-β probe and re-probing it with the *D.viverrinus* cDNA β-globin clone pDG-5.2 (Fig. 4.6.6). This probe clearly hybridized to the 35 and 30 kb PstI fragments but not to the 5.0 kb and 4.2 kb fragments, providing evidence that the S.c-β and adult β-globin loci are both located on the 30 kb and 35 kb PstI fragments.

pDG-5.2 also hybridized to two other fragments (of size 6.6 kb and 2.9 kb)

which did not hybridize to the S.c-β probe. These fragments presumedly hybridized to the 3rd exon region of pDG-5.2 which suggests that a PstI restriction site is located in the adult β-globin gene in a region of DNA encompassing the 2nd intron and 3rd exon. In order to explain the inheritance of the 6.6 kb and 2.9 kb PstI variants, at least one other fragment not detected by the S.c-β probe must be present in M1920. It is likely that this fragment is small and either may not have transferred to the membrane or may have been electrophoresed off the agarose gel which was used in the transfer. Three weakly hybridizing fragments were also detected by the S.c-β probe (Fig. 4.6.6) suggesting that further divergent β-like globin loci may also exist in *S.crassicaudata*.

Restriction analysis of the plasmid subclone pSG-2 which contains the 3.75 kb BamHI / SalI fragment of the phage clone λSG-3, showed that a PstI restriction site was located 1.85 kb upstream of S.c-β (Fig. 4.1.6). This site was not present in the overlapping phage clone λSG-5 at this position (Section 4.5.3). In addition, Southern analysis of animal 407.1b, used in the construction of the genomic DNA library, indicated that this animal contained both the 35 kb and the 30 kb restriction fragments (Section 4.5.4). These results suggest that the 35 kb and 30 kb PstI variants can be ascribed to the presence or absence of a PstI restriction site located 1.88 kb 5' to the CAP site of S.c-β which further suggests that the adult β-globin gene is located to the 3' side of S.c-β, approximately 26 kb downstream (as the PstI site of pSG-2 is $\sim$ 4 kb from the end of the 3rd exon of S.c-β). Since λSG-1 contains 16 kb of 3' flanking sequence of S.c-β, this clone is therefore within 10 kb of the adult β-globin gene.

Fig. 4.6.6

A: Southern analysis of total genomic DNA from the family M1920: detection of PstI restriction fragment length variants.

DNA samples (10μg) from M1920 (P1:- 1542.1b; P2:- 1527.1c; 1:- 1920.1a; 2:- 1920.1b; 3:- 1920.1c; 4:- 1920.1d; 5:- 1920.1f; 6:- 1920.1g) were digested with PstI. Fragments were separated by electrophoresis on a 0.8% w/v agarose gel (20cm long) at 25V for 24 hours, and Southern transferred to Hybond N^+. The filter was probed with the 460bp HindIII fragment of pSG-2 (Section 3.2.7). Shown is an autoradiograph after 5 days exposure at -70°C with an intensifying screen. Note strong hybridization of the probe with 30kb and 35kb PstI fragments.

B: The Southern filter referred to above was stripped and re-probed with the 400bp PstI insert of pDG-5.2 as described in section 3.2.7. Shown is an autoradiograph after 10 days exposure at -70°C with an intensifying screen. The autoradiograph again shows strong hybridization of the probe with 30kb and 35kb PstI fragments.

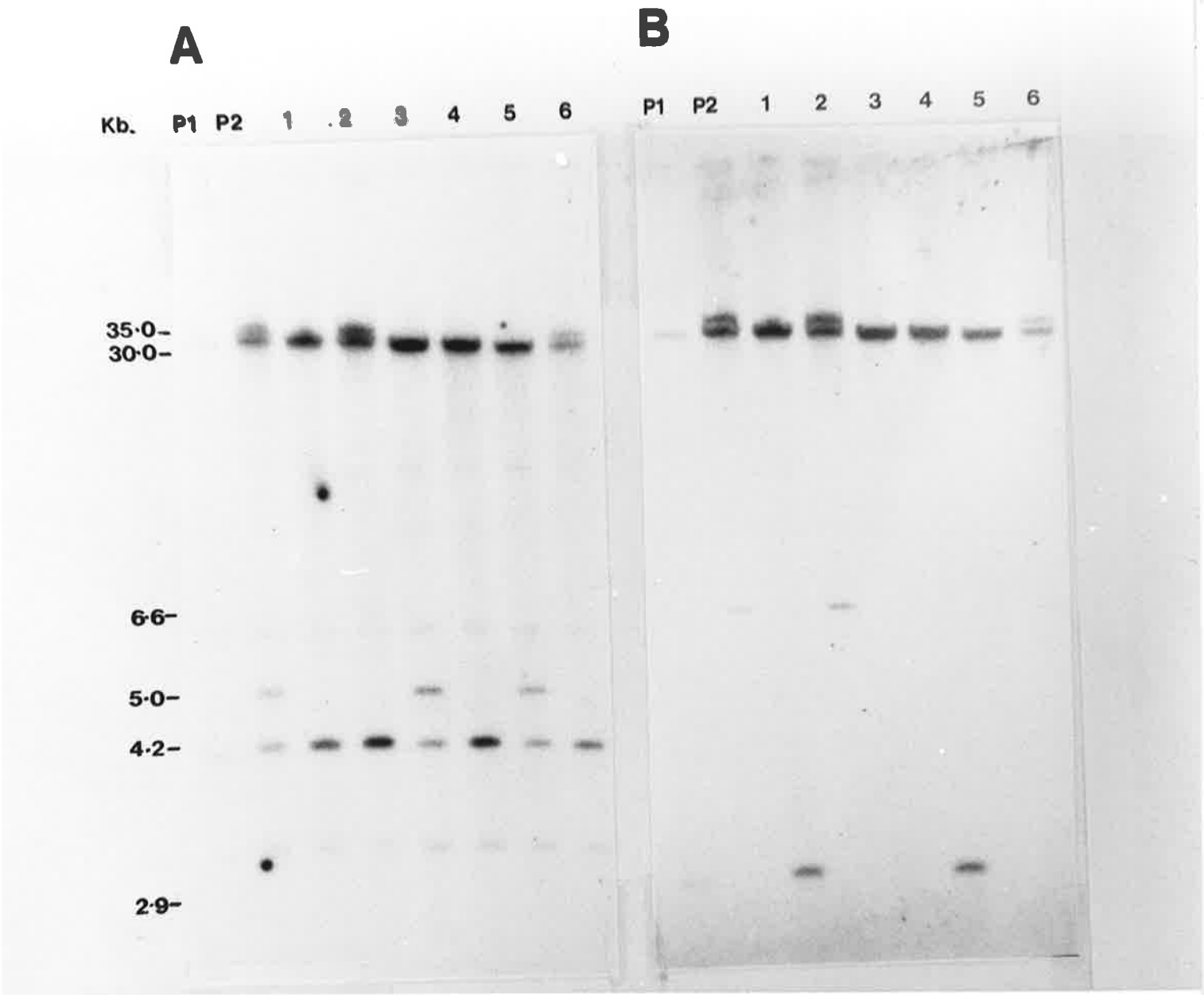

A
B
Kb.
P1
P2
1
2
3
4
5
6
35·0–
30·0–
6·6–
5·0–
4·2–
2·9–

Fig.4.6.7

Inheritance of PstI fragment variants in family M1920. Two linkage phases are possible in animal 1527.1c. Offspring are classified as either parental (no recombination between loci) or recombinant genotypes assuming a set linkage phase of the doubly heterozyous parent 1527.1c.

parents:- 1527.1c x 1542.1b

i) $\frac{\text{P1-35 P2-5.0}}{\text{P1-30 P2-4.2}}$ $\frac{\text{P1-30 P2-4.2}}{\text{P1-30 P2-4.2}}$

or

ii) $\frac{\text{P1-35 P2-4.2}}{\text{P1-30 P2-5.0}}$

linkage phase i)

"genotypes" in offspring			
parental		recombinant	
$\frac{\text{P1-35 P2-5.0}}{\text{P1-30 P2-4.2}}$	$\frac{\text{P1-30 P2-4.2}}{\text{P1-30 P2-4.2}}$	$\frac{\text{P1-35 P2-4.2}}{\text{P1-30 P2-4.2}}$	$\frac{\text{P1-30 P2-5.0}}{\text{P1-30 P2-4.2}}$
0	2	2	2

linkage phase ii)

"genotypes" in offspring			
parental		recombinant	
$\frac{\text{P1-35 P2-4.2}}{\text{P1-30 P2-4.2}}$	$\frac{\text{P1-30 P2-5.0}}{\text{P1-30 P2-4.2}}$	$\frac{\text{P1-35 P2-5.0}}{\text{P1-30 P2-4.2}}$	$\frac{\text{P1-30 P2-4.2}}{\text{P1-30 P2-4.2}}$
2	2	0	2

4.7 Evolutionary studies: relative rate tests of eutherian ϵ-globin genes.

In order to examine whether nucleotide substitution rates have been uniform during the evolution of the ϵ-globin gene in different lineages of eutherian mammals, relative rate tests were carried out using S.c-β as an outgroup.

4.7.1 Rationale.

Evolutionary comparisons of S.c-β with the *D. virginiana* and eutherian β-like globin genes provided evidence that S.c-β is orthologous to the ϵ-globin genes of eutherians (Section 4.2). This conclusion was supported by expression studies (Section 4.3) and by comparisons of the conceptually translated amino acid sequence of S.c-β with a partial amino acid sequence of the adult β-globin chain of *S. crassicaudata* (Section 4.4).

The marsupial and eutherian lineages diverged approximately 130 mya and according to Baverstock *et al.* (1990), the major radiation of eutherian mammals occurred approximately 50 - 55 mya. S.c-β therefore forms an orthologous outgroup of eutherian ϵ-globin genes. The rates of non-synonymous and synonymous nucleotide substitutions per site between S.c-β and eutherian ϵ-globin genes were found in all cases to be less than one (Section 4.2). This indicates that errors in the calculation of non-synonymous and synonymous rates of substitution are likely to be minimal and

should not affect the conclusions drawn from relative rate tests of eutherian ϵ-globin genes.

4.7.2 **Results.**

The relative rate test procedure of Li *et al.* (1985) was used to compare nucleotide substitution rates amongst the human, goat, rabbit and mouse ϵ-globin genes and the human γ-globin gene using S.c-β as an outgroup (Table 4.7). Three other tests were also performed using i) the *D. virginiana* ϵ-globin gene as an outgroup for a comparison of the human and mouse ϵ-globin genes, ii) the mouse ϵ-globin gene as an outgroup for a comparison of the *S. crassicaudata* and *D. virginiana* ϵ-globin genes and iii) the *D. virginiana* β-globin gene as an outgroup for a comparison of the human and mouse β-globin genes (Table 4.7). The statistical significance associated with each test can be determined using a standardized normal test, if the total number substitutions between the two sequences compared is larger than 20 (Wu and Li, 1985). In this study the number of substitutions at two-fold degenerate sites was generally too low to use a standardized normal test.

The results (Table 4.7) indicate that during the evolution of the ϵ-globin gene in eutherians, the mouse ϵ-globin gene has evolved faster than the goat, rabbit and human ϵ-globin genes at both non-synonymous and synonymous sites. In addition, at non-degenerate and four-fold degenerate sites the differences in substitution rates of the mouse and goat ϵ-globin genes were statistically significant (at the 5% level of

probability). The human and rabbit ϵ-globin genes appear to have evolved at similar rates at both synonymous and non-synonymous sites. These results provide evidence that nucleotide substitution rates have not been uniform during the evolution of the ϵ-globin gene in eutherian mammals.

When the human and mouse ϵ-globin genes were compared using the *D. virginiana* ϵ-globin gene as an outgroup, the results were similar to those found using S.c-β as an outgroup, with the mouse ϵ-globin gene evolving faster than the human ϵ-globin gene at each site (Table 4.7). When S.c-β and the *D. virginiana* ϵ-globin gene were compared using the mouse ϵ-globin gene as an outgroup, the results indicated that the *D. virginiana* ϵ-globin gene evolved faster although the difference in substitution rate was not statistically significant (Table 4.7). These results show that the difference in the rate of evolution of S.c-β and *D. virginiana* ϵ-globin gene has little effect on the conclusions reached in relative rate tests of eutherian ϵ-globin genes.

The relative rate test of the mouse and human adult β-globin genes, with the *D. virginiana* β^m-globin gene as an outgroup, shows that the mouse gene evolved faster at non-synonymous sites and slower at synonymous sites. These differences in substitution rates, however, were not statistically significant (Table 4.7).

TABLE 4.7. Comparisons of the number of nucleotide substitutions per site between eutherian ϵ-globin genes, using the relative rate test[1].

Species / Gene			Non-synonymous substitutions				Synonymous substitutions			
			Non degenerate		2-fold degenerate		2-fold degenerate		4-fold degenerate	
1	2	3	K_{12}	$K_{13} - K_{23}$	B_{12}	$B_{13} - B_{23}$	A_{12}	$A_{13} - A_{23}$	K_{12}	$K_{13} - K_{23}$
human ϵ	human γ	S.c-β	0.098	-0.021 (0.021)	0.067	-0.043 (0.031)	0.367	-0.022 (0.011)	0.353	-0.185 (0.138)
human ϵ	mouse ϵ	S.c-β	0.091	-0.026 (0.020)	0.056	-0.031 (0.028)	0.316	-0.034 (0.120)	0.421	-0.208 (0.147)
human ϵ	goat ϵ	S.c-β	0.047	0.025 (0.014)	0.043	-0.029 (0.024)	0.340	-0.069 (0.108)	0.246	0.154 (0.098)
human ϵ	rabbit ϵ	S.c-β	0.082	-0.001 (0.019)	0.022	-0.027 (0.023)	0.336	-0.111 (0.113)	0.389	0.041 (0.121)
mouse ϵ	goat ϵ	S.c-β	0.077	0.052* (0.018)	0.045	0.022 (0.026)	0.444	-0.036 (0.121)	0.544	0.363* (0.156)
mouse ϵ	rabbit ϵ	S.c-β	0.124	0.026 (0.023)	0.069	0.004 (0.031)	0.240	-0.077 (0.103)	0.613	0.249 (0.165)
rabbit ϵ	goat ϵ	S.c-β	0.088	0.026 (0.019)	0.063	-0.002 (0.031)	0.359	0.042 (0.125)	0.265	0.114 (0.096)
human β	mouse β	*D.vir* β	0.140	-0.020 (0.027)	0.041	-0.038 (0.028)	0.264	0.124 (0.121)	0.386	0.121 (0.191)
S.c-β	*D.vir* ϵ	mouse ϵ	0.111	-0.013 (0.022)	0.068	0.001 (0.031)	0.165	-0.050 (0.087)	0.274	-0.056 (0.149)
human ϵ	mouse ϵ	*D.vir* ϵ	0.091	-0.102 (0.020)	0.056	-0.005 (0.029)	0.316	-0.120 (0.106)	0.421	-0.209 (0.165)

1. A_{ij}, B_{ij} and K_{ij} are respectively the number of transitional substitutions, the number of transversional substitutions and the total number of substitutions between species i and j. The numbers in parentheses are the standard errors. The sources of each gene are given in the appendix. * - Standard normal test is significant at the 5% probability level.

4.8 Evolutionary studies: phylogenetic relationships among mammals.

In this section, the evolutionary relationships of monotremes, marsupials and eutherians are investigated using phylogenetic trees constructed for adult α- and β-globin amino acid sequence data. Included in this analysis are α- and β-globin sequences from a representative of the marsupial Order Dasyuromorphia.

Phylogenetic trees were constructed using the "Protpars" computer program contained in the "Phylip" package (version 3.2, obtained from Joseph Felsenstein, Washington University). This program, the algorithm of which is based on the approaches of Eck and Dayhoff (1966) and Fitch (1971), infers the most parsimonious unrooted trees directly from protein seqences. They are therefore consensus minimal trees derived from a complete search of all possible trees (for full details of the program, see Section 3.2.9). Approximate estimates of branch lengths were obtained using a method suggested by Felsenstein (pers. comm.). This involved counting the minimum number of amino acid substitutions along each branch including those substitutions which were unknown (designated as a "?" in the Protpars output, see Section 3.2.9). The total number of substitutions (N) from all branches of the tree was calculated and individual branch lengths were subsequently corrected by dividing by N and multiplying by the original length of the tree (as given in the Protpars output). It should be noted that this approach gives only a crude estimate of branch lengths and therefore no attempt was made to test their statistical significance.

Two sets of data were used for these phylogenetic analyses. The first set

comprised the adult β-globin amino acid sequences from 5 marsupial, 4 eutherian and 2 monotreme taxa (see Appendix). The *S.crassicaudata* partial β-globin amino acid sequence (Section 4.4) was combined with the conceptually translated amino acid sequence of the partial cDNA β-globin clone pDG-5 (Wainwright, 1984), to give a full length polypeptide (146 amino acids) which was assumed to be representative of β-globin sequences of the Order Dasyuromorphia. Data from frog and chicken were used as outgroups for the analysis. The second set of data comprised the adult α-globin amino acid sequences from 3 marsupial, 5 eutherian and 2 monotreme taxa (see Appendix). Data from the frog, alligator and chicken were used as outgroups.

Using β-globin data, a single consensus tree was obtained, which required a total of 388 nucleotide substitutions. The tree was rooted using the frog β-globin sequence as an outgroup (Fig. 4.8.1). This tree suggests that the marsupial β-globin sequences are ancestral to both the monotreme and eutherian sequences (Fig. 4.8.1). Such a branching order is in apparent contradistinction to that generally accepted for the evolution of mammalian groups, where monotremes are regarded as diverging from an ancestral mammalian stock *before* eutherians and marsupials (see Section 2.12).

