Order Number 9432368

Structural/functional characterization of fibrillin, the gene product involved in Marfan syndrome

Pereira, Lygia Veiga, Ph.D.

City University of New York, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

STRUCTURAL/FUNCTIONAL CHARACTERIZATION OF FIBRILLIN, THE GENE PRODUCT INVOLVED IN MARFAN SYNDROME.

by

Lygia V. Pereira

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1994

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

March 7, 1994
Date

Francesco Ramirez, Ph.D.
Chair of Examining Comittee

March 9, 1994
Date

Terry A. Krulwich, Ph.D.
Executive Officer

Robert J. Desnick, Ph.D., M.D.

Harry C. Dietz, M.D.

Christine Eng, M.D.

Bruce Gelb, M.D.

Kevin Kelley, Ph.D.

Supervisory Committee

The City University of New York

ABSTRACT

STRUCTURAL/FUNCTIONAL CHARACTERIZATION OF FIBRILLIN, THE GENE PRODUCT INVOLVED IN MARFAN SYNDROME.

by

Lygia V. Pereira

Adviser: Professor Francesco Ramirez

Marfan syndrome (MFS) is a systemic disorder of connective tissue inherited as an autosomal dominant trait. The three major manifestations of the disease - aortic dilatation, ectopia lentis and skeletal malformations - present a wide clinical variability. For several years the molecular lesion in MFS has been the subject of investigation. Recently fibrillin, a major component of microfibrils, was recognized as the defective gene product in MFS.

The major theme of this study was to provide the basic tools for a thorough molecular investigation of the pathogenesis of MFS and of microfibril assembly. Toward this end this work can be divided in two parts. The first one focuses on the human fibrillin gene, providing its complete coding sequence and intron/exon organization, and identifying novel polymorphic markers in order to help the diagnosis of familial cases of MFS. The second part addresses the question of structure/function correlation of fibrillin, using the technology of homologous recombination in embryonic-stem cells (ES-cells).

Completion of the primary sequence of fibrillin, along with elucidation of the gene organization, shed new light on the structure and function of this protein, as well as the evolutionary origin of the fibrillin gene family. In addition, the results of this study have substantially improved the diagnosis and identification of individuals at risk for MFS.

Completion of the fibrillin transcript sequence will now enable the screening of MFS patients for mutations in its entire length. Moreover, the determination of the sequences of all FBN1 intron/exon junctions will greatly facilitate this process by extending the screening to the genomic level. Additionally, the highly informative set of intragenic polymorphic markers identified in this work will be essential for the pre-natal and pre-symptomatic diagnosis of familial cases of MFS.

Finally, this study has initiated a broader body of work aimed at establishing the transgenic mouse as an animal model for the study of fibrillin biosynthesis and microfibril assembly. Isolation of the mouse fibrillin transcript has revealed a high degree of conservation between the human and murine fibrillin. These data were utilized to design vectors for fibrillin gene targeting in mouse ES-cells. The successful generation of ES-cell lines bearing two different mutations in the fibrillin gene has laid the ground for future investigations aimed at characterizing the mechanisms and factors responsible for microfibril pathogenesis.

ACKNOWLEDGMENTS

In chronological order, I would like to thank Dr. Eliana Abdelhay, who took me from the department of physics into the world of molecular biology. Dr. Desnick, for introducing me to the field of human molecular genetics during a seminar he gave in Rio de Janeiro, and encouraging me to do my Ph.D. at Mount Sinai. I am also thankful to Dr. Krulwich for her dedication to setting a high standard to our graduate program, and for the support given to me during more difficult times.

I am most grateful to Dr. Checco Ramirez for his support, guidance, for giving me the opportunity to develop this project, and mostly, for not letting me settle for less than what I was capable of accomplishing. Although the learning process can be painful some (many) times, once it is finished, if ever, one is able to better appreciate the results. It was definitely worth it.

I would like to thank my friends in the program, specially Hui Zhang, for the emotional and technical support. In addition, I would like to thank my squash partners, the New York Chorus Society, and all the social and cultural life of New York city for keeping my sanity during these years. I cannot stress enough the importance of extra-curricular activities for dealing with the frustrations of the bench work.

Finally, I would like to thank all my friends from Brazil and family members, whose frequent visits were always a joy. In special, my parents and my grandmother Lygia, to whom I dedicate this thesis, for giving me so much love.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	vii
BACKGROUND	1
Marfan Syndrome	1
Elastic Fibers, Microfibrils and Fibrillin	5
Animal Models for Marfan Syndrome	12
Thesis Outline	17
MATERIALS AND METHODS	19
RESULTS	30
Human fibrillin cDNA cloning	30
Start Site of Transcription of FBN1	33
Primary Structure of Human Fibrillin	36
Genomic Structure and Exon Organization	39
Identification of Polymorphic VNTRs in the FBN1 Locus	44
Murine Fibrillin cDNA cloning	47
Gene Targeting	50
DISCUSSION	55
Fibrillin Transcript	55
Fibrillin Protein Structure	57
Fibrillin Gene	62
Mouse Fibrillin Gene	65
APPENDIX	72
FBN1 Coding Sequence	72
Allele size and Frequency for FBN1 Intragenic VNTRs	78
BIBLIOGRAPHY	79

LIST OF FIGURES

		Page
Figure 1:	Head-to-tail assembly of fibrillin.	7
Figure 2:	Folding of an EGF-CB motif.	9
Figure 3:	Mutations reported in the FBN1 gene.	11
Figure 4:	Homologous recombination - sequence replacement vector.	15
Figure 5:	Homologous recombination - hit and run protocol.	16
Figure 6:	BEnT targeting vector.	25
Figure 7:	mgΔ(tgf2) targeting vector.	26
Figure 8:	Fibrillin cDNA clones.	32
Figure 9:	Primer extension and RNase protection.	35
Figure 10:	Amino acid sequence of fibrillin.	37
Figure 11:	PCR amplification of intron/exon junctions.	40
Figure 12:	Genomic organization of FBN1.	41
Figure 13:	Intron/exon sequences.	42
Figure 14:	Sequence of VNTRs.	45
Figure 15:	Segregation of VNTRs in MFS families.	46
Figure 16:	Location of VNTRs within the FBN1 gene.	47
Figure 17:	Amino acid comparison between human and mouse fibrillin.	48
Figure 18:	Mapping of *Fbn1*.	50
Figure 19:	Southern analysis of BEnT clones.	52
Figure 20:	Southern analysis of mgΔ(tgf2) clones.	53
Figure 21:	Folding of consecutive EGF-CB motifs in fibrillin.	57
Figure 22:	Assembly of fibrillin into microfibrils.	60
Figure 23:	Postulated origin of cysteine-rich motifs of fibrillin.	61
Figure 24:	Null *Fbn1* allele.	67

BACKGROUND

I - MARFAN SYNDROME:

A - Clinical features:

Marfan syndrome (MFS) is the most common genetic disease of connective tissue. Inherited as an autosomal dominant trait, MFS has an incidence of about 1 in 10,000 individuals. Thirty percent of these cases are sporadic due to *de novo* mutations (McKusick, 1956). The MFS phenotype is highly penetrant. The major clinical manifestations of this disorder primarily affect the following three organ systems:

Skeletal; clinical findings include long limbs, dolichostenomelia (narrow frame relative to height), arachnodactyly (long and spidery fingers), joint hypermobility, scoliosis and anterior chest deformity.

Ocular; in the eye, ectopia lentis (usually upward), myopia and retinal detachment are typical.

