INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17” x 23” black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Recombination within the human \( \beta \)-globin gene cluster

Anand, Rene, Ph.D.
The Ohio State University, 1989
RECOMBINATION WITHIN THE HUMAN $\beta$-GLOBIN GENE CLUSTER

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Rene Anand, M.S.

The Ohio State University

1989

Dissertation Committee:
C.A. Breitenberger
G.A. Marzluf
R.P. Swenson

Approved by

Advisor

Department of Biochemistry
To my dearest Mum, Joyce
I'd like to take this opportunity to thank my mentor and more especially my good friend Elio Vanin for sharing his wisdom and optimism throughout my career as a graduate student. It was a wonderful feeling to be able to share lunch, friendship and science all at the same time with you. Thank you Vidya Rao for taking care of many an experiment when I was out of town.

Thanks to all those friends from our lunch club for making life as a graduate student fun.

A special thank you to an unmentionable special somebody that has made all of this worthwhile. I'd also like to thank dearest Mum for all her love, support and encouragement all the way from India. I am also grateful to Dr. H.H. Kazazian, Corinne Wong, and Corinna Boehm at the Johns Hopkins University School of Medicine, for the initial characterization of the deletion thalassemia and for providing us with some of the sequence data analyzed in this thesis. Thank you Piri Welcsh for letting me analyze and use your in vitro MAR data. Thanks to Dr. F.S. Collins at University of Michigan, for providing us with a jumping library. Finally, I'd like to acknowledge Dr. M.T. Muller and Jeffrey Spitzner for providing us with chicken topoisomerase II and for their advice on carrying out the cleavage reactions.
VITA

NAME: Rene Anand
DATE OF BIRTH: November 3, 1960
PLACE OF BIRTH: Mysore, India
ADDRESS: Department of Biochemistry
Room 746, Biological Sciences Building
484 W. 12th Avenue
Ohio State University
Columbus, Ohio 43210

EDUCATION:

1978 - 1981 B.S (Chemistry), Loyola College, Madras, India
1981 - 1983 M.S. (Chemistry), Indian Institute of Technology, Madras, India
1983-PRESENT Graduate Student, Department of Biochemistry


FIELD OF STUDY: Biochemistry
# TABLE OF CONTENTS

DEDICATIONS ...................................................................... ii

ACKNOWLEDGEMENTS ................................................................... iii

VITA ................................................................................ iv

LIST OF TABLES ...................................................................... vii

LIST OF FIGURES ................................................................... viii

ABBREVIATIONS ...................................................................... xii

CHAPTER PAGE

I. INTRODUCTION ................................................................. 1
   A. Purpose, significance and organization of the thesis .................................... 1
   B. Research Summary ................................................................................ 3

II. REVIEW OF DELETIONS AND OTHER RECOMBINATIONAL EVENTS IN MAMMALIAN CELLS .............................................. 4
   A. Introduction .................................................................................. 4
   B. The human globin genes and their developmental expression .......................... 4
   C. Deletions within the β-globin gene cluster .............................................. 10
   D. Recombinational events in other mammalian gene systems .......................... 24

III. GENERAL MATERIALS AND METHODS .................................. 30

IV. MOLECULAR CHARACTERIZATION OF A β⁰-THALASSEMIA RESULTING FROM A 1.4 KB DELETION .............................................. 38
   A. Introduction .................................................................................. 38
   B. Materials and Methods ...................................................................... 41
   C. Results and Discussion ...................................................................... 43

V. LOCALIZATION OF A NUCLEAR MATRIX ATTACHMENT REGION WITHIN THE β-GLOBIN GENE ................................................. 72
   A. Introduction .................................................................................. 72
   B. Materials and Methods ...................................................................... 76
   C. Results and Discussion ...................................................................... 79

VI. SEQUENCE CHARACTERIZATION OF TWO REGIONS LOCATED UPSTREAM OF THE β-GLOBIN GENE THAT SHOW EXTENSIVE SEQUENCE POLYMORPHISM ..................................................... 97
   A. Introduction .................................................................................. 97
   B. Materials and Methods ...................................................................... 104
   C. Results and Discussion ...................................................................... 106
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Results of genomic DNA analysis of the 1.4 kb deletion thalassemia</td>
<td>48</td>
</tr>
<tr>
<td>2. Results of densitometric scan analysis of the <em>in vitro</em> nuclear matrix assays</td>
<td>84</td>
</tr>
<tr>
<td>3. List of DNA fragments used for sequencing the -1.4 kb region</td>
<td>107</td>
</tr>
<tr>
<td>4. Summary of the sequence analysis of the -1.4 kb and -550 bp regions</td>
<td>111</td>
</tr>
<tr>
<td>5. List of DNA fragments used in the topoisomerase II cleavage reactions</td>
<td>127</td>
</tr>
<tr>
<td>6. Summary of the Maxam and Gilbert base specific reactions for sequencing end-labeled DNA used in this thesis</td>
<td>185</td>
</tr>
<tr>
<td>7. List of the reaction times and conditions used for the Maxam and Gilbert reactions</td>
<td>187</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Map of the α- and β-globin gene clusters</td>
</tr>
<tr>
<td>2.</td>
<td>Map of the β-globin gene cluster and schematic representation of the various deletions characterized within this locus</td>
</tr>
<tr>
<td>3.</td>
<td>Model of a possible mechanism for generating large deletions by the loss of a chromatin loop during replication</td>
</tr>
<tr>
<td>5.</td>
<td>Autoradiogram of DNA digests hybridized with βIVS-2 probe</td>
</tr>
<tr>
<td>6.</td>
<td>Autoradiogram of DNA digests hybridized with the 1.8 kb Bam HI and cDNA probes</td>
</tr>
<tr>
<td>7.</td>
<td>Autoradiogram of restriction digests of the DNA from the two phage clones isolated from the library, hybridized with the 1.8 kb Bam HI probe</td>
</tr>
<tr>
<td>8.</td>
<td>Autoradiogram of restriction digests of the DNA from the two phage clones isolated from the library, hybridized with the 1.8 kb Bam HI probe</td>
</tr>
<tr>
<td>9.</td>
<td>Restriction enzyme map of the normal and abnormal chromosomes</td>
</tr>
<tr>
<td>10.</td>
<td>Autoradiogram of the sequencing gel containing the sequence across the breakpoint junction</td>
</tr>
<tr>
<td>11.</td>
<td>Sequence across the deletion breakpoint junction of the 1.4 kb β⁰-thalassemia and the corresponding normal 5' and 3'DNAs</td>
</tr>
<tr>
<td>12.</td>
<td>Proposed model for the production of the small deletions found within the human β-globin gene</td>
</tr>
<tr>
<td>13.</td>
<td>Schematic diagram of the β-globin gene and the location of a matrix attachment site as defined by one of the breakpoints of the three deletions</td>
</tr>
</tbody>
</table>
27. Autoradiogram of the 8% sequencing gel containing the fractionated topoisomerase II cleavage products of the p0.8 AccI - Eco RI and the p4.2 Dra I - Sph I fragments ........................................ 138

28. Autoradiogram of 8% sequencing gels containing the fractionated topoisomerase II cleavage products of the p5.5 Acc I - Sph I and p5.5 Hpa I - Acc I fragments ........................................ 140

29. Sequence of the Hpa I - Acc I fragment showing the location of the topoisomerase II cleavage sites mapped ........................................ 142

30. Autoradiogram of 8% sequencing gel containing the fractionated Human, Chicken, and Drosophila topoisomerase II cleavage products of the p5.5 Acc I - Sph I fragment ........................................ 146

31. Schematic diagram of the β-globin gene showing the location of the two strong topoisomerase II sites flanking the 550 bp region shown to exhibit a high degree of sequence polymorphism .............................. 148

32. Schematic diagram showing a possible mechanism by which topoisomerase II could initiate recombination between two homologues to account for the sequence polymorphism observed at certain regions ............... 150

33. Schematic diagram of the β-globin gene cluster showing the approximate location of the 3' breakpoint of the Scotch - Irish deletion thalassemia, along with some of the restriction sites mapped around the 3' breakpoint ........................................ 155

34. Autoradiogram results of the DNA digests from the Scotch - Irish thalassemia and normal DNA hybridized with the pRK 28 probe ........................................ 161

35. Map of the 17 kb Bam HI fragment isolated from the library constructed using thal-3 DNA ............... 164

36. Autoradiogram of dot blots used to identify unique probes 3' to the β-globin gene ....................... 168

37. Schematic diagram of the construction of a partial Mbo I jumping library ........................................ 172
38. Autoradiogram of the DNA isolated from the jumping clones, hybridized to the 3.3 kb Eco RI and 4.0 kb Eco RI probes.......................... 176

39. Autoradiogram of the jumping clones hybridized to $^{32}$P labeled human genomic DNA............................... 179

40. Photographs of 0.8% agarose gels with restriction digests the phage clones BB1 and the plasmid clone containing the 8.0 kb Eco RI fragment.............. 181

41. Map of the jump clone BB1 isolated with the 3.3 kb Eco RI probe......................................................... 183
ABBREVIATIONS

ATP: Adenosine Triphosphate
BSA: Bovine Serum Albumin
bp: Base pairs
cDNA: Complementary DNA
cpm: counts per minute
DNA: Deoxyribonucleic Acid
dpm: Disintegrations per minute
dCTP: Deoxycytidine Triphosphate
EDTA: Ethylenediamine Tetraacetic Acid
Hb: Hemoglobin
HPFH: Hereditary persistence of fetal hemoglobin
IVS: Intervening sequence
kb: Kilobase pairs
ml: Millilitres
mM: Millimolar
mAMSA: [4'-(9-acridinylamino)methanesulfon-m-anisidide]
MAR: Matrix attachment region
ng: Nanogram
RNA: Ribonucleic acid
RFLP: Restriction Fragment Length Polymorphism
SAR: Scaffold attachment region
SDS: Sodium Dodecylsulphate
Thal: Thalassemia
ug: Microgram
CHAPTER I

Introduction

A. Purpose, significance and organization of this thesis

The overall objective of the work described in this thesis has been to obtain a better understanding of the recombinational mechanisms involved in producing genomic deletions within mammalian cells. To this end, the human \( \beta \)-globin gene cluster which has been very well characterized, was chosen as a model system. Molecular aberrations in any of the genes within this locus result in characteristic phenotypes termed either "thalassemias" or hereditary persistence of hemoglobin. Characterization of recombinational events like deletions within this cluster would provide investigators currently involved in studying mammalian recombination with a better understanding of the processes involved which in turn may aid their progress towards providing effective gene therapy in the near future.

Chapter II provides an introduction to the organization of the \( \beta \)-globin gene cluster, the expression of the globin genes during development, a review of all the deletions characterized within the \( \beta \)-globin gene cluster, possible mechanisms for their production, and generalized recombination in mammalian systems. Chapter III contains all the general materials and methods used throughout this project, though, materials and methods specific
to any one aspect of a chapter have been described within that chapter. Chapter IV describes the molecular characterization of a β⁰-thalassemia resulting from a 1.4 kb deletion within an American black individual. Chapter IV also describes a model whereby these small deletions could arise during replication. Chapter V describes the identification of a nuclear matrix attachment point within the second intervening sequence of the β-globin gene cluster, which was a prediction of the deletion-replication model proposed in chapter IV. Described in chapter VI is the sequence analysis of two hypervariable regions, from different chromosomes, found 5' to the β-globin gene. These regions are thought to contain some of the sequences involved in frequent recombination thus being responsible for the linkage equilibrium between the 5' and 3' haplotypes found within the β-globin gene cluster. Chapter VII describes experiments undertaken to investigate the possible role of eucaryotic topoisomerase II in the production of these deletions and the hypervariability observed within the 5' region of the β-globin gene. Finally, chapter VIII describes our attempts to characterize a γβ⁰-thalassemia of Scotch-Irish descent that removes the entire β-globin gene cluster.
B. Research summary

This thesis describes the molecular characterization of a \( \beta^0 \)-thalassemia resulting from a 1.4 kb deletion. Using in vivo methods, the second intervening sequence of the \( \beta \)-globin gene is shown to contain a nuclear matrix attachment site. In addition, two regions located approximately 1.4 kb and 550 bp 5' to the \( \beta \)-globin gene have been shown to exhibit a great degree of sequence polymorphism, which may be the result of the frequent recombination postulated to take place in this region. The role of eucaryotic topoisomerase II has also been examined in generating some of the known thalassemic deletion breakpoints as well as in the generation of the sequence polymorphisms found 5' to the \( \beta \)-globin gene, by localization of the in vitro cleavage sites within these regions of DNA.
CHAPTER II

Review of deletions and other recombinational events in mammalian cells

A. INTRODUCTION

Recombination within the β-globin gene cluster

What follows is a brief introduction to the organization of the human globin genes, their expression and interesting features of their intergenic nucleotide sequence. Most of the emphasis has been placed on the β-globin gene cluster as it was the system used to do the work described in this thesis. The α-globin gene cluster has been introduced only to give the reader a complete picture of the coordinated expression of both the α-like and β-like globin genes and the various forms of hemoglobin which are formed during human development.

B. The human globin genes and their developmental expression

Human hemoglobin is a tetrameric protein which is made up of two α-like and two β-like globin polypeptide chains. The genes that encode the α-like globin proteins are located on the short arm of chromosome 16, span about 30 kb of DNA, and consist of
three expressed genes, ζ, α2 and α1, and two unexpressed pseudogenes ψζ and ψα1 (Bunn et al. 1986) (Figure 1).

Another pseudogene, ψζ2, has recently been identified within this cluster (Hardison et al. 1986). The genes that encode the β-like globin proteins, on the other hand, are located on the short arm of chromosome 11 (Gussela et al. 1981, Lebo et al. 1979, Sanders et al. 1980). The β-globin gene cluster spans about 60 kb of DNA and consists of five expressed genes, the embryonic gene, ε; two fetal genes, Gγ, Aγ; two adult genes, δ, β; and a pseudogene, ψβ (Fritsch et al. 1980) (Figure 1). Early in development, the embryonic hemoglobin z2e2 tetramer is formed and as development progresses the embryonic genes are shut off and the α and γ genes are expressed. During the transition period, both the Hb Portland (ζ2γ2) and the Hb Gower 2 (α2ε2) can be found. The fetal hemoglobin Hb F (α2γ2) is the major hemoglobin found throughout the rest of fetal development. Finally in adulthood, hemoglobin Hb A (α2β2) and a minor amount of Hb A2 (α2δ2) are formed as a result of the expression of the adult globin genes. Thus, in both clusters, the genes are arranged 5' to 3' in the order in which they are expressed during development. The globin genes are not only developmentally expressed but their sites of synthesis also change, the major sites of erythropoiesis shifting from the yolk
Figure 1: Map of the α- and β-globin gene clusters.

The top line shows a scale in kilobases of DNA. The second line shows a schematic representation of the α-like globin genes. The solid boxes represent the exons and the open boxes the introns. Similarly the third line shows a schematic representation of the β-like globin genes.
THE GLOBIN GENES

\[ \begin{align*}
\text{\( \zeta \)} & \quad \psi & \quad \psi_{\alpha_1} & \quad \alpha_2 & \quad \alpha_1 \\
\text{\( \alpha \)} & & & & \\
\text{\( \epsilon \)} & \quad \gamma & \quad \gamma & \quad \psi_{\beta_1} & \quad \beta \\
\text{\( \beta \)} & & & & 
\end{align*} \]
sac during the embryonic stage, to the liver in the fetal stage, and finally to the bone marrow in adulthood (Weatherall et al. 1981). The tissue-specific and developmental expression of the globin genes has been the focus of a number of laboratories, but the factors and regulatory sequences involved in mediating this highly regulated process have still to be fully characterized.

The human β-globin gene cluster

As a result of the combined efforts of a number of laboratories, the entire human β-globin gene cluster has been cloned (Fritsch et al. 1980) and sequenced (Efstradiatis et al. 1980; Poncz et al. 1983; Lawn et al. 1980; Spritz et al. 1980; Slightom et al. 1980). All the globin genes have been found to be structurally very similar in that they all have two intervening sequences that divide the coding region into three exons, the first intervening sequence being much shorter than the second intervening sequence. The introns follow the "GT - AG" rule in that they have the dinucleotides GT and AG at the beginning and the end respectively, of every intron, sequences thought to be essential for proper splicing of introns. The introns also contain the sequence ACTT/CT/C or ATCT/CT/C near the 3' end of each intron. This sequence has been shown to be involved in the formation of a branched RNA intermediate in the
splicing reaction (Ruskin et al. 1984; Keller et al. 1984). Other conserved sequences include the "ATA" box located approximately 30 nucleotides 5' to the cap site which seems important for the accuracy of transcription initiation, and the "CCAAT" box, located approximately 70 nucleotides 5' to the cap site, which has been shown to be important for efficient transcription of the globin genes (Dierks et al. 1981; Grosveld et al. 1982). Aside from the globin genes themselves, the cluster has also been shown to contain a number of repetitive sequences that are found interspersed throughout the human genome. Nine copies of the highly repetitive Alu I family sequences, which are approximately 300 bp in length and are present in 300,000 copies per haploid human genome, have been found within the human β-globin gene cluster. Of these nine, three copies have been localized 5' to the ε-globin gene, one 5' to the Gγ gene, one 5' to the ψβ gene, two 5' to the δ gene and two 3' to the β-globin gene (Fritsch et al. 1980; Coggins et al. 1980). Two copies of long stretches of intermediate repetitive sequences termed Kpn I family sequences, usually 1.5 kb to over 6 kb (Shafit-Zagardo et al. 1982), have also been found within this cluster. These sequences are located 5' to the q globin gene (Jagadeeswaran et al. 1981), and 3' to the β-globin gene (Kaufman et al. 1980; Adams et al. 1980). Other interesting features of the intergenic DNA are described in chapter VI.
C. Deletions within the β-globin gene cluster

Thalassemia is a genetic disease in humans resulting from the altered expression of one or more globin genes in either the α-like or the β-like globin locus. Analysis of thalassemias has shown that they can be caused by a variety of mutations ranging from a point mutation to deletions of >100 kb of DNA. Historically, the globin chain variants resulting from point mutations in the translated regions of the genes have been important in the structure-function analysis of the hemoglobin molecule. On the other hand, point mutations in the non-translated regions have been very important in defining sequence elements important for gene transcription and processing of precursor globin mRNA. These mutations have been extensively reviewed (see Bunn et al. 1986).

As can be seen from Figure 2, a number of different deletions have been reported in the β-globin gene cluster that range in size from about 100 bp to >100 kb. This, together with the fact that the cluster has been so well characterized (approximately 60 kb of sequence spanning the entire cluster is known), has enabled a number of laboratories including our own, to use the β-globin gene cluster as a model system to study recombination in mammalian cells. Some of the terms used in describing these deletions are defined below:
Figure 2: Map of the β-globin gene cluster and a schematic representation of the various deletions characterized within this locus.