The branching order obtained for the marsupial species (Fig. 4.8.1) accords with the relationships of these species inferred from morphological characters (see Kirsch and Calaby, 1977), immunological data (Kirsch, 1977, Maxon *et al.*, 1975; Baverstock *et al.*, 1990) and DNA - DNA hybridization studies (Westerman *et al.*, 1990). Thus the data show that dasyurid and macropodid sequences are more closely

related to each other than either are to didelphids, ie dasyurids and macropodids are monophyletic to the exclusion of didelphids. They also show that the two kangaroo species are monophyletic to the exclusion of potoroids, again in agreement with currently accepted taxonomic relationships (Aplin and Archer, 1987).

For the α-globin data, two consensus minimal trees were obtained each of which required a total of 337 nucleotide substitutions. However, since the two trees differed only by the position of the branch leading to the horse α-globin sequence only one tree is shown (Fig. 4.8.2). This tree was rooted using the frog α-globin sequence as an outgroup. The results are in marked contrast to those obtained with the β-globin data. The monotreme α-globin sequences appear to be ancestral to both the eutherian and marsupial sequences (Fig. 4.8.2). Such a pattern of divergence is in agreement with currently accepted mammalian relationships (Aplin and Archer, 1987). For the three marsupial species used, the α-globin sequence divergence pattern is in accord with that shown for the β-globin sequences above, ie dasyurid and macropodid α-globin sequences are monophyletic to the exclusion of didelphids. Such branching patterns support the existence of two cohorts of marsupials, Cohort Ameridelphia here represented by *D. virginiana* and Cohort Australidelphia here represented by dasyurid and macropodid species.

Fig. 4.8.1

Most parsimonious phylogenetic tree (length: 388 nucleotide substitutions) constructed from marsupial, eutherian and monotreme adult β-globin amino acid sequences. The tree was constructed using the "Protpars" program of Phylip (version 3.2) (J. Felsenstein, Washington University) and was rooted using the frog β-globin sequence as an "outgroup". All sequences were obtained from the NBRF* data base with the exception of the Das (Dasyuromorphia) β-globin sequence (see text for details). Abbreviation of species, common name, and accession number from the data base are:

X.lae:- (*Xenopus laevis*), frog, A02452.
G.gal:- (*Gallus gallus*), chicken, A02432.
M.mus:- (*Mus musculus*), mouse, A02404.
E.cab:- (*Equus caballus*), horse, A02380.
H.sap:- (*Homo sapian*), human, A02352.
B.tau:- (*Bos taurus*), bovine, A02387.
O.ana:- (*Ornithorhyncus anatinus*), platypus, A02431.
T.acu:- (*Tachyglossus aculeatus*), echidna, A02430.
D.vir:- (*Didelphis virginiana*), opossum, A02429.
P.tri:- (*Potorous tridactylus*), potoroo, A02428.
M.ruf:- (*Macropus rufus*), red kangaroo, A02427.
M.gig:- (*Macropus giganteus*), grey kangaroo, A02427.

* George, D.G., Barker, W.C. and Hunt, L.T. (1986) The Protein Identification Resource (PIR). *Nucl. Acid Res.* 14: 11-16.

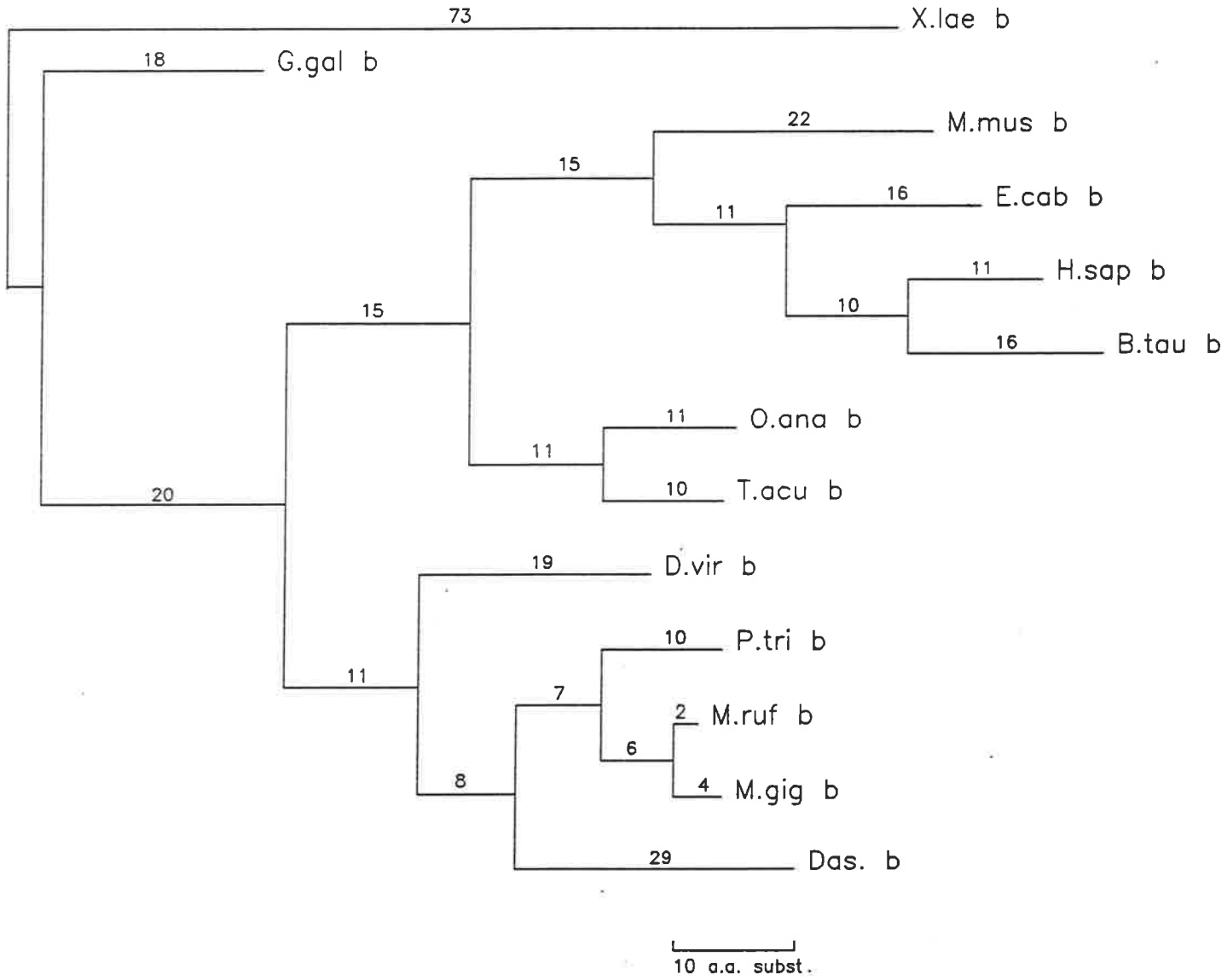
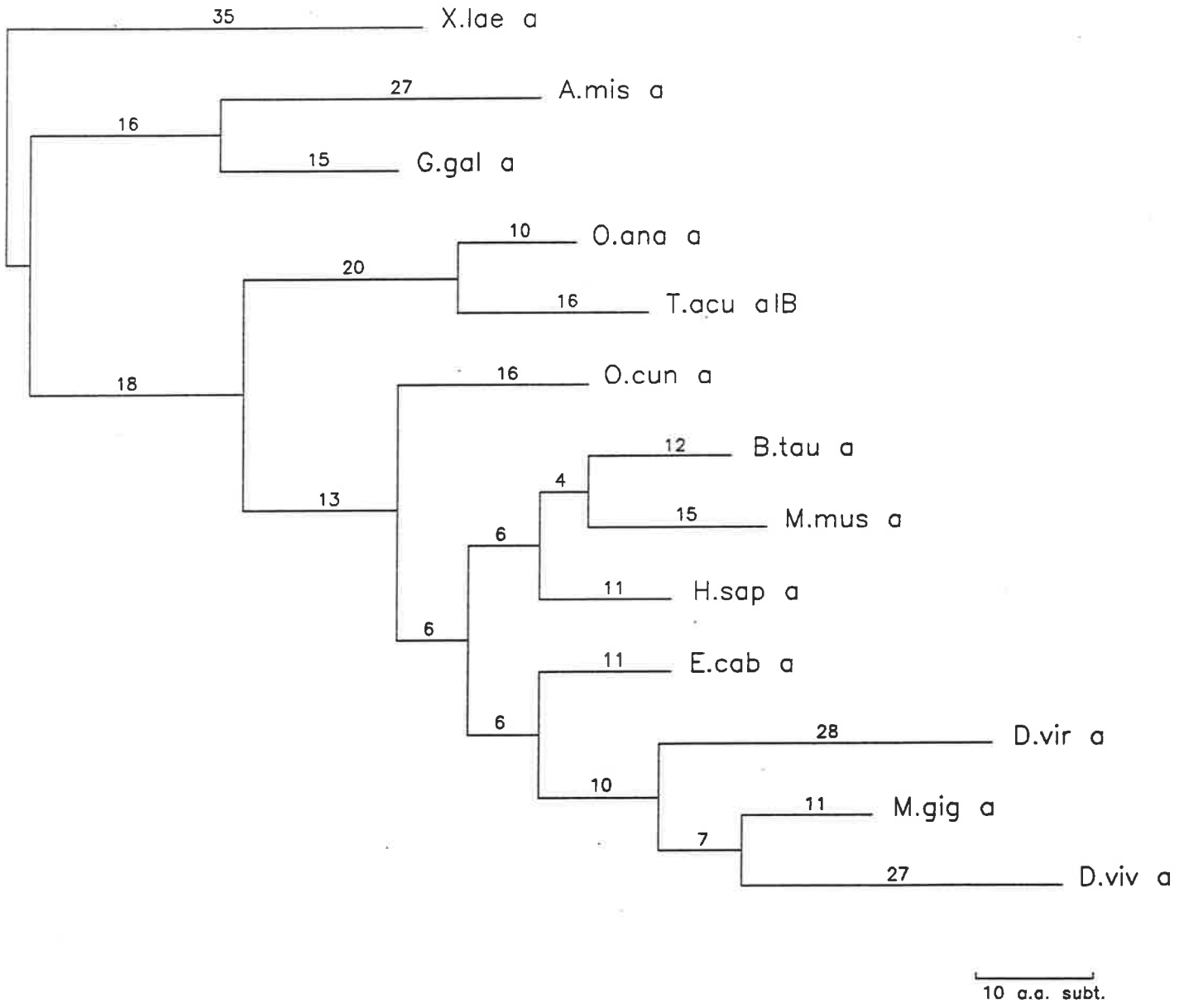

73
X.lae b
18
G.gal b
22
M.mus b
15
16
E.cab b
11
11
H.sap b
15
10
16
B.tau b
11
O.ana b
11
10
T.acu b
20
19
D.vir b
10
P.tri b
11
7
2
M.ruf b
6
8
4
M.gig b
29
Das. b
10 a.a. subst.

Fig. 4.8.2

Phylogenetic tree (length: 337 nucleotide substitutions) constructed from marsupial, eutherian and monotreme adult α-globin amino acid sequences. The tree was constructed using the "Protpars" computer program of Phylip (version 3.2) (J. Felsenstein, Washington University) and was one of two equally most parsimonious trees. The tree was rooted using the frog α-globin sequence as an "outgroup". All sequences were obtained from the NBRF* data base with the exception of the D.viv (*Dasyurus viverrinus*) α-globin sequence (obtained from Wainwright, 1984; Wainwright and Hope, 1985) and the T.acu (*Tachyglossus aculeatus*, echidna) αIB-globin sequence (obtained from Whittaker *et al.*, 1973). Abbreviation of species, common name, and accession number from the data base are:

X.lae:- (*Xenopus laevis*), frog, A02341.
A.mis:- (*Alligator mississippiensis*), alligator, A02320.
G.gal:- (*Gallus gallus*), chicken, A02315.
O.ana:- (*Ornithorhyncus anatinus*), platypus, A02302.
O.cun:- (*Oryctolagus cuniculus*), rabbit, A02270.
B.tau:- (*Bos taurus*), bovine, A02289.
M.mus:- (*Mus musculus*), mouse, A02264.
H.sap:- (*Homo sapian*), human, A02248.
E.cab:- (*Equus caballus*), horse, A02281.
D.vir:- (*Didelphis virginiana*), opossum, A02299.
M.gig:- (*Macropus giganteus*), grey kangaroo, A02298.

* George, D.G., Barker, W.C. and Hunt, L.T. (1986) The Protein Identification Resource (PIR). *Nucl. Acid Res.* 14: 11-16.


35
X.lae a
27
A.mis a
16
15
G.gal a
10
O.ana a
20
16
T.acu aIB
18
16
O.cun a
12
B.tau a
4
13
15
M.mus a
6
11
H.sap a
6
11
E.cab a
6
28
D.vir a
10
11
M.gig a
7
27
D.viv a
10 a.a. subt.

CHAPTER 5

DISCUSSION

5.1 Evidence that S.c-β is a functional β-like globin gene.

A β-like globin gene (S.c-β) was isolated from a bacteriophage genomic DNA library using a *D.viverrinus* cDNA β-globin clone (pDG-5.2) as a probe. The following features of the DNA sequence of S.c-β provide evidence that it is functional:-

i) S.c-β has conserved overall structure consisting of three exons and two introns. The locations at which the introns occur in the gene are identical to the locations of introns in all known vertebrate globin genes.

ii) S.c-β has three conserved promoter elements, a "CACCC" box, "CAAT" box and "ATA" box located 5' to the CAP site. These promoter elements are found in all known mammalian β-like globin genes.

iii) S.c-β has conserved donor and acceptor splice signals at intron / exon junctions.

iv) The exons are open reading frames.

v) S.c-β has a conserved poly-adenylation signal "AATAAA" located in the 3' flanking region of the gene.

vi) Non-synonymous divergence values indicate that S.c-β has a highly conserved coding sequence, which suggests that the gene is under strong selective constraints.

There is no evidence from these data that S.c-β is evolving at a rate greater than or equal to the synonymous substitution rate in functional genes, ie S.c-β is not a pseudogene.

vii) mRNA isolated from an embryo and pouch young (less than 4 days old, *post partum*) hybridizes strongly to the 2nd exon of S.c-β at a single band on a Northern filter, providing evidence that S.c-β is transcribed.

The only feature of the gene which appears anomalous is the size of the 2nd intron, which was significantly larger than the 2nd introns of all known eutherian β-like globin genes. However, such a feature is unlikely to affect the function of the gene as the intron is similar in size to the 2nd intron of a *D. virginiana* adult β-globin gene (1465 bp), which is known to be functional as it conceptually translates into the *D. virginiana* adult β-chain (Koop and Goodman, 1988). The presence of large 2nd introns in β-like globin genes of *D. virginiana* and *S. crassicaudata* provides evidence that an approximately 500 bp DNA fragment was inserted into these genes before the separation of the didelphid / non-didelphid lineages. Alternatively, large 2nd introns were present in β-like globin genes before the divergence of eutherians and marsupials from a common ancestor, and one or more deletion(s) occurred in this intron during the evolution of eutherians. DNA sequence comparisons involving S.c-β, *D. virginiana* and eutherian β-like globin genes provide evidence that a dA/dT rich sequence present in the 2nd intron of S.c-β and *D. virginiana* β-like globin genes but absent in eutherian β-like globin genes may be contained within the inserted / deleted region.

The above proposals are currently being investigated using the polymerase chain reaction to amplify the entire 2nd intron of β-like globin genes from a wide range of marsupials and two monotremes. Preliminary results suggest that large 2nd introns (> 1.3 kb) are a common feature of most marsupial β-like globin genes although some exceptions were found, most notably *Sminthopsis macroura* and *Sminthopsis virginiae* (2nd intron approximately 900 bp in length) and *Dasyurus viverrinus* (2nd intron approximately 700 bp in length). The 2nd intron of a β-like globin gene from an echidna (*Tachyglossus aculeatus*) was found to be approximately 900 bp in length and thus similar in size to 2nd introns of eutherian β-like globin genes. These results, together with the observation that chicken β-like globin genes have 2nd introns of an approximate length of 800 bp (Dolan *et al.*, 1983) suggest that it is more likely that 2nd introns of β-like globin genes in marsupials increased in size by one or more insertion event(s) which occurred prior to the separation of the didelphid / non-didelphid lineages. However, it should be stressed that the mechanisms influencing such changes in the size of the 2nd intron are unknown. It is possible that the variation of intron sizes resulted from a gradual "accumulation" or "diminution" of DNA and not by single deletion / insertion event(s).