Cardiovascular; abnormalities in this system include aortic root dilatation, aortic dissection, regurgitation, and mitral valve prolapse.

The clinical diagnosis of MFS is determined by so-called "major" manifestations of the three affected systems: aortic root dilatation, ectopia lentis and characteristic habitus. Because of the wide clinical variability of MFS, a distinction must be made in the diagnosis of familial and sporadic cases. In the familial cases of MFS, the presence of a major manifestation from one of the affected systems combined with two minor manifestations are sufficient to establish the diagnosis. For the sporadic cases, the diagnosis requires instead the presence of major manifestations in two of the three affected systems plus minor manifestations (Beighton et al., 1988).

The life expectancy for MFS patients is about two thirds of the normal - the average age at death is around 40 years. Cardiovascular failure is the cause of death in approximately 90% of the affected individuals (McKusick, 1956). Early diagnosis followed by medical intervention leads to prolonged life span.

MFS is genetically homogeneous and shares some of its phenotypic features with other disorders. One of these "MFS-related" conditions is dominantly inherited ectopia lentis (EL), presenting only the ocular manifestations of MFS. Interestingly, EL has been mapped to the MFS locus on chromosome 15 suggesting that these two conditions are allelic (Tsipouras et al., 1991). This has been recently confirmed by the identification of an EL patient with a mutation in the MFS gene (Kainulainen et al., 1994). Another interesting example of a MFS-related condition is congenital contractural arachnodactyly (CCA) (Beals and Hecht, 1971). CCA is an autosomal dominant disease of the connective tissue characterized by multiple joint contractures, arachnodactyly, dolichostenomelia and crumpled external ears. Patients with CCA do not present any cardiovascular or ocular abnormalities. Because of its skeletal manifestations, CCA was once thought to be a milder allelic form of MFS. Linkage studies have recently mapped the CCA locus to chromosome 5, thus demonstrating that CCA and MFS are genetically distinct diseases (Lee et al., 1991; Tsipouras et al., 1992).

B - Genetics of Marfan syndrome:

The most common histological finding in MFS patients is necrosis in the wall of the ascending aorta (Byers, 1983). This condition is not present at birth but develops in post-natal life causing dissecting aneurysm of the aorta. Sections from aneurismal aorta display fragmented elastic fibers infiltrated by collagen fibrils and proteoglycans. For this reason, the defect in MFS was originally thought to be in one of the components of the connective tissue, in particular the elastic fiber or some molecule closely associated with it. Several extracellular components of the connective tissue were investigated for their relation to

MFS. They included: elastin, fibronectin, proteoglycans and fibrillar collagens (Abraham et al., 1982; Appel et al., 1979; Boileau et al., 1990; Dalgleish et al., 1987; Francomano et al., 1988; Kainulainen et al., 1990b; Krieg and Muller, 1977; Laitinen et al., 1968; Lamberg, 1978; Nakashima, 1986; Ogilvie et al., 1987; Royce and Danks, 1982; Tsipouras et al., 1986). However, none of these candidate loci showed linkage to MFS.

In 1986, a new 350kD glycoprotein named fibrillin was isolated from the microfibrils of the connective tissue (Sakai et al., 1986). Originally purified from the medium of human fibroblast cell cultures by the use of antibodies raised against a human amnion extract, fibrillin was shown to be the major structural component of microfibrils. Immunohistochemical analysis localized fibrillin to microfibrils of several human tissues, including the skin, aorta, periosteum, perichondrium and ciliary zonules (Sakai et al., 1986). The correlation between the tissue distribution of fibrillin and the MFS phenotype rendered the fibrillin gene a good candidate for the molecular defect underlying MFS. This hypothesis was supported by the finding that anti-fibrillin antibodies cross-reacted less well with the microfibrilar-fiber system of MFS patient skin biopsies than with those of normal control individuals (Hollister et al., 1990). In the skin of unaffected individuals, immunohistochemical staining with anti-fibrillin antibodies produced a characteristic pattern strongly decorating the dermal-epidermal junction in an almost continuous band parallel to the basement membrane. In contrast, analysis of the skin of MFS patients showed both quantitative and qualitative differences in the immunostaining pattern. Samples from affected individuals displayed an overall lower intensity of staining in the dermis; moreover, the immunofluorescence was discontinuously distributed along the basement membrane at the dermal-epidermal junction. Finally, the fibrous meshwork detected in the deeper dermis of control individuals was significantly decreased in MFS patients. Studies of cultured fibroblasts grown to hyperconfluency showed that anti-fibrillin antibodies were able to detect an extensive network of extracellular fibers synthesized and assembled by normal cells. In contrast, fibroblasts from MFS patients had a striking decrease of the

extracellular fibers (Hollister et al., 1990). Further immunohistochemical studies in both skin and cultured fibroblasts of MFS patients and normal control individuals demonstrated the co-segregation of the abnormal immunostaining patterns and the MFS phenotype (Godfrey et al., 1990). However, although the anti-fibrillin antibodies detected a defective microfibrilar system, it remained to be demonstrated whether or not fibrillin was the primary defect in MFS.

Parallel to these biochemical studies, other groups used the alternative approach of positional cloning to identify the MFS gene. In a combined international effort, an exclusion map for MFS was constructed by compiling genetic data from linkage analysis of MFS and 75 informative loci on 18 autosomes (Blanton et al., 1990). As a result, 75% of the loci in the genome were excluded as the MFS locus. Subsequent analysis of Finnish families using markers on the non-excluded chromosomes showed linkage between MFS and 3 polymorphic markers on the long arm of chromosome 15. Linkage studies performed on American families confirmed the assignment of the MFS locus to chromosome 15, narrowing the location of the MFS gene to the chromosomal region 15q15-q21.3 (Dietz et al., 1991b; Kainulainen et al., 1990a; Tsipouras et al., 1991).

Finally, in 1991 the combination of biochemical and genetic analysis led to the elucidation of the basic defect in MFS. Amino acid sequence generated from pepsin-resistant peptides of fibrillin (Maddox et al., 1989) allowed investigators to use degenerate oligonucleotide primers to amplify the fibrillin sequence from cDNA libraries by the polymerase chain reaction (PCR). In this way the fibrillin transcript was partially characterized (Lee et al., 1991; Maslen et al., 1991). The corresponding gene, FBN1, was mapped to the MFS locus, thus rendering fibrillin an even stronger candidate for the primary defect in MFS (Lee et al., 1991). Finally, a missense mutation was detected in the coding sequence of FBN1 in two MFS patients (Dietz et al., 1991a). Fibrillin was therefore recognized as the defective gene product in MFS.

II - ELASTIC FIBERS, MICROFIBRILS AND FIBRILLIN:

A - Elastic fibers:

The physiological function of tissues such as lung, skin and blood vessels requires them to be able to stretch and recoil. This elastic property is provided by elastic fibers present in the extracellular matrix (ECM) of these tissues. Ultrastructural and biochemical analyses have shown that these fibers consist of a 10-12 nm microfibrillar component surrounding an abundant amorphous component consisting of the protein elastin (Fahrenbach et al., 1966; Greenlee et al., 1966; Karrer and Cox, 1961). It is this highly insoluble protein that is mostly responsible for the elastic properties of the elastic fibers (Starcher and Galione, 1976).

During elastogenesis, elastin is secreted into the ECM as a precursor form, tropoelastin (Serafini-Fracassini, 1984). There, the enzyme lysyl oxidase is responsible for extensive cross-linking tropoelastin into what has been described as a rubber-like network. The previously assembled microfibrillar component seems to direct elastin deposition into the elastic fiber.

