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.
Nonhomologous recombination in the human genome: Deletions and insertions in the human factor VIII gene

Woods-Samuels, Patricia, Ph.D.
The Ohio State University, 1989
NONHOMOLOGOUS RECOMBINATION IN THE HUMAN GENOME:
DELETIONS AND INSERTIONS IN THE
HUMAN FACTOR VIII GENE

DISSertation

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

By

Patricia Woods-Samuels, B.S.

* * * * *

The Ohio State University
1989

Dissertation Committee:
G.A. Marzluf
A.H. Burghes
C.A. Breitenberger

Approved by

Adviser
Department of
Biochemistry
DEDICATION

In loving memory of Harry Woods

and James T. Byrne
ACKNOWLEDGEMENTS

I am deeply indebted to Haig Kazazian, Jr. and Stylianos Antonarakis of the Johns Hopkins University for giving me the opportunity to complete my thesis research in their laboratories. I am also forever indebted to Elio F. Vanin of the Ohio State University Biochemistry Department for allowing me to continue as his student in absentia. I would also like to thank all those who have helped me in this endeavor, especially Alan Scott for many enjoyable discussions about my work and science in general. I am grateful to Emily Pasterfield, Pamela Stevens, and especially Diane Darchicourt for their expert assistance in preparation of this manuscript. Special thanks go to Frank and Elizabeth Woods, without whom none of this would be possible and to my husband Brian, the best friend and most enthusiastic supporter a student ever had.
VITA

March 19, 1957

Born - Schenectady, New York

1979

B.S. Chemistry John Carroll University; Cleveland, Ohio

1979 - 1982

Project Associate, Plant Molecular Biology; The Standard Oil Co. of Ohio; Cleveland, Ohio

1982 - Present

Graduate Student, Department of Biochemistry; Ohio State University; Columbus, Ohio

1985 - Present

Researcher, Pediatric Genetics Unit; Johns Hopkins University Medical School; Baltimore, Maryland

Publications


Fields of Study

Major Field: Biochemistry
# Table of Contents

DEDICATION.................................................................................. ii
ACKNOWLEDGMENTS.................................................................. iii
VITA......................................................................................... iv
LIST OF TABLES.......................................................................... vii
LIST OF FIGURES......................................................................... viii
PREFACE.................................................................................... x

## PART 1: MOLECULAR CHARACTERIZATION OF FOUR DELETIONS IN THE HUMAN FACTOR VIII GENE

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction.................................................................2</td>
<td></td>
</tr>
<tr>
<td>A. Juxtaposition............................................................... 3</td>
<td></td>
</tr>
<tr>
<td>B. Breakage and Reunion: End-joining............................... 19</td>
<td></td>
</tr>
<tr>
<td>C. Breakage and Reunion: Enzymes.................................... 25</td>
<td></td>
</tr>
<tr>
<td>D. Research Objective....................................................... 30</td>
<td></td>
</tr>
<tr>
<td>II. Materials and Methods....................................................34</td>
<td></td>
</tr>
<tr>
<td>A. Restriction Endonuclease Analysis..................................34</td>
<td></td>
</tr>
<tr>
<td>B. Genomic Cloning............................................................. 36</td>
<td></td>
</tr>
<tr>
<td>C. Sequence Analysis..........................................................37</td>
<td></td>
</tr>
<tr>
<td>III. Results............................................................................. 38</td>
<td></td>
</tr>
<tr>
<td>A. JH1 Deletion................................................................. 38</td>
<td></td>
</tr>
<tr>
<td>B. JH7 Deletion................................................................. 46</td>
<td></td>
</tr>
<tr>
<td>C. JH37 Deletion............................................................... 53</td>
<td></td>
</tr>
<tr>
<td>D. JH21 Deletion............................................................... 58</td>
<td></td>
</tr>
<tr>
<td>IV. Discussion......................................................................... 68</td>
<td></td>
</tr>
</tbody>
</table>

LIST OF REFERENCES FOR PART 1..................................................82
## PART 2: MOLECULAR CHARACTERIZATION OF L1 INSERTIONS IN THE HUMAN FACTOR VIII GENE

### CHAPTERS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. Introduction</td>
<td>91</td>
</tr>
<tr>
<td>VI. Materials and Methods</td>
<td>104</td>
</tr>
<tr>
<td>A. Restriction Endonuclease Analysis</td>
<td>104</td>
</tr>
<tr>
<td>B. Cloning and Sequence Analysis</td>
<td>105</td>
</tr>
<tr>
<td>VII. Results</td>
<td>106</td>
</tr>
<tr>
<td>A. Restriction Endonuclease Analysis</td>
<td>106</td>
</tr>
<tr>
<td>B. Nucleotide Sequence Analysis</td>
<td>109</td>
</tr>
<tr>
<td>VIII. Discussion</td>
<td>119</td>
</tr>
<tr>
<td>LIST OF REFERENCES FOR PART 2</td>
<td>123</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Altered restriction fragments in factor VIII deletions</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Summary of twenty-eight deletions in the human genome</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES PART 1</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Recombination breakpoints involving Alu I elements</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Slipped-mispairing model of deletion formation</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Chromatin loop model</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Replication loop model of deletion formation</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Breakage and reunion model</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Paired priming model of end-joining</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Gyrase subunit exchange model</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>JH deletions in the factor VIII gene</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Altered restriction fragments in patient JH1</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Restriction map of JH1 deletion breakpoint</td>
</tr>
<tr>
<td>Figure 1.11</td>
<td>JH1 deletion breakpoint sequence</td>
</tr>
<tr>
<td>Figure 1.12</td>
<td>Altered restriction fragments in patients JH7 and JH37</td>
</tr>
<tr>
<td>Figure 1.13</td>
<td>Restriction map of JH7 deletion breakpoint</td>
</tr>
<tr>
<td>Figure 1.14</td>
<td>JH7 deletion breakpoint sequence</td>
</tr>
<tr>
<td>Figure 1.15</td>
<td>Restriction map of JH37 deletion breakpoint</td>
</tr>
<tr>
<td>Figure 1.16</td>
<td>JH37 deletion breakpoint sequence</td>
</tr>
<tr>
<td>Figure 1.17</td>
<td>Screening of JH deletion patients with a 4.7 kb Eco RI fragment containing exons 2 and 3</td>
</tr>
<tr>
<td>Figure 1.18</td>
<td>Identification of the normal 5' Eco RI fragment involved in the JH 21 deletion</td>
</tr>
<tr>
<td>Figure 1.19</td>
<td>Restriction map of JH21 deletion breakpoint</td>
</tr>
<tr>
<td>Figure 1.20</td>
<td>JH21 deletion breakpoint sequence</td>
</tr>
<tr>
<td>Figure 1.21</td>
<td>Paired-priming mechanism for the presence of direct repeats at the JH deletion junctions</td>
</tr>
<tr>
<td>Figure 1.22</td>
<td>Paired-priming mechanism for the generation of duplicated nucleotides at the JH 21 deletion junction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FIGURES PART 2</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Schematic diagrams of Class I and II retrotransposons</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Diagram of two L1 insertions in exon 14 of the human factor VIII gene</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Model for Ty transposition mechanism</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Family JH-25 with an L1 insertion in IVS-10 of the human factor VIII gene</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Diagram of the L1 insertion in IVS-10 of the human factor VIII gene</td>
</tr>
<tr>
<td>Figure 2.6A</td>
<td>Nucleotide lengths of the JH11 elements</td>
</tr>
<tr>
<td>Figure 2.6B</td>
<td>Nucleotide and amino acid sequence comparison of the JH11 elements</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Comparisons of the JH11 sequences to the L1HS consensus sequence</td>
</tr>
</tbody>
</table>
Change is the basis of life. It is therefore not surprising that the DNA molecules which encode the processes of life are in a constant state of change. Alterations in a DNA molecule, can be as small as a single base pair change or as gross as a chromosomal translocation.

Other types of rearrangements include deletions, inversions, and insertions which most often occur by nonhomologous recombination and seem to occur as by-products of the cell's normal processes. Since the ultimate form of therapy for many diseases may be the repair of the mutated gene through homologous recombination with the normal gene, it is important to understand recombination processes which compete with homologous recombination. The rearrangements that occur within the human factor VIII gene, identified because they cause hemophilia A, are the subject of this research. Chapter 1 discusses the possible mechanisms involved in four nonhomologous deletions which have occurred in this gene. The location and sequences of three insertions of L1 repetitive elements are presented in Chapter 2 and discussed as evidence for retrotransposition in the human genome.
PART 1

MOLECULAR CHARACTERIZATION OF FOUR DELETIONS

IN THE HUMAN FACTOR VIII GENE
Deletions are a common type of mutation in the genomes of all organisms. The recombination mechanisms by which these events occur are not well understood and therefore warrant investigation. Initially it was postulated that deletions in eukaryotic genomes were the result of unequal crossing-over between homologous chromosomes (Vogel and Motulsky, 1986). If this were indeed the case one would expect the number of deletions to be approximately equal to the number of duplications, as both would use unequal homologous crossover. A number of chromosomal rearrangements in various gene systems have now been reported and it is obvious that the deletions far outnumber the duplications. In fact, the majority of deletions which have been studied at the sequence level have been found to be the result of nonhomologous breakage and reunion. Of those studied to date, 79% appear to be the result of a nonhomologous breakage and reunion event while 21% seem to be due to unequal homologous crossover (Table 1.2).

Recombination in eukaryotic systems is at present poorly understood. Therefore, in an attempt to better understand this process, we decided to study four different deletions in the
human factor VIII gene. Before presenting the results we obtained from these four deletions I would like to discuss the different mechanisms which have been postulated to produce deletions in the mammalian genome. In discussing recombination, either unequal homologous or nonhomologous, one can divide the process into three steps, 1) juxtaposition of noncontiguous sequences, 2) cleavage of double-stranded DNA and 3) rejoining of the broken ends with subsequent loss of the DNA between the deletion endpoints. The last two steps, known as breakage and reunion, could also proceed by a one-step concerted mechanism if catalyzed by one enzyme, one complex of enzymes or a number of different enzymes. Hopefully, the results of our work will help shed light on the mechanism responsible for the production of deletions. In the following paragraphs I will discuss the various models which have been proposed for each of the three steps outlined above.

A. Juxtaposition

Unequal crossing-over

The classic mechanism for the generation of tandem duplications and deletions is unequal crossing-over between misaligned homologous chromosomes during meiosis (Vogel and Motulsky, 1986). Deletions in the X-linked Duchenne muscular dystrophy (DMD) (Monaco et al., 1987; Bakker et al., 1987; Darras
and Francke, 1987; Lanman et al., 1987; Wood and McGillivray, 1988) and factor VIII genes (Youssoufian et al., 1988 and 1987) have been shown to be the result of germ line mosaicism, demonstrating a mitotic origin for these mutations. Unequal crossing-over between sister chromatids or an intrachromosomal looping-out could be postulated as the mechanism of formation of these deletions. Recently, Hu et al. (1989) has shown evidence that unequal crossing-over between sister chromatids during mitosis is the mechanism behind the generation of a tandem duplication in the DMD locus.