The top line shows a map of the β-globin gene cluster. The various genes are represented by closed boxes. The closed bars represent the extent of each deletion shown, whilst the open bars indicate uncertainty in the exact location of a breakpoint. Breaks in the closed bars indicate that various amounts of DNA have been omitted from this map. Thal stands for thalassemia and HPFH stands for hereditary persistence of fetal hemoglobin. The designation β° indicates the absent synthesis of β-globin from the chromosome bearing the mutation. Similarly Gγ(Aγδβ)° indicates the synthesis of Gγ globin but no synthesis of Aγδβ globins. The Lepore and Kenya deletions cause the production of the globin fusion chain hemoglobins, Hb Lepore and Hb Kenya.
FIGURE 2
Normal 5' DNA (sequence): corresponds to 5' DNA which was originally involved in the deletion, prior to recombination.

Normal 3' DNA (sequence): corresponds to 3' DNA which was originally involved in the deletion, prior to the recombination.

5' and 3' breakpoints: corresponds to the nucleotides at which the breakages occurred on the normal DNA. The 5' and 3' refer to the relative positions of the two breakpoints with respect to each other and in keeping with the transcriptional direction of a nearby gene.

Deletion breakpoint sequence: corresponds to the sequence which was produced by the recombination.

In studying deletions there are two levels of complexity which need to be considered. First, are deletions produced as a result of an unequal homologous crossover event between similar sequences or do they result from non-homologous breakage and reunion events. If the deletions were the result of unequal homologous crossover, one would expect the normal 5' and 3' sequences to share a great deal of homology. If, on the other hand, deletions were the result of non-homologous breakage and reunion events one would expect to see little or no homology.
between the normal 5' and 3' sequences. Furthermore, comparison of the normal 5' and 3' sequences with the deletion breakpoint sequence would indicate whether the recombinational events were "clean" i.e if any spurious nucleotides had been added during the recombinational event. The next section describes such an analysis of all the deletions that have been studied at the sequence level within the β-globin gene cluster (shown in Figure 2).

The large deletion thalassemias γδβ°-thal 1 and thal 2 were shown to remove 99.6 and 95.4 kb of DNA respectively. Thal 1 starts 5' to the β-globin gene and extends in the 5' direction removing all of the β-like globin genes except for the adult β globin gene, whereas thal 2 starts within the β-globin gene and also removes all of the β-like globin genes (Vanin et al. 1983; Taramelli et al. 1986). In the case of an atypical β-thalassemia of Czechoslovakian origin, the deletion starts within β-IVS 2 and removes 4237 bp of DNA consisting of intergenic DNA, the β-globin 5' promoter, mRNA cap site, 5' untranslated region, exon I, β-IVS 1, exon II, and 470 bp of β-IVS 2 (Popovich et al. 1986). A β°-thalassemia of Dutch origin has been described in which a 12.6 kb deletion starts 5' to the β-globin gene and extends in the 3' direction, its 3' breakpoint being located within the Kpn I repeat found 3' to the β-globin gene (Gilman et al. 1987). In the case of a Chinese \(G_\gamma^+ (\gamma\delta\beta)^o\) thalassemia, the deletion has been found to be at least 80 kb long and has its
5' breakpoint in the $\alpha\gamma$-globin gene and extends in the 3' direction (Mager et al. 1985). An Indian $\alpha\gamma\delta\beta$ thalassemia of complex etiology has been reported in which two deletions have occurred, one with the two breakpoints within the $\alpha\gamma$ gene and the other within the $\delta$ gene and extending 3' into the $\beta$ gene, the intermediate DNA between these two deletions being inverted (Jennings et al. 1985). Another Indian $\beta$-thalassemia has been characterized in which the deletion removes 619 bp of DNA starting within $\beta$-IVS 2 and extending in the 3' direction (Spritz et al. 1982). More recently, a Japanese $\delta\beta$ thalassemia has been characterized in which the 5' breakpoint lies 2134 to 2137 bp 3' to the poly A site of the $\alpha\gamma$ globin gene and the deletion extends for an undetermined distance in the 3' direction (Shiokawa et al. 1988). A Sicilian deletion that removes 7,201 bp of DNA starting within the $\delta$ globin gene and extending in the 5' direction has also been recently characterized (Kulozik et al. 1988). Among the HPFHs, HPFH 3 of Indian origin has been characterized and shown to remove 48.5 kb of DNA the deletion starting 3' to the $\alpha\gamma$ globin gene and extending in the 3' direction (Henthorn et al. 1986).

Analysis of the normal 5' and 3' sequences for all of the deletions just described, has revealed that the deletions were the result of nonhomologous breakage and reunion events. Comparison of the normal 5' and 3' sequences to the breakpoint junction sequences for the various deletions has revealed that
the majority of the deletions were clean. The breakpoint junctions that had extraneous nucleotides added include the deletion breakpoint junctions of the Chinese $^A\gamma^\delta^\beta$ thalassemia, which contained a 36 to 41 bp insertion of unknown origin, the 0.6 kb Indian deletion thalassemia, which contained a heptanucleotide insertion of unknown origin, and one of the junctions of the Indian deletion/inversion thalassemias, which contained a hexanucleotide insertion of unknown origin. Thus not all of the nonhomologous recombinational events seem to be clean.

The second question which can be asked about the production of these deletions is how the normal 5' and 3' sequences are brought into close proximity so as to facilitate the recombinational event. A possible mechanism describing how this might happen has been described by Vanin et al. 1983, and has been instrumental in linking the production of these deletions to DNA replication. They showed that in two independent cases of $\gamma^\delta^\beta$ thalassemias (thal 1 and 2) and in two independent cases of HPFHs (HPFH-1 and HPFH-2), the 5' breakpoints were located approximately the same distance apart and in the same order along the DNA as their 3' breakpoints. The simplest interpretation of these results is that the same amount of DNA had been deleted in the two thalassemias, the same being true for the two HPFHs. The deletions were predicted to remove approximately the same amount of DNA in each of the deletions, the size of the deletions being
PLEASE NOTE:

Page(s) not included with original material and unavailable from author or university. Filmed as received.
the size of a chromatin loop. Figure 3 shows the model proposed by the authors as to how these deletions may have arisen. As shown in Figure 3, the breaks which give rise to these deletions occur at the same position with respect to the matrix attachment site but at different times during replication. Rejoining of the breaks would result in the loss of the intervening DNA, thus generating a series of deletions each of which would have deleted the same amount of DNA. This model predicts that the amount of DNA deleted in the two γδβ-thals and also in the two HPFHs would be approximately the size of a DNA loop or a multiple thereof. The authors speculate that these large deletions occur at different times during DNA replication at regions associated with the nuclear matrix. This model has been further strengthened by the fact that γδβ thalassemia 1 and 2 have subsequently been shown to be 99.4 kb and 95.9 kb in length respectively, by linking the normal 5' DNAs of the two thals to the β-globin gene cluster by cosmid walking techniques (Taramelli et al. 1986). In the case of the two HPFHs, the normal 3' DNAs have not yet been linked to the β-globin gene cluster, although, using pulse field gel electrophoresis techniques they have both been shown to be approximately 105 kb in length (Collins et al. 1987).

Mammalian DNA loops have been reported to range in size from 30 kb to 200 kb with an average loop size of 80 kb (Laemuli et al. 1978). Thus the size of the four deletions correlates very well with the average size of a DNA loop. The Chinese Aγδβ deletion
Figure 3: Model of possible mechanism for generating large deletions by the loss of a chromatin loop during replication.

Letters A, B, C, etc. indicate DNA prior to replication and letters A', B', C, etc. indicate DNA after one round of replication. Breakage and reunion points are indicated by asterisks, and the ellipses indicate the matrix attachment points along with the replication machinery (from Vanin et al., 1983.)
thalassemia also shares a similar relationship with the two HPFHs. A similar relationship has also been reported for a group of clustered deletions in the α-globin cluster, that is, the deletion breakpoints are staggered in a similar manner as was found for the large deletion thalassemias and HPFHs, though the deletions range in size from 20-30 kb (Nicholls et al. 1987).

All the deletions that have been discussed so far are produced as a result of nonhomologous breakage and reunion events, but not all deletions within this cluster occur via this mechanism. A smaller number of deletions within the β-globin gene cluster are thought to have arisen as a result of unequal homologous crossover events. The Hb Lepore in which the δ- and β-globin genes are absent but a hybrid δ/β-globin gene is present is thought to have arisen from an unequal homologous crossover between the δ- and β-globin genes, thereby deleting the intervening DNA (Figure 4). The reciprocal product, a β/δ-globin gene, has also been found and termed Hb - Anti-Lepore (Weatherall et al. 1981). Similarly a deletion within Asian individuals that results in Aγ-thalassemia is thought to have arisen by unequal homologous crossover of the two γ-globin genes, Aγ and Gγ (Sukamaran et al. 1983; Nakatsuji et al. 1984). As in the case of the Hb Lepore, individuals having chromosomes with the reciprocal product of such a crossover, i.e. two Gγ genes and one Aγ, have been reported (Trent et al. 1981). A third example of a deletion which is thought to be the
Figure 4: Mechanism of unequal crossing over responsible for the generation of hemoglobins Lepore, anti-Lepore, Kenya, and putative anti-Kenya.

Shown in the top half of the figure is the formation of the Lepore and anti-Lepore hybrid globin genes as a result of a reciprocal unequal homologous crossover between the δ and β-globin genes. The lower half of the figure shows the formation of the Kenya and putative anti-Kenya hybrid globin genes also by an unequal homologous crossover event but between the Ay- and the β-globin genes (adapted from Weatherall et al. 1981).
FIGURE 4
result of an unequal homologous crossover is the Hb Kenya involving a crossover between the $\alpha_\gamma$- and the $\beta$-globin genes (Kendall et al. 1973; Smith et al. 1973). In this case the reciprocal product Hb-Anti-Kenya has yet to be found.

D. Recombinational events in other mammalian gene systems

Deletions

Current models of recombination in mammalian systems have mostly been inferred from the study of recombination within lower eucaryotes like fungi and yeast. Numerous examples of recombinational events like translocations, deletions, and gene conversions have either been implied or shown to occur within mammalian cells. As with the human $\beta$-globin gene cluster, a number of deletions involving the human $\alpha$-globin gene cluster have also been characterized. Among the $\alpha$-thalassemias, a $\alpha^0$-thalassemia of Mediterranean origin has been characterized in which a deletion of 20.5 kb starts within an Alu I family sequence located between the $\zeta_2$ and $\zeta_1$ globin genes and extends in the 3' direction, partially deleting the $\alpha_1$-globin gene (Nicholls et al. 1985). The normal 5' and 3' DNAs seem to have a fair degree of homology that might have been involved in generating the deletion. In the case of the $\alpha^{MED}$ thalassemia, the deletion is approximately 16 kb in length, has
its 5' breakpoint located upstream of the $\psi_{\zeta 1}$ gene and extends in the 3' direction (Nicholls et al., 1987). The deletion breakpoint junction contains an insertion of 134 bp and the deletion seems to have been generated as a result of a homologous crossover event. The $\alpha^{SEA}$ thalassemia, is caused by a deletion of approximately 18 kb that begins just upstream of the $\psi_{\zeta 1}$ gene and extends in the 3' direction, its 3' breakpoint being located within an Alu I family sequence (Nicholls et al., 1987). The deletion was generated as a result of a nonhomologous breakage and reunion event. Another $\alpha$-thalassemia termed ($\alpha$)RA has been shown to delete 62 kb of DNA, its 3' breakpoint being located upstream of the $\psi_{\zeta 1}$ gene and the deletion extending in the 5' direction. Both the 5' and 3' breakpoints in this case lie within Alu I family sequences and the deletion seems to have been generated as a result of a simple homologous recombination between the Alu I sequences (Nicholls et al., 1987). Three other deletions within this cluster that result from homologous recombination are the $-\alpha^{3.7}$ deletion that removes 3.7 kb of DNA and encompasses the $\alpha_2$ gene (Orkin et al., 1979), the $-\alpha^{4.2}$ deletion that deletes 4.2 kb of DNA and also encompasses the $\alpha_2$ gene (Embry et al., 1979, 1980) and the $-\zeta$ deletion that removes approximately 11 kb of DNA, begins within the $\zeta$-globin gene and extends in the 3' direction (Winichagoon et al., 1984). The $-\alpha^{5.2}$ deletion spans the two $\alpha$ globin genes, deleting 5.2 kb
of DNA and appears to be the result of a nonhomologous breakage and reunion event (Pressley et al. 1980). Thus within the the α-globin gene cluster, both homologous and nonhomologous recombinational mechanisms seem to be operational in generating these deletions.

Another mammalian gene system in which deletions have been characterized is the low density lipoprotein (LDL) receptor gene. The LDL receptor gene is about 60 kb in length and contains 18 exons (Lehrman et al. 1985). A total of three deletions have been characterized at the sequence level within this gene. A patient with familial hypercholesterolemia (FH 274) has been shown to have a deletion of 5 kb within this gene. The 5' and 3' breakpoints are located within Alu I family sequences found in intron 15 and exon 18 (Lehrman et al. 1985). The deletion has been postulated to have occurred as a result of intrastrand homologous recombination between the Alu I sequences. In another case (FH 781), a 7.8 kb deletion joins intron 15 to exon 18 and is thought to have occurred as a result of homologous recombination between the Alu I sequences in intron 15 and exon 18 (Lehrman et al. 1987). A smaller deletion has also been described (FH 626), in which a 800 bp deletion removes exon 5 as a result of homologous recombination between Alu I sequences located in introns 4 and 5 (Hobbs et al. 1986). Thus, it seems
that in this gene, most of the deletions are occurring via a homologous recombinational mechanism involving the ubiquitous Alu I family sequences.

The other mammalian gene system in which deletions have been characterized is the Duchenne muscular dystrophy gene locus. Recombinational events that result in a set of deletions between sequences 100 to 300 kb apart have been reported in this locus (Dunven et al. 1987), however the deletions have not yet been characterized at the sequence level. Recently, deletions within DNA sequences at the chromosomal region 3p21 have been observed in all major types of human lung cancer (Kok et al. 1987).

Another interesting finding has been that double minute chromosomes are produced from precursors derived from chromosomal deletions (Carroll et al. 1988).

Translocations

Among the translocations, the well characterized ones include the human leukaemia associated "Philadelphia chromosome" that results from a reciprocal translocation which fuses the c-abl gene on chromosome 9 with the bcr gene on chromosome 22 (Heisterkamp et al. 1985; Shtivelman et al. 1985). Reciprocal translocations have also been detected in Burkitts' lymphoma, a
neoplastic condition involving B cells, where the translocations involve chromosome 8, and chromosomes 2, 14 or 22 (see review Croce et al. 1985). XX maleness has also been shown to be due to an abnormal interchange of the terminal parts of the short arm of the human X and Y chromosomes caused by an Alu-Alu recombination (Rouyer et al. 1987).

Gene conversions

Gene conversion, also known as nonreciprocal recombination, is quite prevalent in mammalian cells. Since gene conversion cannot be followed in mammalian systems due to the poorly characterized genetics, most gene conversions are inferred from sequences analysis of the genes themselves. Gene conversion has been implicated in the human immunoglobulin kappa light chain genes Vκ and the mouse γ2A chain genes (Bentley et al. 1983; Ollo et al. 1983), and also in the mouse histocompatibility antigens genes H-2 (Weiss et al. 1983; Pease et al. 1983). Gene conversion has also been implicated in the human fetal Gγ and Aγ globin genes in order to account for the greater homology between the two genes than between the allelic Aγ gene found on the other homologue (Slightom et al. 1980). The human
embryonic Ψα globin gene is also thought to have undergone gene conversion (Hill et al. 1985).

Thus, a large number of recombinational events have been reported within mammalian systems, some of which have been characterized at the sequence level. The mechanisms by which they occur, the sequence elements that predispose certain regions of a chromosome to frequent recombination and the other factors involved in mediating these processes are being actively pursued by a number of laboratories. The β-globin gene cluster is an ideal system to study recombination in higher eucaryotes as, firstly, most recombinational events have distinct clinically manifestations. Secondly, the extensive characterization of the gene cluster at the molecular level makes it possible to study silent variations that result in sequence polymorphisms in the nucleotide sequence as well as identify regions that are "hot spots" for recombination. The tissue specific expression of the adult globin genes in bone marrow cells has also made it the target of a number of gene transfer experiments aimed at providing gene therapy for thalassemic patients in the near future (Smithies et al. 1985).
CHAPTER III

General Materials and Methods

Human DNA isolation

Human DNA was isolated from placental tissue (obtained from the O.S.U. Hospital) as described follows. Approximately 250 g of placental tissue was cut into small pieces using a scissors and homogenized in a Waring blender at top speed for two 30 sec. periods in the presence of a known amount of T.E. buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA). NaCl and SDS were added to the homogenate to a final concentration of 1 M and 1% respectively. The lysed cells were treated with proteinase K (200 ug/ml) overnight in a water bath at 55°C. The homogenate was then treated with ribonuclease at a final concentration of 100 ug/ml for 1 hour at 55°C followed by proteinase K (100 ug/ml) for a further period of an hour. The homogenate was phenol extracted overnight with an equal volume of phenol followed by extraction with an equal volume of chloroform:isoamylalcohol mixture (24:1). The DNA was then precipitated by the addition of two volumes of ethanol in the presence of 0.3 M Sodium Acetate final concentration. The DNA pellet obtained after centrifugation at 25000xg was washed in 70% ethanol, dried in a vacuum oven and resuspended in T.E. buffer to a final concentration of 1 mg/ml.
When DNA was isolated from blood samples, approximately 30 ml of blood was treated with an equal volume of lysis buffer (3% Sarkosyl, 0.2 M EDTA, 200 μg/ml proteinase K) overnight at 55°C. The samples were then treated essentially as described above.

Plasmids

All plasmids used are derivatives of pBR 322 (Bolivar et al. 1977). The plasmid pAT 153 has been described by Twigg et al. 1980. pUC 19 has been described by Yanisch-Perron et al. 1985. All plasmids were isolated using the alkaline lysis method described by Maniatis et al. 1982.

Bacteriophage

All bacteriophage described were constructed in the laboratory of Dr. Frederick Blattner. Ch 3A Alac has been described by Blattner et al. 1977, Charon 35 by Loenen et al. 1983, and EMBL 3 was obtained from Stratagene, California.
Enzymes

Restriction enzymes were purchased from BRL. Calf intestinal alkaline phosphatase was purchased from Boehringer - Mannheim (Indiana). All other enzymes used were from BRL (Bethesda, Maryland). The enzymes were used as recommended by the manufacturers.

Bacterial Strains

Escherichia coli strains used in these experiments include C600 SF 8, LE 392, P2 392, JM 83, and MC 1061. C600 SF 8 and LE 392 have been completely described in Appendix C of Maniatis et al. 1982. P2 392 is the P2 lysogen of LE 392 obtained from Stratagene, California, and MC 1061 was obtained from Dr. Francis Collins to propagate the jumping library.