B - Microfibrils:

Among the several fibrillar structures present in the ECM, a variety of non-striated fibrils with small cross-sectional diameters can be found. Microfibrils were first defined as fibral matrix structures displaying a diameter of less than 20 nm and lacking the banding periodicity of collagen fibers (Low, 1962). More recently, microfibrils have been divided into the 3-5 nm and the 5-10 nm microfibril groups on the basis of their average diameter (Mecham and Hauser, 1992).

Under electron microscopy (EM) analysis, the 10 nm microfibrils appear as thread-like filaments, regularly beaded in longitudinal sections and tubular in cross-sections

(Keene et al., 1991; Ren et al., 1991). In elastic tissues, microfibrils are first deposited in the matrix; there, elastin is subsequently associated with these fibers. Bundles of microfibrils form fibers that surround elastin cores, extending outwards for variable distances (Mecham and Hauser, 1992). Although they are often found in association with elastin, microfibrils also exist devoid of any elastic element, as, for example, in the suspensory ligaments of the lenses (Mecham and Hauser, 1992).

Microfibrils have been immunolocalized to several different tissues, including the periosteum, perichondrium, pleura, aorta, meninges, cartilage, tendon and muscle (Sakai et al., 1986). They are believed to play three basic roles in the ECM (Mecham and Hauser, 1992). Microfibrils form the scaffold where elastin is later deposited, as, for example, in the tunica media of the aorta. In this structure, rings of microfibrils are first seen during embryogenesis, before elastin deposition. Microfibrils also form a fibrous aggregate that links elastin to other matrix components - for instance, at the dermal/epidermal junction, microfibrils serve as a bridge between elastin bundles and the basement membrane. Finally, in tissues that do not contain elastin, such as the ciliary zone, microfibrils appear to serve an anchoring function linking the lenses to the ciliary body of the eye.

Because of the high insolubility of microfibrils, so far only a few of its several components have been characterized. They are: the microfibril-associated glycoprotein (MAGP) (Gibson et al., 1991), a 34KDa glycoprotein with amino oxidase activity (Serafini-Fracassini et al., 1981), the enzyme lysyl oxidase (Kagan et al., 1986), the 58kDa associated microfibril protein (AMP) (Horrigan et al., 1992), emilin (Bressan et al., 1993), and fibrillin (Sakai et al., 1986)

C - Fibrillin:

Fibrillin was first purified from the medium of human fibroblast cell cultures by Sakai et al. in 1986 (Sakai et al., 1986). Different monoclonal antibodies prepared against a microfibrillar extract recognized the same 350kD protein which was named fibrillin.

Under non-reducing conditions, fibrillin displayed a lower molecular weight, indicating the presence of intrachain disulfide bonds. In addition, this protein was shown to be glycosylated by glucosamine labeling (Sakai et al., 1986).

Further characterization of fibrillin came in subsequent studies in which pepsin digests of this protein were utilized (Maddox et al., 1989). EM rotary shadowing analysis of these peptides suggested that most of the fibrillin molecule is rod-shaped. In addition, another substructure of this protein was identified as a central dense core with flexible arms extending from it, possibly an aggregate of several fibrillin fragments. It was speculated that this region may be a cross-linking domain where fibrillin molecules interact with each other through disulfide bonds (Maddox et al., 1989).

Antibodies raised against the pepsin digestion products of fibrillin were used to study the arrangement of fibrillin into microfibrils (Sakai et al., 1991). Results from the immunolabelling experiments indicated that fibrillin molecules are assembled into microfibers in a head-to-tail configuration. As mentioned before, under EM analysis microfibrils appear as long beaded strings (Gibson et al., 1989; Keene et al., 1991). Accordingly, a model developed from the results of these experiments proposes that the bead structure is the site of interaction between "heads" and "tails" of several fibrillin molecules (Sakai et al., 1991). In addition, the periodicity in which those antibodies decorate microfibers suggested that the beads are separated by only one fibrillin-length. Thus, the strings would be composed of fibrillin molecules assembled in parallel (fig. 1).

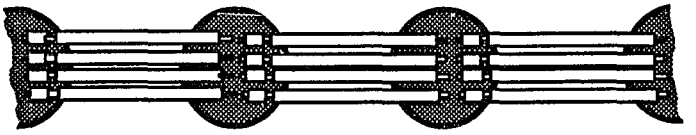


Figure 1: Head-to-tail assembly of fibrillin in the beaded structures of the microfibrils.

As discussed above, degenerate oligonucleotides were derived from the partial amino acid sequence of fibrillin peptides and used to isolate fibrillin sequences from cDNA

libraries (Lee et al., 1991; Maddox et al., 1989; Maslen et al., 1991). Consistent with the molecular weight of the protein, Northern analysis showed that the fibrillin cDNA clones detected an approximately 10kb mRNA (Lee et al., 1991). Taken together, 70% of the full-length fibrillin message was isolated in these studies (Lee et al., 1991; Maslen et al., 1991).

One striking feature noted in this early work was that fibrillin seemed to be mostly composed of several repeated cysteine-rich motifs, homologous to motifs found in the epidermal growth factor (EGF) precursor molecule. The hallmark of this motif is the presence of six cysteine residues spaced in a specific fashion (Davis, 1990). The tertiary structure of human EGF has been elucidated by nuclear magnetic resonance (NMR): disulfide bonds between cysteine residues 1- 3, 2 - 4 and 5-6 stabilize the molecule into a tightly folded structure, with an amino-terminal triple-stranded β-sheet (fig. 2) (Cooke et al., 1987). The EGF-like motifs found in fibrillin contained an additional consensus of D, (D/N), N* and Y/F (where the asterisk denotes a potential hydroxylation site), identifying them with a sub-class of EGF-motifs involved in calcium binding (EGF-CB) (Handford et al., 1990) (fig. 2). This early work also presented some very preliminary data on the organization of the FBN1 gene (Lee et al., 1991). Characterization of three exons suggested that the FBN1 gene follows the same modular arrangement as the fibrillin protein, for each one of those exons encoded one EGF-like motif.

During the search for the fibrillin gene, Lee et al. isolated a cDNA whose predicted amino acid sequence shows 80% homology to fibrillin. The corresponding gene, FBN2, was mapped to chromosome 5 and linked to CCA (Lee et al., 1991). The unexpected finding of a fibrillin-like gene was the first indication of the possible existence of a fibrillin protein family.

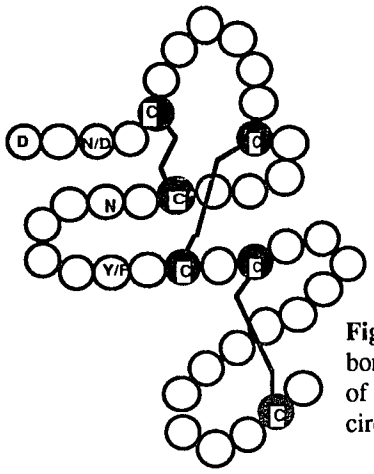


Figure 2: Predicted folding of an EGF-CB motif. Disulfide bonds between cysteines are represented by solid lines. Residues of the calcium binding consensus are indicated in the white circles.

