1997

Characterization of Deletion Sites in a Human Collagen-Like Locus

Marissa A. Michaels
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

CHARACTERIZATION OF DELETION SITES IN A HUMAN COLLAGEN-LIKE LOCUS

A THESIS SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

DEPARTMENT OF BIOLOGY

BY

MARISSA A. MICHAELS

CHICAGO, ILLINOIS

JANUARY 1997

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jeffrey Doering, for all his support and guidance. I also thank my committee members, Drs. John Smarrelli and William Wasserman for their suggestions and comments. I would also like to thank everyone in the Doering lab especially Anne Burket, Sheryl Cammarata, Jennifer Doll and Irnel Wiese for their help and camaraderie.

I am especially indebted to my uncle Dr. Theodore Puck who believed in me and supported my endeavor to follow my true calling. I would also like to thank my mother for everything she has done for me. You made me believe that true desires are worth the fight. I would never be where I am today without your support and guidance, and appreciate all you have done for me.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Osteogenesis imperfecta (OI) is a heterogeneous genetic disorder that affects bone and other connective tissue (Sillence et al., 1979a; Sillence et al., 1979b; Prockop and Kivirikko, 1984a). It is characterized by fragility of bones, blue sclerae, thin skin, joint hypermobility, hearing impairment and dentinogenesis imperfecta (Sillence 1988; Byers 1989; Kuivaniemi et al., 1991). All of the affected tissues are rich in type I collagen (Byers 1990; Sillence 1988; Kuhn 1987). Molecular defects in the structure and synthesis of Type I collagen have been found in many individuals with the OI phenotype (Prockop et al., 1989; Prockop et al., 1994).

Severity of the OI symptoms varies greatly between individuals and very rarely have two cases been found with exactly the same molecular defect (Pope et al., 1985; Rose et al., 1994). There are reports of many cases where asymptomatic individuals carry the same Type I collagen defect as those family members with the OI phenotype (Superti-Furga et al., 1989; Cohn and Byers, 1991; Sykes et al., 1990). There have even been cases of the progressively-deforming, moderately severe form of the disease (Type III)

that show no apparent Type I collagen defects (Byers et al., 1988b; Wenstrup et al., 1990). Thus additional genes must contribute to the OI phenotype other than the Type I collagen genes.

Previous studies have shown the existence of two undescribed collagen-like loci that are different from other already-characterized collagen genes (Cammarata et al., 1991; Doering et al., 1993; Cammarata, 1994). Site 1 is detected as a 7.4kb Hind III restriction fragment which contains a 3.0kb Pst I fragment in normal individuals. Site 2 is detected as a 4.6kb Hind III fragment that overlaps with a 5.2kb Pst I fragment in normal individuals. This site has recently been cloned and localized to chromosome 17 (Doering et al., 1993; Cammarata, 1994). Deletions at the two sites occur more frequently in OI Type III patients than the normal population (Cammarata et al., 1991; Doering et al., 1987). At both of the new sites the deletions are of heterogeneous size and occur spontaneously (Doering et al., 1987). This data for OI patients, when compared to that for normal individuals demonstrates that the frequent occurrence of deletions at these loci <u>is</u> associated with Type III OI. The new loci may code for a component(s) of connective tissue that can contribute to the OI phenotype if defective, or may interact with Type I collagen defects to produce the OI phenotype.

Tandemly-repeated sequences have been shown to be responsible for disorders such as fragile X syndrome and myotonic dystrophy, where length polymorphisms occur as a result of variable numbers of copies of repeats (Caskey et al., 1992). Also, a locus containing uncharacterized collagen-like sequences has been shown to have substantial length polymorphism arising from variable numbers of tandem repeats (Brookes et al., 1989). These studies suggest that tandemly repeated sequences are inherently unstable and prone to further deletion. Variability in the number of copies of the repetitive sequence are what cause such heterogeneity among individuals.

The purpose of this study was to localize and determine the molecular organization of the deletion site at Site 2. Specifically I wished to determine whether tandemly repetitive sequences are responsible for this deletion. Restriction enzymes with sites in the locus were used to digest patient and normal DNAs in order to localize the deletion site. Comparison of the fragment sizes obtained from the normal to those from the patient DNA delineated this localization. The regions surrounding the deletion site in the clone were then sequenced to determine their characteristics. Sequences adjacent to the deletion site that were non-repetitive and non-complementary were used to construct DNA primers for PCR reactions. The polymerase chain reaction was utilized to amplify the patient DNA to

allow sequencing of these regions. The products of the PCR reaction were electrophoresed on agarose gels, and since heterozygotes were used, two different-sized allele fragments were separated from each other and purified by band-stab PCR directly from the gel. The PCR-amplified Site 2 fragment was then directly sequenced using Cycle-Sequencing methods. The characterization of the deletion in the patient was performed by comparing the sequence of this allele to that of the normal allele. This identified deletion endpoints, and indicates that tandemly repetitive sequences are involved in the mechanism causing the deletions.

CHAPTER II

REVIEW OF RELATED LITERATURE

OI and Type I Collagen

Osteogenesis imperfecta (OI) is a heritable connective-tissue disorder in which bone fragility is the main clinical manifestation (Cole et al., 1989; Byers 1989; Bateman et al., 1991). It is one of the most common connective tissue disorders and occurs with a frequency of approximately 1 in 5,000 to 10,000 individuals (Byers and Steiner, 1992). The severity of the symptoms vary and patients may also exhibit blue sclerae, hearing loss, thin skin, joint hypermobility and dentinogenesis imperfecta (Sillence, 1988; Byers, 1989). Since all the tissues affected by OI are rich in Type I collagen (Kuhn, 1987; Sillence, 1988; Byers, 1989), most studies of the disease have focused on the structure of this protein. Many cases of OI have been shown to involve alterations in the structure and/or synthesis of Type I collagen, but only rarely are two cases found with exactly the same defect (Rose, 1994). The phenotype is caused by mutations in either of the genes encoding the $\alpha 1$ or $\alpha 2$ chains of Type I

collagen, COL1A1 or COL1A2 (Rose et al., 1994; Forlino et al., 1994).

Type I collagen is the major protein of mature bone, skin, tendons and ligaments (Kuhn, 1987; Sillence, 1988). Collagens are extracellular proteins composed of three helical polypeptide chains that associate with one another to form a right-handed triple helical structure (Prockop and Kivirikko, 1984a). Type I collagen is a heterotrimer composed of two α1 chains and one α2 chain. Both types of chains are synthesized as procollagen molecules consisting of a triple helical segment flanked by amino and carboxy terminal propeptides. The triple helical domain is characterized by a regular Gly-X-Y repeating unit, where X is often proline and Y is often hydroxyproline, although they can be any amino acid except cysteine, tyrosine or tryptophan (Kuhn, 1987).

Collagen Formation

Helix formation begins at the carboxy end of the chains and proceeds in a zipper-like fashion toward the amino terminus. The glycine residue in every third position facilitates the folding of the three chains into the triple helical conformation, which allows tight

association of the polypeptide chains. During the helix assembly in the rough endoplasmic reticulum, proline and lysine residues are hydroxylated and some prolyl residues are further glycosylated. These post-translational modifications continue until the completion of helix formation (Byers, 1989; Prockop, 1984b), so the extent of modification depends on the rate of helix formation (Bateman et al., 1988). Propeptides are enzymatically removed by proteases after secretion from the cell into the extracellular matrix (ECM). This allows the molecules to spontaneously align into mature collagen fibrils which are stabilized by covalent lysyl-derived crosslinks (Bornstein and Sage, 1980; Kuhn, 1987). The molecular crosslinking also occurs in the mature molecule via disulfide bonding between cysteine residues.

A compact, rigid triple helix is essential for tensile strength in mature collagen fibers (Byers, 1989). Collagen polypeptides are arranged in parallel stacks, and crosslinked to form staggered microfibrils (Prockop, 1984). Because post-translational modification is dependent on the rate of helix formation, mutations in the Type I collagen genes that result in a decrease in the rate of triple helix formation often result in overmodification (Bateman et al., 1986; Bonadio and Byers, 1985). Post-translational modification and intermolecular crosslinking are both important in stabilizing the molecular arrangement

within collagen molecules and between fibrils (Kuivaniemi et al., 1991). In bone, collagen fibrils are densely packed. If a defective molecule is incorporated into the uniform fibril environment, the strength of the tissue will be compromised. It causes not only instability in that molecule, but all the adjacent molecules are affected. It would be less detrimental to produce fewer normal collagen chains than to produce and incorporate a large number of defective ones.

Type I Collagen Defects in OI

As a result of biochemical, molecular genetic, and linkage studies, there is now little question that the vast majority of patients with OI are heterozygous for mutations in Type I procollagen (Sykes et al., 1990; Byers et al., 1988c; Wenstrup et al., 1990). More than 70 different mutations in the two genes that encode the chains of Type I collagen, the major protein of bone, have been identified as the molecular cause of different forms of OI (Byers et al., 1991). The mildest forms of OI generally result from overall production of Type I procollagen being synthesized being decreased due to reduced synthesis of the pro α1(I) chain (Bateman et al., 1988). These defects may be due to non-functional COL1A1 alleles or ones that produce unstable products (Willing et al., 1992). Mutations in Type I collagen genes which result in decreased rates of triple

helix formation, cause overmodification of the protein chains (Bonadio and Byers, 1985) which interferes with the folding, processing and formation of stable collagen molecules. These unstable molecules are often rapidly degraded and result in reduced amounts of collagen (Barsh and Byers, 1981). There are a number of cases that have overmodified chains due to mutations in the sequence of the triple helical domain (Bateman et al., 1988; Deak et al., 1991; de Vries et al., 1986). The mutations include point mutations, insertions, deletions and re-arrangements in the COL1A1 gene that codes for the pro-α1(I) chain and in the COL1A2 gene that codes for the pro-α2(I) chain of Type I collagen (Byers, 1990; Bateman et al., 1991).

The most commonly studied gene defects in OI are point mutations which cause a substitution for a Gly residue in the triple helix (Wallis et al., 1990; Cole et al., 1992). These substitutions can also cause splicing defects which can result in missing coding information (Kuivaniemi et al., 1988). Deletions can also cause splicing defects (Tromp and Prockop, 1988). Occasionally insertions have been observed that are responsible for frameshift mutations (Willing et al., 1990; Genovese at al., 1989). While the detailed nature of the structural defects in the collagen protein have been characterized in an increasing number of OI cases, only occasionally have two cases been found to have exactly the same biochemical defect (Rose et al.,

1994; Prockop et al., 1994).

OI Types

Osteogenesis imperfecta is classified into four different types based on clinical symptoms and the mode of inheritance (Sillence et al., 1979b; Byers et al., 1987; Prockop and Kivirikko, 1984a). OI Type I is a mild dominantly inherited form (Sillence et al., 1979a; Byers et al., 1988c; Prockop, 1984b) that involves defects in the structure or synthesis of Type I collagen. The clinical features associated with Type I OI are bone fragility which is reduced at puberty, blue sclerae, deafness and short stature (Sillence, 1988). The triple helix in Type I cases of OI are as stable as normal collagen or only moderately unstable, and in some cases the helix stability is explained by Type I mutations occurring in the X or Y position of the Gly-X-Y sequence rather than in the glycine position as occurs in Type II OI (Steinmann et al., 1986). This may account for the milder phenotype (Bateman et al., 1988a). There have been Type I OI patients found with mutations that cause splicing defects with exon skipping in the helical domain of the COL1A1 gene. The mild phenotype may be a result of the deleted molecules not being incorporated into the matrix (Superti-Furga et al., 1993; Mottes et al. 1994). In many OI Type I cases, even when no structural defects have been identified, the total amount

of Type I procollagen synthesized is low, due mainly to reduced synthesis of the pro-α1(I) chain (Byers et al., 1988c; Bateman et al., 1988b). Presence of non-functional COL1A1 alleles may be the cause of these defects (Willing et al., 1992; Rowe et al., 1985). By analysis of restriction length polymorphisms (RFLPs), the phenotype has been shown to consistently segregate with one or the other Type I collagen genes (COL1A1 or COL1A2) (Tsipouras et al., 1984; Falk et al., 1986; Sykes et al., 1990). These results clearly show that the defects in the structural genes for Type I procollagen are responsible for this form of OI.

OI Type II is the severe perinatal form that has been well studied. Its clinical characteristics include neonatal dwarfism, severe bone deformity as a result of multiple fractures and death *in utero* or shortly after birth (Sillence, 1988). Most cases are the result of new dominant mutations in the Type I collagen structural genes which affect helix formation (Byers et al., 1988a; Byers et al., 1988b). Mutations have been found in both the COL1A1 and COL1A2 genes which results in substitutions for one of the glycines of the triple helix. The new substituted residue is often cysteine or arginine (Vogel et al., 1987; Bateman et al., 1988b) but valine (Patterson et al., 1989), serine (Lamande et al., 1989) and aspartate (Baldwin et al., 1989) substitutions have also been seen. The

substitutions are found at a number of positions throughout the chains (Lamande et al., 1989; Baldwin et al., 1989; Bateman et al., 1988b). These results indicate that mutations associated with the Type II phenotype are not located in only particular positions of the helix.

Many cases appear to be point mutations that cause a disruption in the regular Gly-X-Y sequence in the triple-helical domain which leads to an unstable helix that is slow in forming and results in poor secretion of the unstable procollagen molecule into the extracellular matrix (Willing et al., 1988; Cohn et al., 1988b). Since modification of collagen chains occurs until a stable triple helix is formed, Type I collagen chains usually contain excess post-translational modification as a result of the slowly forming helices in these cases (Byers et al., 1988; Byers et al., 1988; Bonadio and Byers, 1985). Other mutations such as genomic deletions or point mutations in the helical portion of COL1A1 (Barsh et al., 1985; Chu et al., 1985; Hawkins et al., 1991) or COL1A2 (Willing et al., 1988) can lead to splicing defects and missing exons (Tromp and Prockop, 1988).

OI Type III is the progressively-deforming, moderately severe form of OI. It is deemed progressive in nature due to repeated multiple fractures that result in severe deformity of the limbs and spine. Even though it is one of the most common types, it is the least characterized

(Sillence, 1988). A few autosomal recessive cases exist, yet most are due to spontaneously-occurring dominant mutations (Thompson et al., 1988). RFLP analysis of several recessive pedigrees for Type III OI have shown that this form is unlinked to **either** of the Type I collagen genes (Byers et al., 1987; Aitchison et al., 1988). A few reports exist of substitutions and deletions in the Type I collagen genes of OI Type III patients (Pihlajaniemi et al., 1984; Byers et al., 1983a; Pack et al., 1989). There are also many cases where no such defects have been found (Byers et al., 1988a; Beighton et al., 1988; Wenstrup et al., 1990). For these cases where no Type I collagen mutation is present, defects in genes encoding bone-specific proteins or other ECM components could be responsible for expression of the OI phenotype (Byers, 1990).

OI Type IV is the more severe dominantly inherited form that also involves a variety of Type I collagen defects (Sillence, 1988). Its clinical severity is moderate, with patients having pale blue or white sclerae and variable deformity of the long bones (Sillence, 1988). Amino acid substitutions at glycine positions in the triple helical region of the $\alpha 2(I)$ chain appear to be the basis for many Type IV cases (Byers et al., 1988c; Wenstrup et al., 1988). RFLP analysis has shown that in many cases this form of the disease segregates with the COL1A2 gene

(Superti-Furga et al., 1989; Sykes et al., 1990). The kinds of defects found in Type IV are similar to those seen in Type II OI. This shows that the different forms of OI cannot be distinguished by the collagen chain in which the mutation occurs or even by the position of the mutation in the chain.

It has not been possible to correlate specific defects with particular OI phenotypes because there is still little understanding of how defects in Type I collagen structure cause the various types of OI. Despite the large number of studies of OI cases at the molecular level, the different forms of the disease cannot be distinguished by which of the two Type I collagen chains contains the mutation. Even when mutations map very closely to each other in the same chain, they can be associated with OI phenotypes of widely varying severity (Cohn and Byers, 1991). There is very little understanding of the functional domains of the collagen molecule, which may add to the confusion in determining correlations between defect and phenotype (Sykes et al., 1990). There are many cases where members of a family carry the <u>same</u> Type I collagen defect and show a wide variation of severity, even being asymptomatic (Sykes et al. 1990; Superti-Furga et al., 1989). Thus, the variation in symptoms may involve other collagen types or matrix proteins that interact with the various domains of Type I collagen and thus affect the phenotype.

Collagen Types

At least seventeen kinds of collagen besides Type I have been identified, but little is known about how they contribute to the structural integrity of the extracellular matrix (Mayne and Burgeson, 1987; Sandell and Boyd, 1990; Shaw and Olsen, 1991; von der Mark et al., 1992). They can be classified into three main groups depending on their structure.

Group one consists of the fibrillar collagens (Types I, II, III, V, and XI). All members of the group have similar structural features of being homo or heterotrimers. They all contain three polypeptide chains, with a monomer size greater than 95 kiloDaltons, and long uninterrupted triple-helical domains containing the characteristic Gly-X-Y repeating amino acid sequence. Most of the exons in the helical coding regions of these genes are 54bp or multiples of 54, and all of the exons in the helical coding regions are multiples of 9bp in length, corresponding to integral numbers of the basic Gly-X-Y repeat, beginning with the codon for glycine and ending with a triplet preceding a glycine codon (Sandell and Boyd, 1990; de Wet et al., 1987; von der Mark et al., 1992). The genes for these collagens share a highly conserved exon-intron organization. The coding sequences are divided into nearly identical numbers of exons, and all introns regardless of size, are located in homologous positions in each gene (Sandell and Boyd,

1990; Fleischmajer et al., 1990).

Type III collagen is often present in minor amounts in the same fibrils with Type I. Small amounts of Type V collagen are also associated with Type I fibrils and occasionally with Type II fibrils (Mayne and Burgeson, 1987). Type V and XI are more similar to each other structurally than any other two collagens (Greenspan et al., 1991). Type XI is expressed in all cartilage tissue and may be associated as a minor component with the Type II fibrils that make up those tissues (Mayne and Burgeson, 1987).