Bacteriophage in vitro packaging extracts

Lambda packaging extracts Gigapack Gold were purchased from Stratagene, California and used as described by them.
Restriction enzyme analysis

Genomic, bacteriophage, and plasmid DNAs were analyzed with various restriction enzymes and electrophoresed in agarose and/or polyacrylamide gels, stained with ethidium bromide (10 μg/ml) and photographed under UV illumination. Restriction fragment sizes were determined by measuring the distance migrated on the gel with respect to DNA fragments of known sizes and plotting that data on semi-logarithm graph paper.

Isolation of restriction fragments

The desired fragment of DNA was cut out of 5% - 6% polyacrylamide gels and incubated in a sealed 1 ml pipetman tip with 400 μl of elution solution (500 mM Ammonium acetate, 10 mM Magnesium acetate, 1 mM EDTA, 1% SDS). The eluted DNA was recovered by centrifuging the elution solution through siliconized glass wool and the DNA precipitated by addition of two volumes of 95% ethanol. The pellets that were recovered were washed with 70% ethanol and resuspended in T.E. (10 mM Tris-HCl, pH 7.5, 1 mM EDTA).
Probe synthesis

All restriction fragments used were radiolabeled with α-32P-dCTP (3000Ci/mmol, Amersham) using the random priming method of Feinberg and Vogelstein, 1983. When fragments needed to be 5' end-labeled, γ-32P-ATP (800Ci/mmol, ICN) and polynucleotide kinase were used. Probes were separated from unlabeled nucleotides either by running the mix through a G-100 Sephadex column or by fractionation on a 6% polyacrylamide gel. 3' fill-in reactions were performed essentially as described for the random labelling except the DNA was not denatured, no random primer was included in the labeling buffer, and the appropriate α-32P-dNTP added based on the restriction enzyme site being filled in.

Southern transfer and hybridization

Digested genomic DNA was electrophoresed on a 0.8% agarose gel. Gels were treated essentially as described by Southern (1975), except prior to denaturation, the gels were treated with 0.1 M HCl for 30 minutes to enhance the transfer of large fragments (Wahl et al., 1979). Hybridization of filters was done as described by Vanin et al., 1983, except when total genomic DNA was
used as a probe. When total genomic DNA was used as a probe, poly rA was omitted and only 100,000 cpm/ml of denatured probe was included.

Genomic library construction

Preparation of phage arms

Ch 3A (Δlac) phage DNA was ligated with DNA ligase to covalently circularize the cos ends. The phage DNA was then restricted with Eco RI. The restricted DNA was then dephosphorylated using calf intestinal alkaline phosphatase. EMBL 3 Bam HI arms were purchased from Stratagene, California.

Preparation of genomic DNA

DNA from the patients was digested with the appropriate enzyme and either used directly after phenol extraction and ethanol precipitation or subjected to electrophoresis through 0.8% agarose gels and size selected by visualizing the ethidium bromide stained gels. This DNA was then isolated using the phenol freeze thaw method. The procedure involves crushing the agarose slabs with an equal volume of phenol, freezing the slurry
at -70° C, then thawing the slurry at room temperature, followed by spinning the samples at 25,000 rpm for 30 minutes to pellet all the agarose. The supernatants were then extracted with chloroform - isoamyl alcohol (24:1) mixture and ethanol precipitated with two volumes of ethanol. The pellets were washed with 70% ethanol, vacuum dried and resuspended in an appropriate volume of T.E.

Ligation of phage arms to human DNA

Ligations were done at concentrations of 500 ug/ml. A ratio of 1:1 or 1:2 was used for the molecular ratio of the phage arms to human DNA.

Packaging of ligated DNA molecules

The ligated DNA was packaged into phage particles using the packaging extracts (Gigapack Gold) purchased from Stratagene, California using about 0.5 ug total ligated DNA per extract. The phage library was amplified by propagating the phage on LE 392 and plating out the library on large 25cm x 25cm plates (Gibco laboratories), then screened as described by Blattner et al. 1978.
Isolation of bacteriophage and plasmid DNA

Bacteriophage DNA was isolated from large scale growth of phage (6 litres) by the method described by Slightom et al. 1980. Small scale phage growth (100 ml) was as described by Maniatis et al. 1982. Plasmid DNA was isolated from large scale growth (1 - 6 litres) as described by Maniatis et al. 1982. Plasmid DNA was also isolated from small scale growth (10 ml) as described by Maniatis et al. 1982.

DNA sequencing

DNA sequencing was performed essentially as described by Maxam and Gilbert (1980) using the modifications of Slightom et al. (1980). The modified reactions times used are described in Appendix A of this thesis.
CHAPTER IV

Molecular characterization of a $\beta^o$-thalassemia resulting from a 1.4 Kb deletion

A. Introduction

Many well defined defects in the expression of both the human $\alpha$- and $\beta$-globin genes termed "thalassemias" have been extensively described in chapter I. In several cases, in both the $\alpha$- and $\beta$-globin gene clusters, rearrangement events have produced deletions which have been extensively studied at the sequence level (Vanin et al. 1983; Henthorn et al. 1986; Mager et al. 1985; Nicholls et al. 1985; Spritz et al. 1982; Popovich et al. 1986; Jennings et al. 1985; Nicholls et al. 1987; Gilman et al. 1987; Kulozik et al. 1988; Shiokawa et al. 1988). Nine of these deletions resulted in the removal of large segments (>10kb) of DNA, while six were more localized deletions(<10kb). Twelve of these deletions were shown to be the result of non-homologous breakage and reunion events. In eight of these deletions, four of which were large, the recombinational event was shown to be "clean", that is no spurious bases were added during the recombinational event, while four were shown to contain extra bases at the deletion joint as discussed in chapter II.
Described in this chapter is the characterization and sequence analysis of a third small deletion. Two of the previously characterized small deletions were shown to remove part of the β-globin gene. A deletion which begins in IVS-2 of the β-globin gene and extends in the 3' direction for 619 bp has been reported in an Asian-Indian individual (Spritz et al. 1982). More recently a Canadian individual of Czechoslovakian descent with β°-thalassemia has been studied and was found to have a deletion which also begins in IVS-2 of the β-globin gene but in this case the deletion extends for 4.2 kb in the 5' direction (Popovich et al. 1986). Both deletions were the result of non-homologous breakage and reunion events. The deletion characterized in this chapter begins 484 bp 5' to the β-globin gene transcriptional start site and extends for 1393 bp in the 3' direction into the middle of the β-globin gene IVS-2. As with the other two small deletions, one of the breakpoints for this deletion is also found within IVS-2 of the β-globin gene. Comparison of the parental sequences involved in the deletion indicate that this deletion was also the result of a "clean" non-homologous breakage and reunion event.

Throughout this chapter the exact points at which the breakpoints occur have been referred to as the 5' and 3' breakpoints, with the 5' and 3' being assigned with respect to the β-globin gene. The terms "normal 5' and normal 3' DNAs"
have been used to refer to the DNA sequences, on the normal chromosome, which encompass the 5' and 3' breakpoints. When describing the position of breakpoints or restriction enzyme sites, they have been numbered with respect to the transcriptional start of the β-globin gene, which has been considered as nucleotide 1. The numbering used for the normal β-globin gene has been negative numbers for positions 5' to the transcriptional start site and positive numbers for positions within the β-globin gene.
B. Materials and Methods

Genomic southern blotting and hybridizations

Genomic southern analysis was performed as previously described (Vanin et al. 1983). The probes used in this study were (a) the Bam HI-Eco RI fragment containing the large intervening sequence (IVS-2) of the β-globin gene, (b) the 1.8 kb Bam HI fragment which contains exons 1 and 2, IVS-1 and some 5' flanking DNA of the β-globin gene and (c) a β-cDNA probe (JW 102) (Wilson et al. 1978). These fragments were radiolabeled using the procedure of Feinberg and Vogelstein (1983).

Construction and screening of the genomic library

A recombinant bacteriophage lambda library was constructed by ligating Eco RI digested genomic DNA, from an individual heterozygous for the β-globin deletion, to Charon 3A (de Wet et al. 1980) which had been digested with Eco RI and subsequently dephosphorylated. The ratio of genomic DNA to phage DNA was 1:1 on a microgram basis. The ligated mixture was then packaged into phage particles using the in vitro packaging extracts. A library containing a total of $1 \times 10^4$ recombinant phage was obtained. This was screened using the 1.8 kb Bam HI probe (Blattner et al. 1978).
DNA sequencing

DNA sequencing was performed essentially as described by Maxam and Gilbert 1980 using the modification of Slightom et al. (1980). All sequencing was done using 5' end labeled fragments.
C. Results and Discussion

Genomic DNA analysis

DNA from an American Black, heterozygous for δ\textsuperscript{0}-thalassemia, was digested with a number of different restriction enzymes and the digests analyzed by hybridization using three different DNA probes. Figure 5 shows the results obtained when Bgl II, Eco RI and Hpa I genomic digests were probed with IVS-2 of the β-globin gene. Figure 6 shows the results when Bam HI and Ava II digests were probed with the 1.8 kb Bam HI and cDNA probes respectively. These results as well as additional data obtained are summarized in Table 1.

The simplest interpretation of these data is that the deletion is between 1.1 and 1.5 kb in size, with the deletion beginning within the β-globin gene IVS-2 and extending in the 5' direction. From the results of the Eco RI and Sph I digests we concluded that the 5' breakpoint of this deletion must lie 3' to the Sph I site found at -614 while the 3' breakpoint must lie 5' to the Eco RI site at +1393. The results obtained in the Ava II and Bam HI genomic digests, using both the 1.8 kb Bam HI probe and the cDNA as probes, were consistent with this interpretation. The 8.8 kb fragment seen in the Bam HI digest with both probes must have resulted from the removal of approximately 1.3 kb of
Figure 5: Autoradiogram of DNA digests hybridized with the β IVS-2 probe.

A through C are Eco RI, Bgl II, and Hpa I digests, respectively, of the different DNAs. Lanes 1 and 3 are the restriction digests of DNA from two different normal individuals; lane 2 is the restriction digest of DNA from the affected individual. The bands from the affected individual have been indicated by arrow heads since they were of very light intensity.
FIGURE 5
Figure 6: Autoradiograms of DNA digests hybridized with the 1.8 kb Bam HI and cDNA probes.

The left panel shows the autoradiogram of DNA from the affected individual, digested with Bam HI, and hybridized with the 1.8 kb Bam HI probe. The bands have been highlighted with arrow heads in both panels. The bands marked 6 are due to the cross hybridization of the β-globin probes to the δ globin gene, due to the homology they share. The right panel shows the autoradiogram of the Ava II digests that were hybridized to the cDNA probe. Lanes b and c are of normal individuals, except that the DNA of the individual in lane c is polymorphic for an Ava II site within the β-globin gene. Lane a contains the DNA from the affected individual. The numbers on the side indicate the size of the bands in kb.
Table 1

Results of genomic southern analysis

<table>
<thead>
<tr>
<th>Probe</th>
<th>Enzyme</th>
<th>DNA Fragment Size, kb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Abnormal</td>
</tr>
<tr>
<td>IVS-2</td>
<td>Ava II</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Bgl II</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Eco RI</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>Hpa I</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>Sph I</td>
<td>2.5</td>
</tr>
<tr>
<td>cDNA</td>
<td>Ava II</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Bam HI</td>
<td>8.8</td>
</tr>
<tr>
<td>1.8 kb Bam HI</td>
<td>Bam HI</td>
<td>8.8</td>
</tr>
</tbody>
</table>

The asterisk next to the size of the band from the normal chromosome of this individual is to indicate that it results from the lack of the polymorphic Hpa I site (Kan et al., 1978) 3' to the β-globin gene.
DNA which includes the Bam HI site at +477. Similarly the results of the Ava II genomic digests using both the IVS-2 and cDNA probes were consistent with the fact that the deletion is approximately 1.4 kb in size, with the Ava II sites at +293 and +507 being deleted. The results obtained when the Hpa I digest was probed with the IVS-2 probe can be explained by the absence of the polymorphic Hpa I site (Kan et al. 1978) on the normal chromosome and its presence on the abnormal chromosome. Padanilam et al. 1984 have reported very similar results from a different American Black family.

Isolation and characterization of clones from both chromosomes

In order to isolate clones from both the normal and abnormal chromosomes we decided to make a complete Eco RI library of this individual in Charon 3A. The library was then screened using the 1.8 kb Bam HI probe as we predicted that this probe would be able to identify the abnormal clone from the library. From a total of $10^4$ recombinant clones, two phage were isolated which consistently hybridized to the probe. One was shown to contain a 5.5 kb insert corresponding to the Eco RI fragment from the normal chromosome while the other contained a 4.2 kb Eco RI insert which corresponds to the fragment from the abnormal chromosome (Figure 7). Mapping of the the 4.2 kb Eco RI insert with Sph I, together with Southern analysis using the β-globin
Figure 7: Autoradiogram of restriction digests of the DNA from the two phage clones isolated from the library, hybridized with the 1.8 Bam HI probe.

Panel A contains the normal clone, and panel B contains the abnormal clone. Lanes 1, 2, and 3 are Eco RI, Bam HI - Eco RI, and Bam HI digests of the phage clones. The numbers on the side indicate the size of the fragments in kb.
IVS-2 fragment as a probe, confirmed the results obtained from the genomic Southern analysis; that is, the Sph I site found 5' to the β-globin gene is still present on the abnormal chromosome. Furthermore it was found that the Hpa I site at -815 and the Sph I site at -614 were approximately 0.8 kb and 0.6 kb, respectively, from the Eco RI site at 1393 instead of 2.2 kb and 2.0 kb, respectively, as was seen in the fragment from the normal chromosome (Figure 8). Therefore the deletion was thought to be closer to 1.4 kb. Further mapping of the 4.2 kb Eco RI fragment indicated that the Dra I site at +997 was present in the fragment from the abnormal chromosome. Figure 9 shows a map of the abnormal and normal chromosomes.

As this entire region has been sequenced (Poncz et al. 1983) we were able to develop a fine restriction map of the 4.2 kb Eco RI fragment once it was subcloned into pUC 19. This allowed us to better localize the 5' and 3' breakpoints as well as to provide information concerning the status of the polymorphic sites (Hinf I, Rsa I and HgiA I) on both the normal and abnormal chromosomes. Mapping of the 4.2 kb Eco RI fragment with Xho II and Rsa I enabled us to localize the 5' breakpoint to the -553 and -406 region. In fact the Rsa I site at -553 corresponds to the polymorphic site previously reported by Semenza et al. 1984. The Rsa I digest also allowed us to localize the 3' breakpoint to the region between +822 and +954. Aside from Rsa I the 5.5 kb and 4.2 kb Eco RI fragments from the normal and abnormal
Figure 8: Autoradiogram of restriction digests of DNA from the two clones isolated from the library, hybridized with 1.8 Bam HI probe.

Lane 1 contains DNA from the normal clone and lane 2 contains DNA from the abnormal clone. The restriction enzyme used are E: Eco RI; H: Hpa I; and S: Sph I. The numbers on the side indicate the size of the fragments in kb.
FIGURE 8
Figure 9: Restriction enzyme map of the normal and abnormal chromosomes.

Restriction enzyme map of the 5.5 kb Eco RI fragment and the surrounding region from the normal chromosome as well as the corresponding region on the affected chromosome are shown. The restriction enzymes used are A: Ava II; B: Bam HI; D: Dra I; E: Eco RI; H: Hpa I; and S: Sph I. The solid boxes indicate the exons and the open boxes indicate the introns. The jagged line on the affected chromosome indicates the location of the breakpoint junction. Corresponding restriction enzyme sites on the two chromosomes have been connected with broken lines.
chromosomes, respectively, were also digested with Hinf I and HgiA I in order to determine the presence or absence of these polymorphic sites (Moschonas et al. 1982; Orkin et al. 1982). These digests (data not shown) indicated that the normal chromosome was HgiA I (+):Hinf I (-):Rsa I (-) while the abnormal chromosome was HgiA I (0):Hinf I (+):Rsa I (+); the HgiA I site on the abnormal chromosome is removed as a result of the deletion.

Sequence analysis of the deletion breakpoint

As both the Dra I and Sph I sites were still present following the deletion event (Figure 9) we sequenced this fragment from the Dra I site. Figure 10 shows the results of the Maxam and Gilbert sequencing reactions performed on the Dra I-Sph I fragment. The comparison of the sequence spanning the breakpoint with the normal 5' and 3' sequences is shown in Figure 11. The comparison of these three sequences shows that the recombination was "clean", that is no extra bases were added during the recombinational event and that the deletion results in the removal of 1393 bp. The deletion begins at nucleotides -484 or -485 and extends to nucleotides +908 or +909. The ambiguity is due to the fact that the breakpoint can be at two possible positions (Figure 11). Comparison of the normal 5' and 3' DNAs indicated that the two sequences shared no
Figure 10: Autoradiogram of the sequencing gel containing the sequence across the breakpoint junction.

Shown is a 8% acrylamide sequencing gel containing the Maxam and Gilbert reaction used to sequence a Dra I - Sph I fragment containing the breakpoint junction, from the Dra I site. The arrows indicate the location of the breakpoint junction, and an asterisk has been placed on either side of the A residue that indicates the crossover point.
FIGURE 10

AGGTATTCCCTTAGTAATAATGTAATA

GGCG

A

CC

T

A

T
Figure 11: Sequence across the deletion breakpoint junction of the 1.4 kb $\beta^o$-thalassemia and the corresponding normal 5' and 3' DNAs.

The three aligned sequences are the normal 5' sequence (top), the deletion joint sequence (middle), and the normal 3' sequence (bottom). An interrupted vertical line between sequences indicates positions at which the sequences are identical. Asterisks between the normal 5' and deletion joint sequences indicates the position at which a T to C change produced the polymorphic Rsa I site. Numbers above the normal 5' sequence and below the normal 3' sequence are the distance of those particular nucleotides from the adult $\beta$-globin transcriptional start site.
**FIGURE 11**

<table>
<thead>
<tr>
<th></th>
<th>NORM. 5'</th>
<th>DEL. JOINT</th>
<th>NORM. 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>-550</td>
<td>ATGCAATATATGTATATGTATATGTATATATATACACATATATATA</td>
<td>ATGCAATATATGTATATGTATATGTATATGTATATATACACATATATATA</td>
<td>GCTGCTGATACACTACATACATGGAAATATGGGCTTATTATACTATA</td>
</tr>
<tr>
<td>850</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-500</td>
<td>TTTTTTTTTTTTTTTTACCAGAAGTTTTATCCTAAATAAGGAGAAGATATGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>900</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-450</td>
<td>TTAGAAGTGAGTTAGAGTTTTGATTTACATCTGCTGGTGAAGAATTGGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>950</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
appreciable homology above that expected at random. Random scattering of one of the sequences did not alter the level of homology compared to that obtained for the unscrambled comparison. As can also be seen from Figure 11 the normal 5' and deletion joint sequences are not completely homologous in the region 5' to the breakpoint, as would be expected. One of the differences occurs within the alternating (AT)_nT_n stretch found 5' to the breakpoint and is an A to T change. The other difference observed is a T to C change thereby producing the Rsa I polymorphic site (see previous section).