D - Fibrillin mutations:

Very little is currently known about the biosynthesis of fibrillin and its incorporation into microfibrils. One study of fibrillin synthesis in cell culture suggested that this protein is synthesized as a precursor form, secreted into the media, processed extracellularly to a lower molecular weight product of 320kD, and incorporated into the matrix (McGookey-Milewicz et al., 1992). The same study identified three distinct types of abnormalities in fibrillin metabolism in fibroblasts established from MFS patients (McGookey-Milewicz et al., 1992). One group of patients presented decreased synthesis of the protein. Another group had normal levels of fibrillin synthesis, but the protein was inefficiently secreted from the cells. The third group made normal amounts of fibrillin, secreted it efficiently, but the secreted protein was not incorporated into the ECM.

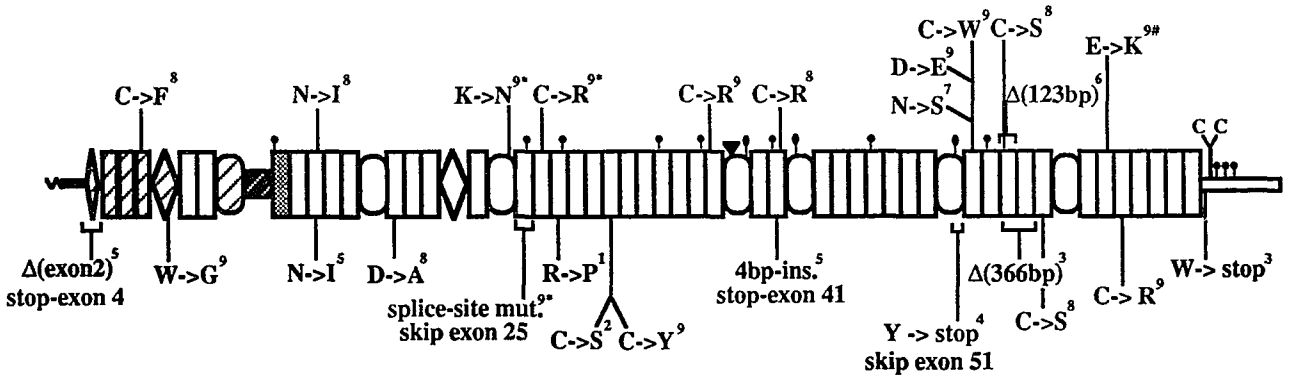
As a result of the cloning of the fibrillin gene, the number of FBN1 mutations characterized in MFS patients is progressively increasing. They can be divided into three general groups. The first consists of 10 missense mutations altering residues of the EGF-CB consensus sequence (Dietz et al., 1992a; Dietz et al., 1992b; Dietz et al., 1993a; Hewett et al., 1993); they are believed to affect either the folding or the calcium binding ability of the EGF-CB motifs. The second group comprises 6 mutations that generate shortened forms of fibrillin by creating early stop-codons or in-frame exon deletions (Dietz et al.,

1993a; Dietz et al., 1993b; Godfrey et al., 1993; Kainulainen et al., 1992). Finally, the third group consists of 3 missense mutations that alter residues others than those of the EGF-CB consensus (Dietz et al., 1991a; Kainulainen et al., 1994). Figure 3 summarizes the current status of the molecular defects identified in the FBN1 gene.

Despite the increasing number of FBN1 mutations being characterized, the precise mechanism by which mutated fibrillin molecules disrupt microfibril formation remains unclear. By analogy to the collagenopathies, the abnormal fibrillin is believed to act in a dominant negative fashion during microfibril assembly (Bonadio et al., 1990; Dietz et al., 1993a). According to this model, mutant fibrillin molecules would interact with their normal counterparts and disrupt the overall organization of these fibers by being incorporated into microfibrils (Dietz et al., 1993a).

It is still unclear whether or not microfibril formation is also susceptible to a gene dosage effect, requiring both FBN1 alleles to be expressed. The dominant-negative mechanism proposed above predicts that a null-allele in heterozygosity would not be deleterious for microfibril assembly. In support of this model, null alleles of FBN1 have not yet been identified in MFS patients. On the other hand, this may be due to the currently used PCR-based strategies which are not able to detect such mutant alleles unless in homozygosity. A recent report described an FBN1 allele with a frame-shift mutation whose level of expression is only 6% of the normal (Dietz et al., 1993a). Interestingly, the patient carrying this almost-null allele has a mild variant of the MFS phenotype. The authors proposed that the patient's mild phenotype may be due to the reduced amount of mutant fibrillin monomers, which allows a preferential interaction among normal fibrillin molecules during microfibril assembly (Dietz et al., 1993a). Although this is a plausible explanation, the effect of quantitative changes in fibrillin awaits more direct evidence.

As previously mentioned, MFS presents a high intrafamilial clinical variability, rendering its diagnosis difficult in many cases (Pyeritz and McKusick, 1979). The elucidation of the molecular defect in MFS brought hopes that the clinical diagnosis of this



(1)Dietz et al., 1991; (2)Dietz et al., 1992; (3)Kainulainen et al., 1992; (4)Dietz et al., 1993a; (5)Dietz et al., 1993b; (6)Godfrey et al., 1993; (7)Hewett et al., 1993; (8)Dietz et al., 1993c; (9)Kainulainen et al., 1994.

Figure 3: Mutations reported in the FBN1 gene. Fibrillin molecule is schematically represented with the EGF-like (rectangles), TGFbp (oval), and Fib (diamond) motifs. Asteriscs denotes mutations found in neonatal-MFS. The mutation found in a EL patient is noted by (#). References are indicated by numbers on top.

disease would be replaced, or at least complemented, by molecular diagnosis. Unfortunately, mutational analysis of FBN1 has so far indicated that no common DNA change is likely to exist in MFS (Sykes, 1993). In addition, due to the large size of the fibrillin transcript, the rate of mutation detection in MFS patients has been relatively slow (Sykes, 1993). These two factors have hampered the development of pre-natal and pre-symptomatic diagnosis of MFS. Theoretically, this could be performed by linkage analysis in familial cases by the use of restriction fragment length polymorphisms (RFLP) and, more recently, variable number of tandem repeat (VNTR) polymorphisms. Briefly, RFLPs are variations of the DNA sequences that either create or destroy a restriction site, allowing the segregation of different alleles to be followed (Botstein et al., 1980). VNTRs, on the other hand, consist of short tandemly repeated DNA sequences, usually CA-dinucleotides (Wyman and White, 1980). Due to their repetitive nature, these sequences are highly polymorphic due to a variability in the number of tandem repeats (thus the name VNTR). Unfortunately, the two RFLPs (Clark et al., 1991; Hewett et al., 1991) and the VNTR (Lee et al., 1991) originally identified in the FBN1 locus proved to be of low informative value for molecular diagnosis.

III - ANIMAL MODELS FOR MFS:

Animal models for human genetic diseases are important tools for the study of the pathologies associated with a specific disorder. In addition, these animals serve as a system in which new therapies can be implemented.

One naturally occurring bovine model for MFS has been described (Besser et al., 1990). Affected cattle present with joint hypermobility, dolichostenomelia, ectopia lentis and aortic dilatation. As in human MFS, the major cause of death in these animals is aortic dissection. A recent biochemical study using an anti-fibrillin antibody indicated that

fibroblasts from affected animals present less immunoreactive fibrillin than normal control cells (Potter et al., 1993). In addition, a decreased fibrillin synthesis and incorporation into the ECM was detected in cells of the affected cattle. Although these findings are similar to those reported in human MFS, it remains to be demonstrated whether or not the bovine MFS is a *bona fide* model for the human disease, i.e., whether it is caused by mutations in the bovine fibrillin gene. Despite its potential usefulness, the bovine model has many practical limitations, including time of gestation, small litter size and maintenance cost.

