COMPARATIVE RESTRICTION ENDONUCLEASE MAPPING OF MACAQUE ADULT NON-ALPHA GLOBIN LOCI

by

De-Ann Margaret Pillers, M.D.

A DISSERTATION

Presented to the Department of Medical Genetics
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

May 1986

APPROVED:

.... [redacted] ....

(Professor in Charge of Thesis)

.... [redacted] ....

(Chairman, Graduate Council)

DEDICATION

To Dr. John D. Hill and the Staff of the Department of Surgery
at the Oregon Regional Primate Research Center
in the Summer of 1975

and

TO THE MEMORY OF

Dr. Edward S. West

For generously sharing their knowledge and interest in the field of scientific research with me through the Summer Scholar program.

TABLE OF CONTENTS

List of Abbreviations viii

List of Tables x

List of Figures xi

Acknowledgements xiii

Abstract xvi

CHAPTER ONE: INTRODUCTION

1.A. Hemoglobin as a Model Evolutionary System 1

1.A.1. Hemoglobin Function 2

1.A.2. Hemoglobin Structure 4

1.A.3. Ontogeny 5

1.A.4. Hemoglobin Electrophoresis 7

1.A.5. Amino Acid Analysis 8

1.A.6. X-Ray Crystallography 9

1.B. Molecular Genetics of Hemoglobin 10

1.B.1. Gene Expression and Protein Synthesis 10

1.B.2. Globin Gene Structure and Organization 13

1.B.3. Globin Pseudogenes 19

1.B.4. Globin Intergenic DNA 23

1.B.4.a. Globin Gene Duplication Units 24

1.B.4.b. Repetitive DNA 25

1.B.4.c. Simple Sequence DNA 29

1.B.5. Restriction Fragment Length Polymorphisms of the Globin Gene Cluster ... 29
1.C. Molecular Phylogeny of Globins ... 32
1.D. Evolution of the Beta Globin Family ... 35
1.D.1. General Considerations ... 35
1.D.2. Specific Examples of Beta-Globin Gene Clusters ... 40
1.D.2.a. Toad ... 41
1.D.2.b. Chicken and Duck ... 41
1.D.2.c. Mouse ... 41
1.D.2.d. Rabbit ... 44
1.D.2.e. Goat ... 45
1.D.2.f. Cow ... 46
1.D.2.g. Lemur ... 47
1.D.2.h. Higher Primates ... 47
1.E. Other General Evolutionary Considerations ... 49
1.E.1. Selectionist versus Neutral Mutation Theories ... 49
1.E.2. Other Methods Used in Molecular Evolution Studies ... 50
1.E.2.a. DNA-DNA Hybridization ... 50
1.E.2.b. Mitochondrial DNA Analysis ... 51
1.E.2.c. Chromosome Evolution ... 53
1.F. Macaque Evolution ... 53
1.G. Outline of Proposal ... 56

CHAPTER TWO: MATERIALS AND METHODS

2.A. Collection and Preparation of Peripheral Blood Samples 59

2.A.1. Selection of Subjects 59

2.A.2. Sample Collection 59

2.A.3. Sample Preparation 60

2.B. Protein Analysis 61

2.B.1. Hemolysate Preparation 61

2.B.2. Cellulose-Acetate Electrophoresis 61

2.B.3. High-Performance Liquid Chromatography Globin Chain Analysis 63

2.C. DNA Analysis 64

2.C.1. High Molecular Weight Genomic DNA Isolation 64

2.C.2. Quantitation of DNA 67

2.D. Probe Preparation 68

2.D.1. Source 68

2.D.2. Mini-Plasmid Preps 69

2.D.3. Plasmid Verification 72

2.D.3.a. Antibiotic Resistance Verification 72

2.D.3.b. Restriction Endonuclease Mapping 73

2.D.4. Large-Scale Plasmid Preparation 73

2.D.4.a. Culture 73

2.D.4.b. DNA Isolation 74

2.D.4.c. Cesium-Chloride Gradient Purification 77

2.E. Southern Blot Analysis 78

2.E.1. Restriction Endonuclease Digestion of Genomic DNA ... 78

2.E.2. Agarose Gel Electrophoresis 80

2.E.3. Southern Transfers 81

2.E.4. Radioactive Labeling of Probe 82

2.E.5. Blot Hybridization 83

2.E.6. Autoradiography 86

2.F. Data Analysis 86

2.F.1. Hemoglobin Electrophoresis 86

2.F.2. HPLC Globin Chain Separation 87

2.F.3. Restriction Fragment Length Determination 87

2.F.4. Restriction Fragment Map Construction 87

CHAPTER THREE: RESULTS

3.A. Protein Analysis 89

3.A.1. Cellulose-Acetate Electrophoresis 89

3.A.2. High-Performance Liquid Chromatography 92

3.B. Probe Preparation 93

3.C. Restriction Endonuclease Maps of the Macaque Adult Non-Alpha Loci 94

CHAPTER FOUR: DISCUSSION

4.A. Protein Data 97

4.B. DNA Data 101

4.B.1. Overall Similarities 101

4.B.2. Differences are Detected across Species Boundaries .. 102

4.B.2.a. Macaca fuscata 102

4.B.2.b. Macaca mulatta 103

4.B.2.c. Macaca nemestrina 104

4.C. Implications 105

4.D. Applications of Restriction Endonuclease Analysis 106

CHAPTER FIVE: SUMMARY AND CONCLUSIONS 108

REFERENCES 111

Tables 142

Figures 150

Appendix A: Animal Demographics 190

Appendix B: Preparation of RNase-Free Dialysis Tubing 193

List of Abbreviations

Abbreviation	Definition
α	alpha
β	beta
δ	delta
ϵ	epsilon
η	eta
γ	gamma
ζ	zeta
ψ	pseudo
A	adenine
C	cytosine
G	guanine
T	thymidine
U	uridine
C	degree Celsius
g	gauge
kb	kilo-base
kbp	kilo-base-pairs
μg	microgram
min	minute
MYA	million years ago
MYr	million years

nm	nanometer
O.D.	optical density
rpm	revolutions per minute
dH_2O	deionized water
DNA	deoxyribonucleic acid
Hb	hemoglobin
HMW	high molecular weight
HPLC	high performance liquid chromatography
IVS	intervening sequence
RFLP	restriction fragment length polymorphism
mRNA	messenger RNA
RNA	ribonucleic acid
SSC	salt-sodium-citrate
U	unit
ORPRC	Oregon Regional Primate Research Center

List of Tables

Table		Page
1.1	Developmental Expression of Hemoglobin in Man	142
1.2	Haplotype Analysis of Sickle Cell Variants	143
2.1	Primate Species Studied	144
2.2	Restriction Endonucleases	145
2.3	Agarose Gel Electrophoresis Patterns	146
3.1	Relative Mobility of Primate Major Adult Hemoglobins	147
3.2	Restriction Fragment Sizes Used to Map Macaque Beta and Delta Globin Loci	148

List of Figures

Figure		Page
1.1	Comparison of Human Beta and Delta Globin Amino Acid Sequences	150
1.2	Codon Usage Charts for Human Alpha and Beta Globins	152
1.3	Globin Gene Organization in the Human Alpha and Beta Globin Clusters	154
1.4	Fine-Structure Mapping of Human Alpha and Beta Globin Genes	156
1.5	Organization of Repetitive DNA in Human Globin Clusters	158
1.6	Restriction Fragment Polymorphisms in the Human Beta Globin Cluster	160
1.7	Non-Alpha Globin Gene Cluster Organization in Various Species	162
1.8	Primate Classifications	164
1.9	Phenogram of Macaque Species	166
3.1	Cellulose-Acetate Electrophoresis of Primate Hemoglobins	168
3.2	HPLC Globin Chain Separation	170
3.3	Genomic Beta Globin Probe pβPstd	172
3.4	Agarose Gel Electrophoresis of Primate DNA	174
3.5	Autoradiography of _Macaca fuscata_ DNA Probed with δ-Probe	176
3.6	Restriction Map and Haplotype Analysis for Adult Non-Alpha Globin Loci of _Macaca fuscata_	178

3.7 Restriction Map of Adult Non-Alpha Loci of *Macaca mulatta* 180

3.8 Restriction Map of Adult Non-Alpha Loci of *Macaca nemestrina* 182

4.1 Common Macaque Restriction Haplotype of the Adult Non-Alpha Globin Loci 184

4.2. Geographical Distribution of Macaque Species Studied: *fascicularis* Group 186

4.3. Geographical Distribution of Macaque Species Studied: *silenus*-*sylvanus* Group 188

ACKNOWLEDGEMENTS

I would especially like to thank Dr. Peter Russell of Reed College for allowing me to find a place to learn molecular techniques when opportunities at the OHSU were limited. Special thanks also go to Sheryl Wagner who taught me at various stages along the way. For companionship, I am grateful to Sheryl, Joanne Pratt, Jennie Russell, and the many Reed College students with whom I have bumped shoulders over the past three years.

I am grateful to Dr. John Hill, Dr. Charles Howard, and Tonya Van Buren at the ORPRC for their assistance in procuring macaque blood samples.

I would like to thank Dr. Mitchell Golbus and his staff at UCSF: George Fry, Monty Montano, and Leigh Nicholaison, for their frequent assistance and helpful advice. I am indebted to them for sharing many of their procedures and protocols.

I am especially grateful to Dr. Morris Goodman for advice and timely expression of support for my work.

Thanks also go to my thesis committee members for their guidance: Dr. Charles Howard of the ORPRC for his continued enthusiasm for my project, Dr. Richard Jones for providing protein expertise and laboratory opportunities, Dr. Michael Litt for anticipating certain equipment needs, Dr. Karin Rodland for always providing expert advice and a sympathetic ear, and Dr. Peter Russell for teaching me the value of a good reputation within the scientific community. Special thanks go to Dr. William Woodward who has been in turn employer, mentor, and

friend: for his help in computer applications.

I am grateful to Dr. Bruce Magun of the Department of Cell Biology and Anatomy and Dr. Karin Rodland for their assistance in sharing equipment and radioactive supplies.

I would like to thank Marilyn Jones for her expert instruction in hemoglobin electrophoresis. In the laboratory of Dr. Richard Jones, thanks go to Charlotte Head, Mike Kavanaugh, and Daniel and Marjorie Shih for many months of instruction in the vagaries of protein analysis.

I would like to thank my advisor, Dr. Robert Koler, for providing me with space and funding to complete my project and for many hours of reading and conference.

I have enjoyed the stimulating conversation and friendship of Marilyn Aberle and Marilyn Jones.

I would like to thank the other graduate students, past and present, for their comaraderie. I am indebted to Steven Lai for his assistance in typing this manuscript. I would especially like to thank Shantharaj Gowda for a special friendship.

Finally, my deepest appreciation goes to my husband and my family for their strength and encouragement throughout my graduate school experience. My husband, Robert Nourse, has spent many late hours throughout the years and more recently helping with the production of the macaque maps via Autocad™. My mother, June Pillers, and my brother Larry for their companionship and assistance in an isolated laboratory during evenings, late nights and weekends. My father, Lauritz Pillers, my brother Mark, my mother, and my husband for help in typing this manuscript. Finally, to Chi-Chi and Monty for providing comic relief.

ABSTRACT

Five species of the Old World monkey genus macaque were studied: *Macaca fascicularis*, *fuscata*, *mulatta*, *nemestrina*, and *nigra*. Preliminary screening by hemoglobin electrophoresis was used to detect those animals representative of each species at the phenotypic level. The adult non-alpha globin genes of three species were then examined: *Macaca fuscata*, *nemestrina*, and *nigra*. Restriction endonuclease mapping of the beta and delta loci was performed. A common haplotype background was determined which appears to be characteristic of the genus *Macaca*. Polymorphic sites were also detected which seem to be distinctive of a particular macaque species. An 0.8 kb deletion between the delta and beta loci of *Macaca nemestrina* was detected in one member of that species. Implications of these data regarding species variation are discussed. Finally, applications of this information to efforts such as breeding programs or distinguishing among ambiguous identities of species are presented.

CHAPTER ONE: INTRODUCTION

1.A. Hemoglobin As A Model Evolutionary System

Hemoglobin is one of the most thoroughly explored molecular systems. Without doubt, this is due to the ready accessibility of blood to the investigator, but there may be other reasons as well. Long before its function was actually known, blood was a source of great fascination to man. The ancients believed it to be one of the four "vital humours" upon which personality or temperament depended (Dickerson and Geis, 1983). Early theories also proposed it as a medium of heredity and some are still propagated today as one hears of the "bloodlines" of a prized horse (Dickerson and Geis, 1983). Loftier sentiments of nobility have also been conveyed, e.g. blueblood. No matter what the driving force for study has been however, blood has been central to many of man's greatest scientific achievements. The field of molecular evolution provides no exception.

Hemoglobin is a major blood protein which is located within red blood cells and gives them their characteristic color[1]. It was the first protein to have a defined physiological purpose. Little over a

---

[1] The term "hemoglobin" was coined by Felix Hoppe in 1862 to describe the protein as a pigment (Hoppe, 1862).

century ago, in 1864, the English mathematician and physicist George Stokes demonstrated that the color changes in the blood were associated with the reversible binding and release of oxygen[2]. Surprisingly something as simple as blood color may help to define an evolutionary system. Hemocyanins, for example, are large, copper containing proteins which circulate in the extracellular hemolymph of certain anthropods and molluscs; they have a rich blue color (Bunn and Forget, 1986) thus delineating one of a family of gene products that provide a solution to the common problem of oxygen-transport.

1.A.1. Hemoglobin Function

Hemoglobin is a widely represented member of a group of gene products that accomplish oxygen transport. It is the respiratory pigment employed by all mammals (Bunn and Forget, 1986).

As detailed by Barcroft in 1928, there are four basic physiological duties of a respiratory pigment. They include: 1) The ability to transport large amounts of oxygen while, 2) maintaining its solubility; 3) appropriate uptake and release of oxygen at the tissue level; and 4) the ability to contribute to blood buffering via a bicarbonate system.

---

[2] In the late 1600's, the English scientist Richard Lower established that the pulmonary circulation acted to aerate the blood and anticipated that its color changes were related to this event (Lower,1932).

Hemoglobin provides an exceptionally successful system of oxygen transport. During its transit through the lung, the pigment collects oxygen for subsequent delivery to peripheral tissues. Each molecule of hemoglobin contains four heme molecules, and each heme molecule is capable of binding a single oxygen molecule. The binding of oxygen to each heme site does not occur independently, rather the binding of any one oxygen molecule is influenced by the binding of the others. This phenomenon is referred to as "cooperativity". Each subsequent oxygen molecule binds with greater affinity. A change in conformation also occurs as the deoxyhemoglobin state converts to oxyhemoglobin.

The buffering capacity of the blood is tied to the proton affinity of deoxyhemoglobin. Under acidic conditions equilibrium favors deoxyhemoglobin. Thus in low pH environments, such as actively working muscles, oxygen dissociation is enhanced. This phenomenon is referred to as the Bohr effect. Other substances can also encourage a shift in the oxygen-equilibrium curve, e.g. carbon dioxide, chloride ions and diphosphoglycerate (2,3-DPG).

Diphosphoglycerate acts as a long term moderator of oxygen affinity. Although all mammals use 2,3 DPG in this regulatory role, other animal groups have used different organophosphates. Teleosts, or bony fishes, predominantly use ATP and GTP. Amphibians and reptiles use ATP and 2,3-DPG, while birds use the inositol organophosphates (Bunn and Forget, 1986).

1.A.2. Hemoglobin Structure

Hemoglobin is composed of two different components: heme molecules and globin chains. The heme molecule is a functional iron-containing prosthetic group essential to the binding of oxygen. The globins are proteins of which there are two genetically coded groups: alpha-like, and non-alpha, or beta-like globins. Each globin chain has a molecular weight of approximatly 16,000 (Dickerson and Geis, 1983). The total molecular weight of the tetramer, including heme groups, is about 64,450.

Each individual hemoglobin molecule is composed of two identical alpha-like globins and two identical beta-like globins. An example is the major human adult hemoglobin, HbA, which is commonly represented as $\alpha_2\beta_2$. There are various types of α- and β-globins which may occur in combination. These hemoglobin types occur at different stages during ontogeny & development (Table 1.1).

The alpha chain of human adult hemoglobin consists of 141 amino acids. The primary structure of the beta globin is similar, though slightly longer at 146 residues. The primary structure, i.e. amino acid sequence, in turn determines the secondary structure. In the case of the globins, as much as 75 % of the protein is found in the alpha-helix conformation (Bunn and Forget, 1986).

The secondary, tertiary, and quaternary structures determine many of

the functional properties of hemoglobin, e.g. the alkaline Bohr effect, cooperativity, and allosteric interactions.

1.A.3. Ontogeny

There are various globin chains which are produced at different stages of life. The earliest known forms are the embryonic hemoglobins of which there are several; hemoglobins Gower-1 and -2 and hemoglobins Portland I and II. These embryonic hemoglobins are produced during the third through eighth week of gestation and appear to be synthesized in the yolk sac (Bunn and Forget, 1986).

Four different globins are implicated in the formation of these embryonic hemoglobins; two alpha and two non-alpha, or beta-like chains. The alpha globin representatives include the adult alpha described above and a purely embryonic form, zeta. The beta globins are represented by a purely embryonic form epsilon (ϵ), and a predominantly fetal form, gamma (γ). The size and primary structures of these globins are similar to their adult counterparts.

The Gower hemoglobins 1 and 2, named after a street in London, were described by Huehns et al. in 1961. The subunit structure of these embryonic hemoglobins has been redefined following the later discovery of hemoglobin Portland by Jones et al. in 1967 (Capp, Rigas & Jones: 1967) and its subsequent characterization (Capp et al., 1967; 1970; Kamuzora et al., 1974). Hemoglobin Gower-1 was demonstrated to be

composed of entirely embryonic subunits ($\zeta_2\epsilon_2$), whereas Gower-2 contains alpha globin ($\alpha_2\epsilon_2$). The Portland hemoglobins consist of the embryonic alpha-like ζ and the fetal gamma globins.

An alternate form of hemoglobin is found during fetal life. This form was discovered as early as 1866 by Korber (Korber, 1926) who demonstrated the resistance of newborn red blood cells to alkaline denaturation. This fetal hemoglobin is designated HbF and is composed of two alpha- and two gamma-globin chains ($\alpha_2\gamma_2$). Two normal forms of gamma-globin chains have been described (Schroeder et.al., 1968). The two globins differ at position 136 where either a glycine ($^G\gamma$) or an alanine ($^A\gamma$) may be found. Seventy-five percent of the HbF during fetal life is composed of the $^G\gamma$ form. During the first ten months of postnatal life this ratio reverses to one in which the $^A\gamma$ globin predominates at a ratio of 60:40 (Schroeder et al., 1971).

There are three hemoglobin molecules identifiable in the human adult. Roughly 97% of the hemoglobin is the major adult hemoglobin described earlier. Another 0-1% is the fetal hemoglobin, HbF. Finally, 2-3% is the minor adult hemoglobin, HbA_2. This hemoglobin is composed of alpha chains and still another beta-like chain, delta (δ).

Hemoglobin A_2 was first described by Kunkel and Wallenius (1955) using the technique of starch block electrophoresis of human hemolysates. The primary sequence of the human delta chain has been determined and is compared with the human beta chain in Figure 1.1.

There is great structural homology between the two chains as delta varies from beta at only 10 of its 146 residues (Dickerson and Geis, 1983).

Though determination of primary and secondary structures of proteins, and in particular globins, has been widely studied and reported, the solution of tertiary and quaternary hemoglobin structures remains the province of only a handful of individuals. Chief among these is Dr. Max Perutz who has used the technique of x-ray crystallography. His work will be described in greater detail in section 1.A.5.

1.A.4. Hemoglobin Electrophoresis

Hemoglobin electrophoresis was first applied by Linus Pauling in 1949 and its efficacy demonstrated by the separation of the sickle hemoglobin, HbS, from adult hemoglobin, HbA. Pauling lauded electrophoresis as a technique which could provide clues to a molecule's, and thus an animal species', past.

A major advantage of electrophoresis is the simplicity of the procedure and the ability to obtain results rapidly. The disadvantage of the technique is that protein molecules are separated based on charge differences. Amino acid substitutions which do not involve changing the charge of the molecule are not likely to be detected.

Human populations have since been screened by hemoglobin electrophoresis and more than 400 mutants have been detected by this and related procedures (Wrightstone, 1986). Other members of the animal kingdom have also been screened using electrophoresis. The data available is therefore substantial. It is difficult, however, to use this technique in other than cursory comparisions as the difference (or lack thereof) in migration between two proteins may have little to do with the total number of amino acid substitutions.

1.A.5. Amino Acid Analysis

Determination of the primary amino acid sequence of proteins has provided a database for evolutionary comparisons. In the case of globins, clearly the most widely sequenced group of proteins, globin samples from such diverse groups as leguminous plants, insects, shellfish and all manner of vertebrates have been sequenced. Although cross-species comparisons have provided considerable evolutionary data, much information has also been garnered by intra-species comparisons of the various globin chains which are thought to reflect divergence from a common orthologous gene.

In general, the assumption has been made that the more similar two amino acid sequences are, the more closely related the corresponding animal species will be. Examining differences may also provide useful information regarding not only species phylogenies but molecular function as well. The prevalence of conservative as opposed to non-

conservative changes is especially useful in this regard.

1.A.6. X-Ray Crystallography

One of the most remarkable biochemical achievements of this century has been the elucidation of the three-dimensional structure of hemoglobin[3]. Although the primary and secondary structure of proteins have provided a wealth of evolutionary information, it is often through examining the tertiary and quaternary structures that keys to the function of molecules are found.

The three dimensional structure of both oxyhemoglobin and deoxyhemoglobin have been solved using the techniques of X-ray crystallography and electron microscopy. X-ray crystallography using Fourier analysis has reached a resolution of 2.0 to 2.5 Å and has detailed not only the globin sub-unit interactions with each other but also the region of the heme pocket and the sites of interaction with the various allosteric effectors (Perutz et al., 1960; Fermi et al., 1984).

X-ray crystallography was an early contributor to the evolutionary data base as the initial structures to be determined were those of the horse hemoglobins. All of the major human hemoglobins have had their tertiary structures determined. In addition, myoglobin structural data indicate that this molecule has descended from a progenitor common with

---

[3]Dr. Max Perutz of Cambridge University is largely responsible for this body of work and has been duly recognized with a Nobel Prize in Chemistry.

all globin genes (Dickerson and Geis, 1983).

Other respiratory proteins have also been analysed by these techniques. Hemerythrin is an intracellular respiratory protein found in four different phyla of marine invertebrates and is structurally distinct from hemoglobin (Bunn and Forget, 1986). Hemocyanins are cupro-proteins with a deep blue color and are found in some arthropods and molluscs. They are also structurally distinct from hemoglobin (Bunn and Forget, 1986). Erythrocruorins, on the other hand, are similar to hemoglobins in that they share a characteristic myoglobin fold. These proteins also share some primary sequence homology with the hemoglobins. Erythrocruorins are found among many varieties of invertebrates including protozoa, platyhelminthes (flatworms), nemertina, annelids (segmented worms), molluscs, arthropods, and echinoderms (Bunn and Forget, 1986).

1.B. Molecular Genetics of Hemoglobin

1.B.1. Gene Expression and Protein Synthesis

The hierarchy involved in the synthesis of hemoglobin can best be decribed by three terms: DNA, RNA, and protein.

As is the case for all proteins, the genetic information which codes for globin proteins is determined at the level of the nucleotide sequence of the DNA. That group of nucleotides which specifically codes

for the amino acid sequence of a protein is referred to as a gene locus. This information will be found on a particular chromosome. In the case of man the information for the alpha globins is encoded on chromosome 16 while that for the non-alpha globins is encoded on chromosome 11 (Maniatis et al., 1980).

Each gene locus also has several levels of organization. The portions of the genes that code for protein sequences are referred to as exons. Exons consist of triplets of nucleotides referred to as codons, each of which codes for a particular amino acid. Although each codon is specific for a single amino acid, the degeneracy of the genetic code is such that most of the amino acids can be specified by more than one codon. Figure 1.2 illustrates the messenger RNA codon usage charts for human alpha and beta globins.

By comparing the two entries in Figure 1.2, α- and β-globin, one can see that the same amino acid is not necessarily encoded by the same triplet codon in each gene locus. For example, phenylalanine in the α-globin is predominantly encoded for by the codon UUC whereas β-globin uses only UUU. In general however, these two proteins are very similar in their codon utilization. Such similarities or differences in codon usage may be of evolutionary significance.

Non-random codon usage is common among protein coding regions. Several factors may affect codon selection. A major consideration is the relative abundance of the tRNA's for each amino acid (Ikemura,

1985). Another consideration is that there are two levels at which codons may be chosen, the first occurring when the amino acids are selected, the second when the particular codon is determined (McLaclan et al., 1984).

There are several steps involved in the synthesis of messenger RNA from DNA. This process, known as transcription, has been thoroughly reviewed (Collins and Weissman, 1984). The initial product is hnRNA (heterogeneous nuclear, or pre-mRNA). This transcript is processed by cleavage of certain portions of the sequence to yield mature messenger RNA. Those parts of the RNA which are spliced out are referred to as introns.

Introns, as opposed to exons, do not appear to have a codon type organization and do not code for the amino acid sequence of the protein. Although their exact function is unknown, some investigators have demonstrated that globin introns may act as regulatory elements (Dobkin et al., 1986). They may also be implicated in limiting gene conversion events within a gene family (Hill et al., 1984).

The procedure by which the genetic information encoded in the mRNA enables synthesis of proteins is known as "translation" and has also been suitably reviewed (Collins and Weissman, 1984). Note that there are a few special codons which are extremely important in this process. The codon AUG serves as an initiator codon for all mammalian systems. As in the case of bacteria, it also codes for the amino acid methionine.

Unlike bacteria, however, the initial methionine is non-formylated and is cleaved from the nascent polypeptide chain during elongation.

There are three chain termination or "nonsense" codons: UAA, UAG, and UGA. These signal the separation of the nascent polypeptide chain from the mRNA complex. At this point the chain assumes its characteristic secondary and tertiary structures and may interact with other polypeptide chains to form quaternary structures.

1.B.2. Globin Gene Structure and Organization

The human alpha globin genes are located on chromosome 16 (Deisseroth, et al. 1977) and the human beta globin genes on chromosome 11 (Diesseroth, et al. 1978). Earlier traditional genetic analyses had determined that there were two different types of globins arranged as family clusters. One group included the alpha globin genes (α,ζ) whereas the other was the non-alpha cluster ($\beta,\delta,\gamma,\epsilon$).

Subsequent studies using Southern blot analysis and in situ hybridization have confirmed the initial chromosome placements and have further localized the beta globin gene cluster to the short arm of chromosome 11 at band 11p15 (Magenis et al., 1985) while the alpha globin gene cluster is on the short arm of chromosome 16 (Koeffler et al., 1981).

Aside from corroborating gene cluster organization and chromosomal

location data, recombinant DNA analyses have allowed construction of genetic maps complete with intergenic distances, a feat not possible using traditional genetic analyses. The currently accepted genetic maps for the alpha and beta globin gene complexes are presented (Figure 1.3).

As expected, single beta and delta loci, and duplicate alpha and gamma loci were found. Some interesting and unexpected characteristics of the globin gene clusters have been determined, however. One may indeed speak of "families" of genes as it is now clear that the embryonic loci zeta and epsilon are linked to the alpha and beta loci, respectively. In addition, regions with structural homology to the globin genes were found which appear to be non-functional thus they have been labeled "pseudo-genes".

A further feature of interest is that the human genes are located in the same order 5′ to 3′ that they are expressed developmentally, e.g.. $5'\epsilon\text{-}^{G}\gamma\text{-}^{A}\gamma-\delta-\beta\ 3'$ (Lawn et al., 1980). Several groups have speculated that the order of gene loci may be important in the control of globin gene expression. Unfortunately this may not prove to be a general rule as notable exceptions have been found in studies of chicken and goat globin loci.

One might also speculate on the role that the intergenic distances play. Those elements which are expressed during the same period of development are found more closely linked than elements which occur at different stages (Bunn and Forget, 1986). These loci are not only

physically closer together but also display greater homology at the amino acid and DNA sequence levels. It has been postulated that gene loci in close proximity are the result of relatively recent gene duplication events whereas larger distances separate those loci which are more distantly related (Bunn and Forget, 1986).

It appears that the globin clusters are areas of the genome which are continuing to evolve. In the case of the alpha-like zeta locus, individuals have been identified with as few as one or as many as three zeta-like loci (Winichagoon et al., 1982). Current theories suggest that cross-over events occur between these closely-related loci. Similar anomalies have been described for the alpha (Goosens et al., 1980; Higgs et al., 1980) and gamma loci (Trent et al., 1981).

Advances have also been made in the determination of the fine structure of the individual globin genes. The DNA sequence is not colinear with the messenger RNA sequence and the amino acid sequence of the protein. Rather, there are non-coding stretches of DNA which are spliced-out during RNA processing. As mentioned earlier, these segments of the gene are referred to as intervening sequences (IVS) or introns.

The globin genes of man and most other mammals are interrupted by introns at two positions: between codons 30 and 31, and 104 and 105 in the beta loci; between codons 31 and 32, and 99 and 100 in the alpha loci. Figure 1.4. illustrates the fine-structure mapping of human globin genes. The globins are relatively small genes with few introns

as compared to such genes as human antithrombin III which is 19 kb in length and is composed of 6 exons and 5 introns (Prochowick et al., 1985) or Factor VIII which is 186 kb long and has 26 exons (Lawn, 1985).

There are certain common features of introns among globin genes. Most of the globins are interrupted by introns at the same codons. There is a conserved sequence GT/AG at the 5' and 3' ends of the intervening sequence which appears to be critical in splicing out the DNA. Finally, the sizes of the introns appear to be highly conservative within the alpha or beta globin groupings, although the two groups differ from one another in the length of IVS-2. In general, IVS-1 is shorter than IVS-2 for both alpha and beta globins, and IVS-2 among beta globins is much larger than that of alpha globins (Bunn and Forget, 1986). Few exceptions to this rule have been noted among the fetal and adult globins. However, the embryonic zeta globin does demonstrate some differences (Bunn and Forget, 1986).

Although introns appear to be highly conserved in length, they are much less conservative at the level of the DNA sequence. Paralogous loci demonstrate significant divergences. Orthologous loci, on the other hand, have been similar enough to delineate the correct ancestors of certain loci. The eta (η) globin story was elicited using such data and will be discussed in greater detail in section 1.D.1.

Little is known about the function of introns although it has been suggested that they may regulate gene expression or the transport of the

RNA from nucleus to cytoplasm (Bunn and Forget, 1986). It is possible that this may be accomplished by adopting left-handed Z-helices (Kilpatrick et al., 1984). Not all globin genes have intervening sequences, e.g. the functional globin genes of the insect *Chironomus thummi thummi* (Antoine and Niessing, 1984).

An alternate hypothesis has been that of "exon-shuffling" which presents the argument that exons code for functional domains of protein which may be rearranged to create new functions (Gilbert, 1978; Blake, 1978). An example of evolution by exon-shuffling, that of proteases involved in blood coagulation and fibrinolysis, has been reviewed by Patthy (1985).

Finally, Go (1981) has postulated that introns divide proteins into modular structural units. Analysis of the distance between α-carbon atoms in β-globins suggested the presence of four, rather than three such domains. Soybean leghemoglobin has since been shown to conform to this expectation with three introns (Brisson, 1982). The third intron is found exactly where Go predicted; thus this may indicate the ancestral conformation of the gene locus (Bunn and Forget, 1986). Another example of exons defining functional domains of a gene is that of the human pancreatic polypeptide gene (Leiter et al., 1985).