5.2 Evidence that S.c-β is orthologous to eutherian and *D.virginiana* ϵ-globin genes.

DNA and protein sequence comparisons provide evidence that S.c-β is orthologous to eutherian and *D.virginiana* ϵ-globin genes. This evidence comes from the following observations:-

i) The 5' promoter region of S.c-β contains a number of presumptive synapomorphs which are common to eutherian embryonic β-like globin genes.

ii) The conceptually translated sequence of S.c-β was found to be more similar to eutherian and *D.virginiana* ϵ-chains than adult β-chains, at both functional sites and sites believed to be important for embryonic haemoglobin function. In particular, the conceptually translated sequence of S.c-β contains phenylalanine at position 3 of the polypeptide chain which is common to all known eutherian ϵ-chains. Koop and Goodman (1988) proposed that phenylalanine reduces the binding of DPG and increases the oxygen affinity of embryonic haemoglobin over that of the maternal haemoglobin, thereby favouring oxygen transfer from mother to embryo across the placenta.

iii) Comparisons of non-synonymous and synonymous divergence values indicate that S.c-β is more similar, at synonymous and non-synonymous sites, to eutherian and *D.virginiana* ϵ-globin genes than it is to a *D.viverrinus* adult β-globin mRNA and eutherian and *D.virginiana* adult β-globin genes. These comparisons also provided evidence that S.c-β is highly conserved. According to Koop and Goodman (1988)

such a feature is a characteristic of embryonic β-like globin genes. They proposed that "stabilizing selection acted more pervasively on embryonic ε than on adult β genes" and further that "less room may exist for variation of internal conditions in the embryonic stage of life than in later stages".

Taken overall, DNA and protein sequence comparisons provide strong evidence that S.c-β is orthologous to *D.virginiana* and eutherian ε-globin genes.

5.3 Evidence that S.c-β is expressed during embryonic development.

Sequence comparisons, however, do not prove that S.c-β is expressed in embryonic tissues. An expression study of a marsupial embryonic β-like globin gene has not previously been reported, nor has an embryonic β-globin chain been isolated and sequenced from a marsupial. It is therefore possible that fortuitous homoplasy between S.c-β and eutherian ε-globin genes was responsible for their sequence similarities and S.c-β is expressed in adults. The expression of S.c-β was therefore investigated by Northern analysis of total mRNA isolated from an embryo, pouch young between 1 and 10 days of age (*post partum*) and adult bone marrow (Section 4.3). The results provide evidence that S.c-β is expressed in embryonic tissues and further that the gene is switched off or expressed at low levels before the 1st or 2nd days (*post partum*) of pouch young development.

Evidence that S.c-β is not expressed at a high level in adults was obtained from a comparison of the conceptually translated sequence of S.c-β with a partial amino

acid sequence of the adult β-chain of *S.crassicaudata* (Section 4.4). The two chains differ by 20 amino acid residues out of 60. Differences of such magnitude are unlikely to result from polymorphic variation at a single β-like globin locus.

Results presented in Section 4.6 provided evidence that just two functional β-like globin genes exist in *S.crassicaudata* (S.c-β and a putative adult β-globin gene). It therefore appears likely that a single switch in β-globin gene expression occurs in *S.crassicaudata*, from S.c-β to an adult β-globin gene, and that this switch occurs within a day or two of birth. Interestingly, in humans, a switch in β-globin expression, from γ to β, also occurs around the time of birth, although in this case it occurs 36 weeks after conception. Recent studies suggest that the switch from γ to β in humans is controlled by a "locus activating region" (LAR) located approximately 11 kb 5' to the ε-globin gene (Enver *et al.*, 1990). In addition, LAR appears to exert a dominant control over the expression of all the β-globin genes in the human cluster, maintaining a high level of β-globin gene expression specifically in cells of erythroid origin. It would be of interest to investigate whether an LAR is also present in *S.crassicaudata* and whether it is involved in the switch in expression from ε to β in this species.

In order to verify the proposal that a switch in expression from S.c-β to an adult β-globin gene occurs within a day or two of birth, a Northern analysis of embryo, pouch young and adult bone marrow RNA should be repeated using a region of the *S.crassicaudata* adult β-globin gene as a probe. Such an analysis would enable a direct comparison to be made between the levels of expression of each gene at

different developmental stages. It would also enable a more precise time to be determined for the switch in β-globin expression.

The results of the Northern analysis provide evidence that embryonic β-globin mRNA is present at least until day 4 (*post partum*) of pouch young development in *S.crassicaudata*. Thus it is likely that embryonic forms of haemoglobin would also be present during early pouch young life, which further suggests that non-adult forms of haemoglobin detected in *M.eugenii* (Holland *et al.*, 1988), *M.robustus, M.giganteus* and *M.rufus* (Richardson and Russell, 1969) and *T.vulpecula* (Hope, 1970) pouch young were embryonic, ie composed of an embryonic β-chain. The presence in early pouch young (< 10 days old *post partum*) of embryonic haemoglobins raises questions concerning the physiological function of such globins during this developmental period, and the time during which they remain in circulation.

Respiratory gas levels in the marsupial pouch have been studied in *Didelphis virginiana* (de Almeida and Rocha, 1932), *Trichosurus vulpecula* (Bailey and Dunnet, 1960), and *Macropus eugenii* (Baudinette *et al.*, 1988; Holland *et al.*, 1988) and have been found to be hypoxic and hypercapnic compared to ambient conditions. In the environment of the pouch, one might predict that a non-adult form of haemoglobin may play an important physiological role in supplying sufficient oxygen to the tissues of the rapidly developing pouch young. Pouch respiratory gas levels have not been studied in *S.crassicaudata* nor any other dasyurid marsupial to date.

Pouch morphology, however, has been well studied in dasyurids and provides clues to the O_2 / CO_2 enviroment which may be experienced by the pouch young.

Some dasyurids, such as *Antechinus stuartii*, have pouches which consist of two lateral folds of skin which do not enclose the mammary area (Woolley, 1974). In these species it would be expected that O_2 and CO_2 levels would be close to ambient air. However, in *S. crassicaudata* (Smith and Godfrey, 1970) the mammary area is covered by a circular fold of skin, suggesting that during early pouch young development in this species, the environment may also be hypoxic and hypercapnic as in *M. eugenii, T. vulpecula* and *D. virginiana*. This environment would be expected to persist until the *S. crassicaudata* pouch young were of a sufficient size to force the pouch to open (within 20 days of pouch young life, Dr P. Woolley pers. comm.). There is reason to believe, therefore, that in *S. crassicaudata*, an embryonic form of haemoglobin may have an important physiological function during early pouch young life.

The question of how long embryonic forms of haemoglobin remain in circulation in *S. crassicaudata* also has not been investigated. They are likely to remain in existence until the erythrocyte containing the embryonic haemoglobin is removed from circulation. In macropodids there is a rapid turnover of erythrocytes during the first 7 to 15 days *post partum* of pouch young development (Richardson and Russell, 1969; Holland *et al.*, 1988). During this time, Holland *et al.* (1988) observed that the four embryonic haemoglobins of *M. eugenii* were replaced by adult haemoglobin. In *S. crassicaudata* there also appears to be a marked reduction in the level of embryonic β-globin RNA during the first 7 days *post partum* of pouch young development. It seems likely that a rapid turnover of erythrocytes during early pouch young development also occurs in *S. crassicaudata* and that just prior to this change

in red cell morphology a switch from ϵ to β expression occurs. Thus it might be expected in *S. crassicaudata* that embryonic haemoglobins would also be removed from circulation within the first 10 days *post partum* of pouch young development. To verify this hypothesis, blood smears should be examined in pouch young between 1 and 10 days of age, to test whether there is a change in red blood cell morphology in *S. crassicaudata* during this developmental period. In addition, it would be of interest to study the haemoglobins of this developmental period using isoelectric focussing.

5.4 The number and arrangement of β-like globin genes in *S. crassicaudata*.

The results presented in Section 4.6 provided evidence that at least 3 β-like globin sequences exist in *S. crassicaudata* including one functional embryonic ϵ-globin gene (S.c-β) and one functional adult β-globin gene. A linkage analysis of restriction fragment length variants at each locus provided evidence that the ϵ- and β-globin genes of *S. crassicaudata* are closely linked to each other but unlinked to the third locus (ϵ_2). ϵ_2 hybridized to S.c-β, but failed to hybridize to a *D. viverrinus* cDNA β-globin clone (pDG-5.2) and is therefore likely to be a pseudogene (see discussion later). A Southern analysis of two PstI restriction fragment variants in the family M1920 indicated that the ϵ-globin and putative adult β-globin genes of *S. crassicaudata* were located together on a 30 kb PstI fragment. A PstI site is located approximately 2 kb 5' to the CAP site of S.c-β in λSG-3 but is absent at an analagous position in λSG-5. This result suggests that the PstI fragment length variation is caused by the presence

or absence of the above site, and further indicates that the adult β-globin gene is located approximately 28 kb 3' to the CAP site, or 26 kb 3' to the end of the 3rd exon of S.c-β. A phage clone, λSG-1, containing 16 kb of 3' flanking sequence of S.c-β is therefore within 10 kb of the adult β-globin gene. A Southern analysis of the phage clones λSG-5, λSG-3 and λSG-1 indicated that no further β-like globin sequences were present either 5' or immediately 3' to S.c-β. Taken overall, the results suggest that the β-globin gene cluster of *S.crassicaudata* spans approximately 30 kb and consists of just two functional genes, one expressed in embryonic tissues (S.c-β) and the other expressed during pouch young and adult life.

The orientation of the two β-like globin genes is: 5'-ϵ-β-3' and is analagous to the orientation of the embryonic and adult sets of β-like globin genes in all eutherian β-globin gene clusters studied to date (see Goodman *et al.*, 1984). It is likely therefore that the ϵ^m-globin and β^m-globin genes of *D.virginiana* are closely linked in a similar orientation. This linkage arrangement and the presence of a single adult β-globin gene and a single embryonic ϵ-globin gene in *S.crassicaudata*, provide support for the hypothesis of Koop and Goodman (1988), that just two progenitors of the five ancestral eutherian β-globin loci (ϵ, γ, η, δ and β) existed at the time the marsupial and eutherian lineages separated. This hypothesis, however, was not supported in comparisons of non-synonymous substitution rates involving S.c-β and human ϵ- and γ-globin genes (Section 4.2). These results were in contrast to those reported by Koop and Goodman (1988), which showed that at non-synonymous sites, the human ϵ- and γ-globin genes are more distantly related to the *D.virginiana* ϵ-globin gene

(13% divergence) than either are to each other (10% divergence).

It is possible that the ϵ and γ gene lineages may have diverged after the separation of the marsupial and eutherian lineages and the human and *S.crassicaudata* ϵ-globin genes appear more similar to each other than either are to the human γ-globin gene due to a slower rate of evolution of the ϵ-genes relative to γ-genes. Relative rate tests involving the human ϵ- and γ-globin genes with S.c-β as an outgroup (Section 4.6) indicate that the human ϵ-globin gene has evolved at a slower rate than the human γ-globin gene, although the differences in substitution rate were not statistically significant. However, these results suggest that comparisons of non-syonymous rates of substitution between γ- and ϵ-globin genes would not lead to a definitive conclusion as to whether marsupials contain descendants of an ancestral γ-globin gene. Therefore, the former conclusion is the most likely, ie that just two progenitors (proto-ϵ and proto-β) of eutherian and marsupial β-like globin genes existed at the time the eutherian and marsupial lineages separated. The results of the present study also lend support to the theory that sometime after the first tandem duplication of an ancestral β-globin gene, approximately 155 to 200 mya, the gene on the 5' side of the two gene cluster became expressed in embryonic tissues only, and the 3' gene became developmentally delayed in its expression.

Despite containing the fewest number of genes of any β-globin gene cluster studied to date (excluding *D.virginiana*) the β-globin gene cluster of *S.crassicaudata* is not the smallest known β-globin cluster. The β-globin gene cluster of the brown lemur (*Lemur macaco (fulvus) mayottensis*) is just 20 kb long and consists of three

functional β-globin genes and one pseudogene (Fig. 2.1). The *S.crassicaudata* β-globin cluster is similar in size to the rabbit cluster (approximately 30 kb) and the four gene ancestral cluster of β-like globin genes which is triplicated in the goat and duplicated in cattle and sheep (Section 2.6). This is despite at least two extra tandem gene duplications in the rabbit and ancestral artiodactyl clusters. It seems probable that there is strong selection to maintain a minimum distance between the ϵ-globin and β-globin genes. For example, competition for a common enhancer element located 3' to the adult β-globin gene and 5' to the ϵ-globin gene in chickens is thought to be responsible for the switch from embryonic to adult β-globin expression. The movement of one of these genes away from or closer to the enhancer element may therefore have a significant effect on the differential expression of these genes. Such a competitive interaction between control elements is also thought to be responsible for the switch from γ-globin to β-globin gene expression in humans (Enver *et al.*, 1990). It is not known how this control operates, but it is possible that the distance and order of genes in relation to these control elements may influence the developmental time in which they are expressed. The β-globin gene cluster of *S.crassicaudata*, consisting of just two functional genes with a single switch in gene expression, provides an excellent model for studying the control of β-globin gene expression.

The Southern analysis of individual and family DNA samples provide evidence that the third β-globin locus (ϵ_2) is a pseudogene and is unlinked to the functional β-globin cluster of *S.crassicaudata*. A third β-globin locus was not detected in

D. virginiana (Koop and Goodman, 1988) and therefore it is likely that ϵ_2 arose after the divergence of the didelphid / non-didelphid lineages. It is difficult to determine from the results of the Southern analysis, whether ϵ_2 contains a full length copy of a β-like globin gene, as it failed to hybridize to the *D. viverrinus* cDNA β-globin probe pDG-5.2. This probe contains a cDNA copy of the 3rd exon and a portion of the 2nd exon of an adult β-globin RNA while the S.c-β probe, which hybridized to ϵ_2, contained the 2nd exon, a portion of the 2nd intron and the entire 1st intron of S.c-β. Thus it is possible that either ϵ_2 is missing regions of the 2nd exon and the 3rd exon of a β-like globin gene, or it has diverged in sequence to such an extent that it will no longer hybridize to pDG-5.2 under the conditions used in this experiment. The present study can not exclude the possibility that ϵ_2 contains an independently derived region of DNA which is common to S.c-β.

It appears more likely that ϵ_2 is a pseudogene and was derived by duplication and silencing of an ancestral ϵ-globin gene. Such a process can be achieved by at least two possible mechanisms:- i) tandem gene duplication of an ϵ-globin gene followed by transposition of ϵ_2 away from the β-globin cluster. ii) reverse transcription of an ϵ-globin RNA followed by insertion of the cDNA copy back into the genome resulting in a processed pseudogene. The isolation of a processed PGK pseudogene from the wallaroo (*Macropus robustus*) indicates this latter mechanism has operated at some time during marsupial evolution (van Daal *et al.*, 1989). Both mechanisms have operated within the α-globin gene family during rodent evolution (Section 2.8) but to date, no example of a β-globin pseudogene which is unlinked to the functional β-

globin gene cluster has been reported. It would be of much interest to study the ϵ_2-globin locus at the molecular level. However, attempts to isolate this "gene" from the *S.crassicaudata* genomic DNA library proved unsuccessful.

5.5 Relative rate tests of eutherian ϵ-globin genes.

Relative rate tests carried out on eutherian ϵ-globin genes (Section 4.7) provide evidence for higher synonymous and non-synonymous nucleotide substitution rates in rodents than in lagomorphs, artiodactyls and primates. These tests were based on a reliable branching order of outgroup and ingroup taxa, in contrast to the relative rate tests reported by Wu and Li (1985) and Easteal (1988).

It is possible that the differences in substitution rates of eutherian ϵ-globin, reported in Section 4.7, were due to chance. Only two tests showed statistically significant rate differences (comparison of goat and mouse ϵ-globin genes at non-degenerate and 4-fold degenerate sites). Wu and Li (1985) claim that the standardized normal test "is probably somewhat too stringent as it is based on the assumption that nucleotide substitution follows a Poisson process". With this in mind and taking into account the observation that in 10 out of 12 relative rate tests that included the mouse ϵ-globin gene, the mouse gene evolved fastest, the results lend support to the conclusions of Wu and Li (1985) of "higher rates of nucleotide substitution in rodents". This conclusion is not supported in relative rate tests involving the human and mouse adult β-globin genes with the *D.virginiana* β-globin gene as an outgroup.

In these tests it was found that the synonymous substitution rate was faster in primates than in rodents. In contrast, relative rate tests reported by Wu and Li (1985) involving the human and mouse β-globin genes reported faster rates of synonymous substitutions in rodents than in primates ($K_{13} - K_{23} = -0.15$, where species 1, 2 and 3 are human, mouse, and goat respectively). The difference between the $K_{13} - K_{23}$ value reported in this study and that reported by Wu and Li (1985) could be due to chance or to an incorrect assumption by Wu and Li that the goat β-globin gene was an outgroup of the human and mouse β-globin genes. It would be of interest to repeat the study by Wu and Li (1985) using marsupial genes as outgroups in relative rate tests of orthologous rodent or primate genes.

The idea that nucleotide substitution rates are influenced by generation times and that there has been a slow-down in the rate of molecular evolution in the hominoid lineage can not be entirely supported by this study. Relative rate tests involving the goat, human and rabbit ϵ-globin genes indicate that nucleotide substitution rates in the primate and lagomorph lineages are similar at both synonymous and non-synonymous sites and slightly faster than those in goats. Generation time can be loosely defined as the period required for the population to turn over, and is weighted for age specific contributions (Sibley and Alquist, 1987). Rabbit females (does) begin breeding when approximately one year old and although they can live over five years, the average expectation of life is approximately one and a half years (Lockley, 1964). Rabbits therefore have a faster population turn over and consequently a shorter generation time than humans and according to Wu and Li

(1985) should have faster rates of sequence evolution. This is not apparent in relative rate tests of ϵ-globin genes reported in this study. It is possible, however, that the generation time of rabbits today may be significantly faster than the generation time of their ancestors or alternatively the generation time of humans may be slower than that in ancestral primates. Taking these possibilities into account, the relative rate tests reported in this and other studies can not provide definitive evidence that nucleotide substitution rates are influenced by generation time.