The recent development of techniques for the manipulation of the mouse genome renders the transgenic mouse the best available model system for the study of human genetic diseases (Capecchi, 1989). The successful germ line propagation of foreign DNA microinjected into the pronucleus of fertilized mouse eggs was first achieved in 1980 (Gordon et al., 1980). The injected exogenous gene was shown to integrate randomly in the genome, and then to be passed onto the mouse offspring in a mendelian fashion. Since that report, pronuclear microinjection has been extensively used to generate animal models for several dominantly inherited diseases (Kivirikko, 1993), including osteogenesis imperfecta (OI) (Stacey et al., 1988). The most severe form of this disease, OI type II, is characterized by poorly mineralized skeleton and reduced amounts of collagen; individuals affected by OI type II die *in utero* or soon after birth. Mutations in the pro-α(1) and pro-α(2) collagen genes (COL1A1 and COL1A2) have been associated with this condition. The introduction into mice of a COL1A1 allele carrying the same mutation as an OI type II individual generated a dominant lethal phenotype in those animals similar to the human disease (Stacey et al., 1988). This method however presents some problems. Because of the random site of transgene integration, this may not be under the control of all the cis-acting elements which normally regulate endogenous gene expression. Thus, the temporal and spatial pattern of expression of the transgene may not mimic that of its endogenous counterpart. In addition, the introduction of a third allele, the mutant transgene, creates an

artificial situation with respect to the ratio of normal to mutant transcripts. This ratio may be critical for conditions susceptible to gene dosage effects.

Recently, the combination of embryonic stem (ES) cell culturing and homologous recombination has created a more powerful method for modifying the mouse genome. ES-cell lines are derived from the inner-cell mass of mouse blastocysts (Evans and Kaufman, 1981). These cells can be maintained in culture while retaining their pluripotency. When injected into mouse blastocysts, ES-cells are able to populate different tissues, including the germ line (Bradley et al., 1984).

Thus, the ability to modify specific regions of the mouse genome by homologous recombination in ES-cells potentially allows the generation of mice with any desired genotype. The most commonly used vector for homologous recombination is the *sequence replacement* vector containing the neomycin resistance (neo^r) gene flanked by sequences homologous to the target gene (Capecchi, 1989) (fig. 4). In addition, the thymidine kinase (tk) gene, cloned outside of the region of homology, provides a selectable marker to enrich the pool of clones for homologous recombination events. Such an event will cause the loss of the tk gene, rendering cells resistant to the drug gancyclovir. Thus, through two reciprocal recombination events, the *sequence replacement* vector mediates the replacement of endogenous DNA with exogenous sequences, allowing the disruption of a gene by insertions or deletions.

Lately, a new type of vector has been designed to allow the introduction of more subtle gene modifications such as point mutations (Hasty et al., 1991). The *hit and run* targeting vector consists of a homology region, containing the desired mutation, adjacent to the neo and tk genes placed in the plasmid backbone (fig. 5). The first targeting event occurs by a single reciprocal recombination which generates a duplication of the homologous sequence. In the second step, a reciprocal intrachromosomal recombination may resolve the duplication, removing the neo and tk genes, the plasmid backbone, and one copy of the duplicated DNA. The clones thus generated are expected to contain either

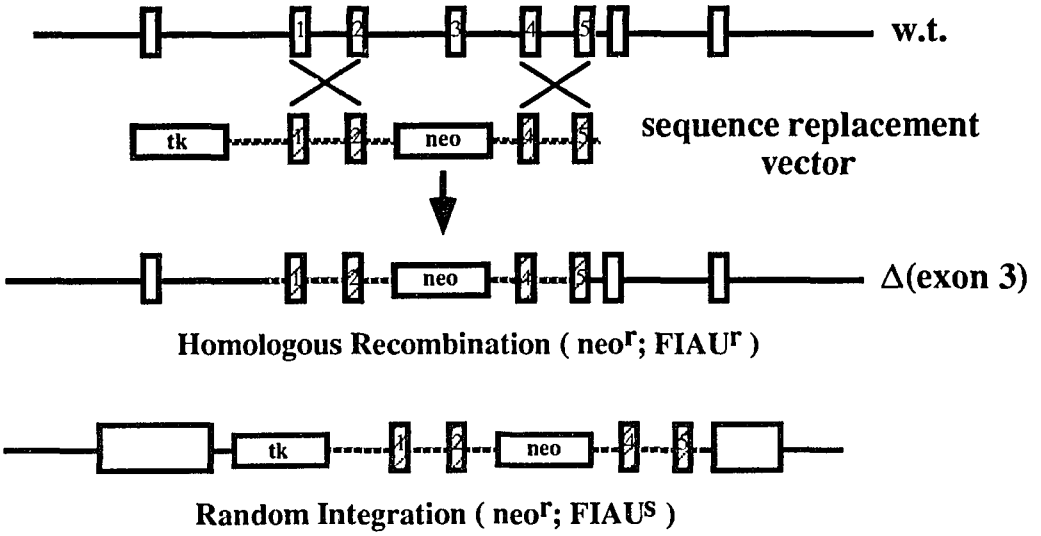


Figure 4: Homologous recombination by the *sequence replacement* vector. Homologous pairing between vector (hatched lines and boxes) and genomic (solid lines and white boxes) sequences initiates a recombination event replacing the genomic sequences with vector sequences including the neo^r gene. Homologous recombination is discern from random integration events by the loss of the tk gene.

the desired mutation or the wild type sequences. By designing a mutation which either destroys or creates a restriction site, these two kinds of clones can be distinguished by Southern blot analysis. Thus, the *hit and run* protocol allows the generation of mice carrying subtle gene modifications.