Although unequal crossing over is an attractive model for the generation of duplication/deletion mutations, the observation of many more deletions than duplications implies that deletions are occurring by more than one mechanism since unequal crossover should generate equal numbers of duplications and deletions. This discrepancy might be due to a bias of ascertainment, at least in mutations analyzed in genes responsible for human disease, if duplications are generally less deleterious in nature than deletions.

Recombination and Alu repetitive elements

It has been suggested that repetitive elements may facilitate the misalignment of chromosomes prior to an unequal crossing-over event (Weatherall and Clegg, 1979; Jeffreys and Harris, 1982; Vogel and Motulsky, 1986). Candidates in the
human genome for such a role are the Alu 1 elements (Table 1.2).

The Alu family of repetitive elements consists of approximately 500,000 members per human genome. The 300bp primate sequence is composed of two imperfect monomers which are tandemly repeated (Schmid and Jelinek, 1982; Deininger et al, 1981). The left hand monomer contains a putative bipartite control region for RNA polymerase III transcription (Paolella et al. 1983) (Figure 1.1 from Lehman et al., 1987).

Lehrman et al. (1987a) have identified a tandem duplication of seven exons in the low density lipoprotein (LDL) gene which apparently was the result of crossing-over between misaligned Alu repetitive elements. Recombination between Alu elements has also been observed for four deletions in the LDL receptor gene (although one was of opposite orientation) (Lehman et al., 1987b; Lehman et al., 1987b; Hobbs et al., 1986; Lehman et al., 1985), a deletion in the α globin locus (Nicholls et al., 1987), a deletion in the adenosine deaminase gene (Markert et al., 1988) and in an X-Y chromosome translocation (Rouyer et al., 1987) (Figure 1.1). The presence of Alu sequences on one side of a deletion junction has been noted for deletions in the α (Nicholls et al., 1987) and β globin clusters (Jagadeeswaran et al., 1982; Ottolenghi et al., 1982; Vanin et al., 1983; Henthorn et al., 1986) as well as the β hexosaminidase α chain gene (Myerowitz et al., 1987), the
Figure 1.1. Recombination breakpoints involving Alu I elements. A consensus Alu I repeat is shown schematically as two directly repeated monomers ending in a A-rich sequence (A\textsubscript{n}). The numbers below the arrow which indicates the 5'-3' direction of the repeat element refer to the consensus of Deininger et al. (1981). Boxes A and B represent the putative RNA polymerase III promoter. The numbers above the arrow indicate the positions of rearrangement breakpoints which have been associated with Alu sequences. The references are as follows: beta globin (1 Ottolenghi et al., 1982; 3 Vanin et al., 1983; 11 Jagadeeswaran et al., 1982; 13 Henthorn et al., 1986), LDL receptor rearrangements (2 Lehrman et al., 1986; 4 and 5 Lehrman et al., 1987a; 6 and 7 Hobbs et al., 1986; 9 and 10 Lehrman et al., 1987b; 8 and 12 Lehrman et al., 1985), alpha globin locus (Nicholls et al., 1987), adenosine deaminase gene (Markert et al., 1988), \(\beta\) hexosaminidase \(\alpha\) chain gene (Myerowitz et al., 1987), X-Y chromosome translocation Rouyer et al., 1987), the factor IX gene (Green et al., 1988), and the factor VIII gene (this paper).
Figure 1.1
factor IX gene (Green et al., 1988) and the factor VIII gene (this paper and Figure 1.1).

Not only do many of the deletions which have been reported involve an Alu family sequence in one way or another, but Lehrman et al. (1987a) also observed that the breakpoints within the Alu sequences were clustered. They also found that one of the breakpoint clusters resided between the putative control regions for RNA polymerase III transcription (Paolella et al., 1983) (boxes A and B Figure 1.1). This observation led to the suggestion that at least some recombination events might be the result of an unusual configuration adopted by Alu sequences during their transcription (Lehrman et al., 1987a).

The Alu model of "homologous" recombination is considered in this discussion of nonhomologous mechanisms because the homology is neither extensive nor exact (Alu family members are about 80% homologous). Also, the number of rearrangements involving only one Alu element suggests that these sequences might serve as a site for nonhomologous breakage and reunion mediated by cellular enzymes in addition to their apparent role in misalignment otherwise nonhomologous stretches of DNA.

Slipped-mispairing of direct repeats

An alternative mechanism for the production of deletions has been termed "slipped-mispairing". This mechanism was first proposed by Streisinger et al. (1966) to explain the formation
of frameshift mutations in the bacteriophage T4. Farabough et al. (1978) has suggested that this mechanism might also be responsible for producing spontaneous deletions in the lac I gene in *E. coli*. They found that spontaneous deletions in this gene occurred predominantly between direct repeats of 5 to 8 bp. Direct repeats of similar size were also present at the junctions of larger (700-1000 bp) deletions in lac I-Z fusion strains of *E. coli* (Albertini et al., 1982). Although occurring more frequently on a RecA+ background, these deletions also take place in RecA- strains. These events, therefore, are nonhomologous since the product of the RecA+ gene is essential for homologous recombination in *E. coli*. Albertini et al. (1982) hypothesized two types of mechanisms based upon these results. One model involves the mispairing of direct repeats in single-stranded (ss) regions generated during replication, the other is an enzymatically mediated double-strand (ds) breakage and reunion model which will be considered later.

Brunier et al. (1988) has pointed out that the key step in the slipped-mispairing model of deletion formation is the switching of DNA polymerase from one repeat to another on a single-stranded template. Template switching is also the key step in copy-choice recombination (Lederberg, 1955). Since the slipped-mispairing or copy-choice mechanism depends on polymerase template switching in a region of ssDNA, Brunier et al. (1988) propagated the *E. coli* transposable element Tn 10 on
a plasmid which could be induced to produce ssDNA in order to test this part of the model. Conversion of the ds plasmid to ss plasmid stimulated the transposon excision thereby supporting template switching as a necessary step in at least some recombination mechanisms.

Efstradiatis et al. (1980) has also implicated this mechanism in the generation of <50 bp deletions in the β-globin locus. As can be seen in Figure 1.2 (Efstradiatis et al., 1980), a direct repeat (R2) within a single stranded region of DNA at a replication fork mispairs with an upstream, homologous repeat which has "slipped" from its normal position (R1) forming a single-stranded loop. The loop is excised and a subsequent round of replication yields a deleted daughter duplex with one copy of the repeat and one normal daughter duplex with two copies of the repeat. The initial mispairing might be assisted by the presence of stem and loop structures which would bring the repeats into close proximity (Albertini et al. 1982; Glickman and Ripley, 1984). The slipped-mispairing replication model for small deletions provides for all three steps involved in deletion formation. However, the breakage and reunion steps in this model are performed by DNA polymerase in a concerted manner that does not actually break the DNA molecule.
Figure 1.2. Slipped mispairing model for deletion formation. A) A region of duplex DNA is shown containing direct repeat sequences (R1 and R2 on one strand R1' and R2' on the other strand). B) As a replication fork moves through this region, the duplex containing the repeats becomes single-stranded. C) The mispairing of the R2 repeat with the complimentary R1' repeat produces a single-stranded loop. D) Excision of the loop by repair enzymes and subsequent replication would generate (E) a normal daughter duplex (to the right) and a duplex containing only one of two original repeats (to the left). Efstradiatis et al., 1980.
Figure 1.2
Models of loop formation

An issue which is not addressed in the slipped-mispairing model of recombination is the juxtaposition of regions of DNA which are linearly far apart (<1kb->100kb) on a mammalian chromosome. Vanin et al. (1983) hypothesized that the 5' and 3' endpoints of two large (>100kb) deletions of the β-globin locus were brought in close physical proximity because of their anchorage to the nuclear matrix. The two 5' endpoints were approximately the same distance apart and in the same order on the chromosome as the 3' endpoints, leading to the speculation that the deletions occurred as the DNA was moving through the matrix attachment site during replication (Figure 1.3). Breakage at a replication fork associated with one anchorage point with subsequent reunion to a replication fork at another anchorage point would then result in the loss of one or more loops. This model is supported by several lines of circumstantial evidence; 1) chromatin is organized into loops of 50-100kb which are attached to a nuclear scaffold or matrix (Benyajati and Worcel, 1976; Paulson and Laemmli, 1977), 2) DNA sequences which associate with the nuclear matrix have been identified (Cockerill and Garrard, 1986; Gasser and Laemmli, 1986), and 3) replication forks are apparently associated with the matrix (Vogelstein et al., 1980).

Since DNA replication has also been associated with the nuclear matrix (Vogelstein et al., 1980), Anand et al. (1988)
Figure 1.3. Chromatin loop model. This is a hypothetical mechanism for the loss of chromatin loops at different stages of replication. Letters ABC etc. indicate sites along the DNA molecule represented by the black line, letters A',A'',B',B'',C',C'' etc. indicate the same sites after one round of replication. Breakage and reunion events, at the sites indicated by asterisks, form the deletions represented at the right. The ellipses represent anchorage points and their associated replication complexes (figure from Vanin et al., 1983).
Figure 1.3
has suggested that this mechanism could also account for the
generation of smaller β-globin deletions (Figure 1.4 from Anand
et al., 1988). These smaller deletions would arise from the
loss of a newly replicated loop of DNA with the size of the
deletion determined by the size of the replication loop at the
time of recombination (Figure 1.4). The loop models for
juxtaposition of nonhomologous stretches of DNA are of general
significance because varying sizes of deletions can be
accounted for by these models. Moreover, the prediction that
replication origins and matrix associating regions might be
involved in nonhomologous recombination provides a hypothesis
which can be tested.

The chromatin and replication loop models for
juxtaposition are similar to the slipped-mispairing model in
that all three models hypothesize that replication is involved
in the generation of deletions. The loop models do not suggest
the means by which the recombination is accomplished whereas
the slipped mispairing model proposes that DNA polymerase
itself is the mediator of recombination. One could speculate
that the occurrence of polymerase template switching between
replication forks juxtaposed by their association with the
matrix might generate the larger deletions frequently seen in
eukaryotes.
Figure 1.4. Replication loop model of deletion formation. DNA before and after replication is represented by a solid line with letters designating sites along the molecule. Sites after one round of replication are indicated as A', A'', B', B'' etc.. The dashed line represents DNA which is passing through an attachment site, indicated by the large open circle. The small closed circles represent DNA polymerase. Loops that are denoted with an asterisk are lost to give rise to the deletion indicated by brackets in the solid line (DNA molecule) to the right (Anand et al., 1988).
Figure 1.4
B. Breakage and reunion: End-joining

To generate a deletion, after two regions of a DNA molecule or molecules have been brought close together, the molecules must be recombined in a manner which results in the loss of the intervening DNA. This process is called breakage and reunion.

In models for breakage and reunion, an initiating double-strand break is followed either by religation or by resection of the broken ends exposing short, homologous repeats which anneal and are subsequently repaired (Figure 1.5). This model has been proposed for lac I deletions which do not have repeats at their junctions (Glickman and Ripley, 1984) and those which do have direct repeats (Albertini et al., 1982). Meselson and Weigle (1961) set out to test which mechanism, breakage and reunion or copy-choice, was responsible for recombinations between bacteriophage. They analyzed recombinants for the presence of parental DNA indicating a break and join mechanism, or the absence of parental DNA, indicating a template-switching mechanism. Their results indicated that the majority of the recombination events were due to breakage and reunion but other mechanisms could not be ruled out.