Recently, several groups have independently derived a "lariat" mechanism of RNA splicing (Keller, 1984). In addition to consensus GT/AG splice sites at the 5′ and 3′ ends of the introns, U1RNA (a snRNP)

has been implicated. Studies in yeast have demonstrated that this splicing reaction may have originated as an autocatalytic process thus prompting the evolutionary question of which came first, the DNA or RNA? This subject has been thoroughly reviewed (Keller, 1984).

Recombinant DNA techniques have also been used to detail the 5'-flanking sequences of the globin genes. Three preserved sequences have been identified which are believed to have important regulatory roles in protein production. The first is the "ATA" box (ATAA) which is located approximately 30 nucleotides from the cap site (Proudfoot et al., 1982; Efstratiadis et al., 1980). The second is the CCAAT sequence located 70 to 80 nucleotides from the cap site. Interestingly, three CCAAT-like sequences are found 5' to the human delta globin gene (Spritz et al., 1980; Efstratiadis et al., 1980). The third structure is of the form GGGGG(C/T)G and is found 80 to 100 nucleotides 5' from the cap site (Collins and Weissman, 1984). It is found duplicated upstream of the human and rabbit β-globins, once before all other β-globins, and is absent upstream of the poorly-expressed human delta. This sequence is presumed to be a distal promoter element. A mutation in this element resulting in β-thalassemia supports that presumption (Orkin et al., 1984).

The 5'-noncoding sequences which become part of the RNA transcript have also been investigated. At the extreme 5' end is the cap site (m^7GppmNmNmN) which is common to all eukaryotic mRNAs and is important for optimal mRNA function (Shatkin, 1985). Among the globins, the cap

site is followed immediately by the dinucleotide AC (AT in the ϵ-globin gene). All eukaryotic RNAs also include the consensus sequence CC(A/G)CCAUGG which includes the start codon and may signal a site for ribosome binding in addition to initiation (Kozak, 1984). A mutation in the initiation codon (AUG -> ACG) in an $\alpha 2$ gene has been identified as a cause of alpha-thalassemia (Piratsu et al., 1984).

Great similarities among eukaryotic mRNA 3′ noncoding sequences are found in the 3′ terminal poly(A) tail. A common feature among globins is the hexanucleotide AAUAAA located 20 nucleotides upstream from the poly(A) tail (Proudfoot and Brownlee, 1976). Both of these loci appear to be important in stabilizing the mRNA as mutations in either site have resulted in thalassemia (Orkin et al., 1985).

1.B.3. Globin Pseudogenes

As mentioned in section 1.B.2, one finding of genomic DNA analysis has been the identification of pseudogenes.[4] In the case of the globins, these are gene-like structures which are not known to code for any functional protein. Close structural analysis of these pseudogenes has generally revealed distinct anomalies which would act to prevent globin gene expression (Little, 1982; Li, 1983).

---

[4] Although pseudogenes had not been previously detected, they were not unanticipated. Nei predicted in 1969 that the genome of higher organisms would contain a number of nonfunctional genes (Nei, 1969; Li, 1983). His prediction was based on a comparative analysis of genome size from bacteria to mammals.

Several pseudogenes have been found within the human globin families. There is a single pseudogene in the beta globin cluster.[5] Formerly referred to as pseudo-β_1, it has recently been renamed pseudo-η to reflect its distinctive history (see section 1.C.). The $\psi\eta$ (pseudo-eta) locus is located between the $^{A}\gamma$ and δ structural loci as illustrated in Figure 1.3.

Pseudogenes have also been detected in the alpha globin gene cluster. Pseudo-alpha-1 (Proudfoot and Maniatis, 1980) is located 5′ to the $\alpha 2$ globin gene (Figure 1.3). This locus has been thoroughly characterized and multiple structural anomalies have been detected including loss of the normal initiation codon and frame-shift mutations causing premature in-phase termination codons. Most pseudogenes have acquired multiple mutational events making it difficult to determine which defect caused the primary silencing event (Li,1983).

A second pseudogene, pseudo-zeta, is found in the alpha cluster 5′ to the $\psi\alpha 1$ locus (Figure 1.3). It has an interesting history in that it is highly homologous to the functional zeta gene. The two differ by only a nonsense mutation in codon 6 and two amino acid substitutions (Proudfoot et al., 1982).

---

[5] A second pseudogene, termed pseudo-β_2, was originally identified in the beta globin gene cluster 5′ to the ϵ locus (Fritsch, 1980). It has since been determined that this locus was not a pseudogene at all but rather an artifact produced when an A-T rich region of the DNA hybridized with an AT-tailed globin cDNA probe (Shen and Smithies, 1982).

Little is known about the function (or lack thereof) of globin pseudogenes although there are strong suggestions as to the mechanism by which they arose. Most globin pseudogenes seem to be the product of gene duplications which have subsequently undergone inactivating mutations (Li, 1983). It is presumed that these inactivated copies are then removed from the effects of selection and therefore accumulate further mutations at a higher rate than functional genes. Eventually they become fixed in the population on the sole basis of random genetic drift (Li, 1983) as they seem to lack any selective bias.

A second and far more common type of pseudogene, referred to as a "processed" pseudogene also exists (Wilde, 1986). These processed pseudogenes are found at chromosomal locations remote from the globin gene clusters and are distinctive in that they do not contain any intron-related structures. In addition, these processed pseudogenes have a poly(dA) stretch at the 3′ end which corresponds to the normal poly(A) tail of messenger RNA. It is assumed that this type of pseudogene is formed when a cDNA copy of an mRNA integrates at random in the genomic DNA.[6] If this mechanism proves to be true, these pseudogenes may have been nonfunctional since their origination (Li, 1983). The best representative of a processed globin pseudogene is the $\psi\alpha 3$ gene which has been identified in the mouse (Little, 1982; Marx, 1982; Flavell, 1982; Li, 1983).

---

[6] The term "retroposon" has also been introduced to describe these loci (Wilde, 1986).

An estimate of an average mutation rate for pseudogenes of $5x10^{-9}$ substitutions per year has been calculated (Li, 1983) making them the fastest changing of any known gene structure. They evolve twice as fast as the third positions of codons and approximately nine times as fast as the first two positions of codons in functional genes (Li, 1983).

It is possible that the rate of nucleotide substitution in pseudogenes may approach the intrinsic mutation rate thus providing a suitable subject of study for comparing Darwinian vs. neutral theories of evolution. Interestingly, the average substitution rate in pseudogenes is higher than that of synonymous substitution (Miyata and Hayashida, 1981). This may be a reflection of the importance of codon bias in evolving proteins (Wilde, 1986) (Section 1.B.1.).

The pattern of nucleotide substitution in pseudogenes has also been examined (Li, 1983; Gojobori et al., 1982). Although it would be expected that transversion mutations (e.g., A -> T, G -> C) would occur twice as often as transition mutations, the transition mutations actually occur more often. Eventually such a sequence without functional constraint would tend to become rich in A and T. This also appears to hold true for major portions of noncoding sequences (Gojobori et al., 1982).

Another consideration in studying pseudogenes is their consistent placement between the adult and embryonic or fetal genes of a cluster (Wilde, 1986). Cleary et al. (1981) have suggested that these

pseudogenes may serve some sort of regulatory spacer function. They may also occupy their intermediate positions because they are orthologues of a gene locus which was susceptible to inactivation, e.g.. eta globin (Wilde, 1986). A third consideration is that it is usually a second adult-gene which achieves pseudogene status (Wilde, 1986). This may be a reflection of some inherent properties of the environment, e.g. chromatin configuration (Hardies et al., 1984). Finally, the pseudogene may provide a sort of "fail-safe" mechanism by ensuring that recombination events with neighboring loci are nonfunctional. Recombinant adult-embryonic or fetal forms of globins might interfere with globin regulation and operation (Wilde,1986).[7]

A number of pseudogenes have been described in other types of gene families, e.g.. ovalbumin. Their particular features have been thoroughly reviewed (Wilde, 1986).

1.B.4. Globin Intergenic DNA

Although the gene structures in any family cluster are of immense importance in an evolutionary study, the intergenic regions also merit study as the majority of DNA is found in these regions. Although the true function of intergenic DNA remains obscure, certain regulatory

---

[7] Hemoglobin Kenya is a Lepore-type fusion-mutant of the human $^{A}\gamma$ and β-globins which exemplifies the results of such an event (Bunn and Forget, 1986). Recent gene mapping studies have verified the loss of the DNA which normally occurs in the inter-$\gamma\delta$ region (Ojwang et al., 1983). This mutant is associated with the syndrome of high-persistant fetal-hemoglobin (HPFH) indicating that there are indeed regulatory elements located within the intergenic regions of the globin clusters.

systems are included in this region.[4,8]

1.B.4.a. <u>Globin Gene Duplication Units</u>

One of the most important mechanisms in the evolution of new genes and biochemical pathways is that of gene duplication (Li, 1983). It is clearly an important method for increasing gene products. The presence of multiple gene families, e.g. the globins, in mammalian and other higher systems indicates that it is also not an infrequent event (Li, 1983). In the globins, for example, it is probable that a gene duplication event 800 million years ago led to the divergence of myoglobin and hemoglobin (Li, 1983).

There are two mechanisms by which gene duplication may occur: tandem gene duplication and genome duplication (Li, 1983). Ohno (1970) argues that genome duplication is the more important process as needed regulatory elements are also duplicated. There does not appear to be a shortage of the results of tandem gene duplication, however.

Gene duplication acts not only to provide higher production of gene products, but also aids in the acquisition of new gene functions. Partial gene-duplication in the form of functional or structural domains

---

[8] Li et al. (1985) have determined the nucleotide sequence of 16 kbp of DNA 5′ to the human ϵ-globin gene. They identified numerous Alu and Kpn repeats but no open reading frames within these regions. In addition, they identified a 2 kb region with unusual features, including a potential stem-loop structure, a unique 39 bp repeat, and a GC-rich region. As this 2 kb region is conserved in goats it may be of functional significance.

of proteins also appears to be an important mechanism in gene elongation (See Section 1.B.2 for a discussion of exons as functional domains).

One may use the degree of homology in regions of DNA surrounding genes to define the size of gene duplication units. As reliable estimates of age or time since divergence of two loci can only be determined by examining those parts of the gene which have not undergone gene conversion events (Wilde, 1986), these flanking regions become more integral to evolutionary analysis.

1.B.4.b. Repetitive DNA

There are a number of short DNA elements which are repetitive and are found interspersed throughout the human genome. One such group is the AluI family of repetitive DNA which is named for the restriction enzyme whose recognition site lies in the middle of each element (Houck et al., 1979; Schmid and Jelinek, 1982; Jelinek and Schmid, 1982).

The AluI repetitive DNA elements are found as single copies or as double copies in an inverted-repeat orientation (Jelinek and Schmid, 1982). They are short (only 300 nucleotides long) but occur some 300,000 times within the human genome (Jelinek and Schmid, 1982) at intervals of approximately 2500 bp (Arnheim, 1983). The AluI elements are not necessarily exact copies. Rather, they display an average of about 80% homology throughout the human genome (Schmid and Jelinek, 1982). Such conservation of repeated sequences scattered throughout the

genome relies on mechanisms implicated in concerted evolution (Section 1.D.1., Footnote 1).

Within the non-alpha globin gene cluster, AluI repeats are found as shown in Figure 1.5. They occur 5′ to the ϵ gene and on either side of the γ gene and δ–β gene pairs. They occupy similar positions in the alpha globin gene clusters (Hess et al., 1983).

Although the exact function of AluI repeats is unknown[9], it has been demonstrated that they can be transcribed _in vitro_ by RNA polymerase III to yield RNA molecules (Duncan et al., 1979; Duncan et al., 1981; Di Segni et al., 1981). It is believed that such transcripts may be implicated _in vivo_ in maintaining the homology of the dispersed members by the processes of concerted evolution (Arnheim, 1983). _In vivo_ studies of the human beta globin gene clusters have also demonstrated transcription of AluI sequences when the adjacent globin gene (ϵ) was being expressed (Allan and Paul, 1984).

The finding that Alu-repeat elements are usually flanked by short direct sequences of DNA is consistent with the hypothesis that the Alu-repeats are the remnants of transposable elements. They may have been generated as cDNA copies by reverse transcriptase from short RNA elements and inserted throughout the genome (Sharp, 1983; Wilde, 1986).

---

[9] Alu-type insertions have been identified in the introns of ruminant fetal and adult beta globin genes. It has been hypothesized that gene conversions between these two loci have been inhibited by the presence of these Alu sequences (Schimenti and Duncan, 1985a).

Processed pseudogenes appear to have a similar evolutionary history and the two groups have been included in the common classification "retroposon" (See Section 1.B.3.). Arguments as to the source of these elements have included reverse-transcription of RNA polymerase III transcripts (Jagadeeswaran, 1981) and 7SL genes (Ullu, 1984). Whatever the origin of these sequences, they predate the divergence of the human and chimpanzee as their positions in the pseudo-α region have been conserved in these two groups (Sawada et al., 1983).

A second family of repetitive elememts is that of the KpnI family of repeats. These pieces of DNA are longer than the AluI sequence and are repeated less frequently. They are also named for the restriction enzyme whose recognition sequence they include (Jelinek and Schmid, 1982; Shafit-Zagardo et al., 1982; Collins and Weissman, 1984). KpnI repeats have been located 5′ to the human ϵ (Li et al., 1985), 5′ to the human $^{G}\gamma$ (Forget, 1981) and 3′ to the human β-globin (Kaufman et al., 1980; Adams et al., 1980) genes (Figure 1.5). KpnI sequences have also been referred to as "LINEs", i.e. long interspersed nucleotide elements (Singer, 1982; Rogers, 1983) as opposed to "SINEs" (short interspersed nucleotide elements).

The KpnI elements differ from the Alu sequences in that they range from 1.5 to over 6 kb in length. Within the Kpn stretch of DNA there may be subsets of repetitive sequences which may be arranged in differing orders depending on the site being studied (Bunn and Forget, 1986).

As in the case of the Alu repeats, the function of KpnI elements is unknown although they are known to be transcribed into RNA molecules (Shafit-Zagardo et al., 1983; Kole et al., 1983; Schmeckpepper, 1984). They also have a poly(dA) sequence (DiGiovanni, 1983) and are flanked by direct repeats (Thayer, 1983; Nomiyama, 1984) which is suggestive of transposition of a cDNA element into the genome (Wilde, 1986).

KpnI sequences have been identified in mouse (Martin, 1984) in addition to man. Nucleotide sequence analysis of such structures has revealed the presence of a long open-reading frame indicating the potential for encoding a protein (Potter, 1984; Martin, 1984).

A third and highly conservative form of repetitive DNA has been identified in the region between δ and β loci of the beta globin gene cluster (Miesfeld et al., 1981). The element is short (250 bp) and is homologous to the nontranscribed spacer DNA of the mouse ribosomal RNA genes. Due to its conservative nature, it may be blocked during Southern blot analysis if salmon-sperm DNA is used to reduce background hybridization. It is also homologous to sequences in such lower eukaryotes as slime mold and yeast indicating that it is conserved thoughout eukaryotic evolution (Miesfeld, 1981).

Of particular interest in this third un-named family of repetitive DNA is the presence of 16 to 17 tandem repeats of the dinucleotide TG (Poncz et al., 1983). A similar sequence has been implicated as a "hot-spot" for recombination between the gamma globin genes (Slightom et al.,

1980; Shen et al., 1981).

1.B.4.c. Simple Sequence DNA

There are regions of DNA located 5′ to the $^{G}\gamma$ and δ-globin genes which are composed of predominantly pyrimidines on one strand and complementary purines on the other (Poncz, 1983). These regions vary from 200 bp (5′ to $^{G}\gamma$) to 40 bp (5′ to δ). A simple repeat sequence, ATTTT is also located 5′ to the β-globin (Spritz, 1981).

Regions of simple sequence DNA have been demonstrated to have increased S1 nuclease sensitivity and an ability to adapt left-handed Z-DNA conformation (Bunn and Forget, 1986). Such a locus is found between the δ- and β-globin genes (Poncz, 1983) and may contribute to the formation of restriction-fragment length polymorphisms in the area. An analogous construction is found in the inter-zeta hypervariable region as described by Higgs et al. (1981).

1.B.5. Restriction Fragment Length Polymorphism of the Globin Gene Cluster

The noncoding DNA does not appear to be subject to functional constraints to the degree to which the coding DNA is, thus allowing such sequences to evolve more rapidly. These changes are therefore not subject to any negative bias of natural selection, thus, they may be maintained in the population. These extragenic (noncoding) changes may

be detected by screening for restriction fragment length polymorphisms (RFLPs).

Restriction fragment length polymorphisms are detected when total genomic DNA is restricted with a particular restriction endonuclease, subjected to agarose gel electrophoresis, and blotted onto nitrocellulose paper or a nylon membrane according to the method of Southern (1975). The membrane-bound DNA is then probed with the radioactive single-stranded DNA of interest. The results are determined by detecting bands on exposed x-ray film. It is the size of the band (as determined by the distance of migration from the origin) which provides the RFLP information. If multiple band sizes are detected using a single probe and restriction enzyme then a polymorphism exists. Such RFLP markers have recently been identified for several genetic disorders, e.g. Duchenne muscular dystrophy (Monaco et al., 1985) and cystic fibrosis (Knowlton et al., 1985; Wainwight et al., 1985; White et al., 1985).

Numerous factors may contribute to the generation of restriction fragment length polymorphisms. The simplest case is that of a single base substitution which either destroys or creates a restriction enzyme recognition site. In the former case the new band detected will be larger than before while in the latter case it will be smaller. If the probe flanks the recognition site either case may affect the total number of bands detected.

An alternate mechanism of generating restriction fragment length polymorphisms is the insertion or deletion of sizeable fragments of DNA. This type of polymorphism may be detected with multiple restriction enzymes. There are several ways in which insertions or deletions may occur. One example is that of a hypervariable region such as that found between the zeta and pseudo-zeta loci of the alpha globins (Goodbourn et al., 1983, Goodbourn et al., 1984). The variability has been ascribed to the number of tandem repeats of a 36 bp sequence. This sequence can be subdivided into 14 bp domains which are closely related to each other and to other tandemly repeated elements found throughout the genome, e.g., in the 5′ flanking DNA of the insulin gene (Bell et al., 1982; Ullrich et al., 1982).

There are multiple polymorphic sites in the human beta globin cluster (Figure 1.6). Many of these are due to the simple alteration of a specific restriction endonuclease site and can be categorized as either being present (+) or absent (-) (Orkin et al., 1983; Orkin and Kazazian, 1984). Several of these sites are located within the KpnI repeat element located 3′ to the beta locus (Bunn and Forget, 1986).

The majority of the non-alpha RFLP sites occur among widely dispersed population and racial groups indicating an early origin of these sites within the human species. The various sites do not occur totally at random, however, as there are distinct linkage patterns which have begun to emerge. These groupings or combinations of RFLPs have been termed "haplotypes".

Haplotype analysis has been applied to the question of the origin of the human sickle cell mutation by many groups (Kan and Dozy, 1978; Kan and Dozy, 1980; Antonarakis et al., 1982; Piratsu et al., 1983; Wainscoat et al., 1983; Antonarakis et al., 1984; and Pagnier et al., 1984). Such studies have demonstrated the extreme heterogeneity of the haplotype background on which the sickle mutation has arisen. Multiple separate mutations must have occurred to generate this degree of divergence. In the case of sickle cell disease, haplotype analysis has also been useful in defining groups which have varying severity of the disease on a geographic basis (Nagel et al., 1985) as demonstrated in Table 1.2.

Figure 1.6. also demonstrates that there are two groups of polymorphisms which tend to occur in linkage disequilibrium: I and II (Bunn and Forget, 1986; Collins and Weissman, 1984). Between these two regions is an area with a high frequency of crossover or recombination events (Orkin et al., 1983; Piratsu et al., 1983; Chakravarti et al., 1984). It has been proposed that there is a "hot spot" for recombination between the 5′ end of the δ-globin gene and the 5′ end of the β-globin gene (Bunn and Forget, 1986). The mechanism for these recombination events may involve the use of a block of tandemly repeated ATTTT elements or TG dinucleotides (Section 1.B.4.b).

1.C. **Molecular Phylogeny of Globins**

The earliest indication that there was a family of related globin

genes came with the comparisons of the three-dimensional structures of sperm-whale myoglobin (Kendrew et al., 1960) and horse α- and β-globins (Perutz et al., 1960). The initial amino acid composition data (sequences were unknown) would not have led anyone to suggest that the two proteins bore any great resemblance to each other (Dickerson and Geis, 1983). Protein crystallography, however, revealed that each of the four chains in the horse hemoglobin molecule was folded exactly like the single sperm-whale myoglobin chain. This triggered the suggestion that hemoglobin was simply four myoglobin molecules which had been assembled (Dickerson and Geis, 1983). Amino acid sequence comparisions provided the first comfirmation (Edmundson, 1965).

This simple model of globin relationship at the structural level spawned theories of globin molecular evolution which have been enriched by ever-increasing molecular data. The model of globin evolution is that the ancestral protein came from a single gene. This gene underwent a series of duplication events; each of the lineages formed from duplicated loci accumulating independent mutations (Dickerson and Geis, 1983).

There are three main families of globin chains: α, β, and myoglobin. Although all of these groups share certain characteristics, they are also distinct. Only five amino acid residues remain totally invariant (Dickerson and Geis, 1983). The overall similarity among globin molecules is greater than these five since many of the amino acid changes have been conservative (Dickerson and Geis, 1983). The

subunits of hemoglobin have nonpolar regions on their side-chains which promote aggregation. In addition, the sites of α–β chain interaction and the heme-binding pockets are essentially totally conserved (Dickerson and Geis, 1983).

The frontiers of globin gene history can be pushed back even further if one considers leghemoglobin, a protein in the nitrogen-fixing root nodules of legumes (Dickerson and Geis, 1983). The globin's function in this context is unusual, however, in that leghemoglobin, unlike all other globins, acts to remove oxygen from the vicinity of the root nodules rather than transporting it to them.

Although it is apparent that there was a globin ancestral gene as early as 600 MYA (million years ago), it is also apparent that the divergence which resulted in the separate α- and β-globin lineages did not occur until sometime later. The best estimate is that the separation occurred 450 MYA during the time of early vertebrate evolution (Hunt et al., 1978; Czelusniak et al., 1982). The myoglobin-hemoglobin duplication occurred earlier, 500-800 MYA (Jeffreys et al., 1984).

Xenopus is the earliest form of vertebrate which has been studied to date. In this group the α- and β-globin genes remain as linked clusters on a single chromosome. In some toad species there is a single linked cluster while in the case of the species Xenopus laevis there are two such linked clusters on different chromosomes (Jeffreys et al., 1980;

Patient et al., 1980; Hosbach et al., 1982; Hoshbach et al. 1983). It is believed that this arrangement arose as a result of tetraploidization (Bunn and Forget, 1986).

In the lineages leading to birds and mammals, the alpha and beta globins have retained their family cluster organization but the clusters have been located on different chromosomes, with the myoglobin locus on yet a third (Bunn and Forget, 1986; Jeffreys et al., 1984). The mechanism by which this arrangement was accomplished is unknown.

Each globin cluster has been formed by continued gene-duplication events. The resultant genes can be divided into two categories. Orthologous genes are those which are common to many species. Of necessity, the duplication events which generated them must have taken place prior to the time of speciation. Paralogous genes, on the other hand, are those which have arisen by gene duplication events within a given species. Each of the globin clusters is composed of both orthologous and paralogous loci. The differences become important only as one compares across species boundaries.

1.D. Evolution of the Beta Globin Family

1.D.1. General Considerations

The evolutionary history of the beta globins has been long and varied. It will be presented in considerable detail in this section.

The history of the alpha globins has been reviewed elsewhere (Bunn and Forget, 1986). Other gene families have also been identified, e.g. histone clusters, which indicates that generation of multigene families is not unique to globins (Perry et al., 1985).

The beta globin gene family has been created through a series of gene duplication events (Section 1.C). The state of the cluster varies considerably from one species to the next. Certain features, such as orthologous gene loci, are shared as one crosses species boundaries. The major differences are determined by paralogous loci or the number of times that the total cluster has been duplicated. Finally, all species have differential expression of globin genes thoughout ontogeny and development. In general, these loci are arranged from 5′ to 3′ in the order which they are expressed, i.e. embryonic through adult forms. Exceptions to this rule do exist and they will be presented in section 1.D.2. An overall scheme of the chromosomal organization of the various species-specific globin gene clusters is presented in Figure 1.7.

The initial attempts at constructing a globin phylogeny were based on the results of amino acid sequence analysis. These attempts were hampered somewhat by the degeneracy of the genetic code. The amino acid substitutions did not always prove to be the result of unambiguous substitutions at the level of the DNA.

Nucleotide substitutions can have two different results. Silent-site substitutions do not change the encoded amino acid sequence whereas

replacement-site substitutions do. Comparisons of the number and type of nucleotide substitutions can give an estimate of the degree of relatedness of the proteins under consideration. Investigators have handled this data in differing ways. One way is to assume that the number and type of substitutions can provide a fixed evolutionary time scale. Estimates for the rate of silent-site substitutions are in the range of 1 per cent per 1.4 million years (Perler et al., 1980) whereas replacement-site substitutions occur at the rate of 1 per cent per 10 million years (Efstratiadis et al., 1980).

Efstratiadis et al. (1980) have used the fixed rates to elucidate a proposed phylogenetic tree for the beta globins. They estimated that the delta and beta globin loci diverged most recently, only 40 million years ago. Gamma and epsilon, the fetal and embryonic forms of beta globin, were the result of a duplication-divergence event which occurred some 100 million years ago. Finally, the divergence between the adult (δ,β) and embryonic-fetal (ϵ,γ) branches took place nearly 200 million years ago.

An alternate method is to pattern phylogenetic trees on the total number of independent mutations which occur between lineages, regardless of the time scale. The smallest number of mutations is used to detect the closest branch points. This is the method of maximum parsimony analysis as espoused by Goodman (Czelusniak et al., 1982).

A major pitfall of both of these methods is the inability to

accomodate the effects of gene conversion events. Gene conversion occurs when intrachromosomal recombination happens between linked genes, resulting in the correction of one gene sequence to that of the other and is believed to be an important mechanism of concerted evolution of gene families such as the globins (Arnheim, 1983)[10]. An excellent example of this process has been described in the fetal globin genes by Slightom et al. (1980; 1985; Shen et al., 1981).

The two fetal globins of man are linked in the configuration ${}^{G}\gamma$-${}^{A}\gamma$. The amino acid sequence of these two proteins are nearly identical; thus, initial estimates dated their divergence as occurring within the recent past. Nucleotide sequence analysis has altered this estimate by incorporating the data available from the larger non-coding intron (IVS-2). In a particular individual, the IVS-2 of the ${}^{A}\gamma$-gene on one chromosome was more similar to the second intron of the ${}^{G}\gamma$ locus on the same chromosome than it was to the ${}^{A}\gamma$-gene on the other homologous chromosome (Slightom et al., 1980; Shen et al., 1981). A gene conversion event can best explain these results (Stephens, 1985). A smaller-scale gene conversion event may account for the findings of individuals with linked ${}^{A}\gamma$-${}^{A}\gamma$ or ${}^{G}\gamma$-${}^{G}\gamma$ loci (Powers et al., 1984).

---

[10] It has been noted that the sequences of paralogous loci are often more homologous within a species than orthologous loci are between species. This phenomenon has been termed "concerted evolution" (Zimmer et al, 1980). A popular molecular mechanism to explain these findings is gene conversion although convergence through parallel mutations is also implicated (Arnheim , 1983). Concerted evolution occurs not only in multigene families, but also in highly repetitive families such as the AluI repeats (Section 1.B.4.b.).

A second example of an evolutionary divergence event occurring in the more distant past than was originally predicted is that of the human delta and beta globin loci. Initial estimates for this date of 40 million years ago (Estradiadis et al., 1980) have been modified by the inclusion of non-coding DNA and cross-species analysis (Hardies et al., 1984; Hutchinson et al., 1984; Hardison, 1984; Hardison and Margot, 1984). These data would indicate that the 5′ flanking and coding sequences of the δ-globin gene were actually *converted* to β-globin-like sequences 40 million years ago. The gene locus itself is clearly much older as δ-like genes are found in prosimii which diverged as long as 75 million years ago (Jeffreys, 1982).

Another interesting locus in the beta-cluster is that now called pseudo-eta ($\psi\eta$). This locus, formerly referred to as $\psi\beta$, is located between the $^{A}\gamma$ and δ-globin genes (Figure 1.3). Dot matrix analysis, whereby the nucleotide sequences are plotted out in a two-dimensional array (Konkel et al., 1979), has revealed that although the locus has considerable homology with the human ϵ-and γ-globin genes, it is actually more closely related to the goat ϵ^{II} gene (Figure 1.7). This would indicate that the human $\psi\eta$ and goat ϵ^{II} genes are related orthologously whereas human $\psi\eta$ and human ϵ and γ are related paralogously (Goodman et al., 1984; Harris et al., 1984).

Other animal species have also contributed to unraveling the eta-globin story. The brown lemur, a prosimian, has a pseudogene (Figure 1.7) which appears to be a fusion between $\psi\eta$ at its 5′ end and δ-globin

at its 3′ end (Barrie et al., 1981; Jeffreys et al., 1982; Chang et al., 1984). This is probably the result of nonhomologous crossing over between pseudo-eta and δ-gene loci (See Section 1.D.2.g.).

An important implication of the discovery of eta-globin as an independent functional locus is that the prototype mammalian beta-globin gene cluster must have been composed of five, rather than four, distinct loci. These five loci could be divided into two groups which probably arose from independent gene duplication events: 1) Proto-ϵ, $-\gamma$, and -η; and 2) proto-δ and -β. The history of these proto-globin genes can be traced for several species lineages.

1.D.2. Specific Examples of Beta-Globin Gene Clusters

Only a handful of species have been examined at the molecular level. Each species has made a unique contribution to the overall story of beta-globin phlyogeny. General comments on the molecular genetics of the beta globin clusters were presented in Section 1.D.1. An overall scheme of beta-globin cluster maps was shown in Figure 1.7.

One objective of the comparative study of globin genes has been an attempt to identify regulatory elements by looking for the most conserved areas or sequences (Edgell et al., 1983). The overall organization of the globin clusters has proven to be surprisingly diverse, which will make determination of orthologous loci and sequences fairly challenging.

1.D.2.a. Toad

The South African Clawed Toad (Xenopus laevis) is the best characterized amphibian with respect to the globin genes. While higher vertebrates have α- and β-clusters on different chromosomes, those of Xenopus laevis are found linked on a single chromosome. There are two clusters of globin genes per chromosome. Each cluster has the overall arrangement of 5'-(larval α-genes)-(adult α-genes)-(adult β-genes)-(larval β-genes)-3' (Meyerhof et al., 1984). There have been a total of 12 globin loci detected in Xenopus.

The fine-structure of the globin genes appears to be conserved from amphibia to mammals as the larval β_I-gene of Xenopus laevis is composed of 3 exons and two introns.

1.D.2.b. Chicken and Duck

The chicken beta globin complex is distinctive among non-alpha globin families in that it lacks a pseudogene (Dolan, et al., 1981; Villeponteau and Martinson, 1981; Roninson and Ingram, 1982). The avian alpha globin complex also lacks such a locus (Engle and Dodson, 1980).