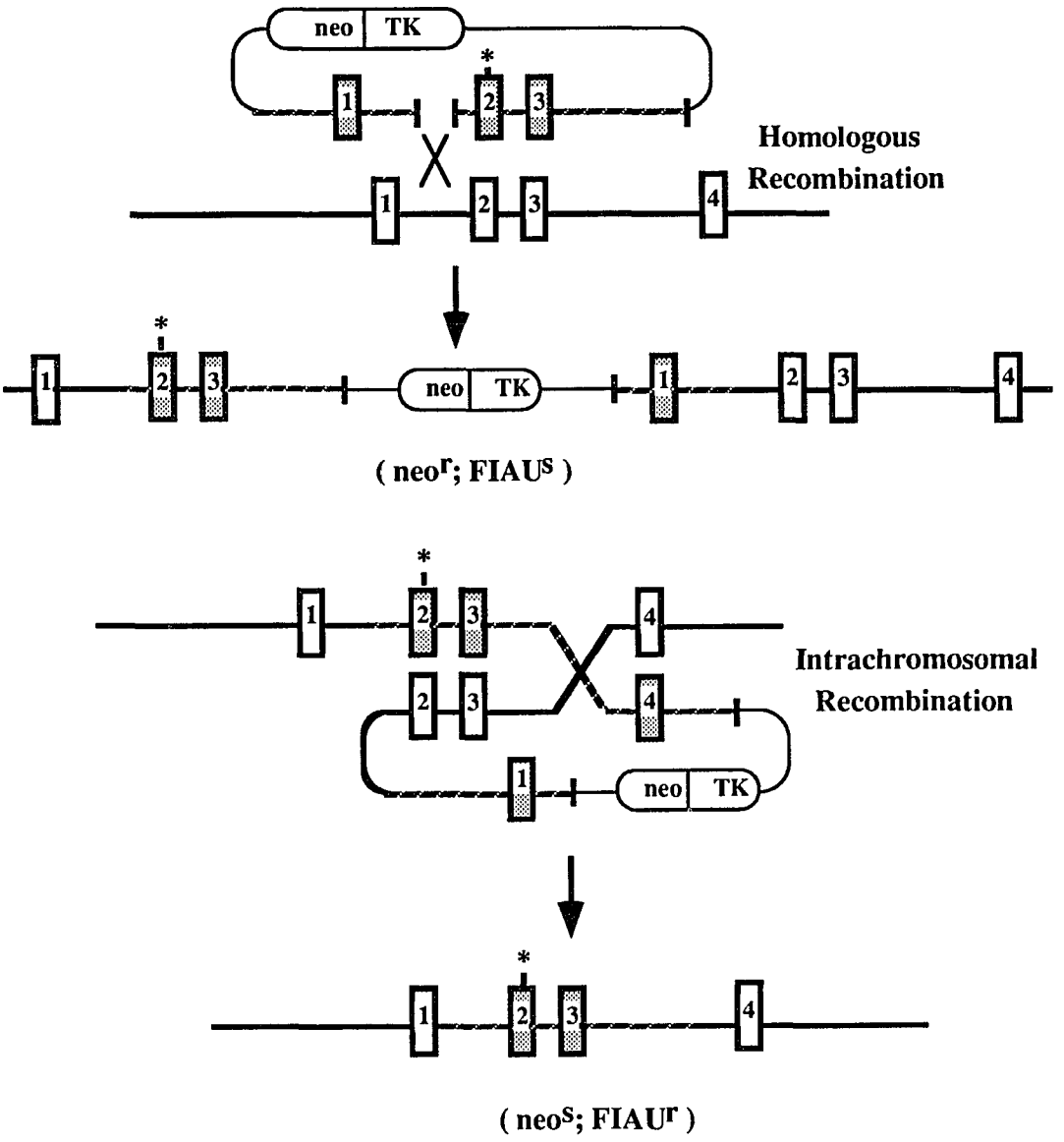


Figure 5: The hit and run procedure. Shaded boxes and striped lines represent vector sequences. The first event consists of a reciprocal recombination between the target gene and the vector (asterisk denotes point mutation). On the second step, an intrachromosomal recombination event resolves the duplication, retaining one copy of the duplicated sequences.

IV - THESIS OUTLINE:

The recent elucidation of the molecular lesion in MFS has raised several issues regarding the role of fibrillin in the biogenesis of microfibrils, and the involvement of the latter in the pathologies associated with the disease. However, a thorough investigation of these processes depends on the availability of complete structural information on the fibrillin gene, as well as of a model system in which to work. In this context, the major theme of this study was to provide these basic tools. Towards this end, the specific aims of this work were the following:

(i) To clone the full-length fibrillin transcript:

The rationale for this aim was to provide some insight in the structure and function of fibrillin by completing the determination of its primary structure. In addition, this information was expected to enable screening of MFS mutations in the full length of the fibrillin message.

(ii) To determine the genomic structure of FBN1:

The characterization of the genomic organization of the FBN1 gene was undertaken for two reasons. First, the determination of the sequences of the intron/exon junctions of FBN1 has the potential to facilitate mutation screening in MFS patients. Second, it can also provide further indication of the evolutionary origin of the newly discovered fibrillin gene family.

(iii) Identification of polymorphic VNTRs in FBN1:

The identification of more informative markers in the FBN1 was undertaken to overcome the difficulties involved in establishing a molecular diagnosis for MFS with the reagents available at the time.

(iv) To develop mutant ES-cell lines for the study of structure/function correlations in fibrillin:

The generation of mice carrying mutations in the fibrillin gene was chosen as an approach to compare the consequences of qualitative and quantitative changes in fibrillin. Specifically, the goal was to investigate the consequences of structural mutations and of gene dosage on microfibril assembly. Toward this end, we created ES-cell lines carrying two different mutant fibrillin alleles.

In summary, we believe that the information and reagents generated by this work will greatly improve the molecular diagnosis and facilitate the mutational screening of MFS. In addition, this study provides the basic tools for a thorough investigation of the biogenesis of microfibrils and their contribution to the MFS phenotype.

MATERIALS AND METHODS

Cloning of the human fibrillin cDNA

Screening of two human placenta cDNA libraries (Clonetech) was performed as described (Sambrook, 1989). Briefly, recombinant phages were plated on six 20-cm plates at a density of 10^4 plaque forming units (pfu) per plate. Duplicate filters were lifted from each plate, denatured for 1 min in 1.5M NaCl, 0.5M NaOH, and neutralized for 6 min in 1.5M NaCl, 0.5M Tris (pH 7.0). After baking at 80°C in a vacuum-oven for 1 hr, the filters were washed with 2XSSC (20XSSC: 3M NaCl, and 0.3 M sodium citrate, at pH 7.0) /0.1% sodium dodecyl sulfate (SDS) at room temperature (r.t.) for 30 min and allowed to air-dry on Whatmann paper. Pre-hybridization and hybridization were performed in 6XSSC, 5X Denhardt's solution (50X Denhardt's solution: 1% Ficoll, 1% polyvinylpyrrolidone, and 1% bovine serum albumin), 10μg/ml salmon sperm-DNA (SS-DNA) and 50% formamide at 65°C for cDNA probes; 6XSSC, 0.1%SDS, 1X Denhardt's solution, 0.005%NaPPi and 50μg/ml SS-DNA at temperatures specific for the calculated melting temperature (T_m) of each oligonucleotide probe. Oligonucleotides were end-labeled with γ-[^{32}P]-ATP (800 Ci/mmol) using 1.0 unit (U) of T4 polynucleotide kinase according to manufacture's protocol (New England Biolabs), while cDNA probes were radiolabeled with α-[^{32}P]-ATP (800 Ci/mmol) using the Random Priming Kit (New England Biolabs). After overnight (o.n.) hybridization, the filters were washed with 2XSSC/0.1%SDS for 15 min/r.t.; 1XSSC/0.1%SDS for 15 min/r.t.; 0.5XSSC/0.1%SDS for 30 min/65°C and a final wash with 0.1XSSC/0.1%SDS for 15 min/65°C. Filters probed with oligonucleotides were washed in 6XSSC/0.005%NaPPi for 30 min/r.t., and then in the same solution for another 30 min, at 5°C below the oligonucleotide's T_m.

PCR amplification of the same cDNA libraries was performed prior to screening in order to determine whether particular fibrillin sequences were present. Accordingly, 1μl

aliquots of human cDNA libraries (from placenta, fibroblasts and an osteosarcoma cell line) containing on the average 10^7 pfu were diluted in 5µl of water, heated at 95°C/5 min, chilled on ice and spun for 10 min/12,000g. The supernatants were used as templates for two different PCR amplifications, both containing a gene-specific primer; additionally each reaction also contained a primer specific for one of the two phage arms. The PCR products were separated on agarose gels, Southern blotted and hybridized to internal oligonucleotides. Positive PCR products were subcloned into pCR1000 (TA-cloning Kit, Invitrogene) according to the manufacturer's instructions. Colony hybridization was performed as described (Sambrook, 1989) using the same probe as for the Southern analysis. Positive clones were sequenced using the Sequenase Sequencing Kit (US Biochemicals) according to manufacturer's protocol.