Recombination junctions between short, direct repeats are also common in eukaryotic systems. Roth et al. (1985) examined 110 junctions, including those generated by transfection, variants of SV 40, SV 40-adenovirus hybrids, integrations and
Figure 1.5. Breakage and reunion model. The solid, parallel lines in this diagram represent a double-stranded DNA molecule which has been broken mechanically, chemically or enzymatically. The arrows overlying the solid lines represent direct repeats within the DNA sequence which are revealed by resection due to nuclease activity. These repeats can then anneal and repair synthesis fills in remaining gaps. In the case of a blunt-end ligation, the steps diagrammed for the pairing of repeats are not necessary, although end-filling or removal of single-stranded overhangs would be necessary if the initiating break were staggered (figure from Brunier et al., 1988).
Figure 1.5
excisions of viral genomes from chromosomes and chromosome
translocation breakpoints. The analysis revealed that, although
40% of the breakpoints showed no homology whatsoever,
junctional homologies of 2 to 5 bp were more common than was to
be expected by chance. Spontaneous deletions in hamster cells
also occur between short repeats of 2-7 bp (Nalbantoglu et al.
1986) as do many spontaneous deletions in humans (see
references in Table 1.2).

Using linearized SV 40 chromosomes with mismatched ends,
Roth et al. (1985) sought to characterize the end-joining step
in nonhomologous recombination. The majority of the
recombination products were flush junctions, the result of
blunt end ligation, a process which occurs very efficiently in
mammalian cells (Wilson et al. 1982). This result indicates
that ligation of blunted, broken double-stranded DNA ends and
reunion may be the most common nonhomologous recombination
mechanism in mammalian cells. However, the frequency with which
junctonal homologies are found and the occurrence of a rare
type of junction called a duplication junction implies a role
for short, direct repeats in the resolution of deletion
junctions. Duplication junctions have nucleotides inserted at
the junction which generate a direct repeat of a sequence near
the junction. Additionally, one to three nucleotides at the
deletion endpoints are homologous to the nucleotides that
define the beginning and end of the repeat (Roth et al., 1985)
(Figure 1.6). To explain duplication junctions and breakpoints
Figure 1.6. Paired priming model for nonhomologous recombination. Two DNA duplexes having broken ends are diagrammed at the top of the figure. The ends are unwound, revealing short homologies (open arrowheads) through which the duplexes can transiently pair. On the left, mispairing occurs between direct repeats near one end (hatched arrowtails) and repair synthesis, primed by the pairing of the short repeats, fills the gap (filled rectangle). Replication of the heteroduplex containing a single-stranded loop would generate one flush junction and one duplication junction. Alternatively, repair against either the top or the bottom (lopped) strand would result in one flush junction or one duplication junction. In the process on the right, pairing between the terminal homologies primes repair DNA synthesis to fill the gap. Removal of the single-stranded tails and and ligation produces a flush junction with a short region of homology corresponding to the terminal homology (Roth et al., 1985). This model is applied to the deletion breakpoint sequences of the JH 37 and JH 21 deletions in Figures 1.21 and 1.22.
Figure 1.6
between short direct repeats, Roth et al. (1985) have suggested a paired-priming model which is essentially a slipped-mispairing mechanism initiated by a double-stranded break (Figure 1.6 from Roth et al., 1985).

C. Breakage and reunion: Enzymes

Breakage and reunion events require enzymes capable of catalyzing them. Likely mediators of such a process are enzymes which are able to break and rejoin DNA such as topoisomerases, (Ikeda, 1986; Bullock et al., 1985) replication origin nicking enzymes (Michel and Ehrlich, 1986) and site specific recombinases (Croce, 1987). Topoisomerases alter the topological state of a DNA molecule by making either single-stranded breaks (topoisomerase I) or double-stranded breaks (topoisomerase II), passing one strand through the break then resealing the break (Wang, 1985).

Topoisomerase II

Type II prokaryotic topoisomerases, E. coli DNA gyrase and T4 topoisomerase, have been shown to stimulate nonhomologous recombination in vitro (Ikeda et al., 1980; Ikeda et al., 1982; Ikeda, 1986). Unlike other prokaryotic rearrangement sites (Albertini et al., 1982) these recombination junctions do not contain short direct repeats.
Eukaryotic topoisomerase II from calf thymus is also able to mediate nonhomologous recombination between phage molecules in vitro (Bae et al., 1988). Of four junctions analyzed in this study, one junction had a 4 bp homology while the other three were completely nonhomologous.

Ikeda et al. (1982) have suggested a subunit exchange model as an explanation for how topo II, which is a homodimer, might facilitate DNA rearrangements. In the model, two enzyme molecules bind to the DNA, make double stranded breaks, then associate as a tetramer. During this association subunits are exchanged and previously noncontiguous (and nonhomologous) DNA breaks are rejoined (Figure 1.7 from Ikeda et al., 1982).

Topoisomerase II is a major component of mitotic chromosomes and is apparently located at the bases of chromatin loops (Earnshaw and Heck, 1985). Matrix association regions (MARs) (Cockerill and Garrard, 1986) and scaffold attached regions (SARs) (Gasser and Laemmli, 1986) are DNA sequences which bind preferentially with the nuclear matrix. These sequences are at least 200 bp long, >70% A-T and are thought to be anchorage sites for chromatin loops. The presence of topoisomerase II consensus cleavage sites in MARs (Cockerill and Garrard, 1986; Gasser and Laemmli, 1986) and its abundance in Drosophila nuclear matrix fractions (Berrios et al., 1985) have led Cockerill and Garrard (1986) to propose that topo II interacts directly with MAR sequences. Although topo II cleavage sites are not always associated with MARs, Sperry et
Figure 1.7. Gyrase subunit exchange model. Each rectangle represents a subunit of DNA gyrase, with two combined rectangles representing the homodimer. Two dimeric enzyme molecules (shaded and unshaded) bind to two different DNA molecules or to juxtaposed points on the same molecule (indicated by parallel lines of differing thicknesses). After cleavage (b), one gyrase-DNA complex assembles with the other complex, forming a tetrameric structure (c). The dissociation of the tetrameric form to dimeric forms results in subunit exchange that leads to exchange of DNA strands (d) (figure from Ikeda et al., 1982).
Figure 1.7

(a) 

\[ \text{Cleavage} \]

(b) 

\[ \text{Assembly} \]

(c) 

\[ \text{Subunit exchange} \]

(d) 

\[ \text{Resealing} \]
al., (1989) have predicted that MARs which have a high affinity for topo II cleavage might be a target for nonhomologous recombination. This prediction is supported by the detection of topo II cleavage sites in MARs associated with chromosomal breakpoints in the mouse and deletion breakpoints in the rabbit immunoglobulin genes (Sperry et al., 1989).

**Topoisomerase I**

Eukaryotic topoisomerase I has been shown to catalyze ligation of nonhomologous DNA in vitro in recombinations between single stranded DNAs and between single stranded and double stranded DNAs (Been et al., 1981; Halligan et al., 1982). Bullock et al. (1984) found that the viral and chromosomal DNA at excision points shared 2-3 bp homologies. Moreover, the trinucleotide, 5'-PyrrTT- 3' is adjacent to the crossover site in the rat genome. This trinucleotide fits the consensus sequence (5' - A/T G/C T/A T - 3') for topoisomerase I cleavage for the rat liver and wheat germ enzymes (Been et al., 1984). Although this site occurs very frequently in random DNA sequence (Been et al., 1984) Bullock et al. (1985) found that rat liver topo I cleavage sites were associated with SV 40 excision sites in a statistically significant manner.

Topo I activity has been associated primarily with transcription but the specific cellular functions of this enzyme are not known (Wang, 1985). Halligan et al. (1982) have
shown that topo I can catalyze recombinations between single-stranded and double-stranded DNA substrates. It is therefore possible to speculate that if the lagging strand (containing single-stranded gaps) at a replication fork was brought close to the earlier synthesized (and un-gapped) end of the same replication loop, the result would be the juxtaposition of a single-stranded gap region with a double-stranded region. This configuration, perhaps occurring in association with the nuclear matrix, could be a substrate for topoisomerase I cleavage and strand transfer, resulting in the loss of the newly replicated loop.

One conclusion which can be drawn from these observations and hypotheses is that the mechanism by which a rearrangement occurs is dependent upon when an event occurs in the cell cycle. For example, slipped-mispairing or template switching are likely to be a replication phenomena, therefore deletions which occur during S phase might be more likely to occur by this pathway. Rearrangements which arise due to the action of enzymes which break, digest and ligate DNA could conceivably occur at any time these enzymes are active during the cell cycle.

D. Research objective

In an effort to understand the mechanisms involved in the generation of spontaneous deletions in the human genome, I have
characterized four intragenic deletions of the human factor VIII gene at the DNA sequence level.

Hemophilia A, which is caused by a deficiency or abnormality in clotting factor VIII:C, is the most common inherited disease of blood coagulation. The gene for factor VIII is located on the long arm of the X chromosome and has been cloned at both the genomic and cDNA level (Gitschier et al., 1984; Toole et al., 1984). A nine kb transcript is divided into 26 exons separated by 25 introns for a total length of 186 kb.

Many deletions of the factor VIII gene have been identified (Higuchi et al., 1989; White and Shoemaker, 1989 for review; Antonarakis and Kazazian, 1988 for review), fourteen of which have been identified in our laboratory (Youssoufian et al., 1988) (Figure 1.8). Of these fourteen, three deletions were chosen for further study based on the availability of genomic probes which could detect junction fragments suitable for cloning. These three deletions are JH 1 (Antonarakis et al., 1985), JH 7 (Youssoufian et al., 1987) and JH 21 (Youssoufian et al., 1988). A fourth deletion, JH 37 (Higuchi et al., 1989) was provided by K. Olek for comparison with JH 7 since both deletions involve exon 14.
Figure 1.8. JH deletions in the factor VIII gene. The human factor VIII gene is diagrammed at the top of this figure with exons indicated by filled boxes, vertical lines and the appropriate numbers. The lengths of fifteen partial deletions are indicated below the gene diagram by stippled rectangles. Uncertainties in the endpoints of these deletions are indicated by the open rectangles flanking some of the deletions (figure from Youssoufian et al., 1988).
Figure 1.8
A. Restriction endonuclease analysis

Genomic DNA was isolated from peripheral blood by standard techniques (Kunkel et al., 1977). Digestion of 5-10 ug of DNA with various restriction endonucleases was carried out according to the manufacturers instructions (Bethesda Research Labs). Subsequent gel electrophoresis (0.8-1% agarose), transfer to nitrocellose, and hybridization with radiolabelled DNA probes were carried out by standard techniques (Maniatis et al., 1982). DNA probes were labelled by the random primer method (Feinberg and Vogelstein, 1983) using kits from either Amersham or Bethesda Research Labs. Table 1.1 summarizes the probes used to identify abnormal Eco RI fragments in patients JH 1, JH 7, JH 37 and in a carrier for deletion JH 21. Factor VIII genomic and cDNA probes utilized in this study are derived from recombinant plasmids and phage provided by D. Pittman and J. Toole from the Genetics Institute (Toole et al., 1984) and a cosmid clone provided by J. Gitschier (Gitschier et al., 1984).
Table 1.1