This chapter describes the characterization of a β°-thalassemia which results from the deletion of approximately 1.4 kb. Phenotypically the individual is unusual in that for a β°-thalassemia heterozygote he has unusually high HbA_2(9.6%) and HbF(7.9%) levels. Usually such heterozygotes have levels of 5-6% for HbA_2 and 1% for HbF. Similar levels were found in the individuals reported by Padanilam et al. 1984. Elevated levels for both HbA_2 and HbF have also been reported in two other studies (Popovich et al. 1986; Gilman et al. 1984). HbA_2 and HbF levels of 8.3% and 5.7%, respectively, were reported in the Canadian family which was found to have a 4.2 kb deletion which started in IVS-2 of the β-globin gene and extended in the 5' direction (Popovich et al. 1986). Above average levels of HbA_2 and HbF were also seen in a number of individuals of Dutch
origin with a deletion which starts 2.4 kb 5' to the β-globin
gene and extends in the 3' direction for 10 kb, thereby deleting
the entire β-globin gene (Gilman et al. 1984). Heterozygous
individuals for this deletion were analyzed and the HbA2 levels
were found to be 6.1 ± 1.5% while the HbF levels varied between 4
and 10%. On the other hand, no such elevated levels of HbA2 and
HbF were found in heterozygous individuals for the 619 bp
deletion β0-thalassemia, which begins in IVS-2 and extends in the
3' direction. Therefore, the elevated levels seem to result
from the removal of the 5' half of the β-globin gene and of the
DNA immediately 5' to the gene. The molecular mechanism(s)
responsible for this elevation of HbA2 and HbF levels is not
known. One possibility is that removal of this DNA causes the
chromatin structure in the remaining β-globin cluster to change,
thereby activating the γ and δ-globin genes. It should be
noted that all three deletions discussed above remove the DNase I
hypersensitive site found immediately 5' to the β-globin gene
(Tuan et al. 1985). Altering the chromatin conformation has
previously been shown to alter the level of expression of the β-
globin gene. Kioussis et al. 1983 found that the β-globin gene
remaining on the affected chromosome of a γδβ-thalassemia was
inactive because the DNA juxtaposed to the gene, as a result of
the deletion event, was conferring an inactive chromatin state on
the gene.
Analysis of this deletion at the sequence level indicated that the deletion begins at -484 or -485, depending upon the positioning of the breakpoint, and extends for exactly 1,393 bp into the middle of IVS-2. Comparison of the normal 5' and 3' sequences indicated that the deletion was the result of a non-homologous breakage and reunion event; that is no appreciable homology was seen between the two sequences. A number of different deletions involving the human β-globin gene cluster have been studied at the sequence level as discussed previously. In all the cases studied to date, excluding Hbs Lepore and Kenya (Baglioni et al. 1962; Flavell et al. 1978; Ojwang et al. 1983), the deletion events have been shown to be the result of non-homologous breakage and reunion events. Some of these deletions were shown to be "clean" while in others extra nucleotides seem to have been added during the recombinational event as discussed in the general introduction. Comparison of the normal 5' DNA with the deletion joint DNA indicates that the two sequences are not completely identical as would have been expected. The first of these differences is due to the fact that the abnormal chromosome is Rsa I+ while the normal 5' sequence we used was from an Rsa I− chromosome. The deletion joint sequence is also different from that reported for the Rsa I+ chromosome (Semenza et al. 1984); the Rsa I+ chromosome sequence contains an extra AT dinucleotide and there is also an A to T change when compared to the deletion joint sequence. Interestingly this region of the
β-globin gene cluster has been identified as a "hot spot" for unequal homologous crossover. In fact Semenza et al. 1984 postulated that the alternating purine:pyrimidine stretch, of which the A:T stretch is a part, may be promoting recombination and has been extensively dealt with in chapter VI.

Aside from the fact that many of these deletions are the result of non-homologous breakage and reunion events, we know very little about the molecular mechanism(s) by which these deletions are produced. Vanin et al. 1983 proposed a model to explain the constant size deletion phenomenon they observed in studying four large deletions involving the human β-globin gene cluster, two of which were shown to be the result of "clean" non-homologous breakage and reunion events as described in chapter II. They suggested that these large deletions were the result of the removal of complete loops of DNA. DNA loops had been shown to be anchored to the nuclear matrix by a number of laboratories (Paulson et al. 1977; Marsden et al. 1979; Pardoll et al. 1980; Vogelstein et al. 1980). Furthermore, Vanin et al. suggested that the variation in the actual breakpoint position could be due to either the breakages occurring while DNA is being moved through the matrix anchorage points, perhaps during replication, or that the anchorage points are not sequence specific and may differ between individuals.

The analysis of this small deletion appears to indicate that these small deletions are produced by the same mechanism; that
is, they also occur during replication and at the anchorage points to the nuclear matrix. The proposed model for the production of these deletions is shown in Figure 12. As can be seen we propose that the anchorage point lies between the 3' breakpoint of the 4.3 kb deletion β°-thalassemia and the 5' breakpoint for the 619 bp Indian β°-thalassemia. Therefore, in each case, one of the breakpoints is defined by the region immediately adjacent to the attachment site. This is illustrated in more detail in Figure 13. The distance between the breakpoints represents the amount of DNA covered by the attachment. Then, as replication proceeds DNA is moved through the attachment site and duplicated. The 619 bp Indian β-thalassemia occurs by recombination when the newly replicated loop is 619 bp while the 4.3 kb and 1.4 kb deletion β-thalassemias are the result of recombination when the newly replicated loop is 4.3 kb and 1.4 kb in size, respectively. Although the normal 5' and 3' DNAs, involved in each deletion, are linearly quite a distance apart they are in close proximity as a result of replication.

As was stated above, one of the breakpoints is defined by the position of the attachment site. The question still remains as to why the break occurs at the other position in these three deletions. An interesting possibility is that these breaks occur
Figure 12: Proposed model for the production of the small deletions found within the human β-globin gene.

Letters A, B, C, etc. represent sites on the DNA prior to replication, and the letters A', B', C' represent sites on the DNA after one round of replication. DNA that is anchored to the matrix attachment site is represented by a broken line whilst all the rest of the DNA is represented by a solid line. Small closed circles represent DNA polymerase α molecules and the large open circles the anchorage point. The asterisk above the newly replicated loop indicates the loop is deleted to give rise to the deletion shown on the right. Square brackets indicate the extent of the deletions. The letter E corresponds to the matrix attachment site and the deletions flank this region.
Figure 13: Schematic diagram of the β-globin gene and the location of a matrix attachment site as defined by one of the breakpoints of the three deletions.

Shown is a model predicting the exact location of a matrix attachment site within the second intervening sequence of the β-globin gene. E1, E2, and E3 represent the three exons of the β-globin gene. The three deletions shown have been described in the text. The region marked MAR represents the core of the matrix attachment site with indefinite boundaries.
**β-GLOBIN GENE**

![Diagram of β-GLOBIN GENE with labeled markers and distances.](image)

- **E1** to **E2**: 1.4 kb β⁰-Thal
- **E2** to **E3**: 4.3 kb β⁰-Thal
- **E1** to **E3**: 0.6 kb β⁰-Thal

**FIGURE 13**
at sequences which cause DNA polymerase α to pause and therefore are more likely to break. Weaver and DePamphilis 1982 studied pause sites for DNA polymerase α using φX174 as a template. They found that the sequence (T\A)GGAG could act as an arrest site for DNA polymerase α. The sequence at the appropriate breakpoints for each of the small deletions is shown below.

<table>
<thead>
<tr>
<th>Breakpoint</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' breakpoint 1.39kb β₀-thalassemia</td>
<td>ATAAGGAGAA</td>
</tr>
<tr>
<td>5' breakpoint 4.2kb β₀-thalassemia</td>
<td>CAGTGGAGGA</td>
</tr>
<tr>
<td>3' breakpoint 619bp β₀-thalassemia</td>
<td>AAAAGGATTC</td>
</tr>
</tbody>
</table>

As can be seen all three deletions have a sequence very similar to the consensus pause sequence at the breakpoint. Some, but not all, of the larger deletions were also found to contain this sequence at or near the breakpoint (Vanin et al. 1983; Mager et al. 1985; Jennings et al. 1985; Nicholls et al. 1987). One possible explanation is that DNA polymerase α may arrest at a number of different sequences, of which (T\A)GGAG is one. Alternatively more than one mechanism may be responsible for producing these different deletions.

The results presented in this chapter together with those of other small deletions within the human β-globin gene cluster begin to indicate possible mechanism(s) responsible for their production.
Localization of a nuclear matrix attachment region within the β-globin gene

A. Introduction

DNA is organized at many different levels within the nucleus of a eucaryotic cell. DNA is first wrapped around the histone octamers to form the nucleosomal unit consisting of 1.75 to 2 turns of DNA per nucleosome. The nucleosomes are arranged in a helical manner to form what has been termed the 30 nm chromatin fibre as it measures 30 nm along the axis of the spiral. The chromatin fibre in turn appears to be organized into DNA loops or domains which are anchored to a proteinaceous networking called the nuclear matrix in metaphase chromosomes (Paulson et al. 1977) or the nuclear scaffold in interphase chromosomes (Berezney et al. 1974; Benyati et al. 1987; Cook et al. 1980; Mirkovitz et al. 1984). High resolution microscopic autoradiographic studies indicate that the sites of DNA replication are preferentially localized in close apposition to the nuclear matrix (Fakan et al. 1971; Fakan et al. 1974). In addition, autoradiographic and biochemical experiments wherein cells were labeled with 3H-thymidine for various periods of time showed that newly
replicated DNA is almost quantitively associated with the nuclear matrix (Vogelstein et al. 1982). Thus these loops are thought to correspond to replication units or active/inactive domains of chromatin (Gasser et al. 1987; Jackson et al. 1986; Nelson et al. 1986) and the attachment site itself is thought to correspond to the site of DNA replication (Jackson et al. 1986; Smith et al. 1984). Recently chromosomal autonomously replicating sequences (ARS) and chromosomal centromere CEN III in yeast have been shown to be associated with the nuclear scaffolding (Amati et al. 1988).

The matrix itself has not been fully characterized, but has been shown to be composed of at least two major proteins, Sc I (170 Kd) and Sc II (135 Kd). Antibodies against topoisomerase II have been shown to immunoprecipitate Sc I and hence this protein is thought to correspond to topoisomerase II (Berrios et al. 1985; Earnshaw et al. 1985). At this point it is not clear whether these proteins are actually involved in anchoring the DNA loops or are just strongly associated with the other protein components of the nuclear matrix. Based on several lines of evidence, another 260 kD protein called chromomatricin has been recently proposed as the actual DNA loop anchoring protein (Ueda et al. 1989). Ueda et al. show that residual DNA on nuclear matrices co-localize with chromomatricin and DNA binding on nitrocellulose filters after Western blotting indicate that chromomatricin is a major DNA binding protein.
Recently, Mirkovich et al. 1984 have described the use of Lithium 3,5-diiodosalicylate, a low ionic strength buffer that is capable of removing the histones while leaving the residual protein framework intact, to isolate DNA that is bound to the nuclear matrix. They have shown that the non-transcribed spacer between the Drosophila H1 and H3 histone genes is enriched in the matrix bound fraction and were able to narrow down the region to a 657 bp region located approximately 360 bp 5' to the H1 histone gene. They also localized matrix attachment sites in the non-transcribed spacer DNA of the hsp 70 heat shock gene family. Cockerill and Garrard (1986) have also used this procedure to study the matrix attachment sites within the constant region of the mouse K immunoglobulin locus. In addition, they have developed an *in vitro* assay for determining the capability of a particular DNA fragment to bind to the nuclear matrix. Using the *in vitro* procedure, they have shown that a 365 bp fragment immediately 5' to the K enhancer is capable of binding the matrix. Furthermore, using the *in vivo* procedure, they show that this same region is bound to the nuclear matrix. The sequence that endows a particular fragment with the capability to bind the nuclear matrix also seems to be evolutionarily conserved in that the MAR between the H1 and H3 histone genes of Drosophila is capable of binding the matrices prepared from mouse cells, *in vitro*. Similarly, Cockerill and Garrard (1986) have also shown that the MAR from the mouse immunoglobulin gene locus is capable
of binding in vitro to the matrices prepared from yeast nuclei. Cockerill et al. (1987) have also identified a MAR surrounding the enhancer region of the mouse heavy chain immunoglobulin gene cluster. From fine mapping studies they have shown that a 350 bp A + T rich segment mediates the attachment to the nuclear matrices. The location of MARs surrounding the chicken lysozyme gene has also been studied. Phi-Van and Stratling (1988) have found two MARs surrounding the lysozyme gene, one between 8.5 and 11.1 kb 5' to the gene and the other between 1.3 to 5.0 kb 3' to the gene. These regions were shown to be both in vitro and in vivo MARs. Interestingly, these regions have also been shown to be the boundaries of the chromatin loop which contains the chicken lysozyme gene.

The replication model for the deletions proposed in chapter III predicts that the second intervening sequence (βIVS-2) contained within a 0.9 kb Bam HI - Eco RI fragment was possibly an origin of replication. Given the compelling evidence that origins of replication may be associated with the nuclear matrix, it was also predicted that this region might be attached to the nuclear matrix. We decided to test this aspect of the model by isolating matrix associated DNA using the in vivo procedure first described by Mirkovitch et al. (1984) and also by using the in vitro procedure described by Cockerill et al. (1986).
B. Materials and Methods

Isolation of cell nuclei

Human placenta obtained from O.S.U Hospital, was washed in isolation buffer containing a final concentration of 3.75 mM Tris-HCl, pH 7.5, 0.05 mM spermine, 0.125 mM spermidine, 0.5 mM EDTA/KOH, pH 7.5, 1% (v/v) thiodyglycol, and 20 mM KCl. The placenta was then cut into small pieces using a scissors, and homogenized in a Waring blender at top speed for two 30 second periods. During the homogenization, isolation buffer containing 0.1% digitonin was added. The nuclei were then collected by repeated washing and centrifugation at 1000xg for 10 min periods. The final nuclear pellet was resuspended in isolation buffer lacking EDTA, but containing 0.1% digitonin. Nuclei that were not used immediately were stored at -20°C in isolation buffer containing 50% (v/v) glycerol. All procedures were carried out on ice or at 4°C and protease inhibitors like Trasylol (10 KIU/ml) and phenylmethylsulfonylfluoride (0.1 mM) were used in the isolation buffer.
Isolation of histone depleted nuclei

The nuclei were resuspended in isolation buffer with 0.1% digitonin and without EDTA such that the absorbance at 260 nm was approximately 3.0 at a ten-fold dilution. 500 ul of this suspension of nuclei was equilibrated at 37°C for 20 min. and 8 ml of low salt extraction buffer was added slowly at room temperature. The low salt extraction buffer contained 5 mM HEPES/HCl pH 7.5, 0.125 mM spermidine, 2 mM EDTA/KOH, pH 7.5, 2 mM KCl, 0.1% digitonin, 25 mM 3,5-diiodosalicylic acid, lithium salt (Sigma). After 5 min at room temperature, the histone depleted nuclei were recovered by centrifugation at 2500 x g for 20 min at room temperature. The pellet was washed four times with 8 ml of digestion buffer by resuspending the pellet gently in the buffer with a pasteur pipet that had the tip broken off. The digestion buffer used was prepared by diluting and using the 10 X buffer supplied for the enzymes Bam HI and Eco RI (buffer 3) by BRL.

Restriction enzymes were then added to the pellet resuspended in 1 ml of buffer at concentrations of 2000 U/ml. The digestions for each enzyme were allowed to proceed for 3 hrs at 37°C in a shaking water bath. Solubilized DNA was then separated from the matrix attached fraction by centrifugation of the digested histone depleted nuclei at 2500 X g for 10 min at 4°C.
DNA analysis

DNA that was recovered in the supernatant and precipitate fractions was treated with proteinase K in the presence of 1% SDS at concentrations of 200 ug/ml of the enzyme at 50°C. The fractions were then phenol extracted and ethanol precipitated. The pellets were resuspended in 30 ul of Tris-EDTA buffer. Three ug of DNA from each fraction were redigested with the respective enzymes prior to electrophoresis on a 0.8% agarose gel and Southern transfer onto nitrocellulose paper. The hybridizations were performed as described in the general materials and methods section.
C. Results and Discussion

The deletion model described in chapter III predicts that the 3' breakpoints of the 4.3 kb Czechoslovakian and 1.4 kb Black deletion, and the 5' breakpoint of the 0.6 kb Indian deletion border a nuclear matrix attachment site and possibly an origin of replication as shown in Figure 13. This region is within the second intervening sequence of the β-globin gene. To test this possibility we decided to determine whether this fragment (1) could bind nuclear matrices in vitro and (2) is bound to the nuclear matrix in vivo.

The in vitro and in vivo procedures used are shown schematically in Figure 14. DNA that was isolated using the in vivo procedure was designated either precipitate (P) if associated with the nuclear matrix or supernatant (S) if not associated, in keeping with the terminology already in use. The in vitro assays were performed by Piri Welcsh. For the in vivo studies we chose to restrict the histone depleted nuclei with Bam HI and Eco RI as these restriction enzymes (1) bordered the proposed MAR and (2) would give a small restriction fragment.

The results of the in vitro experiments are shown in Figure 15 and summarized in Table 2. The in vitro assays were performed by incubating nuclear matrices with end-labeled fragments in the presence of various concentrations of competitor DNA (sheared E.
Figure 14: Schematic representation of the two procedures used to identify DNA attached to the nuclear matrix (from Cockerill et al. 1986).
NUCLEI

"MATRIX"

Extract Histones and DNA

Blind $^{32}$P-labeled Gene Fragments, Centrifugation

Purify DNA, Gel Electrophoresis, Autoradiography

In Vitro MAR

"HALO"

Restriction Enzyme Digestion, Centrifugation

Purify DNA, Gel Electrophoresis, Southern Analysis

In Vivo MAR

FIGURE 14
Figure 15: Autoradiogram of the in vitro nuclear matrix assays performed on the 1.8 Bam HI and the 0.9 Bam HI - Eco RI fragments (kindly provided by Piri Welcsh).