1.D.2.c. Mouse

There are four known expressed β-globins in the Mus musculus haplotype [Hbb]^d (Edgell et al., 1983): two are adult β-globins and two

are non-adult. At the level of the DNA sequence, however, there appear to be seven structures with homology to the mouse adult β-globin (Edgell et al., 1983; Jahn et al., 1980). Of these seven structures, three can be demonstrated to code for the functional β-globin proteins: d-major (dmaj), d-minor (dmin), and a non-adult protein η (Edgell et al., 1983; Konkel et al., 1979).

There is a second beta globin haplotype present in Mus musculus; [Hbb]^s (Edgell et al., 1983; Weaver et al., 1981). It also has two adult coding loci although only a single β-globin protein is found. Nucleotide sequence data indicate that a gene-conversion event has taken place on the [Hbb]^s cluster resulting in two loci similar to the β^{dmaj} of the [Hbb]^d haplotype.

The intervening sequences of β^{dmaj} and β^{dmin} exhibit considerable divergence despite the homology of the coding sequence (Edgell et al., 1983). This may reflect lower selection pressures on noncoding DNA. It may also depend on the inherent nature of the DNA to undergo mutation.

As in other genera, the mouse globin loci also contain a family of repeat sequences which occur in at least nine locations. Overall, they occur some 20-50,000 times within the mouse genome. The repeats are 90% homologous but are unusual in that the repeats seem to terminate at random points with respect to each other. These sites of repetitive DNA may play a role in the phenomenon of gene conversion (Edgell, et al., 1983).

There are four other loci with homology to the adult globin locus in the mouse which are not expressed. A pseudogene, βh3, is present and appears to be a product of an illegitimate recombination between two different ancestral genes (Hardies, et al., 1984).[11] It is characterized by an internal deletion which would lead to the hypothesis that βh3 was formed from β and βh2 by a Lepore-type mechanism (Wilde, 1986). There are also numerous other mutations and deletions which make this a non-functional locus. The βh0 and βh1 are assumed to be normal functional embryonic genes. The locus βH2 appears to be a second pseudogene and is accumulating nucleotide changes at a rate compatible with that supposition although it does retain the structural features of a protein (Edgell, 1983). It has only 72% homology with the β coding loci but only two overtly deleterious mutations (Phillips, et al., 1984) leading to the conclusion that it must have diverged as a functional gene for some portion of its history (Wilde, 1986).

Examination of the 3' flanking DNA of the mouse loci lends credence to the theory that β-globin is a primitive gene which has recently undergone gene-conversion in the primate lineage (Edgell, 1983). The βH3 3' sequences appear to be orthologous to the human δ-globin gene. Mouse βH2 pseudogenes are also δ-orthologous (Wilde, 1986).

Although it appears that the common mammalian ancestor had a five

---

[11] The mouse beta globin complex has a distinct heritage in this regard as there are no pseudogenes within the mouse alpha globin cluster (Leder, et al., 1980). There are mouse alpha globin pseudogenes but they are processed pseudogenes which are located at remote sites within the genome (Section 1.B.3.).

member beta-globin linkage group; 5′ $\epsilon-\gamma-\eta-\delta-\beta$ 3′, this type of organization has not been conserved in the mouse (Wilde, 1986). Mouse non-alpha globins have been derived from the four globin gene set; 5′ $\epsilon-\gamma-\delta-\beta$ 3′. The eta globin locus was lost at an early stage of rodent evolution.

1.D.2.d. <u>Rabbit</u>

There are four beta loci found in the rabbit (Hardison et al., 1979; Lacy et al., 1979): two embryonic ($\beta 4$, $\beta 3$); and two adult ($\beta 2$, $\beta 1$). Of these four loci, a single pseudogene ($\beta 2$) is present and is characterized by multiple mutational changes suggesting that it is of ancient origin. Lacy and Maniatis (1980) have analyzed the sequence of this locus and suggested that it probably evolved as a functional gene for 22 MYr (million years) after the initial gene duplication event prior to becoming a pseudogene some 33 million years ago. As mentioned in section 1.B.4.a. it is important to examine noncoding DNA in making such analyses. In the case of the rabbit $\psi\beta 3$ it appears that it was actually converted 50-55 MYA and originated much earlier (85 MYA) thus predating the mammalian radiation (Hardison and Margot, 1984).

As is the case among rodents, lagomorphs also appear to have descended from a common ancestor who had a four globin gene linkage set: 5′ $\epsilon-\gamma-\delta-\beta$ 3′. The eta globin locus which is known to have been present in the common ancestor of all mammals appears to have been lost prior to any lagomorph radiation (Wilde, 1986).

1.D.2.e. Goat

The goat globins tell an interesting tale in that not only have tandem duplications of single gene loci occurred, but tandem duplications of the entire β-globin cluster have occurred as many as three times (Townes, Fitzgerald, et al., 1984; Townes, Shapiro, et al., 1984). The basic organization of the cluster units is similar to that of other beta-globin clusters as there are two embryonic and two adult genes (Wilde, 1986).

Pseudogenes have been identified in two of the three clusters. The two loci, $\psi\beta^{X}$ and $\psi\beta^{Z}$, are extremely homologous at the level of the nucleotide sequence (90%) to the point of sharing several deleterious mutational events, e.g. in-phase terminators and altered ATA promoter boxes (Cleary et al., 1980; Cleary et al., 1981). This would indicate that the progenitor of these sequences must have been a pseudogene prior to gene duplication. As in the case of other beta-globin pseudogenes, the age at divergence of these two loci appears to have been obscured by a gene-conversion event (Hardies et al., 1984; Hardison and Margot, 1984; Wilde, 1986).

Unlike primate pseudogenes the goat $\psi\beta^{X}$ and $\psi\beta^{Z}$ loci appear to be orthologues of the primate minor adult globin δ rather than the primitive eta locus (Wilde, 1986). The eta orthologues in this group are also unusual in that they appear to be functional, e.g. the goat ϵ^{II} gene (Goodman et al., 1984).

1.D.2.f. Cow

Goats and cows share a common ancestor as recent as 18-20 million years ago. Cows (Bos taurus) are like humans in their developmental expression of hemoglobins (embryonic -> fetal -> adult) whereas goats (Capra hircus) have an additional preadult or juvenile form . The entire bovine beta globin cluster has recently been characterized (Schimenti and Duncan, 1985b). Cows have a duplicated four-gene set composed of a fetal and an adult globin cluster. The beta cluster has a single pseudogene while the fetal cluster has two.

Cows, like goats, provide exceptions to the rule of 5' to 3' developmental expression of globin genes as γ is located 3' to β. The fetal and adult globin loci occupy comparable positions in the goat and cow. The cow γ and β genes are 90% homologous while they are even more homologous to their orthologues in the goat. Nucleotide sequence analysis of the bovine ϵ genes has identified the following relationships: bovine ϵ^3, ϵ^4, ϵ^1 and ϵ^2 genes are orthologous to goat $\epsilon^{I}/\epsilon^{III}$, $\epsilon^{II}/\epsilon^{IV}$, ϵ^{V} and ϵ^{VI} genes, respectively (Schimenti and Duncan, 1985b). The cow ϵ^2 and ϵ^4 appear to have undergone a recent gene duplication event (Schimenti and Duncan, 1985a).

The three bovine beta pseudogenes have also been analyzed. Multiple defects in both regulatory and structural regions have been identified which could have made these genes nonfunctional (Schimenti and Duncan, 1985b). These defects are common to both cows and goats thus it appears

that these loci must have already been pseudogenes in the common ruminant ancester. The duplicate pseudogene in the fetal cluster arose after the cow diverged.

1.D.2.g. Lemur

The β-globin cluster in lemurs is the smallest yet descibed among primates (Wilde, 1986). There is a total of four loci of which only three are functional (ϵ, γ, β). The remaining locus is the $\psi\beta\delta$ hybrid pseudogene (Jeffreys et al., 1982). Nucleotide sequence analysis of the 5′ lemur pseudogene and its flanking regions indicates that it is 75% homologous to the the human $\psi\beta$ ($\psi\eta$) locus (Jagadeeswaran et al., 1983; Wilde, 1986). Analysis of the 3′ end and flanking regions, on the other hand, seem to be more homologous to the human delta globin (Jeffreys et al., 1982). Thus the lemur beta-globin cluster provides one more example of an event which appears to be not uncommon in the beta family; a Lepore-type unequal-crossover fusion event (Baglioni,1962).

1.D.2.h. Higher Primates

A number of higher primates in addition to man have now been analyzed at the nucleotide level within the beta globin cluster (Wilde, 1986). As mentioned in Section 1.D.1. it has recently been determined that the ancestral primate β-globin gene cluster comprised a five member linkage group: 5′ $\epsilon-\gamma-\psi\eta-\delta-\beta$ 3′ (Goodman et al., 1984; Koop et al., 1986). Primates appear to be the only analyzed group in which

representatives of all of the proto-beta globin genes have been retained (Wilde, 1986). Mouse and rabbit have lost the η group while goats have lost the γ orthologues.

The primate δ-globin gene has also had an interesting history. As previously mentioned (Sections 1.D.2.c. & d.) the δ-globin gene is orthologous to the mouse pseudogene βh2 and the rabbit $\psi\beta 2$. Delta-globin enjoys limited expression among primates and in some instances, e.g. Old World Monkeys, a δ-globin gene product has never been identified (Martin et al., 1980; Martin et al., 1983). It appears that the delta locus may be predisposed to inactivation (Wilde, 1986). In the great apes and man the δ-globin locus appears to have undergone a recent gene conversion event with its 5′ end and flanking sequences converted to β-like sequences (Martin et al., 1983; Wilde, 1986). Presumably this has allowed gene expression at its current low level.

Although a delta globin gene product has never been found in Old World monkeys, a gene locus at the DNA level has been identified for several species, e.g. colobus, langur, green monkey, talapoin, baboon, and rhesus (Martin, 1982). As delta globin is expressed in New World monkeys (Boyer et al., 1969) it appears that some sort of gene-silencing event must have occurred early on in the Old World monkey lineage. The nature of this silencing event has yet to be elucidated. An alternative hypothesis is that the gene was silenced in the common Old World monkey - Anthropoid ancestor and later rescued by gene conversion in the Anthropoid line (Wilde, 1986).

1.E. Other General Evolutionary Considerations

1.E.1. Selectionist vs Neutral Mutation Theories

The neutral theory was first proposed by Kimura as an attempt to account for the difference between expected and observed mutation rates (Fitch and Langley, 1976). The basic premise of the theory is that most changes at the molecular level are due to the random fixation of selectively neutral or nearly neutral mutations (Kimura, 1983). It places emphasis on mutation and random-drift as the causative factors of molecular evolution. The neutral theory has been strongly supported by King and Jukes (1969) and has been thoroughly reviewed by Kimura (1983).

The Darwinian or selectionist theory places its emphasis on the concept of natural selection. The basic tenet of this theory is that organisms adapt to their environment as a result of positive selection for beneficial mutants. A strong proponent of the neo-Darwinian position is Goodman and his position has been thoroughly reviewed (Goodman et al., 1975).

The evidence Goodman mounts in support of his position is the nonconstancy of the rate of mutant substitutions among globin genes. In the early stages of globin evolution, as new functions were being developed, the mutation fixation rate was high due to positive selection. As the globin function became more thoroughly developed with greater intermolecular interactions subsequent changes became

disadvantageous. Such disadvantageous mutations were selected against and the mutation fixation rate slowed down.

The concept of a molecular clock is often related to the theory of neutral evolution, in as far as the two require that the fixation of mutations should occur at regular intervals. The neutral mutation theory and the molecular clock hypothesis remain as independent concepts despite any similarities.

The term "molecular evolutionary clock" was first introduced by Zuckerkandl and Pauling in 1965 (Kimura, 1983). They proposed that there was a near-constancy of the amino acid substitution rate throughout evolution. Since its description, it has remained a subject for great debate.

1.E.2. **Other Methods Used in Molecular Evolution Studies**

1.E.2.a. DNA-DNA Hybridization

The best known example of the application the technique of DNA-DNA hybridization is the work of Sibley and Ahlquist (1983, 1984a, 1984b, 1986). They have redefined the traditional avian classification by studying thousands of birds over the past ten years.

The advantage of the technique of DNA-DNA hybridization is that it enables one to compare a large number of genetic units (Sibley and

Ahlquist, 1986) as opposed to a few single gene loci. It is the single-copy DNA which is compared as the repetitive DNA is removed prior to analysis. A basic assumption of the method is the existence of a uniform mutation rate, or "molecular clock" (Wilson et al., 1977). This assumption appears to be valid for bird species and the median rate is a reduction of one degree Celsius per 4.5 million years since two lineages shared their last common ancester.

1.E.2.b. Mitochondrial DNA Analysis

Mitochondrial DNA (mtDNA) has become an important subject for study in the field of molecular evolution of multicellular organisms due to its relative simplicity in comparision with the nuclear genome (Ferris et al., 1981a; 1981b; Brown, 1983). Mitochondrial DNA consists of a small closed-circular duplex DNA molecule (16.6 kbp) which contains gene sequences that occur only once per mitochondrial genome (Avise and Lansman, 1983). These gene sequences have no intergenic regions and are devoid of introns (Brown, 1983). Although mtDNA does not appear to be susceptible to sequence rearrangements it does mutate at a fairly rapid rate making it an ideal subject for comparision of recently diverged species (Brown et al., 1979). It also has the advantage (or disadvantage as the particular case may be) of reflecting purely uniparental contributions as it appears to be clonally inherited from maternal sources (Brown, 1983). A thorough review of all aspects of mitochondrial DNA variation has been written by Brown (1983).

Although mtDNA appears to change rapidly at the level of the nucleotide sequence, the overall genome organization has not changed since the divergence of the mammals and amphibia some 350 MYA (Brown, 1983; Avise and Lansman, 1983). This may provide a clue as to the independent nature of the mitochondrial genetic code.

As mentioned, the mtDNA evolves more rapidly than single-copy nuclear DNA by a factor of 3-4 (Brown et al., 1979). Restriction endonuclease cleavage site mapping confirmed this rapid rate of mutation but also demonstrated that among primates the initial rate of divergence is faster by a factor of ten (Brown et al., 1974). Thus closely related species, i.e. those with a divergence time of less than 15 million years, can readily be compared using mtDNA. The main disadvantage of this system is that silent sites become rapidly saturated with mutations thereby decreasing resolution of divergences (Avise and Lansman, 1983).

Mitochondrial DNA does bear some similarities to genomic DNA. One example is the occurrence of modified bases: 5-methylcytosine (m^5C) in particular. Pollack et al. (1984) have demonstrated that m^5C occurs particularly in the doublet CpG. This doublet appears to be under-represented within the mitochondrial DNA. DNA methylation has been implicated as a regulatory event in gene expression.

1.E.2.c. Chromosome Evolution

Considerable work has been done in comparative studies at the chromosomal level. Nonhuman primate chromosomes have proven to have great similarities in both qualitative and quantitative differences thus making comparisons relatively straight-forward (DeGrouchy et al., 1978; Mai, 1983). The most significant event in the vertebral evolutionary series has been the Robertsonian translocation (Miller, 1977). Since such structural rearrangements are both rare and complicated, it is unlikely that two lineages would share the same changes as a result of independent mutations (Mai, 1983).

More than 30 different structural rearrangements have occurred within the hominoid lineage. The majority of these have been inversions. The events implicated in human, chimpanzee, gorilla, and orangutan have been summarized (Mai, 1983; DeGrouchy et al., 1978).

1.F. Macaque Evolution

The primates are a distinct and well-varied group of genera; of these the genus macaque is a prominant member of the category known as "Old World Monkeys". The relationship of macaques with respect to other families of primates is portrayed in Figure 1.8.

The field of primate molecular anthropology actually had its beginnings in 1904 when Nuttall tested the serum of various animals

species with antisera to the whole serum of other animals (Lasker, 1976). His preciptin test results have since been confirmed by others using more modern methods (Goodman, 1962). Studies of the macaques did not make any significant contributions until 1940, however, when Landsteiner and Wiener demonstrated a rhesus blood group substance on human red cells. They used an antiserum to macaque red cells to study human cells (Lasker, 1976).

There are some 13 to 19 macaque species within the cercopithecinae subfamily of Old World monkeys and they represent the Asian members of this group. Their geographical range is immense, including southern and eastern Asia and northwestern Africa, ranking second only to _Homo_ among the primates (Fooden, 1980; Lindburg, 1980). The overall range which they occupy is but one measure of their success. The physical niches which they inhabit are also extremely varied: ranging from arboreal splendor to near deserts, and snowy mountains to Indian cities (Fooden, 1980). They have also proven to be extremely useful as subjects for medical research; the best known example being the rhesus macaque (_Macaca mulatta_).

Macaques are usually divided into four groups of living species: (1) _silenus_-_sylvanus_ group, (2) _sinica_ group, (3) _fascicularis_ group, and (4) _arctoides_ group (Fooden, 1980). The physical characteristic used in defining these groups is the shape of the glans penis. Behavioral characteristics such as copulatory patterns assist in defining these groups. The geographic distribution of these groups is sympatric within

the Indochinese Peninsula, although the distribution of the component species is strictly allopatric. This may account for the relative rarity of inter-specific hybrids (Fooden, 1980).

Little is known about the earliest stages of the cercopithecine radiation although it is believed that they may have separated from the ancestral colobines some 15 million years ago (Delson, 1980). The papionins diverged into a distinct group from the early cercopithecines some 5 million years later, during the Miocene epoch. The Macaca subtribe diverged somewhat later within the Papionini at a date some 5 to 7 million years ago (Delson, 1980). A review of the geographic dispersion with time of these groups has been published by Delson (1980).

Although penile morphology and copulatory behavior have been used to define the macaque subgroups, these are not considered standard paleontological criteria. Unfortunately, the fossil-data have been little more than ambiguous (Delson, 1980). Thus, the macaques represent an excellent group of animals for the application of molecular techniques in an attempt to define both evolution and systematics (Cronin et al., 1980).

A useful form of molecular data is the elucidation of the karyotype. In the case of the cercopithecines, the species in the genera Papio, Macaca, Theropithecus, and Cercocebus all share the same diploid number of 42 chromosomes (Cronin et al., 1980). The karyotype is composed of

13 submetacentric pairs, 6 metacentrics, one large achromatic pair with a secondary constriction, and the sex chromosomes (Cronin et al., 1980). Detailed studies using G and C banding have not defined any great differences between the genera (Cronin et al., 1980).

The information regarding macaque systematics at the macromolecular level has been more enlightening. Such studies as immunodiffusion, protein sequences, immunological comparisons, DNA annealing, and various electrophoretic techniques have been used (Cronin et al., 1980). Although the data have been informative, they have not been definitive. The time since divergence of any of the individual macaque species does not exceed one to three million years making the expected and observed differences very small (Cronin et al., 1980). A phenogram for the macaque species based on protein electrophoretic data is presented in Figure 1.9.

1.G. Outline of Proposal

The globin system has long been in the forefront of biomedical research. Phenotypic studies provided evidence for the applicability of Mendelian genetics to mammals. Protein studies led to greater understanding of the mechanisms involved in gene expression and function in addition to elucidating plausible phylogenetic relationships among living organisms. More recently, with the advent of recombinant DNA technology, the globin system has maintained its lead as correlations between defects at the DNA level and phenotypic expression at the

disease level have been elucidated.

An additional application of recombinant DNA technology to the globin gene system has been an extension of previous phylogenetic correlations. Studies at the DNA level have been carried out on the globin systems of mice, frogs, chickens, ducks, rabbits, goats, sheep, cows and primates (Section 1.D.). Of particular interest to this project has been the work conducted by Martin (1982) and Zimmer (1981) in the labs of Allan Wilson and Y.W. Kan. In addition to generating a globin gene based phylogeny of primates, their studies demonstrated that despite an absence of delta globin gene expression in Old World Monkeys, a delta globin coding sequence did indeed occur at the DNA level.

The purpose of this thesis project has been to extend this type of study throughout the Old World Monkey genus macaque. Using those macaque species readily available at the Oregon Regional Primate Research Center, studies at the both the DNA and protein levels were done.

The beta-delta globin region of the non-alpha globin cluster has been the focus in this study. One locus, beta, is functional, whereas delta is not. Rather, delta is essentially a pseudogene and may not be subjected to the stringency of selection to which beta globin is. This may be a useful distinction when comparing closely related species.

Preliminary studies at the protein level included hemoglobin

electrophoresis. Previous studies of macaque hemoglobins have utilized the technique of starch-gel electrophoresis. This project has used the newer cellulose-acetate technique. As either of these electrophoretic techniques are rather inexact, further protein work utilized the technique of high-performance-liquid-chromatography (HPLC).

The purpose of the protein studies was to ensure that in the macaque species studied, no protein was detected which behaves in a similar fashion to human delta globin under the same conditions. In addition, these studies were used to determine the overall phenotype of the monkeys so that subsequent studies at the DNA level could make use of animals which were representative of the species norm at the globin loci.

Once the protein studies were completed, restriction enzyme maps of the beta-delta globin gene cluster were generated for the appropriate representative animals. Such maps serve not only to define the presence or absence of delta globin coding sequences in additional macaque species but may also provide useful information as to the degree of relatedness between the various macaque species and other primates, including man.

CHAPTER TWO: MATERIALS AND METHODS

2.A. Collection and Preparation of Peripheral Blood Samples

2.A.1. Selection of Subjects

Five Macaque species housed at the Oregon Regional Primate Research Center (ORPRC) in Beaverton, Oregon, were selected for study (Table 2.1).

Ten normal, healthy adult animals were selected at random from each of the available Macaque colonies. An exception was made for the *Macaca nemestrina* colony. At the time of sampling this colony consisted of a total of only eight adult females. All eight animals were studied. Please see Appendix A for characteristics of individual animals from each colony.

In addition ten human subjects were selected to provide control data.

2.A.2. Sample Collection

Ten to twenty ml samples of venous blood were obtained from peripheral veins. The blood was collected into 5ml Vacutainer™ (Becton-

Dickinson Co., Rutherford, New Jersey) tubes containing 7.5 mg Na_2EDTA by ORPRC staff and delivered to OHSU by courier on the same day. Repeat samples were collected as necessary.

2.A.3. Sample Preparation

The blood was mixed well in the Vacutainer™ tubes then transferred to sterile 15 ml polystyrene centrifuge tubes. Ten ml aliquots of the blood samples were centrifuged for 30 minutes at room temperature and 1,625 x g in a Model CL International Clinical Centrifuge with a fixed No. 213 rotor (2,800 rpm; International Equipment Co., Needham Hts., Mass.). The spun samples were stored overnight at 4 C.

The plasma was removed by gentle suction using sterile glass Pasteur pipettes. Congealed buffy coats containing the nucleated white cells were removed using sterile wooden sticks, while those not congealed were removed using sterile Pasteur pipettes. Buffy coats were stored frozen at -70 C in sterile 1.5 ml Eppendorf tubes. The red cells were washed three times by resuspending them in an equal volume of cold reticulocyte saline (0.13 M NaCl, 5 mM KCl, 7.4 mM $MgCl_2$ $6H_2O$; 4 C) and centrifuged at 1,625 x g (2800 rpm) for 10 minutes at room temperature. Washed red blood cells were used immediately or stored frozen at -70 C.

2.B. Protein Analysis

2.B.1. Hemolysate Preparation

The red cells isolated in Section 2.A.3. were hemolyzed by a modification of the procedure described by Riggs (1981). Cells were washed three times in an equal volume of 0.15 M NaCl in 15 ml conical polystyrene centrifuge tubes. The cells were centrifuged in a Model CL International Clinical Centrifuge at 1,625 x g (2,800 rpm) for 10 minutes. After washing, the cells were mixed with one-half volume deionized water (dH_2O) and vortexed. The cells were alternately frozen, at -20 C, and thawed, at room temperature, three times. Vortex mixing was performed after each freeze-thaw cycle. The red cell solution was centrifuged in a 50 ml round-bottom polypropylene centrifuge tube at 1100 x g (3000 rpm, SS-34 rotor) for 20 minutes. The supernatant fluid was filtered through one layer of Whatman No. 1 filter paper (Whatman Ltd., Maidstone, England) and stored at 4 C. The final concentration of the hemolysate was in the range of 2 to 3 g/dl by this procedure.

2.B.2. Cellulose - Acetate Electrophoresis

Hemolysates were electrophoresed on Titan III-H cellulose-acetate plates with mylar backing (Helena Laboratories, Beaumont, Texas) in TEB buffer (85 mM Tris(hydroxymethyl)aminomethane (Tris), 1.5 mM K_2EDTA, 50 mM Borate; pH 8.4) according to the method outlined by the Center for Disease Control (1976). A mixture of human hemoglobins A, F, S, and C

provided standard markers.

The cellulose-acetate plates were soaked in TEB buffer for at least 20 minutes prior to use. A standard Helena electrophoresis chamber was used with TEB buffer in the two outer chambers and Whatman 3MM chromatography paper (Whatman Ltd., Maidstone, England) used as wicks to the two inner chambers. The plates were blotted evenly between paper towels to absorb excess moisture. Hemolysates were applied 2 cm from one end of the plate using a Helena Zip-Zone™ applicator. The cellulose-acetate plate was placed mylar side up with the applied hemolysate at the cathode end. A 2 x 3 inch glass microscope slide was placed over the plate to insure adequate contact with the wicks. The hemolysate was then electrophoresed at 300 V for 20 minutes at room temperature.

The plate was removed at the end of electrophoresis and stained for 5 minutes with Ponceau S (0.5% Ponceau S in 5% Trichloroacetic acid). The plate was destained in three 5 minute washes of 5% acetic acid then allowed to air-dry at room temperature.

Samples were loaded in concentrated (see section 2.B.1.) and dilute (1:1 with dH_2O) hemolysate form in order to identify individual hemoglobin phenotypes. When results were ambiguous, a larger-scale electrophoresis was done using Sepraphore III (Gelman Instrument Co., Ann Arbor, Michigan) cellulose polyacetate electrophoresis strips under the same conditions as described above. The Sepraphore™ strips were

counterstained with Benzidine stain[1] until the bands were defined and washed by soaking in dH_2O acidified with a few milliliters of glacial acetic acid. The Sepraphore strips were dried between sheets of Whatman 3MM paper under a glass plate to prevent shrinkage.

2.B.3. High-Performance Liquid-Chromatography Globin Chain Analysis

Hemolysate samples prepared as described in Section 2.B.1. were prefiltered through disposable 0.45 μm filters prior to high-performance liquid-chromatography (HPLC) in order to remove any residual stroma or other debris. A sample of 0.1 mg hemoglobin was analyzed for representative animals of each species. Human hemoglobins A and A_2 were also run separately for comparison.

The equipment used included a Beckman-Altex Model 332 HPLC System with a Model 420 microprocessor control unit. Samples were injected through a Model 210 syringe-loading injection valve with a 20 μl sample loop. The column was a Vydac large pore C_4 column (330Å, 4.6 x 250 mm). The absorbance of the effluent at 220 nm was monitored using a Model 100-30 Altex/Hitachi UV-Vis variable wavelength detector. Results were recorded using the printer-plotter of an HP 5830A (Hewlett-Packard) gas

---

[1] Benzidine stain is prepared as follows:

Mix 1 gram benzidine hydrochloride with 50 ml of glacial acetic acid until completely dissolved. Add 1 gram sodium nitroprusside. Add dH_2O to 1 liter. Immediately prior to use pour desired amount into a small container and add a few drops of 30 volume % hydrogen peroxide.

Store benzidine stain at 4 C.

chromatograph.

The separation gradient was prepared with 0.1% aqueous TFA (trifluoroacetic acid) and 0.1 % TFA in acetonitrile. All reagents and H_2O were of HPLC-Grade. Two developers were used: 1) Mixture A was 80 % aqueous TFA and 20 % of 0.1 % TFA in acetonitrile; and 2) Mixture B was 40 % aqueous TFA and 60 % of 0.1 % TFA in acetonitrile. The samples were developed using a gradient from 44 to 56.5 % Mixture B in 60 min at a flow rate of 1 ml/min at ambient temperature. Prior to use the column was equilibrated for 30 min at 44 % Mixture B. Between uses the column was stored in 100 % acetonitrile.

2.C. DNA Analyses

2.C.1. High Molecular Weight Genomic DNA Isolation

The frozen buffy coats were thawed slightly then washed in 10 ml of sterile Krebs-Ringer-Phosphate buffer (KRP: 0.127 M NaCl, 0.01 M Na_2HPO_4/NaH_2PO_4 pH 7.4, 5 mM KCl, 1 mM $MgSO_4$ $7H_2O$, 1 mM $CaCl_2$). The cells were centrifuged in sterile 50 ml polypropylene round-bottom centrifuge tubes at 14,500 x g (11,000 rpm, SS-34 rotor) for 10 minutes at 4 C to obtain a nuclear pellet. The supernatant was discarded.

DNA was isolated from the buffy coats according to a protocol described by A. Dozy (personal communication). The cells were transferred to sterile 125 ml screw-cap Erlenmeyer flasks (Kimax) in a

total volume of 25 ml STE (0.1 M NaCl, 0.05 M Tris-HCl pH 7.5, 1 mM Na_2EDTA pH 7.4). Ten percent sodium dodecyl sulfate (SDS) was added to a final concentration of 0.5 %. Proteinase K (10 mg/ml in 10 mM Tris-HCl, pH 7.5; Boehringer-Mannheim Biochemicals, Indianapolis, Indiana) was added to a final concentration of 100 μg/ml. The mixture was incubated overnight at 50-55 C with moderate shaking in a Gyrotory waterbath (New Brunswick Scientific Co., New Brunswick, New Jersey).

No evidence of solid material was present when lysis was sufficient. If solid material remained after an overnight incubation, an additional 0.25 ml proteinase K (10 mg/ml) was added. Shaking at 50-55 C was then continued for a further two hours. All refractory samples were lysed by this extension.

The extraction mix was allowed to equilibrate to room temperature. One-half volume of redistilled phenol (International Biotechnologies, Inc., New Haven, Connecticut) and one-half volume of Sevag's reagent (chloroform : isoamyl alcohol ; 24:1) were added. The samples were mixed gently at room temperature on an orbital shaker for 10 minutes. The samples were then incubated on wet ice for 10 minutes.

The mixture was separated into aqueous and organic phases by centrifuging at 1,400 x g, 4 C, for 10 minutes in a Beckman Model TJ-6 centrifuge with a fixed TH-4 rotor (2,800 rpm; Beckman Instruments Inc., Palo Alto, California) in sterile 50 ml polypropylene conical centrifuge tubes. The supernatant aqueous layer was transferred to fresh 50 ml

centrifuge tubes using wide-bore plastic pipettes. The protein interface and organic phase were discarded.

The DNA samples were precipitated by adding one-tenth volume 2 M sodium acetate and 2-1/2 volumes of absolute ethanol at -20 C. The samples were mixed well and stored at -20 C overnight. The DNA was pelleted by centrifuging at 1,100 x g, 4 C, for 15 minutes (2,500 rpm; TJ-6). The pellet was washed once with 70 % EtOH at -20 C, the centrifugation step repeated, and then the pellet drained on paper towels until dry. The DNA for each sample was resuspended in a total volume of 10 ml 10 mM TE (10 mM Tris-HCl pH 7.5, 10 mM Na_2EDTA pH 7.4) at 37 C. DNase-free RNase (10mg/ml, DNase heat-inactivated; Cooper Biomedical Co., Garden Grove, California) was added to a final concentration of 100 μg/ml then incubated for two hours at 37 C in a dry incubator. Proteinase K (10 mg/ml) to a final concentration of 100 μg/ml and 10 %(w/v) SDS to a final concentration of 0.5% were added. The mixture was incubated in a Gyrotory water bath for 1 hour at 50-55 C with gentle shaking.