<table>
<thead>
<tr>
<th>Patient/ Deletion</th>
<th>Altered Eco RI Fragment (kb)</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>JH 1</td>
<td>15</td>
<td>a) .7kb BamHI-Hind II fragment containing exon 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) Probe 4: 1.3kb Bgl II - Hind III genomic fragment containing exon 10</td>
</tr>
<tr>
<td>JH 7</td>
<td>4.7</td>
<td>2.5kb EcoRI Bam HI exon 14 fragment subcloned from cDNA probe BC</td>
</tr>
<tr>
<td>JH 21</td>
<td>2</td>
<td>4.7kb Eco RI genomic fragment containing exons 2 and 3</td>
</tr>
<tr>
<td>JH 37</td>
<td>4</td>
<td>2.5kb Eco RI-Bam HI exon 14 fragment subcloned from cDNA probe BC</td>
</tr>
</tbody>
</table>
B. Genomic Cloning

A sample of 30-60 ug of genomic DNA from each individual was digested to completion with Eco RI and subsequently size selected by preparative agarose gel electrophoresis. Appropriately sized fractions (see Table 1.1) were ligated with GT10 arms (Promega Biotec) and packaged with Gigapack Gold extracts (Stratagene). Eco RI digested DNA from patient JH 1 was ligated with EMBL 4 arms without size selection (Stratagene) and packaged as before. Approximately 300,000 independent recombinant phage from each library were screened with the appropriate probes (Table 1.1) using standard techniques (Blattner et al., 1978). Probe 4 was utilized for JH 1 library screening and the library for JH 37 was screened with a 4.5 kb genomic Eco RI fragment which contains most of exon 14 (Figure 1.15). Hybridizing recombinant phages were purified and restriction mapped. Junction fragments were subcloned into pEMBL 18 (Dente et al., 1983) and M13 phage vectors (Messing and Viera 1982) for further mapping and sequencing. Corresponding 5' and 3' normal sequences were subcloned from both cosmid and phage clones of the human factor VIII gene. (Toole et al., 1984; Gitschier et al., 1984).

The 5' breakpoint of the JH 21 deletion occurs in a region of IVS 1 which has not been cloned. In order to obtain the 5' normal sequence, a human genomic library made from a 48, XXY
individual, provided by D. Pittman, Genetics Institute Inc., was screened with a 27 nt oligonucleotide which had been end-labelled with $^32P$ ATP (Maniatis et al., 1982). The oligonucleotide (5'-GTGCTTGGTAAAAATAAGCTTATC-3') was derived from the sequence of the JH 21 junction fragment clone. Three overlapping phage were isolated and mapped. A 1.5 kb normal Eco RI fragment that had been identified by hybridization of genomic Southern blots with the JH 21 junction fragment was identified in all three phages. This fragment was subcloned into pEMBL 18 and sequenced using junction fragment specific oligonucleotides, confirming its identity as the 5' normal fragment altered in deletion JH 21.

C. Sequence Analysis

DNA sequencing with both single stranded M13 clones and double stranded plasmid clones was performed using the dideoxy chain termination method (Sanger et al., 1977) and the modified DNA polymerase, Sequenase (U.S. Biochemicals). Both universal primers and specifically synthesized oligonucleotides were utilized. Sequence compilation and analysis were accomplished using commercially available software (Queen and Korn, 1984).
CHAPTER III
RESULTS

A. JH 1 Deletion

Restriction mapping and sequence analysis reveals that the 5' end of the deletion is within probe 4 approximately 1 kb 3' to exon 10 (Figure 1.10). The 3' breakpoint is located in IVS 18 approximately 1.2 kb 5' to exon 19 and occurs within an Alu repetitive element at nt +100-102 of the Alu consensus sequence (Deininger et al., 1981). Identification of an altered Eco RI fragment in this patient required the isolation of a 700 bp Hind III - Bam HI genomic fragment from the 3' end of the deletion because probe 4 could not differentiate between the approximately 15 kb altered Eco RI fragment and the normal 15 kb RI fragment (Table 1.1, Figures 1.9 and 1.10). No significant sequence homology was detected between the 5' and 3' normal sequences around the breakpoints although a 3 bp direct repeat is present at the breakpoint (Figure 1.11). This deletion removes 57 kb from the factor VIII gene, resulting in a frameshift mutation (assuming normal splicing between exons 10 and 19) as well as a loss of 4461 nt of coding sequence.
The 5' normal sequence involved in the deletion junction contains a putative matrix associating region (MAR) (Cockerill and Garrard, 1986) (see Figure 1.11 legend) identified by the >70% A+T content of the sequence.
Figure 1.9. Altered restriction fragments in patient JH 1.

Five micrograms of genomic DNA from normal individuals (N) and the JH 1 patient (P) were digested to completion with Eco RI, run on a 0.8% agarose gel and blotted to nitrocellulose. The probe is described in the figure and diagrammed in Figure 1.10. The 15 kb deletion junction band in the patient also hybridizes to probe 4 (see Figure 1.10). The 9.3 kb band is detected by a contaminating 0.7kb fragment which co-purifies with the Bam HI/Hind III fragment. Note that both the 3.8 and 9.3 kb bands are missing in the JH 1 deletion patient as would be expected since the 9.3 kb Eco RI fragment is 5' and adjacent to the 3.8 kb Eco RI fragment (see Figure 1.10).
ivs 18 probe: 0.7 kb BamHI / HindIII fragment
Figure 1.10. Restriction map of JH 1 deletion breakpoint (JH 1 BP). The 5' and 3' normal regions of the factor eight gene are depicted in this figure. The 15 kb Eco RI fragment containing exons 7-10 of the factor VIII gene is shown as a bold line, exons are depicted as filled boxes. Probes utilized in restriction analysis and library screening are presented as striped boxes, dashed lines indicate the junction sites. The arrow shows the location and orientation of an Alu repetitive element which is involved in the deletion breakpoint. The restriction sites are indicated as follows; Bam HI (B), Eco RI (R), Bgl II (Bg) and Hind III (H).
Figure 1.10
Figure 1.11. JH 1 deletion breakpoint sequence. The 5' and 3' normal sequences are shown aligned with the junction fragment sequence. Homology between the normal and junction sequences is indicated by shading. The 3bp direct repeat (CAA) present at the breakpoint is boxed. The first nucleotide of this repeat in the 3' normal sequence corresponds to nucleotide +100 in the Alu repeat consensus sequence (Deininger et al, 1981). The Alu repeat is indicated by the arrow. The putative MAR sequence having an A+T content of 72% is indicated by a bracket on the right. The A/T rich sequence motifs (5'-A/TATTTT-3') (Cockerill and Garrard, 1986) for both strands are underlined with mismatches indicated by an open square. Sequences which show homology of 12/15 or greater to the Drosophila topo II consensus cleavage sites (5'-GINA/TAC/TATTNINNA/G-3') (Sander and Hsieh, 1985) for both strands are indicated by circles, filled circles denote a match, open circles indicate a mismatch. Topo I sites near the deletion junctions (5'-PyrrTT-3' or 5'-GTT-3')(Been et al., 1984) for both strands are indicated by stars.
B. JH 7 Deletion

This partial deletion of exon 14 was a de novo mutation in the germ line of the patient's maternal grandfather (Youssoufian et al., 1986). An altered 4.7 kb Eco RI fragment was identified using a cDNA probe specific for exon 14 (Figure 1.12). The 3' deletion breakpoint occurs in exon 14 between nucleotides 2816 and 2817 of the factor VIII coding sequence (Wood et al., 1984). The deletion extends for 3 kb in the 5' direction, ending in IVS 13 (Figure 1.13). Loss of the exon 14 acceptor splice site generates a frameshift mutation (Youssoufian et al., 1986).

This deletion event is also nonhomologous in nature and like JH 1 possesses a short (2 bp) direct repeat at the breakpoint (Figure 1.14). The 5' normal sequence contains nt -167 to -111 of an Alu repeat (Deininger et al., 1981), 279 bp 5' to the breakpoint (data not shown).
Figure 1.12. Altered restriction fragments in patients JH 7 and JH 37. A subclone of the factor VIII cDNA probe EC which detects only exon 14 sequences was utilized to detect the deletion junction fragments in patients JH 7 and JH 37. Five micrograms of normal (N), heterozygote (H), and patient (P) DNA were digested with Eco RI, run in 0.8% agarose and blotted to nitrocellulose. The exon 14 probe identified an altered 4.7 kb fragment in patient JH 7 and a 4.0 kb band in patient JH 37.
Figure 1.12

exon 14 probe: 2.5 kb Bam HI / Eco RI fragment
Figure 1.13. Restriction map of JH 7 deletion breakpoint (JH 7 BP). Three Eco RI fragments from IVS 13, encompassing exon 14 and part of IVS 14 are depicted. IVS sequences are shown as a continuous line, the filled box represents exon 14. The striped box represents the fragment of probe BC which was used for restriction analysis and library screening. The endpoints of the deletion are indicated by dashed lines. The filled circle over the most 5' Eco RI site shown designates which site is present in the altered fragment. Restriction sites are symbolized: Bam HI (B), Bgl I (Bg), Eco RI (R), Kpn I (K), Pst I (P) and Sst I (S).
Figure 1.13

FVIII Gene

JH-7 4.7 kb altered Eco R1 fragment

exon 14 probe

1 kb
Figure 1.14. JH 7 deletion breakpoint sequence. Homology between the aligned normal and breakpoint sequences is indicated by shading. The 2 bp homology (CT) is boxed at the breakpoint. The A+T content for the 5' and 3' normal sequence is 62.3% and 60.4% respectively. Topo I and II cleavage sites are indicated as described in the legend for Figure 1.11.
Figure 1.14
C. JH 37 Deletion

JH 37, like JH 7, is a partial deletion of exon 14 which occurred as a de novo mutation in the maternal grandfather of the patient. Genomic cloning of the altered 4 kb Eco RI fragment (Figure 1.12) and subsequent sequence analysis places the 3' deletion junction between nucleotides 3859 and 3861 in the coding sequence of the factor VIII gene. The 5' breakpoint is located 2.5 kb upstream in IVS 13 (Figure 1.15).