Lanes A through H contain increasing amounts of competitor E. coli DNA, starting with no DNA in lane A and increasing in increments of 10 ug. The left panel contains the P32 labeled 1.8 kb Bam HI fragment and the right panel contains the P32 labeled 0.9 kb Bam HI - Eco RI fragment.
<table>
<thead>
<tr>
<th>Lane</th>
<th>Area under peak</th>
<th>Normalized Area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.8 B</td>
<td>0.9 B-E</td>
</tr>
<tr>
<td>A.</td>
<td>0.687</td>
<td>1.179</td>
</tr>
<tr>
<td>B.</td>
<td>0.733</td>
<td>1.139</td>
</tr>
<tr>
<td>C.</td>
<td>6.423</td>
<td>5.836</td>
</tr>
<tr>
<td>D.</td>
<td>1.267</td>
<td>9.735</td>
</tr>
<tr>
<td>E.</td>
<td>5.45</td>
<td>15.700</td>
</tr>
<tr>
<td>F.</td>
<td>7.70</td>
<td>28.571</td>
</tr>
<tr>
<td>G.</td>
<td>6.33</td>
<td>58.333</td>
</tr>
</tbody>
</table>

Table 2

Results of densitometric scans of the in vitro MAR assay autoradiograms shown in Figure 15
coli DNA). Analysis of the *in vitro* assay results was performed by normalizing the binding of the test fragment to that of the plasmid vector which was given a value of 1. The results of such an analysis shows that both the 0.9 kb Bam HI - Eco RI fragment and the 1.8 kb Bam HI fragment immediately 5' to it had stronger affinity for the matrices than did the plasmid vector. However, the 0.9 kb Bam HI - Eco RI fragment seems to have a stronger affinity for the matrices judging from the fact that the normalized ratio for the 0.9 kb Bam HI - Eco RI fragment continues to increase whilst that of the 1.8 kb Bam HI fragment decreases at higher concentrations of competitor DNA concentrations. These results indicated that both the 0.9 kb Bam HI - Eco RI fragment as well as the 1.8 kb Bam HI fragments were capable of binding to the nuclear matrices *in vitro*.

In order to determine whether these fragments were actually bound to the nuclear matrix *in vivo*, histone depleted nuclei were treated with Eco RI and the P and S fractions were isolated. Following electrophoresis of equal amounts of the P and S fraction DNAs (*Shown on Figure 16*) and transfer of the DNAs onto nitrocellulose, the blots were hybridized to the 0.9 kb Bam HI - Eco RI probe, which was expected to hybridize to a 5.5 kb Eco RI fragment. We found that the probe hybridized to the 5.5 kb fragment in both the S and P lanes (*data not shown*). This result indicated that there was a MAR somewhere within the 5.5 kb Eco RI fragment. To further sublocalize this MAR, the experiment
Figure 16: Photograph of a 0.8% agarose gel containing equal amounts of DNA from the supernatant and precipitate fractions obtained using the in vivo MAR assay.

M indicates standard lambda size markers, P the precipitate fraction and S the supernatant fraction. The numbers on the side indicate sizes of the fractionated DNA in kb.
FIGURE 16
was repeated except that the histone depleted nuclei were restricted with Bam HI and Eco RI. As before, DNAs from both the S and P fractions were electrophoresed, transferred, and probed with the 0.9 kb Bam HI - Eco RI fragment. The probe hybridized to a 0.9 kb Bam HI - Eco RI fragment in both the S and P lanes. Densitometric scanning of the bands indicated that of the total hybridization, 40% was in the S fraction whilst 60% was in the P fraction.

We believe this distribution is indicative of the attachment of this fragment to the nuclear matrix and is very similar to the typical distribution patterns obtained in similar experiments done by other groups. The same DNA blot was also probed with the 1.8 kb Bam HI fragment. The results which are shown in Figure 17 indicate that this fragment is also bound to the nuclear matrix. Initially, this result seemed rather puzzling as we had digested the matrix bound DNA with Bam HI and Eco RI and we would have expected to see enrichment of a single fragment. A plausible explanation for these results is that the matrix attachment site starts upstream of the Bam HI site located in the second exon of the β-globin gene, and extends into the second intervening sequence. This interpretation is consistent with the results obtained from the in vitro MAR assays in which we found that both the 0.9 kb Bam HI - Eco RI and 1.8 kb Eco RI fragments were able to bind to the nuclear matrices though with differing affinities.
Figure 17: Autoradiograms of the DNAs in the supernatant and precipitate fractions hybridized to the 0.9 Bam HI - Eco RI fragment (βIVS 2) and the 1.8 Bam HI fragment.

The DNAs in both lanes were digested with Bam HI - Eco RI prior to fractionation. The numbers on the side indicate the size of the fragments hybridizing to the various probes in kb.
FIGURE 17
Using the same DNA blots, we also tested the distribution of a 4.0 kb Eco RI fragment, which was located approximately 95 kb 5' to the β-globin gene and contained the 5' breakpoint for γδβ thal 2. This fragment had been shown not to bind the nuclear matrices in vitro (data not shown). The results of the hybridization with the 4.0 kb Eco RI fragment are shown in Figure 18. As can be seen, the probe hybridized to a 4.0 kb Eco RI fragment as expected and almost 90% of the fragment segregated into the supernatant fraction. Thus these results were both consistent with the in vitro results and served as a control for determining the segregation pattern expected of a fragment that is not an in vivo MAR. We also tested a 3.3 kb Eco RI fragment located 60 kb 5' to the β-globin gene. This fragment is located within a region known to contain DNase I hypersensitive sites and was possibly in close proximity to another matrix attachment site within the β-globin gene cluster. The results obtained indicate a 90% distribution of the probe in the supernatant fraction, once again acting as a control for the in vivo procedure used to identify matrix attachment regions (Figure 18). Figure 19 shows the location of the MAR identified within the β-globin gene. Similar results to those discussed above were obtained by Higgs et al. (1988), although, they were only able to show that the entire 5.5 kb Eco RI fragment containing part of the β-globin gene and the 5' region of the β-globin gene, was an in vivo MAR.
Figure 18: Autoradiograms of DNAs in the supernatant and precipitate fractions hybridized to the 4.0 kb Eco RI and 3.3 kb Eco RI fragments.

The DNAs in both lanes were digested with Bam HI - Eco RI prior to fractionation. The numbers on the side indicate the size of the fragments hybridizing to the various probes in kb. The blots were not stripped of the previous probes to retain as much of the DNA as was present when the blot was first hybridized.
Figure 19: Map of the β-globin gene and surrounding regions showing the location of the nuclear matrix attachment region.

The enzymes indicated are E: Eco RI; B: Bam HI. The solid boxes indicate exons and the open boxes indicate introns. The broken bar indicate the location of the matrix attachment site.
Analysis of the deletions within the β-globin gene cluster indicates that some if not all of the deletions are probably occurring during replication at sequences that are involved in the nuclear organization of the DNA. Analysis of some of these deletions has allowed us to pinpoint regions of DNA that might be serving as origins of replication in addition to giving us insight into the nuclear architecture of a eucaryotic cell. Another aspect of the deletion-replication model is that the nuclear matrix attachment site probably contains an origin of replication, a notion that needs to be tested in the near future.
Sequence characterization of two regions located upstream of the β-globin gene that show extensive sequence polymorphism

A. Introduction

DNA polymorphisms, also known as restriction fragment length polymorphisms (RFLP), are frequently produced by single nucleotide substitutions, though some polymorphisms are produced by variations in the number of tandem repeats of short sequences of DNA. Two or more alleles are considered to be polymorphic if they are present in a population at frequencies of greater than 0.01 (Antonarakis et al. 1982). Linkage analysis using DNA polymorphisms has been useful in the past for prenatal diagnosis of genetically inheritable diseases such as the thalassemias (Kan et al. 1978; Kan et al. 1980; Little et al. 1980), and recently for Cystic Fibrosis (Estivill et al. 1988) and can be used for any number of genetic diseases that show linkage to a particular restriction enzyme polymorphism.

The first DNA polymorphism within the β-globin gene cluster was observed by Kan et al. 1978. Subsequent to that at least 17 common polymorphisms have been observed (see review Orkin et al. 1984). Each of these sites can either be present (+) or absent (-), and, in theory for n polymorphic restriction sites within a
given segment of DNA, there are $2^n$ possible combinations and each of these possible combinations has been termed a haplotype. These sites can be randomly or non-randomly associated with each other, and the way in which one determines random versus nonrandom association of these sites is best demonstrated in the following example: the probability of two sites being present together on a random basis being the product of the individual probabilities of the presence of the two sites. For example, if polymorphic site I is present in 60 percent of the chromosomes examined in a given population, and site II is present in 30 percent of the same chromosomes, then the theoretical probability of the presence of both sites in these chromosomes is $60 \times 30 = 18$ percent. If, in the chromosomes examined, one finds that the theoretical values deviate greatly from the actual values observed, then the sites are said to be non-randomly associated with each other. This non-randomness of association has been termed linkage disequilibrium as opposed to linkage equilibrium if they are randomly associated with each other. In such an analysis of the β-globin gene cluster, Antonarakis et al. 1982 have shown that in Italian, Greek, Indian, and Turkish populations, of the 32 possible combinations expected in theory for the five polymorphic sites (Hinc II sites numbered 2, 6, 7 and Hind III sites numbered 3, and 4 in Figure 20), only 3 combinations account for 92% of 89 chromosomes examined. The observed frequencies for the three haplotypes were 0.64, 0.15,
Figure 20: Schematic diagram of the β-globin gene cluster with the various polymorphic restriction enzyme sites (Adapted from Orkin et al. 1984).

The numbers 1 through 17 indicate the polymorphic sites Taq I, Hinc II, Hind III, Hind III, Pvu II, Hinc II, Hinc II, Rsa I, Taq I, Hinf I, Rsa I, Hgi Al, Ava II, Hpa I, Hind III, Bam HI, and Rsa I, respectively. The 5' and 3' haplotypes have been indicated below.
5' Haplotype

3' Haplotype

FIGURE 20
and 0.13, in contrast to the expected frequencies of 0.2, 0.006, and 0.005, indicating that these sites are in linkage disequilibrium. Two other sites (Bam HI site numbered 16, and Ava II site numbered 13) located at the 3' end of the cluster have similarly been shown to be in linkage disequilibrium. Further analysis of these sites indicate that sites numbered 2 through 9 show linkage disequilibrium with respect to each other as do sites numbered 12 through 17 to form the 5' and 3' partial haplotypes shown in Figure 20 (Orkin et al., 1984).

Thus the β-globin gene cluster has been subdivided into the 5' haplotype consisting of a 34 kb region extending from the e-globin gene to the 5' end of the δ-globin gene and the 3' haplotype consisting of the β-globin gene extending up to 18 kb downstream (Orkin et al., 1985) (Figure 20). Since the three common 5' haplotypes and three common 3' haplotypes observed in different populations within this locus are in linkage equilibrium with respect to each other, it has been proposed that a 9.1 kb region that lies between the 5' and 3' haplotypes is highly recombinogenic serving to randomize the 5' and 3' regions of the β-globin cluster (Orkin et al., 1985). This randomization of the 5' and 3' partial haplotypes with respect to each other has been termed linkage equilibrium. The intervening 9.1 kb region displays random association with both partial haplotypes as demonstrated by the fact that the Hinf I polymorphic site located 1 kb 5' to the β-globin gene and the Rsa
I polymorphic site located 550 bp 5' to the β-globin gene show no association with either the 5' or the 3' segments constituting the two partial haplotypes of the cluster (Kazazian et al. 1984).

Limited nucleotide sequence analysis of the 9.1 kb segment of DNA between the 5' and 3' haplotypes (Spritz et al. 1981; Moschanas et al. 1982; Ponc et al. 1983) has led to the speculation that the variable ATTTT pentanucleotide repeat located 1.4 kb 5' to the β-globin gene, a series of 16-17 TG dinucleotide repeats located 2.7 kb 5' to the β-globin gene, and the alternating purine - pyrimidine tract found approximately 550 bp 5' to the β-globin gene cluster, might play a role in the increased frequency of recombination within this region. Further evidence for increased recombination with a 3.1 kb Eco RI fragment 5' to the δ-globin gene has been observed (Maeda et al. 1983).

During the sequencing of the DNA from the clone derived from the chromosome bearing the 1.4 kb deletion we noticed that the sequences of the region that was not deleted did not match up exactly to the sequence of the same region from a normal chromosome at at least two nucleotides as discussed in chapter IV. One of the mismatches happened to lie within a stretch of polypurine - pyrimidine residues that has been shown to show some variability by Semenza et al. 1984. In an attempt to further analyze at the sequence level the sequence variability that
occurs within this region, the elements responsible for the observed linkage equilibrium, and the extent to which this randomization occurs within different populations, we sequenced the region containing the variable number of ATTTT pentanucleotide repeats located 1.4 kb 5' to the β-globin gene and the alternating purine - pyrimidine segment located approximately 550 bp upstream to the β-globin gene, from clones that were isolated from different chromosomes. In the discussion of the results, a similar numbering system as that used in chapter IV has been used.
B. Materials and Methods

Construction and screening of the size-selected libraries

Recombinant bacteriophage lambda libraries were constructed by ligating Eco RI digested and 4 kb to 7 kb size-selected genomic DNA from the various individuals to Charon 3A DNA, which had been digested with Eco RI and subsequently dephosphorylated. The ratio of genomic DNA to phage DNA was 1:1 on a microgram basis. The ligated mixtures were then packaged into phage particles using \textit{in vitro} packaging extracts. Libraries of approximately 100,000 recombinant phage were then screened with the 0.9 kb Bam HI - Eco RI probe containing the second intervening sequence of the \(\beta\)-globin gene.

Identity of the clones used

The following clones were used for sequencing the two regions:
1. p 4.2 RI Brit: A clone obtained from a Eco RI-Ch3A lambda library made using DNA obtained from an American Black suffering from a 1.4 kb deletion thalassemia. The cloned region originated from the chromosome bearing the deletion. The isolation of the clone has been described in chapter IV.
2. p 5.5 RI Brit: This clone was also obtained from the same library as p 4.2 RI Brit, except that the clone contains DNA from the normal chromosome.

3. p GB H3: This clone was obtained from a library made from the DNA of a person of Dutch origin.


5. p 5.5 RI MCI: This clone was obtained from an Eco RI Ch3A lambda library constructed using size-selected genomic DNA isolated from the blood donated by a normal individual of Korean origin, and the isolation has been described in the above section.

6. p 5.5 RI MC2: This clone was obtained from the same library as p 5.5 RI MCI and probably is from the other homologue.

7. p 5.5 RI ME: This clone was obtained from an Eco RI Ch3A library constructed in a similar way as for clones p 5.5 RI MCI and MC2, except that the DNA used was from an normal individual (despite the fact that he seems crazy at times), of Indian origin, who by the way is the candidate defending this thesis.
C. Results and Discussion

Both of the regions which we wished to sequence are contained within a 5.5 Kb Eco RI fragment which begin 4.1 kb 5' to the transcriptional start site of the β-globin gene and extended into the beginning of exon 3 of the β-globin gene. Therefore we chose to make complete Eco RI size selected libraries using DNA from different individuals, in Charon 3A lambda cloning vectors. These libraries were then screened with the 0.9 kb Bam RI - Eco RI fragment, which contains the second intervening sequence of the β-globin gene. In each case approximately $10^5$ recombinant phage plaques were screened. Positive hybridizing plaques were purified and the Eco RI inserts subcloned into pUC 19.

The sequence of the two regions of interest were then obtained from the fragments indicated in Table 3, using the partial chemical modification sequencing technique of Maxam and Gilbert (1980). The region containing the $(AT)_nT_n$ repeats was sequenced by using a 240 bp Acc I - Sph I fragment end-labeled at the Acc I site (at -372 ). The only exception to this was for the p4.2 RI Brit clone, in which this region was sequenced by using a Dra I - Sph I fragment end-labeled at the Dra I site. This was necessary as the clone was derived from the chromosome bearing the 1.4 kb deletion discussed in chapter IV and the Acc I
## Table 3

List of DNA fragments used for sequencing the -1.4 kb region

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Source</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bam HI - Rsa I</td>
<td>p 5.5 RI Brit. clone</td>
<td>345 bp</td>
</tr>
<tr>
<td>2. Bam HI - Rsa I</td>
<td>p 4.2 RI Brit. clone</td>
<td>345 bp</td>
</tr>
<tr>
<td>3. Bam HI - Dra I</td>
<td>pHβ PstI clone</td>
<td>430 bp</td>
</tr>
<tr>
<td>4. Bam HI - Dra I</td>
<td>p Gβ H3 clone</td>
<td>430 bp</td>
</tr>
<tr>
<td>5. Acc I - Sph I</td>
<td>pHβ PstI clone</td>
<td>240 bp</td>
</tr>
<tr>
<td>6. Bam HI - Rsa I</td>
<td>p 5.5 RI MC1 clone</td>
<td>345 bp</td>
</tr>
<tr>
<td>7. Bam HI - Rsa I</td>
<td>p 5.5 RI MC2 clone</td>
<td>345 bp</td>
</tr>
<tr>
<td>8. Bam HI - Rsa I</td>
<td>p 5.5 RI ME clone</td>
<td>345 bp</td>
</tr>
</tbody>
</table>
site was located within the sequence deleted in that chromosome. Sequences of the region containing the (ATTTT)$_n$ repeats were obtained from either a 345 bp Bam HI - Rsa I fragment or from a 430 bp Bam HI - Dra I fragment end-labeled at the Bam HI site ( - 1460 ). The only exception in this case was the pH8 Pst I clone, in which the sequence was obtained from a 240 bp Acc I - Sph I fragment which had been end-labeled at the Acc I site.

Sequence analysis of the 550 bp and 1.4 Kb regions

The sequence of the region containing the (ATTTT)$_n$ repeats and that of the region containing the (AT)$_n$T$_n$ repeats from each of the chromosomes analyzed are shown in Figure 21. Out of a total of 14 chromosomes analyzed, the two American Black sequences are from the same individual, as are the two Korean sequences. The number of repeats, (ATTTT)$_n$, (AT)$_n$ and T$_n$, as well as restriction enzyme polymorphism data is summarized in Table 4. The polymorphism sites listed in Table 4 are the Rsa I site at - 553 and the Hgi Al site at + 57. For ease of discussion the individuals listed in Table 4 will be referred to as 1 through 14. Sequence of the clones labeled 6, 7, 8, 9, 10, and 11 were obtained from the laboratory of Dr. Haig Kazazian, Johns Hopkins University School of Medicine, as part of a collaboration between our laboratories. Comparison of the number
Figure 21: Sequence of the variable regions located 1.4 kb and 550 bp 5' to the β-globin gene.