If a slow-down in the rate of molecular evolution of the ϵ-globin gene occurred in the hominoid lineage, then a corresponding slow-down must have occurred in the the lagomorph and artiodactyl lineages. It seems more likely that the rates of molecular evolution of the ϵ-globin gene became faster in the rodent lineage and have remained relatively stable in lagomorphs and primates. An exception to the rate slow-down in the hominoid lineage was also reported by Shaw *et al.* (1989), in a study of α-globin genes. Such exceptions suggest that generation time is not the only mechanism which can influence the rate of molecular evolution. Britten (1986) argues that "changes in repair mechanisms are a likely source of the differences in mutation rates" and also suggests that transposable elements and changes in the biochemistry of DNA replication may also play a role. Further analyses of DNA sequence data from a wider range of taxonomic groups may help to distinguish the underlying mechanisms responsible for mutation rate differences.

5.6 Inter-relationships of marsupials, monotremes and eutherians; inferred from phylogenies of α-globin and β-globin amino acid sequences.

A phylogeny derived from β-globin amino acid sequences suggests that the marsupial sequences are ancestral to the eutherian and monotreme sequences. In contrast, phylogenies derived from α-globin sequences depict monotreme sequences as being ancestral to marsupial and eutherian sequences. These results have been verified independently using the phylogenetic tree construction program "Penny's Min. tree", which also uses an algorithm based on the maximum parsimony procedure (P.G. Martin, pers. comm.), and are also concordant with "geneological reconstructions" reported by Goodman *et al.* (1985, 1987). The inclusion in the present study, of representative α- and β-globin sequences of the Order Dasyuromorphia, did not alter previous conclusions that α-globin sequences "depict the traditional arrangement of Prototheria (monotremes) as the sister group of Theria" and β-globin sequences "depict monotremes as an ancient branch of Eutheria" (Goodman *et al.*, 1987). These results, however, lend support to the proposal that the three mammalian infra-classes diverged at much the same time (see Clemens, 1989; Czelusniak *et al.*, 1990; Westerman and Edwards, 1991)

Some liberties were taken in combining the partial adult β-globin amino acid sequences of *S.crassicaudata* and *D.viverrinus* to form a "representative" Dasyuromorphia adult β-globin sequence. However, the branching order within marsupials inferred from this sequence was concordant with those inferred from

morphological characters (Kirsch and Calaby, 1977), immunological data (Maxon *et al.*, 1975; Kirsch, 1977; Baverstock *et al.*, 1990), and DNA - DNA hybridization studies (Westerman *et al.*, 1990) and was also concordant with the branching order within marsupials inferred from α-globin data. Thus, the α- and β-globin phylogenies support the existence of two cohorts of marsupials, Cohort Ameridelphia and Cohort Australidelphia (Aplin and Archer, 1987).

In order to resolve the evolutionary relationships of the three mammalian infra-classes much more sequence data are required, particularly from monotremes and marsupials. It might be sensible in this analysis to adopt a more coordinated approach and include for marsupial sequence data one representative of each of the Orders Dasyuromorphia and Diprotodontia of the Cohort Australidelphia and one representative of the Order Didelphimorphia of the Cohort Ameridelphia. This approach may help to overcome errors in phylogenies due to non-uniformity of rates of molecular evolution in different lineages.

5.7 Concluding remarks.

The main aim of this Ph.D project was to characterize at the molecular level the β-globin gene family of *S.crassicaudata*. This aim has been achieved, although much of the work is still preliminary in nature. The results provide evidence that at least three β-like globin genes exist in *S.crassicaudata* including one gene expressed during embryonic development (S.c-β), and a single adult β-globin gene. There is no

evidence from this study that *S.crassicaudata* contains β-like globin genes specifically adapted to the oxygen transport requirements of pouch young. However, the physiological role of the embryonic ε-chain during this developmental period remains to be determined. Southern analysis of *S.crassicaudata* genomic DNA, provides evidence that S.c-β (ε) and the adult β-globin gene (β) are closely linked in the order: 5'-ε-β-3' over a region of DNA spanning approximately 30 kb. A third β-like globin sequence (ϵ_2), detected by Southern analysis, appears to be unlinked to the functional gene cluster and is likely to be a pseudogene.

In order to complete the characterization of the β-globin gene family of *S.crassicaudata* it will be necessary to isolate the adult β-globin gene and the putative pseudogene ϵ_2. At least two procedures could be used to isolate and characterize the adult β-globin gene:

1. amplify regions of the gene, from genomic DNA or mRNA, using the "polymerase chain reaction". This approach would enable the DNA sequence and structure of the gene to be determined directly, without the need for time consuming cloning procedures. In addition, amplified regions of the gene could be used as probes in Northern analyses of embryo, pouch young and adult mRNA, in order to verify the proposal that a switch in β-globin gene expression (from embryonic to adult) occurs within a day or two of birth.

2. isolate the gene from a genomic DNA library. This approach would enable the entire β-globin gene cluster of *S.crassicaudata* to be isolated in a series of "over-lapping" phage clones, and would also facilitate a study of DNA sequences, which

may be important in the regulation of gene expression. A similar approach (approach 2) could also be used to isolate the putative pseudogene ϵ_2.

In light of the results presented in this thesis, it would be particularly interesting to study the β-globin gene family of a monotreme. Given that the ϵ-globin of marsupials and eutherians resulted from a duplication event which occurred between 155 and 200 mya, it might be expected that monotremes would also have β-like globin genes expressed during embryonic development. Such studies, of distantly related mammalian species, will provide useful information, not only on how the β-globin gene family has evolved during mammalian evolution, but also on questions concerning: i) how member genes of α- and β-globin gene clusters are coordinately and differentially expressed during development; ii) why pseudogenes and poorly expressed globin genes accumulate in the middle of the gene cluster; and iii) how and why embryonically expressed globin genes evolved and why they are almost always located on the 5' side of globin gene clusters.

BIBLIOGRAPHY

AIR, G.M. and THOMPSON, E.O.P. (1969) 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. and 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 (Lond.) **229**: 391-394.

ALMEIDA, M.O. de, and ROCH, A. (1932) Sur la composition de l'air de la poche mammaire du gamba (*Didelphis*). C R Soc. Biol. **109**: 131.

ANAGNOU, N.P., KARLSSON, S., MOULTON, A.D., KELLER, G. and NIENHUIS, A.W. (1986) Promoter sequences required for function of the human γ globin gene in erythroid cells. The EMBO Journal **5** (1): 121-126.

ANTOINE, M. and NIESSING, J. (1984) Intron-less globin genes in the insect *Chironomus thummi thummi*. Nature (Lond.) **310**: 795.

ANTONIOU, M., DEBOER, E., HABETS, G. and GROSVELD, F. (1988) The human β-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. The EMBO Journal, **7** (2): 377-384.

APLIN, K.P. and ARCHER, M. (1987) Recent advances in marsupial systematics with a new syncretic classification. In "Possums and Opossums: Studies in Evolution" (Ed. M. Archer) Surrey Beatty and Sons, with Royal Zoological Society of New South Wales: Chipping Norton, N.S.W.

ARCHER, M., FLANNERY, T.F., RITCHIE, A. AND MOLNAR, R.E. (1985) First Mesozoic mammal from Australia - an early Cretaceous monotreme. Nature (Lond.) **318**: 363-366.

BAILEY, S.W. and DUNNET, G.M. (1960) The gaseous environment of the pouch young of the brush-tailed possum, *Trichosurus vulpecula* Kerr. CSIRO Wildl. Res. **5**: 149-151.

BARALLE, F.E., SHOULDERS, C.C. and PROUDFOOT N.J. (1980) The primary structure of the human ϵ-globin gene. Cell **21**: 621-626.

BARRIE, P.A. and JEFFREYS, A.J., SCOTT, A.F. (1981) Evolution of the β-globin gene cluster in Man and the Primates. J. Mol. Biol. **149**: 319-336.

BAUDINETTE, R.V., RUNCIMAN, S.I.C., FRAPPELL, P.F. and GANNON, B.J. (1988) Development of the Marsupial Cardiorespiratory System. In "The Developing Marsupial. Models for Biomedical Research". (Eds. C.H. Tyndale-Biscoe and P.A. Janssens). pp. 132-147.

BAVERSTOCK, P.R., KRIEG, M. and BIRRELL, J. (1990) Evolutionary relationships of Australian marsupials as assessed by albumin immunology. Aust. J. Zool. **37**: 273-287.

BEHRINGER, R.R., HAMMER, R.E., BRINSTER, R.L. PALMITER, R.D. and TOWNES, T.M. (1987) Two 3' sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. Proc. Natl. Acad. Sci. USA **84**: 7056-7060.

BENNETT, J.H., HAYMAN, D.L. and HOPE, R.M. (1986) Novel sex differences in linkage values and meiotic chromosome behaviour in a marsupial. Nature (Lond.) **323**: 59-60.

BENNETT, J.H., BREED, W.G., HAYMAN, D.L. and HOPE, R.M. (1990) Reproductive and genetical studies with a laboratory colony of the dasyurid marsupial *Sminthopsis crassicaudata*. Aust. J. Zool. **37**: 207-22.

BENOIST, C., O'HARE, K., BREATHNACH, R. and CHAMBON, P. (1980) The ovalbumin gene-sequence of putative control regions. Nucl. Acids Res. **8**: 127-142.

BENTON, W.D. and DAVIS, R.W. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**: 180-182.

BERG, P.E., WILLIAMS, D.M., QIAN, R-L., COHEN, R.B., CAO, S-X., MITTELMAN, M. and SCHECHTER, A.N. (1989) A common protein binds to two silencers 5' to the human β-globin gene. Nucl. Acids Res. **17**: 8833-8852.

BIRNBOIM, H.C. and DOLY, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**: 1513.

BLAKE, C.C.F. (1978) Do genes-in-pieces imply proteins-in-pieces? Nature (Lond.) **273**: 267.

BLANCHETOT, A., WILSON, V., WOOD, D. and JEFFREYS, A.J. (1983) The seal myoglobin gene: an unusually long globin gene. Nature (Lond.) **301**: 732-734.

BODINE, D.M. and LEY, T.J. (1987) An enhancer element lies 3^1 to the human $^A\gamma$ globin gene. The EMBO Journal **6(10)** 2997-3004.

BOGUSZ, D., APPLEBY, C.A., LANDSMANN, J., DENNIS, E.S., TRINICK, M.J. and PEACOCK, W.J. (1988) Functioning haemoglobin genes in non-nodulating plants. Nature (Lond.) **331**: 178-180.

BRIEHL, R. (1963) The relation between the oxygen equilibrium and aggregation of subunits in lamprey haemoglobin. J. Biol. Chem. **238**: 2361-2366.

BRITTEN, R.J. (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science **231**: 1393-1398.

BROWN, B.A., PADGETT, R.W., HARDIES, S.C., HUTCHISON III, C.A. and EDGELL, M.H. (1982) β-globin transcript found in induced murine erythroleukemia cells is homologous to the βho and βh1 genes. Proc. Natl. Acad. Sci. USA **79**: 2753-2757.

BRUNS, G.A. and INGRAM, V.M. (1973) The erythroid cells and haemoglobins of the chick embryo. Phil. Trans. R. Soc. London B **266**: 225-305.

CAO, S.X. GUTMAN, P.D., DAVE, H.P.G., SCHECHTER, A.N. (1989) Negative control of the human ϵ-globin gene. Prog. Clin. Biol. Res. **316A** (Hemoglobin switching, Pt. A): 279-89.

CHADA, K., MAGRAM, J., RAPHAEL, K., RADICE, G., LACY, E., COSTANTINI, F. (1985) Specific expression of a foreign β-globin gene in erythroid cells of transgenic mice. Nature (Lond.) **314**: 377-380.

CHAO, M.V., MELLON, P., CHARNAY, P., MANIATIS, T. and AXEL, R. (1983) The regulated expression of β-globin genes introduced into mouse erythroleukemia cells. Cell **32**: 483-493.

CHARNAY, P., TREISMAN, R., MELLON, P. CHAO, M., AXEL, R. and MANIATIS, T. (1984) Differences in human α- and β- globin gene expression in mouse erythroleukemia cells: The role of intragenic sequences. Cell **38**: 251-263.

CHOI, O-R.B. and ENGEL, J.D. (1988) Developmental regulation of β-globin gene switching. Cell **55**: 17-26.

CHOMCZYNSKI, P. and SACCHI, N. (1987) Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry **162**: 156-159.

CLEGG, J.B., NAUGHTON, M.A, and WEATHERALL, D.J. (1966) Abnormal human haemoglobins. Separation and characterization of the α and β chains by chromatography and the determination of two new variants, Hb Chesapeake and Hb J (Bangkok). J. Mol. Biol. **19**: 91-108.

CLEMENS, W.A., RICHARDSON, B.J. and BAVERSTOCK, P.R. (1989) Biogeography and phylogeny of the Metatheria. In "Fauna of Australia. Mammalia." (Eds. D.W. Walton and B.J. Richardson) Vol. **1B**: 527-548. (Canberra: Australian Government Publishing Service).

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

COOPER, D.W. JOHNSTON, P.G., VANDEBERG, J.L. and ROBINSON, E.H. (1990) X chromosome inactivation in marsupials. Aust. J. Zool. **37**: 411-17.

COOPER, S.J.B. (1985) The construction of a *Sminthopsis crassicaudata* genomic DNA library and isolation of β-globin clones. Unpublished Honours Thesis. The University of Adelaide.

CZELUSNIAK, J., GOODMAN, M., HEWETT-EMMETT, D., WEISS, M.L., VENTA, P.J. and TASHIAN, R.E. (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature (Lond.) **298**: 297-300.

CZELUSNIAK, J., GOODMAN, M., KOOP, B.F., TAGLE, D.A., SHOSHANI, J., BRAUNITZER, G., KLEINSCHMIDT, T.K., DeJONG, W.W. and MATSUDA, G. (1990) Perspectives from amino acid and nucleotide sequences on cladistic relationships among higher taxa of Eutheria. In "Current Mammalogy" (Ed. H.H. Genoways) Vol. **2**: 545-572. (Plenum Press: New York).

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

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

DEISSEROTH, A., NIENHUIS, A., LAWRENCE, J., GILES, R., TURNER, P. and RUDDLE, F.H. (1978) Chromosomal localization of human β-globin gene on human chromosome 11 in somatic cell hybrids. Proc. Natl. Acad. Sci. USA, **75**: 1456-1460.

DIERKS, P., VAN OOYEN, A., COCHRAN, M.D., DOBKIN,C., REISER, J. and WEISSMANN, C. (1983) Three regions upstream from the Cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**: 695-706.

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

DOLAN, M., DODGSON, J.B. and ENGEL, J.D. (1983) Analysis of the adult chicken β-globin gene. J. Biol. Chem. **258**: 3983-3990.

DOLAN, M., SUGARMAN, B.J., DODGSON, J.B. and ENGEL, J.D. (1981) Chromosomal arrangement of the chicken β-type globin genes. Cell **24**: 669-677.

EASTEAL, S. (1985) Generation time and the rate of molecular evolution. Mol. Biol. Evol. **2**: 450-453.

EASTEAL, S. (1987) The rates of nucleotide substitution in the human and rodent lineages: a reply to Li and Wu. Mol. Biol. Evol. **4**: 78-80.

EASTEAL, S. (1988) Rate constancy of globin gene evolution in placental mammals. Proc. Natl. Acad. Sci. USA **85**: 7622-7626.

EATON, W.A. (1980) The relationship between coding sequences and function in haemoglobin. Nature (Lond.) **284**: 183-185.

ECK, R.V. and DAYHOFF, M.O. (1966) Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. Science **152**: 363-366.

EDGELL, M.H., HARDIES, S.C., BROWN, B., VOLIVA, C., HILL, A., PHILLIPS, S., COMER, M., BURTON, F., WEAVER, S. and HUTCHISON III, C.A. (1983) Evolution of the mouse β-globin complex locus. In "Evolution of Genes and Proteins" (Eds. M. Nei and R.K. Koehn) Sunderland, Mass: Sinauer Assoc.

EFSTRATIADIS, A., POSAKONY, J.W., MANIATIS, T., LAWN, R.M., O'CONNELL, C., SPRITZ, R.A., DeRIEL, 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**: 653-668.

EMERSON, B.M., NICKOL, J.M., JACKSON, P.D. and FELSENFELD, G. (1987) Analysis of the tissue-specific enhancer at the 3' end of the chicken adult β-globin gene. Proc. Natl. Acad. Sci. USA **84**: 4786-4790.

ENGEL, J.D. and DODGSON, J.B. (1980) Analysis of the closely linked adult chicken α-globin genes in recombinant DNA's. Proc. Natl. Acad. Sci. USA **77**: 2596-2600.

ENVER, T., RAICH, N., EBENS, A.J., PAPAYANNOPOULOU, T., COSTANTINI, F. and STAMATOYANNOPOULOS, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature (Lond.) **344**: 309-313.

ERHART, M.A., SIMONS, K.S. and WEAVER, S. (1985) Evolution of the mouse β-globin genes: a recent gene conversion in the Hbb^s haplotype. Mol. Biol. Evol. **2**(4): 304-320.

EVANS, T. and FELSENFELD, G. (1989) The erythroid-specific transcription factor Eryfl: a new finger protein. Cell. **58**: 877-885.

EVANS, T., REITMAN,M. and FELSENFELD, G. (1988) An erythrocyte specific DNA binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. USA **85**: 5976-5980.