The samples were re-extracted as described above using one-half volume redistilled phenol and one-half volume Sevag's reagent. DNA was precipitated from the supernatant aqueous layer using one-tenth volume 2 M sodium acetate or one-half volume 7.5 M ammonium acetate and 2-1/2 volumes absolute ethanol at -20 C. After mixing well the samples were stored at -20 C overnight. The DNA was pelleted as described above then resuspended in 1.0 ml TE-4 buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM

Na_2EDTA).

The resuspended DNA was dialyzed against 1 liter of 1:0.1 TE (1 mM Tris-HCl, pH 7.5; 0.1 mM Na_2EDTA, pH 7.5) at 4 C. The sample was placed in rinsed nuclease-free dialysis tubing (Spectrapor 2 MWCO 12-14,000, Spectrum Medical Industries, Los Angeles, California). Appendix B outlines the method for tubing preparation. The ends were sealed with sterile plastic clamps (Spectrum Medical Industries). Up to four samples were dialyzed at a time for a period of 6 hours to overnight in 3 changes of buffer. The dialysis buffer was stirred with a battery-operated magnetic stirrer (Cole-Parmer Co., Chicago, Illinois). The DNA was transferred to a sterile 1.5 ml Eppendorf tube after dialysis by removing the clamps and "milking" the DNA solution from the tubing.

2.C.2. Quantitation of DNA

The amount of DNA isolated from each buffy coat was determined by spectrophotometric analysis. A 1:20 dilution of the DNA preparation was made using TE-4 buffer. The dilute DNA sample was scanned over the range of 220 to 310 nm using the computer-assisted scanning mode of an Hitachi spectrophotometer model 100-80 (Hitachi Instrument Co., Mountain View, California). The results were plotted with the Hitachi plotter. The general shape of the curve was noted and those samples displaying unusual conformations, e.g. peaks at 250 nm, were re-dialyzed (see Section 2.C.1.) to remove residual phenol and contaminants.

Spectrophotometric plots were used to determine optical density readings at 230 nm (OD_{230}), 260 nm (OD_{260}), and 280 nm (OD_{280}). The concentration of DNA was determined directly from the OD_{260} reading. The concentration of double-stranded DNA at 260 nm is 50 μg/ml for 1.0 OD (Maniatis et al.,1982). With a dilution factor of 1:20, the DNA concentration was determined by multiplying the OD_{260} by 1,000.

In addition to visual inspection of the location of the peak of the curve on the spectrophotometric plots to determine purity, ratios of the optical density readings were also determined. A ratio of OD_{260}:OD_{280} was used to determine RNA contamination (ratios of 1.7 to 1.8 were considered acceptable). A ratio of OD_{260}:OD_{230} was used to determine the amount of protein contamination (ratios >2.0 were considered acceptable).

2.D. Probe Preparation

2.D.1. Source

Genomic probe pβPstd was kindly provided by Dr. Tom Maniatis of Harvard University. Plasmid pβPstd contains a 4.4 kb PstI fragment of human genomic beta-globin isolated from lambda HβGI and sub-cloned into the vector pBRd, a pBR322 derivative. The host cell for pβPstd is *E. coli* DHI. The probe pβPstd was obtained as a stab culture by air mail.

Genomic probe δ-probe was kindly provided by Dr. Mitchell Golbus of

the University of California at San Francisco. Plasmid δ-probe contains a 2.3 kb Pst I fragment of human genomic delta-globin which was subcloned in the vector pBR322. Delta-probe was obtained as a purified insert and plasmid DNA by air express on dry ice. The plasmid was transformed into E. coli HB101 competent cells (Bethesda Research Laboratories, Bethesda, Maryland) according to the manufacturer's small-scale transformation protocol.

2.D.2. Mini-Plasmid Preps

Plasmid DNA was prepared according to the method of Karen Zachow PhD (personal communication). Five 8 ml cultures (40 ml Luria Broth + 400 μl 10% Glucose + 40 μl 1,000X Tetracycline[2]) were prepared in 15 ml polypropylene tubes (Sarstedt, Princeton, New Jersey). One culture was reserved as a control. The remainder were inoculated with single colonies from the tetracycline plate described in section 2.D.3. below.

The cultures were incubated at 37 C overnight with shaking at 100 rpm in a dry Gyrotory incubator. Aliquots were removed for glycerol cultures (250 μl 80 % glycerol + 500 μl overnight culture) and stored at -20 C. The remaining culture was incubated for a further 24 to 48 hours.

After the cultures were grown the cells were pelleted by

---

[2] 1000X Tetracycline is 12.5 mg/ml in 50 % EtOH. It should be prepared immediately prior to use.

centrifugation at 7,300 x g (5,000 rpm, Sorvall Model RC3b) for 15 minutes. The supernatant was decanted and the cells kept on ice as they were resuspended in 200 μl Sucrose-Tris (25 % sucrose, 0.05 M Tris-HCl pH 8.1). After the suspended cells were transferred to a fresh 4 ml polypropylene tube (Sarstedt), 40 μl of freshly prepared cold Lysozyme solution (10 mg Lysozyme per ml Sucrose-Tris) was added. The cells were incubated on ice for 5 minutes then 80 μl of 0.25 M $Li_{3.6}EDTA$, pH 8.0[3] was added. After a further 5 minutes incubation on ice, 100 μl of 5 M LiCl and 40 μl 10 % SDS were added. The tube was immediately vortexed then placed at 4 C for 30 minutes or up to overnight.

The resultant viscous solution was then centrifuged at 35,600 x g (17,000 rpm, SM-24 rotor) for 60 minutes. The supernatant was decanted into 1.5 ml Eppendorf tubes and incubated at 65 C for 30 minutes. The remaining cellular debris was pelleted after centrifugation at 6,200 x g for 10 min in a microfuge (10,000 rpm; Beckman Model 12, Beckman Instruments Inc., Palo Alto, California). The supernatant was transferred to fresh 4 ml Sarstedt tubes where 50 μl Pronase (10 mg/ml; Boehringer-Mannheim Biochemicals, Indianapolis, Indiana) was added prior to incubation at 37 C for 3 hours. Six volumes (3.6 ml) of ethanolic

---

[3] 0.25 M $Li_{3.6}EDTA$, pH 8.0 is prepared by neutralizing the free acid of EDTA with solid LiOH H_2O to pH 8.0.

Per liter:
37.6 g LiOH H_2O (Aldrich Chemical Co., Milwaukee, Wi.)
73.2 g Free Acid EDTA (Sigma Chemical Co., St. Louis, Mo.)

perchlorate reagent (EPR)[4] were added and the DNA precipitated at 4 C for 30 minutes to overnight.

The DNA was pelleted by centrifugation at 7,300 x g (5,000 rpm, Sorvall RC3b) for 15 minutes. The supernatant was drained and the DNA resuspended in 1 ml of Low-Salt-Buffer (LSB: 1 mM Tris-HCl pH 7.4, 1 mM NaCl, 0.1 mM Na_2EDTA). Four μl of RNase (10 mg/ml with DNase heat inactivated; Cooper Biomedical, Garden Grove, California) were added and the solution was incubated at 37 C for 3 hours to overnight.

The RNase was removed by Sevag's reagent (24:1, chloroform:isoamyl alcohol) extraction. One ml of Sevag's reagent was added, the mixture rocked gently by hand for 5 min, and the phases separated by centrifugation at 7,300 x g (5,000 rpm, RC3b) for 15 min. The aqueous phase (top layer) was transferred to a fresh 4 ml Sarstedt polypropylene tube using a sterile Pasteur pipette. The DNA was precipitated by adding one-half volume (500 μl) of 7.5 M ammonium acetate and 2 volumes (3 ml) of 100 % EtOH and incubating overnight at -20 C. The DNA was pelleted by centrifuging for 25 min at 5000 rpm (Sorvall RC3b). The

---

[4] Ethanolic Perchlorate Reagent (EPR) was prepared as follows:

Heat 53.6 ml dH_2O to 60 C. Add 122.44 g anhydrous $NaClO_4$ (Aldrich Chemical Company, Milwaukee, Wisconsin). Stir on magnetic stir plate under hood until dissolved. Filter solution through 0.2 μm glass-fiber filter (Millipore Corporation, Bedford, Massachusetts), followed by 381 ml prewarmed 60 C 100 % EtOH.
Store under hood at room temperature.

NOTE: EPR is explosive. Never heat directly or exceed temperature of 60 C. Wear gloves and goggles when handling.

supernatant was decanted and the pellet dried by vacuum-dessication for 1 hour. The DNA was suspended in 40 μl LSB.

Yields of DNA from mini-preps were determined by running 4 μl aliquots of each isolate on 1 % agarose (IBI: International Biotechnologies Inc., New Haven, Connecticut) gels at 40 V overnight. Known amounts of salmon-sperm DNA (Sigma Chemical Co., St. Louis, Missouri) were run on the same gel as standards. The gels were stained with a solution of Ethidium Bromide (5 μg/ml) and the intensity of UV-fluorescence visually compared to determine concentrations.

The plasmid isolate which had the greatest intensity of plasmid bands as compared to bacterial genomic DNA was selected for large-scale purification.

2.D.3. Plasmid Verification

2.D.3.a. Antibiotic Resistance Verification

Three types of 1.5 % agar Luria-Bertani plates were prepared. Viability of the stab culture was verified by streaking on plates without antibiotics. Tetracycline resistance was determined by streaking on plates containing 15.0 μg/ml of Tetracycline (Sigma Chemical Co., St.Louis, Missouri). Ampicillin sensitivity was determined by streaking on plates containing 50 μg/ml of Ampicillin (Sigma Chemical Co.).

2.D.3.b. <u>Restriction Endonuclease Mapping</u>

A map of the plasmid pβPstd was kindly provided by Dr. Tom Maniatis. Verification of this map was attempted by performing a series of single and double digests using the restriction enzymes BamHI, EcoRI, HindIII, and PstI (International Biotechnologies Inc.) at 5 U/μg. Fragment sizes were determined by agarose gel electrophoresis through a 1 % agarose gel as described in section 2.E.2. and the sizes compared to those expected. A similar protocol was followed in order to verify the delta globin probe.

2.D.4. <u>Large-Scale Plasmid Preparation</u>

A suitable plasmid isolate was obtained as described in sections 2.D.2. and 3. above.

2.D.4.a. <u>Culture</u>

The plasmid preparation began with a starter culture. In a sterile 300 ml culture flask (Pyrex) were placed 75 ml Luria broth [10 g Bacto-tryptone (Difco), 5g Bacto-yeast extract (Difco), 5g NaCl per liter], 750 μl 10 % α-D+-glucose and 75 μl of 12.5 mg/ml Tetracycline (Sigma Chemical Co.). Five μl of glycerol culture (Section 2.D.2.) was added and the culture incubated with shaking (100 rpm, New Brunswick Scientific Psychrotherm Incubator, New Brunswick, New Jersey) at 37 C overnight.

Six liters of Luria broth in three Fernbachs flasks (Pyrex) were steam sterilized. Twenty ml of separately autoclaved 10 % glucose was added to each flask in addition to 2 ml of tetracycline (12.5 mg/ml in 50 % EtOH, prepared freshly). An aliquot of the culture media was removed to serve as a zero point for the Klett meter (Klett-Summerson Photoelectric Colorimeter Model 800-3 Klett Mfg. Co., NY, NY.). The cultures were grown to a Klett reading of 70 by shaking at 100 rpm at 37 C. After the cultures had reached a Klett of 70, the plasmid replication was amplified by the addition of 4 ml CAM [10 per cent Chloramphenicol (Sigma Chemical Co.) in absolute ethanol]. The cultures were left shaking at 37 C overnight.

2.D.4.b. <u>DNA Isolation</u>

The cells were pelleted by centrifuging in sterile 1-liter polypropylene bottles (Nalgene) at 7,300 x g for 15 min (5,000 rpm, Sorvall RC3b). The supernatant was decanted and treated with 50 % household bleach prior to disposal. The cells were maintained on wet ice. A total of 48 ml of 25 % sucrose, 0.05 M Tris pH 8.1 at 0 C was used to resuspend the <u>E. coli</u>. The cells were pooled in a fresh 250 ml polypropylene centrifuge bottle (Nalgene) while on ice. Immediately prior to use, 0.06 g Lysozyme (Sigma Chemical Co.) was dissolved in 3 ml cold Sucrose-Tris. This was added to the pooled cells and left on ice for an additional 5 min. The lysozyme was followed by 19.2 ml 0.25 M $Li_{3.6}$EDTA pH 8.0 (see footnote 2, Section 2.D.2.) and left on ice for 5 min. Finally, 24 ml 5 M LiCl and 9.6 ml 10 % SDS were added

simultaneously and the mixture was shaken vigorously by hand to produce a viscous solution which was stored at 4 C overnight.

The viscous solution was divided among sterile 50 ml round-bottom polypropylene centrifuge tubes and centrifuged at 35,000 x g (17,000 rpm; SS-34 rotor) for 60 min at 4 C. The supernatant was decanted into fresh tubes and incubated at 65 C for 30 min. The tubes were then centrifuged at 10,000 rpm for 10 min and the resultant supernatant was decanted into a fresh sterile 1-liter polypropylene centrifuge bottle to which 10.5 ml of Pronase (2 mg/ml, Boehringer-Mannheim Biochemicals, Indianapolis, Indiana) was added. The solution was incubated at 37 C for 3 hours before adding 2 volumes of 100 % ethanol and storing at -20 C overnight.

After overnight precipitation, the pronase digest was pelleted by centrifugation at 7,300 x g for 25 min (5,000 rpm, Sorvall RC3b). The supernatant was decanted and the pellet was allowed to air-dry. The pellet was then dissolved in 60 ml Low-Salt-Buffer (LSB: 1 mM Tris-HCl pH7.4, 1mM NaCl, 0.1 mM EDTA) and 83.4 g sodium perchlorate (anhydrous, Aldrich Chemical Co.) was added. The mixture was incubated with occasional shaking at 50 C in a water bath in a fume hood until dissolved. After cooling to room temperature, 159 ml of room temperature absolute ethanol was added and the solution was incubated at 4 C for 30 minutes.

The precipitated DNA was pelleted by centrifuging 20 min at

7,300 x g (5,000 rpm, Sorvall RC3b). After the supernatant was decanted and the pellet dried, the DNA was dissolved in 30 ml LSB. A total of 150 μl RNase (10 mg/ml, Cooper Biomedical Co., Garden Grove, California) was added and the whole incubated at 37 C for 3 hours with gentle shaking. The reaction mix was left at 4 C overnight or extracted immediately with Sevag's reagent. After 5 min of gentle shaking by hand, the phases were separated by centrifugation at 7,300 x g for 10 min (5,000 rpm, RC3b) in 50 ml polypropylene centrifuge tubes. The aqueous phase was removed to fresh tubes and the process repeated until no precipitated interphase was present. The aqueous phases from all tubes were pooled in a sterile 250 ml polypropylene centrifuge bottle and 2 volumes of 100 % ethanol at -20 C were added. The DNA was allowed to precipitate at -20 C for 30 min to overnight.

The DNA was pelleted by centrifuging at 7,300 x g for 25 min (5,000 rpm, RC3b), decanted, drained, and resuspended in 30 ml LSB. One-half volume of 7.5 M ammonium acetate and twice the final volume of 100 % EtOH at -20 C was added. After precipitating overnight at -20 C, the DNA was pelleted by centrifuging at 16,300 x g for 15 min (10,000 rpm, HB4 rotor). The supernatant was decanted and the pellet dried before it was resuspended in 4 to 24 ml High-Salt-Buffer (HSB: 25 mM Tris-HCl pH 7.4, 100 mM NaCl, 2 mM Na_2EDTA). An estimate of the DNA concentration was obtained as described in section 2.D.2.

2.D.4.c. Cesium-Chloride Gradient Purification

The plasmid DNA isolated according to the large-scale plasmid preparation protocol was further purified by cesium-chloride gradients. One to five mg of plasmid DNA in HSB was placed in a tared 15 ml polystyrene centrifuge tube. The solution weight was adjusted to 4.25 g using HSB. To each tube, 3.97 g of CsCl was added and allowed to dissolve completely. Ethidium-bromide (216 μl of 10 mg/ml) was added and the whole transferred to Quick-Seal™ polyallomer ultracentrifuge tubes (Beckman Instruments Inc., Palo Alto, California) using Pasteur pipettes. The tubes were balanced against each other then heat-sealed using a Beckman heat-sealer.

The plasmid DNA was separated from the bacterial genomic DNA by centrifugation at 184,400 x g (45,000 rpm) and 20 C, for 24 hr in a Beckman model L8-70M ultracentrifuge using a VTi65.2 rotor. The plasmid DNA was visualized using a short-wave ultraviolet light box (Ultra-Violet Products Inc., San Gabriel, California). The lower band, which represented the plasmid DNA, was removed using a 1 ml tuberculin syringe with a 25 g 5/8 inch needle and placed in a new polyallomer tube. Stock solution (4.25 g HSB + 3.97 g CsCl + 216 μl EtBr) was added to a total weight of 4.25 g and the process repeated.

The plasmid DNA recovered was dialyzed as described in section 2.C.1. and quantitated as in Section 2.C.2.

2.E. Southern Blot Analysis

2.E.1. Restriction Endonuclease Digestion of Genomic DNA

The restriction enzymes used in this study and their sources are indicated in Table 2.2. Five to 6.5 μg of high-molecular-weight (HMW) DNA was allowed per digest based on the yield of the individual DNA preparation. The digests were completed in two phases.

In the first phase all single digests were performed in addition to the first digest of all double digests. In general, the restriction enzyme requiring the lower salt concentration for reactivity was chosen as the first digest. For each individual, all first digests requiring the same enzyme were pooled in a single 1.5 ml Eppendorf tube and digested according to a batch process following the manufacturer's recommendations. Five to seven units of enzyme were used per microgram of DNA to be digested. In the case of BamHI and KpnI enzymes this amount was increased to at least 15 U/μg. The total volume of the reaction was based on a volume twenty times that of the total enzyme volume. The reactions were carried out in a 37 C dry incubator for 4 hours to overnight.

At the end of the digest period, 0.5 to 1.0 μg of DNA was removed from each digest and electrophoresed on an 0.8 % agarose gel. See Section 2.E.2. below for details of this procedure. If the representative DNA demonstrated that the digest was complete, the

remaining DNA was re-aliquoted among the appropriate number of tubes and precipitated with 1/2 volume 7.5 M ammonium acetate and twice the total volume of 100 % ethanol at -20 C overnight or at -70 C for 1 hour. The DNA was then pelleted by centrifuging 8,800 x g for 15 min in an Eppendorf centrifuge model 5413 (Brinkmann Instruments, Westbury, New York). The supernatant was decanted and the pellet allowed to air-dry prior to being resuspended in 20 μl TE-4 (10 mM Tris-HCl pH7.5, 0.1 mM Na_2EDTA).

In the event that the digest was incomplete, the amount of enzyme was increased and the volume readjusted. The process was repeated until the test gels demonstrated complete digests.

The panel of second enzyme digests was carried out on an individual tube basis. One microgram of undigested DNA was restricted in a separate tube for each enzyme in order to demonstrate effectiveness of the digest on agarose gel electrophoresis. When the digests were complete, the DNA was precipitated and resuspended in TE-4 as outlined above.

When appropriate, the specificity and nuclease-contamination of the restriction enzymes were tested using lambda DNA (International Biotechnologies, Inc.) and digesting under the same conditions (i.e. enzyme concentration, reaction volume, and time of incubation) as described above.

2.E.2. Agarose Gel Electrophoresis

Table 2.3 demonstrates the pattern of agarose gel electrophoresis of the completed digests. Two gels were run for each individual carrying single and double digests in the presence of two size markers: lambda restricted with HindIII (Bethesda Research Laboratories; BRL) and lambda restricted with HindIII and EcoRI (International Biotechnologies, Inc.).

The samples were electrophoresed through 0.8 % agarose gels in order to maximize separation of fragments in the size range of 0.6 to 9 kb (Maniatis et al.,1982). The electrophoresis chambers were BRL model H4 with a 20 cm wide by 25 cm long tray. The comb was of Dupont DeLrin™ with 20 teeth 9 mm wide by 1 mm deep. The electrophoresis buffer was made up fresh by a 1:20 dilution of 20X stock buffer (8 M Tris-Acetate, 10 M Na-Acetate, 0.2 M Na_2EDTA; pH 7.80).

The agarose gels were prepared using Type I: LEEO low-melting point agarose (Sigma Chemical Company, St. Louis, Missouri) and 1X electrophoresis buffer (0.4 M Tris-Acetate, pH 7.8; 0.5 M NaAcetate, 10mM Na_2EDTA). The gel was boiled for 5 min in a Sears Kenmore microwave oven at full power. Evaporative water loss was replaced before gels were poured into a Time-tape™ sealed tray. A total volume of 300 ml agarose solution was used to pour gels which were 5 mm thick.

After the gel had set the tray was placed in the BRL electrophoresis chamber and submerged under 7 mm of 1X electrophoresis buffer. The 20 μl samples of restricted DNA were loaded using 10 μl Sugelpot as the

loading dye (60 % sucrose, 0.1% SDS, 5 mM Na_2EDTA). The gels were run for 36 hours at 20 V and room temperature. The level of the electrophoresis buffer was monitored to assure that it did not drop below the level of the gel surface.

Upon completion of electrophoresis the agarose gel was stained for 15 min in 500 ml ethidium bromide (5 μg/ml made up fresh from 1000X stock). It was destained for an equivalent time period in dH_2O. The gel was viewed on a short-wave UV-light box (Ultraviolet Products Inc., San Gabriel, California). Migration of the lambda markers was determined by placing parafilm™-wrapped transparent rulers adjacent to the appropriate lanes and photographing the result. The camera used was a Polaroid MP4 and the film Polaroid Type 57. Photographs were taken using a Kodak Wratten 21 gelatin filter. Optimal resolution was obtained at a setting of f16 for 8 seconds.

2.E.3. Southern Transfers

DNA was blotted onto Genescreen-Plus™ nylon membrane (New England Nuclear, Boston, Massachussetts) from agarose gels using the alkaline transfer method as outlined by Chromczynski and Qasba (1985). The method involves capillary transfer using Transfer Solution (TS: 0.4 N NaOH, 0.6 M NaCl).

After photographing the gel (see Section 2.E.2.), it was denatured for 30 min in TS with gentle agitation. The gel was trimmed to a length

of 20 cm from the base of the wells and was covered with a 20 x 20 cm sheet of Genescreen-Plus™ with the concave side B facing the gel. The gel was supported by Whatman 3MM chromatography paper (Whatman Ltd., Maidstone, England) extending over a glass plate into a liter of TS. The membrane was covered by 4 layers of Whatman 3MM paper cut to 20 x 20 cm pieces and wet with TS. The sides of the gels were sealed using Handi-Wrap™ (Dow Chemical Company, Indianapolis, Indiana). Three inch stacks of paper towels were used to promote capillary transfer. A glass plate and a 500 g weight were placed on top of the assembly.

The transfers were complete after 16 hours and the membranes were neutralized by soaking for 15 min in neutralizing solution (0.5 M Tris-HCl pH 7.0, 1 M NaCl) with gentle agitation. The membranes were dried at room temperature between layers of Whatman 3MM paper.

2.E.4. Radioactive Labeling of Probe

The probes were routinely labelled to 10^8 dpm/μg using a BRL Nick-Translation Kit (Bethesda, Maryland). Manufacturer directions were followed in order to label a 1 μg aliquot of probe DNA. High specific activities were obtained by doubling the recommended quantity of solution C (DNA Polymerase I and DNase I) in the presence of 100 μCi α-^{32}P dCTP (Karin Rodland PhD, personal communication). Radioactive nucleotides were purchased from New England Nuclear (Boston, Massachussetts).

Specific activities and percentage incorporation of ^{32}P were determined by Cerenkov counting using the ^{3}H channel (Berger and Krug, 1985) on a Packard Tri-Carb 460C scintillation counter (United Packard Technologies Instrument Company Inc., Downers Grove, Illinois). The DNA aliquots were spotted on Whatman GF/B glass-fiber discs and counted in Kimax polyethylene scintillation vials in the absence of scintillation fluid.

Unincorporated nucleotides were removed by precipitation of the probe with 1/2 volume 7.5 M ammonium acetate and twice the final volume of 100 % EtOH at -20 C. Twenty μg of sheared salmon-sperm DNA (Sigma Chemical Co.) was added prior to precipitation. After 2 hr to overnight at -20 C, the probe was pelleted by centrifuging at 8,800 x g for 10 min in an Eppendorf Model 5143 centrifuge (Brinkmann Instrument Co.). The supernatant was drained, the pellet was air-dried, and resuspended in 20 μl of 1 M NaOH. After 5 min the probe was resuspended in an additional 400 μl of TE buffer. Prior to use, the probe was boiled for 10 min in a Temp-Blok heater (American Scientific Products, McGaw Park, Illinois) then quenched on ice. Immediately prior to use the probe was centrifuged for a few seconds to pellet any contaminants.

2.E.5. Blot Hybridization

The blots were hybridized according to the manufacturer's recommendations for Genescreen-Plus™ (New England Nuclear). The formamide method was used.

The blots were stored in 10 inch Dazey[R] Seal-A-Meal™ pouches. Nineteen ml of prehybridization solution[5] was added. Bubbles were removed by rolling the hybridization bag with a 5 ml pipette then the bag was double-sealed using a Dazey[R] Seal-A-Meal™ heat sealer. The blots were prehybridized at 42 C with gentle agitation (30 rpm) in an Orbit Microprocessor Shaker Bath (Lab-Line Instrument Co., Melrose Park, Illinois) for 2 to 6 hours.

Denatured radioactive probe was added to the reserved 1 ml of prehybridization fluid. Sufficient probe was used to obtain a final concentration in the bag of 4 x 10^5 dpm/ml. The probe was drawn-up into a 1 ml tuberculin syringe with a 25 g x 5/8 in needle. When prehybridization was completed, the bags were tilted such that the prehybridization fluid collected in one corner. The probe was then injected into the pocket of fluid (Leigh Nicholaison, personal communication) and the bag re-sealed beyond the injection site. The bag was then returned to the shaker bath at 42 C with continued agitation for an additional 6 to 24 hours.

At the completion of hybridization, the blots were removed from the Seal-A-Meal™ bags and washed according to the manufacturer's protocol

---

[5] Prehybridization solution was prepared using a modification of the recommended technique as follows:

For 20 ml:
Mix 10 ml deionized formamide (Sigma Chemical Co.), 4 ml 50 % dextran sulfate (Sigma Chemical Co.), and 2 ml 10 % SDS by inversion. Incubate at 42 C for 10 to 15 minutes. Add 4 ml 5 M NaCl to the tube and mix by inversion. Incubate for a further 10 to 15 minutes. Centrifuge the buffer for 5 min at 2800 rpm.

with minor modifications: 1) 2 x 200 ml of 2XSSC (0.3 M NaCl, 0.03 M NaCitrate) at room temperature for 5 min with moderate agitation; 2) 4 x 500 ml of a solution containing 2XSSC and 1.0% SDS at 55-58 C for 15 min with constant agitation (80 rpm); and 3) 2 x 500 ml of 0.1XSSC (0.015 M NaCl, 1.5 mM NaCitrate) at room temperature for 30 min with gentle agitation.

See Section 2.E.6. for autoradiography set-up.

Blots were rehybridized after the following manufacturer's protocol: 1) The Genescreen-Plus™ was incubated in 400 ml of 0.4 N NaOH at 42 C for 30 min with gentle agitation (30 rpm); then 2) incubated in 500 ml of 0.1XSSC, 0.1% SDS, 0.2 M Tris-HCl, pH7.5, at 42 C for an additional 30 min. If an overnight autoradiograph showed that sufficient probe had been removed, the blot was prehybridized and hybridized as described above.[6]

If an overnight autoradiograph did not show that sufficient probe had been removed an alternate stripping protocol was followed, that of Gatti et al (1984). One liter of 0.1 X SSC with 0.1 % SDS was prepared. One-half of the solution was heated to 95 C and poured on the blotting membrane which was then gently shaken at room temperature for 20 min. The wash solution was then discarded and the procedure repeated. The membrane was then pre-hybridized as detailed above.

---

[6] NOTE:Care was taken to prevent the membrane from drying out completely at any stage of the prehybridization through rehybridization process to prevent irreversible binding of the probe to the membrane.

2.E.6. Autoradiography

The membranes were removed from the final 0.1X SSC wash and the excess moisture blotted onto Whatman 3MM paper. The moist blots were then wrapped in Handi-Wrap™ and taped to paper backing. The paper was labelled using an Ult-Emit™ pen (New England Nuclear).

The blots were exposed to two sheets of Kodak 8"x10" XAR-5 film between two Dupont Cronex Lightning-Plus™ or two Kyokko Super HS™ intensifying screens in Kodak 8"x10" film holders at -70 C. After overnight exposure, one piece of film was removed to estimate total exposure necessary. Blots were exposed for 3 to 7 days. The x-ray film was developed using a Kodak RP X-OMAT autoprocessor.

2.F. Data Analysis

2.F.1. Hemoglobin Electrophoresis

The mobility patterns for each animal within a given species were visually compared to determine characteristic species and individual patterns. Human hemoglobins A,F,S, and C were run as standard markers. The mobility of monkey major hemoglobins was compared to that of human hemoglobin A. Those bands which were anodic to human A were referred to as "fast" while those more cathodic were "slow". Interspecies comparisons were performed when representative hemolysates were loaded on the same cellulose-acetate plate. Observed differences were

quantified according to the method of Schneider (1978, 1981).

2.F.2. HPLC Globin Chain Separation

The pattern of effluent peaks at 220 nm was compared for each species and for human hemoglobins A and A_2. The identity of individual globin peaks was based on comparisons to those identified for *Macaca nemestrina* by subsequent tryptic peptide analysis (D.M.Pillers, M.D., unpublished data). Relative percentages of globin chain expression were determined by comparing the area under the curves of the appropriate peaks.

2.F.3. Restriction Fragment Length Determination

The size of the individual restriction fragments as determined by autoradiography was calibrated using cubic-spline and least-squares linear regression analysis. These calculations were performed using the IBI-Pustell Sequence Analysis System and Database Manager in Cyborg Environment Program™ (International Biotechnoligies Incorporated). Autoradiography data was entered into the computer using an IBI Gel-Reader™ digitizer.

2.F.4. Restriction Fragment Map Construction

Restriction endonuclease maps of the monkey beta and delta globin genes were generated by determining the relative order along the genome

of the restriction fragments obtained in section 2.F.3. The accuracy of the maps was determined by comparing them to those previously generated for orthologous human (Collins and Weissman, 1984) and rhesus loci (Martin, 1980).

CHAPTER THREE: RESULTS

3.A. Protein Analysis

3.A.1. Cellulose-Acetate Electrophoresis

Hemolysates were prepared for each animal as described in section 2.B.1. All hemolysates were analyzed by cellulose-acetate electrophoresis against human standards (Hbs A,F,S,and C). The results of a representative cellulose-acetate plate are reproduced in Figure 3.1.

Relative electrophoretic mobilities of human mutant hemoglobins have been characterized by Schneider and Barwick (1978; Barwick and Schneider, 1980). The method is based on a comparison of the mobilities of human Hbs A and C which are run as standards, where HbA is assigned a value of "0" and HbC is assigned a value of "10". The distance between HbA and C under standard electrophoretic conditions is measured, as is the distance between HbA and the mutant. The two distances are related according to the following equation:

$$N = \frac{D\ (\ A \text{ to } X)}{D\ (\ A \text{ to } C)} \times 10$$

where N represents the ratio of the two values, and D the distances. Hemoglobins which have mobilities anodic to human HbA are designated (+)

while cathodic ones are (-).