Shown in the top line is the immediate 5' region of the β-globin gene and the two variable regions at 1.4 kb and 550 bp. Indicated in the next line are the pentanucleotide repeating unit, the purine-pyrimidine stretch and the variable thymidine stretch. The rest of the figure contains the actual sequence of these regions from the various clones. The ethnic origin of the individual from which the DNA was isolated is indicated in bold print above each of the sequences. The sequence of the Kurdish - Jew was obtained from Poncz et al. (1983). The question marks indicate that the sequence of that region has not been determined. The sequence of the Laotian, the two Iranians, the northern European and the two Blacks were obtained as part of a collaboration between Dr. Haig Kazazian's laboratory and our laboratory.
KURDISTAN JEW:

American Black:

Dutch:

UNKNOWN:

LAOTIAN:

IRANIAN:

N. EUROPEAN:

BLACK:

KOREAN:

INDIAN:

FIGURE 21
### Table 4

Summary of the sequence analysis of the -1.4 Kb and -550 bp regions

<table>
<thead>
<tr>
<th>Origin</th>
<th># of (ATTTT)</th>
<th>Rsa I</th>
<th># of (AT)</th>
<th># of T</th>
<th>HgiA I</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. KURDISH - JEW</td>
<td>5</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>+</td>
<td>A*</td>
</tr>
<tr>
<td>2. AMERICAN BLACK</td>
<td>5</td>
<td>+</td>
<td>6</td>
<td>9</td>
<td>N.P.</td>
<td>B</td>
</tr>
<tr>
<td>3. AMERICAN BLACK</td>
<td>5</td>
<td>-</td>
<td>8</td>
<td>4</td>
<td>+</td>
<td>C</td>
</tr>
<tr>
<td>4. DUTCH</td>
<td>4</td>
<td>+</td>
<td>9</td>
<td>5</td>
<td>+</td>
<td>D</td>
</tr>
<tr>
<td>5. UNKNOWN</td>
<td>4</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>+</td>
<td>E*</td>
</tr>
<tr>
<td>6. LAOTIAN</td>
<td>5</td>
<td>+</td>
<td>7</td>
<td>7</td>
<td>+</td>
<td>F</td>
</tr>
<tr>
<td>7. IRANIAN</td>
<td>4</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>+</td>
<td>E*</td>
</tr>
<tr>
<td>8. IRANIAN</td>
<td>7 (+ATT)</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>G</td>
</tr>
<tr>
<td>9. N.EUROPEAN</td>
<td>5</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>+</td>
<td>A*</td>
</tr>
<tr>
<td>10. BLACK</td>
<td>5</td>
<td>-</td>
<td>6</td>
<td>9</td>
<td>+</td>
<td>H</td>
</tr>
<tr>
<td>11. BLACK</td>
<td>4</td>
<td>-</td>
<td>8</td>
<td>9</td>
<td>-</td>
<td>I</td>
</tr>
<tr>
<td>12. KOREAN</td>
<td>5</td>
<td>+</td>
<td>9</td>
<td>5</td>
<td>+</td>
<td>J</td>
</tr>
<tr>
<td>13. KOREAN</td>
<td>5</td>
<td>+</td>
<td>10</td>
<td>3</td>
<td>+</td>
<td>K</td>
</tr>
<tr>
<td>14. INDIAN</td>
<td>5</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>L</td>
</tr>
</tbody>
</table>
of (ATTTT)_n repeats indicates that the most common number of repeats is 5 (present in chromosome numbers 1, 2, 3, 6, 9, 10, 12, 13, and 14). Four (ATTTT)_n repeats were found in four chromosomes (chromosomes 4, 5, 7 and 11). Chromosome 8 showed a unique number of repeats in that seven repeat units were followed by what appears to be a truncated repeat consisting of a ATT trinucleotide. Spritz et al. (1982) have reported individuals who have four, five and six such repeats. We also determined the polymorphic status of each chromosome from which each clone originated, with respect to the restriction enzyme Rsa I site at -553. This polymorphism was found to be caused by a T to C change (shown by an asterisk in Figure 21) resulting in the sequence GTAT (Rsa I^-) or GTAC (Rsa I^+). The status was determined from our sequence analysis of each chromosome. Of the 14 chromosomes, six were found to be Rsa I^+, and there seemed to be no correlation between the presence or absence of the Rsa I site and the number of (ATTTT) repeats. For example, chromosomes which have five (ATTTT) repeats can either be Rsa I^+ or Rsa I^-. The same is true for chromosomes which have four (ATTTT) repeats. This is surprising considering that these two regions are approximately 800 bp apart. The sequence analysis also enabled us to determine the number of (AT) repeats and the number of T residues immediately 3' to the (AT) repeats. The counting of the (AT) repeats begins immediately 3' to the C residue
indicated by the first vertical arrow and the counting of the number of T residues ceases immediately 5' to the C residue indicated by the second vertical arrow in Figure 21. As can be seen from the tabulated results (Table 4), by far the most common chromosome contains seven (AT) repeats followed by seven T residues. This accounts for half the chromosomes analyzed. Three chromosomes have nine (AT) repeats followed by five T residues and two chromosomes have six (AT) repeats followed by nine T residues. Each of the other chromosomes are represented only once (Table 4). The status of the HgiA I polymorphic site (Orkin et al. 1982) for each chromosome is also included in Table 4. This site is the 5' most site of the 3' haplotype; that is, this site is in linkage disequilibrium with other polymorphic sites 3' to it. As can be seen from the results of this analysis of the 14 chromosomes analyzed (Table 4), twelve were found to be different as designated by the letters A through L assigned on an arbitrary basis. These results indicate that this region of the β-globin gene cluster must be undergoing frequent recombinational events. This is substantiated by the fact that the polymorphisms in this region are in linkage equilibrium with one another. That is, the Hinf I polymorphism shows no linkage to the Rsa I site which itself shows no linkage with the HgiA I. It has also been found that the Hinf I site shows no linkage to the 5' haplotype indicating recombination occurs frequently between these sites.
The question remains as to which sequences are actually involved in the recombinational events which give rise to this linkage equilibrium. An excellent candidate for such a sequence between the Rsa I and HgiA I sites is the \((AT)_nT_n\) repeat region which was first suggested by Semenza et al. (1984). The data concerning the number of each repeat, \((AT)_n\) and \(T_n\), is summarized in Table 4. Until this point, the expansion of such simple repeats has been thought to be due to strand slippage and mispairing between the tandem repeat units on the leading and trailing strands during replication. If one used such a mechanism to explain the changes in \((AT)_n\) and \(T_n\) repeats then one would have to postulate two separate events, one for the change in the number of \((AT)\) repeats and one for the change in the number of \(T\) residues. Although, we believe that this mechanism is probably operational in generating some of these changes, it is not the only mechanism by which one could generate these changes. Some of the changes are probably the result of unequal homologous crossover in this region. The unequal homologous crossover event shown in Figure 22 indicates how such an event between two \((AT)_7T_7\) repeats could give rise to a repeat of either \((AT)_6T_9\) or \((AT)_8T_5\). Therefore a single recombinational event can be used to explain a change in both the number of \((AT)\) repeats as well as the number of \(T\) residues. However, such a recombinational event would result in a reduction of the number
Figure 22: Model of the crossover event between two misaligned \((AT)_7T_7\) chromosomes to generate the \((AT)_8T_5\) and \((AT)_6T_9\) chromosomes.

The two \((AT)_7T_7\) chromosomes are shown during a crossover event in the top third of the figure. For clarity, the sequence of one of the two misaligned chromosomes has been bolded. The product of the crossover is shown in the middle third of the figure. The bottom third shows the final products after repair of the mismatched bases to give either a \((AT)_8T_5\), \((AT)_6T_9\) or \((AT)_7T_7\) repeat chromosome.
of (AT) repeats by 1 with a concomitant increase in the number of T residues by two. If one analyzes the six different chromosomes (with respect to the (AT)$_n$T$_n$ repeat structure; Table 4) such a mechanism can account for four of those chromosomes. The other two repeat structures, the (AT)$_8$T$_9$ and (AT)$_9$T$_4$, could be produced as a result of slippage and repair during replication from the (AT)$_6$T$_9$ and (AT)$_9$T$_5$ repeats, respectively. As was shown in Figure 22, the (AT)$_7$T$_7$ repeat can give rise to the (AT)$_6$T$_9$ repeat structure. This same mechanism can be used to describe how the (AT)$_9$T$_5$ repeat can give rise to the (AT)$_{10}$T$_3$ repeat structure. If such unequal homologous events are occurring within this region then one can imagine the occurrence of many more equal homologous crossover events, which could not be detected as there would be no net change in the nucleotide sequence of this region. The high frequency of such crossover events explains the linkage equilibrium observed between the Rsa I and HgiA I polymorphic sites.

Such multiallelic variations have previously been observed in human minisatellite regions (Jeffrey et al. 1985), c-Ha-ras-1 oncogene (Capon et al. 1983), the human insulin gene (Bell et al. 1982), and in the α-globin genes (Higgs et al. 1981). The tandemly repeating unit size varied in these cases from 11 to 60 nucleotides in length. It is also quite probable that the 550 bp region ending in the variable thymidine residues represents the 3' most border of the 9.1 kb region that serves to randomize the 5' and 3' partial haplotypes.
CHAPTER VII

Role of eucaryotic DNA topoisomerase II in recombination within the β-globin gene cluster

A. Introduction

This chapter describes our attempts to study the possible role of eucaryotic DNA topoisomerase II in generating the deletions observed in several naturally occurring β-thalassemias and in generating the nucleotide sequence polymorphisms observed at the hot spots for recombination found 5' to the β-globin gene (described in chapter VI). DNA topoisomerase II is an enzyme that transiently breaks and rejoins double stranded DNA in the presence of ATP. The mechanism involves passage of one strand around the other during the temporary cleavage of double stranded DNA. The cleavage occurs such that a four base pair staggered break is produced. In vitro, topoisomerase II has been shown to be capable of catalyzing the relaxation of positive and negative supercoils of DNA, as well as catenation and decatenation of covalently closed circular DNA (see review Wang 1985). This enzyme has been isolated from a wide variety of sources including yeast, Drosophila, Xenopus, and mammalian cells (Goto et al. 1982; Miller et al. 1981; Sander et al. 1983; Shelton et al. 1983; Halligan et al. 1985). The mammalian topoisomerase II has
been shown to be a homodimeric protein of $M_w = 170$ KD (Miller et al. 1981; Halligan et al. 1985).

Although the exact role of topoisomerase in higher eucaryotes is not known, it has been shown to be essential for mitosis in S. pombe and S. cerevisiae (Clemura et al. 1987; Uemura et al. 1984; Holm et al. 1985; Dinardo et al. 1984). During this stage of mitosis it has been shown to be required for the decatenation of daughter chromosomes after DNA replication. Several lines of evidence in other systems studied indicate that topoisomerase II might be involved in the organization of chromatin domains, in replication, and in the introduction of torsional stress into transcriptionally active chromatin. For example, using $^3$H-thymidine pulse labeled DNA and by teniposide-induced trapping of topoisomerase II-DNA complexes, it has been shown that newly replicated DNA in cultured rat prostatic adenocarcinoma cells is associated with topoisomerase II (Nelson et al. 1986). Additional interesting observations are that topoisomerase II is a major component of the nuclear matrix (Earnshaw et al. 1985) and matrix associated regions of DNA in the mouse K immunoglobulin gene have been shown to contain topoisomerase II binding sites (Cockerill et al. 1986). In addition, this enzyme has also been localized to the base of DNA loops of mitotic chromosomes (Earnshaw et al. 1985), suggesting a functional link between chromosomal loop attachment sites,
topoisomerase II and modulation of torsional stress. Thus, from information presently available, it seems likely that topoisomerase II is involved in controlling or mediating topological changes in chromatin structure by its ability to make and reseal double stranded cleavages and it may also be involved in DNA replication itself. Its role in recombination, if any, within mammalian cells has not been established. However, topoisomerase II isolated from calf thymus has been shown to be capable of mediating nonhomologous recombination between lambda DNA molecules in vitro (Bae et al. 1988). Thus, we decided to examine the role of topoisomerase II in recombination within mammalian cells.

There were two main objectives for mapping in vitro topoisomerase II cleavage sites within the β-globin gene cluster. The first was to see if the cleavage sites correlated with deletion breakpoints observed within naturally occurring deletion β-thalassemias. The second objective was to see if the cleavage sites correlated to highly recombinogenic regions within the β-globin gene cluster.
B. Materials and Methods

**In vitro cleavage Reactions**

Topoisomerase II cleavage reactions were performed in a final volume of 20 ul in a standard cleavage buffer containing the following:

- Tris pH 7.6: 30 mM
- KCl: 60 mM
- β-mercaptoethanol: 15 mM
- MgCl₂: 8 mM
- ATP: 3 mM
- BSA: 30 ug/ml

DNA fragments used as substrates were 5' end-labeled with γ-^{32}P-ATP using T4 polynucleotide kinase in the standard protocol described by Maniatis *et al.* 1982. The reactions were performed at 30° C using approximately 4 units of topoisomerase II for 30 mins with 1-10×10^4 dpm labeled DNA. Where indicated 1 ul of a stock of 1 mg/ml of m-AMSA was added to the reaction. Both the chicken topoisomerase II and mAMSA used were obtained from M.T. Muller's laboratory (O.S.U.). Reactions with topoisomerase II were terminated by addition of 40 ul of a 1.5 % SDS, 250 mM EDTA
solution. This was followed by proteolysis using 50 ug/ml of proteinase K at 55°C for 30 mins. The samples were then precipitated by 2 volumes of ethanol after the addition of 0.1 vol of 3 M sodium acetate. The pellets were centrifuged at 13000xg for 20 mins following freezing the samples at -70°C for 10 mins. The pellets were dried in a vacuum oven and resuspended in sequencing gel loading buffer at 1x10^4 dpm/ul. The samples were heated for 5 mins at 90°C and quick cooled on ice prior to loading on a 6-8% sequencing bis-acrylamide-urea gel.

Isolation of human placental Topoisomerase II

A complete human placenta (obtained from the O.S.U hospital) was cleaned of all clotted blood by suspending the placenta in TEMP buffer (10 mM Tris pH 7.5, 1 mM EDTA, 5 mM MgCl₂, 1 mM PMSF). The umbilical cord and any connective tissue was dissected away. The placenta was then chopped into little bits using scissors and washed again in the TEMP buffer. The minced placenta was blended in a Waring blender for two 30 sec pulses at highest speed in the presence of ice cold TEMP buffer. The homogenate was filtered through two, four, and eight layers of cheese cloth successively and the filtrate was collected. The cell nuclei were pelleted by centrifugation at 9000xg for 20 mins and resuspended in an ice cold buffer containing 50 mM Tris pH
7.5, 50 mM NaHSO₃ pH 7.6, 60 mM β-mercaptoethanol, 1 mM PMSF. Solid NaCl was then added to a final concentration of 0.4 M and the cell suspension was stirred at 4° C for 1 hour. Polyethyleneimine was added to a final concentration of 0.3 % and the cell suspension was stirred for a further 20 mins at 4° C. The suspension was centrifuged at 9000xg for 15 mins. Ammonium sulfate was added slowly to the supernatant until 70 % saturation. Following stirring for 40 mins at 4° C, the suspension was then centrifuged at 9000xg for 25 mins. The pellet obtained was resuspended in buffer A containing 0.1 M potassium phosphate pH 7.1, 10 % (v/v) glycerol, 25 mM β-mercaptoethanol, 10 mM NaHSO₃, 0.5 mM PMSF. The suspension was loaded onto a 50 ml Biorex 70 column that had been equilibrated with 0.2 M buffer A. The procedure following this point was carried out by J. Spitzner from M.T. Muller's laboratory (O.S.U.) as part of a collaboration. The eluate was loaded onto a 4 ml phenylSepharose column equilibrated with buffer B containing 20 % (v/v) ethylene glycol, 25 mM potassium phosphate pH 7.1, 10 mM NaHSO₃, 25 mM β-mercaptoethanol, 0.5 mM PMSF, then washed with 10 column volumes of buffer B. The enzyme was eluted off the column with buffer B containing 60% ethylene glycol. Topoisomerase II activity was assayed using decatenation of kinetoplast DNA, 1 unit being defined as the amount of enzyme required to decatenate 1 ug of DNA in 30 mins at 30° C.
C. Results and Discussion

Correlation of in vitro topoisomerase II cleavage sites to β-thalassemia breakpoints

To test whether some of the β-thalassemia deletions were the result of cleavages catalyzed by DNA topoisomerase II, we undertook to map the in vitro cleavage sites on fragments of DNA within certain regions of the β-globin gene cluster that were known to contain some of the breakpoints of previously characterized β-thalassemias. The fragments tested are shown in Table 5, together with their approximate size and the deletion breakpoint which they contained. Figure 23 shows the location of the fragments tested. The 5'-end-labeled fragment was subject to topoisomerase II cleavage in the presence and absence of mAMSA as described in the materials and methods. Following treatment with NaOH/SDS and proteinase K, the reactions were phenol-extracted and the DNA ethanol-precipitated. The cleavage products were electrophoresed adjacent to G and G+A base-specific reactions of the same fragments. Figure 25 shows an example of the results obtained when the reactions were performed with the 0.9 Bam HI - Eco RI fragments. The results obtained from the various fragments tested are shown in Figures 25, 27, 28 and 30. The position of the cleavage sites within the actual sequence are
Figure 23: Schematic representation of the β-globin gene showing the location of the fragments used for the topoisomerase II cleavages.

The fragments labeled 1, 2, 3, and 4 are the Acc I - Sph I, Eco RI - Acc I, Bam HI - Eco RI and Eco RI - Bam HI fragments labeled at the sites marked with an asterisks. The restriction sites indicated are A: Acc I; B: Bam HI; E: Eco RI; and S: Sph I.
### Table 5

List of DNA fragments used in the topoisomerase II cleavage reactions

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Size</th>
<th>Breakpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. p0.8 Acc I - Eco RI</td>
<td>500bp</td>
<td>5'breakpoint γδβ° Thal</td>
</tr>
<tr>
<td>2. p4.2 Dra I - Sph I</td>
<td>210bp</td>
<td>breakpoint jn. 1.4kb β°-Thal</td>
</tr>
<tr>
<td>3. p5.5 Acc I - Sph I</td>
<td>240bp</td>
<td>5'breakpoint 1.4kb β°-Thal</td>
</tr>
<tr>
<td>4. p5.5 Hpa I - Acc I</td>
<td>440bp</td>
<td>5'breakpoint 1.4kb β°-Thal</td>
</tr>
<tr>
<td>5. p0.9 Bam HI - Eco RI</td>
<td>900bp</td>
<td>3'breakpoint 1.4kb β°-Thal</td>
</tr>
<tr>
<td>6. p0.9 Eco RI - Bam HI</td>
<td>900bp</td>
<td>5'breakpoint 0.6kb β°-Thal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3'breakpoint 4.3kb Czech.Thal</td>
</tr>
</tbody>
</table>
Figure 24: Schematic diagram of the in vitro procedure used to identify topoisomerase II cleavage sites.

The solid bars indicate single strands of DNA. The asterisks indicate the 5' $^{32}$p label. Prot. K stands for proteinase K. The open circles indicate the covalently linked topoisomerase II to the DNA strands.
5' End Labeled DNA

Topo II
mAMSA

SDS
NaOH

Prot.K

FRACTIONATE ON
SEQ. GEL

FIGURE 24
Figure 25: Autoradiogram of 8% sequencing gel containing the fractionated topoisomerase II cleavage products of the p0.9 Bam HI - Eco RI fragments.