A short direct repeat of 3 bp (Figure 1.16) is present at the JH 37 deletion breakpoint as is the case with deletions JH 1 and JH 7. The JH 37 deletion was also generated by non-homologous recombination since no significant sequence homologies were detected.
Figure 1.15. Restriction map of JH 37 deletion breakpoint (JH 37 BP). The normal restriction map for the area involved in this deletion is the same as in Figure 1.13. Although the exon 14 probe represented by the striped box was used for restriction analysis, the 4.7 Eco RI fragment which contains most of exon 14 was used for library screening in this case. The filled circle over the IVS 13 Eco RI site indicates which Eco RI site is present in the altered fragment.
Figure 1.15
Figure 1.16. JH 37 deletion breakpoint sequence. The shaded areas indicate the regions of homology between the aligned 5' and 3' normal sequences and the breakpoint sequence. The normal sequences share a 3 bp (TTC) at the deletion junction which is indicated by a box. A+T content is 61.1% for the 5' side and 62.2% for the 3' side of the break. Topo I and II sites are indicated as described in figure legend 1.11.
Figure 1.16
D. JH 21 Deletion

The JH 21 factor VIII gene deletion extends from IVS 1 to IVS 3, removing exons 2 and 3 and generating a frameshift mutation. The JH 21 deletion was originally identified by the loss of a 4.7 kb Eco RI fragment which contains exons 2 and 3 (Youssoufian et al., 1988). When the 4.7 kb Eco RI fragment was used as a probe against a panel of deletion patients an abnormal 2 kb Eco RI fragment was identified in a heterozygote with the JH 21 deletion (Figure 1.17). The 2 kb deletion junction fragment was cloned and hybridized to Eco RI digests of normal and heterozygote DNA. A 1.5 KB Eco RI fragment was identified (Figure 1.18) which is not present in the published map of the factor VIII gene (Gitschier et al., 1984). This result indicates that the 5' breakpoint of the JH 21 deletion resides in a region of IVS 1 which had previously been mapped by genomic restriction mapping but not by cloning (Gitschier et al., 1984).

Sequence from the junction fragment for this deletion provided a 27 nt oligonucleotide probe which has been used to isolate overlapping clones from the IVS 1 "gap" (Figure 1.19). Although the 5' side of the gap has been closed, the 3' side will require further library screening with an oligonucleotide probe generated from the 6.4 kb RI fragment of phage J311 (Figure 1.19). Since the 6.4 kb RI fragment of phage J311
does not hybridize to phage X554, the deletion is greater than 20 kb in length (Figure 1.19).

The junction sequence of JH 21 differs from the other three sequences in that 2 nucleotides have been inserted at the break, thereby generating two adjacent 4 bp direct repeats (Figure 1.20). The 5' and 3' sequences juxtaposed in this event do not show significant homology to each other. Although Alu sequences were not present at the breakpoint, an Alu repeat is present in the 5' normal sequence 116 bp 5' to the deletion junction.

A possible MAR region has also been identified in association with this deletion in the 3' normal sequence (see Figure 1.20).
Figure 1.17. Screening of JH deletion patients with a 4.7 kb Eco RI fragment containing exons 2 and 3. Five micrograms of genomic DNA from each individual carrying a deleted factor VIII gene was digested with Eco RI, run on a 0.8% agarose gel, blotted to nitrocellulose and hybridized with the 4.7 kb Eco RI fragment containing exons 2 and 3. The numbers above each lane refer to the deletions diagrammed below. Only the JH 21 deletion shows an altered fragment with this probe.
Digest: Eco RI
Probe: 4.7 kb Eco RI fragment spanning exons 2 and 3

Figure 1.17
Figure 1.18. Identification of the normal 5' Eco RI fragment involved in the JH 21 deletion. Five micrograms of normal (N) and heterozygote (H) DNA were digested with Eco RI, blotted to nitrocellulose and hybridized with the 2 kb Eco RI junction fragment isolated from the individual who is heterozygous for the JH 21 deletion. The probe identifies the 4.7 kb Eco RI fragment in both the normal and heterozygous individual as would be expected since the 4.7 kb fragment was used to isolate the junction fragment. The junction fragment probe also identifies a 1.5 kb fragment which is not present in the published map of factor VIII (Gitschier et al., 1984) but is present in three overlapping phage isolated using an oligonucleotide probe from the junction fragment (see Figure 1.19). The 2 kb fragment in the heterozygote lane is the junction fragment.
Digest: Eco RI

Probe: JH 21 2kb deletion junction fragment

Figure 1.18
Figure 1.19. Restriction map of JH 21 deletion (JH 21 BP). This restriction map links the published map of IVS 1 of the factor VIII gene (Gitschier et al, 1984) with the Eco RI (R) - Bam HI (B) map derived from phage clones J111, B111 and J311, represented by open boxes. Phage clones, X761 and X554, (Toole et al., 1984) from Genetics Institute are also presented.

Restriction fragments of estimated size are indicated by parentheses as are restriction sites which are cloning sites. Fragment lengths are given in kilobases. The Eco RI fragment which contains the 5' end of the deletion is indicated by a shaded box, exons 2 and 3 are represented as filled boxes. The striped box indicates the 4.7 Eco RI fragment which was utilized for restriction analysis and library screening. Filled and open circles denote the Eco RI sites which define the JH 21 deletion breakpoint fragment. According to this map IVS 1 is at least 24 kb larger than originally thought and the deletion is 24 kb or larger.
Figure 1.19
Figure 1.20) JH 21 deletion breakpoint sequence. Homology between the normal sequences and the deletion junction sequence is indicated by shading. Two nucleotides (TG) are inserted at the breakpoint generating a direct repeat of four nucleotides (GTTT) which are boxed. The arrow denotes the orientation and location of an Alu repeat in the 5' normal sequence. The 5' normal sequence is 61.7% A+T. The candidate MAR sequence present in the 3' normal sequence is 74% A+T and is indicated by a bracket on the right. A/T rich sequence motifs, topo I and II sites are shown as in Figure 1.11.
CHAPTER IV
DISCUSSION

Many deletions at genetic loci in the human genome are examples of eukaryotic nonhomologous recombination. Table 1.2 summarizes the sequence characteristics found in 28 deletions (56 breakpoints) characterized at eight genetic loci in humans for which both the 5' and 3' normal sequences are known. Deletions which result from homologous recombination between regions of DNA sharing extensive homology, such as certain α-globin deletions (Nicholls et al., 1987), have not been included in Table 1.2.

The deletions presented in Table 1.2 are primarily germ line mutations, including three of the four retinoblastoma gene deletions. One retinoblastoma deletion is a somatic mutation which gave rise to a tumor (Canning and Dryja, 1989). The analysis of the factor VIII gene deletions at the genetic and sequence level indicates that nonhomologous recombination mechanisms are the products of cellular processes common to both mitotically and meiotically dividing cells. Haplotype analysis reveals that both exon 14 deletions are de novo events occurring in males (Youssoufian et al., 1987; Higuchi et al., 1989) who are hemizygous for the X-linked factor VIII gene.
Table 1.2. Summary of 28 deletions in the human genome. Twenty-eight deletions for which sequence data is available for both the 5' and 3' breakpoints are categorized by sequence characteristics found at the deletion junction. The references are: beta globin; Vanin et al., 1983; Mager et al., 1985; Spritz and Orkin, 1982; Popovich et al., 1986; Gilman, 1987; Anand et al., 1988; Kolozik et al., 1988; Henthorn et al., 1986; Jennings et al., 1985, alpha globin; Nicholls et al., 1987, LDL receptor; Lehrman et al., 1985; Lehrman et al., 1986; Hobbs et al., 1986; Lehrman et al., 1987b, adenosine deaminase; Markert et al., 1988; hexosaminidase chain; Myerowitz and Hogikyan, 1987, factor IX; Green et al., 1988, factor VIII; this paper, retinoblastoma gene; Canning and Dryja, 1989.
Table 1.2

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>0-1 BP</th>
<th>2-6 BP</th>
<th>NO REPEAT</th>
<th>WITH REPEAT</th>
<th>ALU-ALU HOMOLOGY</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BETA GLOBIN</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>ALPHA GLOBIN</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LDL RECEPTOR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>ADENOSINE DEAMINASE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HEXOSAMININDASE</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FACTOR IX</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FACTOR VIII</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>RETINOBLASTOMA GENE</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>%</td>
<td>11</td>
<td>43</td>
<td>11</td>
<td>14</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>
This means that even if the deletions arose during meiosis, the factor VIII locus is not paired at the synapse of homologous pairs, the event which differentiates meiotically dividing cells from mitotically dividing cells. Therefore the exon 14 deletion events are either the result of unequal sister chromatid exchange as has been suggested for duplications in the DMD locus (Hu et al., 1989) or an intrachromosomal looping-out mechanism.

Sequence analysis of four inherited factor VIII deletions has shown that these germ line mutations are similar to the somatic retinoblastoma mutation (Canning and Dryja, 1989) and deletions in cultured cells (Nalbantoglu et al., 1986) in that short direct repeats are present at the deletion junctions of both. The analysis of factor VIII deletions, therefore, does not support the hypothesis that nonhomologous recombination occurs by the different mechanisms in somatic versus germ line cells. An alternate conclusion from these observations is that sister chromatid exchange can occur in both mitotically and meiotically dividing cells as would be expected.

Six of the 28 deletions presented in Table 1.2 occur between Alu repeats or between an Alu repeat and an exon having homology to Alu sequences (Lehrman et al., 1986). Alu-Alu recombination appears to play a major role in the formation of deletions and duplications in the low density lipoprotein receptor (IDL) locus (Lehrman et al., 1987a). The majority of the other deletions in Table 1.2 are nonhomologous events,
although 17 of the 56 breakpoints (30%) do involve an Alu sequence on one side of the break. Lehrman et al. (1987a) have suggested that the frequent occurrence of deletion endpoints in Alu elements indicates a role for these elements in genomic rearrangements. Of eight factor VIII deletion endpoints analyzed in this study only the 3' normal endpoint of JH 1 involves an Alu element. Alu sequences are present in the 5' normal sequence of deletions JH 7 and JH 21 within 279 and 116 bp from the junctions, respectively. The presence of Alu repeats near recombination junctions is of unknown significance since Alu repeats occur in the human genome about every 5 kb (Schmid and Jelinek, 1982). The current data for factor VIII deletion breakpoints does not support an important role for Alu repeats in the generation of deletions at this locus.

Forty-three percent of the deletions summarized in Table 1.2 have 2-6 bp of homology between the 5' and 3' normal sequences at the junction. In an analysis of homologies at SV 40 sites of excisions and integrations and chromosomal translocation junctions, breakpoints having 2-5 bp of homology occurred more frequently than would be expected by chance (Roth et al., 1985), suggesting a facilitating role for short repeats in the joining of broken ends. Short direct repeats (2-7 bp) are also present at deletion junctions of the aprt locus in hamster cells (Nalbantoglu et al., 1986) and Efstradiatis et al. (1980) have implicated direct repeats of 3 to 11 bp in the
generation of small deletions (<50 bp) during replication of the β-globin locus. Spontaneous deletions of 9-900 bp in the E. coli lac I gene (Farabough et al., 1978, Albertini et al., 1982) occur predominately between short homologies of 5 to 8 bp. The three factor VIII deletions which have 2-3 bp of homology at their junctions, JH 1, 7 and 37, fall within this class.

Twenty-five percent of the deletions contain varying numbers of inserted nucleotides (1-41 bp). Nearly 60% of these "insertion junctions" generate a direct repeat at the breakpoint. The JH 21 deletion has 2 nucleotides inserted at its junction generating a 4 bp direct repeat consisting of GTTTGTGT at the point of insertion. In the SV40 study mentioned previously (Roth et al., 1985), a small number of recombination junctions were found to contain inserted nucleotides which generate a direct repeat of a sequence found close to the junction. Roth et al. (1985) called these junctions duplication junctions (Figure 1.6). The JH 21 deletion is of this type (see Figure 1.20) as are six other deletion breakpoints in Table 1.2.