Schneider's method was applied to the macaque samples analyzed. The values determined for the hemolysates electrophoresed in Figure 3.1 are given in Table 3.1.

Cellulose-acetate electrophoresis was used as a preliminary screening technique to identify those animals which had phenotypes representative of their species. Results for each species are summarized below:

1) Macaca fascicularis

Ten individuals were studied and two major adult hemoglobins were identified (Schneider's N = +0.7 and +5.0). Five individuals were homozygous for the slower band and five animals were heterozygotes.

In addition to the major hemoglobin bands, several minor bands were also detected when the samples were electrophoresed on Sepraphore™ cellulose polyacetate strips (Section 2.8.2.). From one to four additional bands were identified by this technique. Interestingly, three of the ten crab-eating macaques had a minor hemoglobin with a mobility like that of human HbA_2 (data not shown).

The Macaca fascicularis subjects displayed from two to six bands when both major and minor components were identified. No consistent pattern emerged as all ten animals had these hemoglobins in varying

combinations. On this basis, the species Macaca fascicularis was not analyzed at the DNA level due to a lack of "representative" animals.

2) Macaca fuscata

Ten individuals were studied. Only one major adult hemoglobin was identified (Schneider's N = 0). A single minor band with a mobility similar to human HbF was identified in all animals when hemolysates were electrophoresed on cellulose-polyacetate strips. All animals were deemed representative of the species at this level. Five animals were selected at random using the random number generator program of a Ti 57 calculator (Texas Instruments) for further study.

3) Macaca mulatta

Ten animals were studied and as in the case of the Macaca fuscata, only one major (Schneider's N = +0.4) and one minor hemoglobin were identified. The minor hemoglobin had a mobility similar to human HbF. Five animals were selected at random for further study.

4) Macaca nemestrina

Eight pig-tailed macaques were studied. A single major adult hemoglobin was identified (Schneider's N = +5.7). A single minor hemoglobin, also faster than human HbA (Schneider's N = +1.9), was identified in all animals. Five animals were selected at random for further study.

5) Macaca nigra

Ten Celebes apes were studied. Two major adult hemoglobins were identified (Schneider's N = -0.7 and +10.0). Five animals were homozygous for the slow variant, two for the fast, and three were heterozygous. A single minor band was also identified which resolved poorly from the major variants. Five animals were selected at random to represent the species. A representative of each phenotype was included in this number.

3.A.2. High-Performance Liquid-Chromatography

An implicit assumption made during this study was that no minor hemoglobin analogous to human HbA_2 was expressed in Old World monkeys. In light of the finding in Macaca fascicularis of a few individuals with a minor hemoglobin with mobility like that of HbA_2 (Section 3.A.1.), further characterization of these animals was undertaken.

High-performance liquid-chromatography was used due to its ability to quickly separate hemoglobins into globin chains. The patterns of separated globin chains were then examined in comparison with that of human adult hemolysate. A representative chromatograph of a Macaca fascicularis monkey is shown in Figure 3.2. The identity of the α- and β-globin peaks was identified by analogy to those of Macaca nemestrina which had been subjected to further structural analysis (D.M.Pillers, M.D., unpublished data). Aside from α- and β-globin peaks a smaller third peak eluting just prior to the beta globin peak was found. It was

present consistently in all macaque species studied. Although the protein isolated from this peak in Macaca nemestrina can be identified as a non-alpha globin upon analysis of tryptic peptides, it has no tryptic peptides that are distinctive for either human gamma or delta as opposed to human beta globins (D.M.Pillers, M.D., unpublished data).

3.B. **Probe Preparation**

The pβPstd isolate obtained by this laboratory was subjected to restriction enzyme analysis. The map obtained is presented in Figure 3.3 in comparison with the putative map provided by the Harvard laboratory.[1]

Two major differences are noted: The beta-globin insert is ligated in the reverse orientation, and the vector size of 4.4 kb is compatible with the use of pBR322 as the cloning vector rather than its 3.3 kb derivative, pBRd. One disadvantage of this larger vector size is that it was not possible to separate the vector from insert DNA by gel electrophoresis in order to obtain purified insert DNA.

The identity of the probe was confirmed by using it to probe human genomic DNA by Southern blot analysis. The results of these restriction mapping experiments have been comparable to the published maps of the beta locus (data not shown).

---

[1] Subsequent interactions with personnel in the Maniatis laboratory have confirmed that the pβPstd map provided was inaccurate.

3.C. Restriction Endonuclease Maps of the Macaque Adult Non-Alpha Loci

Restriction endonuclease mapping experiments were performed as described in section 2.E. A preliminary step involved in this procedure is agarose gel electrophoresis of DNA which has been digested with restriction endonucleases. Figure 3.4 shows examples of agarose gel electrophoresis for each of the species examined by DNA analysis. Although very little specific information can be obtained at this stage, it is clear that the macaque species as a group can be readily differentiated from humans by comparing the patterns produced by digestion of repetitive DNA.

Two of the original five species, Macaca fasicularis and Macaca nigra, were discontinued from the study prior to completion of this phase of the analysis. Macaca fascicularis was eliminated due to the high degree of heterogeneity identified at the protein level by hemoglobin electrophoresis (see section 3.A.). Macaca nigra was discontinued at a later date due to the development of an immunodeficiency syndrome among the animals at the ORPRC. Yields of DNA from stored buffy coats were low and illness in the colony precluded obtaining subsequent samples.

Restriction endonuclease mapping experiments were performed on five animals of each of the remaining species : Macaca fuscata, mulatta, and nemestrina. Figure 3.5 demonstrates a representative autoradiograph

from which restriction fragment data were obtained. Table 3.2 summarizes the restriction fragment sizes obtained for each species.

Results of restriction endonuclease experiments for the <u>Macaca fuscata</u>, or Japanese macaques, are presented as a common map in Figure 3.6. The overall organization of the restriction sites surrounding the beta and delta globin loci was extremely conservative: of the 30 enzyme recognition sites determined only five loci could be identified as being polymorphic. These polymorphic sites included BamHI and SacI sites immediately 5′ to the delta globin locus, and EcoRI, SacI, and PvuII sites 3′ to the beta locus. Haplotype analysis for the <u>Macaca fuscata</u> at these loci is also presented in Figure 3.6.

Results of the restriction mapping experiments for <u>Macaca mulatta</u> are summarized in Figure 3.7. These maps are highly conservative within the species yet ten out of thirty-two loci are polymorphic. Polymorphic sites include HindIII and BamHI sites immediately 5′ to the delta locus, a HindIII and PvuII site within the delta locus, a BamHI site 5′ to, and a second within the beta locus, and EcoRI, PvuII, and SacI sites 3′ to the beta locus.

A distinctive and unusual feature of the rhesus macaques was the presence an anomalous 2.2 and /or 2.4 kb band upon digestion with PvuII.[2] This band was present only when using the probe pβPstd. It was

---

[2] The anomalous band was present in SacI digests as a 4.4 kb fragment. When PvuII was used as a second digest the 4.4 kb fragment was replaced by the familiar 2.2 kb PvuII band.

not detected with the δ-probe, regardless of wash stringency. All other beta-globin bands demonstrated some degree of cross-hybridization with δ-probe when reduced stringency conditions were used. The band was present regardless of which enzyme was used to perform second digests. Finally, other fragment sizes detected in PvuII digests would preclude the presence of such restriction enzyme recognition sites within the vicinity of the genome detected by pβPstd. Due to the unusual features of this fragment and its total absence in the other macaque species studied, it was designated "anomalous".

A summary of the maps generated for Macaca nemestrina is presented in Figure 3.8. These maps are extremely conservative as only four enzyme sites were identified as being polymorphic. The polymorphic sites detected included HindIII and EcoRI sites in the 3' flanking region of the delta globin gene and two BamHI sites located in analogous positions 5' to the beta and delta loci. One of the pig-tailed macaques was also the carrier of an 0.8 kb deletion within the intergenic DNA between the beta and delta loci.

CHAPTER FOUR: DISCUSSION

4.A. Protein Data

Hemoglobin electrophoresis is an inexpensive and rapidly performed method of analyzing hemoglobin phenotypes. Single amino acid substitutions can readily be detected by a change in the mobility patterns observed. The main disadvantage of the technique is that the substitutions must involve changes in the overall charge structure of the molecule although it cannot differentiate the number of changes present (Jolly & Barnicot, 1966).

A preliminary phase of this thesis involved the use of hemoglobin electrophoresis as a screening technique. Animals were tested to ensure that they provided adequate representation of their species at the protein level. Data were collected for five different macaque species: *Macaca fascicularis*, *Macaca fuscata*, *Macaca mulatta*, *Macaca nemestrina*, and *Macaca nigra*.

Due to its convenience as a screening technique, many monkeys have been tested by investigators both in the laboratory setting and in the wild, e.g. Hoffman and Gottlieb (1969) and Ishimoto et al. (1970). Electrophoretic data have been published for all of the macaque species studied. The majority of the published studies have used starch-gel

electrophoresis as the method of choice. The results presented in this thesis were obtained using the newer technique of cellulose-acetate electrophoresis.

In spite of their presumed close relationships, the macaque species demonstrated variable phenotypes by electrophoretic methods. Both the Macaca mulatta and the Macaca fuscata[1] hemolysates demonstrated mobility patterns similar to that found for human hemoglobin A. The results were much more variable among the other species.

The Macaca nemestrina, like the Japanese and rhesus macaques, displayed a single hemoglobin phenotype.[2] Unlike the other species, however, Macaca nemestrina has a major hemoglobin which behaves differently on hemoglobin electrophoresis. It has a fast hemoglobin (Schneider's N = +5.7) which is similar in mobility to one of the major hemoglobin components of Macaca fascicularis (N = +5.0). Macaca nemestrina demonstrates that closely related species need not have the same protein electrophoretic mobility. Species-differences can indeed be identified at the protein level.

---

[1] Previous studies of Macaca fuscata have demonstrated as many as six different hemoglobin types (Massa, 1981; Ishimoto, 1975). No variation between animals was detected in the hemoglobins of the ORPRC colony used in the current study.

[2] It has been shown that some variation in hemoglobin structure does occur in Malaysian Macaca nemestrina and is due to variant alpha chains (Nute & Pataryas, 1974; Nute, 1974; Mahoney and Nute, 1979; Nute et. al., 1981).

Electrophoretic results for the _Macaca nigra_ also provide interesting insights. Two major adult hemoglobins were detected. The differences between the two hemoglobins have previously been shown to be the result of a single amino acid substitution due to a two-nucleotide change in one codon of the beta globin chain (Murata and Thompson, 1976). The co-segregation of these hemoglobins in the ORPRC colony has previously been described (Howard,1982). This species demonstrates that such variations can occur, and achieve polymorphic status without noticeable effect on the health status of the population.

Macaca fascicularis demonstrated the most heterogeneity at the protein level. Two major adult hemoglobins were identified.[3] The slower hemoglobin was found both in homozygous and heterozygous individuals. The faster hemoglobin was found only in heterozygotes. Several minor hemoglobins were also identified which occurred in various combinations

---

[3] The major hemoglobins of _Macaca fascicularis_ have been studied previously by electrophoretic techniques (Tuttle et. al., 1961). Three major hemoglobins were detected by that study. Type A had a mobility similar to that of human HbA. The other major hemoglobins were both fast and were designated as types "B" and "C". The mobilities of these two hemoglobins are distinct but similar. As the electrophoretic techniques vary considerably, it is not possible to determine which of these fast types is the same as the fast variant described in the current study.

with the major hemoglobins.[4] No individual could be identified who was clearly "representative" of the species at all band positions.

It has been assumed that closely related species will share highly conserved proteins. If this is the case, then a single species should demonstrate conservation in hemoglobin phenotype. Man provides an example of this assumption. Literally thousands of individuals have been examined at the protein level using hemoglobin electrophoresis. The vast majority of individuals have demonstrated a single hemoglobin phenotype. Variant individuals have been sufficiently rare that they are designated as "mutants" regardless of the effects of the substitution upon their state of well-being.

The *Macaca fascicularis* results indicate that not all species are characterized by a single dominant hemoglobin phenotype. Many variations can actually co-exist within a healthy population.

The protein data combine to suggest that not only do variations occur between closely related species but that a significant number of these variants can be readily detected by hemoglobin electrophoresis.

---

[4] Minor hemoglobins have previously been described in *Macaca fascicularis* (Wade et. al., 1970). Two of these minor variants have mobilities similar to that of human HbA_2. The slower of these two variants, designated $Hb\text{-}X^{mi}$, was characterized and found to differ from $Hb\text{-}A^{mi}$ in the alpha-globin chain. The other HbA_2-like electrophoretic mobility variant was not characterized. The uncharacterized variant appears to be the one involved in the ORPRC animals studied. Exact comparisons are uncertain however, due to differences in the electrophoretic techniques used.

4.B. DNA DATA

4.B.1. Overall Similarities

Restriction endonuclease mapping of globin genes is a relatively recent addition to the armamentarium of the molecular evolutionist. Only a handful of species have been thoroughly characterized at this level (Section 1.D.2.). Although all of the globin systems function in similar ways, the degree of variability which has been detected at the gene level has been remarkable both with respect to the number of gene loci and with respect to the functional gene organization. The more similar two species are, the more similar the organization of their globin gene clusters appears to be.

The results presented in Chapter Three for the three macaque species studied at the restriction map level have demonstrated great similarity, some divergence and a few polymorphisms. All of the animals share a common chromosomal background. There is a general haplotype organization which is conserved across species boundaries yet is unique to the genus _Macaca_. The common macaque restriction haplotype is presented in Figure 4.1. It is compared to the maps previously determined by Martin (1982) for other Old World monkeys. The similarities within the cluster are substantial as one compares the macaques to the other genera although it is clear that each genus has a distinct and characteristic restriction enzyme pattern.

4.B.2. Differences are Detected across Species Boundaries

As indicated (Section 4.B.1.), there are remarkable similarities at the restriction map level in comparing both across species and across genus boundaries. Within the macaques studied here, significant differences were detected which appear to be species-specific. Surprisingly, each of the three species studied: *Macaca fuscata*, *Macaca mulatta*, and *Macaca nemestrina* has a somewhat different tale to tell.

4.B.2.a. *Macaca fuscata*

As shown in Figure 4.2, the *Macaca fuscata* are a member of the *fascicularis* group of macaques. Their geographic distribution is restricted to the Japanese Islands. The Japanese macaques housed at the Oregon Regional Primate Research Center comprise a single native colony which was a gift by the Japanese government when their natural habitat was surrendered to make way for a major airport. Thus, the colony is representative of what one would expect to find in a mating colony in the wild.

The *Macaca fuscata* share the common haplotype background identified for the three macaque species studied. In addition to these common sites, five additional restriction sites were identified as being variable within this group. A total of 30 restriction enzyme sites were mapped for the species. The polymorphic sites included BamHI and SacI sites immediately 5′ to the delta globin locus, and EcoRI, SacI, and

PvuII sites located 3′ to the beta locus. Significantly, none of these polymorphic sites was located within the coding or noncoding regions of either locus.

A current paradigm of molecular evolution suggests that DNA in coding regions is more highly conserved than that in noncoding regions. These results support this paradigm at the level of the restriction endonuclease map. Interestingly, no polymorphic sites were identified within the delta globin locus, which is believed to be nonfunctional and therefore removed from any functional selective constraints

4.B.2.b. Macaca mulatta

The Macaca mulatta, or rhesus macaque, is also represented as a member of the fascicularis group of macaques demonstrated in Figure 4.2. In contrast to the Japanese macaque, their geographic distribution is both wide and diverse. The demographics of the rhesus monkeys used in this study are given in Appendix A. Unlike the Macaca fuscata which represent the type of mating colony one would expect to find in the wild, the Macaca mulatta represent much more diverse geographic origins. Unfortunately, many of the animals which were either zoo or colony born do not have records to indicate the geographic areas they represent.

The Macaca mulatta sample studied shares the common haplotype background identified in Figure 4.1. They display much greater intraspecies diversity than the Japanese macaques; ten of thirty-two

restricted loci could be identified as polymorphic.

Two sites, (HindIII and PvuII) were located within the delta globin locus. A single BamHI site was located in the beta globin locus. All three of the rhesus intra-genic polymorphic sites were also present in the *Macaca fuscata* but were fixed rather than variable sites.

Whereas the Japanese macaques represent a single colony from an island population the rhesus monkeys represent animals from multiple geographic locations. This may also be implicated in the degree of polymorphism which has been detected.

4.B.2.c. *Macaca nemestrina*

The *silenus*-*sylvanus* group of macaques is presented in Figure 4.3. Although *Macaca nemestrina* is usually associated with the Malay peninsula and Sumatra, the animals used in this study were identified as originating from Sulawesi (Celebes Island).

The data representing *Macaca nemestrina* demonstrate much less diversity as only four polymorphic sites were identified. Polymorphic restriction sites were identified using BamHI, EcoRI and HindIII. All sites were in non-coding DNA. Of particular interest in this species was the identification of an 0.8 kb deletion in extragenic DNA between the beta and delta loci.

Studies on Cheetah populations have identified an extremely low level of variation among members of the species. The theory that a recent population bottleneck resulted in the extreme homozygosity of the species has been advanced (O'Brien et al., 1985). It is possible that such a mechanism may account for the homogeneity detected at the restriction map level in the *Macaca nemestrina*.

Macaca nemestrina is characterized by having several ambiguous sub-species designations. Such sub-species may be of recent origin and, as such, may also be demonstrating homogeneity due to a founder effect. A further consideration is that the smaller, non-random sample used may have provided an inaccurate representation of the species. Further study of pig-tailed macaques representing more diverse geographic origins might help resolve some of these possibilities.

4.C. Implications

The data presented have several implications. A major finding of this study is that restriction-endonuclease polymorphisms do occur in non-human primates. They can readily be detected by comparing the restriction maps of several individuals. No particular care was taken to optimize the detection of polymorphisms in this study although BamHI, HindIII, PstI, and PvuII have been shown to detect polymorphisms in the human globin clusters (Collins and Weissman, 1984). Studies with enzymes such as TaqI and MspI which are believed to recognize CpG "hot-spots" for mutation (Wyman and White, 1980) might optimize detection of

restriction fragment length polymorphisms.

A second observation supported by this study is that diverging populations do demonstrate differences at the DNA level which can be detected by restriction endonuclease analysis. By studying several animals from each of various species, it is possible to identify restriction sites characteristic of the genus as opposed to those of the particular species. Haplotype analysis may become a useful tool in categorizing animals when species boundaries are ambiguous at the morphologic level.

Restriction endonucease analysis allows the detection of intraspecies variation at the level of the individual. The greater degree of diversity detected among the rhesus macaques may be a reflection of their diverse geographic origins. Considerable evidence has accumulated through studies of both normal and mutant human globin loci indicating that racial and ethnic differences occur at the restriction map level (Wainscoat et al., 1986). The mechanisms implicated in generating such human divergence seem also to be operating upon Old World monkeys and other non-human primates.

4.D. Applications of Restriction Endonuclease Analysis

There are several useful applications for continued study of macaque or other populations by restriction endonuclease analysis. First, species designations may be improved for otherwise ambiguous

individuals. The mechanisms of speciation may be determined by detecting population isolates which are diverging, especially those diverging due to geographic isolation, such as island populations.

Second, breeding programs for animals in captivity might also be improved. Restriction fragment data may be useful at two levels: to ensure existing heterozygote variability while preserving species identities.

The data presented also relate to humans. Several groups have identified geographical correlates of disease at the restriction enzyme level (Antonarakis et al., 1982; 1984; Elbein et al., 1985; Thein et al., 1985; Maggio et al., 1986). Knowledge of the mechanisms of population divergence at this level might help investigators anticipate racial and ethnic distinctions which could confuse linkage analysis of the restriction-fragment length polymorphisms used in preclinical and prenatal diagnosis.

CHAPTER FIVE: SUMMARY AND CONCLUSIONS

The thesis of this study was to examine a group of closely related species, namely the Old World monkey genus *Macaca*. Preliminary studies included cellulose-acetate electrophoresis of hemolysates in order to both characterize the species and to identify those individuals who were deemed "representative" for further study. Five species were studied: *Macaca fascicularis*, *fuscata*, *mulatta*, *nemestrina*, and *nigra*. DNA analysis was performed on the adult non-alpha globin loci of *Macaca fuscata*, *mulatta*, and *nemestrina*. The results were distinctive for each species.

Macaca fuscata was used as a representative of a single mating colony. Thirty restriction sites were mapped using the human genomic probes δ-probe and pβPstd, and eight restriction enzymes: BamHI, BglII, EcoRI, HindIII, KpnI, PstI, PvuII, and SacI. Of the thirty sites identified, five were polymorphic. These sites were limited to extragenic (non-coding) DNA.

Macaca mulatta was studied using the same protocol. Thirty-two restriction sites were mapped of which ten were polymorphic. Three of these sites were located in coding DNA. The high degree of divergence detected is presumed to reflect the diverse geographical origins of the animals studied.

Macaca nemestrina was also studied. Four sites were identified as polymorphic. In addition, a deletion of 0.8 kb in the region between delta and beta loci was detected.

The three species demonstrated conservation of a common haplotype at the restriction map level. This map is characteristic of the macaques as opposed to other Old World monkeys. The polymorphisms and deletion sites clearly fit into distinct categories corresponding to species boundaries.

Several conclusions can be made based on the results of this study. One is that restriction fragment length polymorphisms do occur in non-human primates. They do not appear to correspond to any sites yet described for the orthologous human non-alpha globins. This is not surprising in light of the finding that such markers differ significantly among the more closely-related macaque species.

Secondly, although individual polymorphic sites are not necessarily conserved from one species to another, a characteristic back-ground haplotype appears to be conserved within the genus. Thus, haplotype analysis can be used as a tool for phylogenetic analysis. The polymorphic loci may also serve to help determine species identity when morphology is ambiguous.

Such species determinations may prove useful in colony-breeding programs when it is important to maintain both heterozygosity and

species identity within a breeding population. Racial and ethnic haplotype correlates in the severity of human disease and linkage of restriction markers to genetic defects have already been determined. Similar correlates may also prove useful in non-human primates. Species and geographic variation in monkeys may provide clues to the understanding of human racial variants.

REFERENCES

Adams JW, Kaufman RE, Kretschmer PJ, Harrison M, and Nienhuis AW: A family of long reiterated DNA sequence, one copy of which is next to the human β globin gene. Nucleic Acids Res. 8:6113, 1980

Allan M, and Paul J: Transcription in vivo of an Alu family member upstream from the human ε-globin gene. Nucleic Acids Res. 12:1193, 1984

Antoine M, and Niessing J: Intron-less globin genes in the insect _Chironomus thummi thummi_. Nature 310:795, 1984

Antonarakis SE, Boehm CD, Giardina PJV, and Kazazian Jr HH: Nonrandom association of polymorphic restriction sites in the β-globin gene cluster. P.N.A.S.(USA) 79:137-141, 1982

Antonarakis SE, Boehm CD, Serjeant GR, et al: Origin of the β^S-globin gene in blacks: the contribution of recurrent mutation or gene conversion or both. P.N.A.S.(USA) 81:853-856, 1984

Arnheim N: Concerted evolution of multigene families. _In_: Evolution of Genes and Proteins pp. 38-61, Nei M and Koehn RK, eds., Sinauer Associates Publishers, Sunderland, Mass., 1983

Avise JC and Lansman RA: Polymorphism of mitochondrial DNA in populations of higher animals. In: Evolution of Genes and Proteins pp. 147-164, Nei M and Koehn RK, eds., Sinauer Associates Publishers, Sunderland, Mass., 1983

Barcroft J: The Respiratory Function of the Blood. Part II. Haemoglobin. London, Cambridge University Press, 1928

Barrie PA, Jeffreys AJ, and Scott AF: Evolution of the β globin gene cluster in man and the primates. J. Mol. Biol. 149:319, 1981

Barwick RC, and Schneider RG: Computer assisted differentiation of hemoglobin variants. Texas Rep. Biol. Med. 40:143, 1980

Bell GI, Selby MJ, and Rutter WJ: The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31, 1982

Berger SL and Krug MS: A simple method for quantifying ^{32}P in submicroliter samples. BioTechniques 3(1):38, 1985

Blake CCF: Do genes-in-pieces imply proteins-in-pieces? Nature 273:267, 1978

Boyer SH, Crosby EF, Thurmon TF, et al: Hemoglobins A and A_2 in New World primates: comparative variation and its evolutionary implications. Science 166:1428, 1969

Brisson N, Pombo-Gentile A, and Verma DPS: Organization and expression of leghaemoglobin genes. Canadian J. Biochem. 60:272, 1982

Brown WM: Evolution of animal mitochondrial DNA. In: Evolution of Genes and Proteins pp. 62-88, Nei M and Koehn RK, eds., Sinauer Associates Publishers, Sunderland, Mass., 1983

Brown WM, George Jr M, and Wilson AC: Rapid evolution of animal mitochondrial DNA. P.N.A.S.(USA) 76(4):1967, 1979

Brown WM and Vinograd J: Restriction endonuclease cleavage maps of animal mitochondrial DNAs. P.N.A.S.(USA) 71(11):4617, 1974

Bunn HF and Forget BG: Hemoglobin: Molecular, Genetic and Clinical Aspects. Dyson J, ed., W. B. Saunders Co., Philadelphia, 1986

Capp GL, Rigas DA, and Jones RT: Hemoglobin Portland 1: A new human hemoglobin unique in structure. Science 157:65, 1967

Capp GL, Rigas DA, and Jones RT: Evidence for a new haemoglobin chain (ζ-chain). Nature 228:278, 1970

Center for Disease Control: Laboratory Methods for Detecting Hemoglobinopathies. U.S. Department of Health and Human Services, 1984

Chakravarti A, Buetow KH, Antonarakis SE, et al.: Non-uniform recombination within the β-globin gene cluster. Am. J. Hum. Genet. 36:1239, 1984

Chang LYE and Slightom JL: Isolation and nucleotide sequence analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. J. Mol. Biol. 180:767, 1984

Cleary ML, Haynes JR, Schon EA, and Lingrel JB: Identification by nucleotide sequence analysis of a goat pseudoglobin gene. Nucleic Acids Res. 8:4791, 1980

Cleary ML, Schon EA, Lingrel JB: Two related pseudogenes are the result of a gene duplication in the goat β-globin locus. Cell 26:181, 1981

Collins FS and Weissman SM: The molecular genetics of human hemoglobin. Prog. Nucl. Acid Res. Mol. Biol. 31:315, 1984

Cronin JE, Cann R, Sarich VM: Molecular evolution and systematics of the genus Macaca. In The Macaques: Studies in Ecology, Behavior and Evolution. Lindburg DG, Ed., Van Nostrand Reinhold, New York, 1980

Czelusniak J, Goodman M, Hewett-Emmett D, et al: Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature 298:297, 1982

Dayhoff MO (ed): Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington, D.C., 1978

DeGrouchy J, Turleau C, and Finaz C: Chromosomal phylogeny of the primates. Annual Review of Genetics 12:289, 1978

Deisseroth A, Nienhuis A, Turner P, et al.: Localization of the human α-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. Cell 12:205, 1977

Deisseroth A, Nienhuis A, Lawrence J, et al.: Chromosomal localization of human β globin gene on human chromosome 11 in somatic cell hybrids. P.N.A.S.(USA) 75:1456, 1978

Delson E: Fossil macaques, phyletic relationships and a scenario of deployment. In The Macaques: Studies in Ecology, Behavior and Evolution. Lindburg DG, Ed., Van Nostrand Reinhold, New York, 1980

DiGiavoni L, Haynes SR, Misra R, and Jelinek WR: Kpn I family of long-dispersed repeated DNA sequences of man: Evidence for entry into genomic DNA of DNA copies of poly (A)-terminated Kpn I RNAs. P.N.A.S.(USA) 80:6533, 1983

DiSegni G, Cararra G, Tocchini-Valentini GR, Shoulders CC, and Baralle FE: Selective in vitro transcription of one of the two Alu family repeats present in the 5'-flanking region of the human ϵ globin gene. Nucleic Acids Res. 9:6709, 1981

Dickerson RE and Geis I: Hemoglobin: structure, function, evolution, and pathology. Hagopian P, ed., The Benjamin Cummings Publishing Co., Menlo Park, CA, 1983

Dobkin C, Clyne J, Metzenburg A, and Bank A: Expression of a cloned Lepore globin gene. Blood 67:168, 1986

Dolan M, Sugarman BJ, Dodgson JB, and Engel JD: Chromosomal arrangement of the chicken β-type globin genes. Cell 24:669, 1981

Duncan C, Biro PA, Choudary PV, et al.: RNA polymerase III transcriptional units are interspersed among human non-α globin genes. P.N.A.S.(USA) 76:5095, 1979

Duncan CH, Jagadeeswaran P, Wang RRC, and Weissman SM: Structural analysis of templates and RNA polymerase III transcripts of Alu family sequences interspersed among the human β-like globin genes. Gene 13:185, 1981

Edgell MH, Hardies SC, Brown B, et al: Evolution of the mouse β globin complex locus. In: Evolution of Genes and Proteins pp. 1-13, Nei M and Koehn RK, eds. Sinauer Associates Publishers, Sunderland, Mass., 1983

Edmundson AB: Amino-acid sequence of sperm whale myoglobin. Nature 205:883, 1965

Efstratiadis A, Posakony JW, Maniatis T, et al.: The structure and evolution of the human β-globin gene family. Cell 21:653, 1980

Elbein S, Rotwein P, Permutt MA, et al.: Lack of association of the polymorphic locus in the 5'-flanking region of the human insulin gene and diabetes in American blacks. Diabetes 34:433, 1985

Engel JD, and Dodgson JB: Analysis of the closely linked adult chicken α-globin genes in recombinant DNAs. P.N.A.S.(USA) 77:2596, 1980

Elbein S, Rotwein P, Permutt MA et al: Lack of association of the polymorphic locus in the 5'-flanking region of the human insulin gene and diabetes in American blacks. Diabetes 34:433, 1985

Fermi G, Perutz MF, Shaanan B, and Fourme B: The crystal structure of human deoxyhemoglobin at 1.7 Å resolution. J. Mol. Biol. 175:159, 1984

Ferris SD, Brown WM, Davidson WS, and Wilson AC: Extensive polymorphism in the mitochondrial DNA of apes. P.N.A.S.(USA) 78:6319, 1981

Ferris SD, Wilson AC, and Brown WM: Evolutionary tree for apes and humans based on cleavage maps of mitichondrial DNA. P.N.A.S.(USA) 78:2432, 1981

Fitch WM and Langley CH: Evolutionary rates in proteins: Neutral mutations and the molecular clock. In: Molecular Anthropology: Genes and Proteins in the Evolutionary Ascent of the Primates, Goodman M and Tashian RE eds., Plenum Press, New York, 1976

Flavell RA: The mystery of the mouse α-globin pseudogene. Nature 295:370, 1982

Fooden J: Classification and distribution of living macaques (Macaca lacepede, 1799). In The Macaques: Studies in Ecology, Behavior and Evolution. Lindburg DG, Ed., Van Nostrand Reinhold, New York, 1980

Forget BG, Tuan D, Biro PA, Jagadeeswaran P, and Weissman SM: Structural features of the DNA flanking the human non-α globin genes: Implications in the control of fetal hemoglobin switching. Trans. Assoc. Am. Physicians 94:204, 1981

Fritsch EF, Lawn RM, and Maniatis T: Molecular cloning and characterization of the human β-like globin gene cluster. Cell 19:959, 1980