Group two consists of the large nonfibrillar Types IV, VI and VII. They all have monomer chain sizes of greater than 95 kiloDaltons with large globular domains, and a number of interruptions in the repeating Gly-X-Y repeat within the triple helical domains (Mayne and Burgeson, 1987; Sandell and Boyd, 1990; Parente et al., 1991). These interruptions allow the molecules flexibility which is not possible in the fibrillar collagens (Hoffman et al., 1984). Types IV and VI are heterotrimers, while Type VII is believed to be a homotrimer (Fleischmajer et al., 1990). This group is not well characterized; only the genes for the two chains of Type IV and three chains of Type VI have been studied in detail. Their exon sizes and overall intron-exon patterns are completely different from the 54bp exon pattern of the fibrillar collagens (Soininen et al.,

1988; Soininen et al., 1989). These collagens aggregate through end-to-end associations of monomer units rather than by lateral association (Timpl et al., 1981).

The function of Type IV collagen in the matrix is not understood. It is found in basement membranes and interacts weakly with cell membranes, laminin and proteoglycans. Type VI collagen may serve as a link between other matrix components and cells, yet does not appear to be physically associated with the major fibrillar elements in extracellular matrices. Type VII seems to link the basal surface of epithelial cells with the dermis as part of the anchoring fibrils, but the biochemical basis for this is unclear (Mayne and Burgeson, 1987; Fleischmajer et al., 1990).

Group three consists of the short-chain Types VIII, IX, X, XII, XIV and XVI. These collagens all have monomer chains of less than 95 kiloDaltons with many interruptions in the Gly-X-Y repeat of their triple helical domains (Sandell and Boyd, 1990). This collagen group shows great variation in gene structure. Two chains of Type VIII and one of Type X have been characterized, and they show common structural features, that consist of 8 interruptions of the helix with similar sequence, length and relative location (Yamaguchi et al., 1989; Muragaki et al., 1991). Both Types IX and X contain short collagenous domains, yet the 5kb gene for pro-α1(X) collagen contains only 3 exons and

no introns in its triple helical coding sequences (Sandell and Boyd, 1990). The genes for pro-α1(IX) and pro-α2(IX) have been partially characterized and only a few 54bp exons have been found in either of them (Sandell and Boyd, 1990; Ninomiya et al., 1990). Type VIII collagen is found in basement membrane of corneal endothelial cells and Type X in the endochondral ossification zones of cartilage (Muragaki et al., 1991; Reichenberger et al., 1992).

Types IX, XII, XIV and XVI have been grouped into a subgroup called FACIT collagens (fibril-associated collagens with interrupted triple helices) to emphasize their homologous structures and possible function (Shaw and Olsen, 1991). The molecules in this class contain more than one triple-helical domain separated by non-triple-helical segments (Shaw and Olsen, 1991). The FACIT proteins are associated with fibrils composed of fibrillar collagens. FACITs all share structural features, yet their sizes and primary structures vary greatly (Pan et al., 1992). They may be important in the organization and stability of extracellular matrices.

Type IX collagen is found in cartilage and is arranged in a periodic fashion along the surface of the major fibrils made up of Types II and XI (Shaw and Olsen, 1991). The globular domain of Type IX projects out of the surface of the major fibril possibly facilitating interactions between the major fibril and other molecules of the matrix

(Shaw and Olsen, 1991). Type XII, with a structure homologous to that of Type IX, is found primarily in tissues where Type I is the major fibrillar collagen. Although Type XII co-localizes with Type I in many tissues, it is **not** found in skin or bone (Shaw and Olsen, 1991). Another recently discovered collagen, Type XIV, is found in skin and, since it is homologous to Types IX and XII it may interact with the major fibrils composed of Types I and V (Shaw and Olsen, 1991; Dublet and van der Rest, 1991). Both types XII and XIV localize near the surface of banded collagen fibrils and influence interaction between adjacent fibrils as-well-as participate in extracellular matrix deformability (Nishiyama et al., 1994). Type XVI collagen is made by fibroblasts and keratinocytes (Pan et al., 1992). It has structural similarities to Types IX and XII (Pan et al., 1992).

New collagen loci continue to be discovered and characterized. The newly identified Type XIII gene has a helical-coding domain made up of several 54bp exons and a number of other exons with unrelated sizes (Tikka et al., 1991). This gene has both fibrillar and non-fibrillar gene characteristics. The collagen structure it codes for resembles that of the FACIT collagens. More new collagen types include XV, XVII and XVIII that are characterized by many short triple helical domains, unlike the long, contiguous domains found in the other collagen types. Type

XV is found in kidney and pancreas and Type XVII is found in hemidesmosomes, yet the function of these collagens is still unknown (Muragaki et al., 1994). BPAG2 is yet another new collagen and is found in stratified squamous epithelia. It contains both collagenous and non collagenous domains. Its exon organization is different from any other fibrillar or nonfibrillar collagen genes (Li et al., 1991).

Newly-Identified Collagen-Like Loci

Recent studies have identified two undescribed collagen-like loci that are different from other already characterized collagen genes (Doering et al., 1993). The loci were detected with a pro-α2(I) cDNA probe (Doering et al., 1993). The normal restriction fragments containing Site 1 are a 7.4kb Hind III that contains a 3.0kb Pst I fragment. Site 2 is detected in normal individuals as a 4.6kb Hind III fragment that overlaps a 5.2kb Pst I fragment and is localized to chromosome 17 (Doering et al., 1993). This site has been cloned from a chromosome 17-specific lambda phage library and a restriction map has been constructed (Cammarata et al., 1991). The 4.6kb HindIII fragment was isolated and the triple helical coding region within it was characterized by restriction mapping and sequencing (Cammarata, 1994). Determining the detailed molecular organization of this locus is the inital step

necessary to characterize the structure, and function, of the putative protein that is coded for by the new locus.

Recent studies of the new collagen-like loci suggest that deletion sites could be contributing to the OI phenotype. Deletions that are heterogeneous in size and occur in a spontaneous fashion have been detected at these sites in genomic DNA of Type III OI patients (Breslin et al., 1995; Cammarata et al., 1991; Doering et al., 1987). While these deletions do occur in normal individuals they are much less frequent there and less heterogeneous in size (Doering et al., 1990; Cammarata et al., 1993). Because of a very low frequency of heterozygote deletions at Site 2, the alleles there are not in Hardy-Weinberg equilibrium. This is important because it suggests that Site 2 is at a locus that is expressed and that its genotype may result in a disadvantageous phenotype that selection will operate against. The new loci may code for a connective tissue component(s) which, when defective, contributes to the OI phenotype (Doering et al., 1993). It is suggested that the disease phenotype requires an interaction between Type I collagen and the products of other loci, as suggested by the many examples of asymptomatic individuals carrying the same Type I collagen as their family members who have OI (Cohn and Byers, 1991; Constantinou et al., 1989; Hollister et al., 1988). The deleted alleles at Site 2 may not cause the disease directly, but the defective product of the

locus may interact with other mutant connective tissue components to produce the OI phenotype. The deletions in patients then may act independently or by modulating the effect of Type I collagen defects.

Length polymorphisms where insertions/deletions change the length of DNA loci have been shown to be caused by various kinds of repetitive sequences (Sykes et al., 1985; Lee et al., 1991; Sykes et al., 1990). Tandemly-repeated sequences normally cause the length polymorphisms (Caskey et al., 1992). Direct repeats have also been shown to be involved in causing deletions within collagen genes (Chu et al., 1985; Sykes et al., 1985; Hawkins et al., 1991). Fascioscapulo-humeral muscular dystrophy (FSHD) has recently been shown to be associated with deletions in a tandemly repeated sequence (van Deutekom et al., 1993; Winokur et al., 1994). In fragile X syndrome, myotonic dystrophy and Huntington's disease, variable numbers of tandemly-repeated sequence cause longer polymorphisms in the alleles of affected individuals. There is also a range of premutation alleles in a range between that of normal and affected individuals seen in individuals with these diseases (Caskey et al., 1992; Sutherland and Richards, 1995). These alleles may cause mild or no disease symptoms in the individual carrying them, but long alleles arise spontaneously in their affected children (Caskey et al., 1992). The parental deletions we have observed in OI

families may give rise to further spontaneous deletions in their children, which then could contribute to the OI phenotype. The mechanism responsible for the instability and frequent deletions at Site 2 may be similar to that of these other diseases that are associated with unstable tandem repeats.

By determining the molecular organization of deletions at Site 2 in OI patients it may be possible to determine the mechanism responsible for the frequent deletions. The deletion location in Site 2 was determined and then sequenced in deleted and undeleted alleles to define deletion sizes, endpoints and the involvement of tandemly repeated sequences. These studies will ultimately provide understanding as to how the abnormalities caused by the deletions contribute to the OI phenotype.

CHAPTER III

MATERIALS AND METHODS

Experimental Approach

The purpose of this study was to localize and determine the molecular organization of the deletion site in one of the new collagen-like loci (Site 2). To characterize the deletions in the Site 2 locus (4.6kb HindIII fragment), the location of the deletion was determined. The Site 2 clone was used to probe blots of restriction digested normal and patient DNAs, to approximately localize the deletions within the 4.6kb site. Sequencing of the cloned normal locus in the region where deletions occur was done to determine the nature of the sequence. Primers for PCR reactions were then chosen from non-repetitive, non-complementary sequences in this region. PCR was then done on patient DNAs to isolate the deletion region and to allow its sequencing. The molecular nature of the deletions was then determined by comparing the sequence of a short allele in an OI patient to the normal allele. This was done in order to determine the deletion endpoints and possible mechanism(s) responsible for the deletions.

Restriction Enzyme Digests

To localize the deletion site within the locus, DNA was restriction enzyme digested by established methods using single and double enzyme combinations (Doering, 1977; Peterson et al., 1980; Rosenthal et al., 1984). Restriction enzymes with sites in the 4.6kb (Site 2) clone were used to digest DNAs from a normal individual and from a patient with a homozygous 200bp deletion at Site 2. A comparison of the fragment sizes obtained in the control to that of the patient was done to localize the deletion. The smallest region where the fragment sizes are different between the normal and patient allowed this localization. All restriction enzymes were purchased from Bethesda Research Laboratories (Bethesda, MD) and used in the supplier's recommended buffers. Restriction digests were stopped by adding 0.5% SDS and 10mM EDTA. Gel electrophoresis was carried out on 0.9% agarose gels (Rosenthal and Doering, 1983) and the DNA transferred to a nylon membrane (Gene Screen Plus hybridization membranes; NEN Research Products, Boston, MA) using the alkaline transfer method (Reed and Mann, 1985). The DNA size markers used were lambda phage DNA digested with HindIII and ϕ-X 174 phage DNA digested with HaeIII.

Nick Translation and Hybridizations

The method for locating the deletion involved using the Site 2 radiolabeled clone as a probe to compare restriction fragment sizes obtained in a normal control to that in a patient with the deletion. The EcoRI/SstI fragment of the Site 2 clone P4.6 (Cammarata, 1994), which contains no repetitive sequences was ^{32}P-labeled by nick translation (Rigby et al., 1977). Prehybridization of the membrane was carried out for at least 5 hours at 37°C with agitation (Doering et al., 1982; Rosenthal and Doering, 1983) in 40ml of hybridization solution (50% formamide, 0.9M NaCl, 50mM Tris pH 7.5, 1% sodium dodecyl sulfate, 10μg/ml denatured E. coli DNA). Denatured nick translated probe (6ng/ml) was added to the prehybridized membranes and incubation proceeded at 37°C for a minimum of 16 hours with agitation. Membranes were washed twice at room temperature for 30 minutes in 2X SSC (1X SSC is 150mM NaCl, 15mM sodium citrate, 0.1 mM EDTA); twice at 60°C for 30 minutes in 2X SSC with 1% SDS; and lastly, twice at room temperature for 30 minutes in 0.5X SSC. Autoradiography was then done using Kodak XAR x-ray film and Dupont Lightning-Plus intensifying screens as necessary.

DNA Sequencing

To determine the sequence in the deletion site, DNA sequencing of the Site 2 subclone P4.6ME (Cammarata, 1994) was done by the dideoxy method (Sanger et al., 1977) using the "AmpliTaq" kit (Perkin Elmer Cetus, Norwalk, CT). The double-stranded template and primer were denatured and reannealed as follows prior to extension/termination reactions. Three micrograms of template DNA and 0.5pmol primer DNA were combined in a total volume of 20.0μl. NaOH and EDTA were added to a final concentration of 0.2M and 0.2mM respectively, and the mixture was incubated for 5 minutes at 85°C. The tube was placed on ice, and 6.0μl of 3M NaOAC, pH 5.2 and 60μl 100% EtOH were added to precipitate the denatured DNA. This solution was placed at -20°C for 20 minutes. The DNA was pelleted by centrifugation at 9,500rpm in an SHMT rotor (Sorvall/E.I. duPont, Newtown, CT) for 20 minutes and washed carefully with 70% EtOH. The dried DNA pellet was resuspended in 14μl distilled H_2O and reannealed for 5 minutes at 37°C followed by 7 minutes at room temperature.

The extension and termination reactions were carried out according to the manufacturer's instructions for sequencing double stranded DNA using the "Amplitaq" sequencing kit (Perkin Elmer Cetus, Norwalk, CT). The DNA was tagged by adding ^{35}S-labelled dATP nucleotides during the extension reaction. The T7 primer was used and

extension began from the Eco RI site, or in the case of the nested deletion clones, sequences adjacent to this site, and proceeded to the left through the helical coding region.

The sequence that was determined was highly repetitive and internal primers would not allow reading far enough into the sequence. Subclone deletions of the DNA fragment of interest were generated in order to facilitate reading far enough into the sequence to locate the deletion region.

Generation of Nested Deletions

Nested deletions were generated using the "Erase-a-Base" system (Promega Corporation, Madison, WI) according to the manufacturer's instructions. In order to sequence the region where the deletions occur, the Site 2 subclone, P4.6ME, was cloned into a pGEM7Zf^{+} vector (Promega Biotech, Madison, WI), and digested by exonuclease III digestion from the Eco RI site, with samples being removed at timed intervals. S1 nuclease treatment was done to remove single stranded tails left from digestion. The S1 nuclease was neutralized and heat inactivated, and Klenow DNA polymerase was added to generate blunt ends which were ligated to circularize the plasmids for transforming competent cells (Sambrook et al., 1989). The deleted subclones were screened by doing plasmid mini-preps (Mierendorf et al., 1978) to determine subclones of the appropriate size (see

Figure 3). Overlapping subclones at intervals that were far enough into the region were sequenced (P4.6ME-9, P4.6ME-13, P4.6ME-18). Sequence analysis was done with standard computer programs (Vasios et al., 1987). Sequences in the region of the deletion site allowed the construction of DNA primers for PCR reactions. The sequences necessary for primers are non-repetitive, non-complementary sequences of approximately 20-25bp long (Innis et al, 1990).

Polymerase Chain Reaction

Sequence analysis of the P4.6ME subclone sequences was done with the aid of the Genetics Computer Group sequence analysis program, version 7.1 (Devereux et al., 1984), and 20bp primers (4.6-1: GGAGACACGAGCCACTCAAG, 4.6-2: AGAGAACCCCCTCCAAGAGC) that flanked the region of the deletion were generated by the Macromolecular Analysis Facility (Loyola University Chicago, Maywood, IL). The primers were used in the polymerase chain reaction (PCR) to amplify the sequence around the site of the deletion region. PCR reactions were optimized using the Bios Optimization System (Bios Laboratories, New Haven, CT) and consisted of 10μl of 10x Optitaq buffer G (40mM NaCl, 10mM Tris-HCl, pH 8.4, 0.01% gelatin, 0.1% Triton X-100, 1.50mM $MgCl_2$) or O (10mM $(NH_4)_2SO_4$, 10mM KCl, 10mM Tris-HCl, pH 8.4, 0.01% gelatin, 0.1% Triton X-100, 3.75mM $MgCl_2$) (Bios

Laboratories, New Haven, CT) , 10μM of each primer, 1.06μl dNTP mix which contained 20mM of each deoxyribonucleotide: dGTP, dATP, dTTP and dCTP (Bios Laboratories, New Haven, CT), 4ng template for p4.6EP and 0.5ug template for patient DNA, 2U Taq DNA polymerase (Perkin Elmer Cetus, Norwalk, CT) and dH_2O to 100μl. PCR was carried out in an MJ Research Programmable Thermal Controller using a profile of: 94°C for 1 minute (denature), 64°C for 1 minute (anneal), 72°C for 1 minute (extend) for a total of 35 cycles.

In PCR products of patient DNAs, two bands resulted in those patients who were heterozygous for the deletion. Only one of the regions was the area of interest for sequencing. In order to isolate that one band, band-stab PCR was utilized (Kadokami et al., 1994; Bjourson et al., 1992). PCR reactions consisted of 10 μl of 10x Optitaq buffer G or O (Bios Laboratories, New Haven, CT), 10μM of each primer, 1.06μL dNTP mix (as above), 2U Taq DNA polymerase and dH_2O to 100 μl. The template for these reactions was acquired as follows: an aliquot of the two-banded PCR product was run out on a 1.5% agarose gel. The gel was stained with ethidium bromide, and viewed on a UV transilluminator (300nm). Excess water was removed from the surface of the gel with Whatman 3MM paper and the appropriate band was stabbed or lifted out of the gel with a hypodermic needle which was then dipped with gentle

agitation into a fresh PCR reaction mixture. PCR was carried out under the same conditions as above.

Cycle Sequencing

PCR products were prepared for sequencing using the "wizard" PCR clean-up method (Promega Corporation, Madison, WI). Thermal cycle sequencing using the "fmol" kit (Promega Corporation, Madison, WI) and the dsDNA Cycle Sequencing system (Life Technologies, Inc., Gaithersburg, MD) were done following the manufacturer's instructions which involved a two-step process to prepare the samples for PCR.