The direct repeats and the insertion/ duplication found at the four JH deletion breakpoints are consistent with the model for nonhomologous recombination suggested by Roth et al. (1985). In this mechanism, diagrammed in Figure 1.6, double stranded DNA breaks are repaired after the broken ends are unwound or resected to reveal short direct repeats. The broken
Figure 1.21. Paired-priming mechanism for the presence of direct repeats at JH deletion junctions. A) A double stranded break is made in the 5' normal DNA on the right and the 3' normal DNA on the left. B) The ends are unwound or resected by digestion revealing the direct repeats. C) The repeats pair and DNA repair or replication, indicated by the dashed arrow, generates the deletion junction of JH 37.
Figure 1.21
Figure 1.22. Paired-priming mechanism for the generation of duplicated nucleotides at the JH 21 deletion junction. A) A double stranded break is made in the 5' normal DNA on the right and the 3' normal DNA on the left. B) The ends are unwound or resected by digestion revealing the direct repeats. C) The AA dinucleotide "slips" and mispairs with an upstream TT dinucleotide forming a single stranded loop. DNA repair generates the TG dinucleotide insertion and replication using the bottom strand as the template generates the GITT repeat at the junction (D). The stars indicate the nucleotides of the recombination endpoints which are homologous to the beginning and end of the repeat as pointed out by Roth et al., 1985.
Figure 1.22
ends then pair at these repeats and gaps are filled in, with the recombined DNA molecule retaining one copy of the repeat. Roth et al., (1985) incorporated some aspects of the slipped-mispair mechanism of recombination in order to explain the occurrence of duplication junctions like the JH21 deletion junction (Figure 1.6). Figures 1.21 and 1.22 presents the paired-priming mechanism of Roth et al. (1985) as it applies to the factor VIII deletions.

Streisinger et al. (1966) proposed the slipped-mispairing model for the generation of frameshift mutations. This model has also been suggested as an explanation for the presence of direct repeats at deletion junctions (Farabough et al., 1978; Efstradiatis et al., 1980; Albertini et al., 1982) and at SV 40 recombination junctions (Roth et al., 1985). Although slipped-mispairing of short direct repeats in deletions that are less than 1 kb seems reasonable, the JH deletions vary in size from 2.5 kb to 57 kb and the deletions summarized in Table 1.2 vary in length from 600 nt to greater than 100 kb. If some type of slipped mispair recombination is involved in larger deletions then one would have to postulate an additional mechanism for the juxtaposition of noncontiguous stretches of single stranded DNA as has been suggested by Brunier et al., 1988. Vanin et al. (1983) and Anand et al. (1988) have postulated that deletions of varying size occur due to breakage and reunion of DNA between replication forks which are brought into close proximity by attachment to the nuclear matrix. Since
both slipped mispairing and the loop model are postulated to be replication phenomena, perhaps DNA polymerase can switch between replication forks assisted by mispairing between short repeats.

The processes diagrammed in Figure 1.21 and 1.22 take place after an initiating double-stranded DNA break. Although the enzymes which catalyze the breaking and rejoining of DNA in nonhomologous recombination are not known, the activities of the topoisomerases make them likely candidates. Topoisomerases control the topological state of a DNA molecule by making single stranded (topo I) or double stranded (topo II) breaks, passing one strand through the break, and then resealing the break (Wang, 1985). Both eukaryotic topoisomerase I and II have been implicated in genetic rearrangements (Halligan et al., 1982; Been et al., 1981; Bae et al., 1988).

Both topo I and II have the ability to break and rejoin DNA but differences in cellular distribution might have an effect on their involvement in the generation of deletions. Topo II is a major component of the Drosophila nuclear matrix (Berrios et al., 1985) and sequences shown to associate with the matrix contain topo II consensus cleavage sites as well as being at least 200 bp long and >70% A+T (Cockerill and Garrard, 1986; Gasser and Laemmli, 1986). Sperry et al. (1989) have suggested that matrix associating regions (MARS) that are associated with topo II cleavage sites might be responsible for chromosomal translocation and deletions based upon in vitro
topo II assays performed for MARs located near these
breakpoints.

The JH1 5' and JH21 3' deletions endpoints reside in
"candidate" MAR sequences. The putative JH1 5' MAR is 400 bp in
length, 72% A + T and contains 1 perfect and 2 imperfect A/T
rich sequence motifs in addition to three Drosophila topo II
consensus sites with greater than or equal to 12/15 bp
homology, one of which is 14 bp 3' to the deletion junction
(Figure 1.11). JH21 5' normal sequence has a 300bp, 74% A+T
region which contains three perfect and one imperfect A/T rich
sequence motifs often found in MARs (Cockerill and Garrard,
1986) and four Drosophila topo II consensus sequences having a
match of 12/15 6bp or better (Figure 1.20). The MARs described
for JH1 and JH21 are referred to as candidates because they
have not been assayed for matrix association. Topo II cleavage
and matrix associating assays of the JH1 and JH21 deletion
endpoints would therefore be of great interest.

Bullock et al. (1984) found the trinucleotide, 5'-PyrTT-
3' to be associated with SV 40 excision sites in the rat
genome. This trinucleotide fits the consensus sequence (5'
- A/T G/C T/A T - 3') for topoisomerase I cleavage for the rat
liver and wheat germ enzymes (Been et al., 1984). All of the
factor VIII deletion junctions in this study either contain or
are adjacent to a 5'-PyrTT-3' or 5'-GTT-3' trinucleotide
(Figures 1.11,14,16,20), suggesting that topo I cleavage might
be involved in these rearrangements.
In conclusion, it seems likely that genomic rearrangements in general and deletions specifically are the result of more than one set of factors which may vary from locus to locus. For the factor VIII deletions, the most likely sequence of events is the double-stranded breakage of the DNA, followed by an end-joining reaction as diagrammed in Figures 1.21 and 1.22, perhaps in association with the matrix and/or replication. The identity of "breaking agent" is not clear although topoisomerase II seems the most likely candidate because this enzyme makes double-stranded DNA breaks.
LIST OF REFERENCES FOR PART 1


PART 2

MOLECULAR CHARACTERIZATION OF L1 INSERTIONS

IN THE HUMAN FACTOR VIII GENE
CHAPTER V
INTRODUCTION

Transposable elements are classified according to their mode of transposition. Elements that transpose via reverse transcription of an RNA intermediate are called retrotransposons. Retrotransposons can be further classified based on their similarity to retroviruses (Fanning and Singer, 1987). The generalized structures of Class I and II retrotransposons are depicted in Figure 2.1 (modified from Finnegan, 1989).

Copia-like elements in Drosophila and Ty elements in Saccharomyces cerevisiae are examples of type I retrotransposons in that they are flanked by terminal direct repeats (TIR's), putative primer binding sites (PBS and PRS), and contain two open reading frames with similarities to the gag gene and the reverse transcriptase, protease, integrase, and ribonuclease of retroviruses (Finnegan, 1989; Boeke, 1989). Ty elements are known to transpose through an RNA intermediate (Boeke et al., 1985) and encode reverse transcriptase activity (Garfinkel et al., 1985; Mellor et al., 1985), making a functional correlation between these elements and retroviruses. Moreover, virus-like particles (VLP's) containing Ty RNA
molecules, Ty encoded reverse transcriptase and capsid proteins have been found in *S. cerevisiae* cells undergoing transposition (Boeke, 1989).

Class II retrotransposons (Figure 2.1) are more distantly related to retroviruses in that they lack LIR's and are often deleted at the 5' end (Fanning and Singer, 1987). Sequence analysis of the Class II open reading frames indicates that the first open reading frame may encode a nucleic acid binding protein (gag-like) and the second (RT) open reading frame contains similarities to ribonuclease H, a protease and an RNA binding site (Doolittle et al., in press). This class of elements is represented by the F,I,G, and jockey elements of *Drosophila*. Class II elements are apparently transcribed by RNA polymerase II because their 3' ends are A-rich. Mizrohki et al., (1988) has shown that the jockey element contains an internal promoter for RNA polymerase II.

LI elements or LINES (long interspersed elements) are Class II retrotransposons present at greater than 10^4 copies per genome in all mammals that have been studied (Fanning and Singer, 1987a; Burton et al., 1986). Full length human elements (which are designated L1Hs for homo sapiens) are about 6 kb but the majority of LI elements are 5' truncated and some are rearranged (Fanning and Singer, 1987a; Voliva et al., 1984; Grimaldi et al., 1984). Most LI elements have adenine-rich 3' ends and are flanked by direct repeats (target site duplications) of 7 to 16 bp (Fanning and Singer, 1987a; Voliva et al.,
Figure 2.1. Schematic diagrams of Class I and II retrotransposons. Class I elements are retrovirus-like with direct repeats at each end (LTRs). PBS refers to a potential tRNA primer-binding site just after the left LTR and FRS refers to a purine-rich sequence just before the right LTR. These sequences are thought to be involved in the reverse transcription of the RNA molecule. The thin lines above the are between the two LTRs indicate the locations of open reading frames (ORFs) contained in these elements. The first ORF has some similarity to the retroviral gag gene, the second ORF encodes a potential reverse transcriptase. Some elements have a third ORF (3 in the figure) located in a position similar to the env genes of retroviruses but not related to them in sequence. Examples of Class I retrotransposons are copia-like elements in Drosophila (Mount and Rubin, 1985) and Ty in yeast (Boeke et al., 1985). Class II retrotransposons do not have LTRs and usually have two ORFs indicated as before in the figure. Most of these elements have an A-rich sequence (An) at the 3' end of one strand and are often deleted at the 5' end by varying amounts as indicated by the varying lengths of the thick lines below the ORFs. 11 elements in mammals (Fanning and Singer, 1987a), F,I,G, and jockey elements in Drosophila (Dinocera and Casari, 1987; Fawcett et al., 1986; Dinocera et al., 1986; Priimagi et al., 1988), ingi in Trypanosoma brucei (Kimmel et al., and R2 in Bombyx mori are examples of this class of retrotransposon (figure from Finnegan, 1989).
Figure 2.1
LI elements from all mammals are similar in overall structure. Consensus sequences derived from a number of LIHs elements identify two open reading frames (5' and 3' ORFs) separated by a 49 bp non-coding region (Hattori et al., 1986; Scott et al., 1987). Regions of sequence similarity with the polymerase domain of reverse transcriptases are present in the 3' ORF (Hattori et al., 1986; Loeb et al., 1986; Fanning and Singer, 1987b).

LI transcripts have been identified in various cell types (Schmeckpeper et al., 1984; Kole et al., 1983; Shafit-Zagardo et al., 1983), including full-length polyadenylated RNAs from a teratocarcinoma cell line, NTera2D1 (Skowronski and Singer, 1985). Poly A+ RNA from this cell line was used to prepare a cDNA library from which several LI cDNAs were isolated and characterized (Skowronski et al., 1988).