Gatti RA, Concannon P, and Salser W: Multiple use of Southern blots. BioTechniques 2(3):148, 1984

Gilbert W: Why genes in pieces? Nature 271:501, 1978

Go M: Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90, 1981

Gojobori T, Li WH, Graur D: Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. 18:360, 1982

Goodbourn SEY, Higgs DR, Clegg JB, and Weatherall DJ: Molecular basis of length polymorphism in the human zeta-globin gene complex. P.N.A.S.(USA) 80:5022, 1983

Goodbourn SEY, Higgs DR, Clegg JB, and Weatherall DJ: Allelic variation and linkage properties of a highly polymorphic restriction fragment in humans. Mol. Biol. Med. 2:223, 1984

Goodman M: Evolution of immunologic species specificity of human serum protein. Human Biology 34:104, 1962

Goodman M, Koop BF, Czelusniak J, Weiss ML, and Slightom JL: The eta(η)-globin gene: its long evolutionary history in the β-globin gene family of mammals. J. Mol. Biol. 180:803, 1984

Goodman M, Moore GW, and Matsuda G: Darwinian evolution in the genealogy of haemoglobin. Nature 253:603, 1975

Goosens M, Dozy AM, Embury SH, et al: Triplicated α-globin loci in humans. P.N.A.S.(USA) 77(1):518, 1980

Hardies SC, Edgell MH, and Hutchison III CA: Evolution of the mammalian β-globin gene cluster. J. Biol. Chem. 259(6):3748, 1984

Hardison RC: Comparison of the β-like globin gene families of rabbits and humans indicates that the gene cluster $5'\epsilon-\gamma-\delta-\beta$ 3' predates the mammalian radiation. Mol. Biol. Evol. 1(5):390-410, 1984

Hardison RC, Butler ET III, Lacy E, and Mantiatis T: The structure and transcription of four linked rabbit β-like globin genes. Cell. 18:1285, 1979

Hardison RC and Margot JB: Rabbit globin pseudogene $\psi\beta2$ is a hybrid of δ- and β-globin gene sequences. Mol. Biol. Evol. 1(4):302, 1984

Harris S, Barrie PA, Weiss ML, and Jeffreys AJ: The primate pseudo-β gene: an ancient β-globin pseudogene. J. Mol. Biol. 180:785, 1984

Hess JF, Fox M, Schmid C, and Shen CK: Molecular evolution of the human adult α-globin-like gene region: Insertion and deletion of Alu family repeats and non-Alu DNA sequences. P.N.A.S.(USA) 80:5970, 1983

Higgs DR, Old JM, Pressley L, Clegg JB, and Weatherall DJ: A novel α-globin gene arrangement in man. Nature 284:632, 1980

Higgs DR, Goodbourn SEY, Wainscoat JS, Clegg JB, and Weatherall DJ: Highly variable regions of DNA flank the human α globin genes. Nucleic Acids Res. 9(17):4213, 1981

Hill A, Hardies SC, Phillips SJ, et al.: Two mouse early embryonic β-globin gene sequences: Evolution of the nonadult β-globins. J. Biol. Chem. 259:3739, 1984

Hoffman HA and Gottlieb AJ: Investigations of nonhuman primate hemoglobin-electrophoretic variation. Ann. N. Y. Acad. Sci. 162:205, 1969

Hoppe F: Über das Verhalten des Blutfarbestoffes im Spectrum des Sonnenlichtes. Virchows Arch. Path. Anat. Physiol. 23:446, 1862

Hosbach HA, Widmer HJ, Andres AC et al: Expression and organization of the globin genes in Xenopus laevis. In: Embryonic Development, Part A:Genetic Aspects pp 115-125, Alan R. Liss, Inc., New York, 1982

Hosbach HA, Wyler T, and Weber R: The Xenopus laevis globin gene family: Chromosomal arrangement and gene structure. Cell 32:45, 1983

Houck CM, Rinehart FP, and Schmid CW: A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. 132:289, 1979

Howard CF Jr: Correlations of hemoglobin A_{1C} and metabolic status in nondiabetic, borderline diabetic, and diabetic Macaca nigra. Diabetes 31(12):1105, 1982

Huehns ER, Flynn FV, Butler EA, and Beaven GH: Two new haemoglobin variants in the very young human embryo. Nature 189:496, 1961

Hunt TL, Hurst-Calderone S, and Dayhoff MO: Globins. In: Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, Dayhoff M.O. ed., Washington D.C., National Biomedical Research Foundation, pp. 229-251, 1978

Hutchinson CA III, Hardies SC, Padgett RW, Weaver S, and Edgell MH: The mouse globin pseudogene βh3 is descended from a premammalian δ-globin gene. J. Biol. Chem. 259:12881, 1984

Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms (review). Mol. Biol. Evol. 2(1):13, 1985

Ishimoto G, Kuwata M, Shotake T: Abnormal hemoglobin found in Japanese macaques. Japanese Journal of Human Genetics 20:33, 1975

Ishimoto G, Tanaka T, Nigi H, and Prychodko W: Hemoglobin variation in macaques. Primates 11(3):229, 1970

Jagadeeswaran P, Forget BG, and Weissman SM: Short interspersed repetitive DNA elements in eucaryotes: Transposable DNA elements generated by reverse transcription of RNA Pol III transcripts? Cell 26:141, 1981

Jagadeeswaran P, Pan J, Forget BG, and Weissman SM: Sequences of human repetitive DNA, non-α globin genes and major histocompatibility locus genes. Part II. Sequences of non-α globin genes in man. Cold Spring Harbor Symp. Quant. Biol. 47:1079, 1983

Jahn CL, Hutchison III CA, Phillips SJ, et al: DNA sequence organization of the β-globin complex in the BALB/c mouse. Cell 21:159, 1980

Jeffreys AJ, Barrie PA, Harris S, et al.: Isolation and sequence analysis of a hybrid δ-globin pseudogene from the brown lemur. J. Mol. Biol. 156:487, 1982

Jeffreys AJ, Wilson V, Blanchetot A, et al: The human myoglobin gene: a third dispersed globin locus in the human genome. Nucleic Acids Res. 12(7):3235, 1984

Jeffreys AJ, Wilson V, Wood D, et al: Linkage of adult α- and β-globin genes in Xenopus laevis and gene duplication by tetraploidization. Cell 21:555, 1980

Jelinek WR, and Schmid CW: Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. 51:813, 1982

Jolly CJ and Barnicot NA: Serum and red-cell protein variations of the Celebes black ape. Folia Primata. 4:206, 1966

Kamuzora H. Jones RT, and Lehmann H: The ζ-chain, an α-like chain of human embryonic haemoglobin. FEBS Lett. 46:195, 1974

Kan YW, and Dozy AM: Polymorphism of DNA sequence adjacent to human β-globin structure gene: Relationship to sickle mutation. P.N.A.S.(USA) 75:5631, 1978

Kan YW, and Dozy AM: Evolution of the hemoglobin S and C genes in world populations. Science 209:388, 1980

Kaufman RE, Kretschmer PJ, Adams JW, et al.: Cloning and characterization of DNA sequences surrounding the human γ, δ and β-globin genes. P.N.A.S.(USA) 77:4229, 1980

Keller W: The RNA lariat: a new ring to the splicing of mRNA precursors. Cell 39:423, 1984

Kendrew JC, Dickerson RE, Strandberg BE, et al.: Structure of myoglobin: a three dimensional Fourier synthesis at 2Å resolution. Nature 185:422, 1960

Kilpatrick MW, Klysik J, Singleton CK, et al.: Intervening sequences in human fetal globin genes adopt left-handed Z helices. J. Biol. Chem. 259(11):7268, 1984

Kimura M: The neutral theory of molecular evolution. In: Evolution of Genes and Proteins pp. 208-233, Nei M and Koehn RK, eds., Sinauer Associates Publishers, Sunderland, Mass., 1983

King JL and Jukes TH: Non-Darwinian evolution. Science 164:788, 1969

Knowlton RG, Cohen-Haguenauer O, Cong NV, et al: A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. Nature 318:380, 1985

Koefler HP, Sparkes RS, Stang H, Mohandas T: Regional assignment of genes for human α-globin and phosphoglycollate phosphatase to the short arm of chromosome 16. P.N.A.S.(USA) 78:7015, 1981

Kole LB, Haynes SR, and Jelinek WR: Discrete and heterogeneous high molecular weight RNAs complementary to a long dispersed repeat family (a possible transposon) of human DNA. J. Mol. Biol. 165:257, 1983

Konkel DA, Maizel JV, and Leder P: The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18:865, 1979

Koop BF, Goodman M, Xu P, Chan K, and Slightom JL: Primate eta(η)-globin DNA sequences and man's place among the great apes. Nature 319:234, 1986

Korber E: Inaugural dissertation: "Über differenzen Blutfarbstoffes." Dorpat, 1866. Cited by Bischoff H: Z. Exp. Med. 48:472, 1926

Kozak M: Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12:857, 1984

Kunkel HG and Wallenius G: New hemoglobins in normal adult blood. Science 122:288, 1955

Lacy E, Hardison RC, Quon D, Maniatis T: The linkage arrangement of four rabbit β-like globin genes. Cell 18:1273, 1979

Lacy E and Maniatis T: The nucleotide sequence of a rabbit β-globin pseudogene. Cell 21:545, 1980

Lasker G: What is molecular anthropology? In: Molecular Anthropology: Genes and Proteins in the Evolutionary Ascent of the Primates. Goodman M and Tashian RE, eds., Plenum Press, New York, 1976

Lawn RM: The molecular genetics of hemophilia: blood clotting VIII and IX. Cell 42:405, 1985

Lawn RM, Efstratiadis A, O'Connell C, and Maniatis T: The nucleotide sequence of the human β-globin gene. Cell 21:647, 1980

Leder P, Hansen JN, Konkel D, et al.: Mouse globin system: a functional and evolutionary analysis. Science 209:1336, 1980

Leiter AB, Montminy MR, Jamieson E, and Goodman RH: Exons of the human pancreatic polypeptide gene define functional domains of the precursor. J. Biol. Chem. 260(24):13013, 1985

Li Q, Powers PA, and Smithies O: Nucleotide sequence of 16-kilobase pairs of DNA 5' to the human ϵ-globin gene. J. Biol. Chem. 260(28):14901, 1985

Li WH: Evolution of duplicate genes and pseudogenes. In: Evolution of Genes and Proteins pp. 14-37, Nei M and Koehn RK, eds., Sinauer Associates Publishers, Sunderland, Mass., 1983

Lindburg DG (ed.): The Macaques: Studies in Ecology, Behavior, and Evolution. Van Nostrand Reinhold Co., New York, 1980

Little PFR: Globin pseudogenes. Cell 28:683, 1982

Lower R: Tractus de Corde. Translated by KJ Franklin. In Gunther RT: Early Science at Oxford, Vol. IX. Clarendon Press, 1932

Magenis RE, Donlon TA, and Tomar DR: Localization of the β-globin gene to 11p15 by in situ hybridization: utilization of chromosome 11 rearrangements. Human Genetics 69:300, 1985

Maggio A, Acuto S, Lo Gioco P, et al.: β^{A} and β^{thal} DNA haplotypes in Sicily. Human Genetics 72:229, 1986

Mahoney WC and Nute PE: Hemoglobin α-chain variation in macaques: primary structures of the α^{I} and α^{II} chains from the adult hemoglobins of Malaysian Macaca nemestrina. Arch. Biochem. Biophys. 196:64, 1979

Mai LL: A model of chromosome evolution and its bearing on cladogenesis in the hominoidea. In: New Interpretations of Ape and Human Ancestry. Eds. RL Ciochon and RS Corrucini, Plenum Press, New York, 1983

Maniatis T: The molecular genetics of human hemoglobins. Annual Review of Genetics 14:145, 1980

Martin SL: PhD Thesis. University of California. Evolutionary Silencing of the Delta Globin Gene in Old World Monkeys. 1982

Martin SL, Vincent KA, and Wilson AC: Rise and fall of the delta globin gene. J. Mol. Biol. 164:513, 1983

Martin SL, Voliva CF, Burton FH, Edgell MH, and Hutchison CA III: A large interspersed repeat found in mouse DNA contains a long open reading frame that evolves as if it encodes a protein. P.N.A.S.(USA) 81:2308, 1984

Martin SL, Zimmer EA, Kan YW, and Wilson AC: Silent δ-globin gene in Old World monkeys. P.N.A.S.(USA) 77(6):3563, 1980

Marx JL: Is RNA copied into DNA by Mammalian Cells? Science 216:969, 1982

Massa ER, Guerrasio A, and Ricco G: Hemoglobin pattern in some primates. Antropologia Contemporanea 4(1-2):117, 1981

McLachlan AD, Staden R, and Boswell DR: A method for measuring the non-random bias of a codon usage table. Nucleic Acids Res. 12:9567, 1984

Meyerhof W, Klinger-Mitropoulos S, Stalder J, Weber R, and Knochel W: The primary structure of the larval β_1-globin gene of Xenopus laevis and its flanking regions. Nucleic Acids Res. 12(20): 7705, 1984

Miesfeld R, Krystal M, and Arnheim N: A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β globin genes. Nucleic Acids Res. 9(22):5931, 1981

Miller DA: Evolution of primate chromosomes. Science 198:1116, 1977

Miyata T and Hayashida H: Extraordinarily high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons. P.N.A.S.(USA) 78(9):5739, 1981

Monaco AP, Bertelson CJ, Middlesworth W, et al: Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. Nature 316:842, 1985

Murata M and Thompson PE: Two-nucleotide codon change in a hemoglobin polymorphism of the Celebes black ape (Macaca nigra). Biochem. Genet. 14(1-2):183, 1976

Nagel RL, Fabry ME, Pagnier J: Hematologically and genetically distinct forms of sickle cell anemia in Africa: The Senegal type and the Benin type. N. Engl. J. Med. 312(14):880, 1985

Nei M: Gene duplication and nucleotide substitution in evolution. Nature 221:40, 1969

Nomiyama H, Tsuzuki T, Wakasugi S, Fukda M, and Shimada K: Interruprion of a human nuclear sequence homologous to mitochondrial DNA by a member of the Kpn I 1.8 kb family. Nucleic Acids Res. 12:5225, 1984

Nute PE: Multiple hemoglobin α-chain loci in monkeys, apes, and man. Annals N. Y. Acad. Sci. 241:39, 1974

Nute PE, Mahoney WC, and Boosman A: Hemoglobin α-gene duplication in macaques: individual Macaca nemestrina with three structurally different α chains. Arch. Biochem. Biophys. 206(2):346, 1981

Nute PE and Pataryas HA: Multiple fetal and adult hemoglobins in Malaysian Macaca nemestrina: consequences of α-chain variation. Am. J. Phys. Anthrop. 40:17, 1974

O'Brien SJ, Roelke ME, Marker L, et al: Genetic basis for species vulnerability in the cheetah. Science 227:1428, 1985

Ohno S: Evolution by Gene Duplication. Springer-Verlag, New York, 1970

Ojwang PJ, Nakatsuji T, Gardiner MB et al.: Gene deletion as the molecular basis for the Kenya-$^{G}\gamma$-HPFH condition. Hemoglobin 7(2):115, 1983

Orkin SH, Antonarakis SE, and Kazazian HH: Polymorphism and molecular pathology of the human beta-globin gene. Prog. Hematol. 13:49, 1983

Orkin SH, Antonarakis SE, and Kazazian Jr HH: Base substitution at position -88 in a β-thalassemic globin gene. J. Biol. Chem. 259(14):8679, 1984

Orkin SH, Cheng T-C, Antonarakis SE, and Kazazian HH Jr: Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human β-globin gene. EMBO J. 4:453, 1985

Orkin SH, Kazazian HH,Jr: The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Ann. Rev. Genet. 18:131, 1984

Pagnier J, Mears GS, Dunda-Belkhodja O, et al.: Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. P.N.A.S.(USA) 81:1771,1984

Patient RK, Elkington JA, Kay RM, and Williams JG: Internal organization of the major adult α- and β-globin genes of Xenopus laevis. Cell 21:565, 1980

Patthy L: Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. Cell 41:657, 1985

Pauling L, Itano HA, Singer SJ, and Wells IC: Sickle cell anemia, a molecular disease. Science 110:543, 1949

Perler F, Efstratiadis A, Lomedico P, et al: The evolution of genes: the chicken preproinsulin gene. Cell 20:555, 1980

Perry M, Thomsen GH, Roeder RG: Genomic organization and nucleotide sequences of two distinct histone gene clusters from Xenopus laevis: identification of novel conserved upstream sequence elements. J. Mol. Biol. 185:479, 1985

Perutz MF, Rossman MG, Cullis AF, Muirhead H, Will G, and North ACT: Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. Nature 185:416, 1960

Phillips SJ, Hardies SC, Jahn CL, Edgell MH, and Hutchison III CA: The complete nucleotide sequence of a β-globin-like structure, βh2, from the $[Hbb]^d$ mouse BALB/c. J. Biol. Chem. 259(12): 7947, 1984

Pirastu M, Doherty M, Galanello R, et al.: Frequent crossing over in human DNA generates multiple chromosomes containing the sickle and β-thalassemia genes and increases HbF production. Blood 62(Suppl.1):75a, 1983

Pirastu M, Saglio G, Chang JC, Cao A, and Kan YW: Initiation codon mutation as a cause of α-thalassemia. J. Biol. Chem. 259: 12315, 1984

Pollack Y, Kasir J, Shemer R, Metzger S, and Szyf M: Methylation pattern of mouse mitochondrial DNA. Nucleic Acids Res. 12: 4811, 1984

Poncz MP, Schwartz E, Ballantine M, and Surrey S: Nucleotide sequence analysis of the $\delta\beta$-globin gene region in humans.
J. Biol. Chem. 258(19):11599, 1983

Potter SS: Rearranged sequences of a human KpnI element.
P.N.A.S.(USA) 81:1012, 1984

Powers PA, Altay C, Huisman THJ, and Smithies O: Two novel arrangements of the human fetal globin genes: $^{G}\gamma$-$^{G}\gamma$ and $^{A}\gamma$-$^{A}\gamma$.
Nucleic Acids Res. 12(18):7023, 1984

Prochownik EV, Bock SC, and Orkin SH: Intron structure of the human antithrombin III gene differs from that of other members of the serine protease inhibitor superfamily. J. Biol. Chem. 260(17): 9608, 1985

Proudfoot NJ, and Brownlee GG: 3′ Noncoding region sequences in eukaryotic messenger RNA. Nature 263:211, 1976

Proudfoot NJ, Gil A, and Maniatis T: The structure of the human ζ-globin gene and a closely linked, nearly identical pseudogene.
Cell 31:553, 1982

Proudfoot NJ and Maniatis T: The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication.
Cell 21:537, 1980

Riggs A: Preparation of blood hemoglobins of vertebrates. Methods in Enzymology 76:5, 1981

Rogers J: A straight LINE story. Nature 306:113, 1983

Roninson IB and Ingram VM: Gene evolution in the chicken β-globin cluster. Cell 28:515, 1982

Sawada I, Beal MP, Shen CKJ, et al: Intergenic DNA sequences flanking the pseudo alpha globin genes of human and chimpanzee. Nucleic Acids Res. 11(22):8087, 1983

Schimenti JC and Duncan CH: Concerted evolution of the cow $\epsilon 2$ and $\epsilon 4$ β-globin genes. Mol. Biol. Evol. 2(6):505, 1985

Schimenti JC and Duncan CH: Structure and organization of the bovine β-globin genes. Mol. Biol. Evol. 2(6):514, 1985

Schmeckpeper BJ, Scott AF, and Smith KD: Transcripts homologous to a long repeated DNA element in the human genome. J. Biol. Chem. 259(2):1218, 1984

Schmid CW and Jelinek W: The Alu family of dispersed repetitive sequences. Science 216:1065, 1982

Schneider RG, and Barwick RC: Measuring relative electrophoretic mobilities of mutant hemoglobins and globin chains. Hemoglobin 2:417, 1978

Schroeder WA, Huisman THJ, Brown AK, et al: Postnatal changes in the heterogeneity of human fetal globin. Pediatr. Res. 5:493, 1971

Schroeder WA, Huisman THJ, Shelton JR, et al: Evidence for multiple structural genes for the γ chain of human fetal hemoglobin. P.N.A.S.(USA) 60:537, 1968

Shafit-Zagardo B, Brown Fl, et al.: KpnI families of long interspersed repetitive DNAs associated with the human β-globin gene cluster. Gene 20:397,1982

Shafit-Zagardo B, Brown FL, Zavodny PJ, Maio JJ: Transcription of KpnI families of long interspersed DNAs in human cells. Nature 304:277, 1983

Sharp PA: Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature 301:471, 1983

Shatkin AJ: mRNA cap binding protein: Essential factors for initiating translation. Cell 40:223, 1985

Shen S-H, Slightom JL, and Smithies O: A history of the fetal globin gene duplication. Cell 26:191, 1981

Shen S-H and Smithies O: Human globin pseudo-β_2 is not a globin-related sequence. Nucleic Acids Res. 10(23):7809, 1982

Sibley CG and Ahlquist JE: Phylogeny and classification of birds based on the data of DNA-DNA hybridization. Current Ornithology 1:245, 1983

Sibley CG and Ahlquist JE: The relationships of the starlings (Sturnidae: Sturnini) and the mockingbirds (Sturnidae: Mimini). The Auk 101:230, 1984a

Sibley CG and Ahlquist JE: The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. J. Mol. Evol. 20:2, 1984b

Sibley CG and Ahlquist JE: Reconstructing bird phylogeny by comparing DNA's. Scientific American 254(2):82, 1986

Singer MF: SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes. Cell 28:433, 1982

Slightom JL, Blechi AE, and Smithies O: Human fetal Gγ- and Aγ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21:627, 1980

Slightom JL, Chang LYE, Koop BF, and Goodman M: Chimpanzee fetal Gγ and Aγ globin gene nucleotide sequences provide further evidence of gene conversions in hominine evolution. Mol. Biol. Evol. 2(5):370, 1985

Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503, 1975

Spritz RA: Duplication/deletion polymorphism 5'- to the human β globin gene. Nucleic Acids Res. 9(19):5037, 1981

Spritz RA, DeRiel JK, Forget BG, and Weissman SM: Complete nucleotide sequence of the δ-globin gene. Cell 21:639, 1980

Stephens JC: Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion.
Mol. Biol. Evol. 2(6):539, 1985

Thayer RE and Singer MF: Defining the beginning and end of Kpn I family segments. EMBO J. 3:1753, 1984

Thein SL, Wainscoat JS, Old JM et al: Feasibility of prenatal diagnosis of β-thalassemia with synthetic DNA probes in two Mediterranean populations. The Lancet ii:345, 1985

Thomas M, Davis RW: Studies on the cleavage of bacteriophage lambda DNA with EcoRI restriction endonuclease. J. Mol. Biol. 91:315, 1975

Townes TM, Fitzgerald MC, and Lingrel JB: Triplication of a four-gene set during evolution of the goat β-globin locus produced three genes now expressed differentially during development. P.N.A.S.(USA) 81:6589, 1984

Townes TM, Shapiro SG, Wernke SM, and Lingrel JB: Duplication of a four-gene set during the evolution of the goat β-globin locus produced genes now expressed differentially in development.
J. Biol. Chem. 259(3):1896, 1984

Trent RJ, Bowden DK, Old JM et al.: A novel rearrangement of the human β-like globin gene cluster. Nucleic Acids Res. 9:6723, 1981

Tuttle AH, Newsome FE, Jackson CH, and Overman RR: Hemoglobin types of _Macaca irus_ and _Macaca mulatta_ monkeys. Science 133:578-9, 1961

Ullrich A, Dull TJ, Gray A, Philips JA III, and Peter S: Variation in the sequence and modification state of the human insulin gene flanking region. Nucleic Acids Res. 10:2225, 1982

Ullu E and Tschudi C: Alu sequences are processed 7SL RNA genes. Nature 312:171, 1984

Villeponteau B and Martinson H: Isolation and characterization of the complete chicken β-globin region: frequent deletion of the adult β-globin genes in lambda. Nucleic Acids Res. 9(15):3731, 1981

Wade PT, Barnicot NA, and Huehns ER: Structural studies on the major and minor haemoglobin of the monkey _Macaca irus_.
Biochim. Biophys. Acta 221:450, 1970

Wainscoat JS, Bell JI, et al.: Multiple origins of the sickle mutation: Evidence from beta S globin gene cluster polymorphisms. Mol. Biol. Med. 1:191, 1983

Wainscoat JS, Hill AVS, Boyce AL, et al.: Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491, 1986

Wainwright BJ, Scrambler PJ, Schmidtke J, et al.: Localization of cystic fibrosis locus to human chromosome 7cen-q22. Nature 318:384, 1986

Weaver S, Comer MB, Jahn CL, Hutchison III CA, and Edgell MH: The adult β-globin genes of the "single" type mouse C57BL. Cell 24:403, 1981

White R, Woodward S, Leppert M, et al.: A closely linked genetic marker for cystic fibrosis. Nature 318:382, 1986

Wilde CD: Pseudogenes. CRC Crit. Rev. Biochem. 19(4):323, 1986

Wilson AC, Carlson SS, and White TJ: Biochemical Evolution. Ann. Rev. Biochem. 46:573, 1977

Wilson AC: The molecular basis of evolution. Sci. Am. 253(4):64, 1985

Winichagoon P, Higgs DR, Goodbourn SEY, et al: Multiple arrangements of the human embryonic zeta globin genes. Nucleic Acids Res. 10:5853, 1982

Wrightstone RN: International Hemoglobin Information Center Variants List. Hemoglobin 10(3):261, 1986

Wyman AR and White R: A highly polymorphic locus in human DNA. P.N.A.S.(USA) 77:6754, 1980

Zimmer EA: PhD thesis. University of California. Evolution of Primate Globin Genes. 1981

Zimmer EA, Martin SL, Beverley SM, Kan YW, and Wilson AC: Rapid duplication and loss of genes coding for the α chains of hemoglobin. P.N.A.S.(USA) 77(4):2158, 1980

TABLE 1.1 Developmental Expression of Hemoglobin in Man

Stage of Development	Hemoglobin	Globin Chains: Alpha	Beta	Subunit Composition
Embryonic	Hb Gower 1	zeta	epsilon	$\zeta_2 \varepsilon_2$
	Hb Gower 2	alpha	epsilon	$\alpha_2 \varepsilon_2$
	Hb Portland I	zeta	gamma	$\zeta_2 \gamma_2$
	Hb Portland II	zeta	beta	$\zeta_2 \beta_2$
Fetal	Hb F	alpha	gamma	$\alpha_2 \gamma_2$
Adult	Hb A	alpha	beta	$\alpha_2 \beta_2$
	Hb A_2	alpha	delta	$\alpha_2 \delta_2$

TABLE 1.2 Haplotype Analysis of Sickle Cell Variants

Restriction Enzyme	Group	Benin	Senegal
Hinc II	I	–	–
Hind III	I	–	+
Hind III	I	–	–
Pvu II	I	+	+
Hinc II	I	–	+
Hinc II	I	+	+
Hinf I	Inter δ–β	–	+
Hgi AI	II	+	+
Ava II	II	+	+
Hpa I	II	–	+
Bam HI	II	+	+

TABLE 2.1 Primate Species Studied

Latin Name	Common Name	Study Abbreviation
Macaca fascicularis	Crab-eating macaque	MFa
Macaca fuscata	Japanese macaque	MFu
Macaca mulatta	Rhesus macaque	MMu
Macaca nemestrina	Pig-tailed macaque	MNe
Macaca nigra	Celebes black macaque	MNi
Homo sapiens	Man	HSa

TABLE 2.2 Restriction Endonucleases

Enzyme	Source	Cleavage Site	Supplier
BamHI	*Bacillus amyloliquefaciens H*	G / GATCC	BRL[1]
BglII	*Bacillus globigii*	A / GATCT	IBI[2]
EcoRI	*Escherichia coli RY13*	G / AATTC	IBI
HindIII	*Haemophilus influenzae R*	A / AGCTT	IBI
KpnI	*Klebsiella pneumoniae*	GGTAC / C	B-M[3]
PstI	*Providencia stuartii*	CTGCA / G	IBI
PvuII	*Proteus vulgaris*	CAG / CTG	IBI
SacI	*Streptomyces achromogenes*	GAGCT / C	IBI

1. Bethesda Research Laboratories, Bethesda, Maryland
2. International Biotechnologies, Inc., New Haven, Connecticut
3. Boehringer-Mannheim Biologicals, Indianapolis, Indiana

TABLE 2.3 Agarose Gel Electrophoresis Patterns

Restriction Digest Panel A		Restriction Digest Panel B	
Well No.	Digests	Well No.	Digests
1	λ HindIII + EcoRI	1	λ HindIII + EcoRI
2	λ HindIII	2	λ HindIII
3	EcoRI	3	KpnI
4	BamHI + EcoRI	4	KpnI + BamHI
5	BamHI	5	PvuII + Bam HI
6	SacI + EcoRI	6	BamHI
7	SacI + BamHI	7	HindIII + BamHI
8	SacI	8	BglII + BamHI
9	SacI + BglII	9	HindIII
10	SacI + PvuII	10	HindIII + BglII
11	λ HindIII + EcoRI	11	HindIII + EcoRI
12	BglII	12	BglII
13	PvuII + BglII	13	λ HindIII + EcoRI
14	PvuII	14	BglII + EcoRI
15	PvuII + EcoRI	15	EcoRI
16	PvuII + BamHI	16	PvuII + EcoRI
17	HindIII + PvuII	17	PstI + EcoRI
18	HindIII	18	PstI
19	λ HindIII	19	λ HindIII
20	λ HindIII + EcoRI	20	λ HindIII + EcoRI

TABLE 3.1 Relative Mobility of Primate Major Adult Hemoglobins

Species	Schneider's N
Macaca fascicularis	+ 0.7 + 5.0
Macaca fuscata	0.0
Macaca mulatta	+ 0.4
Macaca nemestrina	+ 5.7
Macaca nigra	- 0.7 +10.0

TABLE 3.2 Restriction Fragment Sizes Used to Map Macaque Beta and Delta Globin Loci (in kb)

Enzyme(s)	Macaca fuscata	Macaca mulatta	Macaca nemestrina
BamHI	2.3,2.7,4.7, 8.7,19.0	2.7,5.0,10.6	2.3,2.7,4.8,19.6, 23
BglII	5.2,8.9	5.2,9.4	4.1,5.1,8.6,8.9
EcoRl	2.5,5.7	2.4,5.7,7.2	2.3,5.7
HindIII	5.3,5.4	5.3,18.2,23	4.3,5.1,16.2,17.5
KpnI	>20	>20	4.8
PstI	4.5,4.6	4.6,4.9	4.6
PvuII	15.3,19.3	2.2,2.3,14.2	16.8,17.3,19.3
SacI	10.9,19.5	4.6,11.3,19.8	19.2
BamHI+BglII	2.1,2.9,4.8	1.6,2.9,5.0, 5.5	2.0,2.8,4.7,5.3
BamHI+EcoRI	1.1,1.4,1.8, 2.7	1.2,1.5	2.3,2.7,4.1

BamHI+ HindIII	1.3,2.2,3.9, 4.1,4.8	4.4,5.5,10.4	3.8,4.0
BamHI+PvuII	2.3,2.7,4.8, 7.3,7.8,15.4	2.2,2.3,4.9, 5.4	2.6,4.5,5.2,19
BamHI+SacI	1.4,2.3,2.8, 4.8,7.1	2.6,4.3,4.9	2.7,4.6,5.8
BglII+EcoRI	1.3,2.4,2.7, 5.1	2.2,3.1	2.3,2.9,5.0,20
BglII+ HindIII	5.2,8.9	4.5,4.6,8.4	4.3,4.5,9.1
BglII+PvuII	2.9,5.2,5.6	2.2,2.3,5.2, 5.9	2.8,4.5,4.9,5.4, 19
BglII+SacI	5.2,7.2	4.5,5.1,7.3	5.0,6.8
EcoRI+ HindIII	2.2,2.3,3.2	2.4,4.6,7.7	2.2,3.3,4.5,7.5
EcoRI+PstI	2.3,3.5	2.3	1.8,4.0,4.6
EcoRI+PvuII	1.1,1.4,3.7, 5.7	1.2,1.4,2.3, 5.6	5.6,6.4
EcoRI+SacI	2.4,5.8	2.4,4.5,5.7	5.8
HindIII+ PvuII	5.4,6.3	2.2,2.3,4.7, 5.4,14.9	4.3,5.1,17.2
PvuII+SacI	1.4,11.1,18.2, 18.8,19.1	2.2,2.3,11.2, 12.6,18.5	10.5,19.2

<u>Figure 1.1.</u> Comparison of Human Beta and Delta Globin Amino Acid Sequences.