FARACE, M.G., BROWN, B.A., RASCHELLA, G., ALEXANDER, J., GAMBARI, R., FANTONI, A., HARDIES, S.C., HUTCHISON III, C.A. and EDGELL, M.H. (1984) The mouse βh1 gene codes for the Z chain of embryonic hemoglobin. J. Biol. Chem. **259**: 7123-7128.

FERMI, G. (1975) Three-dimensional Fourier synthesis of human deoxyhaemoglobin at 1.5Å resolution: refinement of the atomic model. J. Mol. Biol. **97**, 237-256.

FITCH, W.M. (1970) Distinguishing homologous and analogous proteins. Syst. Zool. **19**: 99-113.

FITCH, W.M. (1976) Molecular Evolutionary Clocks. In "Molecular Evolution" (Ed. F.J. Ayala) pp. 160-178. (Sinauer, Sunderland, Mass.)

FRISCHAUF, A-M., LEHRACH, H., POKSTKA, A. and MURRAY, N. (1983) Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**: 827-842.

FRITSCH, E.F., LAWN, R.M. and MANIATIS, T. (1980) Molecular cloning and characterization of the human β-like globin gene cluster. Cell **19**: 959-972.

GARNER, K.J. and LINGREL, J.B. (1988) Structural organization of the β-globin locus of B-haplotype sheep. Mol. Biol. Evol. **5**: 134 - 140.

GIGLIONI, B., CASINI C., MANTOVANI, R., MERLI, S., COMI, P., OTTOLENGHI, S., SAGLIO, G., CAMASCHELLA, C. and MAZZA, U. (1984) A molecular study of a family with Greek hereditary persistance of fetal haemoglobin and β-thalassemia. The EMBO Journal: **3**(11): 2641-2645.

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

GILMAN, J.G. (1976) Mouse haemoglobin beta chains. Biochem. J. **159**: 43-53.

Gō, M. (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature (Lond.) **291**: 90-92.

GOODMAN, M. (1981) Decoding the pattern of protein evolution. Prog. Biophys. Molec. Biol. **37**: 105-164.

GOODMAN, M., BARNABAS, J., MATSUDA, G. and MOORE, G.W. (1971) Molecular evolution in the descent of man. Nature (Lond.) **223**: 604-613.

GOODMAN, M., CZELUSNIAK, J. and BEEBER, J.E. (1985) Phylogeny of Primates and other eutherian orders: a cladistic analysis using amino acid and nucleotide sequence data. Cladistics **1(2)**: 171-185.

GOODMAN, M., CZELUSNIAK, J., KOOP, B.F., TAGLE, D.A. and SLIGHTOM, J.L. (1987) Globins: A case study in molecular phylogeny. Cold Spring Harbor Symposia on Quantitative Biology, **LII**: 875-890.

GOODMAN, M., KOOP, B.F., CZELUSNIAK, J., WEISS, M.L. and SLIGHTOM, J.L. (1984) The η-globin gene. Its long evolutionary history in the β-globin gene family of mammals. J. Mol. Biol. **180**: 803-823.

GOODMAN, M., MOORE, G.W. and MATSUDA, G. (1975) Darwinian evolution in the genealogy of haemoglobin. Nature (Lond.) **253**: 603-608.

GOODMAN, M., PEDWAYDON, J., CZELUSNIAK, J., SUZUKI, T., GOTOH, T., MOENS, L., SHISHIKURA, F., WALZ, D. and VINOGRADOV, S. (1988) An evolutionary tree for invertebrate globin sequences. J. Mol. Evol. **27**: 236-249.

GRAVES, J.A.M., HOPE, R.M. and COOPER, D.W. (1990) True beasts from pouches and eggs. Aust. J. Zool. **37**: 143-146.

GROSVELD, F., Van ASSENDELFT, G.B., GREAVES, D.R. and KOLLIAS, G. (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell **51**: 975-985.

HANSEN, J.N., KONKEL, D.A. and LEDER, P. (1982) The sequence of a mouse embryonic β-globin gene. J. Biol. Chem. **257**: 1048-1052.

HARDIES, S.C., EDGELL, M.H., HUTCHISON III, C.A. (1984) Evolution of the mammalian β-globin gene cluster. J. Biol. Chem. **259**: 3748-3756.

HARDISON, R.C. (1983) The nucleotide sequence of the rabbit embryonic globin gene β4. J. Biol. Chem. **258**: 8739-8744.

HARDISON, R.C. (1984) Comparison of the β-like globin gene families of rabbits and humans indicates that the gene cluster 5' - ϵ - γ - δ - β - 3' predates the mammalian radiation. M. Biol. Evol. **1**(5): 390-410.

HARDISON, R.C. and MAROGOT, J.B. (1984) Rabbit globin pseudogene $\psi\beta_2$ is a hybrid of δ- and β-globin gene sequences. Mol. Biol. Evol. **1(4)**: 302-316.

HARDISON, R.C., BUTLER III, 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**: 1285-1297.

HARRIS, S., BARRIE, P.A., WEISS, M.L. and JEFFREYS, A.J. (1984) The primate $\psi\beta$1 gene. An ancient β-globin pseudogene. J. Mol. Biol. **180**: 785-801.

HARRIS, S., THACKERAY, J.R., JEFFREYS, A.J. and WEISS, M.L. (1986) Nucleotide sequence analysis of the lemur β-globin gene family: evidence for major rate fluctuations in globin polypeptide evolution. Mol. Biol. Evol. **3**: 465-484.

HAYMAN, D.L., MOORE, H.D and EVANS, E.P. (1988) Further evidence of novel sex differences in chiasma distribution in marsupials. Heredity **61**: 455-8.

HESSE, J.E., NICKOL, J.M., LIEBER, M.R. and FELSENFELD, G. (1986) Regulated gene expression in transfected primary chicken erythrocytes. Proc. Natl. Acad. Sci. USA **83**: 4312-4316.

HILL, A., HARDIES, S.C., PHILLIPS, S.J., DAVIS, M.G., HUTCHISON III, C.A. and EDGELL, M.H. (1984) Two mouse early embryonic β-globin gene sequences: evolution of the nonadult β-globins. Biol. Chem. **259**: 3739-3747.

HODGSON, C.P. and FISK, R.Z. (1987) Hybridization probe size control: optimized "oligolabelling". Nucl. Acids Res. **15**: 6295.

HOLLAND, R.A.B., RIMES, A.F., COMIS, A. and TYNDALE-BISCOE, C.H. (1988) Oxygen carriage and carbonic anhydrase activity in the blood of a marsupial, the Tammar Wallaby (*Macropus eugenii*), during early development. Respir. Physiol. **73**(1): 69-86.

HOLLIS, G.F., HIETER, P.A., MCBRIDE, O.W., SWAN, D. and LEDER, P. (1982) Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. Nature (Lond.) **296**: 321-325.

HOPE, R.M. (1970) Genetic variation in natural and laboratory populations of *Trichosurus vulpecula* and *Sminthopsis crassicaudata* (Marsupialia). Ph.D. Thesis, University of Adelaide.

HOPE, R., COOPER, S. and WAINWRIGHT, B. (1990) Globin macromolecular sequences in marsupials and monotremes. Aust. J. Zool. **37**: 289-313.

HOPSON, J.A. (1970) The classification of non-therian mammals. J. Mammalogy **51**: 1-9.

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

HSU, S.L., MARKS, J., SHAW, J.P., TAM, M., HIGGS, D.R., SHEN, C-C., SHEN, C-K.J. (1988) Structure and expression of the human θ_1, globin gene. Nature (Lond.) **331**: 94-96.

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

HUNKAPILLER, M.W. and HOOD, L.E. (1978) Direct microsequence analysis of polypeptides using an improved sequenator, a non-protein carrier (polybrene) and high pressure liquid chromatography. Biochem. **17**: 2124-2133.

HUNT, T.L., HURST-CALDERONE, S. and DAYHOFF, M.O. (1978) Atlas of Protein Sequence and Structure. Ch. 13. Globins. (Ed. M.O. Dayhoff). National Biomedical Research Foundation, Washington, D.C., Vol. 5, Supplement 3, pp 229-251.

HYLDIG-NIELSEN, J.J., JENSEN, EØ., PALUDAN, K., WIBORG O., GARRETT, R., JØRGENSEN P. and MARCKER K.A. (1982) The primary structures of two leghemoglobin genes from soybean. Nucl. Acids Res. **10**: 689-701.

JAHN, C.L., HUTCHISON III, C.A., PHILLIPS, S.J., WEAVER, S., HAIGWOOD, N.L., VOLIVA, C.F. and EDGELL, M.H. (1980) DNA sequence organization of the β-globin complex in the BALB/c mouse. Cell **21**: 159-168

JEFFREYS, A.J. (1982) Evolution of globin genes. In "Genome Evolution" (Eds. G.A. Dover and R.B. Flavell), Academic Press, London.

JEFFREYS, A.J., BARRIE, P.A., HARRIS, S., FAWCETT, D.H., NUGENT, Z.J. and BOYD, A.C. (1982) Isolation and sequence analysis of a hybrid δ-globin pseudogene from the brown lemur. J. Mol. Biol. **156**: 487-503.

JEFFREYS, A.J., WILSON, V., WOOD, D., SIMONS, J.P., KAY, R.M. and WILLIAMS, J.G. (1980) Linkage of adult α- and β-globin genes in *X. laevis* and gene duplication by tetraploidization. Cell **21**: 555-564.

JENSEN, E.O., PALUDAN, K., HYLDIG-NIELSEN, J.J., JORGENSEN, P. and MARCKER, K.A. (1981) The structure of a chromosomal leghaemoglobin gene from soybean. Nature (Lond.) **291**: 677-679.

JEPPSSON, J.O. and SJOQUIST, J. (1967) Thin-layer chromatography of PTH amino-acids. Analyt. Biochem. **18**: 264-269.

JHIANG, S.M., GAREY, J.R., and RIGGS, A.F. (1988) Exon-intron organization in genes of earthworm and vertebrate globins. Science **240**: 334-336.

KARN, J., BRENNER, S., BARNETT, L. and CESARENI, G. (1980) Novel bacteriophage λ cloning vector. Proc. Natl. Acad. Sci. USA. **77**: 5172-76.

KEMP, T.S. (1983) The relationships of mammals. Zool. J. Linn. Soc. **77**: 353-84.

KERMACK, K.A. and KIELAN-JAWOROWSKA, Z. (1971) Therian and non-therian mammals. In "Early Mammals". (Eds. D.M. Kermack and K.A. Kermack) Zoological Journal of the Linnean Society 50, Supplement **1**: 103-115.

KIELAN-JAWOROWSKA, Z., CROMPTON, A.W. and JENKINS, F.A.Jr. (1987) The origin of egg laying mammals. Nature (Lond.) **326**: 871-873.

KILMARTIN, J.V. (1976) Interaction of haemoglobin with protons, CO_2 and 2,3-diphosphoglycerate. Br. Med. Bull. **32**: 209-212.

KIMURA, M. (1981) Was globin evolution very rapid in its early stages?: a dubious case against the rate-constancy hypothesis. J. Mol. Evol., **17**: 110-113.

KIMURA, M. (1983) The neutral theory of molecular evolution. (Cambridge Univ. Press: London).

KIRSCH, J.A.W. (1977) The comparative serology of Marsupialia, and a classification of marsupials. Aust. J. Zool., Suppl. Ser. No. **52**.

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).

KNÖCHEL, W., KORGE, E., BASNER, A. and MEYERHOF, W. (1986) Globin evolution in the genus *Xenopus*: comparative analysis of cDNAs coding for adult globin polypeptides of *Xenopus borealis* and *Xenopus tropicalis*. J. Mol. Evol. **23**: 211-223.

KOLLIAS, G., HURST, J., DeBOER, E. and GROSVELD, F. (1987) The human β-globin gene contains a downstream developmental specific enhancer. Nucl. Acids Res. **15** (14): 5739-5747.

KONKEL, D.A., MAIZEL, J.V. Jr. and LEDER, P. (1979) The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell **18**: 865-873.

KONKEL, D.A., TILGHMAN, S.M. and LEDER, P. (1978) The sequence of the chromosomal mouse β-globin major gene: homologies in capping, splicing and poly(A) sites. Cell **15**: 1125-1132.

KOOP, B.F. and GOODMAN, M. (1988) Evolutionary and developmental aspects of two β-haemoglobin genes (ϵ^m and β^m) of opossum. Proc. Natl. Acad. Sci. USA **85**: 3893-3897.

KOOP, B.F., GOODMAN, M., XU, P., CHAN, K. and SLIGHTOM, J.L. (1986) Primate η-globin DNA sequences and man's place among the great apes. Nature (Lond.) **319**: 234-237.

KOOP, B.F., TAGLE, D.A., GOODMAN, M. and SLIGHTOM, J.L. (1989) A molecular view of primate phylogeny and important systematic and evolutionary questions. Mol. Biol. Evol. **6**: 580-612.

LACY, E. and MANIATIS, T. (1980) The nucleotide sequence of a rabbit β-globin pseudogene. Cell **21**: 545-553.

LACY, E., HARDISON, R.C., QUON, D. and MANIATIS, T. (1979) The linkage arrangement of four rabbit β-like globin genes. Cell **18**: 1273-1283.

LADNER, R.C., HEIDNER, E.J. and PERUTZ, M.F. (1977) The structure of horse methaemoglobin at 2.0 Angstrom units resolution. J. Mol. Biol. **114**, 385-414.

LAMB, P., WATT, P., PROUDFOOT, N.J. (1989) Negative regulation of the human embryonic globin genes ζ and ϵ. Prog. Clin. Biol. Res. **316A** (Hemoglobin switching, Pt. A): 269-77.

LANDSMANN, J., DENNIS, E.S., HIGGINS, T.J.V. APPLEBY,C.A., KORTT, A.A. and PEACOCK, W.J. (1986) Common evolutionary origin of legume and non-legume plant haemoglobins. Nature (Lond.) **324**: 166-168.

LAUER, J., SHEN, C-K.J. and MANIATIS, T. (1980) The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions. Cell **20**: 119-130.

LAWN, R.M., EFSTRATIADIS, A., O'CONNELL, C. and MANIATIS, T. (1980) The nucleotide sequence of the human β-globin gene. Cell **21**: 647-651.

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

LEE, J.S., BROWN, G.G. and VERMA, D.P.S. (1983) Chromosomal arrangement of leghemoglobin genes in soybean. Nucl. Acids Res. **11**: 5541-5553.

LEUNG, S., PROUDFOOT, N.J. and WHITELAW, E. (1987) The gene for θ-globin is transcribed in human fetal erythroid tissues. Nature (Lond.) **329**: 551-554.

LI, S.L. and RIGGS, A. (1970) The amino-acid sequence of haemoglobin V from the lamprey, *Petromyzon marinus*. J. of Biol. Chem. **245**: 6149-6169.

LI, W-H., GOJOBORI, T. and NEI, M. (1981) Pseudogenes as a paradigm of neutral evolution. Nature (Lond.) **292**: 237-239.

LI, W-H. and TANIMURA, M. (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature (Lond.) **326**: 93-96.

LI, W-H., WU, C-I. and LUO, C-C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**: 150-174.

LLOYD, J.A., LEE, R.F., MENON, A.G., LINGREL, J.B. (1989) Sites I and II upstream of the $^{A}\gamma$ globin gene bind nuclear factors and affect gene expression. Prog. Clin. Biol. Res. **316A** (Haemoglobin Switching, Pt. A): 139-48.

LOCKLEY, R.M. (1964) The private life of the rabbit. Macmillan, New York.

LOVE, W.E., KLOCK, P.A., LATTMAN, E.E., PADLAN, E.A., WARD, K.B.Jr. and HENDRICKSON, W.A. (1971) The structure of lamprey and bloodworm hemoglobins in relation to their evolution and function. Cold Spring Harbor Symposia on Quantitative Biology **36**: 349-357.

MAGRAM, J., CHADA, K. and COSTANTINI, F. (1985) Developmental regulation of a cloned adult β-globin gene in transgenic mice. Nature (Lond.) **315**: 338-340.

MANIATIS, T., FRITSCH, E.P. and SAMBROOK, J. (1982) Molecular cloning. A laboratory manual. (Cold Spring Harbour Press: USA).

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**, 163-182.

MANTOVANI, R., MALGARETTI, N., NICOLIS, S., GIGLIONI, B., COMI, P., CAPELLINI, N., BERTERO, M.T., CALIGARIS-CAPPIO, F. and OTTOLENGHI, S. (1988) An erythroid specific nuclear factor binding to the proximal CACCC box of the β-globin gene promoter. Nucl. Acids Res. **16**: 4299-4313.

MANTOVANI, R., SUPERTI-FURGA, G., GILMAN, J. and OTTOLENGHI, S. (1989) The deletion of the distal CCAAT box region of the $^{A}\gamma$-globin gene in black HPFH abolishes the binding of the erythroid specific protein NFE3 and of the CCAAT displacement protein. Nucl. Acids Res. **17**: 6681-6691.

MARKS, J., SHAW, J-P., SHEN, C-K.J. (1986) Sequence organization and genomic complexity of primate $\theta 1$ globin gene, a novel α-globin-like gene. Nature (Lond.) **321**: 785-788.

MARTIN, D.I.K., TSAI, S-F. and ORKIN, S.H. (1989) Increased γ-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. Nature (Lond.) **338**: 435-438.

MARTIN, P.G. and JENNINGS, A.C. (1983) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphoshate carboxylase. I. Biochemical methods and the patterns of variability. Aust. J. Bot. **31**: 395-409.