The left panel contains the results of the Maxam and Gilbert sequencing reactions and the topoisomerase II reactions performed on the Bam HI - Eco RI fragment 5' end-labeled with $^{32}$P at the Bam HI site. The left set of lanes is the first loading and the right set of lanes is the second loading of the reactions. The arrow marked I indicates the 5' breakpoint of the 1.4 kb deletion thalassemia as discussed in the text.

The right panel contains the results of the Maxam and Gilbert sequencing reactions and topoisomerase II reactions performed on the Eco RI - Bam HI fragment 5' labeled with P32 at the Eco RI site. The arrow marked I indicates the 5' breakpoint of the 0.6 kb Indian deletion. The arrow marked II indicates the 3' breakpoint of the 4.3 kb deletion, and III indicates the 3' breakpoint of the 1.4 kb deletion. The arabic numerals indicate the cleavage sites that were mapped on these fragments.
Figure 26: Sequence of the Bam HI - Eco RI fragment showing the location of the topoisomerase II cleavage sites mapped.

The topoisomerase II sites are marked with a small bar above the bases within which the cleavages occurred. The arabic numbers correspond to the cleavage sites marked with the same numbers in figure 32. The uncertainty of two nucleotides in the cleavage site is due to the fact that only G and G + A base specific sequencing reactions were used to map the cleavages. The Roman numerals I, II and III correspond to the 5' breakpoint of the 0.6 kb, the 3' breakpoint of the 4.2 kb deletion and the 3' breakpoint of the 1.4 kb thalassemias respectively.
FIGURE 26 continued
summarized in Figures 26 and 29. The results summarized in Figure 29 are those which were obtained from the *Acc I - Sph I and the *Hpa I - Acc I fragments shown in Figure 28. Due to the resolution of the gels used, only cleavage sites within approximately 200 bp of the labeled end could be positioned with any certainty. The various breakpoints within the sequences tested are indicated by Roman numerals. As can be seen from Figures 25 and 26, topoisomerase II cleavage sites do flank the 3' breakpoint for the 1.4 kb deletion β0-thalassemia. On the other hand, the 3' breakpoint for the 4.3 kb deletion β0-thalassemia is not as close to any topoisomerase II sites. The same is true for the 5' breakpoint for the 0.6 kb deletion β0-thalassemia and also for the 5' breakpoint of the 1.4 kb deletion β0-thalassemia (Figures 25, 26, 28 and 29). We also subjected a Eco RI - Acc I fragment which contained the 5' breakpoint for the γεβ-thal 1 to topoisomerase II cleavage. As with the others, the breakpoint does not coincide with a topoisomerase II cleavage site (Figure 27). Thus, not only does there seem to be no correlation between breakpoints and topoisomerase II cleavage sites but each of the fragments tested had numerous topoisomerase II cleavage sites. In addition, the strongest cleavage sites were not the ones nearest to the breakpoints.
Correlation between *in vitro* topoismerase II cleavage sites and the hypervariable region 5' to the β-globin gene

Two of the fragments tested in the above experiments, the Hpa I - Acc I and the Acc I - Sph I fragments, contained the hypervariable \((AT)_nT_n\) region discussed in the previous chapter. In the previous chapter we showed that this stretch of DNA undergoes frequent crossover events as deduced from the sequence hypervariability of this region observed in chromosomes of different individuals. It was postulated that frequent crossover may in fact be responsible for the linkage equilibrium observed between the 5' and 3' haplotypes observed within the β-globin gene cluster and this variable stretch of DNA may be the 3' border of the area of randomization within this cluster. Thus, it was of interest to see if this region of DNA contained strong topoisomerase II cleavage sites. As can be seen from Figure 28, we localized two very strong cleavage sites in addition to several other cleavages of weaker strength within this region. The band marked 1 in the Hpa I - Acc I fragment is one of the strong cleavage sites, and the band marked 4 in the Acc I - Sph I fragment is the other cleavage site as shown in Figure 28. The strength of the cleavage by topoisomerase II at two sites flanking this variable sequence seems significant in view of the fact that the region flanked by the two strong cleavages is the 550 bp region 5' to the β-globin gene that has been shown to be...
Figure 27: Autoradiogram of 8% sequencing gels containing the fractionated topoisomerase II cleavage products of the p0.8 Acc I - Eco RI and the p4.2 Dra I - Sph I fragments.

The left panel contains the results of the Maxam and Gilbert sequencing reactions and the topoisomerase II reactions performed on the 500 bp p0.8 Acc I - Eco RI fragment. The arrow marked II indicates the location of the 5' breakpoint of the γδβ0 deletion thalassemia 1.

The right panel also contains the results of the Maxam and Gilbert sequencing reactions and the topoisomerase II reactions performed on the 210 bp p4.2 Dra I - Sph I fragment. The arrow marked I indicates the location of the breakpoint junction of the 1.4 kb deletion thalassemia.
FIGURE 27
Figure 28: Autoradiogram of 8% sequencing gels containing the fractionated topoisomerase II cleavage products of the p5.5 Acc I - Sph I and p5.5 Hpa I - Acc I fragments.

The left panel contains the results of the Maxam and Gilbert sequencing reactions and the topoisomerase II reactions performed on the 240 bp Acc I - Sph I fragment. The arrow marked I indicates the location of the 5' breakpoint of the 1.4 kb deletion thalassemia.

The right panel contains the results of the Maxam and Gilbert sequencing reactions and the topoisomerase II reactions performed on the 240 bp p5.5 AccI - Sph I fragment. The arabic numerals besides each band are used to indicate cleavage sites that were mapped on these fragment.
Figure 29: Sequence of the Hpa I - Acc I fragment showing the location of the topoisomerase II cleavage sites mapped.

The topoisomerase II cleavage sites are marked by a small bar above the nucleotides within which the cleavages occur. The arabic number correspond to the sites shown with the same numbers in figure 35. The uncertainty of two nucleotides in the cleavage site is due to the fact that only the G and G+A base specific sequencing reactions were used to map the cleavage sites. The Roman numeral I indicates the location of the 5' breakpoint of the 1.4 kb deletion. The dashed bars over the sequence indicates the sequence located 550 nucleotides 5' to the β-globin gene that was shown to exhibit sequence polymorphism. The asterisks indicate strong cleavage sites.
FIGURE 29
hypervariable in chapter VI (Figure 29). Most fragments we analyzed do not exhibit such strong cleavages. The upstream cleavage site is located about 200 ± 5 bp 5' to the polypurine/pyrimidine tract whilst the downstream cleavage site is located 82 ± 1 bp 3' to the stretch of thymidine residues downstream to the polypurine/pyrimidine tract (shown schematically in Figure 31). It is also interesting to note from the Hpa I - Acc I fragment (Figure 28) that this region also seems to contain multiple cleavage sites as indicated by the broad band in that region of the sequence. As this band was at the top of the gel we were not able to map it precisely.

Therefore topoisomerase II cleavages at either one of the two strong sites or the sites within the repeat could be the first event which initiates strand invasion between two homologues as shown in Figure 32. If the topoisomerase II cleavage occurred at either of the two sites, one would have to postulate branch migration in order to explain the expansion/contraction of the repeat. If on the other hand the cleavage was within the repeat region, then the misalignment would occur as a result of strand invasion. Following resolution this would result in expansion/contraction of the repeat structure.
Comparison of Chicken and Human topoisomerase II

Since all of the cleavages on the human β-globin gene sequences were carried out using chicken topoisomerase II, we were not sure of the significance of the cleavages in a heterologous system. The chicken topoisomerase II cleavages differed from the consensus sequence published for the Drosophila enzyme (Sanders et al. 1985). Given the differences, we decided to isolate topoisomerase from human placenta and determine the specificity of cleavage in comparison to the chicken topoisomerase II. The enzyme was isolated as described in the materials and methods section. Cleavage reactions were performed on the 240 bp p5.5 Acc I - Sph I fragment from the 5' region of the β-globin gene. This fragment had been shown to contain one of the strong cleavage sites with the chicken enzyme and was an ideal fragment to use for the comparisons. Fig 30 shows the autoradiogram of the 8% sequencing gels containing the fractionated cleavage products using the chicken, human, and Drosophila topoisomerase II. The Drosophila and chicken enzyme used were provided by M.T. Muller. The chicken enzyme shows virtually identical cleavages as the human enzyme. The relative strength of the cleavages also seem to correlate well between the chicken and human enzyme. Interestingly, the Drosophila enzyme shares a subset of cleavage sites with the two vertebrate enzymes but does have unique cleavages, though the strong cleavage sites
Figure 30: Autoradiogram of 8% sequencing gel containing the fractionated Human, Chicken and Drosophila topoisomerase II cleavage products of p5.5 Acc I - Sph I fragment.

Shown is the comparative cleavages between Human, Chicken and Drosophila topoisomerase II performed on the 240 bp p5.5 Acc I - Sph I fragment. The arabic numerals beside each band are used to indicate either unique or common cleavage sites between the three enzymes, 1, 2, and 3 being unique to Drosophila and 4 being common to all three enzymes.
Figure 31: Schematic diagram of the β-globin gene showing the location of the two strong topoisomerase II sites mapped flanking the 550 bp region shown to exhibit a high degree of sequence polymorphism.

The topoisomerase II sites are marked with arrows. The closed box represents the β-globin gene. The two regions shown to exhibit sequence polymorphism are shown below.
Figure 32: Schematic diagram showing a possible mechanism by which topoisomerase II could initiate recombination between two homologues to account for the sequence polymorphisms observed at certain regions.

A pair of solid lines represents one chromosome and the pair of broken lines represent the homologous chromosome. Topoisomerase II is represented by open circles. Panel A shows the initial cleavage reaction followed by strand invasion between homologues. Panel B shows mutual exchange of short segments of DNA following a crossover between invading strands, topoisomerase being still covalently linked to the DNA. Panel C shows the sealing of the initial cleavage by topoisomerase II followed by dissociation of the enzyme. Panel D shows the recombinant chromosomes after replication.
observed with the vertebrate enzymes is also a relatively strong cleavage site for the Drosophila enzyme. Thus, the chicken and human enzyme seem to have conserved the sequences at which they cleave. The cleavages identified with the chicken enzyme on the human sequences thus seem relevant.

It is possible to envisage a mechanism in which topoisomerase II may initiate frequent recombination events in this region by making single stranded breaks flanking this hypervariable region thus increasing its propensity to participate in meiotic or mitotic crossover events as shown. The mechanism shown involves the initial cleavage of one of the two homologues during a period when they are aligned next to each other. The next step involves invasion of one of the strands into the homologue forming a classical Holiday junction. Following the strand invasion, the next step is crossover at an appropriate point followed by resolution of the junction and religation of the breaks by topoisomerase II. Chromosomal segregation followed by DNA replication would result in the final product shown in the figure. It has also been shown that the in vivo cleavage sites for topoisomerase II are a subset of the in vitro cleavage sites (Yang et al. 1985; Udvardy et al. 1986). So it is quite possible that these sites actually represent in vivo cleavages, as a result of which frequent recombination occurs in this region and is reflected in the sequence polymorphism observed this region of the β-globin locus. Future
experiments need to be focused on determining if these strong cleavages do correlate with the in vivo cleavage sites of this enzyme to add strength to this observation, and for it to take on greater physiological significance.
CHAPTER VIII

Characterization of a deletion γδβ-thalassemia of Scotch-Irish descent

A. Introduction

This section describes our attempt to characterize at the sequence level a large deletion that removes the entire β-globin gene cluster. Pirastu et al. 1983 have described four generations of a family of Scotch-Irish descent in which a deletion, encompassing the entire β-globin cluster, results in a γδβ-thalassemia. Affected individuals exhibit normal HbA₂ and HbF levels and a decreased β/α globin chain ratio. Hybridization with DNA probes upstream and downstream from the β-globin gene cluster demonstrated that the deletion starts over 100 kb 5' to the β-globin gene cluster and extends to approximately 7 kb 3' to the ε-globin gene. The extent of the deletion was obtained as a result of the results discussed in the next segment. Using a probe γδ-1 (Van der Ploeg et al. 1980), located approximately 100 kb 5 to the ε-globin gene (Probe 1, Figure 33), they obtained the following results.

Hybridization of this probe to the Taq I digested DNAs from the father, mother and daughter, yielded a 2.3 kb fragment from the affected father, a 3.6 kb fragment from the mother, and a 3.6
Figure 33: Schematic diagram of the β-globin gene cluster showing the extent and approximate location of the 3' breakpoint of the Scotch - Irish deletion thalassemia, along with some of the restriction sites mapped around the 5' and 3' breakpoints.

The arrow heads above the gene locus indicate the location of Alu I family repeat sequences. The broken horizontal line indicates the location of a Kpn I family repeat sequence. The solid bar below the locus indicates the extent of the deletion. The open bar indicates uncertainty in the exact location of the 3' breakpoint. Thal 3 stands for the thalassemia of Scotch - Irish origin. Probes 1, 2, and 3 indicate the location of the ggdd-1, 0.9 Bam HI - Eco RI and pRK 28 probes. The restriction enzymes shown are B: Bam HI, H: Hind III, and T: Taq I. Sites that are marked with an asterisk are polymorphic.
A 3.6 kb and 2.3 kb fragments obtained are the result of a Taq I polymorphism as shown in Figure 33. The fact that the affected daughter inherited only the 3.6 kb Taq I fragment from the mother indicated that on the chromosome carrying the deletion, the DNA corresponding to the probe had been deleted. Hybridization of a βcDNA probe (Probe 2, Figure 33) to Bam HI digested DNAs from the three individuals yielded a 9.3 kb and 1.8 kb fragment from the mother, a 22 kb and 1.8 kb from the father, and a 9.3 kb and 1.8 kb from the daughter. The common 1.8 kb fragment is obtained as the cDNA probe spans a Bam HI site within the β-globin gene as shown in Figure 33. The authors concluded that the absence of the 22 kb fragment in the daughter was a result of the deletion encompassing the β-globin gene. Hybridization of the pRK 28 probe (Kaufman et al. 1980) (Probe 3, Figure 33) to Hind III digests of the three individuals yielded a 13.3 kb and 11.5 kb band from the mother, a 11.5 kb band from the father, and both the 13.3 kb and 11.5 kb band from the daughter. The two fragments are obtained as a result of a polymorphic Hind III site 7 kb 3' to the β-globin gene. This indicated that the deletion on the affected chromosome of the daughter did not encompass the polymorphic Hind III site 7 kb 3' to the β-globin gene and the 3' breakpoint of the deletion must lie between this Hind III site and the 3' end of the β-globin gene.
B. Materials and Methods

Preparation of dot blots

100 to 200 ng of DNA from each fragment was blotted onto nitrocellulose, following which the DNA was denatured by treating the nitrocellulose with 0.4 M NaOH, and then with 1 M Tris - HCl, pH 7.5. Hybridizations were performed as described in the general materials and methods section.

Screening of the jumping library

A partial Mbo I jumping library obtained from Dr. Francis Collins, at the University of Michigan, Ann Arbor, was screened with the following probes: 1) a 0.9 kb Bam HI - Eco RI probe containing β-IVS II 2) a 3.3 kb Eco RI fragment located 20 kb 5' to the ξ globin gene and 3) a 4.0 kb Eco RI fragment located 55 kb 5' to the ξ globin gene. A total of 800,000 clones were screened the first time. All the hybridizations were done as described in the general materials and methods section, except 5 to 10 ug/ml of pBR 328 DNA was added to prevent hybridization of residual plasmid sequences in the probe to pBR 328 DNA inserted in the library during its construction at a frequency of 1/1000 clones.
Construction of Bam HI EMBL 3 library containing thal 3 DNA

A recombinant bacteriophage lambda library was constructed by ligating Bam HI genomic DNA, from an individual heterozygous for the large thal 3 deletion, to Bam HI EMBL 3 arms purchased from Stratagene in a ratio of 1:1 on a microgram basis. The ligated mixture was then packaged into phage particles using the in vitro packaging extracts. A library of 100,000 clones was screened using the pRK 28 probe (a 1.1 kb Eco RI - Bgl II fragment) located 17 kb 3' to the β-globin gene.

C. Results and Discussion

Cloning of an appropriate fragment located 3' to the deletion

The first approach was to identify and clone restriction fragments using available distal unique probes within the 3' region of the β-globin gene such that they would encompass the breakpoint junction. To do this, we decided to use the pRK 28 probe located 17 kb downstream to the β-globin gene. From preliminary data available at the time regarding the restriction sites 3' to the β-globin gene we decided to use Bam HI. We decided to determine the polymorphic status of the normal and abnormal chromosome with respect to this enzyme since the Bam HI
site located approximately 7 kb 3' to the β-globin gene had been shown to be polymorphic (Kan et al. 1980). Using the pRK 28 probe, this enzyme was expected to yield a normal 23 kb fragment due to the absence of the polymorphic Bam HI site located 7 kb 3' to the β-globin gene, and an abnormal fragment if the chromosome bearing the deletion also lacked the polymorphic Bam HI site.

Upon restriction of the patient DNA with Bam HI and Southern analysis using the probe pRK 28 located 17 kb downstream to the β-globin gene we obtained two bands (Figure 34), one 23 kb and one 17 kb in length. The 23 kb fragment corresponds to the normal chromosome as seen in control blots containing normal DNA. The 17 kb fragment also looked normal in that its size corresponded to the size of a 17 kb band in the control lane. As we had the affected father's DNA, we knew that the normal chromosome lacked the polymorphic Bam HI site and hence the 23 kb fragment had to have originated from this chromosome. Thus, the 17 kb fragment had to have originated from the affected chromosome.

To eliminate the trivial but real possibility that the deletion had occurred in such a manner as to bring in new DNA that contained a Bam HI site close to the original site so as to give a similar band, we decided to clone the 17 kb Bam HI fragment. We constructed a Bam HI EMBL 3 library using the conditions described in the methods section. Approximately
Figure 34: Autoradiogram of the DNA digests from the Scotch-Irish thalassemia and normal DNA hybridized to the pRK 28 probe.

Lanes a and b contain DNA from a normal individual, and lanes c and d contain DNA from the affected individual. Bands not marked, are from contaminating plasmid. Lanes a and c are Kpn I - Bam HI digests and lanes b and d are Bam HI digests.
FIGURE 34
100,000 recombinant plaques were screened using the pRK 28 probe, and a single hybridizing plaque was identified. The plaque was purified and the DNA isolated from the purified plaque. Restriction of the DNA indicated that a 17 kb Bam HI fragment had been cloned. The 17 kb Bam HI insert was then subcloned into pUC 19 plasmid vector. The plasmid bearing the 17 kb Bam HI insert was analyzed with a series of restriction enzymes that include Sma I, Xho I, Bam HI, Eco RI, Bgl II, and Hind III to restriction map the fragment (Figure 35). The enzymes chosen to map this fragment were selected with the idea of being able to use one of them if they were located so as to encompass the pRK 28 probe and also give a clonable abnormal restriction fragment. In other words, the sites needed to be located 3' to the pRK 28 probe and the given restriction enzyme should not have any more cleavage sites 5' to the probe. This would increase the probability that the next restriction site would lie within the new DNA brought in as a result of the deletion. Another requirement was that if such a restriction site was identified, it should also have to yield a fragment that was clonable. The restriction pattern obtained firstly indicated that the clone contained normal DNA. Secondly all the restriction sites were located in a manner that would not yield an abnormal band using the pRK 28 probe. Thus the 3' boundary of this deletion lies 5' to the polymorphic Bam HI site located 7 kb downstream of the β-globin gene.
Figure 35: Map of the 17 kb Bam HI fragment isolated from the library constructed using thal-3 DNA.