The amino acid sequences of human beta and delta globins are presented. The primary sequences of the two chains differ at only 10 of 146 residues. The amino acids that vary are presented in italics and underlined. (Dayhoff, 1978)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BETA	Val	His	Leu	Thr	Pro	Glu	Glu	Lys	Ser	Ala	Val	Thr	Ala	Leu	Trp
DELTA	Val	His	Leu	Thr	Pro	Glu	Glu	Lys	*Thr*	Ala	Val	*Asn*	Ala	Leu	Trp

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
BETA	Gly	Lys	Val	Asn	Val	Asp	Glu	Val	Gly	Gly	Glu	Ala	Leu	Gly	Arg
DELTA	Gly	Lys	Val	Asn	Val	Asp	*Ala*	Val	Gly	Gly	Glu	Ala	Leu	Gly	Arg

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
BETA	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Glu	Ser	Phe
DELTA	Leu	Leu	Val	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Glu	Ser	Phe

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
BETA	Gly	Asp	Leu	Ser	Thr	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val
DELTA	Gly	Asp	Leu	Ser	*Ser*	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
BETA	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Gly	Ala	Phe	Ser	Asp	Gly	Leu
DELTA	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Gly	Ala	Phe	Ser	Asp	Gly	Leu

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
BETA	Ala	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ala	Thr	Leu	Ser	Glu
DELTA	Ala	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	*Ser*	*Gln*	Leu	Ser	Glu

	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
BETA	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn	Phe	Arg	Leu
DELTA	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn	Phe	Arg	Leu

	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
BETA	Leu	Gly	Asn	Val	Leu	Val	Cys	Val	Leu	Ala	His	His	Phe	Gly	Lys
DELTA	Leu	Gly	Asn	Val	Leu	Val	Cys	Val	Leu	Ala	*Arg*	*Asn*	Phe	Gly	Lys

	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
BETA	Glu	Phe	Thr	Pro	Pro	Val	Gln	Ala	Ala	Tyr	Gln	Lys	Val	Val	Ala
DELTA	Glu	Phe	Thr	Pro	*Gln*	*Met*	Gln	Ala	Ala	Tyr	Gln	Lys	Val	Val	Ala

	136	137	138	139	140	141	142	143	144	145	146
BETA	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His
DELTA	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His

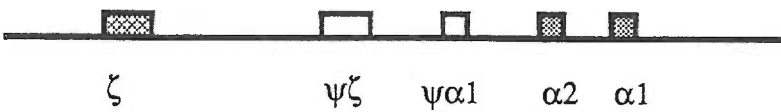
Figure 1.2. Codon Usage Charts for Human Alpha and Beta Globins.

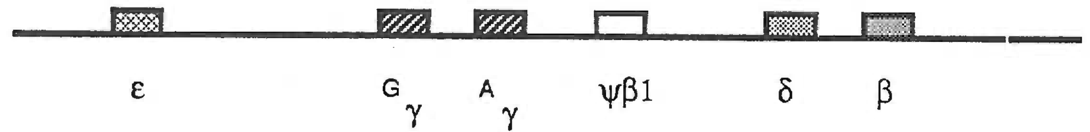
Amino acids are coded for by triplets of nucleotides. Each triplet is called a "codon". A codon usage chart indicates the number of times each amino acid is coded for by a particular nucleotide triplet. Some residues are represented by multiple codons, e.g. Leucine. Different proteins may demonstrate different codon preferences. In alpha-globin the codon UUC is used for most phenylalanine residues whereas only UUC is used in beta-globin. (Adapted from Collins and Weissman, 1984).

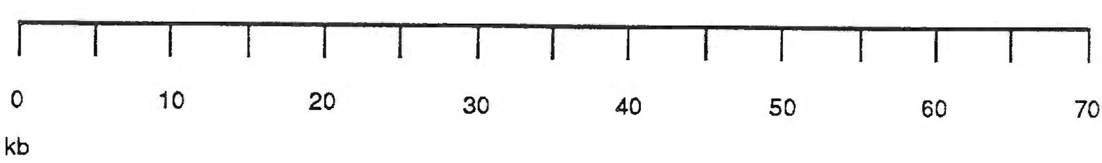
	U		α	β	C		α	β	A		α	β	G		α	β
U	Phe	UUU	0	5	Ser	UCU	3	1	Tyr	UAU	1	2	Cys	UGU	0	2
		UUC	7	3		UCC	4	2		UAC	2	1		UGC	1	0
	Leu	UUA	0	0		UCA	0	0	Term	UAA	1	1	Term	UGA	0	0
		UUG	0	0		UCG	0	0		UAG	0	0	Trp	UGG	1	2
C	Leu	CUU	1	0	Pro	CCU	2	5	His	CAU	0	2	Arg	CGU	1	0
		CUC	2	3		CCC	3	0		CAC	10	7		CGC	0	0
		CUA	1	0		CCA	0	2	Gln	CAA	0	0		CGA	0	0
		CUG	14	15		CCG	2	0		CAG	1	3		CGG	1	0
A	Ile	AUU	0	0	Thr	ACU	0	3	Asn	AAU	0	1	Ser	AGU	0	2
		AUC	0	0		ACC	9	3		AAC	4	5		AGC	4	0
		AUA	0	0		ACA	0	1	Lys	AAA	1	3	Arg	AGA	0	0
	Met	AUG	2	1		ACG	0	0		AAG	10	8		AGG	1	3
G	Val	GUU	1	3	Ala	GCU	2	6	Asp	GAU	0	5	Gly	GGU	2	4
		GUC	3	2		GCC	12	9		GAC	8	2		GGC	5	8
		GUA	0	0		GCA	0	0	Glu	GAA	0	2		GGA	0	0
		GUG	9	13		GCG	7	0		GAG	4	6		GGG	0	1

Figure 1.3. Globin Gene Organization in the Human Alpha and Beta Globin Clusters.

The human alpha and beta globin gene clusters are presented. The alpha globin gene cluster has been localized to HC 16 (human chromosome 16) and the beta globin cluster to HC 11. Embryonic, fetal, pseudo-, and adult globin genes are represented as coded for by the key. The pseudo-β1 locus has also been called pseudo-η to indicate its relationship to the newly-described eta-globin orthologues (Goodman et al., 1984). (Figure adapted from Bunn and Forget, 1986).







Embryonic

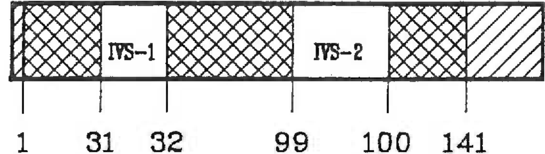
Fetal

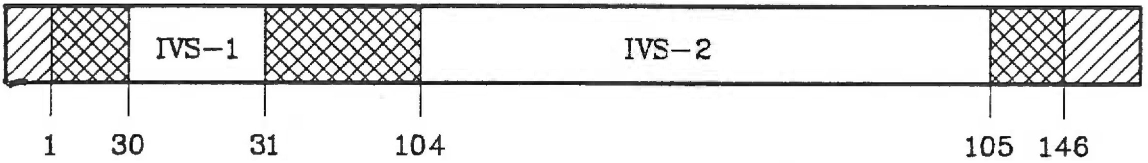
Pseudogene

Adult

Figure 1.4. Fine-Structure Mapping of Human Alpha and Beta Globin Genes.

Alpha and beta globin fine-structure mapping is presented. Each gene shares the characteristics of two introns (white boxes) and three exons (hatched boxes). Noncoding sequences flanking the genes are also indicated (diagonal lines). Although sizes of the exons are comparable, sizes of the introns vary greatly. In general, IVS-2 is larger than IVS-1. (Adapted from Bunn and Forget, 1986).


IVS-1
IVS-2
1
31
32
99
100
141


IVS-1
IVS-2
1
30
31
104
105
146

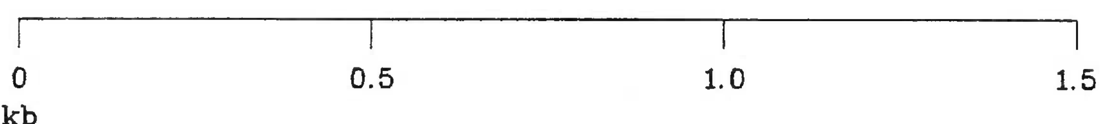
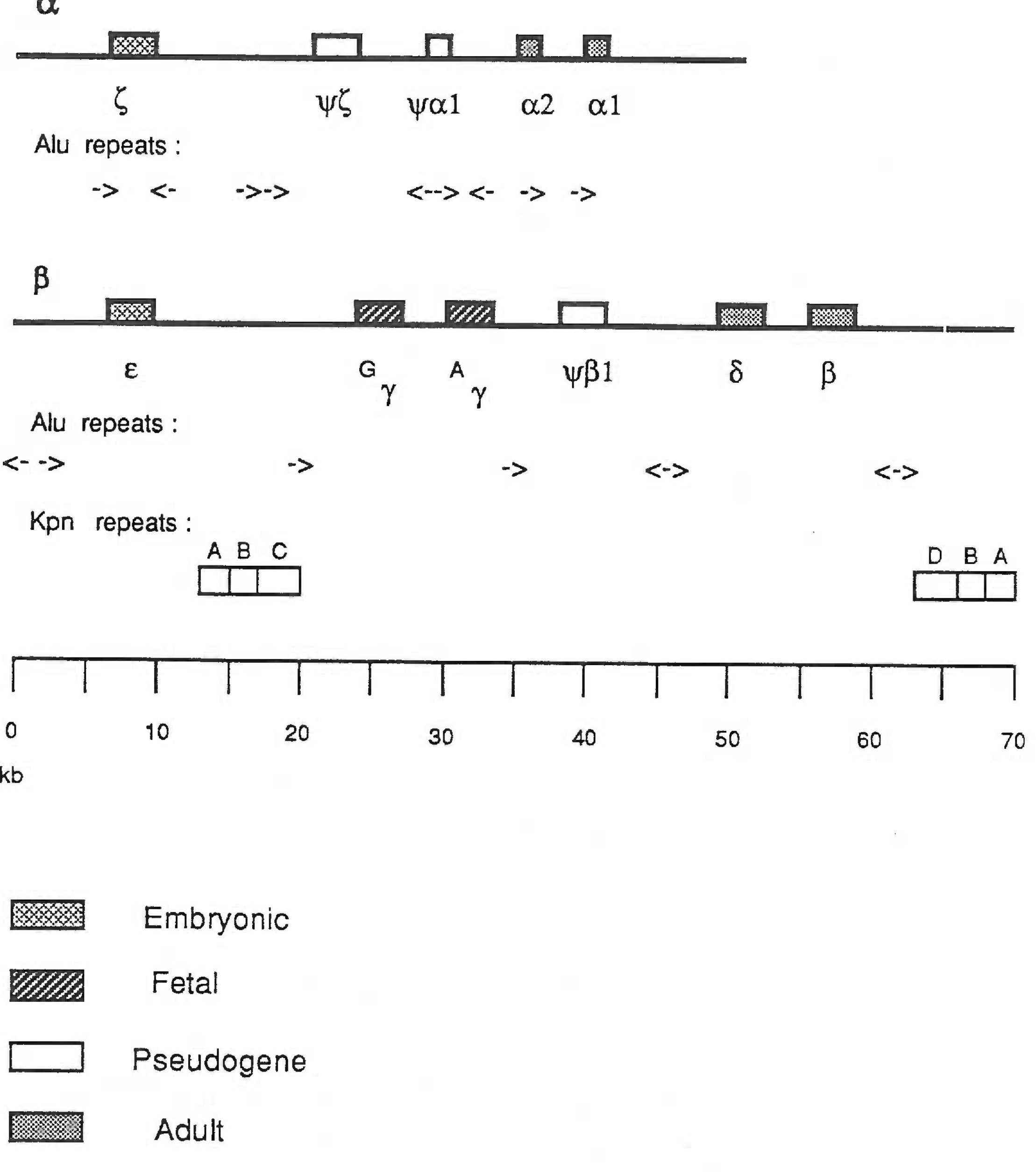

0
0.5
1.0
1.5
kb

Figure 1.5. Organization of Repetitive DNA in Human Globin Clusters.

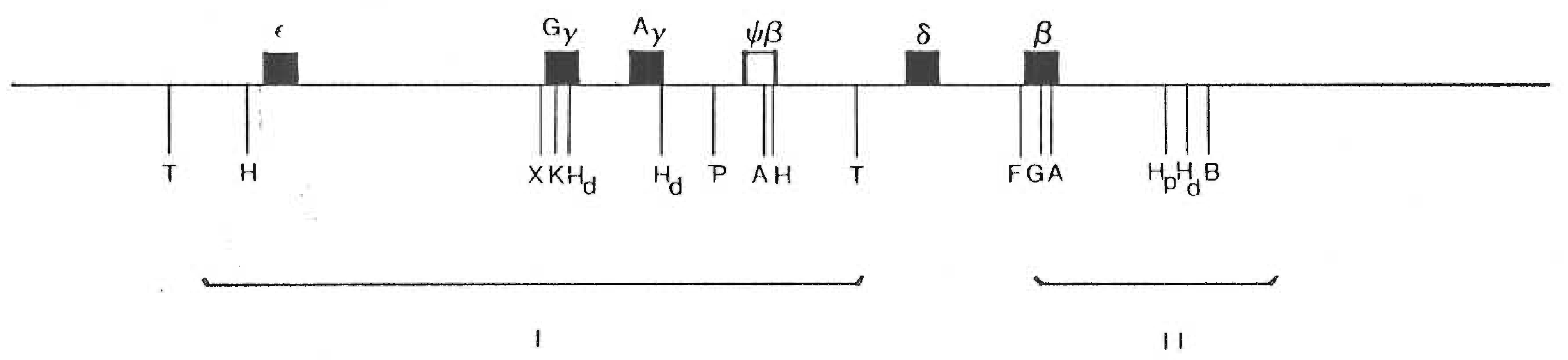
DNA elements have been identified which are repetitive and are found interspersed thoughout the genome. Two such groups of elements are represented. AluI repeats are short (300 nt) elements which occur as single or double copies in inverted-repeat orientation. Alu repeats are represented by arrows. KpnI repeats are longer (1.5 to 6kb) and are composed of differing subsets of repetitive elements. Kpn repeats are represented by boxes and the subsets designated by letters. (Adapted from Bunn and Forget, 1986).



α
ζ
ψζ
ψα1
α2
α1
Alu repeats :
-> <- ->-> <--> <- -> ->
β
ε
G γ
A γ
ψβ1
δ
β
Alu repeats :
<- -> -> -> <-> <->
Kpn repeats :
A B C
D B A
0 10 20 30 40 50 60 70
kb
Embryonic
Fetal
Pseudogene
Adult

Figure 1.6. Restriction Fragment Polymorphisms in the Human Beta Globin Clusters.

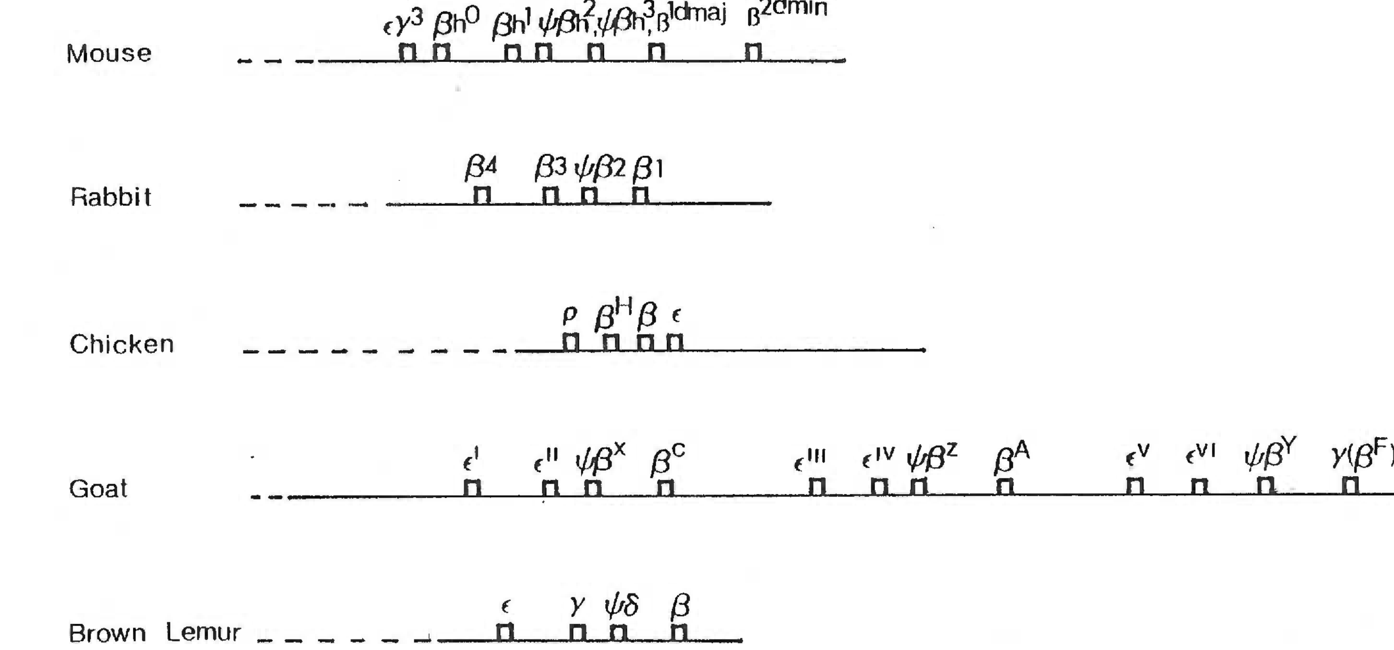
Polymorphic sites have been extensively mapped in the vicinity of the human beta globin cluster. Two groups of restriction sites have been identified which seem to be inherited as independent linkage groups. Group I includes loci in the vicinity of the embryonic, fetal, and pseudo-beta genes. Group II includes sites within and 3' to the beta globin gene. The restriction sites in the region between the two groups segregate in a random manner. (Adapted from Bunn and Forget, 1986).



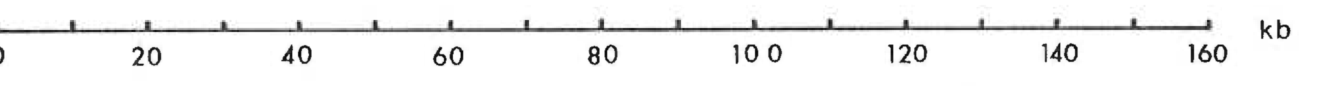
ε
Gγ
Aγ
ψβ
δ
β
T
H
X K H_d
H_d
P
A H
T
F G A
H_p H_d B
I
II

Figure 1.7. Non-Alpha Globin Gene Cluster Organization in Various Species.

Representative non-alpha clusters for several species are presented. Each species has a unique gene organization which is discussed in Section 1.D. (Modified from Bunn and Forget, 1986).



Mouse
εy3 βh0 βh1 ψβh2 ψβh3 β1dmaj β2dmin
Rabbit
β4 β3 ψβ2 β1
Chicken
ρ βH β ε
Goat
εI εII ψβX βC εIII εIV ψβZ βA εV εVI ψβY γ(βF)
Brown Lemur
ε γ ψδ β
Human
ε Gγ Aγ ψβ δ β



0
20
40
60
80
100
120
140
160
kb

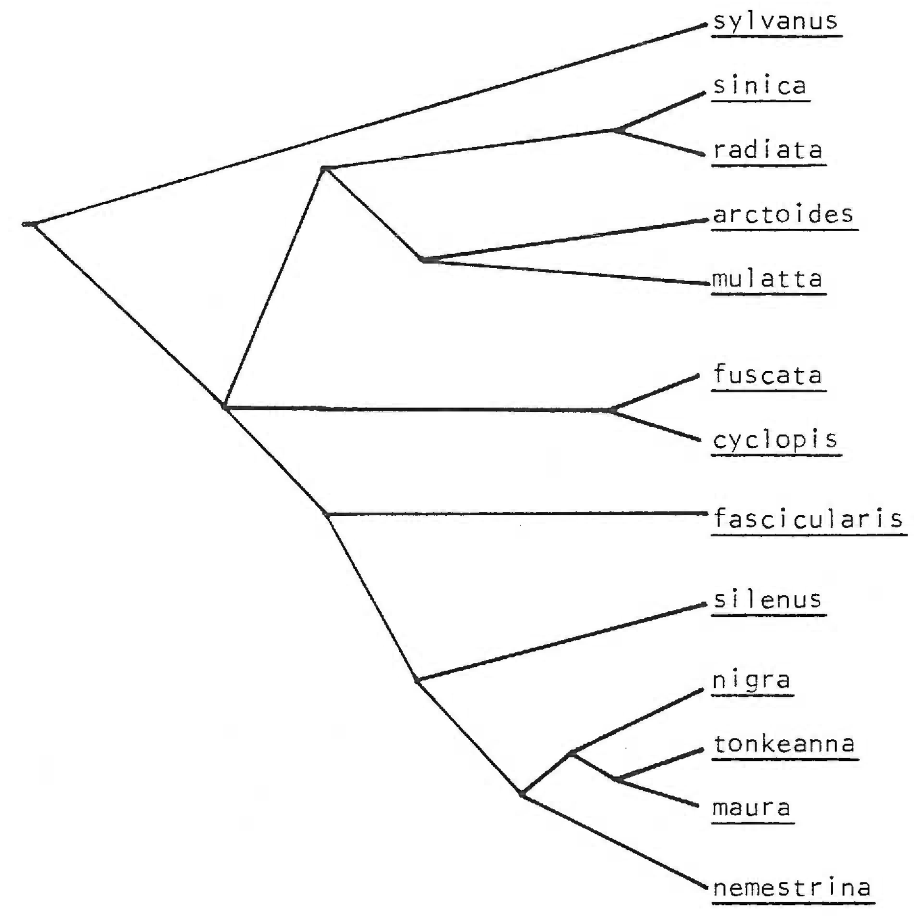
Figure 1.8. Primate Classifications

The placement of the macaques within the order Primates is indicated. Several species of the genus *Macaca* are listed. These species were utilized in this project.

- Order PRIMATES
  - Suborder ANTHROPOIDEA
    - Superfamily CERCOPITHECOIDEA
      - Family CERCOPITHECIDAE
        - Subfamily CERCOPITHECINAE
          - *Macaca*
            - *fascicularis* (Cynomologous)
            - *fuscata* (Japanese)
            - *mulatta* (Rhesus)
            - *nemestrina* (Pig - tailed)
            - *nigra* (Celebes Black)
    - Superfamily HOMINOIDEA
      - Family HOMINOIDAE
        - *Homo*
          - *sapiens* (Man)

Figure 1.9. Phenogram of Macaque Species

A phenogram of macaque species is presented based on accumulated protein data. The speciation radiation occurred relatively recently in the genus Macaca as indicated by the time-scale.



MACAQUE MOLECULAR PHYLOGENY

Millions of Years

Figure 3.1. Cellulose-Acetate Electrophoresis of Primate Hemoglobins.

Monkey hemolysates were electrophoresed on cellulose-acetate plates using Tris-EDTA-borate buffer at pH 8.4. Human standards of Hbs A,F,S, and C were also run. The samples originated at the cathode (-). Abbreviations are used as follows: MFa = *Macaca fascicularis*, MFu = *Macaca fuscata*, MMu = *Macaca mulatta*, MNe = *Macaca nemestrina*, MNi = *Macaca nigra*.

C S F A

32

HSaAFSC

MFa02

MFu01

MMu01

MNe01

MNi02

MNi03

HSa08

— +

Figure 3.2. HPLC Globin Chain Separation

A representative chromatograph of the globin chain separation for a *Macaca fascicularis* monkey is presented. An initial tall spike of heme-protein is followed by three broad peaks. The identities of the beta and alpha peaks are indicated.

OVEN&?

3.34

6.43

heme

8.99

?

14.58

β

17.57

α

27.63

38.05

42.46

44.28

45.59

47.88

min

STOP

hp 5830A

AREA %

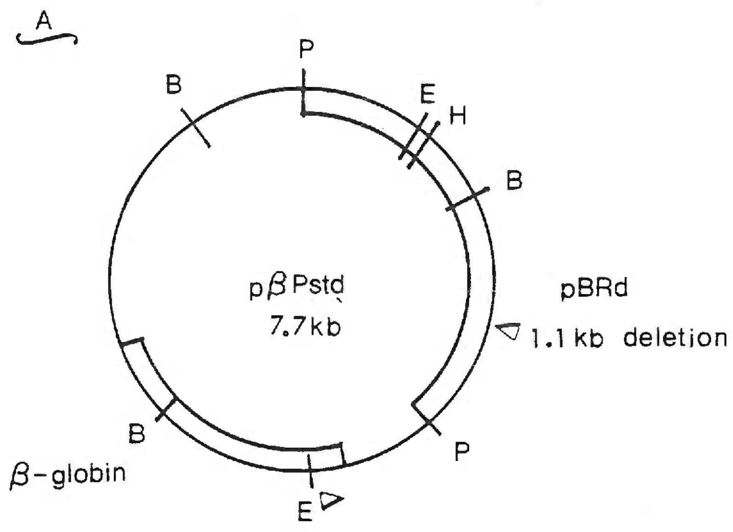
RT	AREA	AREA %
1.43	49250	7.623
3.54	73220	11.334
8.99	86460	13.383
17.57	217300	33.636
27.63	219800	34.023

1.0000 E+ 0

<u>Figure 3.3.</u> Genomic Beta Globin Probe pβPstd.

A. A putative restriction map of the probe as supplied by the Maniatis laboratory is presented.

B. A restriction map of the isolate of pβPstd received by this laboratory is presented. Note that the insert is in reverse orientation and that the vector size is larger than expected.


A
P
B
E
H
B
pβPstd
7.7kb
pBRd
1.1 kb deletion
B
P
β-globin
E

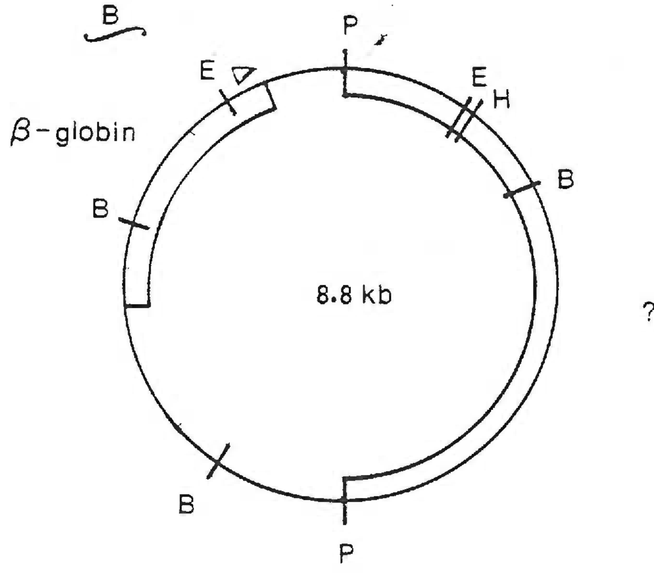
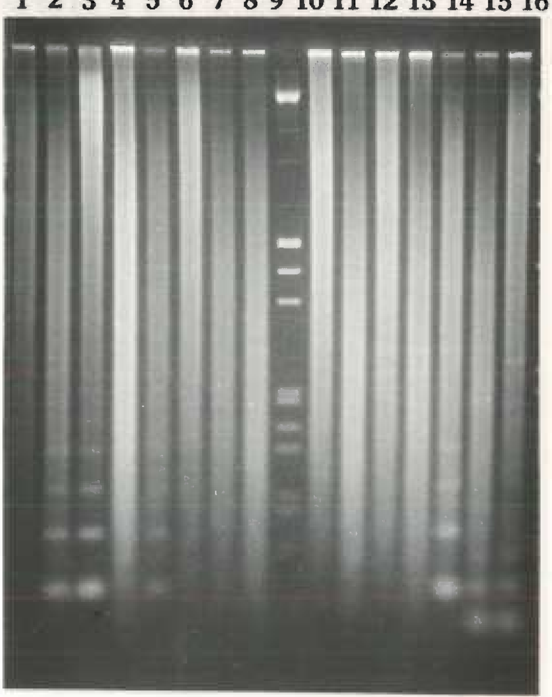

B
P
E
E
H
β-globin
B
B
8.8 kb
?
B
P

Figure 3.4. Agarose Gel Electrophoresis of Primate DNA.

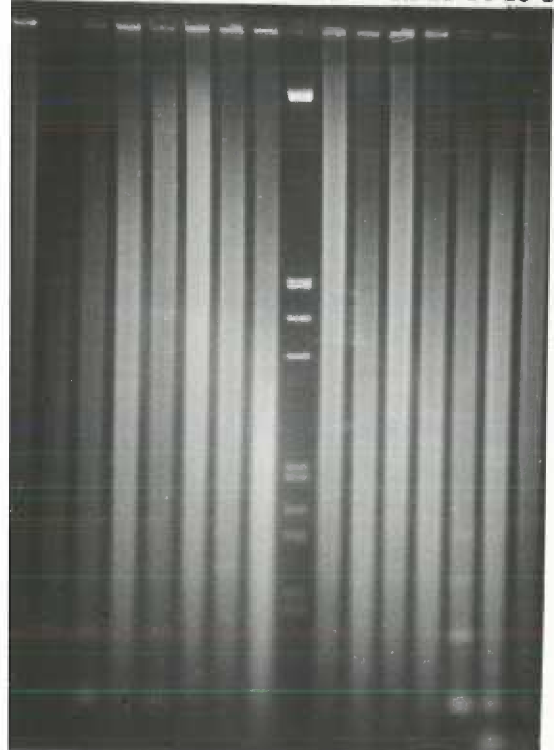
Agarose gels representing panel A (Table 2.3.) are presented. Note the light bands which correspond to repetitive DNA in each of the monkey samples. Lane 9 contains a lambda standard digested with restriction enzymes HindIII and EcoRI. Abbreviations: HSa = *Homo sapiens*, MFu = *Macaca fuscata*, MMu = *Macaca mulatta*, and MNe = *Macaca nemestrina*.