MAXAM, A.M. and GILBERT, W. (1980) Sequencing end-labelled DNA with base specific chemical cleavages. In "Methods in Enzymology" (Ed. L. Grossman and K. Moldave). Vol. **65(1)**, pp. 499-560. (Academic Press: New York).

MAXON, L.R., SARICH, V.M. and WILSON, A.C. (1975) Continental drift and the use of albumin as an evolutionary clock. Nature (Lond.) **255**: 397-400.

MESSING, J. (1983) New M13 vectors for cloning. In "Methods in Enzymology" (Ed. R. Wu, L. Grossman and K. Moldave) **101**: 20. (Academic Press: New York).

MESSING, J., CREA, R. and SEEBURG, P.H. (1981) A system for shotgun DNA sequencing. Nucl. Acids Res. **9**: 309-321.

MIYATA, T. and HAYASHIDA, H. (1981) Extraordinarily high evolutionary rate of pseudogenes: Evidence for the presence of selective pressure against changes between synonymous codons. Proc. Natl. Acad. Sci. USA **78**: 5739-5743.

MIYATA, T. and YASUNAGA, T. (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J. Mol. Evol. **16**: 23-26.

MONOD, J., WYMAN, J. and CHANGEUX, J.P. (1965) On the nature of allosteric transitions: A plausible model. J. Mol. Biol. **12**: 88-118.

MOUNT, S.M. (1982) A catalogue of splice junction sequences. Nucl. Acids Res. **10**: 459-472.

MURPHY, W.S., METCALFE, J., HOVERSLAND, A.S. and DHINDSA, D.S. (1977) Postnatal changes in blood respiratory characteristics in an American opossum (*Didelphis virginiana*). Respir. Physiol. **29**: 73-80.

MYERS, R.M., TILLY, K., MANIATIS, T. (1986) Fine structure genetic analysis of a β-globin promoter. Science **232**: 613-618.

NEI, M. and GOJOBORI, T. (1986) Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol. Biol. Evol. **3**(5): 418-426.

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**: 2806-2809.

ORKIN, S.H. (1978) The duplicated human α globin genes lie close together in cellular DNA. Proc. Natl. Acad. Sci. USA **75**: 5950-5954.

ORKIN, S.M. and KAZAZIAN, H.H. Jr. (1984) The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Ann. Rev. Genet. **18**: 131-171.

PADGETT, R.W., LOEB, D.D., SNYDER, L.R.G., EDGELL, M.H., and HUTCHISON III, C.A. (1987) The molecular organization of the beta-globin complex of the deer mouse, *Peromyscus maniculatus*. Mol. Biol. Evol. **4**(1): 30-45.

PATIENT, R.K., ELKINGTON, J.A., KAY, R.M. and WILLIAMS, J.G. (1980) Internal organization of the major adult α- and β- globin genes of *X. laevis*. Cell **21**: 565-572.

PERLER, F., EFSTRATIADIS, A., LOMEDICO, P., GILBERT, W., KOLODNER, R. and DODGESON, J. (1980) The evolution of genes: the chicken preproinsulin gene. Cell **20**: 555-566.

PERUTZ, M.F. (1970) Stereochemistry of cooperative effects in haemoglobin. Nature (Lond.) **228**: 726-739.

PERUTZ, M.F. (1976) Structure and mechanism of haemoglobin. Br. Med. Bull. **32**: 195-208.

PERUTZ, M.F. and LEHMANN, H. (1968) Molecular pathology of human haemoglobin. Nature (Lond.) **219**: 902.

PERUTZ, M.F., MUIRHEAD, H., COX, J.M. and GOAMAN, L.C.G. (1968) Three dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Angstrom units resolution: the atomic model. Nature (Lond.) **219**, 131-9.

PESCHLE, C., MAVILIO, F., CARE, A., MIGLIACCIO, G., MIGLIACCIO, A.R., SALVO, G., SAMOGGIA, P., PETTI, S., GUERRIERO, R., MARINUCCI, M., LAZZARO, D., RUSSO, G. and MASTROBERARDINO, G. (1985) Haemoglobin switching in human embryos: asynchrony of $\zeta \rightarrow \alpha$ and $\epsilon \rightarrow \gamma$-globin switches in primitive and definitive erythropoietic lineage. Nature (Lond.) **313**: 235-238.

PROUDFOOT, N.J., GIL, A. and MANIATIS, T. (1982) The structure of the human zeta-globin gene and a closely linked nearly indentical 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**: 537-544.

RASHIN, A.A. (1981) Location of domains in globular proteins. Nature (Lond.) **291**: 85-87.

RICHARDSON, B.J. and RUSSELL, 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. Exp. Biol. Med. Sci., **47**: 573-80.

ROHRBAUGH, M.L. and HARDISON, R.C. (1983) Analysis of rabbit β-like globin gene transcripts during development. J. Mol. Biol. **164**: 395-417.

RONINSON, I.B. and INGRAM, V.M. (1982) Gene evolution in the chicken β-globin cluster. Cell **28**: 515-521.

ROSENBERG, N. and COURT, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. **13**: 319-353.

RUNNEGAR, B. (1984) Derivation of the globins from type b cytochromes. J. Mol. Evol. **21**: 33-41.

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T. (1989) Molecular Cloning. A Laboratory Manual. 2nd ed. (Cold Spring Harbor Laboratory Press: USA).

SANGER, F., COULSON, A.R., BARNELL, B.G., SMITH, A.J.H. and ROE, B.A. (1980) Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**: 161-178.

SARICH, V.M. and WILSON, A.C. (1967) Rates of albumin evolution in primates. Proc. Natl. Acad. Sci. USA **58**: 142-148.

SCHIMENTI, J.C. and DUNCAN, C.H. (1985) Structure and organization of the bovine β-globin genes. Mol. Biol. Evol. **2(6)**: 514-525.

SCHMID, C.W. and JELINEK, W. (1982) The *Alu* family of dispersed repetitive sequences. Science **216**: 1065-70.

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**: 359-369.

SCOTT, A.F., HEATH, P., TRUSKO, S., BOYER, S.H., PRASS, W., GOODMAN, M., CZELUSNIAK, J., CHANG, L-Y.E., SLIGHTOM, J.L. (1984) The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. Mol. Biol. Evol. **1(5)**: 371-389.

SHAPIRO, S.G., SCHON, E.A. TOWNES, T.M. and LINGREL, J.B. (1983) Sequence and linkage of the goat ϵ^1 and ϵ^{11} β-globin genes. J. Mol. Biol. **169**: 31-52.

SHAW, J-P., MARKS, J. and SHEN, C-K.J. (1987) Evidence that the recently discovered θ1-globin gene is functional in higher primates. Nature (Lond.) **326**: 717-720.

SHAW, J.P., MARKS, J., SHEN, C.C. and SHEN, C-K.J. (1989) Anomalous and selective DNA mutations of the old world monkey α-globin genes. Proc. Natl. Acad. Sci. USA, **86**: 1312-16.

SHEN, C.C., BAILEY, A., KIM, J.H., YU, C.Y., MARKS, J., SHAW, J.P., KLISAK, I., SPARKES, R. and SHEN, C-K.J. (1989) The human α_2-α_1-θ1 globin locus: some thoughts and recent studies of its evolution and regulation. Prog. Clin. Biol. Res. **316B** (Hemoglobin Switching, Pt. B), 19-32.

SHEN, S., SLIGHTOM, J.L. and SMITHIES, O. (1981) A history of the human fetal globin gene duplication. Cell **26**: 191-203.

SIBLEY, C.G. and AHLQUIST, J.E. (1987) DNA hybridization evidence of homonoid phylogeny: results from an expanded data set. J. Mol. Evol. **26**: 99-121.

SLIGHTOM, J.L., BLECHL, A.E. and SMITHIES, O. (1980) Human $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**: 627-638.

SMITH, C.A.B. (1968) Linkage scores and corrections in simple two- and three-generation families. Ann. Hum. Genet. Lond. **32**: 127-150.

SMITH, M.J. and GODFREY, G.K. (1970) Ovulation induced by gonadotrophins in the marsupial, *Sminthopsis crassicaudata* (Gould). J. Reprod. Fert. **22**: 41-47.

SOUTHERN, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., **98**: 503-517.

STAMATOYANNOPOULOS, G. and NIENHUIS, A.W. (1987) In "Molecular Basis of Blood Diseases" (Eds. G. Stamatoyannopoulos, A.W. Nienhuis, P. Leder and P. Majerus). pp. 66-93. (W.B. Saunders, Philadelphia, USA).

STONEHOUSE, B. (1977) Introduction: The marsupials. In "The Biology of Marsupials", (Eds. B. Stonehouse and D. Gilmore). Macmillan Press Ltd. London.

TAGLE, D.A., KOOP, B.F., GOODMAN, M., SLIGHTOM, J.L., HESS, D.L. and JONES, R.T. (1988) Embryonic ϵ- and γ-globin genes of a prosimian primate (*Galago crassicaudatus*): nucleotide and amino-acid sequences, developmental regulation and phylogenetic footprints. J. Mol. Biol. **203**: 439-455.

TALBOT, D., COLLIS, P., ANTONIOU, M., VIDAL, M., GROSVELD, F. and GREAVES, D.R. (1989) A dominant control region from the human β-globin locus conferring integration site-independent gene expression. Nature (Lond.) **338**: 352-355.

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**: 1199-217.

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.

TOWNES, T.M., FITZGERALD, M.C. and LINGREL, J.B. (1984) Triplication of a four-gene set during evolution of the goat β-globin locus produced three genes now expressed differentially during development. Proc. Natl. Acad. Sci. USA, **81**: 6589-6593.

TUAN, D., SOLOMON, W., LI, Q. and LONDON, I.M. (1985) The "β-like-globin" gene domain in human erythroid cells. Proc. Natl. Acad. Sci. USA **82**: 6384-6388.

TUAN, D.Y.H., SOLOMON, W.B., LONDON, I.M. and LEE, D.P. (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "β-like-globin" genes. Proc. Natl. Acad. Sci. USA **86**: 2554-2558.

UEDA, S. and SCHNEIDER, R.G. (1969) Rapid differentiation of polypeptide chains of hemoglobins by cellulose acetate electrophoresis of hemolysates. Blood **34**: 230-235.

Van DAAL, A., COOPER, D.W. and MOLLOY, P.L. (1989) A marsupial phosphoglycerate kinase (PGK) processed pseudogene. Genomics **5**: 264-269.

VANIN, E.F., GOLDBERG, G.I., TUCKER, P.W. and SMITHIES, O. (1980) A mouse α-globin-related pseudogene lacking intervening sequences. Nature (Lond.) **286**: 222-226.

VAN SANTEN, V.L. and SPRITZ, R.A. (1985) mRNA precursor splicing *in vivo*: sequence requirements determined by deletion analysis of an intervening sequence. Proc. Natl. Acad. Sci. USA **82**: 2885-2889.

WAHL, G.M., STERN, M. and STARK, G.R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzloxymethal-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. **76**: 3683.

WAINWRIGHT, B.J. (1984) Globin gene mapping in the marsupial *Dasyurus viverrinus*. Ph.D. thesis, University of Adelaide.

WAINWRIGHT, B. and HOPE, R.M. (1985) Cloning and chromosomal locations of the α- and β-globin genes from a marsupial. Proc. Natl. Acad. Sci. USA **82**: 8105-8108.

WAKABAYASHI, S., MATSUBARA, H. and WEBSTER, D.A. (1986) Primary sequence of a dimeric bacterial haemoglobin from *Vitreoscilla*. Nature (Lond.) **322**: 481-483.

WEATHERALL, D.J. and CLEGG, J.B. (1979) Recent developments in the molecular genetics of human hemoglobin. Cell **16**: 467-479.

WEAVER, S., COMER, M.B., JAHN, C.L., HUTCHISON III, C.A. and EDGELL, M.H. (1981) The adult β-globin genes of the "single" type mouse C57B1. Cell **24**: 403-411.

WEISLANDER, L. (1979) A simple method to recover intact high molecular weight RNA and DNA after electrophoretic spepartion in low gelling temperature agarose gels. A. Biochem. **98**: 305-309.

WEISSBLUFF, M. (1974) Hemoglobin: co-operativity and electronic properties. (Springer-Verlag, Berlin, Heidelberg).

WELLER, P., JEFFREYS, A.J., WILSON, V. and BLANCHETOT, A. (1984) Organization of the human myoglobin gene. The EMBO Journal **3(2)**: 439-446.

WESTERMAN, M. and EDWARDS, D. (1991) The divergence between echidna and platypus - new data from DNA studies. Aust. Mammal. **14**: in the press.

WESTERMAN, M., JANCZEWSKI, D.N. and O'BRIEN, S.J. (1990) DNA-DNA - hybridisation studies and marsupial phylogeny. Aust. J. Zool. **37**: 315-323.

WHITTAKER, R.G., FISHER, W.K. and THOMPSON, E.O.P. (1972) Studies on monotreme proteins. I. Amino acid sequence of the β chain of haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus.* Aust. J. Biol. Sci. **25**: 989-1004.

WHITTAKER, R.G., FISHER, W.K. and THOMPSON, E.O.P. (1973) Studies on monotreme proteins. II. Amino acid sequence of the α-chain in haemoglobin of from the echidna *Tachyglossus aculeatus aculeatus.* Aust. J. Biol. Sci. **26**: 877-888.

WHITTAKER, R.G., FISHER, W.K. and THOMPSON, E.O.P. (1980) Monotreme haemoglobin and myoglobin amino acid sequences and their use in phylogenetic divergence point estimations. Aust. J. Zool. **20**: 57-68.

WHITTAKER, R.G. and THOMPSON, E.O.P. (1975) Studies on monotreme proteins. VI. Amino acid sequence of the β-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus*. Aust. J. Biol. Sci. **28**: 353-365.

WILSON, A.C., CARLSON, S.S. and WHITE, T.J. (1977) Biochemical evolution. Ann. Rev. Biochem. **46**: 573-639.

WITTENBERG, J.B. (1970) Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. Physiol. Rev. **50**: 559-636.

WOOLLEY, P. (1974) 3. The pouch of *Planigale subtilissima* and other dasyurid marsupials. J. of the Royal Society of Western Australia **57**: 11-15.

WRIGHT, S., ROSENTHAL, A., FLAVELL, R. and GROSVELD, F. (1984) DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell **38**: 265-273.

WU, C-I., and LI, W-H. (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82**: 1741-45.

YAMAMOTO, K.R., ALBERTS, B.M., BENZINGER, R., LAWHORNE, L. and TREIBER, G. (1970) Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. Virology **40**: 734.

YOUNG, P.R., SCOTT, R.W., HAMER, D.H. and TILGHMAN, S.M. (1982) Construction and expression *in vivo* of an internally deleted mouse α-fetoprotein gene: presence of a transcribed *alu*-like repeat within the first intervening sequence. Nucl. Acids. Res. **10**: 3099-3116.

ZUCKERKANDL, E. and PAULING, L. (1965) Evolutionary divergence and convergence in proteins. In "Evolving Genes and Proteins" (Eds. V. Bryson and H.J. Vogel). pp 97-116. (Academic Press, New York).

APPENDIX

Fig. A.1

5' flanking DNA sequence of S.c-β determined from one strand only. The sequences are given in the order: 5' - 3'and are the same DNA strand as the transcribed strand of S.c-β. Sequence from clone a (see Fig. 4.1.10a) was determined from the PstI (CTGCAG) site of pSG-2. The sequence given for clone b is complementary to the DNA sequence actually determined from this clone (Fig. 4.1.10). A HindIII site (AAGCTT) was present in clone b and c (Fig. 4.1.10), but was not sequenced due to technical difficulties.

5' Flanking region of S.c-β from the PstI site of pSG-2.

a

5'

CTGCAGGTCA TACGGATCAG AACAAAAGAG AAAAGCCACA AAAGTCATTT
TTTTTTGAGA GTGGAAATAA TATGCTTTGA TCTGCATTCA GATTCCATAG
ATGTTTCTCT GAATGTGTAT TTTCCATCAT GTGTCTTTTG AAACTTTCTT
GGATCATAAC TGAAAAAAGC TAAGTCACTC ATAGTTGATT ATTGAACAAT
GCTGTTGGGC CCTTTTTTCT ATTTAGTATT TTATTTTTCC ACATTTACAT
ATAAAACAGT TTTTAAATAT ATATTTTTTT AAATTTTTAA ATTTTAAATT
TTCTTCCTTC TTCCCCTCTT ACCACCCCCT CACTGAGAAT GCAAGCAATT
CAAATAAGGT AACATAGAGT CATCGAGACT TCCT

b

AAGCACATAT ATTTCATGAT AATCATATAC CACAGTTATT CAGCCATCCC
CAATGATGAG CATTTTCATA TTCTTAATTC TTCTTGCCCT CAGAAAGAAT
TGCTAGAAAT ATTTTTGTAC ATATAGGTCC TTTCCTTTTT TTTTTTTTTT
AAATCTCCTT TGGGATAGAG ACCTAGTAGT AGTATTGTTA GGTCAAATAG
TATTCACAGT TTAATTGCCT TTTGGGCATA GTTCCAAATT GCTCTCTAGA
ACAGTGGGAT ACATTCACAA CTTCATCAGC AATACATTAG TGACCTATTT
TCCTACATCT TTTCAAACAT TTATAATTTC TCTTTTTTTT CATATTACCA
ATCTGATGAA GTAATACCTC AGAATTGTTT TAGTTTGATT TATTAAACAA
TAAGAGAATT TTTTTTCATA TAGTTA.... AAGCTT

c

TTACCAGAAA TACCAACTAA AATTGTTCCC CAGATTTATG CTTTCCTTTC
AGTCATGGTT GTGTTTTTTT TGGTTTGTGT AAAATCATTT TAATTTAATG
TAATAGAATT AATGAAATCA AAATGTTGCA CTTTATAATG TTTTCTATCT
CGTTTGGCCA TAAATTCTTC CCTTATCCAT AGATCT

Fig. A.2

3' flanking DNA sequence of S.c-β determined from one strand only. The sequences are given in the order: 5' - 3'; and are the same DNA strand as the transcribed strand of S.c-β.