The dsDNA Cycle Sequencing system first step is labeling the primer: 1pmol of primer, 2pmol of γ^{32}P ATP, 1.0 μl 5x Kinase buffer, and 1.0 unit of T-4 kinase and dH_2O to 5μl were mixed and incubated in a 37°C waterbath for 10 minutes; the kinase was then deactivated at 55°C for 5 minutes. The reaction mixture consisted of: 5.0 μl labeled primer, 4.5μl 10x sequencing buffer, .033ng (50fmol) of template DNA, 1.23 units Taq polymerase and dH_2O to 36μl. 8μl of this mixture were added to tubes containing 2μl/tube of appropriate termination mix: ddGTP, ddATP, ddTTP or ddCTP. The tubes were overlaid with silicone oil and spun. After PCR was completed 5μl of stop solution was added and the tubes spun.

The fmol kit started by labeling the primer: 5pmol primer, 5pmol γ^{32}P ATP, 0.5μl 10x kinase buffer, and 2.5 units of T-4 kinase to 5μl was mixed and incubated in a 37°C waterbath for 30 minutes; the kinase was then deactivated at 90°C for 2 minutes. The reaction mixture consisted of: 1.5μl labeled primer, 5.0μl 5x fmol buffer, 235.3ng (500fmol) of template DNA and dH_2O to 16μl. To this tube 1.25 units of Taq polymerase were added, then 4μl were added to tubes containing 2μl/ tube of appropriate termination mix: ddGTP, ddATP, ddTTP or ddCTP. The tubes were overlaid with silicone oil and spun. After PCR was completed 3μl of stop solution was added and the tubes spun.

PCR was carried out for either method in an MJ Research Programmable Thermal Controller using a profile of: a 95°C "hot-soak" for 3 minutes, 95°C for 30 seconds (denature), 62°C for 30 seconds (annealing), 72°C for 1 minute (extension/ termination), repeating the denature/annealing/ extension/termination steps for 19 cycles, then 95°C for 30 seconds (denature), 70°C for 1 minute (extension/ termination), repeating this for 9 more cycles. The above two methods were utilized to sequence the deletion sites in the P4.6 clone and a patient with a short deletion. Comparison of the sequences of the short allele and the normal allele were then made.

Sequencing Gels

Sequencing gels containing 8% acrylamide/bis (37.5:1) and 7M urea were used to obtain the sequencing data. To prepare the gel for casting, glass plates were cleaned using "clear Ivory" dishwashing detergent and rinsed with tap water. Excess water was removed by rinsing the plates with 95% ethanol. The larger plate was then silanized using either Rain-X (Unelko Corporation, Scottsdale, AZ) or Gel Slick (AT Biochem, Malvern, PA) following the manufacturer's instructions to allow for easy separation following electrophoresis. One hundred twenty five mls of the 8% acrylamide/bis and urea mixture was filtered and degassed. Immediately before pouring, 800μl of fresh 10% ammonium persulfate and 20μl of TEMED were added to the gel mixture. The gel was poured between the plates using a 60ml syringe and 18 1/2 gauge needle. The gel was run using 1xTBE buffer and 55W of constant power. After electrophoresis, the sequencing gel was soaked for 15 minutes in 10% glacial acetic acid/10% methanol to remove the urea. The gel was then transferred to Whatman 3MM filter paper by established methods (Kozak et al., 1991), dried in a gel dryer (Bio-Rad laboratories, Hercules, CA) for 3 hours, 20 minutes and autoradiographed. The sequence was read a minimum of 3 times from independent gel runs in order to confirm the sequence accuracy. Sequence analysis was done with the aid of the VAX network system (Loyola

University Chicago), the Genetics Computer Group sequence analysis program, version 7.1 (Devereux et al., 1984), and the Hibio DNAsis DNA sequence analysis system, seventh version (Hitachi, 1990).

CHAPTER IV

RESULTS

Localization of Deletion in Site 2

The first step in characterizing the Site 2 locus (4.6kb Hind III fragment) was to approximately localize the region where the deletions occur. A genomic clone containing Site 2 (p4.6) has previously been isolated (Cammarata, 1994), and its restriction map is shown in Figure 1. The 1.4kb EcoRI/SstI fragment of this clone was ^{32}P-labeled by nick translation and used to probe blots of restricted genomic DNAs from two different normal heterozygous individuals and a patient with a homozygous 200bp deletion at Site 2. Single and double digest combinations were performed in order to delimit the deletion site. A single digest using Hind III was done, as well as various double digests, working inward toward the helical coding region (Figure 1). Smaller and smaller regions were digested to narrow down the location of the deletion. The goal was to locate where the fragment in the patient and the normal become the same size. The Hind III digest resulted in one fragment of 4.4kb in the patient and two fragments of 4.6kb and 4.3kb respectively in the normal (Figure 2A, lanes 9 and 10). The Hind III+Pst I digest resulted in one fragment of 3.7kb in the patient and

two fragments of sizes 3.9kb and 3.6 kb in the normal individual (Figure 2A, lanes 7 and 8). The Hind III+PvuII digest resulted in one fragment of 2.8kb in the patient and two fragments of sizes 3.0kb and 2.7kb in the normal individual (Figure 2A, lanes 5 and 6). In the second set of digests, done to confirm the data, one fragment of 2.8kb in the patient and two fragments of sizes 3.0kb and 2.7kb were also seen in the normal (Figure 2B, lanes 3 and 4). The Hind III+EcoRI digest resulted in one fragment 2.3kb in size in both the patient and normal (Figure 2A, lanes 3 and 4). The second set of digests resulted in one fragment of 2.3kb in both the patient and normal (Figure 2B, lanes 1 and 2). The HindIII+BglI digest resulted in one fragment 2.0kb in size for both the patient and normal (Figure 2A, lanes 1 and 2). The deletion was determined to be in the region between the EcoRI and PvuII sites. This is the region where the second allele, which contains the deletion is lost. The PvuII digest still had two bands and a smaller patient band, but in the Eco digest the bottom normal band is lost and the normal and patient band become the same size. This region contains the triple helical-like coding region of Site 2 (see Figure 1) (Cammarata, 1994). Because of the collagen-like nature of this region, and the fact that deletions can occur here, the region was sequenced to determine its characteristics.

Sequencing of the 4.6 Subclone

At Site 2, deletions of heterogeneous size are significantly more prevalent in OI Type III patients than in unaffected individuals (Cammarata et al., 1991; Doering et al., 1987). Because the above results showed the deletion was located in the collagen-like helical coding region of Site 2 (see Figure 1) the detailed sequence of this region is important to know. Previous characterization of 735bp of this region found it to be repetitive in nature and very GC rich (Cammarata, 1994). This led to problems in sequencing the entire region. A "ladder" effect from all the secondary structure resulted when trying to sequence farther in than 200bp from the primer. Internal primers based on the new sequence usually allows for further sequence determination and helps reduce the premature termination that results from secondary structure. However, the repetitive nature of the sequence (Cammarata, 1994) did not allow for identification and use of internal primers. To allow for sequencing, the EcoRI/SmaI fragment was isolated and subcloned as a 1.5kb EcoRI/SmaI fragment into a pGEM7Zf+ vector (P4.6ME) (see Figure 3C), so that nested deletions could be generated to allow sequencing of the entire region extending beyond 735bp to 899bp, which is the entire repeat region.

Generation of Nested Deletions

The "Erase-A-Base" system (Promega Corporation, Madison, WI) allowed serial removal of nucleotides from the EcoRI digested end of the P4.6ME subclone by employing exonuclease III. A series of clones is created by removing samples at timed intervals which bring the 3' sequences closer to the primer. This method allowed the creation of a series of subclones containing overlapping sequences that encompassed the remaining portion of the triple helical coding region (see Figure 3D). The deletion subclones P4.6ME-9 (325bp), P4.6ME-13 (225bp), and P4.6ME-18 (175bp) were used to generate the nucleotide sequence from position 535 through to position 899 which includes the remainder of the helical coding region. Sequence was obtained using the "Ampli-Taq" sequencing kit (Perkin-Elmer Cetus). This kit was utilized because it uses Taq polymerase, which is a very heat stable enzyme (45-72°C), which was needed to reduce secondary structure in the GC rich areas. Figure 4 shows a representative example of the sequence information obtained from 8% acrylamide sequencing gels of the subclones. The sequence generated by the above methods will be referred to as "NT899" (see Figure 5).

Results of the Sequencing Data

The nucleotide sequence. Analysis of the sequence shows that from position 128 to 775 there are 34 direct repeats composed of the consensus 18 nucleotide repeating unit: AAG ACC AGC AGC CCA GAC. Each of the repeating units have an average of 93.3% identity to this sequence, yet have substantial heterogeneity. There are three interruptions that are 9 nucleotides long, and these insertions, which disrupt the direct repeats, begin with a PvuII restriction site (indicated by carats and underlined in Figure 6). Although these insertions interrupt the direct repeats, their presence results in the coding of two perfect triple helical 9-mers with the amino acid sequence $(Gly\text{-}X\text{-}Y)_2$ in the reverse translation at these locations. There are also thymidine insertions which interrupt the direct repeats at various locations in the NT899 sequence (see Figure 6).

The basic repeating unit is not itself a strict triplet repeat sequence, but by substituting "C" for "A", in the fourth position it becomes one. Very few repeats contain this substitution though. On the occasions when it does occur, a "T" insertion also occurs in that repeat making it 19bp long. In the repeats with "T" insertions subsequent to a PvuII site, two perfect triplet repeats adjacent to each other also occur. These are also in the reverse translation.

Sequence analysis. Computer analysis of the sequence shows there is 63% GC content, which is consistent with

collagen sequences (Bernard et al., 1983). Analysis of the helical coding region sequence was performed with the aid of the GCG "bestfit","overlap" and "fasta" programs (Devereux et al., 1984). This allowed sequences to be chosen for DNA primers for PCR reactions. The desired primers would flank the repetitive region to allow amplification of the deletion region (see Figure 6). The sequences necessary for primers are non-repetitive, non-complementary sequences of approximately 20-25bp in length (Innis et al., 1990).

Fifteen primers were originally chosen from the regions flanking the repeat clusters of NT899 for analysis. The bestfit program was used to make an optimal alignment of the best segment of similarity between each primer and NT899. It does this by inserting gaps to maximize the number of matches. You want a primer with the smallest number of bestfits, so it will not be complementary with multiple places in the sequence. Both strands of the sequence are searched using this method. You can thus also determine that the primer will amplify in the proper orientation (5' or 3'), most efficaciously.

Once the primers are narrowed down by bestfit, you run them on the overlap program. It compares two sets of DNA sequences to each other in both orientations looking for overlaps between sequences rather than simply regions of similarity (see Figure 7). The lower the percentage of overlap found, the safer it would be to say that you have

only one site that would bind the primer. Anything below 50% overlap is considered acceptable (Devereux et al., 1984). If possible, overlap should not be on the end of the primer in the orientation of amplification, to avoid the risk of misamplification. Also, if the overlap is in the complementary orientation, it is not that big a concern because you will only be amplifying or sequencing in the direction for which the primer was chosen.

The final check of the primers was to run a "fasta" analysis. This searches for similarities between one sequence and any other sequence, aligning the search set against the query set for display Figure 7). It optimizes the length of overlap and minimizes gaps to only a few nucleotides, finding the portion of sequence with the highest identity. The resulting primers 4.6-1: GGAGACACGAGCCACTCAAG and 4.6-2: AGAGAACCCCCTCCAAGAGC flanked the region of the deletion (see Figure 6). These primers were chosen because they had the lowest amount of overlap to the NT899 sequence in the orientation for amplification, had no repetitive sequence, and had the least match to the NT899 sequence- in the complementary orientation.

Primer 4.6-1 was located between bp88 and 107 of NT899 (see bold underlined primer in Figure 6). Analysis of the first primer showed it had only 28-30% overlap to any site in the NT899 sequence in the forward direction, except for one site with 40% overlap. None of the regions of overlap were

on the end of the primer. It had one region of 30% overlap in the reverse (non-amplifying) direction. This was acceptable since PCR and DNA sequencing would only be in the forward direction (left to right in Figure 6) and not the reverse direction with this primer. Primer 4.6-2 was located between bp781 and 800 of NT899 (see bold underlined primer in Figure 6). Analysis of this primer showed it had only 30-35% overlap to any sites in the NT899 sequence in its amplification direction (right to left in Figure 6). It had one region of 30% overlap on the opposite strand (non-amplifying direction) that was within the primer, not on the end. This was acceptable since this primer would only be used to generate PCR products and DNA sequence in the amplification direction. The primers had no overlap with each other (data not shown).

PCR of the p4.6EP Subclone Repeat Region

The primers were used in the polymerase chain reaction (PCR) to amplify the sequence around the site of the deletion region in the p4.6EP subclone. PCR reactions were optimized using the Bios Optimization System (Bios Laboratories, New Haven, CT) which varies Mg^{++} concentrations (range 0.75-3.75mM) in three different buffers based on KCl, NaCl or $(NH_4)_2SO_4$. Mg^{++} concentration is a key variable in controlling specificity and yield of PCR. The ionic environment provided by the buffer is also critical to the specificity of

amplification. Amplification proceeded with a thermal profile specific to the locus and products were electrophoresed on agarose to select the buffer yielding the optimum balance of specificity and yield (data not shown). Buffers G(40mM NaCl, 10mM Tris-HCl, pH 8.4, 0.01% gelatin, 0.1% Triton x-100, 1.50mM $MgCl_2$) or O(10mM $(NH_4)_2SO_4$, 10mM KCl, 10mM Tris-HCl, pH 8.4, 0.01% gelatin, 0.1% Triton X-100, 3.75mM $MgCl_2$) were used to amplify the repeat region. Reactions contain 4ng of the p4.6EP clone DNA, which is amplified with Taq polymerase using the following profile: 94°C for 1 minute, 64°C for 1 minute, 72°C for 1 minute, for 35 cycles.

The p4.6 site 2 clone was amplified first to make sure the primers were specific and resulted in the proper size fragment. The resulting fragment was approximately 790bp long (see Figure 8, lane 3). This is the same size as the tandem repeat region flanked by the chosen primers. Genomic DNA from an unaffected individual was then amplified. Two fragments resulted, one 790bp in size and one 500bp in size (see Figure 8, lane 2). The smaller fragment is the one containing the deleted region. This verified the results from the Southern blots in Figure 1A and B, that there are two alleles in the normal individual and one contains an approximately 300bp deletion in the triple helical coding region.

PCR of Patient DNA

The constructed primers were then used to amplify the repeat region from the genomic DNAs of five additional OI patients known to have Site 2 deletions, three of which are shown here (see Figure 8, lanes 4, 5 and 6). Four out of the five patient DNAs that were PCR amplified were either heterozygous or homozygous for deletions in the repeat region. A possible explanation for the patient with only one full-length band could be that the deletion region expanded outside the tandem repeat region or may be located elsewhere in the Site 2 gene. Deletions at the site result in shorter than normal fragments being generated by the PCR. As a result, comparing the sizes of the DNA fragments generated from each sample permits identification of individuals with a deletion and indicates what size deletion they have. One patient (lane 6) was homozygous for a deletion of approximately 200bp. The range of sizes of the deletion in the three heterozygous patients from 60-340 bp in length. In all cases but one, shorter than normal fragments resulted, indicating that Site 2 deletions are usually confined to the region containing the tandem repeat clusters.

Direct Sequencing of PCR Products

In order to sequence PCR products, they must be free of contaminating amplification primers and nucleotides. The wizard PCR preps DNA purification system (Promega, Madison,

WI) was used to prepare PCR products for sequencing. It purifies double-stranded PCR amplified DNA away form contaminants, including primer-dimers and amplification primers. The DNA is eluted from the Magic PCR preps resin (Promega Corporation, Madison, WI) in water free of salt or macromolecular contaminants. Samples were then ready for direct sequencing.

Because the sequence is very GC rich and repetitive in nature, sequence extension can be difficult. There are also problems with secondary structure causing premature termination. Thus, two methods were investigated for direct cycle sequencing to determine which would give better results. The first was the dsDNA Cycle Sequencing system (Life Technologies, Inc., Gaithersburg, MD) which uses 50 fmol of template and an end-labeled primer with Taq polymerase. The second was the Fmol kit (Promega Corporation, Madison, WI) which uses 500 fmol of template, an end-labeled primer and Taq polymerase. The kits were very similar except for the 100-fold difference in the starting template (see Materials and Methods). The fmol kit was chosen to sequence a majority of the patient DNA because it resulted in clearer, more readable sequence further in. Optimization experiments were done increasing the dsDNA method starting template amount to see if that improved the quality of the results, It did not. Fmol allowed for clear sequencing of the repeat region of patient DNA far enough in

to narrow down the region where the deletion was located. However, due to the repetitive nature of the DNA and lack of restriction sites as landmarks, it made determining the exact endpoints of the deletion impossible.

PCR sequencing reactions were carried out for either method using the following profile: 95°C for 3 minutes, 95°C for 30 seconds, 62°C for 30 seconds, 72°C for 1 minute, repeating all but the first step for 19 cycles, then 95°C for 30 seconds, 70°C for 1 minute, for 9 more cycles.

Sequencing of the p4.6EP Subclone Repeat Region

Cycle sequencing was first performed on the p4.6EP subclone to verify that the region of interest was being properly amplified, and to optimize methods (see Figure 10A). Both direct cycle sequencing methods were used, with the Fmol method generating most of the sequence. The sequence obtained was an exact match to NT899 (see Figure 5). This verified that the sequence that would be obtained from the tandem repeat region of the patient DNA would be reliable.

Sequencing of the Repeat Region in Patient DNA

In order to sequence the patient DNA, the two fragments must be separated to isolate only the band containing the deleted allele. This was done using bandstab PCR (Kadokami et al., 1994; Bjourson et al., 1992). An aliquot of the PCR amplified patient DNA was run out on a high percentage

agarose gel (1.5%). The gel was stained with ethidium bromide, and viewed on a UV transilluminator (300nm). Excess water was removed from the surface of the gel and the lower band was stabbed from the gel with a hypodermic needle which was then agitated into a normal PCR reaction mixture that was prepared without any template. PCR was carried out under the same conditions as above. An aliquot of the reaction was run out on an agarose gel to verify amplification. Figure 9 shows these results. Lane 2 is an aliquot of the original patient genomic amplification, showing the two bands of approximately 790bp and 700bp in size. Bandstab results are shown in lanes 3 and 4. The top band is in lane three and the bottom band (containing the deletion) is in lane 4. This shows the deletion to be approximately 90bp in size.