Target site duplications, 3' poly A tracts and the identification of LI-specific transcripts are all indirect evidence that LI elements are retrotransposons and are dispersed in the genome by reintegration of an RNA or cDNA copy (Fanning and Singer, 1987b; Temin, 1985; Loeb et al., 1986). Additional evidence for LI transposition has been provided by observations of allelic variation (Burton et al., 1985; Iakshmikumaran et al., 1985) and somatic mutation (Katzir et al., 1985) due to LI element insertion. Two recent LIHs retrotranspositions have been observed in exon 14 of the human factor VIII gene, both of which cause hemophilia A (patients
JH-27 and JH-28 of Kazazian et al., 1988) (Figure 2.2). Both insertions are de novo events, present in the mother’s germ cells but not in her peripheral blood. If the insertions occurred in the patients, the event would have to have been at the one-cell stage since the patients do not appear to be mosaic in genotype and exhibit severe phenotype.

We now report the characterization of a third LI insertion in intron 10 of the factor VIII gene of a hemophilia A patient (JH-25). This LI insertion is not the cause of hemophilia A in the patient because it is also present in the maternal grandfather who does not have the disease.

Most LIHs elements lack the ORFs which are present in the consensus sequence because of one or more mutations which disrupt the coding sequence. In contrast, the nucleotide sequences of the three JH LI elements contain open reading frames.

Approximately 300 patients with hemophilia A have been analyzed in our laboratory, one percent of whom have recent LI insertions. This high rate of transposition fits well with the prediction that a relatively high rate of transposition is necessary to account for the large copy number and relative sequence similarity of these elements (Scott et al., 1987).

Ty element expression and subsequent transposition is regulated in yeast, which helps control the copy number (approximately 30) and possibly deleterious effects of insertional mutagenesis. The NDTera 2 cell line is
Figure 2.2. Diagram of two L1 insertions in exon 14 of the human factor VIII gene. Two L1 insertions, represented by shaded boxes, are shown with their insertion sites relative to the factor VIII gene map at the top of the figure. The orientation of the L1 element's open reading frames are indicated by arrows within the boxes (the JH 28 insertion is rearranged). The poly A tail lengths (Aₚ) are also shown within the boxes. The bold numbers above the boxes refer to nucleotide positions in the L1 consensus sequence (Scott et al., 1987). The sequences of the target site duplications are given with dashed lines indicating ambiguity in the precise boundaries of the site. The numbers above the target sites refer to the nucleotide positions in the factor VIII cDNA sequence (Wood et al., 1984). (Figure from Kazazian et al., 1988).
Patient JH-27

Patient JH-28

Figure 2.2
phenotypically embryonic and can be induced to differentiate. The full-length transcript present in the embryonic stage disappears after differentiation. This observation is interesting in light of the fact that the de novo hemophilia insertions occurred in the germ cells of the mother or during the patients' early development. One could speculate that the cellular factors necessary for efficient L1 element transcription and subsequent transposition are present primarily in totipotent cell types. The occurrence of a somatic L1 element insertion in a canine tumor (Katzir et al., 1985) and the presence of L1 transcripts in other cell types (see references above), however, indicate that other scenarios for L1 transcription and transposition exist.

The proposed mechanism of Ty transposition is diagrammed in Figure 2.3 from Boeke (1989). Transcription starts within the LTR indicated by arrowheads but does not include the flanking 5 bp parental direct repeats represented by arrows. The RNA molecule is then packaged into VLP's with some form of primer molecule and reverse transcriptase because purified Ty-VLP's can carry out reverse transcription without addition of exogeneous reagents. Integration of Ty elements is presumably through the action of Ty integrase which makes a cut in a nonspecific parental target site, with a 5 bp overhang which is subsequently repaired to regenerate flanking direct repeats of parental DNA.
Figure 2.3. Model for Ty transposition mechanism. The open box flanked by boxed, filled triangles represents a Ty element flanked by LTRs. The small arrows above the filled boxes indicate the 5 bp direct repeats which flank the parental element. The parental element is transcribed in the nucleus and the RNA (wavy line) is packaged into a cytoplasmic Ty VLP particle in which reverse transcription takes place. The progeny Ty element (now a ds DNA molecule) forms a hypothetical intermediate, complexed with the integrase protein (squiggly line) which presumably makes a staggered 5 bp cut in the target DNA (boxed T). The target site is shown with an arbitrary 5' overhang to which the ends of the element are joined. Repair synthesis of DNA fills the gaps generating 5 bp direct repeats which flank the progeny Ty element (figure from Boeke, 1989).
Figure 2.3
How much of the Ty process is analogous to L1 (Class II) transposition is unknown. There is currently no evidence for L1 particles although the analysis of Doolittle et al. (in press) implies that an L1 RNA protein complex may exist. It has been suggested that blocks of DNA may be transferred between different L1 elements giving rise to patchwork sequences (Scott et al., 1988; Jubier-Maurin et al., 1985). Template switching by reverse transcriptase between packaged RNA molecules has been shown for poliovirus particles (Kirkegaard and Baltimore) and hybrid transcripts have been noted in Ty elements (Boeke, 1989).

Integration of L1 elements appears to also be initiated by a staggered double-strand break in the target site since L1 elements are flanked by direct repeats of varying length. These repeats have been shown to be parental target site duplications in the recently inserted JH1 elements (Kazazian et al., 1988, Figures 2.2 and 2.5) and interestingly are more conserved in length (12 - 13 bp not counting ambiguity Figures 2.2 and 2.5). Since DNA insertion necessitates breakage and repair it is possible that the process might lend to other types of nonhomologous DNA rearrangements such as chromosome translocation. Retroviral insertion has been associated chromosome translocation in mice (Mahen et al., 1988). L1 integration may however, be a passive process in that ds DNA breaks are caused by some other factor and L1 cDNA copies are inserted as part of the repair process in a manner analogous to
processed pseudogene formation (Vanin et al., 1985). Kazazian et al. (1988) has pointed out that the JH Ll element target sites are A rich, thereby providing a complimentary sequence for the poly T region of a cDNA copy of a polyadenylated Ll transcript.

Transposable elements exist in many systems (see Berg and Howe, 1989) serving an unknown function, if any, in the genomes of their hosts. Although the Ll element insertions in the human genome have proven to be either deleterious or neutral, these elements are providing another means of change within the DNA molecule, a necessary precursor to adaptive evolution.
CHAPTER VI
MATERIALS AND METHODS

A. Restriction endonuclease analysis

Genomic DNA from members of family JH-25 was isolated from leukocytes contained in 5 to 15 ml of peripheral blood as previously described (Kunkel et al., 1977). Five to ten micrograms of genomic DNA were digested to completion with several enzymes. Subsequent gel electrophoresis, transfer to nitrocellulose, hybridization, washing and autoradiography were carried out by standard techniques (Maniatis et al., 1982). The probes used are (i) a 0.8 kb Hind III-Bam HI cDNA fragment containing exons 8-12 of the human factor VIII gene, isolated from probe A, a 1.7 kb Rpn I-Sst I cDNA clone spanning exons 1 to 12 of the factor VIII gene, and (ii) a 0.4 kb Bgl II genomic fragment containing exon 11 of the factor VIII gene. The 0.4kb Bgl II fragment was isolated from pF8ex11(2.5)R, a subclone of X031 (Toole et al., 1984). Both probe A and X031 were provided by D. Pittman and J. Toole from the Genetics Institute.
B. Cloning and sequence analysis

A recombinant phage containing the abnormal 3.2 kb EcoRI fragment from the hemophilia A patient (individual IV-1 in figure 1) was isolated from a size selected library constructed in GT10 (Huynk et al., 1985; Blattner et al., 1978). The presence of an L1Hs element in the 3.2 kb EcoRI fragment was detected by hybridization with the 1.4 kb BamHI-Hind III fragment from pRK20 (Adams et al., 1980).

Fragments which had been ligated by standard techniques (Maniatis et al., 1982) into either pEMBL (Dente et al., 1983) or M13 (Messing and Viera, 1982) phage vectors were sequenced by the enzymatic chain termination method (Sanger et al., 1977). Both universal primers and L1-specific oligonucleotides were used. Several of the L1-specific primers were supplied by T. Fanning and M. Singer. Sequences were analyzed using commercially available software (Queen and Korn, 1984).
A. Restriction endonuclease analysis

Figure 2.4 shows a Southern blot of genomic ENAs from family JH-25 digested with EcoRI and probed with a 0.4 kb Bgl II genomic DNA fragment which includes exon 11 of the factor VIII gene. An abnormal 3.2 kb EcoRI fragment was detected in individuals from four generations of this family including the patient with severe hemophilia A (factor VIII activity <1 unit/dl). The maternal grandfather (individual II-2) also had the abnormal fragment. His factor VIII activity was 66 units/dl (normal = 62 to 224 units/dl) and his factor VIII antigen level was 135 units/dl (normal = 66 to 224 units/dl). The presence of normal factor VIII levels and the abnormal 3.2 kb EcoRI fragment in a male family member indicates that this mutation in the factor VIII gene is not the cause of hemophilia A in the patient.

Restriction enzyme analysis of the patient's DNA indicated that the observed alteration was due to the insertion of approximately 700 bp of DNA between the EcoRI sites that flank exon 11 (data not shown). This insertion was not seen in any
Figure 2.4. Family JH-25 with an L1 insertion in IVS-10 of the human factor VIII gene. Five μg of genomic DNA from each individual were digested with Eco RI, subjected to electrophoresis and blotted to nitrocellulose. The probe is a 0.4 kb Bgl II fragment which includes approximately 200 bp of IVS-10 5' to exon 11 and 143 bp of exon 11. An insertion of approximately 700 bp into the normal 2.5 kb Eco RI fragment results in a 3.2 kb Eco RI fragment which has been inherited in this family in a Mendelian, X-linked fashion. Individual IV-1 has hemophilia A (indicated by the filled square), individual II-1 does not have hemophilia A (indicated by the open square). The carrier status of females (open circles) in this family cannot be determined.
Figure 2.4
other of the more than 300 hemophilia A patients that have been analyzed in this study. A recombinant phage containing the 3.2 kb EcoRI fragment was isolated from a GT10 library made from the patient's DNA. Hybridization with pRK20 detected the presence of an L1Hs element in the cloned fragment. Oligonucleotides complementary to the 3' end of the L1Hs consensus sequence and to the 5' and 3' boundaries of exon 11 were utilized in the polymerase chain reaction (Saiki et al., 1988) to determine the location of the L1 insertion within the normal 2.5 kb EcoRI fragment. Analysis of the amplification products (data not shown) indicated that the L1 element was inserted into IVS-10 less than 200 bp from exon 11 in the 5' to 3' orientation with respect to the gene (Figure 2.5).