A: DNA sequence from the BglII restriction site (AGATCT) located immediately adjacent to the end of the 3rd exon of S.c-β (Fig. 4.1.10b) and comprises the sequence of a Bal31 clone (clone e, starts from base with an *) which overlapped with sequence determined from the BglII site of pSG-12. Also underlined is the tandem repeat referred to in Fig. 4.1.14.

B: Partial sequence of the 3.3 kb EcoRI fragment of pSG-12 from clone f (Fig. 4.1.10b). Shown is the complementary DNA sequence to that actually determined from clone f; underlined is the EcoRI site.

C: DNA sequence of an unknown Bal31 clone located 3' to S.c-β. Comparisons involving this sequence and that of pBR322 native DNA indicate this sequence was determined from a region of the *S.crassicaudata* genomic insert of pSG-12.

A

3' Flanking sequence of S.c-β from BglII site of pSG-12.

AGATCTATCT GTGGTGGGGG TCTATGGGCA AATGGAAGCC CAACCATTCA
TACTGCCTTG TAGTCCTAGC CCTTAACTAA TAAAGTGCTT TTGGCCTTGC
AGTTCATTGT CATTGTGTTT TACTTTTAAT CATCCTTACT GGGAGATGGT
GACATAGGTG GGGATAAGAC CTAAGGAAGT CTTTTGTCCT AGCCCTTCCT
TCATGCTTAT TCTAAGAATG AAGAACAAAT CTGGAGTATC TAAGGGATAC
AAAGTTTGGA ACAGAGTGTA AGATAGAAAC ACTGGGAAAT AAGGTTGAAC
AAGTTGACTG TGTGCAAATT GTGGCAGATT TTTAATGACA ATCTGAGAAA
TTCTGCCAAT TAATCAATCA GCAAACATTT GTTAAACACC TACTATGTGC
TAGGCATTAT GGAGAGGCAA AAGACAGTCC CTGCTCTGGT GGAGCTACAG
TCCCTGCTCT AGTGGAGCTT ACAATATAAT TGTGTGTGTG TTGGCCAGGG
TACGGTAGGG GAGAGGTGGG GAGGGAGAAA CCACATGCAA AAAATATTAT
ACAATGAACT AGAATTTAGA GGGGGAAGAA AGGCTTCCTA TAAAAGATGA
GTTTATGGTT GGGACTTAAA AGAAGCCAAA TAAGTCAGTA GGTAGACTTG
AGAAGGGAGA GCATTTGTAG GGGACTGATA GAGTAAGTGC CCAAAACAAG
AAATGGGCCT CTTGCTTGTG CTTCATTGCT GATGAATGAA TGACTATGCA
AGATTGTTGA ATAGGAATTA TGATGCAACT AGAACCTTGC TTTGGAATAT
TAGAAGGGAT TGGATGAGGA CGCTTGAGAC CTAATGTAAT ACTTCTACCA
TAGACTATAG GATCTAAATT GGAAAAGATC TATAGTTC

B

3' Flanking sequence: partial sequence of 3.3 kb EcoRI fragment of pSG-12.

ATGTACCTCA TATCAGTTCA TTTGATTAAT ATAGATTCTT TTTTCCTTTA
CTTTCTTCTG CAAAGTTCCT TCTATTCTGT ATTTTAAATT TCAGATTTGT
TAGTTTCTTT CTATGTTTCT GTTTTCTCTC TCATTGTCTC TGTCTCTCAT
TCCACTACTA ACTAAAGATA CTTGACAAGA TTTTCTATTT TCCATTTTGT
ATTCACTTTT TTCTGTATTG TTGGTGGGAA TTTTTTGAAT CTGAATTGAA
AATTTTTTGA GTCTGAATTA AGTCTGACTT ACCTTCATTT TTAATTTCCT
CTCTGACAAG TTTTATTTCC CTTGGGAATC ATAATGGAAT TC

C

DNA sequence of an unknown Bal31 clone located 3' to S.c-β.

AATTCAGAAC TGGTTAAATG CTTAGAATCC AAGGGCAGTC ATTAATGATT
TAAAGTAACC TTGGGAGGAA GATTCCAGTG GAGTTCCAGG GGCTTTGCTT
GGTTTAAGGA GGACAGAGTA GAGATTTAAA AATGTTAAAA CAAAAAAAGT
TTCGAGACTC TATAGTATTG AATTGAATCT AAGAAGATGA AATTTAGTAA
AGATAAATAT AAAGTCTTAA ACTTGGGTTA ATAACATACA ACTTCCCAAG
TACAAGAATC ATTAGACAGA ACTTTAAAAA TTAAAAAAAA AAGAGATCTG
AAAGTTTTAA TGAATTTCAA GCTCACTAGG GTCTAGGTCG TACAGTGAAT
AGAGTACCAG GTCTGTAGTC AAAAGGATTC ATCTTTCTGA GTTCAAATCT
GGCATCAAGC ACTTATTAGA TGTGTGACCC TTAACCCTGT TTGCCTCAGT
TCCCTAATCT GAACCAGAAA AGAAAATGGT GAATCACCCC AGTATCTCCA
CCAAGAAAAC CCTAAATGAG GTCATGAAGA ATTGGACATG ACTAAACAAT
TGAACAACAA AAGTCATCAG TGTGGGTAGA A

Fig. A3

DNA sequence of coding regions of eutherian and marsupial β-like globin genes used in the estimation of non-synonymous and synonymous divergence values presented in section 4.2.4. The format of each sequence is that used in a computer program for estimating divergence values (kindly provided by Dr.'s W-H. Li, C-I. Wu and C-C. Luo). The abbreviation and source of each sequence are: SMIN:- S.c-β (present study, see Fig 4.1.16); OPEM and OPAD:- *D.virginiana* ϵ^m- and β^m-globin genes respectively (Koop and Goodman, 1988); NCAD:- partial sequence of a β-globin mRNA (Wainwright, 1984); MOEM:- mouse βh1-globin gene from Hill *et al.*, 1984); MOAD:- mouse βmaj-globin gene (Konkel *et al.*, 1978, 1979); HUAD:- human adult β-globin gene (Lawn *et al.*, 1980); HUEM:- human ϵ-globin gene (Baralle *et al.*, 1980); HUGA:- human γ-globin gene (Slightom *et al.*, 1980); CHAD:- chicken adult β-globin gene (Dolan *et al.*, 1982).

441 GLOBIN SMIN

*** GTG CAC TTC ACA GCT GAG GAG AAG AAC GCC ATC ACT ACC ATT TGG GGC AAG GTC AAT
GTG GAA GAG ACT GGT GGG GAA GCT TTA GGC AGA CTG TTG GTT GTC TAC CCC TGG ACC CAG
AGG TTC TTT GAT TCC TTT GGC AAC CTC TCT TCT GCC TCT GCT ATC CTG GGA AAC CCC AAG
GTT AAG GCC CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAT GCT GTC AAG AAC CTG GAC
AAC CTC AAG GGT ACC TTC TCC AAG CTG AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAC
CCT GAG AAC TTC CGG CTC CTT GGG AAT GTT CTG GTT ATT GTT ATG GCT GCT CAT TTC AAC
AAG GAA TTC ACT CCT GAA GTG CAG GCT GCT TTT CAG AAG CTG GTG ACT GGT GTG GCC AAT
GCT TTG GCT CAC AAG TAC CAC

441 GLOBIN OPEM

*** GTG CAT TTC ACC CCT GAG GAC AAG ACC AAT ATC ACT TCT GTG TGG ACC AAG GTC GAT
GTG GAA GAC GTT GGT GGG GAA TCT TTG GCC AGG CTG CTG GTA GTC TAT CCC TGG ACT CAG
AGG TTC TTT GAT TCC TTT GGC AAC CTA TCT TCT GCA TCT GCT GTT ATG GGG AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAA GGT GTC AAG AAC ATG GAC
AAC CTC AAG GGT ACC TTT GCC AAG CTG AGT GAG CTG CAC TGT GAC AAA CTG CAT GTA GAC
CCT GAG AAC TTC AGG CTC CTC GGG AAT GTG CTG ATT ATT GTT TTG GCT TCT CGT TTT GGC
AAG GAG TTC ACT CCA GAA GTG CAG GCT TCT TGG CAG AAG CTG GTG TCT GGT GTG TCC AGT
GCC TTG GGC CAC AAG TAC CAC

441 GLOBIN OPAD

*** GTG CAC TTG ACT TCT GAG GAG AAG AAC TGC ATC ACT ACC ATC TGG TCT AAG GTG CAG
GTT GAC CAG ACT GGT GGT GAG GCC CTT GGC AGG ATG CTC GTT GTC TAC CCC TGG ACC ACC
AGG TTT TTT GGG AGC TTT GGT GAT CTG TCC TCT CCT GGC GCT GTC ATG TCA AAT TCT AAG
GTT CAA GCC CAT GGT GCT AAG GTG TTG ACC TCC TTC GGT GAA GCA GTC AAG CAT TTG GAC
AAC CTG AAG GGT ACT TAT GCC AAG TTG AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAC
CCT GAG AAC TTC AAG ATG CTG GGG AAT ATC ATT GTG ATC TGC CTG GCT GAG CAC TTT GGC
AAG GAT TTT ACT CCT GAA TGT CAG GTT GCT TGG CAG AAG CTC GTG GCT GGA GTT GCC CAT
GCC CTG GCC CAC AAG TAC CAC

441 GLOBIN NCAD

*** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
*** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
*** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
GTC AGA GCC CAT GGC GCT AAG GTG CTG GTC TCC TTT GGT GAT GCT GTC AAG AAC CTG GAC
AAC CTG AAG GGT ACC TTT GCC AAA CTG AGT GAG CTC CAC TGT GAC AAG CTG CAC GAG GAC
CCT GAG AAC TTC AAG CTC CTG GGC AAC ATC CTG GTG ATC TGC CTG GCT GAG CAT TTT GGC
AAA GAA TTC ACC CCT GAG GTT CAG GCT GCC ACC CAG AAG ACT GTG GCT GGT GTG GCC AAC
GCT CTG GCC CAC AAG TAC CAC

441 GLOBIN MOEM

*** GTT AAC TTT ACT GCT GAG GAA AAA ACC CTC ATC AAT GGC CTG TGG AGT AAG GTC AAT
GTT GAA GAG GTT GGT GGT GAA GCC TTG GGA AGG CTT CTT GTT GTG TAC CCA TGG ACC CAG
AGA TTC TTT GAC AGC TTT GGG AAC TTG TCC TCT GCC TCT GCC ATA ATG GGC AAC CCA AGG
GTC AAA GCC CAT GGC AAG AAG GTG CTG ACT GCT TTT GGA TTG GAG TCC ATT AAC ATG GAC
AAC CTC AAG GAG ACC TTT GCT CAT CTC AGT GAG CTG CAC TGT GAC AAG CTT CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGC AAC ATG TTG GTG ATT GTC CTT TCT ACT CAT TTT GCC
AAG GAA TTC ACC CCA GAG GTG CAG GCT GCC TGG CAG AAG CTG GTG ATT GGA GTG GCC AAT
GCT CTG TCC CAC AAG TAC CAT

441 GLOBIN MOAD

*** GTG CAC CTG ACT GAT GCT GAG AAG GCT GCT GTC TCT TGC CTG TGG GGA AAG GTG AAC
TCC GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCT TGG ACC CAG
CGG TAC TTT GAT AGC TTT GGA GAC CTA TCC TCT GCC TCT GCT ATC ATG GGT AAT GCC AAA
GTG AAG GCC CAT GGC AAG AAG GTG ATA ACT GCC TTT AAC GAT GGC CTG AAT CAC TTG GAC
AGC CTC AAG GGC ACC TTT GCC AGC CTC AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAT ATG ATC GTG ATT GTG CTG GGC CAC CAC CTT GGC
AAG GAT TTC ACC CCC GCT GCA CAG GCT GCC TTC CAG AAG GTG GTG GCT GGA GTG GCC ACT
GCC TTG GCT CAC AAG TAC CAC

441 GLOBIN HUAD

*** GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG GGC AAG GTG AAC
GTG GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTG GTC TAC CCT TGG ACC CAG
AGG TTC TTT GAG TCC TTT GGG GAT CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG
GTG AAG GCT CAT GGC AAG AAA GTG CTC GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAC
AAC CTC AAG GGC ACC TTT GCC ACA CTG AGT GAG CTG CAC TGT GAC AAG CTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTC TGT GTG CTG GCC CAT CAC TTT GGC
AAA GAA TTC ACC CCA CCA GTG CAG GCT GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT
GCC CTG GCC CAC AAG TAT CAC

441 GLOBIN HUEM

*** GTG CAT TTT ACT GCT GAG GAG AAG GCT GCC GTC ACT AGC CTG TGG AGC AAG ATG AAT
GTG GAA GAG GCT GGA GGT GAA GCC TTG GGC AGA CTC CTC GTT GTT TAC CCC TGG ACC CAG
AGA TTT TTT GAC AGC TTT GGA AAC CTG TCG TCT CCC TCT GCC ATC CTG GGC AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAG GTG CTG ACT TCC TTT GGA GAT GCT ATT AAA AAC ATG GAC
AAC CTC AAG CCC GCC TTT GCT AAG CTG AGT GAG CTG CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGT AAC GTG ATG GTG ATT ATT CTG GCT ACT CAC TTT GGC
AAG GAG TTC ACC CCT GAA GTG CAG GCT GCC TGG CAG AAG CTG GTG TCT GCT GTC GCC ATT
GCC CTG GCC CAT AAG TAC CAC

441 GLOBIN HUGA

*** GGT CAT TTC ACA GAG GAG GAC AAG GCT ACT ATC ACA AGC CTG TGG GGC AAG GTG AAT
GTG GAA GAT GCT GGA GGA GAA ACC CTG GGA AGG CTC CTG GTT GTC TAC CCA TGG ACC CAG
AGG TTC TTT GAC AGC TTT GGC AAC CTG TCC TCT GCC TCT GCC ATC ATG GGC AAC CCC AAA
GTC AAG GCA CAT GGC AAG AAG GTG CTG ACT TCC TTG GGA GAT GCC ATA AAG CAC CTG GAT
GAT CTC AAG GGC ACC TTT GCC CAG CTG AGT GAA CTG CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGA AAT GTG CTG GTG ACC GTT TTG GCA ATC CAT TTC GGC
AAA GAA TTC ACC CCT GAG GTG CAG GCT TCC TGG CAG AAG ATG GTG ACT GCA GTG GCC AGT
GCC CTG TCC TCC AGA TAC CAC

441 GLOBIN CHAD

*** GTG CAC TGG ACT GCT GAG GAG AAG CAG CTC ATC ACC GGC CTC TGG GGC AAG GTC AAT
GTG GCC GAA TGT GGG GCC GAA GCC CTG GCC AGG CTG CTG ATC GTC TAC CCC TGG ACC CAG
AGG TTC TTT GCG TCC TTT GGG AAC CTC TCC AGC CCC ACT GCC ATC CTT GGC AAC CCC ATG
GTC CGC GCC CAC GGC AAG AAA GTG CTC ACC TCC TTT GGG GAT GCT GTG AAG AAC CTG GAC
AAC ATC AAG AAC ACC TTC TCC CAA CTG TCC GAA CTG CAT TGT GAC AAG CTG CAT GTG GAC
CCC GAG AAC TTC AGG CTC CTG GGT GAC ATC CTC ATC ATT GTC CTG GCC GCC CAC TTC AGC
AAG GAC TTC ACT CCT GAA TGC CAG GCT GCC TGG CAG AAG CTG GTC CGC GTG GTG GCC CAT
GCC CTG GCT CGC AAG TAC CAC

441 GLOBIN MOEY

*** GTG AAC TTT ACT GCT GAG GAA AAA ACC CTC ATC AAT GGC CTG TGG AGT AAG GTC AAT
GTT GAA GAG GTT GGT GGT GAA GCC TTG GGA AGG CTT CTT GTT GTG TAC CCA TGG ACC CAG
AGA TTC TTT GAC AGC TTT GGG AAC TTG TCC TCT GCC TCT GCC ATA ATG GGC AAC CCA AGG
GTC AAA GCC CAT GGC AAG AAG GTG CTG ACT GCT TTT GGA GAG TCC ATT AAG AAC CTA GAC
AAC CTC AAG TCT GCC TTG GCC AAG CTC AGT GAA CTG CAC TGT GAC AAG CTA CAT GTG GAT
CCT GAG AAC TTC AAA CTC TTG GGT AAT GTG CTG GTG ATT GTT TTG GCT AGT CAC TTC GGC
AAT GAA TTC ACA GCT GAG ATG CAG GCT GCC TGG CAG AAG CTG GTG GCT GGG GTG GCC ACT
GCG CTG TCC CAC AAG TAC CAC

Fig. A4

DNA sequence of coding regions of eutherian and marsupial β-like globin genes used in the relative rate tests presented in Section 4.7. The format of each sequence is that used in a computer program for estimating divergence values (kindly provided by Dr.'s W-H. Li, C-I. Wu and C-C. Luo). The abbreviation and source of each sequence are: SMIN:- S.c-β (present study; see Fig. 4.1.16); OPEM:- *D.virginiana* ϵ^m-globin gene (Koop and Goodman, 1988); MOEM:- mouse ϵy2-globin gene (Hansen *et al.*, 1982); HUEM:- human ϵ-globin (Baralle *et al.*, 1980); HUGA:- human γ-globin gene (Slightom *et al.*, 1980); GOEM:- goat ϵ^1-globin gene (Shapiro *et al.*, 1983); RAEM:- rabbit β4-globin gene (Hardison, 1983). The source and sequence of the *D.virginiana* β^m-, mouse β- (maj) and human β-globin genes, which were also used in relative rate tests are given in Fig. A3.