The so-called "rapid amplification of cDNA ends" (RACE) protocol (Frohman, 1988) was used for the cloning of the foremost 5' of the fibrillin message. First-strand cDNA synthesis from the osteosarcoma cell line MG-63 mRNA was primed with oligonucleotide mf-122, at position +126 to +145 (because of the uncertainty regarding the exact point of the start-site of transcription, the numbering of the FBN1 message (+1) begins at the A residue of the putative initiator ATG). The extended product was tailed with dATP by terminal transferase, and amplified by PCR with the forward primer mf-123 (+32 to +51) and an oligo-dT reverse primer. PCR products were cloned into pCR1000 vector (TA-cloning Kit, Invitrogene) and colonies were screened with mf-134, an oligonucleotide corresponding to a sequence immediately upstream of mf-123. Positive clones were sequenced as described (Sequenase Sequencing Kit, US Biochemicals).

PCR amplification of the fibrillin gene and cDNA

PCR was performed in a final volume of 100µl containing 1.25mM of each dNTP, 1 X reaction buffer (Promega), 50 ng of each sense and anti-sense primers and 2.5 U Taq Polymerase (Promega) for 35 cycles of 94°C/1min30sec, 50-55°C/2min30sec and

72°C/2min Alternatively, 200ng of each primer per reaction were used for 30 cycles of 94°C/1min, 50-55°C/1min, and 72°C/2min PCR amplification of mouse genomic DNA was performed as described above, using primers mf-70 (+3090 to +3109) and mf-19 (+3380 to +3399).

For the VNTRs amplification, PCR was performed in a final volume of 25µl containing 1X reaction buffer (Cetus), 1.5mM $MgCl_2$, 400µM of each primer, 400µM of each dNTP, and 1.5 U Taq Polymerase (Cetus). Reactions went through denaturation at 94°C for 8min followed by 30 cycles of 94°C/30 sec, 58°C/30 sec and 72°C/30 sec

Primer Extension

Approximately 50ng of the 20-mer mf-122 were end-labeled with 100µCi of γ-$[P]^{32}$-ATP by incubation in 1X kinase buffer at 37°C/1hr with 1U of T4-polynucleotide kinase (New England Biolabs). The labeled oligonucleotide was precipitated with 0.3 M sodium acetate/2 volumes of ethanol, and resuspended in 50µl of formamide. Isotope incorporation was measured in a scintillation counter. Annealing of primer and RNA was performed in a final volume of 30µl containing 1X annealing buffer (40mM PIPES, pH 6.4; 1mM EDTA, pH 8.0; 0.4M NaCl), 1µg of poly-A^+ RNA (from either MG-63 or HT-1080 cell lines), $1x10^6$ cpm of the labeled primer, and 80% formamide. After denaturation at 85°C/3min, samples were incubated at 30°C/o.n., diluted to 300µl with diethyl pyrocarbonate (DEPC)-treated water, and precipitated as described above. Reverse transcriptase reaction was performed in a final volume of 20µl. This contained the annealed primer-RNA, 2.5 U of RNasin, 5mM $MgCl_2$, 1mM each dNTP, 1x reaction buffer (10mM Tris-HCl, pH 8.8; 50mM KCl; 0.1% Triton X-100), and 15U of reverse transcriptase (Promega), at 42°C/1hr. The reactions' volume was then increased to 300µl, extracted with phenol/chloroform, chloroform, ethanol precipitated and resuspended in 5µl of 1x loading buffer (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05%

xylene cyanol). Samples were heated at 94°C/3min prior to loading on a 5% polyacrylamide gel (60W/2hours). Products were visualized by autoradiography.

RNase protection

A phage clone containing exon 1, HG-3, was used to generate the riboprobes for the RNase protection experiments. Briefly, a 354-bp fragment extending from position -14 to -358 was subcloned into the pT7/T3α–19 vector (Gibco-BRL), and sequenced to determine its orientation. In order to generate riboprobes from this fragment in the sense and anti-sense orientations, 10μg of the plasmid were linearized with HindIII and EcoRI, respectively. The plasmid DNA was extracted with phenol/chloroform, chloroform, ethanol precipitated and resuspended in 10μl of DEPC-treated water. Riboprobe reactions were performed in 1X reaction buffer (MAXIscript, Ambion), 10mM DTT, 2.5U of RNasin, 0.5mM of ATP, GTP and UTP, 50μM CTP, 50μCi $[P]^{32}$-CTP, 1μg of template plasmid, and 10U of either T3 or T7 RNA polymerase in a final volume of 20μl. After incubation for 1 hr at 37°C, 2.5 U RNasin and 1 U of DNase were added to each reaction followed by another 15 min incubation at 37°C. The volume of each reaction was then increased to 100μl with DEPC-water and reactions were extracted with phenol/chloroform and chloroform. Half volume of 7.5 M ammonium acetate and 2.5 volumes of ethanol were added to the reactions. After 30 min in dry ice, tubes were centrifuged at 12,000g/20min at 4°C. Pellets were resuspended in 100μl of 1M ammonium acetate and reprecipitated with 300μl of ethanol. Pellets were resuspended in 50μl of 100% formamide. Counts were measured from a 1μl aliquot of each reaction.

Riboprobe-RNA hybridization was performed in a final volume of 30μl containing 1X annealing buffer (as described above), 10μg of total RNA, $5x10^5$ cpm of the riboprobe and 80% formamide. After denaturation at 85°C/5min, samples were incubated at 45°C/o.n. The next day, reactions were cooled down to r.t., digested with RNase in a 180μl nuclease mix [RNase T1 (20μg/ml), RNase A (2μg/μl), and 176μl 1X buffer

(300mM NaCl; 10mM Tris-HCl, pH 7.5; 5mM EDTA)] at r.t/30min The reaction was stopped by adding 5µl of 20% SDS and incubating samples at 37°C/30min with 50µg of Proteinase K. After phenol/chloroform and chloroform extractions, samples were precipitated with 2 volumes of ethanol and resuspended in 6µl of loading buffer (as above). Samples were denatured and separated through polyacrylamide gel as described above for the primer extension.

Analysis of the genomic structure of FBN1

Approximately 0.5 µg of human genomic DNA was PCR amplified using sense primers at the predicted 3'-end of one exon, and anti-sense primers at the predicted 5'-end of the following exon. PCR products were fractionated through 0.8% agarose gels and visualized by ethidium bromide staining. Products were cloned into pCR1000 (TA-cloning Kit, Invitrogene) and sequenced as described above.

Regions that could not be PCR amplified were isolated from a human genomic library (Clonetech) and from a phage genomic library derived from DNA of YAC-15. This is a yeast artificial chromosome clone containing a human DNA insert of 230kb that includes the 110kb long FBN1 gene. This clone was isolated and characterized by Dr. Bryan Sykes in collaboration with our group. Different oligonucleotides from the fibrillin mRNA sequence were end-labeled and used to screen these libraries according to the methods described above. Purified genomic clone DNAs were digested with different restriction enzymes, separated on agarose gels, Southern blotted and hybridized with different end-labeled oligonucleotide probes corresponding to the exons not amplified by PCR. Hybridizing bands were subcloned into the plasmid pUC-19 and sequenced using the hybridizing oligonucleotide as primer.