HSa07

MFu07



MMu06



MNe06

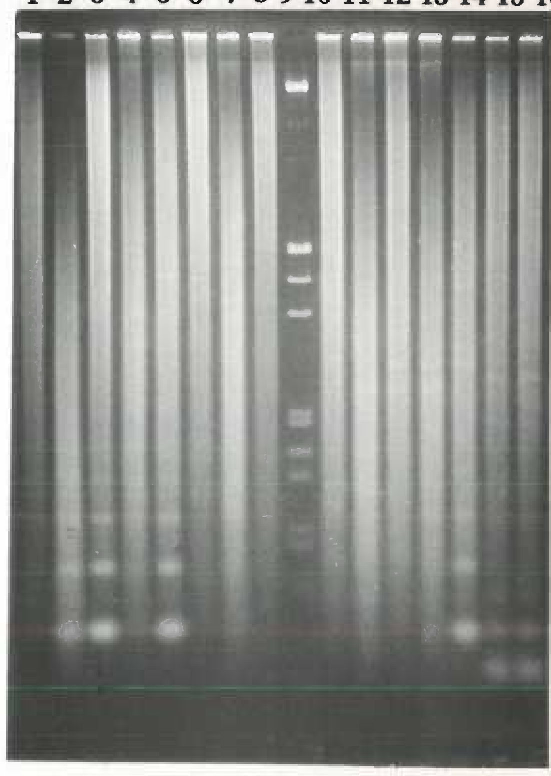


Figure 3.5. Autoradiography of Macaca fuscata DNA Probed with δ-Probe.

A representative autoradiograph is presented. Five to 6.5 μg of monkey DNA was loaded into each lane. Lambda DNA digested with HindIII and EcoRI was used as a standard and fragment sizes are indicated. The enzymes used in single and double digests are given above the appropriate lanes.

MFu05 Blot A

EcoRI
BamHI + EcoRI
BamHI
SacI + EcoRI
SacI + BamHI
Sac I
Sac I + Bgl II
Sac I + PvuII
λH3 + RI
Bgl II
Pvu II + Bgl II
Pvu II
Pvu II + EcoRI
Pvu II + BamHI

Size of Lambda Standard (kb)

21.70
5.15
5.00
4.27
3.48
1.98
1.90
1.59
1.37
0.94
0.83
0.56

Figure 3.6. Restriction Map and Haplotype Analysis for Adult Non-Alpha Globin Loci of *Macaca fuscata*.

Thirty restriction enzyme sites were mapped using probes pβPstd and δ-probe. Restriction enzymes are abbreviated as follows: B = BamHI, G = BglII, E = EcoRI, H = HindIII, P = PstI, V = PvuII, and S = SacI. Polymorphic sites are represented by an asterisk (*). Haplotypes for the five Japanese macaques studied are indicated: (+) indicates the presence of a restriction site; (-) its absence.

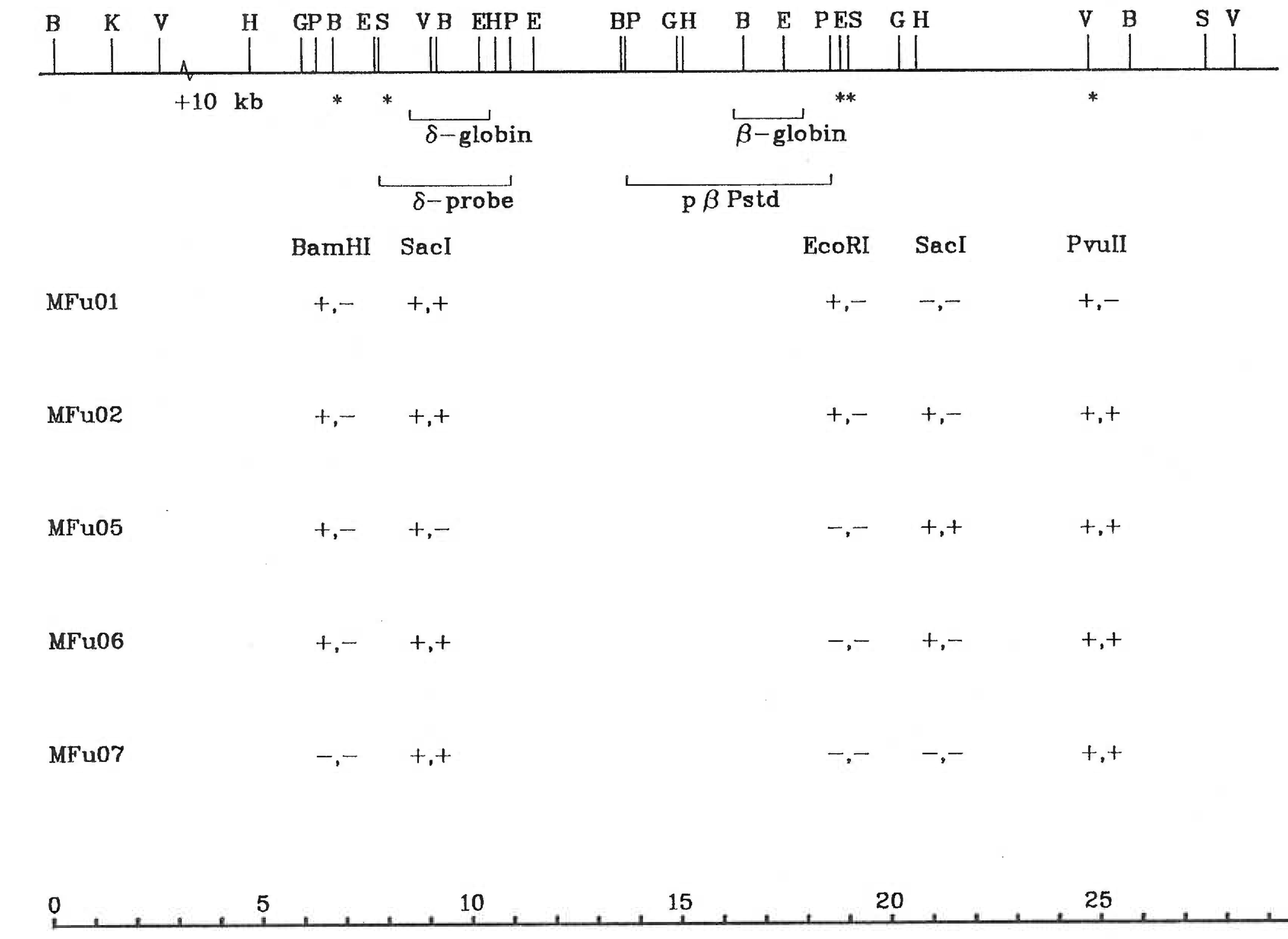
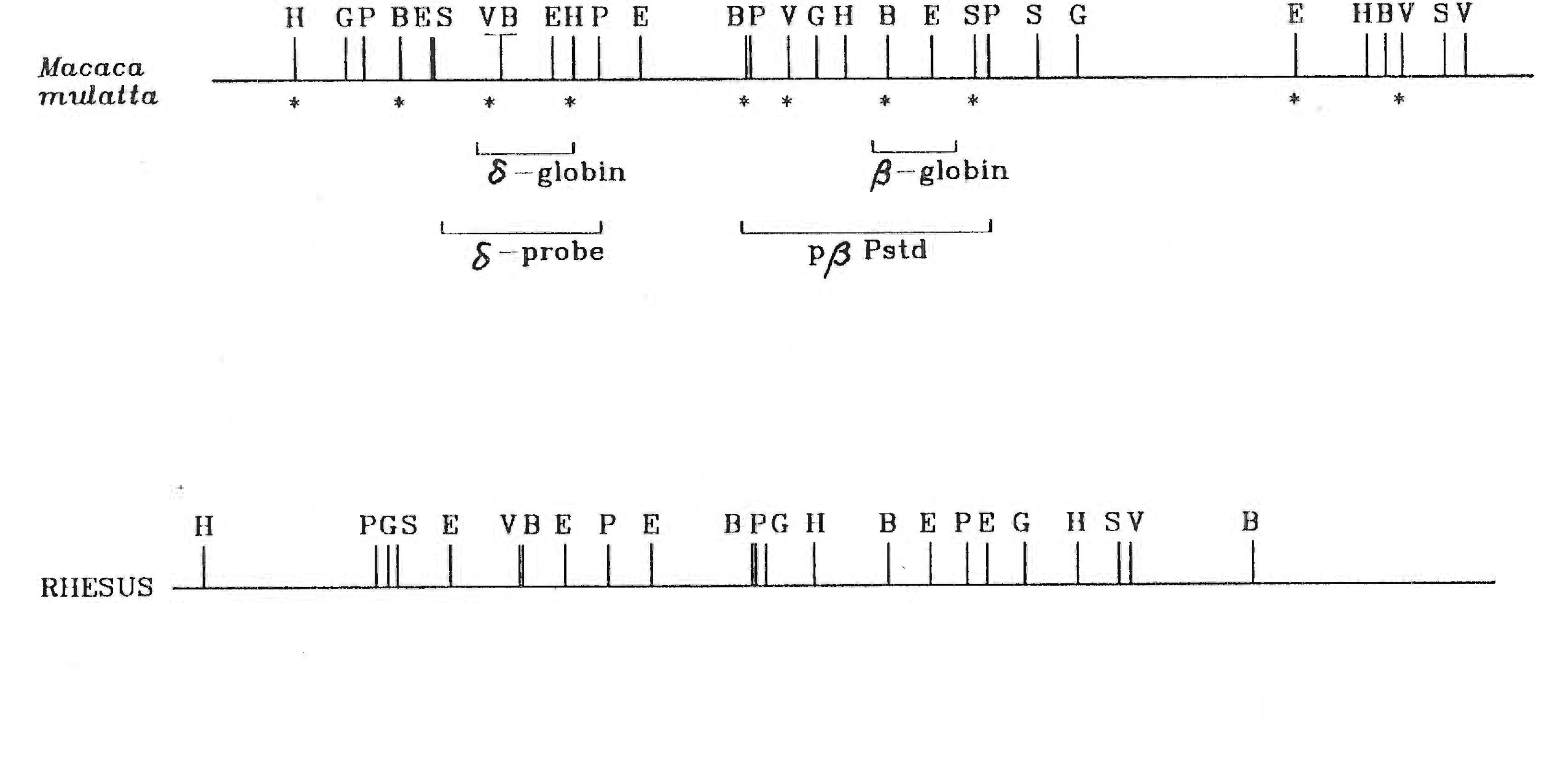

B K V H GP B E S V B EHP E BP GH B E P ES GH V B S V
+10 kb
*
*
**
*
δ−globin
β−globin
δ−probe
p β Pstd
BamHI
SacI
EcoRI
SacI
PvuII
MFu01 +,− +,+ +,− −,− +,−
MFu02 +,− +,+ +,− +,− +,+
MFu05 +,− +,− −,− +,+ +,+
MFu06 +,− +,+ −,− +,− +,+
MFu07 −,− +,+ −,− −,− +,+
0
5
10
15
20
25
30

Figure 3.7. Restriction Map of Adult Non-Alpha Loci of *Macaca mulatta.*

Beta and delta globin loci were probed using pβPstd and δ-probe. A composite map of the rhesus data is compared to the map of Martin (1982). Restriction enzymes are abbreviated as follows: B = BamHI, G = BglII, E = EcoRI, H = HindIII, P = PstI, V = PvuII, and S = SacI. Polymorphic sites are represented by an asterisk (*).


Macaca mulatta
H GP BES VB EHP E BP V G H B E SP S G E HBV SV
δ-globin
β-globin
δ-probe
pβ Pstd
RHESUS
H PGS E VB E P E BPG H B E PE G H SV B
0
5
10
15
20
25
30

<u>Figure 3.8.</u> Restriction Map of Adult Non-Alpha Loci of *Macaca nemestrina*.

Beta and delta globin loci were probed using pβPstd and δ-probe. Restriction enzymes are abbreviated as follows: B = BamHI, G = BglII, E = EcoRI, H = HindIII, P = PstI, V = PvuII, and S = SacI. Polymorphic sites are represented by an asterisk (*). A deletion site is represented by a triangle.

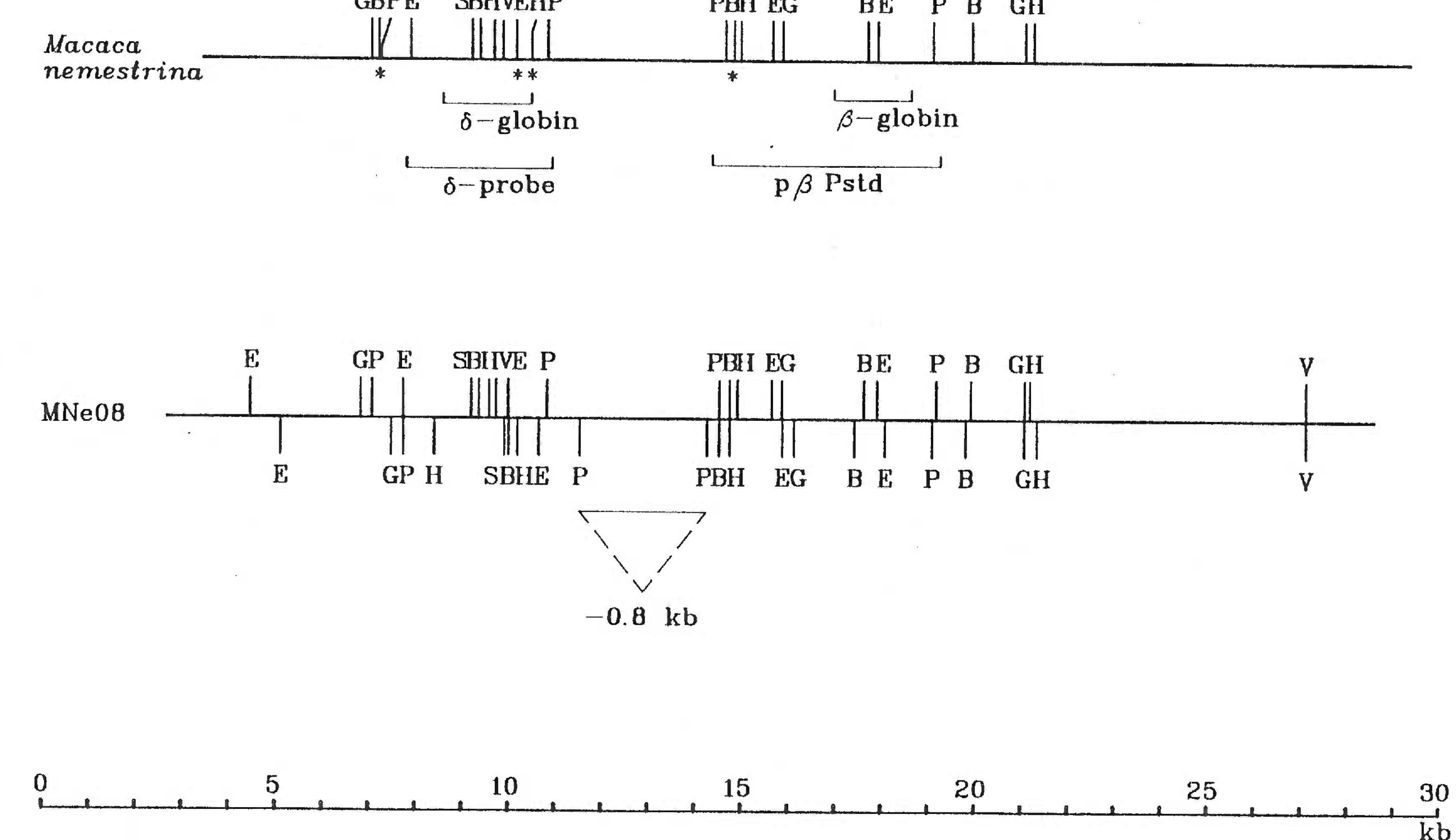

Macaca nemestrina
GBPE
SBHVEHP
PBHI EG
BE
P
B
GH
δ−globin
β−globin
δ−probe
pβ Pstd
MNe08
E
GP E
SBHVE P
PBHI EG
BE
P
B
GH
V
E
GP H
SBHE P
PBH
EG
B E
P B
GH
V
−0.8 kb
0
5
10
15
20
25
30
kb

Figure 4.1. Common Macaque Restriction Haplotype of the Adult Non-Alpha Globin Loci.

The three macaque species *Macaca fuscata*, *mulatta*, and *nemestrina* demonstrated a number of common restriction sites which are presented. The common restriction haplotype of macaques is compared to the restriction maps of other Old World Monkeys as determined by Martin (1982). Abbreviations: B = BamHI, G = BglII, E = EcoRI, H = HindIII, P = PstI, V = PvuII, and S = SacI.

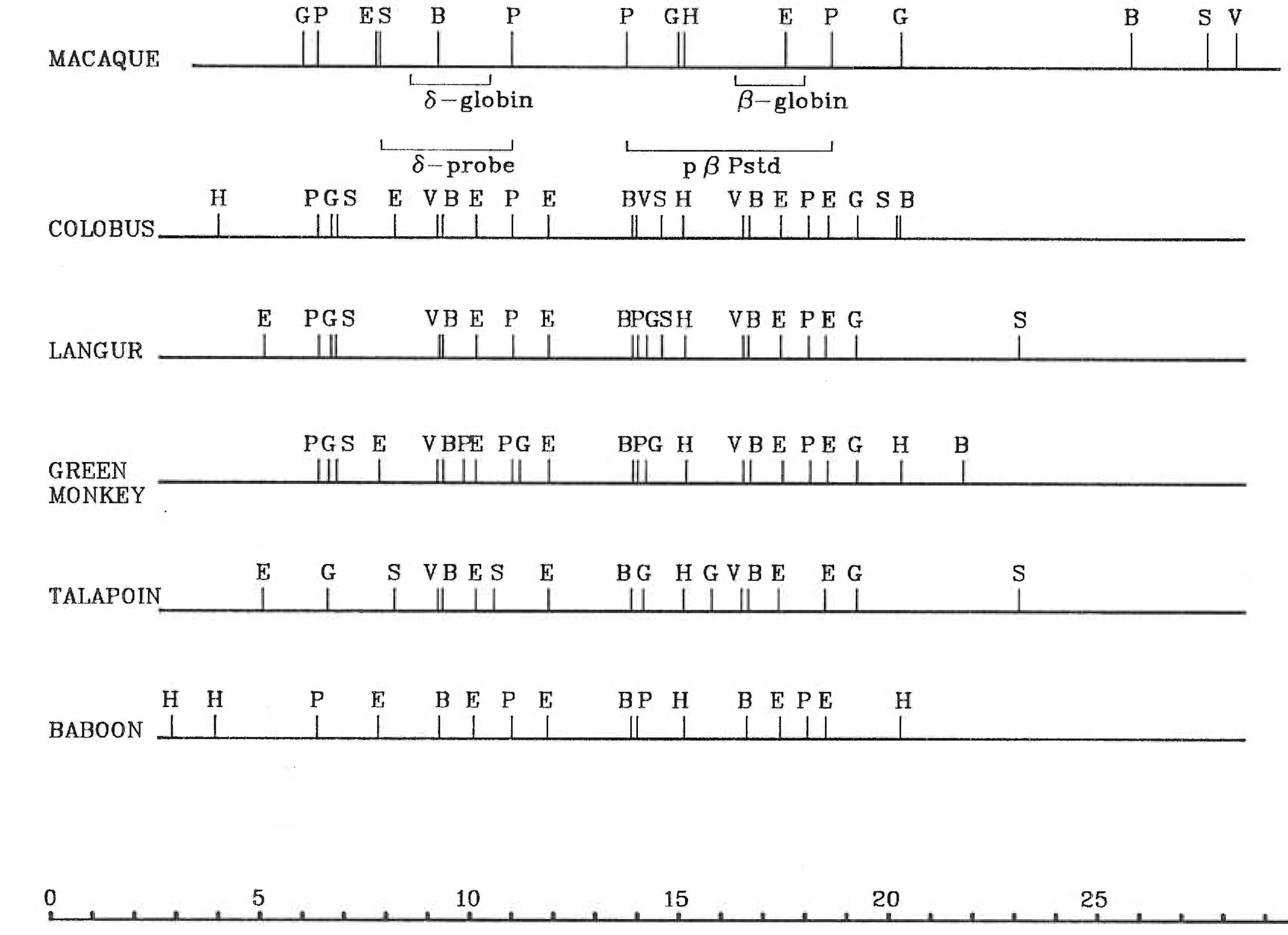
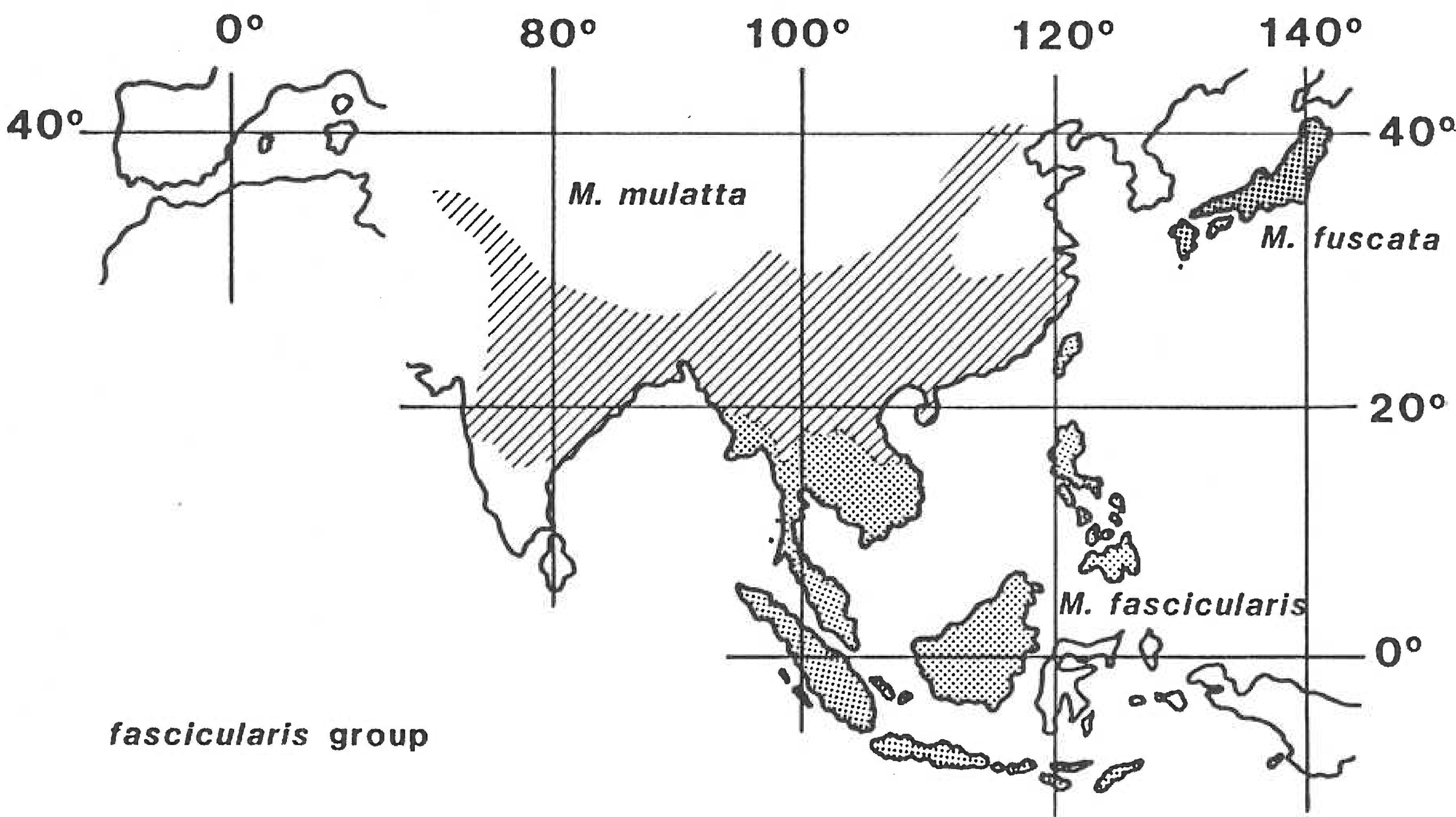

MACAQUE
G P E S B P P G H E P G B S V
δ−globin
β−globin
δ−probe
p β Pstd
COLOBUS
H P G S E V B E P E B V S H V B E P E G S B
LANGUR
E P G S V B E P E B P G S H V B E P E G S
GREEN MONKEY
P G S E V B P E P G E B P G H V B E P E G H B
TALAPOIN
E G S V B E S E B G H G V B E E G S
BABOON
H H P E B E P E B P H B E P E H
0 5 10 15 20 25 30

Figure 4.2. Geographical Distribution of Macaque Species Studied: *fascicularis* Group.

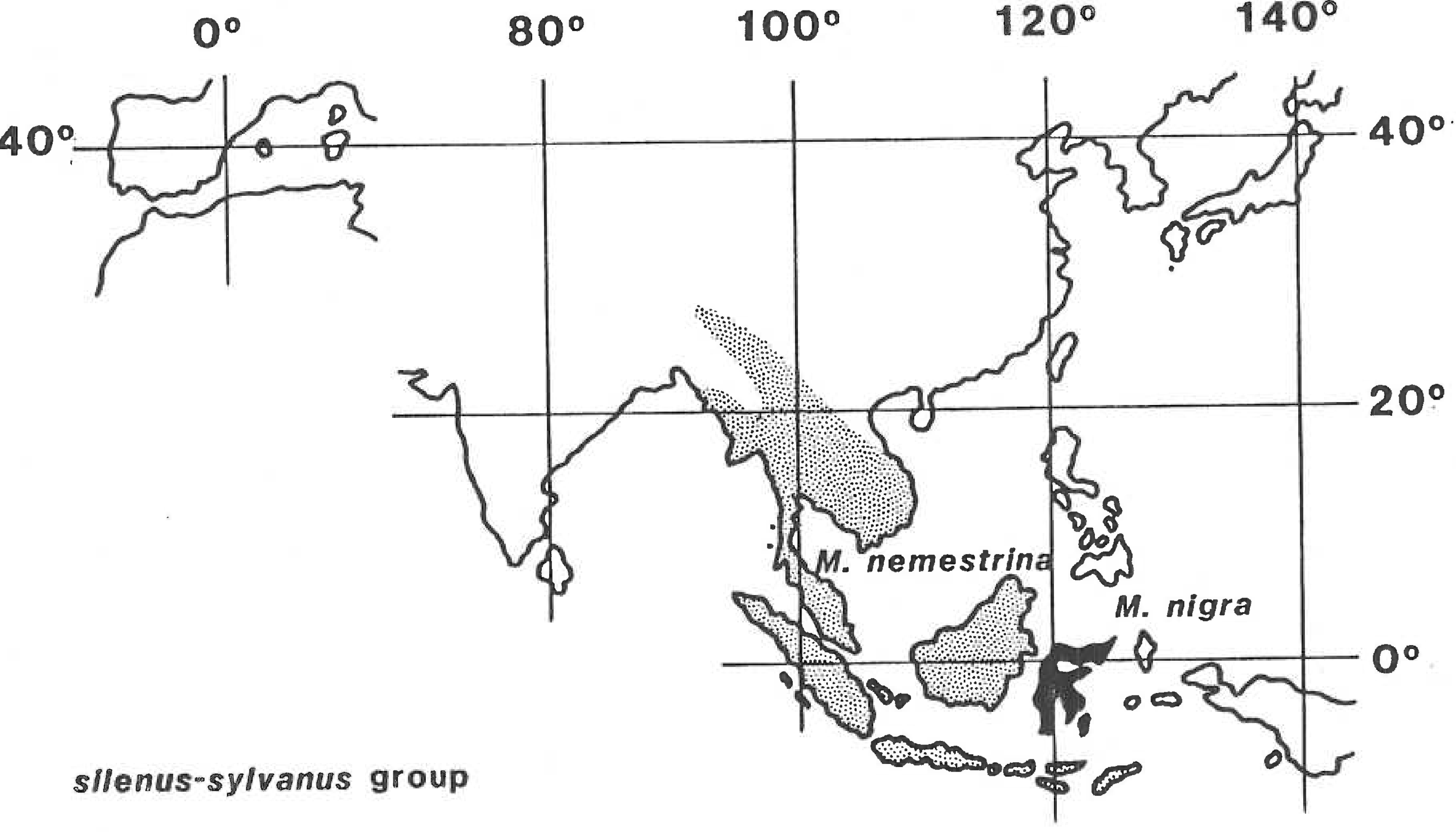
Three macaque species are included in the *fascicularis* group: *Macaca fascicularis* (stippled), *Macaca mulatta* (lined), *Macaca fuscata* (hatched). (Adapted from Fooden, 1980).



Geographical Distribution of Macaque Species Studied

Figure 4.3. Geographical Distribution of Macaque Species Studied: *silenus-sylvanus* Group.

Two macaque species are included in the *silenus*-*sylvanus* group: *Macaca nemestrina* (stippled) and *Macaca* *nigra* (black). (Adapted from Fooden, 1980)



Geographical Distribution of Macaque Species Studied

Macaca mulatta	MMu01	2380	F	Colony Born: ORPRC
	MMu02	4752	F	Colony Born: ORPRC
	MMu03	6750	F	Unknown
	MMu04	7691	F	India
	MMu05	7934	F	Colony Born: ORPRC
	MMu06	8230	F	Colony Born: ORPRC
	MMu07	8453	F	India
	MMu08	8517	F	India
	MMu09	8549	F	Colony Born:ORPRC
	MMu10	9002	F	India
Macaca nemestrina	MNe01	12530	F	Sulawesi/Indonesia
	MNe02	12533	F	Sulawesi/Indonesia
	MNe03	12528	F	Sulawesi/Indonesia
	MNe04	10833	F	Unknown
	MNe05	9468	F	Unknown
	MNe06	12531	F	Sulawesi/Indonesia
	MNe07	12532	F	Sulawesi/Indonesia
	MNe08	12529	F	Sulawesi/Indonesia
Macaca nigra	MNi01	12978	M	Unkown
	MNi02	13065	M	Colony Born: Yerkes
	MNi03	12981	M	Unknown
	MNi04	10757	M	Unknown
	MNi05	13034	F	Unknown

	MNi06	13038	M	Colony Born: Yerkes
	MNi07	13037	F	Colony Born: Yerkes
	MNi08	12972	M	Sulawesi/Indonesia
	MNi09	11720	M	Colony Born: Seattle Zoo
	MNi10	12958	F	Colony Born: Tulsa Zoo
Homo sapiens	HSa01	N/A	M	India
	HSa02	N/A	M	Wyoming, U.S.A.
	HSa03	N/A	M	Missouri, U.S.A.
	HSa04	N/A	F	California, U.S.A.
	HSa05	N/A	F	Oregon, U.S.A.
	HSa06	N/A	F	Oregon, U.S.A.
	HSa07	N/A	F	United Kingdom
	HSa08	N/A	M	Washington, U.S.A.
	HSa09	N/A	F	Unknown
	HSa10	N/A	M	California, U.S.A.

APPENDIX B: PREPARATION OF RNase-FREE DIALYSIS TUBING

The following protocol was provided by L. Nicolaisen (personal communication).

Spectra-por cellulose dialysis tubing with a molecular weight cut off of 12,000 to 14,000 was used. The tubing was handled with gloves and sterile solutions were used throughout.

Dialysis tubing was cut in 7-9 inch lengths. It was placed in a 2 l Pyrex beaker and boiled in 2 % Ammonium-bicarbonate for 10 minutes under a hood then drained well. A 95 % Ethanol wash was also performed by boiling. This was followed by boiling twice in deionized water. A final boiling step was performed in 10 mM EDTA pH 7.4.

The tubing was stored in 10 mM EDTA, pH 7.4, 30% EtOH at 4 C. It was rinsed with sterile water prior to use.