The bandstab PCR reaction for the deleted allele was then cleaned up in order to remove contaminants and extraneous primer by the "wizard" PCR clean-up method (Promega Corporation, Madison, WI). DNA cycle sequencing reactions were then performed on this fragment and run on 8% acrylamide sequencing gels. Figure 10B is an example of the patient sequence obtained by this method. The region shown is the same region as that in Figure 10A from the p4.6EP subclone. The sequences match exactly. The PCR sequencing reactions accurately amplified the patient template, generating sequence of the repeat/deletion region. This again verified that this was an effective, reliable method of

determining the sequence of patient DNA.

The deleted allele for this patient (3C2) was then sequenced to the extent possible as above and compared to the entire NT899 sequence and the p4.6 PCR product sequence. The sequence is shown in Figure 11. Nucleotides sequenced are shown aligned along the NT899 sequence for comparison. The available regions of patient sequence matched NT899 exactly. No point mutations were present in the patient sequence. The region of the deletion is represented by the cross hatches. The deletion region of approximately 125bp is shown in the repeat aligned sequence of NT899 in Figure 12. This is the maximum size of the deletion region. Since the sequence could not be reliably read any further in from the primers. The exact deletion endpoints could not be determined due to the repetitive nature of the sequence and lack of restriction sites to use as landmarks. It was difficult to tell where one repeat started and ended as you got further into the deletion region. PCR of the genomic patient DNA gave a deletion size of approximately 90bp in length (see Figure 9). Thus the sequencing results and the PCR data indicate that the deletion in this patient is approximately 90-125bp in length.

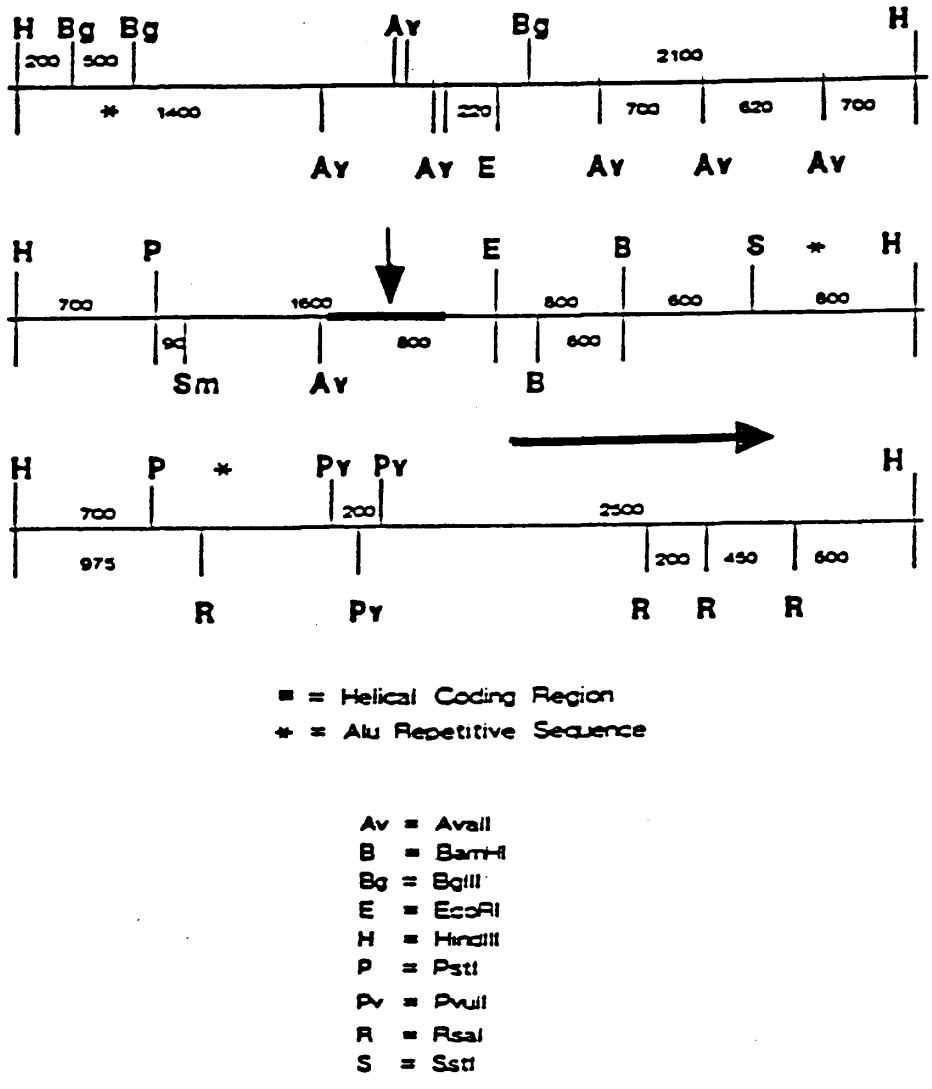


Figure 1. Restriction map of 4.6kb cloned fragment showing the deletion location.
The new collagen-like locus on chromosome 17 was cloned from a chromosome specific library as a 4.6kb HindIII restriction fragment. Restriction sites for the indicated enzymes are shown. Fragment sizes in base pairs are indicated. The location of Alu repetitive sequences are also indicated (*). The locus contains a short triple helical coding region that cross-hybridizes the the pro-α2(I) cDNA, Hf32 (bold line). The region where deletions occur (as indicated by the bold arrow) is located between the EcoRI and PvuII restriction sites. This is the same sequence where the 18bp tandem repeats are located. The direction of transcription in this locus is indicated by the horizontal arrow.

Figure 2. Localization of deletion region.

A: Restriction digests of DNAs from a patient with a homozygous 200bp deletion at Site 2 and from a normal heterozygous individual were hybridized to the EcoRI/SstI fragment of p4.6.

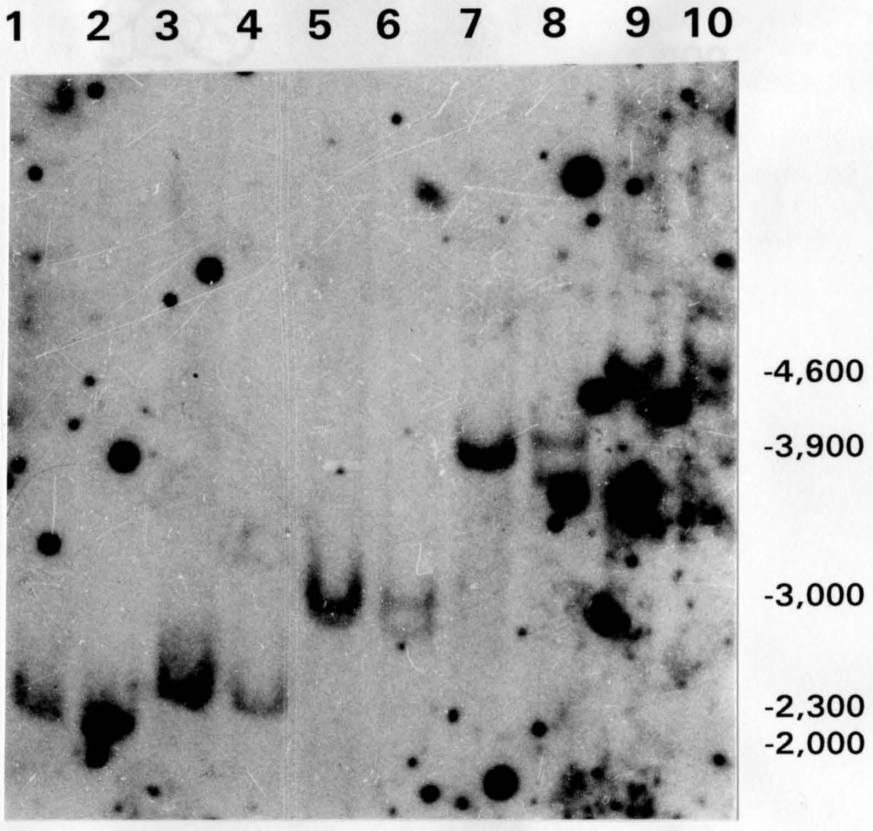
Lane 1, patient DNA digest: HindIII+BglI
Lane 2, normal DNA digest: Hind III+BglI
Lane 3, patient DNA digest: HindIII+EcoRI
Lane 4, normal DNA digest: HindIII+EcoRI
Lane 5, patient DNA digest: HindIII+PvuII
Lane 6, normal DNA digest: HindIII+PvuII
Lane 7, patient DNA digest: HindIII+PstI
Lane 8, normal DNA digest: HindIII+PstI
Lane 9, patient DNA digest: HindIII
Lane 10, normal DNA digest: HindIII

B: Restriction digests of DNAs from a patient with a homozygous 200bp deletion at Site 2 and from a different normal heterozygous individual were hybridized to the EcoRI/SstI fragment of p4.6.

Lane 1, patient DNA digest: HindIII+EcoRI
Lane 2, normal DNA digest: HindIII+EcoRI
Lane 3, patient DNA digest: HindIII+PvuII
Lane 4, normal DNA digest: HindIII+PvuII

Comparing the fragments obtained in the patient to that in the normal control allowed localization of the deletion to the small region between PvuII and EcoRI sites.
This is seen as the loss of the second band in the normal heterozygous individual, and both normal and patient having one band of the same size.

A.



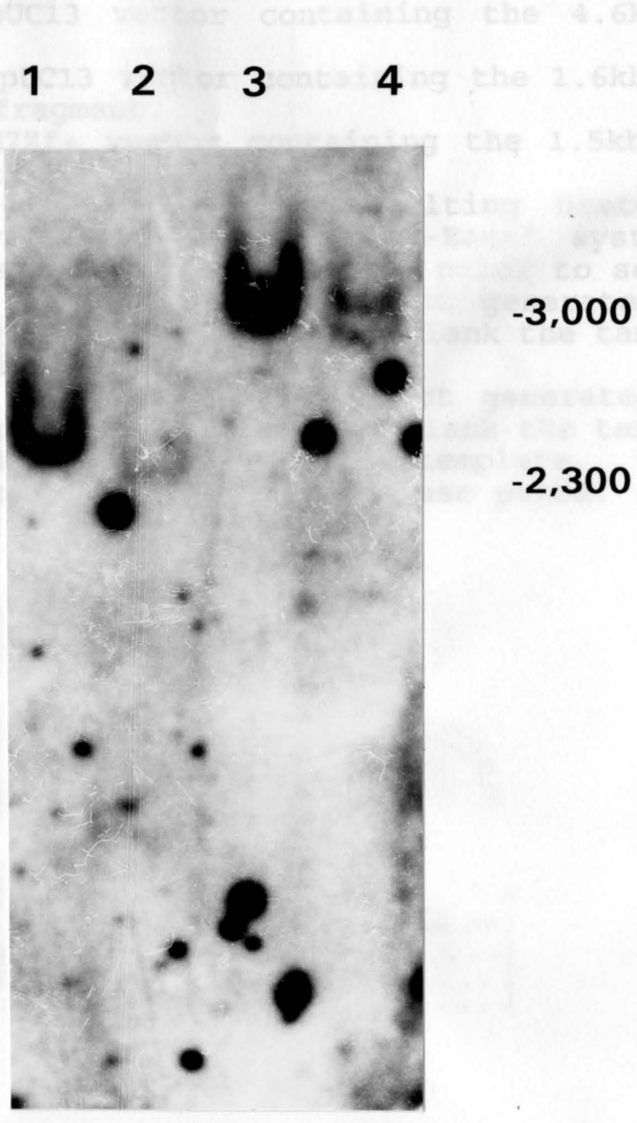

B.
1
2
3
4
-3,000
-2,300

Figure 3. Schematic diagram of the clones, subclones and PCR products used to characterize the deletion site. Only a portion of the vectors are shown in the diagrams.
A: p4.6, the pUC13 vector containing the 4.6kb Hind III fragment.
B: p4.6EP, the pUC13 vector containing the 1.6kb EcoRI/PstI helical coding fragment.
C: p4.6ME, pGEM7Zf+ vector containing the 1.5kb EcoRI/SmaI triple helical coding fragment.
D: Linearized p4.6ME and the resulting nested deletion subclones generated by the "Erase-A-Base" system. These deleted subclones were recircularized prior to sequencing.
E: PCR product 4.6EP, The PCR product generated using the constructed pair of PCR primers that flank the tandem cluster on the p4.6 EcoRI/PstI template.
F: PCR product Pt3C2, The PCR product generated using the constructed pair of PCR primers that flank the tandem cluster on a heteozygous patient genomic DNA template.
Sizes of fragments are indicated in base pairs.

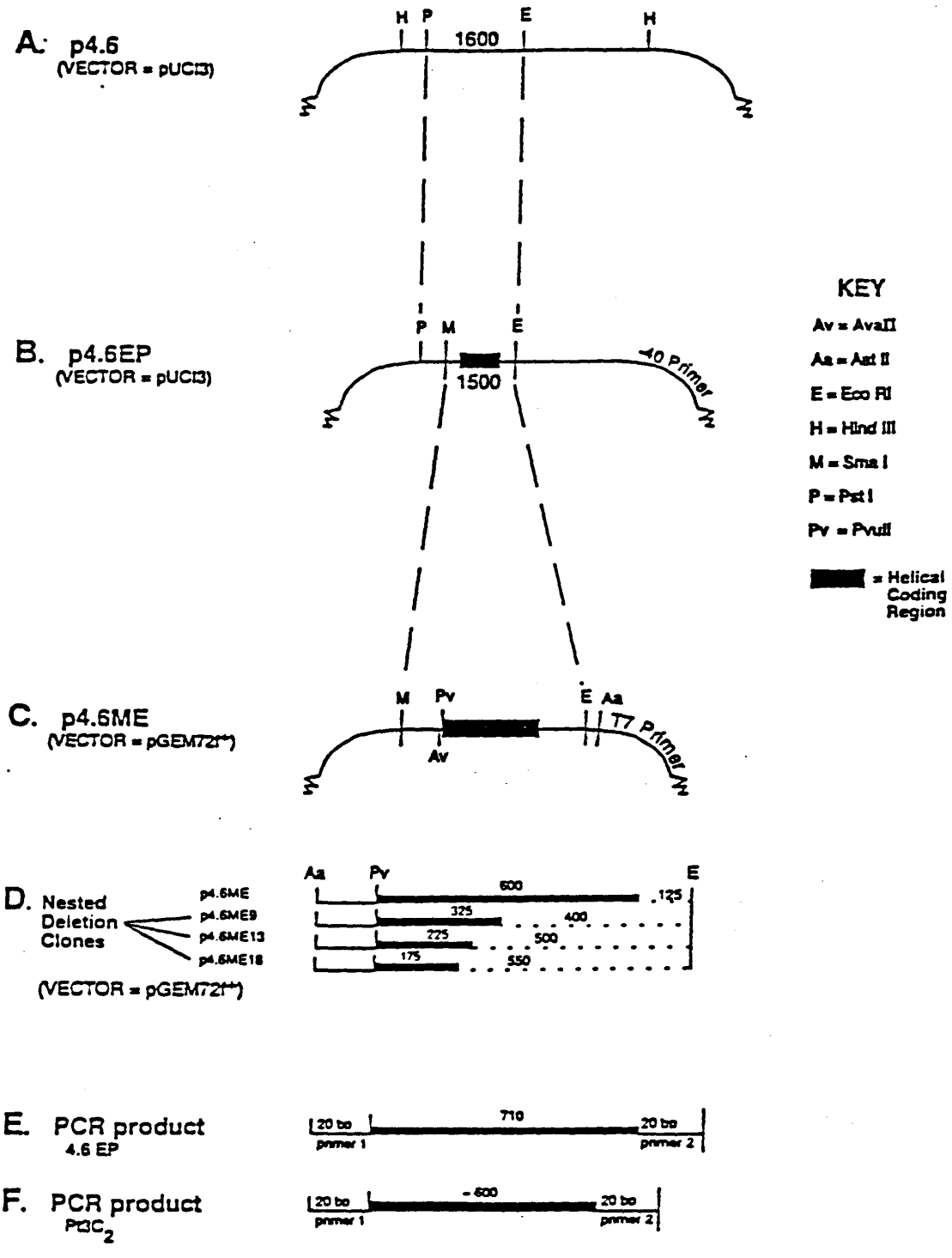
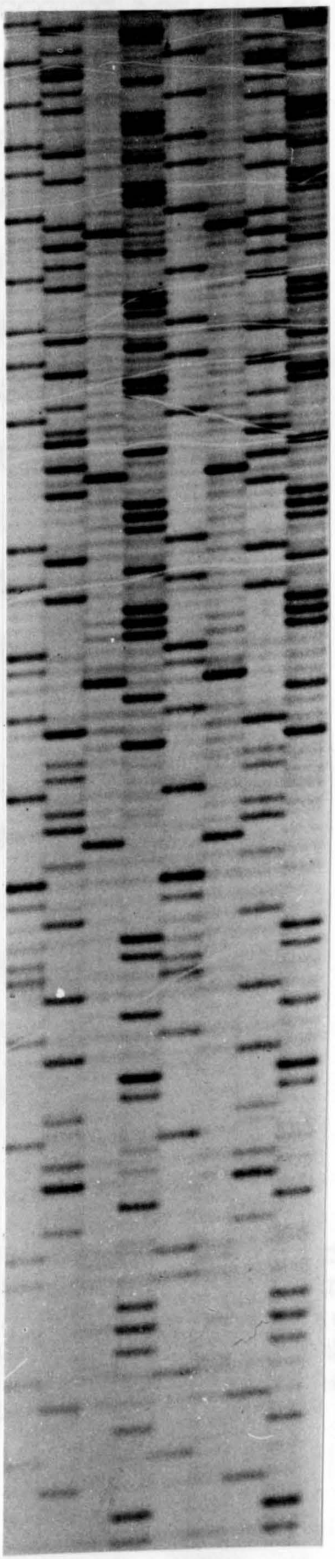

A. p4.6
(VECTOR = pUC13)
H P E H
1600
B. p4.6EP
(VECTOR = pUC13)
P M E
1500
40 Primer
KEY
Av = AvaII
Aa = Ast II
E = Eco RI
H = Hind III
M = Sma I
P = Pst I
Pv = PvuII
= Helical Coding Region
C. p4.6ME
(VECTOR = pGEM72f**)
M Pv E Aa
Av
T7 Primer
D. Nested Deletion Clones
p4.6ME
p4.6ME9
p4.6ME13
p4.6ME18
(VECTOR = pGEM72f**)
Aa Pv E
600
125
325
400
225
500
175
550
E. PCR product
4.6 EP
20 bp
710
20 bp
primer 1
primer 2
F. PCR product
PUC2
20 bp
~600
20 bp
primer 1
primer 2

Figure 4. 8% acrylamide sequencing gel of Site 2 nested deletions.
A typical example of the sequence generated using the "Ampli-Taq" kit. 102 nucleotides can be read as pictured here (position 643 to 745 of the NT899 sequence). The PvuII site sequence at position 680 is indicated. The letters at the top indicate the nucleotide represented in the corresponding lane.