The restriction enzymes used were B: Bam HI; Bg: Bgl II; S: Sma I; E: Eco RI; H: Hind III. The location of the Bgl II - Eco RI fragment pRK 28 probe is shown by the broken bar.
FIGURE 35

2 Kb

B  Bg  S  E  E  H  H  B

Bg  Bg  Bg

prK28
Identification of unique DNA probes immediately 3' to the \( \beta \)-globin gene

The second approach involved the identification of unique DNA fragments localized between the 3' end of the \( \beta \)-globin gene and the approximate 3' boundary of this deletion. As no convenient restriction sites were available using the pRK 28 probe, the idea behind this approach was to obtain single copy probes that would again yield clonable abnormal fragments spanning the breakpoint junction. As shown in Figure 33, the 3' end of the \( \beta \)-globin gene is located adjacent to a series of repeat elements like the Alu I and Kpn I family sequences. To be able to identify unique sequences that could be used as a probe in the analysis, a 3.6 kb Eco RI fragment located within 3 kb 3' to the \( \beta \)-globin gene was digested with a series of restriction enzymes that included Alu I, Cfo I, Dde I, Hae III, Hinf I, Sau 3A, Tha I, and Taq I. Most of the enzymes were chosen because they have four or five nucleotide recognition sequences and by using this approach we had hoped to be able to identify small fragments within these regions which were unique. The 3.6 kb Eco RI fragment was analyzed since it borders the approximate location of the 3' breakpoint of the deletion. The restriction fragments were either run on a 2.5% agarose gel and blotted onto nitrocellulose using standard procedures described in the
"Material and Methods" section, or were fragment isolated after fractionation on a 6% acrylamide gel and then dot blotted onto nitrocellulose. The blots were then hybridized to randomly \(^{32}\text{P}\) labeled human genomic DNA. Fragments that were unique to the 3.6 kb Eco RI insert and that failed to hybridize to the labeled human genomic DNA were identified. Under the conditions used, only highly repetitive sequences hybridize to the labeled human DNA. The results of such an analysis of the 3.6 kb Eco RI fragment is shown in Figure 36, panels B and C. Fragments that failed to hybridize to human DNA were presumed to be either unique or of low copy number (fragments labeled numbers 3 and 4 in panel C). These fragments were isolated and subsequently labeled using the random primer extension procedure of Feinberg et al. 1983. The labeled fragments were then hybridized to restricted human genomic DNA that had been transferred onto nitrocellulose after fractionation on a 0.8% agarose gel. In all cases the labeled fragments did not hybridize to a single band, but instead hybridized extensively to the genomic DNA. Therefore these fragments must contain low to moderately repetitive sequences, despite the fact that initially they seemed to be low copy number using labeled genomic DNA. A similar approach was used to analyze a 1.6 kb Bgl II fragment that partly overlaps with the 3.6 kb Eco RI fragment but extends further 3'. This fragment was restricted with Dde I, Hae III, Hinf I, and Sau 3A. The results of such an analysis is shown in Figure 14, Panel A.
Figure 36: Autoradiograms of dot blots used to identify unique probes 3' to the β-globin gene.

Panel A contains various fragments isolated from a 1.6 kb Bgl II fragment located 3' to the β-globin gene that were probed with P32 labeled human genomic DNA. Spots 1 through 5 are Hinf I fragments, spots 6 through 8 are Hae III fragments, and spots 9 through 12 are Sau 3A fragments. Panel B and C contain various fragments from a 3.2 kb Eco RI fragment also located 3' to the β-globin gene as described in the text. Spots 1 through 6 are Dde I fragments, spots 7, 8 are Hae III fragments, spots 9 through 14 are Hinf I fragments and spot 15 is the total 3.2 Eco RI fragment. The fragments were spotted in duplicate in panels B and C.
Fragments that were thought to be either low copy or unique were fragments labeled numbers 3, 7, 9, and 10 in panel A of Figure 36. As before all fragments that seemed to be low copy number using labeled genomic DNA, extensively hybridized to genomic DNA when labeled themselves. These results were very much in keeping with the highly repetitive nature of the DNA found 3' to the β-globin gene.

Isolation of unique probes located upstream of the 5' breakpoint of the deletion using a jumping library

Since we were unable to identify unique probes 3' to the β-globin gene, we decided to try and isolate unique probes sufficiently 5' to the β-globin gene cluster that they would lie upstream of the 5' breakpoint of the γδβ thal-3 deletion. To be able to do this, we decided to screen a jumping library using available probes 5' to the β-globin gene cluster, and thus be able to "jump" 80 kb to 130 kb. Once these clones were isolated and used as probes against blots of thal-3 DNA we expected to get two possible results. First, we could get bands that were identical to those on control blots, in which case the results would indicate that the probes were located in a region that had not been deleted in thal-3. A second possibility was that abnormal bands were obtained as compared to blots containing normal DNA, in which case the probe was hybridizing to a fragment
that spanned the breakpoint or was the result of the presence of a polymorphic restriction site. In the event that no band was obtained using a given probe, it would indicate that this region had also been deleted in thal-3. The way in which one "jumps" as opposed to "walks" is best shown in Figure 37, and involves using a given probe and isolating DNA that is contiguous with the probe but is located anywhere from 80 kb to 100 kb 5' to the probe, lacking all the intervening DNA. A jumping library constructed by Dr. Francis Collins and his colleagues was obtained. The jumping library was constructed by restricting high molecular weight genomic DNA partially with Mbo I such that most of the DNA fragments were in the size range of approximately 100 kb. The DNA was then ligated in the presence of a selectable marker (E. coli sup F gene) under conditions that favour the formation of large circular pieces of DNA. These circular DNAs were then restricted with Eco RI and the resulting linear pieces of DNA ligated to Charon 16A lambda arms. The ligated DNAs were then packaged into phage particles and propagated on a sup F- E. coli strain. Thus only those arms that had ligated to an insert containing a selectable marker sup F gene would be propagated as viable phage particles in the library. Present in the sup F gene is an Ava I site that usually separates the two halves of the jump clones enabling one to identify the "jump" fragment provided there are no other Ava I sites within the jump clone. The library was propagated on MC1061 bacteria and plated out onto
Figure 37: Schematic diagram of the construction of the partial Mbo I jumping library.

The closed boxes represent the probe used to isolate the jump clone, whilst the open boxes represent the jump fragment. The jagged line represent the selectable marker (Sup F gene). The closed box represents the probe used to isolate a given jump clone and the open box the location of the jump fragment with respect to the probe. (adapted from Collins et al, 1984).
Very high $M_r$ CNA

Partial Mbo I digest
Selection of desired size range

Low concentration ligation
in presence of excess sub F

Eco RI digestion

Ligate to lambda arms (e.g. λCh16A)
Package, plate on sub- host

Library of junction fragments

Screen with

Junction fragment clone

FIGURE 37
large plates (25X25 cm) as described in the "Materials and Methods" section. The bacterial strain MC1061 only permits those phage that contain the bacterial suppressor tRNA gene SupF to grow. The library was then screened using standard procedures except that pBR 328 plasmid DNA was included in the hybridization at a concentration of 10 µg/ml to prevent hybridization of trace quantities of plasmid sequences in the labeled probe to the pBR 328 sequences inadvertently cloned into the library during its construction (present at a frequency of 1/1000 clones). Three probes were used to screen the library. The first was the 0.9 kb Bam HI - Eco RI (βIVS II) fragment so as to identify sequences 100 to 130 kb 5' to the β-globin gene. These sequences would be located just 5' to the two other probes used to screen the library and would serve to link any probes thus isolated to the β-globin region. The other two probes used were a 3.3 kb Eco RI fragment and a 4.0 kb Eco RI fragment located 20 kb and 55 kb 5' to the ε-globin gene, respectively.

Using the 0.9 kb Bam HI - Eco RI probe, four clones were isolated and plaque purified. The clones were labeled H1, C2, C4, and C5. Of these four clones, C2, C4, and C5 were found to be identical based on the restriction patterns obtained with Eco RI, Ava I and Eco RI - Ava I double digests. They all contained two Eco RI inserts of 2 kb and 2.2 kb in size. Clone H1 also showed two Eco RI inserts but these were found to be 5.5 kb and 4.0 kb in length. Restriction of these clones with Eco RI and
hybridization to the the 0.9 Bam HI -Eco RI probe, yielded the following results. The probe hybridized to the 5.5 kb fragment in the H1 clone. In the case of the C2, C4, and C5 clones, the resolution of the 2.0 kb and 2.2 kb fragments on a 0.8% agarose gel was insufficient to permit identification of the hybridizing band. Upon hybridization of the Eco RI and Eco RI - Ava I restricted DNA of these clones with labeled human DNA the inserts hybridized to the probe, indicating that the jump fragment contained highly repetitive DNA. Hence these clones were not analyzed any further as they were of limited value in the analysis of the large deletion thalassemia.

Using the 3.3 kb Eco RI probe one clone designated BB1 was isolated, whilst using the 4.0 Eco RI probe, two clones, designated CC1 and DD1, were isolated. These clones were plaque purified and the DNA restricted with Eco RI to release the cloned insert DNA. The restricted DNA was transferred onto nitrocellulose and hybridized to each of the two probes as shown in Figure 38. Clone BB1 contained a 8 kb Eco RI insert and a 2 kb insert of which the 3.3 kb Eco RI probe hybridized to the 8 kb Eco RI fragment. Clones CC1 and DD1 showed identical sized inserts and identical restriction patterns when restricted with Eco RI, Ava I, and Eco RI + Ava I. They were assumed to be identical clones based on these criteria. These two clones contained a single insert of 7.5 kb in size which hybridized to the 4.0 kb Eco RI probe in both clones. Clones BB1, CC1 and DD1
Figure 38: Autoradiogram of DNA isolated from the jumping clones, hybridized to the 3.3 Eco RI and 4.0 Eco RI probes.

The left panel was probed with the 3.3 Eco RI fragment and the right panel was probed with the 4.0 Eco RI fragment. Lane 1 contains DNA from clone DD1, lane 2 contains DNA from clone CC1 and lane 3 contains DNA from clone BB1 restricted with Eco RI. The numbers on the side indicate the size of the fragment hybridizing in kb.
FIGURE 38
were restricted with Eco RI, Ava I, and Eco RI + Ava I, fractionated on a 0.8% agarose gel and transferred to nitrocellulose. The blots were then hybridized to P-32 labeled human DNA to identify repetitive sequences present in the isolated clones. The results of the hybridizations (Figure 39) indicated that clone BB1 did not contain highly repetitive sequences. On the other hand, both the Eco RI inserts of clones CC1 and DD1 hybridized to the labeled human genomic DNA indicating that the jump fragment contained highly repetitive DNA. Clone BB1 was selected for further analysis. The 8 kb Eco RI insert was subcloned into pUC19 and amplified. The jump fragment was identified using restriction analysis (Figure 40) and was found to be a 1.8 kb Eco RI - Ava I fragment. Figure 41 shows the map of the 8.0 kb Eco RI fragment isolated from the clone BB1. The jump fragment had to be further characterized to show that it was contiguous with the probe used to isolate it i.e. it was from chromosome 11. This was necessary because it was possible that during the construction of the library, fragments that were not necessarily contiguous with the probe could have been ligated to the selectable marker and then to the fragment containing the probe. To test this possibility, the jump fragment was P-32 labeled and hybridized to blots containing restricted total human genomic DNA, total mouse genomic DNA and DNA from a hybrid cell line (Hu 11) containing Human chromosome 11 translocated on to the mouse X chromosome (Zavodny
Figure 39: Autoradiogram of jumping clones hybridized to P32 labeled human genomic DNA.

Lanes 1, 2 and 3 contain DNA from clones BB1, CC1, and DD1 respectively, restricted with the different enzymes. Restriction enzymes used were E: Eco RI; and A: Ava I.
PROBE

Human genomic DNA

<table>
<thead>
<tr>
<th>E</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A−E</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

FIGURE 39
Figure 40: Photographs of 0.8% agarose gels with restriction digests of the phage clone BB1 and the plasmid clone containing the 8.0 Eco RI fragment.

Panel A: lanes 1 and 2, contain DNA from the phage clone, lanes 3, 4, and 5 contain DNA from the plasmid bearing the subcloned 8.0 Eco RI insert and lane 6 contains sized lambda DNA fragments. Lanes 2 and 5 are Eco RI digests, lanes 1 and 3 are Eco RI - Ava I digests, and lane 4 is an Ava I digest.

Panel B: lanes 2, 3, and 4 contain DNA from the plasmid containing the 8.0 Eco RI fragment. Lane 1 contains sized lambda DNA fragments. Lane 2 is a Eco RI digest, lane 3 is an Ava I digest and lane 4 is an Eco RI - Ava I digest.

The Roman numeral I indicates the 8.0 Eco RI fragment. II indicates the plasmid pUC 19. III indicates the "jump" fragment. IV indicates the sup F gene fragments.
Figure 41: Map of the jump clone BB1 isolated with the 3.3 Eco RI probe.

The restriction enzymes shown are E: Eco RI; and A: Ava I. The broken bar indicates multiple copies of the sup F gene in the clone.
1 Kb

Jump \([\text{Sup F}]_n\) 3-3 Probe

FIGURE 41
et al. 1983). In this cell line the only human DNA present is from chromosome 11. Unfortunately the jump probe repeatedly failed to hybridize to either the human or the hybrid cell line DNA. The identity of this jump fragment is not known but probably is not of human origin. The library was screened two more times plating out $1 \times 10^6$ phage each time and screened with the same probes. A single clone identical to CC1 and DD1 was obtained. Since these clones contained highly repetitive DNA, no attempt was made to further analyze them.

Our attempt to identify a suitable restriction enzyme that would yield an abnormal clonable fragment using the pRK 23 probe was unsuccessful due to the unsuitable location of all the restriction sites mapped in the 3' region of the $\beta$-globin gene. To circumvent this problem, we attempted to isolate single copy probes in the immediate vicinity of the 3' breakpoint of the $\gamma\delta\beta$ thal-3 deletion. The nature of the DNA surrounding the 3' breakpoint of the deletion (highly repetitive) resulted in our not being able to accomplish this task. Attempts to isolate single copy probes, upstream to the 5' breakpoint of the deletion using the jumping library approach was also unsuccessful due to the fact that the jump probes isolated also contained highly repetitive DNA. Thus the molecular analysis of the $\gamma\delta\beta$ thal-3 will have to await the availability of other jumping libraries such that appropriate single copy probes can be isolated.
APPENDIX A

TABLE 6

Summary of the Maxam and Gilbert base specific reactions for sequencing and labeled DNA used in this thesis

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G + A</th>
<th>G + T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ul DMS buffer</td>
<td>12 ul H2O</td>
<td>12 ul H2O</td>
<td>15 ul 5M NaCl</td>
<td></td>
</tr>
<tr>
<td>3 ul carrier DNA</td>
<td>3 ul carrier DNA</td>
<td>3 ul carrier DNA</td>
<td>3 ul carrier DNA</td>
<td></td>
</tr>
<tr>
<td>4 ul 32-P DNA</td>
<td>6 ul 32-P DNA</td>
<td>6 ul 32-P DNA</td>
<td>4 ul 32-P DNA</td>
<td></td>
</tr>
<tr>
<td>1 ul DMS</td>
<td>2 ul Pip. Pcr.</td>
<td>20 Hydrat.</td>
<td>30 ul Hydrat.</td>
<td></td>
</tr>
<tr>
<td>50 ul DMS Stop</td>
<td>Freeze</td>
<td>200 ul H2O Stop</td>
<td>200 ul H2O Stop</td>
<td></td>
</tr>
<tr>
<td>750 ul Ethanol</td>
<td>Lyophilize</td>
<td>750 ul Ethanol</td>
<td>750 ul Ethanol</td>
<td></td>
</tr>
<tr>
<td>Chill -70°C, 5 min</td>
<td>Chill -70°C, 5 min</td>
<td>Chill -70°C, 5 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrifuge, 5 min</td>
<td>Centrifuge, 5 min</td>
<td>Centrifuge, 5 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250 ul 0.2M NaAc</td>
<td>250 ul 0.2M NaAc</td>
<td>250 ul 0.2M NaAc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>750 ul Ethanol</td>
<td>750 ul Ethanol</td>
<td>750 ul Ethanol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chill -70°C, 5 min</td>
<td>Chill -70°C, 5 min</td>
<td>Chill -70°C, 5 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrifuge, 5 min</td>
<td>Centrifuge, 5 min</td>
<td>Centrifuge, 5 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process Description</td>
<td>Process Description</td>
<td>Process Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethanol rinse</td>
<td>Ethanol rinse</td>
<td>Ethanol rinse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 ul H₂O</td>
<td>20 ul H₂O</td>
<td>20 ul H₂O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 ul, 1.0 M</td>
<td>100 ul, 1.0 M</td>
<td>100 ul, 1.0 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90°C, 30 min</td>
<td>90°C, 30 min</td>
<td>90°C, 30 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 ul H₂O</td>
<td>20 ul H₂O</td>
<td>20 ul H₂O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td>Vacuum dry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loading dye</td>
<td>Loading dye</td>
<td>Loading dye</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: DMS: Dimethylsulfate; HZ: Hydrazine; Pip.For: Piperidinium Formats

Note: For a more detailed procedure refer to Maxam and Gilbert (1980)
<table>
<thead>
<tr>
<th>Fragment size (bp)</th>
<th>G</th>
<th>A</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 250</td>
<td>6</td>
<td>25</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>250 - 350</td>
<td>3 - 4</td>
<td>20</td>
<td>3 - 4</td>
<td>3 - 4</td>
</tr>
<tr>
<td>350 - 450</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>450 - 650</td>
<td>9 - 10</td>
<td>8 - 10</td>
<td>9 - 10</td>
<td>9 - 10</td>
</tr>
<tr>
<td>&gt; 650</td>
<td>6 - 8</td>
<td>8</td>
<td>6 - 8</td>
<td>6 - 8</td>
</tr>
</tbody>
</table>

* These reactions are carried out on ice for the stated times.

The A reactions were carried out at 37°C instead of 20°C as per the Maxam and Gilbert protocol.
LIST OF REFERENCES