B. Nucleotide sequence analysis

The JH-25 L1 insertion is 677-681 bp in length with a 3' 66 bp poly A tract. Target site duplications of 13 to 17 bp flank the insertion. Ambiguity in the target site length is due to sequence identities between the 5' end of the L1 element and the 3' end of the target site in IVS-10 (Figure 2.5). As was previously reported (Kazazian et al., 1988), the A-rich target site duplications of the JH-27 and JH-28 L1 elements are 12-15 bp and 13-14 bp. The JH-25 L1Hs target site is of similar length and is also A-rich.
The complete nucleotide sequences of the JH-25, JH-27 and JH-28 L1 elements were determined. Figure 2.6A depicts the lengths of the JH L1 sequences relative to the full length human element. Although the JH-25 and JH-27 L1 elements are colinear with the L1 consensus sequence, the JH-28 L1 element is rearranged presumably during reverse transcription. JH-28L1Hs contains two blocks of sequence; nucleotides 4019-5114 and 5115-6150 in a head-to-head arrangement without loss of L1 sequences (Kazazian et al., 1988). A reconstruction of the JH-28 L1 sequence is presented in figure 2.6A and discussed in this manuscript in order to compare its sequence to other L1 sequences. Nucleotide and amino acid sequence comparisons indicate a high degree of similarity between the three elements and between the JH L1 elements and the L1Hs consensus (Figure 2.6B). The majority of L1 elements which comprise the L1Hs consensus sequence differ from the consensus by less than 5% (Scott et al., 1987). Presumably, the variation in percent divergence is a manifestation of the varying ages of the L1Hs elements. It is interesting that the three recently transposed JH L1Hs elements differ from consensus by 1.1% or less (Figure 2.6B). This observation implies that the consensus sequence closely approximates the sequence of a functional L1 element. Figure 2.7 shows the nucleotides at which the JH L1 elements differ from each other and/or the L1Hs consensus. The sequence of cD11, a nearly full length L1 cDNA (Skowronski et al., 1988), is shown at the same positions for comparison. All
three elements include portions of the 3' open reading frame predicted by the LlHs consensus sequence (Scott et al., 1987). The open reading frame of JH-27 LlHs contains the entire region which has sequence similarity to the polymerase domain of reverse transcriptases (Hattori et al., 1986). Within this domain no substitutions were observed at the conserved amino acid residues identified by Hattori et al. (Hattori et al., 1986). Similarly, the ORF of JH-28 LlHs contains about 60% of the reverse transcriptase region without substitutions at conserved amino acid residues (Figure 2.7).
Figure 2.5. Diagram of the Ll insertion in IVS-10 of the human factor VIII gene. The 677-681 bp Ll insertion is flanked by a 13-17 bp target site duplication shown in brackets. The ambiguity in the length of the direct repeat is due to the identity between the first four nucleotides at the 5' end of the Ll element (nucleotides 5470-5473 bracketed with a dashed line) and the last four nucleotides at the 3' end of the target site. The open box represents the Ll element with the arrow indicating the 5' → 3' orientation. The 3' end of the target site duplication is 154 bp 5' to the first nucleotide of exon 11. This Ll element is referred to as JH-25 LlHs to indicate the pedigree in which it was discovered (JH-25) and its species of origin (Hs for Homo sapiens). The factor VIII map is derived from Gitschier et al (Gitschier et al., 1984).
Figure 2.5
Figure 2.6. A) Nucleotide lengths of the JH LI elements. The three LI insertions are represented as solid lines below a schematic diagram of the organization of the human LI element. The rearrangement of the JH-28 LI element has been corrected in this figure (see text). The numbers in the diagram refer to nucleotide positions in the LIHs consensus sequence (Scott et al., 1987). Regions of non-coding sequence are denoted by open boxes. Boxes with diagonal lines correspond to the open reading frames. Nucleotide positions which mark the 5' endpoints of the JH LI elements are indicated. Note that JH-27 LIHs is 3 nucleotides shorter than the length predicted by its 5' endpoint (3788 nt) because there are 6 A residues rather than 9 A residues between nt 3689 and nt 3699 of the LIHs consensus. The variation in the number of A residues in this region has been previously noted in the consensus (Scott et al., 1987). The 3' trailer region (Skowronski and Singer, 1986) begins at the end of the 3' ORF (nt 5943) and ends at the start of the 3' poly A tract (nt 6150) and is indicated by a checked box. The region of sequence similarity with the polymerase domain of reverse transcriptases (RT) is represented by a filled box.

B) Nucleotide and amino acid sequence comparison of JH LI elements. The percent identity between the JH LI insertions and the LIHs consensus is presented in matrix form. Nucleotide comparisons are above the diagonal line, amino acid comparisons are below the line.
Figure 2.6
Figure 2.7. Comparisons of the JH LI sequences to the LIHs consensus sequence. Nucleotide positions at which the JH LI elements differ from each other and/or the LIHs consensus are shown. The location of these nucleotides relative to the organization of LIHs elements is indicated by the diagram at the bottom. The solid line represents the 3' ORF, the dashed line denotes the 3' trailer region. The region of sequence similarity between the LIHs elements and the polymerase domain of reverse transcriptase (RT) is represented by a filled box. The cDNA and Ta subset cDNA consensus for the 3' trailer region are also presented (Skowronski and Singer, 1986). Note that cD11, a nearly full length LI cDNA (Skowronski et al., 1988), matches the cDNA consensus whereas JH-27 LIHs and JH-25 LIHs match the Ta consensus at the four underlined nucleotides which define the Ta subset. JH-28 LIHs matches at three out of four positions. The G residue at nucleotide position 6142 which is present in the JH LI elements is also present in the only two members of the Ta subset which contain that region (M. Singer, unpublished results). Circles indicate nucleotide differences between one or more JH LI elements and the LIHs consensus which alter a codon. Filled circles denote an amino acid replacement, while open circles represent nucleotide substitutions which do not alter the encoded amino acid. Silent nucleotide substitutions occur at 2 conserved amino acids (hatched arrows) and 2 invariant amino acids (closed arrows) within the regions of sequence similarity between LI
Figure 2.7 (cont.) elements and the polymerase domain of reverse transcriptases (Hattori et al., 1986). The other 6 nucleotide substitutions within the RT region result in amino acid changes; however, these changes are in gaps which separate the blocks of RT similarity. The nucleotide changes at positions 4395 and 4396 are bracketed to denote that these two changes occur within one codon.

The complete sequences of JH-25 LIHs, JH-27 LIHs and JH-28 LIHs have been deposited with GenBank.
Figure 2.7
Two previously reported L1 insertions into exon 14 of the factor VIII gene have proven to be deleterious by disruption of coding sequences (Kazazian et al., 1988). The discovery of an insertion into IVS-10 of the factor VIII gene raises the possibility of deleterious mutation by insertion into an intron of a eukaryotic gene. L1HS elements have a potential acceptor splice site (Shapiro and Senapathy, 1987) from nt 5835 to nt 5850 (TTCATGCGTTTGTAGG) of the L1HS consensus sequence (Scott et al., 1987). If an L1 element containing this acceptor splice site were incorporated into an RNA transcript in the functional orientation, any incorrectly spliced RNA molecules would contain premature stop codons. The JH-25 L1 insertion places the potential L1 acceptor splice site 538 bp 5' to the natural acceptor site of exon 11 and in the 5' to 3' orientation with respect to the factor VIII gene. This seemed to be a reasonable hypothesis since the maize transposable elements Spm and Ds have been shown to bring in cryptic splice sites when they insert (Kim et al., 1987; Wessler et al., 1987). However, the presence of the insertion in the unaffected maternal grandfather indicates that incorrect
splicing either does not occur or occurs so infrequently that it is not biologically significant. LL elements are normally present in the introns of functional genes, such as the factor VIII gene (J. Gitschier, personal communication, and PWS et al., unpublished observations) and the dystrophin gene (Monaco et al., 1987). At present there is no evidence that insertion of LL elements into the introns of functional genes is associated with disruption of gene expression. LL cDNAs have been isolated from a human teratocarcinoma cell line, NTera2D1 (Skowronski et al., 1988). Of the 19 cDNAs isolated no two cDNAs are identical (Skowronski et al., 1988). Skowronski et al. (Skowronski et al., 1988) have found subsets of related sequences in their cDNAs by comparison to a consensus sequence of the 208 bp 3' trailer region (Figure 2.6A). When the cDNA consensus is compared to the LLHs consensus (Scott et al., 1987), 16 nucleotide differences are observed among these 208 bp (Skowronski and Singer, 1986). Four of the 19 cDNAs (defined as the Ta subset) had four additional nucleotide changes which were unique to them (Skowronski and Singer, 1986) (Figure 2.7). The 3' trailer sequences of JH-27 LLHs, JH-25 LLHs and the cDNA Ta consensus are identical. JH-28 LLHs differs from the other two JH LL elements and the cDNA Ta consensus at nucleotide 6059 which is one of the nucleotides defining the Ta subset (Figure 2.7). The JH LL elements, therefore, are closely related to members of the Ta subset of LL cDNAs.
Fanning and Singer (Fanning and Singer, 1987a) have proposed two classes of retrotransposons. Class I retrotransposons, such as copia in Drosophila (Mount and Rubin, 1985) and Ty in yeast (Boeke et al., 1985), are more similar to retroviruses in that they possess long terminal repeats (LTRs) and generate constant length target site duplications. The L1 elements of mammals as well as the L1-like elements, ingi in Trypanosoma brucei (Kimmel et al., 1987), R2 in Bombyx mori (Burke et al., 1987), and the F, I, G and jockey elements of Drosophila (DiNocera and Casari, 1987; Fawcett et al., 1986; DiNocera et al., 1986; Priimagi et al., 1988) are members of the Class II retrotransposon family. Class II retrotransposons apparently transpose through an RNA intermediate because they have 3'poly A tracts. Class II elements and other elements thought to move by retrotransposition, including Alu sequences (Lin et al., 1988) and processed pseudogenes, lack both LTRs and target site duplications of specific lengths. However, Class II retrotransposons contain open-reading frames (ORFs) which may encode proteins necessary for their transposition. This characteristic differentiates them from Alu sequences and processed pseudogenes (Schmid and Jelinek, 1982; Temin, 1985; Vanin, 1985).

Most L1 elements are thought to be nonfunctional because of mutations in their coding regions, truncation, deletion, or rearrangement (Fanning and Singer, 1987a; Scott et al., 1987). If a functional element is defined as having a transcribable
and translatable reading frame like an active gene, then reintegrated copies of functional elements should have undisrupted 5' and 3' ORFs. JH-27 LIHs, JH-28 LIHs, and JH-25 LIHs have open reading frames of 1192aa, 642aa and 157aa, respectively (Figure 2.6A). There are 42 nucleotide positions in the 3' ORF at which the LIHs consensus sequence differs from one, two, or three of the JH LI elements. At 24 of these 42 positions the nucleotide substitution in the JH LI element(s) does not result in a change in the encoded amino acid (Figure 2.7). Within the region containing sequence similarity to the polymerase domain of reverse transcriptases, 4 of the 10 changes are conservative with regard to encoded amino acid. All 4 silent nucleotide substitutions are within the conserved blocks of amino acid homology (Hattori et al., 1986), while the 6 replacement substitutions occur in the regions between these blocks. Coding sequence conservation and the maintenance of open reading frames are characteristics of functional genes and as such suggest that the recently inserted JH LI elements represent the reintegrated transcripts of a set of functional elements.
LIST OF REFERENCES FOR PART 2


Kimmel, B. E., Ole-Moijoi, O. K. and Young, J. R. (1987). Ingi, a 5.2 kb dispersed sequence element from Trypanosoma brucei that carried half of a smaller mobile element at either end and has homology with mammalian LINES. Mol. Cell Biol. 7:1465-1475.