G A T C G T A C

-745



-PvuII

-643

```
      EcoRI
  1   GAATTCCAGT CACTCCCTCC AGGAGACACA CCAGCAGTGG TCAGGAAGCA

 51   AGTCCACCAC AGCCTTCTGG AAAGGTCTCT AGTCTCAGGA GACACGAGCC
                                      *
101   ACTCAAGCCT CCCACATGCC ACCAAAGAAG ATAAGCTGGC CCAGAGAAGA

151   CCAGCAGCCC AGAGAAGACC AGCAGCCTGG ACAAGACCAA CAGCCCAGAC
                            AvaII                AvaII
201   ATAGACCAGC AGCCCAGAAA GGACCAGCAG CCCAGACAGG ACCAGCAGCC

251   CAGGCAAGAC CAGCAGCCTG GACAAGACCA GCAGCCTGGA CAAGACCAAC
                                        AvaII
301   AGCCCAGACA TAGACCAGCA GCCCAGACAG GACCAACAGC CCAGACATAG
                      AvaII
351   ACCAGCAGCC CAGACAGGAC CAGCAGCCCA GACAAGACCA GCAGCCTGGA
        PvuII
401   CAAGAACAGC TGGCCCAGCA GCCCATACAA GACCAGCAGC CCAGATATAG

451   CCCAGCAGCC CAGACATAGC CCAGCAGCCC AGACAAGACC AGCAGCCCAG
                                                    PvuII
501   ACAAGACCAG CAGCCTGGAC AAGACCAGCA GCCAGGATAA GAACAGCTGG

551   CCCAGCAGCC CATACAAGAC CAGCAGCCCA GATATAGCCC AGCAGCCCAG

601   ACATAGCCCA GCAGCCCAGA CAAGACCAGC AGCCCAGACA AGACCAGCAG
                                        PvuII
651   CCCGGACAAG ACCAGCAGCC AGGATAAGAA CAGCTGGCCC AGCAGCCCAT

701   ACAAGACCAG CAGCCCAGAC ATAGCCCAGC AGCCCAGACA AGACCAGCAG
                                      *
751   CCCAGACAAG AACCAATCTG CCCAGGCTGC TCTTGGAGGG GGTTCTCTCA

801   GGATTCTCAG GCTCAAAGTA GGCCGAGCCT GGGTCATCAC CACTCCGGCA
            AvaII
851   GCCCAAGTGG GACCTCCTGG GACAGGCTTC ATCTCCGCAC TTGCAGGAG
```

Figure 5. Complete nucleotide sequence of NT899. Selected restriction sites are indicated above their recognition sequence. Asterisks indicate the beginning and end of the tandem repeat clusters. Numbers on the left margin indicate the nucleotide position at the start of the line. A space is inserted after every tenth nucleotide for easier reading.

Figure 6. Sequence of NT899 aligned to illustrate the 18bp nucleotide direct repeats.
There are 34 repeats with a minimum of 83.4% identity to the basic repeat and 3 brief insertions of 9 nucleotides each. Single thymidine insertions also occur periodically. The carats indicate the exact location of the insertions which have been removed and placed above the sequence they interrupt. The primers used to amplify the region are indicated by the bold underline. Numbers to the left indicate the respective repeat. Numbers to the right indicate the nucleotide position. Select restriction sites are indicated.

GAATTCCAGTCACTCCCTCCAGGAGACACACCAGCAGTGGTCAGGAAGCA 50
AGTCCACCACAGCCTTCTGGAAAGGTCTCTAGTCTCA**GGAGACACGAGCC** 100
ACTCAAGCCTCCCACATGCCACCAAAG 127

T

1 A A G A T A A G CvG G C C C A G A G 146

2 A A G A C C A G C A G C C C A G A G

3 A A G A C C A G C A G C C T G G A C 182

4 A A G A C C A A C A G C C C A G A C
T
5 AvA G A C C A G C A G C C C A G A A 219
AVAII
6 A G G A C C A G C A G C C C A G A C
AVAII
7 A G G A C C A G C A G C C C A G G C 255

8 A A G A C C A G C A G C C T G G A C

9 A A G A C C A G C A G C C T G G A C 291

10 A A G A C C A A C A G C C C A G A C
T
11 AvA G A C C A G C A G C C C A G A C 328
AVAII
12 A G G A C C A A C A G C C C A G A C
T
13 AvA G A C C A G C A G C C C A G A C 365
AVAII
14 A G G A C C A G C A G C C C A G A C

15 A A G A C C A G C A G C C T G G A C 401
PVUII
CAGCTGGCC
16 A A G A AvC A G C A G C C C A T A C

17 A A G A C C A G C A G C C C A G A T 446
T
18 AvA G C C C A G C A G C C C A G A C
T
19 AvA G C C C A G C A G C C C A G A C 484

20 A A G A C C A G C A G C C C A G A C

21 A A G A C C A G C A G C C T G G A C 520

22 A A G A C C A G C A G C C A G G A T
PVUII
CAGCTGGCC
23 A A G A AvC A G C A G C C C A T A C 565

24 A A G A C C A G C A G C C C A G A T
T
25 AvA G C C C A G C A G C C C A G A C 602
T
26 AvA G C C C A G C A G C C C A G A C

27 A A G A C C A G C A G C C C A G A C 639

28 A A G A C C A G C A G C C C C G G A C

29 A A G A C C A G C A G C C A G G A T 675
PVUII
CAGCTGGCC
30 A A G A AvC A G C A G C C C A T A C

31 A A G A C C A G C A G C C C A G A C 720
T
32 AvA G C C C A G C A G C C C A G A C

33 A A G A C C A G C A G C C C A G A C 757

34 A A G A A C C A A T C T G C C C A G

GCT**GCTCTTGGAGGGGGTTCTCT**CAGGATTCTCAGGCTCAAAGTAGGC 823
CGAGCCTGGGTCATCACCACTCCGGCAGCCCAAGTGGGACCTCCTGGG 871
ACAGGCTTCATCTCCGCACTTGACGGAG 899

A.

```
PRIMER #1 FROM NT899 SEQUENCE
TO: NT899.seq;1
P4.6 M/E deletion subclone sequence

SCORES      Init1: 80 Initn: 80 Opt: 80
            100.0% identity in 20 bp overlap

                                              10        20
PRIMER
                                     GGAGACACGAGCCACTCAAG
NT899                                ||||||||||||||||||||
CACAGCCTTCTGGAAAGGTCTCTAGTCTCAGGAGACACGAGCCACTCAAGCCTCCCACAT
 60         70         80         90        100        110


GCCACCAAAGAAGATAAGCTGGCCCAGAGAAGACCAGCAGCCCAGAGAAGACCAGCAGCC
 120        130        140        150        160        170
```

B.

```
PRIMER #2 FROM NT899 SEQUENCE
TO: NT899.seq;1
P4.6 M/E deletion subclone sequence

SCORES      Init1: 80 Initn: 80 Opt: 80
            100.0% identity in 20 bp overlap

                                              10        20
PRIMER
                                     GCTCTTGGAGGGGGTTCTCT
NT899                                ||||||||||||||||||||
AGCCCAGACAAGAACCAATCTGCCCAGGCTGCTCTTGGAGGGGGTTCTCTCAGGATTCTC
       760        770        780        790        800        810

AGGCTCAAAGTAGGCCGAGCCTGGGTCATCACCACTCCGGCAGCCCAAGTGGGACCTCCT
       820        830        840        850        860        870
```

Figure 7. Primer pattern analysis results.
A: Fasta computer analysis of Primer 1 showing complete identity only with the region that exactly matches the primer selected.
B: Fasta computer analysis of Primer 2 showing complete identity only with the region that exactly matches the primer selected.

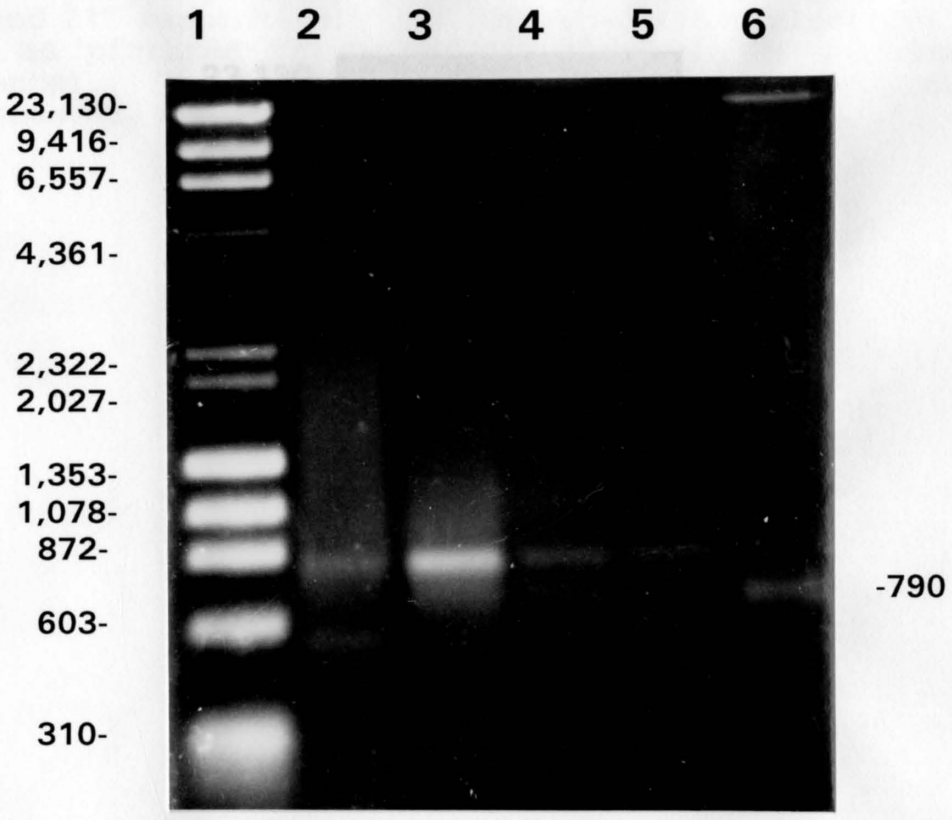


Figure 8. PCR amplification of repeat region.
Using primers synthesized from areas surrounding the tandem repeats in Site 2, PCR amplification of the deletion regions was possible. Lane 1: molecular weight marker, Lane 2: amplification of tandem repeat region in a unaffected individual (BenVen), Lane 3: amplification of tandem repeat region in Site 2 clone (p4.6EP), Lane 4, 5, 6: amplification of tandem repeat region in three OI patients with Site 2 deletions (Pt3C2, Pt20C2, Pt530).

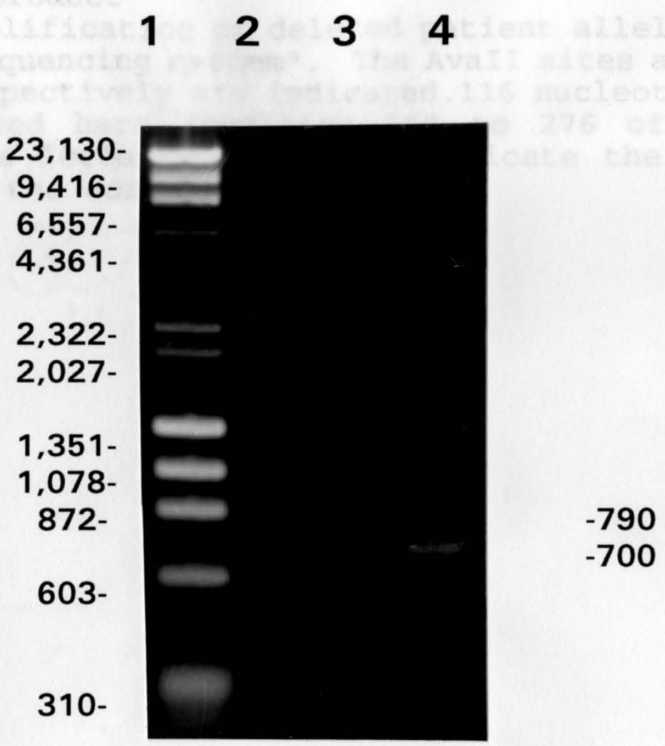
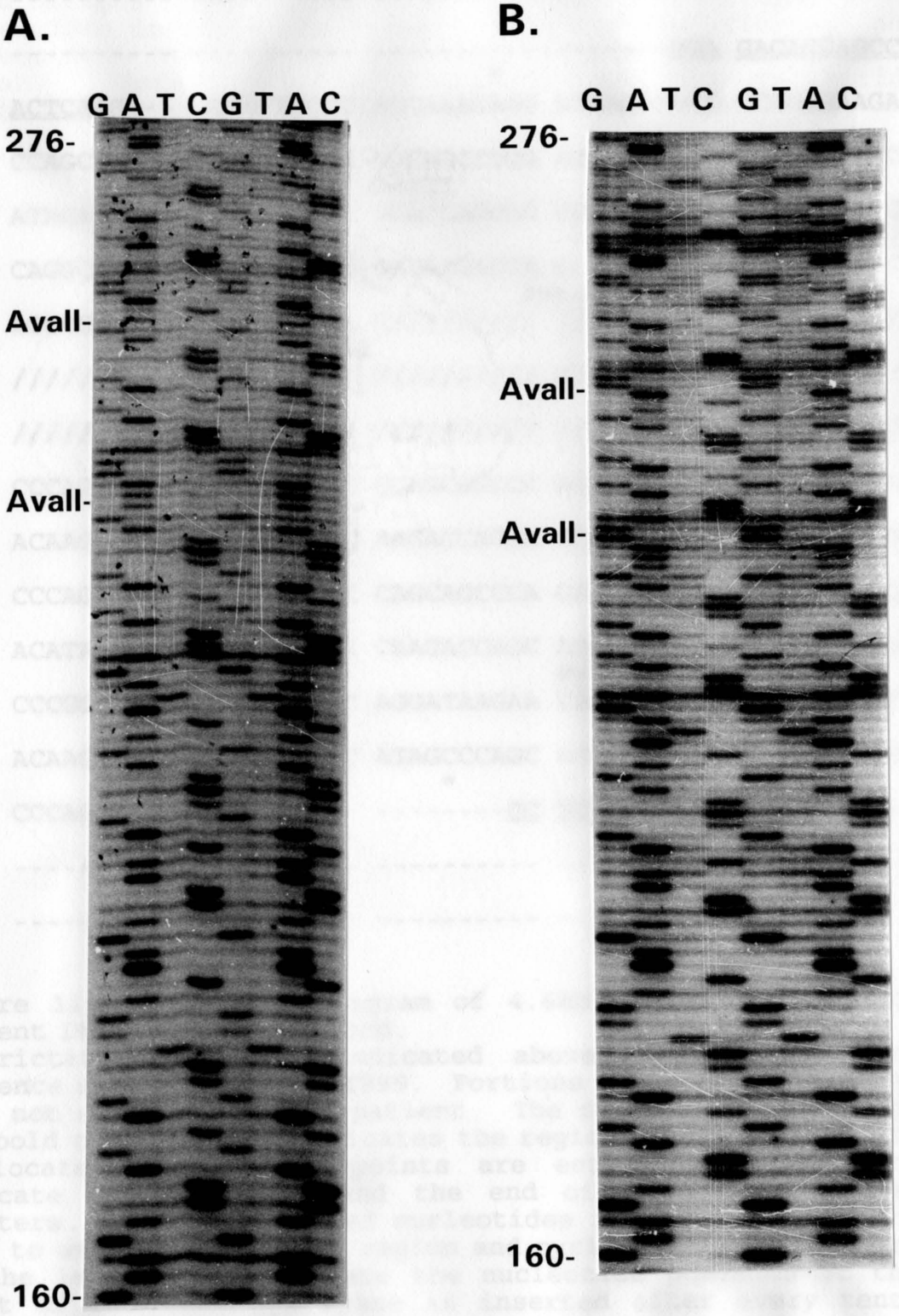


Figure 9. Bandstab amplification of deleted region.
In order to sequence the deletion region in patient DNA, it was necessary to determine which allele contained the deletion. Bandstab PCR was used to amplify each individual allele and direct sequencing of the PCR-amplified fragment from the deleted allele was performed. Lane 1: molecular weight marker, Lane 2: amplification of tandem repeat region in OI patient, Lane 3: bandstab amplification of larger patient allele, Lane 4: bandstab amplification of smaller patient allele.

Figure 10. PCR direct sequencing acrylamide sequencing gels. 8% acrylamide sequencing gel of:
A: p4.6EP PCR product
B: bandstab amplification of deleted patient allele using the "dsDNA cycle sequencing system". The AvaII sites at positions 220 and 237 respectively are indicated.116 nucleotides can be read as pictured here (position 160 to 276 of the NT899 sequence). The letters at the top indicate the nucleotide represented in the corresponding lane.