441 GLOBIN SMIN

*** GTG CAC TTC ACA GCT GAG GAG AAG AAC GCC ATC ACT ACC ATT TGG GGC AAG GTC AAT
GTG GAA GAG ACT GGT GGG GAA GCT TTA GGC AGA CTG TTG GTT GTC TAC CCC TGG ACC CAG
AGG TTC TTT GAT TCC TTT GGC AAC CTC TCT TCT GCC TCT GCT ATC CTG GGA AAC CCC AAG
GTT AAG GCC CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAT GCT GTC AAG AAC CTG GAC
AAC CTC AAG GGT ACC TTC TCC AAG CTG AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAC
CCT GAG AAC TTC CGG CTC CTT GGG AAT GTT CTG GTT ATT GTT ATG GCT GCT CAT TTC AAC
AAG GAA TTC ACT CCT GAA GTG CAG GCT GCT TTT CAG AAG CTG GTG ACT GGT GTG GCC AAT
GCT TTG GCT CAC AAG TAC CAC

441 GLOBIN OPEM

*** GTG CAT TTC ACC CCT GAG GAC AAG ACC AAT ATC ACT TCT GTG TGG ACC AAG GTC GAT
GTG GAA GAC GTT GGT GGG GAA TCT TTG GCC AGG CTG CTG GTA GTC TAT CCC TGG ACT CAG
AGG TTC TTT GAT TCC TTT GGC AAC CTA TCT TCT GCA TCT GCT GTT ATG GGG AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAA GGT GTC AAG AAC ATG GAC
AAC CTC AAG GGT ACC TTT GCC AAG CTG AGT GAG CTG CAC TGT GAC AAA CTG CAT GTA GAC
CCT GAG AAC TTC AGG CTC CTC GGG AAT GTG CTG ATT ATT GTT TTG GCT TCT CGT TTT GGC
AAG GAG TTC ACT CCA GAA GTG CAG GCT TCT TGG CAG AAG CTG GTG TCT GGT GTG TCC AGT
GCC TTG GGC CAC AAG TAC CAC

441 GLOBIN MOEM

*** GTG AAC TTT ACT GCT GAG GAA AAA ACC CTC ATC AAT GGC CTG TGG AGT AAG GTC AAT
GTT GAA GAG GTT GGT GGT GAA GCC TTG GGA AGG CTT CTT GTT GTG TAC CCA TGG ACC CAG
AGA TTC TTT GAC AGC TTT GGG AAC TTG TCC TCT GCC TCT GCC ATA ATG GGC AAC CCA AGG
GTC AAA GCC CAT GGC AAG AAG GTG CTG ACT GCT TTT GGA GAG TCC ATT AAG AAC CTA GAC
AAC CTC AAG TCT GCC TTG GCC AAG CTC AGT GAA CTG CAC TGT GAC AAG CTA CAT GTG GAT
CCT GAG AAC TTC AAA CTC TTG GGT AAT GTG CTG GTG ATT GTT TTG GCT AGT CAC TTC GGC
AAT GAA TTC ACA GCT GAG ATG CAG GCT GCC TGG CAG AAG CTG GTG GCT GGG GTG GCC ACT
GCG CTG TCC CAC AAG TAC CAC

441 GLOBIN HUEM

*** GTG CAT TTT ACT GCT GAG GAG AAG GCT GCC GTC ACT AGC CTG TGG AGC AAG ATG AAT
GTG GAA GAG GCT GGA GGT GAA GCC TTG GGC AGA CTC CTC GTT GTT TAC CCC TGG ACC CAG
AGA TTT TTT GAC AGC TTT GGA AAC CTG TCG TCT CCC TCT GCC ATC CTG GGC AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAG GTG CTG ACT TCC TTT GGA GAT GCT ATT AAA AAC ATG GAC
AAC CTC AAG CCC GCC TTT GCT AAG CTG AGT GAG CTG CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGT AAC GTG ATG GTG ATT ATT CTG GCT ACT CAC TTT GGC
AAG GAG TTC ACC CCT GAA GTG CAG GCT GCC TGG CAG AAG CTG GTG TCT GCT GTC GCC ATT
GCC CTG GCC CAT AAG TAC CAC

441 GLOBIN HUGA

*** GGT CAT TTC ACA GAG GAG GAC AAG GCT ACT ATC ACA AGC CTG TGG GGC AAG GTG AAT
GTG GAA GAT GCT GGA GGA GAA ACC CTG GGA AGG CTC CTG GTT GTC TAC CCA TGG ACC CAG
AGG TTC TTT GAC AGC TTT GGC AAC CTG TCC TCT GCC TCT GCC ATC ATG GGC AAC CCC AAA
GTC AAG GCA CAT GGC AAG AAG GTG CTG ACT TCC TTG GGA GAT GCC ATA AAG CAC CTG GAT
GAT CTC AAG GGC ACC TTT GCC CAG CTG AGT GAA CTG CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGA AAT GTG CTG GTG ACC GTT TTG GCA ATC CAT TTC GGC
AAA GAA TTC ACC CCT GAG GTG CAG GCT TCC TGG CAG AAG ATG GTG ACT GCA GTG GCC AGT
GCC CTG TCC TCC AGA TAC CAC

441 GLOBIN GOEM

*** GTG CAT TTT ACT GCC GAG GAG AAG GCT GCT ATC ACT GGC CTG TGG GGC AAA GTC AAT
GTG GAA GAG GCT GGA GGC GAG GCT CTG GGC AGG CTC CTG GTT GTC TAC CCC TGG ACC CAG
AGG TTC TTT GAT AGC TTT GGC AAC CTG TCC TCT GCC TCT GCC ATA ATG GGA AAC CCC AAG
GTC AAG GCC CAC GGC AAG AAG GTG CTG ACC TCC TTT GGA GAA GCT ATT AAG AAT TTG GAC
AAC CTC AAA GGT GCC TTC GCT AAG CTG AGT GAG CTG CAC TGT GAC AAG TTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAT GTG ATT GTG ATT ATT CTG GCT ACT CAT TTT GGC
AGA GAA TTC ACC CCT GAC GTG CAG GCT GCC TGG CAG AAG CTG GTG TCT GGT GTT GCC ACT
GCT CTG GCC CAC AAG TAC CAC

441 GLOBIN RAEM

*** GTG CAC TTT ACT CCT GAG GAA AAG TGC ATT ATC AGT AAG CAG TGG GGC CAG ATG AAC
ATC GAT GAG ACT GGA GGT GAA GCT TTG GGC AGG CTC CTT GTT GTC TAC CCC TGG ACC CAA
AGA TTC TTT GAC AAC TTT GGC AAC TTG TCC TCT TCC TCT GCC ATA CTG GGA AAC CCC AAG
GTC AAG GCA CAT GGC AAG AAG GTG CTG AAC TCC TTT GGA GAT GCC ATC AAG AAC ATG GAC
AAC CTT AAG GGT GCC TTT GCT AAG CTG AGT GAG CTG CAC TGT GAC AAA CTG CAT GTG GAT
CCT GAG AAC TTC AAA CTC CTG GGA AAC GTG CTG CTG ATC GTT CTG GCC ACT CAT TTT GGC
AAA GAA TTC ACT CCG GAG GTG CAG GCT GCT TTG CAG AAG CTC GTG TCT GGT GTT GCC ATT
GCT CTG GCC CAC AAG TAT CAC

TABLE A1. BamHI and EcoRI fragment variants at β-like globin loci of *S.crassicaudata* family members[1].

Family	N	Animal	S.c-β locus		β-globin locus	ϵ_2-globin locus
			BamHI	EcoRI	EcoRI	EcoRI
M1355	5	1076.1a	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.2 E2-3.4	E3-6.8 E3-6.8
		1040.1f	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1355.1a	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1355.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1355.1c	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1355.1d	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1355.1e	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
M1940		1520.1b	B1-8.9 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-7.6 E3-6.8
		1443.2e	B1-4.5 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
M1940.1	8	1940.1a	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1940.1b	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1940.1c	B1-8.9 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-7.6 E3-6.8
		1940.1d	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1940.1f	B1-8.9 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-6.8 E3-6.8
		1940.1g	B1-8.9 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-6.8 E3-6.8
		1940.1h	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1940.1i	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-7.6
M1940.2	9	1940.2a	B1-8.9 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-7.6 E3-7.6

Family	N	Animal	S.c-β locus		β-globin locus	ϵ_2-globin locus
			BamHI	EcoRI	EcoRI	EcoRI
		1940.2b	B1-8.9 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-6.8 E3-6.8
		1940.2c	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1940.2d	B1-8.9 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-6.8 E3-6.8
		1940.2e	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-7.6 E3-7.6
		1940.2f	B1-8.9 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-7.6 E3-6.8
		1940.2g	B1-8.9 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.4	E3-7.6 E3-7.6
		1940.2h	-	-	-	-
		1940.2i	B1-8.9 B1-4.5	-	-	-
M1382	5	1220.1c	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1073.1f	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
		1382.1a	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-3.4	E3-7.6 E3-6.8
		1382.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1382.1c	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
		1382.1d	B1-8.9 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1382.1e	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-7.6 E3-6.8
M1746	7	1494.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
		1471.1d	B1-8.9 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-7.6 E3-6.8
		1746.1a	B1-8.9 B1-4.5	-	-	-
		1746.1b	B1-8.9 B1-4.5	-	-	-

Family	N	Animal	S.c-β locus		β-globin locus	ε$_2$-globin locus
			BamHI	EcoRI	EcoRI	EcoRI
		1746.1c	B1-8.9 B1-4.5	-	-	-
		1746.1d	B1-8.9 B1-4.5	-	-	-
		1746.1e	B1-8.9 B1-4.5	-	-	-
		1746.1g	B1-8.9 B1-4.5	-	-	-
		1746.1h	B1-4.5 B1-6.5	-	-	-
M1593	6	1316.1c	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-7.6$_*$
		1137.1f	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1593.1a	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-7.6 E3-6.8*
		1593.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1593.1c	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-7.6 E3-6.8*
		1593.1d	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-7.6 E3-6.8
		1593.1e	B1-4.5 B1-4.5	-	-	-
		1593.1f	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
M1361	8	953.1c	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-6.8 E3-6.8
		1073.1g	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	-
		1361.1a	B1-4.5 B1-6.5	-	-	-
		1361.1b	B1-8.9 B1-4.5	-	-	-
		1361.1c	B1-4.5 B1-6.5	-	-	-
		1361.1d	B1-4.5 B1-4.5	-	-	-

Family	N	Animal	S.c-β locus		β-globin locus	ε$_2$-globin locus
			BamHI	EcoRI	EcoRI	EcoRI
		1361.1e	B1-4.5 B1-6.5	E1-4.4 E1-4.4	-	-
		1361.1f	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-6.8 E3-6.8
		1361.1g	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-6.8 E3-6.8
M1625	8	1157.1a	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1211.1d	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
		1625.1a	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8
		1625.1b	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1625.1c	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1625.1d	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-6.8 E3-6.8
		1625.1e	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1625.1f	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-6.8 E3-6.8
		1625.1g	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1625.1h	B1-8.9 B1-4.5	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
M1877	7	1492.1b	B1-8.9 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-7.6 E3-6.8
		1508.1e	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1877.1a	-	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-7.6
		1877.1b	-	E1-4.4 E1-4.4	E2-3.2 E2-3.4	E3-7.6 E3-7.6
		1877.1c	-	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1877.1d	-	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8

Family	N	Animal	S.c-β locus		β-globin locus	ε,-globin locus
			BamHI	EcoRI	EcoRI	EcoRI
		1877.1e	-	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1877.1f	-	E1-4.4 E1-4.4	E2-3.2 E2-3.4	E3-7.6 E3-6.8
		1877.1g	-	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-6.8 E3-6.8
M1920	6	1542.1b	B1-6.5 B1-4.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-6.8 E3-6.8
		1527.1c	B1-4.5 B1-8.9	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1920.1a	B1-8.9 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-6.8 E3-6.8
		1920.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-6.8 E3-6.8
		1920.1c	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1920.1d	B1-8.9 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-3.4	E3-6.8 E3-6.8
		1920.1f	B1-4.5 B1-6.5	E1-4.4 E1-4.4	E2-3.4 E2-4.0	E3-7.6 E3-6.8
		1920.1g	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-4.0 E2-4.0	E3-6.8 E3-6.8
M1939	4	1650.1a	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1571.1f	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1939.1a	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-3.2	E3-6.8 E3-6.8
		1939.1b	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-4.0	E3-7.6 E3-6.8
		1939.1c	B1-4.5 B1-4.5	E1-4.4 E1-4.4	E2-3.2 E2-3.2	E3-7.6 E3-6.8
		1939.1d	B1-4.5 B1-8.9	E1-4.4 E1-11.4	E2-3.2 E2-4.0	E3-6.8 E3-6.8

1. Details of BamHI and EcoRI variants are given in Sections 4.6.3 and 4.6.4 respectively. N designates the number of offspring in each family.

TABLE A2. Genotypes of parents and the proportion of recombinant and non-recombinant offspring in *S.crassicaudata* families, for pairs of β-like globin loci[1].

S.c-β (ε) and ϵ_2 loci.

Family	Parent 1	Parent 2	R	NR
M1382	1073.1f	1220.1c		
	B1-4.5 E3-6.8 / B1-4.5 E3-6.8	B1-8.9 E3-7.6 / B1-4.5 E3-6.8	2	3
		B1-8.9 E3-6.8 / B1-4.5 E3-7.6	3	2
M1625	1211.1d	1157.1a		
	B1-4.5 E3-6.8 / B1-8.9 E3-6.8	B1-4.5 E3-7.6 / B1-6.5 E3-6.8	5	3
		B1-4.5 E3-6.8 / B1-6.5 E3-7.6	3	5
M1920	1542.1b	1527.1c		
	B1-6.5 E3-6.8 / B1-4.5 E3-6.8	B1-4.5 E3-7.6 / B1-8.9 E3-6.8	4	2
		B1-4.5 E3-6.8 / B1-8.9 E3-7.6	2	4
M1939	1650.1a	1571.1f		
	B1-4.5 E3-6.8 / B1-4.5 E3-6.8	B1-4.5 E3-6.8 / B1-8.9 E3-7.6	3	1
		B1-4.5 E3-7.6 / B1-8.9 E3-6.8	1	3

β and ϵ_2 loci.

Family	Parent 1	Parent 2	R	NR
M1382	1073.1f	1220.1c		
	E2-3.2 E3-6.8 / E2-4.0 E3-6.8	E2-3.4 E3-7.6 / E2-4.0 E3-6.8	2	3
		E2-3.4 E3-6.8 / E2-4.0 E3-7.6	3	2
M1625	1211.1d	1157.1a		
	E2-3.2 E3-6.8 / E2-4.0 E3-6.8	E2-3.4 E3-7.6 / E2-4.0 E3-6.8	3	5
		E2-3.4 E3-6.8 / E2-4.0 E3-7.6	5	3

TABLE A2 (continued)

S.c-β (ϵ) and β loci.

Family	Parent 1	Parent 2	R	NR
M1382	1073.1f	1220.1c		
	B1-4.5 E2-3.2 B1-4.5 E2-4.0	B1-8.9 E2-3.4 B1-4.5 E2-4.0	0	5
		B1-8.9 E2-4.0 B1-4.5 E2-3.4	5	0
M1940	1520.1b	1443.2e		
	B1-8.9 E2-3.2 B1-8.9 E2-3.4	B1-4.5 E2-3.4 B1-8.9 E2-4.0	15	0
		B1-4.5 E2-4.0 B1-8.9 E2-3.4	0	15
M1355	1040.1f	1076.1a		
	B1-4.5 E2-3.2 B1-8.9 E2-3.2	B1-4.5 E2-3.2 B1-6.5 E2-3.4	0	5
		B1-4.5 E2-3.4 B1-6.5 E2-3.2	5	0

1. The above tables give information on only those families for which linkage data could be obtained (see Table 4.6.2). The proportion of recombinant (R) and non-recombinant (NR) offspring in each family was determined using the information given in Table A1. Information on the restriction fragment variants B1 at the S.c-β (ϵ locus), E2 at the adult β-globin locus, and E3 at the ϵ_2-globin locus are given in Sections 4.6.3 and 4.6.4.

TABLE A3

One letter and three letter codes for each amino acid (Dayhoff, 1975).

amino acid		
alanine:	Ala	A
arginine:	Arg	R
asparagine:	Asn	N
aspartic acid:	Asp	D
cysteine:	Cys	C
glutamine:	Gln	Q
glutamic acid:	Glu	E
glycine:	Gly	G
histidine:	His	H
isoleucine:	Ile	I
leucine:	Leu	L
lysine:	Lys	K
methionine:	Met	M
phenylalanine:	Phe	F
proline:	Pro	P
serine:	Ser	S
threonine:	Thr	T
tryptophan:	Trp	W
tyrosine:	Tyr	Y
valine:	Val	V