A.
G A T C G T A C
276-
AvaII-
AvaII-
160-
B.
G A T C G T A C
276-
AvaII-
AvaII-
160-

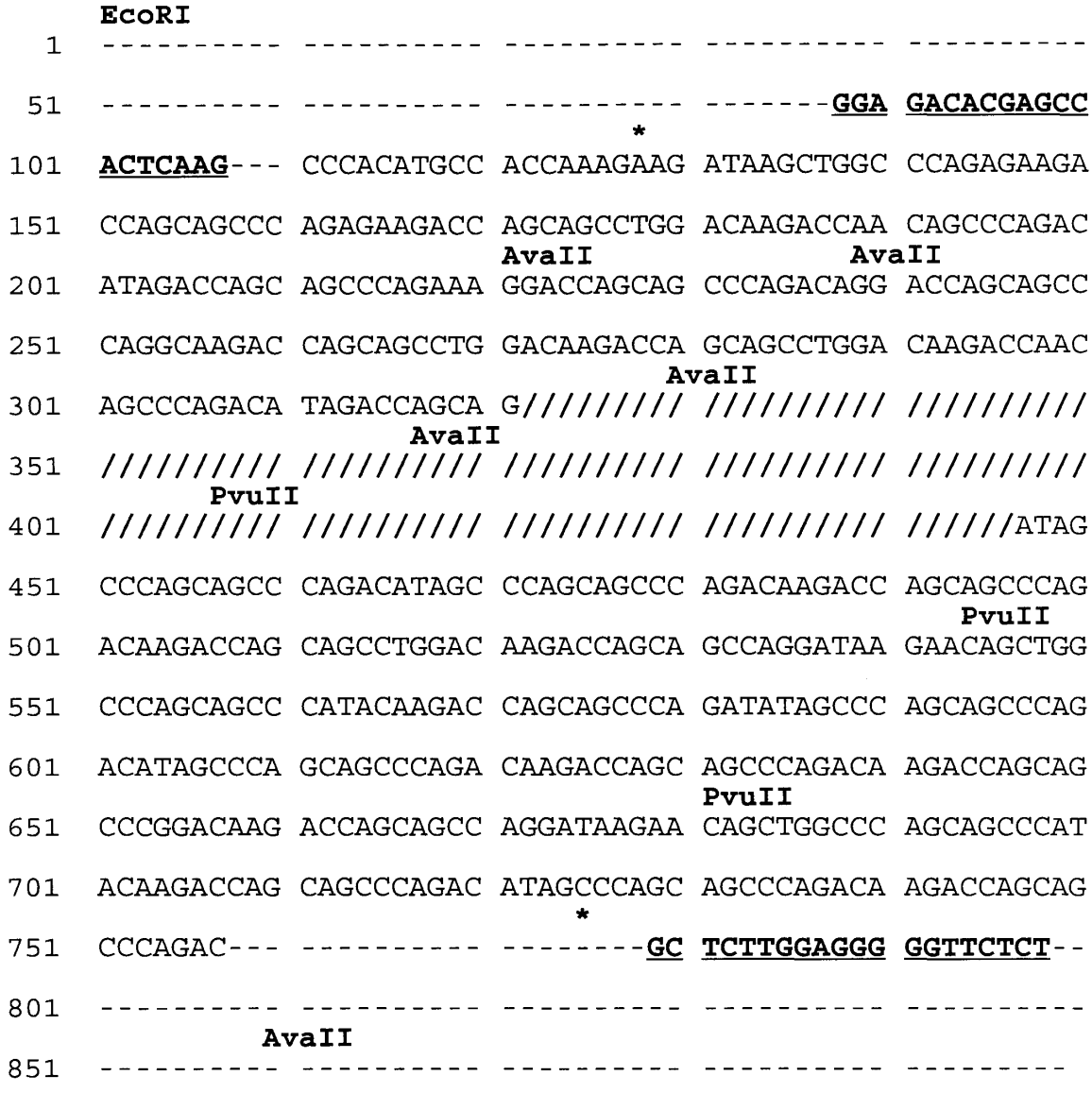
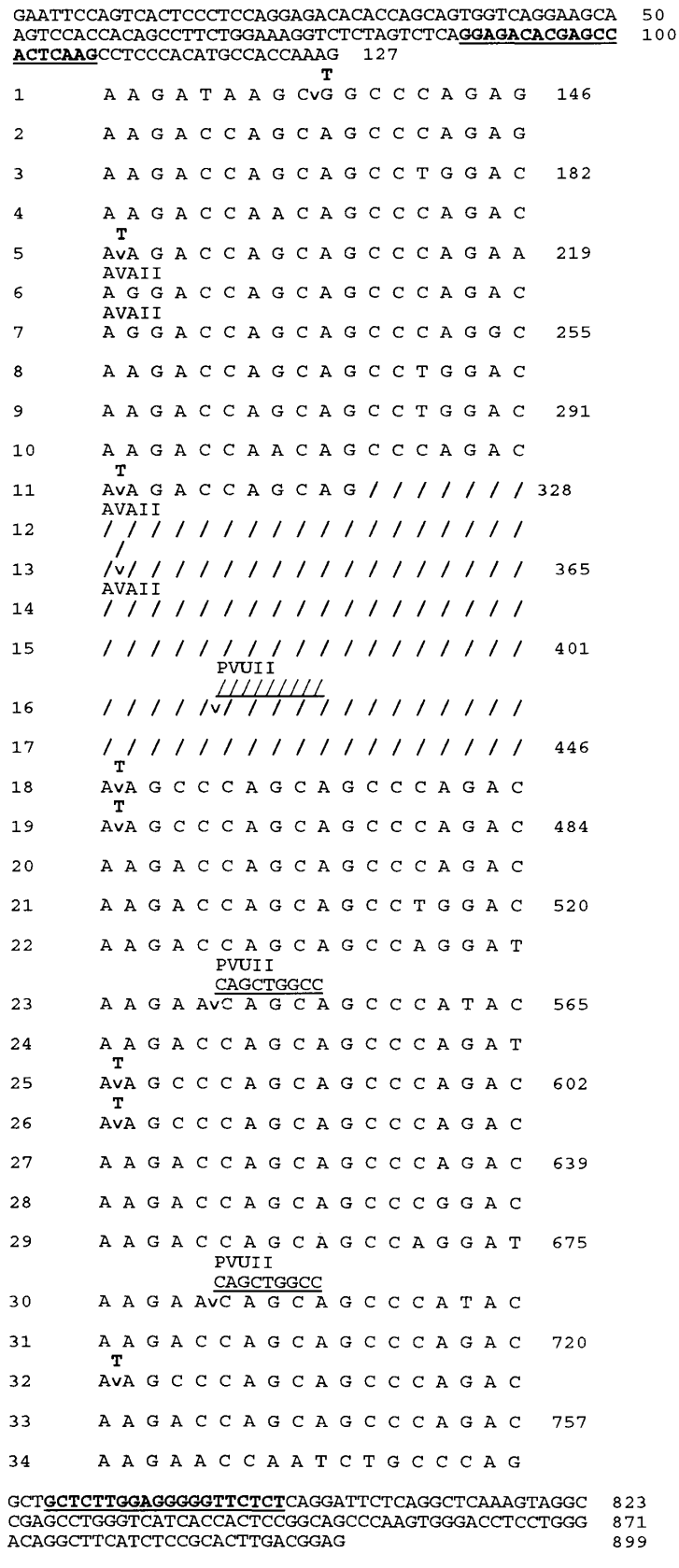


Figure 11. Schematic diagram of 4.6kb clone sequence and patient DNA sequence aligned.
Restriction sites are indicated above their recognition sequence as they were in NT899. Portions represented by lines were not sequenced in the patient. The region represented by the bold cross hatches indicates the region where the deletion is located. Actual endpoints are estimated. Asterisks indicate the beginning and the end of the tandem repeat clusters. Bold, underlined nucleotides indicate the primers used to amplify the repeat region and cycle sequence. Numbers on the left margin indicate the nucleotide position at the start of the line. A space is inserted after every tenth nucleotide for easier reading.

Figure 12. Schematic diagram of direct repeats in NT899 indicating deletion region.
Cross hatches indicate the deletion region. Bold, underlined nucleotides indicate the primers used for amplification and sequencing. Numbers to the left indicate the respective repeat. Numbers to the right indicate the nucleotide position. Select restriction sites are indicated.


GAATTCCAGTCACTCCCTCCAGGAGACACACCAGCAGTGGTCAGGAAGCA 50
AGTCCACCACAGCCTTCTGGAAAGGTCTCTAGTCTCAGGAGACACGAGCC 100
ACTCAAGCCTCCCACATGCCACCAAAG 127
T
1 A A G A T A A G CvG G C C C A G A G 146
2 A A G A C C A G C A G C C C A G A G
3 A A G A C C A G C A G C C T G G A C 182
4 A A G A C C A A C A G C C C A G A C
T
5 AvA G A C C A G C A G C C C A G A A 219
AVAII
6 A G G A C C A G C A G C C C A G A C
AVAII
7 A G G A C C A G C A G C C C A G G C 255
8 A A G A C C A G C A G C C T G G A C
9 A A G A C C A G C A G C C T G G A C 291
10 A A G A C C A A C A G C C C A G A C
T
11 AvA G A C C A G C A G / / / / / / / 328
AVAII
12 / / / / / / / / / / / / / / / / /
/
13 /v/ / / / / / / / / / / / / / / / 365
AVAII
14 / / / / / / / / / / / / / / / / /
15 / / / / / / / / / / / / / / / / / 401
PVUII
16 / / / / /v/ / / / / / / / / / / /
17 / / / / / / / / / / / / / / / / / 446
T
18 AvA G C C C A G C A G C C C A G A C
T
19 AvA G C C C A G C A G C C C A G A C 484
20 A A G A C C A G C A G C C C A G A C
21 A A G A C C A G C A G C C T G G A C 520
22 A A G A C C A G C A G C C A G G A T
PVUII
CAGCTGGCC
23 A A G A AvC A G C A G C C C A T A C 565
24 A A G A C C A G C A G C C C A G A T
T
25 AvA G C C C A G C A G C C C A G A C 602
T
26 AvA G C C C A G C A G C C C A G A C
27 A A G A C C A G C A G C C C A G A C 639
28 A A G A C C A G C A G C C C G G A C
29 A A G A C C A G C A G C C A G G A T 675
PVUII
CAGCTGGCC
30 A A G A AvC A G C A G C C C A T A C
31 A A G A C C A G C A G C C C A G A C 720
T
32 AvA G C C C A G C A G C C C A G A C
33 A A G A C C A G C A G C C C A G A C 757
34 A A G A A C C A A T C T G C C C A G
GCTGCTCTTGGAGGGGGTTCTCTCAGGATTCTCAGGCTCAAAGTAGGC 823
CGAGCCTGGGTCATCACCACTCCGGCAGCCCAAGTGGGACCTCCTGGG 871
ACAGGCTTCATCTCCGCACTTGACGGAG 899

CHAPTER V

DISCUSSION

This work involves characterization of the deletions in a newly discovered human collagen-like locus associated with osteogenesis imperfecta. Spontaneous deletions of heterogeneous size occur at the Site 2 locus more frequently in OI Type III patients than in normal individuals (Doering et al., 1987). This suggests that the defect in the structure of the protein coded for by Site 2 can contribute to the OI phenotype. The goal of this study was to locate the deletion site in the 4.6kb Hind III Site 2 fragment, and analyze the molecular structure of this region in affected individuals. This allowed the deletion to be more precisely localized and suggested a mechanism for the frequent occurrence of deletions at this site. The PCR methods developed here will also be useful in eventual screening programs for identifying individuals carrying deletions at Site 2.

Region of Deletion

Initially the deletion in the Site 2 locus was localized by Southern blotting to the region between the EcoRI and

PvuII sites in one patient (see Figure 1). This is the collagen-like helical coding region. The full sequence of this region revealed 34 direct repeats composed of an 18bp consensus repeating unit (see Figure 7). To determine whether this repeat region is the major site for deletions, primers were synthesized that flanked the region, and PCR was performed on a normal individual, and multiple OI Type III patients. Four out of five patients amplified were either heterozygous or homozygous for a deletion in this region. The deletions were of heterogeneous size from patient to patient ranging from 60-340bp in length. The one patient that only had one normal-sized fragment by PCR could have a Site 2 deletion that extends outside the tandem repeats. Deletions are often found that extend from tandemly repeated sequences into adjacent regions (Willing et al., 1988; Zuliani et al., 1990). In any case, these PCR results demonstrate that Site 2 deletions are primarily localized to the cluster of 18bp repeats. Direct sequence data obtained from one patient indicates that the deletion is actually in the tandem repeat cluster and not in immediately adjacent sequences (see Figure 11 and 12). The patient sequence also contained no point mutations.

Repetitive Nature of the Sequence

Repetitive sequences with relatively short basic repeat units (2-25 nucleotides) are known to be responsible for

deletions which lead to length polymorphisms (Stoker et al., 1985; Lee et al., 1991; Sykes et al., 1990). For example, collagen-like sequences on chromosome 11 have substantial length polymorphisms from variable numbers of a tandem repeat (Brooks et al., 1989). There are reports of a region adjacent to the COL2A1 gene which exhibits a high degree of length polymorphism due to the variation in the number of copies of a simple tandemly repeating dinucleotide sequence (Sykes et al., 1985; Stoker et al., 1985). The COL3A1 gene contains a polymorphic block of dinucleotide repeats in one intron (Lee et al., 1991). Direct repeats have been shown to cause deletions in the helical coding portions of the COL1A1 and COL1A2 collagen genes leading to the OI phenotype (Chu et al., 1985; Hawkins et al., 1991; Willing et al., 1988). A 38bp insertion/deletion is a common polymorphism within an intron of the COL1A2 gene (Dalgleish et al., 1986). Recombination between direct Alu sequence repeats can cause clinical pathologies due to deletion of coding information (Stoppa-Lyonnet et al., 1990).

The organization of the 18bp tandem repeat of NT899 is much like that of the direct repeats which cause polymorphisms seen in other collagen and non-collagen genes (Richards et al., 1992; Zuliani et al., 1990). Each repeating unit, AAG ACC AGC AGC CCA GAC, has an average of 93.3% identity to this sequence. Since there is high identity between the repeats of NT899, and they extend over

at least 617 base pairs, misalignment during DNA replication not only seems possible, but highly probable. Also, since most of the deletions seen in the OI patients are in the tandem repeat cluster, it strongly suggests that this repetitive organization facilitates the deletions. The absence of point mutations in the patient sequence eliminates them as the cause of the deletions. It is therefore likely that the direct repeat organization is the cause of the length variations seen at the Site 2 locus.

Deletion Mechanism

Direct repeats are frequently involved in causing deletions, including within collagen genes (Chu et al., 1985; Hawkins et al., 1991; Richards et al., 1992). The probable mechanism involved is slipped mispairing (Richards et al., 1992), where during replication one direct repeat misaligns with another on the complementary strand causing looping out and subsequent excision of the intervening DNA (Efstradiadis et al., 1980). Also, homologous chromosomes may misalign within the tandem cluster during meiosis, causing some of the DNA to loop out, creating deletions and frameshift mutations.

Such repeat-mediated deletions appear to be occurring at Site 2. Since the deletions occur spontaneously (Doering et al, 1987) the slipped mispairing must be occurring during meiosis in the parent of an OI child or during early cell divisions in the embryo that will develop OI. The number of

tandem repeats may be a factor in the occurrence of slipped mispairing and deletions. Large clusters would facilitate more frequent slipped alignments. The number of tandem 18bp clusters found in NT899 is apparently large enough to make the region deletion prone. These tandem repeat clusters are thus unstable and can change each time they are transmitted from parent to child (Sutherland and Richards, 1994). Also, deletions are often found that extend from tandemly repeated sequences into adjacent regions (Zuliani et al., 1990; Willing et al., 1988). Thus, in some cases, as in one patient described in the present study, Site 2 deletions could be extending outside the 18bp tandem repeats. Slipped-mispairing would not be the cause of this deletion. The specific location or extent of deletion within Site 2 may play a role in the expression of the OI phenotype. It is possible that individuals heterozygous for small deletions may be less affected, if at all, than individuals who are either heterozygous for larger deletions or homozygous for deletions at this locus.

Relation to Other Repeat Disorders

Fascioscapulo-humeral muscular dystrophy (FSHD) is associated with deletions in a tandemly repeated sequence (Winokur et al., 1994). The cluster of 3.2kb tandemly repeated unit is located on chromosome 4 and deletions in this region of at least 6kb show tight linkage to the

disease. The repeat cluster is actually adjacent to the FSHD gene, and how the deletions affect its expression is not yet known. Slipped-strand mispairing can also result in the duplication or amplification of repeated sequences. Triplet repeat clusters which expand in size are the sites of mutation in three heritable disorders, myotonic dystrophy, fragile X syndrome, and spinal and bulbar muscular atrophy. In these disorders, premutation or precursors of the full mutation alleles contain triplet-repeat sequences which, when amplified beyond a crucial threshold number, result in mutated alleles that cause the expression of the disease (Caskey et al., 1992). Parents with repeat cluster sizes below a specific threshold have mild or no disease, but give rise to significantly amplified repeats in progeny, who have the full disease phenotype (Caskey et al., 1992). The sequence amplifications that result in the disease phenotype in these patients is analogous to the Site 2 deletions associated with Type III OI. The parental deletions we have observed in OI families may be analogous to the premutation alleles. They could then give rise to the OI phenotype in offspring by further slipped-mispairing deletions. Further work will investigate if such premutation alleles exist in OI patients.

An Expressed Collagen-Like Gene

There is strong genetic evidence that Site 2 is expressed, and that some of its deleted genotypes can lead to a detrimental phenotype against which selection operates Doering et al., 1993). The association of Site 2 deletions with OI Type III is one example of such a phenotype.

Computerized analysis of the Site 2 sequence shows the presence of long open reading frames in the reverse translation frames (left to right in Figure 1). Analysis of potential reading frames, codon preference, and GC bias suggests that the reverse translation frames contain coding information that could occur in a gene that is expressed (Cammarata, 1994). The data are consistent with the presence of an expressed, collagen-like sequence being present on the reverse strand.

Dot blot hybridizations of total RNA from normal diploid human fibroblasts to p4.6 sequences confirms that Site 2 is transcribed (Breslin et al., 1995). Transcription experiments indicate that the gene is transcribed in the left to right direction on the restriction map as seen in Figure 1 (Breslin et al., 1995). This is consistent with all the computer data that the Site 2 locus is an expressed collagen-like gene.

Possible Function of the Site 2 Locus

The level of GC richness (63%) seen in the NT899 sequence is similar to that of other collagen and connective tissue (ECM) genes (Bernard et al., 1983).Previous sequence comparisons of NT899 (Cammarata, 1994), have identified high similarity to a number of ECM components and transmembrane glycoproteins, which are important in directing the composition and structure of the extracellular matrix and effecting cellular interactions with the matrix (Alberts et al., 1994). This supports a function in which the putative protein encoded by the Site 2 locus could modulate the interaction of Type I collagen fibrils with the cell by acting as a communication bridge between them (Shaw and Olsen, 1991). This is possible via a signaling function which directs the secretion of ECM components or in a structural capacity, somehow connecting or associating the major fibril with other cell-secreted ECM components (Muller-Glauser et al., 1986).

The sequence pattern in the Site 2 gene is an interrupted collagen triplet sequence which contains a substitution for a glycine rather than a missing glycine. This pattern suggests that this molecule has the potential for flexibility while still maintaining a collagen-like α-helical structure (Sandell et al., 1990). Cysteine residues allow inter- and intra-molecular crosslinking of the collagen polypeptide and are important in determining the structure of

the terminal globular domains of the Type IX molecules (Ninomiya et al., 1990). The abundance of these residues in p4.6 suggest that the gene product has the capacity for forming other higher order structures in addition to a typical α-helix.

Abnormal Type I collagen incorporated into fibrils has a drastic effect on tissue integrity, resulting in the OI phenotype (Prokop et al., 1989). Since the Site 2 collagen-like gene is associated with the OI phenotype (Doering et al., 1987), it is likely that defects in its protein product, which result from deletion, will also contribute to the OI phenotype. Because the exact structure of the protein has not yet been identified, it is difficult to say specifically how these deletions affect its structure.

If the Site 2 gene product does modulate Type I fibril interactions, then it has the potential to affect the strength and structural integrity of Type I collagen containing tissue. Thus, defects in p4.6 that disrupt this interactive capacity, could in turn effect the expression of the OI phenotype where a Type I collagen defect is or is not present. For example, there are many cases described where substitutions for glycine in similar positions of the same Type I collagen chain can produce drastic differences in OI severity between individuals (Cohn et al., 1991; Rose et al., 1995; Forlino et al., 1994). It is possible that the additional presence of a Site 2 mutation could cause the more

severe condition, such as that seen in Type III OI. In some cases a single defect at the Site 2 locus could in itself result in the Type III OI phenotype.

The deleted allele at Site 2 may not cause OI directly, but rather the defective product of the locus may interact with other mutant connective tissue components (such as Type I collagen) in producing the OI phenotype. There are many examples of asymptomatic individuals carrying the same Type I collagen defect as their family members that have OI (Cohn and Byers, 1991; Constantinou et al., 1989; Hollister et al., 1988). This suggests that the disease phenotype requires an interaction between Type I collagen and the products of other loci. Site 2 is a very good candidate for such a locus. Future work will determine if the size or location of Site 2 deletions can be correlated with severity of symptoms in patients.

Importance as a Diagnostic Tool

Site 2 does not code for a true collagen molecule, but rather a protein with short collagen-like domains (Thiel and Reid, 1989). Whatever its product, the newly discovered locus is of importance, since our studies on OI patients and their families demonstrate that deletions at these sites are clearly associated with the disease and might serve as reliable diagnostic/genetic counseling markers for Type III OI. PCR screening of the repeat/deletion region could be a

simple, effective diagnostic indicator. This could be of real value for the many Type III cases not associated with Type I collagen defects, and could help in predicting severity of cases that do contain Type I defects.

The PCR primers that flank the tandem repeat region in Site 2 amplify a 711bp sequence, which is a much more sensitive means for detecting deletions than the Southern blot protocol (Doering et al., 1993), which used the entire 4.6kb Site 2 fragment. Being able to isolate the repeat region in a small fragment should detect short deletions and heterogeneity that previously went undetected. Deletions at the site would result in shorter than normal fragments being generated by the PCR. Simple comparison of the sizes of the DNA fragments generated from each sample would permit identification of individuals with a deletion, and what size the deletion is. Any sample that shows a Site 2 deletion by Southern blotting but does not show a deletion by PCR, would indicate that the deletion occurs outside the tandem repeat region.

Future Studies

We have already demonstrated the use of cycle sequencing to directly sequence PCR-amplified fragments in the Site 2 deletion region. It will also be important to examine the detailed sequences of deletions in some patients and their asymptomatic parents. Comparing the detailed

structures of the deletions in a patient and the asymptomatic parent could further elucidate the deletion mechanism and indicate whether the parent's deletion is an intermediate in the process of converting a normal allele to the patient's allele. Any correlation between the deletion size and severity of symptoms could eventually be a useful diagnostic tool. Also deletions in the normal population should be studied to see how their deletion sizes compare to those in patients and their parents. Comparing the normal deletions to those in asymptomatic parents would show if there is a premutation not found in the normal population, which could serve as a marker to identify individuals predisposed to having a child with the OI phenotype.

REFERENCES

Aitchison, K., Ogilvie, D., Honeyman, M., Thompson, E. and Sykes, B. (1988). Homozygous osteogenesis imperfecta unlinked to collagen I genes. Hum. Genet. 78,233-236.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994). Molecular Biology of the Cell. (Garland Publishing Inc, New York). pp.978-984.

Baldwin, C.T., Constantinou, C.D., Dumars, K.W. and Prockop, D.J. (1989). A single base mutation that converts glycine 907 of the a2 I chain of Type I procollagen to aspartate in a lethal variant of osteogenesis imperfecta. The single amino acid substitution near the carboxyl terminus destabilizes the whole triple helix. J. Biol. Chem. 264,3002-3006.

Barsh, G.S. and Byers, P.H. (1981). Reduced secretion of structurally abnormal Type I procollagen in a form of osteogenesis imperfecta. Proc. Natl. Acad. Sci. USA 78,5142-5146.

Barsh, G.S., Roush, C.L., Bonadio, J., Byers, P.H. and Gelinas, R.E. (1985). Intron-mediated recombination may cause a deletion in an a1 type I collagen chain in a lethal form of osteogenesis imperfecta. Proc. Natl. Acad. Sci. USA 82,2870-2874.

Bateman, J.F., Chan, D., Lamande, S., Mascara, T. and Cole, W.G. (1988a). Biochemical heterogeneity of Type I collagen mutations in osteogenesis imperfecta. Ann. N.Y. Acad. Sci. 543,95-105.

Bateman, J.F., Lamande, S.R., Dahl, H-H.M., Chan, D. and Cole, W.G. (1988b). Substitution of arginine for glycine 664 in the collagen a1 (I) chain in lethal perinatal osteogenesis imperfecta. Demonstration of the peptide defect by in vitro expression of the mutant cDNA. J. Biol. Chem. 263,11627-11630.

Bateman, J.F., Hannagan, M., Chan, D. and Cole, W.G. (1991). Characterization of Type I collagen a2 (I) glycine-586 to valine substitution in osteogenesis imperfecta Type IV. Biochem. J. 276,765-770.

Bjourson, A.J. and Cooper, J.E. (1992). Band-stab pcr: a simple technique for the purification of individual pcr products. Nucl. Acids Res. 20,4675.

Bonadio, J. and Byers, P.H. (1985). Subtle structural alterations in the chains of Type I procollagen produce osteogenesis imperfecta Type II. Nature 316,363-366.

Bornstein, P. and Sage, H. (1980). Structurally distinct collagen types. Ann. Rev. Biochem. 4,957-1003.

Breslin, P.W., Michaels, M.A., Cammarata, S.A., Burket, A.E. and Doering, J.L. (1995). Characterization of deletions in a new human collagen-like gene. Am. J. Hum. Genet. 57 (Suppl.),141.

Brookes, A.J., Hedge, P.H. and Solomon, E. (1989). A highly polymorphic locus on chromosome 11 which has homology to a collagen triple-helix coding sequence. Nucl. Acids Res. 17,1792.

Byers, P.H., Bonadio, J.F., Steinmann, B., Barsh, G.S., Holbrook, K.A., Greenberg, C., Rowe, D.W. and Gelinas, R. (1983a). Molecular heterogeneity in perinatal lethal osteogenesis imperfecta (OI Type II). Am. J. Hum. Genet. 35,39A.

Byers, P.H., Shapiro, J. R., Rowe, D.W., David, K.E. and Holbrook, K.A. (1983b). Abnormal a2-chain in Type I collagen form a patient with a form of osteogenesis imperfecta. J. Clin. Invest. 71, 689-697.

Byers, P.H., Wenstrup, R.J., Bonadio, J.F., Starman, B. and Cohn, D.H. (1987). Molecular basis of inherited disorders of collagen biosynthesis: Implications for prenatal diagnosis. Curr. Probl. Derm. 16,158-174.

Byers, P.H., Tsipouras, P., Bonadio, J.F., Starman, B.J. and Schwartz, R.C. (1988a). Perinatal lethal osteogenesis imperfecta OI TypeII: A biochemically heterogeneous disorder usually due to new mutations in the genes for Type I collagen. Am. J. Hum. Genet. 42,237-248.

Byers, P.H., Bonadio, J.F., Cohn, D.H., Starman, B.J., Wenstrup, R.J. and Willing, M.C. (1988b). Osteogenesis imperfecta: The molecular basis of clinical heterogeneity. Ann. N.Y. Acad. Sci. 543,117-128.

Byers, P.H., Starman, B.J., Cohn, D.H. and Horwitz, A.L. (1988c). A novel mutation causes a perinatal lethal form of osteogenesis imperfecta. An insertion in one a1 (I) collagen allele COL1A1. J. Biol. Chem. 263,7855-7861.

Byers, P.H. (1989). Inherited disorders of collagen gene structure. Am. J. Med. Genet. 34,72-80.

Byers, P.H. (1990). Brittle bones-fragile molecules: Disorders of collagen gene structure and expression. Trends in Genetics 6,293-300.

Byers, P.H., Wallis, G.A. and Willing, M.C. (1991). Osteogenesis imperfecta: Translation of mutation to phenotype. J. Med. Genet. 28,433-442.

Cammarata, S.A., Burket, A.E. and Doering, J.L. (1991). Characterization of a new collagen locus associated with osteogenesis imperfecta. J. Cell Biol. 115,106a.

Cammarata, S.A. (1994). Cloning and characterization of a new human collagen locus. M.S. Thesis, Loyola University Chicago.

Caskey, C.T., Pizzuti, A., Fu, Y-H., Fenwick, R.G. and Nelson, D.L. (1992). Triple repeat mutations in human disease. Science 256,784-789.

Chu, M-L., Gargiulo, V., Williams, C.J. and Ramirez, F. (1985). Multiexon deletion in an osteogenesis imperfecta variant with increased Type III collagen mRNA. J. Biol. Chem. 260,691-694.

Cohn, D.H., Apone, S., Eyre, D.R., Starman, B.J., Andreassen, P., Charbonneau, H., Nucholls, A.C., Pope, F.M. and Byers, P.H. (1988a). Substitution of cysteine for glycine within the carboxyl-terminal telopeptide of the a1 chain of Type I collagen produces mild osteogenesis imperfecta. J. Biol. Chem. 26,14605-14607.

Cohn, D.H., Wenstrup, R. and Willing, M. (1988b). General strategies for isolating the genes encoding Type I collagen and for characterizing mutations which produce osteogenesis imperfecta. Ann. N.Y. Acad. Sci. 543,129-135.

Cohn, D.H. and Byers, P.H. (1991). Cysteine in the triple helical domain of the pro a2 (I) chain of Type I collagen in nonlethal forms of osteogenesis imperfecta. Hum. Gen. 87,167-172.

Cole, W.G., Jaenisch, R. and Bateman, J.F. (1989). New insights into the molecular pathology of osteogenesis imperfecta. O.J. Med. 261,1-4.

Constantinou, C.D., Nielsen, K.B. and Prockop, D.J. (1989). A lethal variant of osteogenesis imperfecta has a single base mutation that substitutes cysteine for glycine 904 of the a1(I) chain of Type I procollagen. The asymptomatic mother has an unidentified mutation producing an overmodification and unstable Type I procollagen. J. Clin. Invest. 83,574-584.

Dalgleish, R., Williams, G. and Hawkins, J.R. (1986). Length polymorphism in the pro α2(I) collagen gene: an alternative explanation in a case of marfan syndrome. Hum, Genet. 73,91-92.

Deak, S.B., Scholz, P.M., Amenta, P.S., Constantinou, C.D., Levi-Minzi, S.A., Gonzalez-Lavin, L. and Mackenzie, J.W. (1991). The substitution of arginine for glycine 85 of the a1(I) procollagen chain results in mild osteogenesis imperfecta. The mutation provides direct evidence for three discrete domains of cooperative melting of intact Type I collagen. J. Biol. Chem. 266,21827-21832.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A Comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12,387-395.

de Vries, W.N. and de Wet, W.J. (1986). The molecular defect in an autosomal dominant form of osteogenesis imperfecta. Synthesis of Type I procollagen containing cysteine in the triple-helical domain of pro-a1(I) chains. J. Biol. Chem. 261,9056-9064.

de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M-L., Dickson, L., Weil, D. and Ramirez, F. (1987). Organization of the human pro-a2 (I) collagen gene. J. Biol. Chem. 26, 160032-16036.

Doering, J. (1977). The structure of X. borealis oocyte and somatic 5s DNAs. Carnegie Inst. Wash. Yrbk 76,102-105.

Doering, J.L., Jelachich, M.L. and Hanlon, K.M. (1982). Identification and genomic organization of human tRNA lys genes. FEBS lett. 146,47-51.

Doering, J.L., Burket, A.E. and Vogel, L.C. (1987). Collagen gene deletions in osteogenesis imperfecta patients. J. Cell Biol. 105,213a

Doering, J.L., Burket, A.E. and Vogel, L.C. (1990). Collagen gene deletions associated with osteogenesis imperfecta. First regional meeting, The American Society for Cell Biology. 96a.

Doering, J.L., Burket, A.E. and Vogel, L.C. (1993). Length polymorphisms in new human collagen-like loci. FEBS lett. 334, 237-240.

Dublet, B. and van der Rest, M. (1991). Type XIV collagen, a new homotrimeric molecule extracted from fetal bovine skin and tendon, with triple helical disulfide-bonded domains homologous to Type IX and Type XII collagens. J. Biol. Chem. 266,6853-6858.

Efstratiadis, A., Posakony, J.W., Maniatis, T. et al. (1980). The structure and evolution of the human β-globin gene family. Cell. 21,653-668.

Falk, C.T., Schwartz, R.C., Ramirez, F. and Tsipouras, P. (1986). Use of molecular haplotypes specific for the human proa2 (I) collagen gene in linkage analysis of the mild autosomal dominant forms of osteogenesis imperfecta. Am. J. Hum. Genet. 38,269-279.

Fleischmajer, R., Olsen, B.R. and Kuhn, K., eds. (1990). Structure, Molecular Biology, and Pathology of Collagen. Ann. N.Y. Acad. Sci. Vol. 580.

Forlino, A., Zolezzi, F., Valli, M., Pignatti, P.F., Cetta, G., Brunelli, P.C. and Mottes, M. (1994). Severe (type III) osteogenesis imperfecta due to glycine substitutions in the central domain of the collagen triple helix. Hum. Molec. Genet. Vol.3,12, 2201-2206.

Greenspan, D.S., Cheng, W., and Hoffman, G.G. (1991). The pro- a1(V) collagen chain. Complete primary structure, distribution of expression, and comparison with the pro-a1(XI) chain. J. Biol. Chem. 266,24727-24733.

Hawkins, J.R., Superti-Furga, A., Steinmann, B. and Dalgleish, R. (1991). A 9-base pair deletion in COL1A1 in a lethal variant of osteogenesis imperfecta. J. Biol. Chem. 266, 22370-22374.

Hoffman, H., Voss, T. and Kuhn, K. (1984). Localization of flexible sites in thread-like molecules from electron micrographs: Comparison of interstitial, basement membrane and intima collagens. J. Mol. Biol. 172,325-343.

Hollister, D.W., Gaudette, N.D. and Rao, V.H. (1988). Collagen studies in an osteogenesis imperfecta Type III family. Ann. N.Y. Acad. Sci. 543, 62-72.

Innis, M. and Gelfand, D. (1990). In PCR Protocols: A Guide to Methods and Applications. (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) P.9, Academic Press, San Diego.

Kadokami, Y. and Lewis, R.V. (1994). Repeated pcr of a gel band can be used to obtain a single pcr band. Biotechniques. 17,438.

Kozak, K., Foster, L. and Ross, I. (1991). A method for transferring sequencing gels from glass to adsorbent filter paper. Biotechniques. 11,54.

Kuhn, K. (1987). The classical collagens: Types I, II, and III. In Structure and Function of Collagen Types. Mayne, R. and Burgeson, R.E., eds. (Academic Press, New York). pp1-42.

Kuivaniemi, H., Tromp, G. and Prockop, D.J. (1991). Mutations in collagen genes: Causes of rare and some common diseases in humans. FASEB. 5,2052-2060.

Lamande, S.R., Dahl, H.H.M., Cole, W.G. and Bateman, J.F. (1989). Characterization of point mutations in the collagen COL1A1 and COL1A2 genes causing lethal perinatal osteogenesis imperfecta. J. Biol. Chem. 264,15809-15812.

Lee, B., D'Alessio, M., Vissing, H., Ramirez, F., Steinmann, B. and Superti-Furga, A. (1991). Characterization of a large deletion associated with a polymorphic block of repeated dinucleotides in the Type II procollagen gene (COL3A1) of a patient with Ehlers-Danlos syndrome Type IV. Am. J. Hum. Genet. 48,511-517.

Li, K., Sawamura, D., Giudice, G.J., Diaz, L.A., Mattei, M-G., Chu, M-L. and Uitto, J. (1991). Genomic organization of collagenous domains and chromosomal assignment of human 180-kDa bullous pemphigoid antigen-2, a novel collagen of stratified squamous epithelium. J. Biol. Chem. 266,24064-24069.

Mayne, R. and Burgeson, R.E., eds. Structure and Function of Collagen Types. Academic Press, New York. (1987).

Mierendorf, R.C. and Pfeffer, D. (1987). Direct sequencing of denatured plasmid DNA. In Guide to Molecular Cloning Techniques. Berger, S.L., Kimmel, A.R., eds. (Promega, Madison, WI). 152, pp.556-562.

Mottes, M., Sangalli, A., Valli, N., Forlino, A., Gomez-Lira, M., Antoniazzi, F., Constantinou-Deltas, C.D., Cetta, G. and Pignatti, P.F. (1994). A base substitution at IVS-19 3'-end splice junction causes exon 20 skipping in proα2(I) collagen mRNA and produces mild osteogenesis imperfecta. Hum. Genet. 93,681-687.

Muller-Glauser, W., Humbel, B., Glatt, M., Strauli, P. and Winterhalter, K. (1986). On the role of Type IX collagen in the extracellular matrix of cartilage: Type IX collagen is localized to intersections of collagen fibrils. J. Cell Biol. 102,1931-1939.

Muragaki, Y., Kimura, T., Ninomiya, Y. and Olsen, B.R. (1991). The complete primary structure of two distinct forms of human A1 (IX) collagen chains. Eur. J. Biochem. 192,703-708.

Muragaki, Y., Jacenki, O., Apte, S., et al. (1991). The α2(VIII) collagen gene. A novel member of the short chain collagen family located on the human chromosome 1. J. Biol. Chem. 266,7721-7727.

Muragaki, Y., Abe, N., Ninomiya, Y., Olsen, B.R. and Ooshima, A. (1994). The human α1(XV) collagen chain contains a large amino-terminal non-triple helical domain with a tandem repeat structure and homology to α(XVIII) collagen. J. Biol. Chem. 269,4042-4946.

Ninomiya, y., Castagnola, P., Gerecke, D., Gordon, M., Jacenko, O., Lu Valle, P., McCarthy, M., Muragaki, Y., Nishimura, I., Oh, S., Rosenblum, N., Sato, N., Sugrue, R., Taylor, R., Vasios, G., Yamaguchi, N. and Olsen, B. R. (1990). The molecular biology of collagens with short triple-helical domains. In Extracellular Matrix Genes. Sandell, L.J. and Boyd, C.D. eds. (Academic Press, New York). pp80-114.

Nishiyama, T., McDonough, A.M., Bruns, R.R. and Burgeson, R.E. (1994). Type Xii and XIV collagens mediate interactions between banded collagen fibers in vitro and may modulate extracellular matrix deformability. J. Biol. Chem. 269,28193-28199.

Pack, M., Constantinou, C.D., Kalia, K., Nielsen, K.B. and Prockop, D.J. (1989). Substitution of serine for a-1 (I)- glycine-884 in a severe variant of osteogenesis imperfecta minimally destabilizes the triple helix of Type-I procollagen. The effects of glycine substitutions on thermal stability are either position or amino acid specific. J. Biol. Chem. 264,19694-19699.

Pan, T., Zhang, R., Mattei, M., Timpl, R. and Chu, M. (1992). Cloning and chromosomal location of human α1(XVI) collagen. Proc. Natl. Acad. Sci. USA. 89,6565-6569.

Parente, M.G., Chung, L.C., Ryynane, J., et al. (1991). Human Type VII collagen: cDNA cloning and chromosomal mapping of the gene. Proc. Natl. Acad. Sci. U.S.A. 88,6931-6935.

Patterson, E., Smiley, E. and Bonadio, J. (1989). RNA sequence analysis of a perinatal lethal osteogenesis imperfecta mutation. J. Biol Chem. 264,10083-10087.

Peterson, R.C., Doering, J.L. and Brown, D.D. (1980). The characterization of two Xenopus somatic 5s DNAs and one minor oocyte-specific 5s DNA. Cell 20,131-141.

Pihlajaniemi, T., Dickson, L.A., Pope, F.M., Korhonen, V.R., Nicholls, A., Prockop, D.J. and Myers, J.C. (1984). Osteogenesis imperfecta: Cloning of a pro-a2 (I) collagen gene with a frameshift mutation. J. Biol. Chem. 259,12941-12944.

Pope, F.M., Nicholls, A.C., McPheat, J., Talmud, P. and Owen, R. (1985). Collagen genes and proteins in osteogenesis imperfecta. J. Med. Genet. 22,466-478.

Prockop, D.J. and Kivirikko, K.I. (1984a). Heritable diseases of collagen. N. Engl. J. Med. 311,376-386.

Prockop, D. (1984b). Osteogenesis imperfecta: Phenotypic heterogeneity, protein suicide, short and long collagen. Am. J. Hum. Genet. 34 60-67.

Prockop, D.J., Constantinou, D.D., Dombrowski, K.E., Hojuma, Y., Kadler, K.E., Kuivaniemi, H., Tromp, G. and Vogel, B.E. (1989). Type I procollagen: The gene-protein system that harbors most of the mutations causing osteogenesis imperfecta and probably more common heritable disorders of connective tissue. Am. J. Med. Genet. 34,60-67.

Prockop, D.J., Kuivaniemi, H. and Tromp, G. (1994). Molecular basis of osteogenesis imperfecta and related disorders of bone. Clinics in Plastic Surgery 21,407-413.

Reed, K.C. and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. 13, 7207.

Reichenberger, E., Beier, F., LuValle, P., Olsen, B.R., von der Mark, K. and Bertling, W.M. (1992). Genomic organization and full-length cDNA sequence of human collagen X. FEBS Lett. 311,305-310.

Richards, A.J., Lloyd, J.C., Narcisi, P., Nicholls, A.C., Depaepe, A. and Pope, F.M. (1992). A 27-bp deletion from one allele of the Type III collagen gene (COL3A1) in a large family with Ehlers-Danlos syndrome Type IV. Hum. Genet. 88,325-330.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labelling deoxyribonucleic acid to high specific activity in-vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113,237-251.

Rose, N.J., Mackay, K., Byers, P.H., and Dalgleish, R. (1995). A Gly238Ser substitution in the α2 chain of Type I collagen results in osteogenesis imperfecta Type III. Hum. Genet. 95,215-218.

Rose, N.J., Mackay, K., DePaepe, A., Steinmann, B., Punnett, H.H. and Dalgleish, R. (1994). Three unrelated individuals with perinatally lethal osteogenesis imperfecta resulting from identical Gly502Ser substitutions in the α2-chain of type I collagen. Hum. Genet. 94,497-503.

Rosenthal, D.S. and Doering J.L. (1983). The genomic organization of dispersed tRNA and 5s RNA genes in Xenopus laevis. J. Biol. Chem. 258,7402-7410.

Rosenthal, D.S., Doering, J.L., Fokta, F.J. and Jeske, J.B. (1984). Two new tDNA families in Xenopus laevis. J. Cell Biol. 99,253a.

Rowe, D.W., Shapiro, J.R., Poirier, M. and Schlesinger, S. (1985). Diminished Type I collagen synthesis and reduced alpha 1(I) collagen messenger RNA in cultured fibroblasts from patients with dominantly inherited (Type I) osteogenesis imperfecta. J. Clin. Invest. 76,604-611.

Sambrook, J., Fritsch, E. and Maniatis, T. (1989). Molecular Cloning, A Laboratory Manual. Second edition. (Cold Spring Harbor Press, Cold Spring Harbor, NY). pp.2.60-2.111.

Sandell, L.J. and Boyd, C.D. (1990). Conserved and divergent sequence and functional elements within the collagen genes. In Extracellular Matrix Genes. Sandell, L.J. and Boyd, C.D. eds. (Academic Press, New York). pp1-56.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74,5463-5467.

Shaw, L.M. and Olsen, B.R. (1991). FACIT collagens: Diverse molecular bridges in extracellular matrices. TIBS 16,191-194.

Sillence, D.O., Rimoin, D.L. and Danks, D.M. (1979a). Clinical variability in osteogenesis imperfecta: Variable expressivity or genetic heterogeneity. Birth Defects XV,113-129.

Sillence, D.O., Senn, A.S. and Danks, D.M. (1979b). Genetic Heterogeneity in osteogenesis imperfecta. J. Med. Genet. 16,101-116.

Sillence, D.O. (1988). Osteogenesis imperfecta nosology and genetics. Ann. N.Y. Acad. Sci. 543,1-15.

Soininen, R., Huotari, M., Hostikka, S.L., Prockop, D.J. and Tryggvason, K. (1988). The structural genes for a1 and a2 chains of human Type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. J. Biol. Chem. 263,17217-17220.

Soininen, R., Huotari, M., Ganguly, A., Prockop, D.J. and Tryggvason, K. (1989). Structural organization of the gene for the a1 chain of human Type IV collagen. J. Biol. Chem. 264,13565-13571.

Steinmann, B., Nicholls, A., Pope, F.M. (1986). Clinical variability of osteogenesis imperfecta reflecting molecular heterogeneity: Cysteine substitutions in the α1(I) collagen chain producing lethal and mild forms. J. Biol. Chem. 261,8958-8964.

Stoker, N.G., Cheah, K.S.E., Griffin, J.R. and Pope, F.M. and Solomon, E. (1985). A highly polymorphic region 3' to the human Type II collagen gene. Nuc. Acids Res. 13,4613-4622.

Stoppa-Lyonnet, D., Carter, P.E., Meo, T. and Tosi, M. (1990). Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. Proc. Natl. Acad. Sci. 87,1551-1555.

Superti-Furga, A., Pistone, F., Romano, C. and Steinmann, B. (1989). Clinical variability of osteogenesis imperfecta linked to COL1A2 and associated with a structural defect in the Type I collagen molecule. J. Med. Genet. 26,358-362.

Superti-Furga, A., Raghunath, M., Pistone, F.M., Romano, C. and Steinmann, B. (1993). An intronic deletion leading to skipping of exon 21 of COL1A2 in a boy with mild osteogenesis imperfecta. Connec. Tissue Res. 29,31-40.

Sutherland, G.R. and Richards, R.I. (1995). Simple tandem repeats and human genetic disease. Proc. Natl. Acad. Sci. USA 92,3636-3641.

Sykes, B.C., Ogilvie, D.J. and Wordworth, B.P. (1985). Lethal osteogenesis imperfecta and a collagen gene deletion. Length polymorphism provides an alternative explanation. Hum. Genet. 70,35-37.

Sykes, B., Ogilvie, D., Wordsworth, P. et.al. (1990). Consistent linkage of dominantly inherited osteogenesis imperfecta to the Type I collagen loci: COL1A1 and COL1A2. Am. J. Hum. Genet. 46,293-307.

Thiel. S. and Reid, K.B.M. (1989). Structures and functions associated with the group of mammalian lectins containing collagen-like sequences. FEBS Lett. 250,78-84.

Thompson, E.M., Young, I.D., Hall, C.M. and Pembrey, M.E. (1988). Genetic counselling in perinatally lethal and severe progressively deforming osteogenesis imperfecta. Ann. N.Y. Acad. Sci. 543,142-156.

Tikka, L., Elomma, O., Pihlajaniemi, T. and Tryggvason, K. (1991). Human A1 (XIII) collagen gene. Multiple forms of the gene transcripts are generated through complex alternative splicing of several short exons. J. Biol. Chem. 266,17713-17719.

Tiller, G., Rimoin, D., Murray, L. and Cohn, D. (1990). Tandem duplication within a Type II collagen gene (COL2A1) exon in an individual with spondyloepiphyseal dysplasia. Proc. Natl. Acad. Sci. USA 87,3889-3893.

Timpl, R., Wiedemann, H., van Delden, V., Furthmayer, H. and Kuhn, K. (1981). A network model for the organization of Type IV collagen molecules in basement membranes. Eur. J. Biochem. 120,203-211.

Tromp, G. and Prockop, D.J. (1988). Single base mutation in the Proa2(I) collagen gene that causes efficient splicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame proa2(I) chain. Proc. Natl. Acad. Sci. USA 85,5254-5258.

Tsipouras, P. Borresen, A-L., Dickson, L.A., Berg, K., Prockop, D.J. and Ramirez, F. (1984). Molecular heterogeneity in the mild autosomal dominant forms of osteogenesis imperfecta. Am. J. Hum. Genet. 36,1172-1179.

van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H. and Frants, R.R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2kb tandemly repeated unit. Hum. Mol. Genet. 2,2037- 2042.

Vasios, G., Ninomiya, Y. and Olsen, B.R. (1987). Analysis of collagen structure by molecular biology techniques. In Structure and Function of Collagen Types. Mayne, R. and Burgeson, R.E., eds. (Academic Press, New York). pp283-309.

Vogel, B.E., Minor, R.R., Freund, M. and Prockop, D.J. (1987). A point mutation in a Type I procollagen gene converts glycine 748 of the a1 chain to cysteine and destabilizes the triple helix in a lethal variant of osteogenesis imperfecta. J. Biol. Chem. 262,14737-14744.

von der Mark, K., von der Mark, H. and Goodman, S. (1992). Cellular responses to extracellular matrix. Kidney International 41,632-640.

Wenstrup, R.J., Cohn, D.H., Cohen, T. and Byers, P.H. (1988). Arginine for glycine substitution in the triple-helical domain of the products of one a2 (I) collagen allele COL1A2 produces the osteogenesis imperfecta Type IV phenotype. J. Biol. Chem. 263,7734-7740.

Wenstrup, R.J., Willing, M.C., Starman, B.J. and Byers, P.H. (1990). Distinct biochemical phenotypes predict clinical severity in nonlethal variants of osteogenesis imperfecta. Am. J. Hum. Genet. 46,975-982.

Willing, M., Cohn, D., Starman, B., Holbrook, K., Greenberg, C. and Byers, P. (1988). Heterozygosity for a large deletion in the a2 (I) collagen gene has a dramatic effect on Type I collagen secretion and produces perinatal lethal osteogenesis imperfecta. J. Biol. Chem. 263,8398-8404.

Willing, M.C., Pruchno, C.J., Atkinson, M. and Byers, P.H. (1992). Osteogenesis imperfecta Type I is commonly due to a COL1A1 null allele of Type I collagen. Am. J. Hum. Genet. 51,508-515.

Winokur, S.T., Bengtsson, U., Feddersen, J., Mathews, K.D., Weiffenbach, B., Bailey, H., Markovich, R.P., Murray, J.C., Wasmuth, J.J., Alther, M.R. et al. (1994). The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chrom. Res. 2,225-234.

Yamaguchi, N., Benya, P.D., van der Rest, M. and Ninomiya, Y. (1989). The cloning and sequencing of a1 (VIII) collagen cDNAs demonstrate that Type VIII collagen is a short chain collagen and contains triple-helical and carboxy-terminal non-triple-helical domains similar to those of Type X collagen. J. Biol. Chem. 264,16022-160229.

Zuliani, G. and Hobbs, H.H. (1990). A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. Am. J. Hum. Genet. 46,963-969.

VITA

The author, Marissa Ann Michaels, was born on March 22, 1967 in Highland Park, Illinois.

In August 1985, she entered Loyola University Chicago where she received a Bachelor of Science with a major in Biology and a minor in Theology in May 1989.

In August 1991, she entered the Department of Biology of the Graduate School at Loyola University Chicago. In August 1992, she received a fellowship in biology assisting her with her work. A Master of Science degree was completed in April 1996.

In May 1993, she presented at the Graduate Research Forum sponsored by the Loyola University Chicago Chapter of Sigma Xi, where she placed 2nd for oral presentations. In October 1993, she was awarded an Illinois State Academy of Sciences Student Research Grant that assisted with the funding of her research.

PUBLICATIONS

Breslin, P.W., Michaels, M.a., Cammarata, S.A., Burket, A.E. and Doering, J.L. (1995). Characteriztion of deletions in a new human collagen-like gene. Am. J. Hum. Genet. 57 (Suppl.),141.

Michaels, M.A., Cammarata, S.A., Burket, A.E. and Doering, J. (1993). Characterization of deletion sites in a new human collagen-like locus. Am. J. Hum. Genet., 53 (Suppl.), 1625.

Michaels, M.A., Cammarata, S.A., Burket, A.E. and Doering, J. (1993). Characterization of deletion sites in a new human collagen-like locus. New York Academy of Sciences, DNA: The Double Helix, Forty Years: Perspective and Prospective, Poster Session.

Michaels, M.A., Cammarata, S.A., Burket, A.E. and Doering, J.L. (1992). Characterization of a new Human Collagen Locus Associated with Osteogenesis Imperfecta. Sigma Xi, St. Albert's Day Poster Session.

Victor, T.A., Bergman, A., Knop, R.H. and Michaels, M.A. (1991). Measurement of Polyunsaturated Fatty Acid Content of Normal and Cancerous Human Breast Tissue by ^{13}C NMR. Society of Magnetic Resonance in Medicine.

APPROVAL SHEET

The thesis submitted by Marissa A. Michaels has been read and approved by the following committee:

Dr. Jeffrey L. Doering, Director
Professor, Biology
Loyola University Chicago

Dr. John Smarrelli
Professor, Biology
Loyola University Chicago

Dr. William Wasserman
Associate Professor, Biology
Loyola University Chicago

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirement for the degree of Master of Science.

4-29-96
Date

Jeffrey L. Doering
Director's Signature

Marissa A. Michaels

Loyola University Chicago

CHARACTERIZATION OF DELETION SITES IN A HUMAN COLLAGEN-LIKE LOCUS

Osteogenesis imperfecta (OI) is a highly heterogeneous genetic disorder that affects bone and other connective tissues rich in Type I collagen. Previous studies have shown the existence of two previously-undescribed collagen-like loci, with sequence similarity to pro-α2(I) collagen. At these two sites deletions of heterogeneous size are found to be significantly more prevalent in OI type III patients than in the unaffected population.

This study determined the molecular organization of the deletion site in one of the new collagen-like loci (Site 2). Sequencing reveals that the locus contains a region with 34 copies of an 18bp tandemly-repetitive sequence. Using blots of restriction digested DNA from unaffected and OI patients, and PCR primers that flank the cluster of tandem repeats, I demonstrated that the deletions observed at this locus in OI patients usually occur within the tandem cluster. The molecular nature of the deletions was then determined by comparing the sequences of a short allele in an OI patient to the normal allele. The probable mechanism creating the

deletions is slipped mispairing of misaligned repeats during DNA replication.