Identification of VNTRs

The PCR clones generated during the characterization of the genomic organization of FBN1, together with the phage genomic clones from the same locus, cover most of this gene. These DNAs were cleaved with the frequent cutter restriction enzymes Sau3A, HaeIII and ApaI, and subjected to Southern blot hybridization with a labeled poly(dA-dC).poly(dT-dG) oligonucleotide (Pharmacia). Pre-hybridization in 0.5M sodium phosphate (pH 7.0), 7% SDS, 1% BSA was performed at 65°C/2 hrs. After hybridization in the same solution at 65°C/o.n., filters were washed with 2XSSC/0.1%SDS for 15 min/r.t.; 1XSSC/0.1%SDS for 15 min/r.t. and a final wash with 0.1XSSC/0.1%SDS for 15 min/55°C. Bands were visualized by autoradiography. Positive fragments were gel purified, cloned into the pUC-19 vector and sequenced using universal primers. Primers for PCR amplification were design from sequence flanking the repetitive sequences. PCR was performed as described above. After fractionation through 4% Nu-Sieve agarose gels (FMC-BioProducts), PCR products were visualized by ethidium bromide staining. The results of this analysis were confirmed and extended by our collaborator Dr. Hal Dietz. Briefly, PCR primers were end-labeled with 100μCi of γ-[P]^{32}-ATP by incubation in 1X kinase buffer at 37°C/1hr with 1U of T4-polynucleotide kinase (New England Biolabs). The labeled oligonucleotides were precipitated with 0.3 M sodium acetate/2 volumes of ethanol, and resuspended in water. PCR products were radioactively labeled by using one end-labeled PCR primer in each set. The products were fractionated through denaturing 8% polyacrylamide gels at 60 watts, and visualized by autoradiography.

Targeting vectors for homologous-recombination

Overlapping cDNA clones spanning the full-length human FBN1 sequence were used to screen a murine NIH-3T3 cDNA library as described above. Positive clones were sequenced as described and confirmed to represent the mouse counterpart of FBN1, *Fbn-1*, by sequence homology. These experiments primarily performed by our collaborator Dr.

Bonadio at the University of Michigan. In order to isolate genomic clones of *Fbn1* from the same mouse strain as the ES-cells used for the homologous recombination experiments, a 129/Sv-mouse genomic library (a gift from Konstantinos Andrikopoulos) was screened using the murine cDNA probes as described above. Genomic clones were mapped by restriction digests and oligonucleotide hybridization.

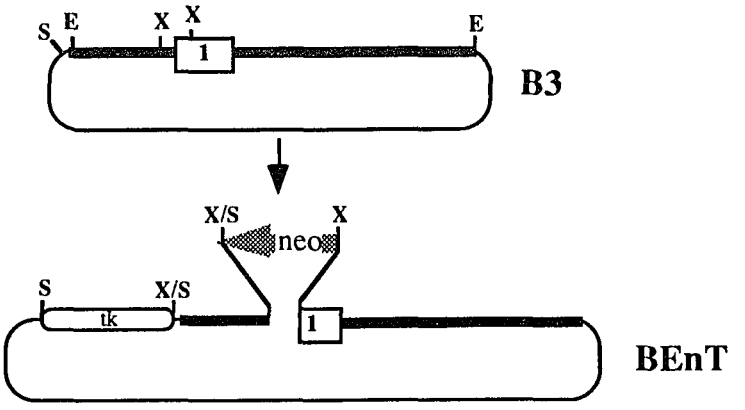


Figure 6: Construction of the targeting vector BEnT (see text below). Restriction sites S (SalI), E (EcoRI), and X (XhoI) are shown. Arrow represents neo-cassete clone in the reverse orientation.

A genomic clone containing exon 1 (ΦB3) was isolated and characterized in detail in order to the generate the null-allele targeting vector BEnT. A 9.5kb EcoRI fragment of ΦB3, including exon 1, was first subcloned into pUC-19 (B3E). A 0.7-kb XhoI fragment of B3E (from nt -650 to +50) was then replaced with a PGK-neo-poly(A) cassette. This cassette contains the neomycin resistance gene (neo^r) with a phosphoglycerate kinase (PGK-1) derived-polyadenylation signal under the control of the PGK-1 promoter (a gift from Dr. R. Jeanisch). The BEnT vector contains 6 kb and 3.5 kb of sequence homology 5' and 3' of the neo-cassette, respectively. As a marker for homologous recombination, a PGK-tk-poly(A) cassette was cloned in the polylinker of pTα19 (fig. 6). In order to generate a probe for the detection of homologous recombination events, a 1.3 kb EcoRI/SalI fragment of ΦB3 (B3ES) lying 3' of B3E was subcloned into pUC-19. In

Southern blot analysis of 129/Sv genomic DNA digested with BamHI, B3ES detects a 12 kb fragment corresponding to the *Fbn1* wild type allele. Introduction of another BamHI site from the neo^r gene was expected to yield a 9 kb band corresponding to the homologously recombined allele.

For the exon-24 deletion construct (mgΔ[tgf2]), a genomic phage clone containing exons 24 through 28 was isolated and characterized (Φmg). The targeting vector mgΔ[tgf2] was constructed by subcloning a 3.0 kb EcoRI fragment of Φmg containing exons 25, 26, 27 and 28 3' of the neo-cassete gene in pGEM7. A 7 kb HindIII fragment of Φmg 5' to and not inclusive of exon 24 was cloned 5' to the neo-cassete . As for B3EnT, the tk-cassette was cloned at the polylinker of pGEM7 (fig. 7). The probe for the Southern analysis of recombinant clones was generated from a 1.5 kb SalI/HindII fragment (mgHS) from the 5' end of Φmg. This probe detects a fragment larger than 15 kb in Southern blot analysis of BamHI digested 129/Sv genomic DNA. Like the first construct, introduction of another BamHI site from the neo^r gene was expected to yield a 10 kb band specific for the mutated allele.

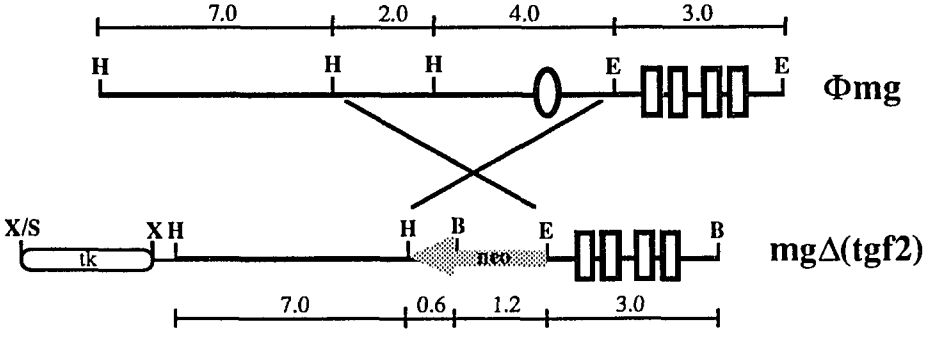


Figure 7: Targeting vector mgΔ(tgf2) (see text above). Restriction sites are as follows: H (HindIII), E (EcoRI), B (BamHI), S (SalI), and X (XhoI). Arrow represents neo-cassete cloned at the reverse orientation.

Culture and electroporation of embryonic stem-cells

J1 ES-cells (generously provided by Dr. R. Jaenisch) were grown in ES-media [DMEM supplemented with 15% heat inactivated fetal calf serum (FCS), 0.1mM non-essential amino acids, 0.1 mM β-mercaptoethanol, and 5U of penicillin and streptomycin] on a layer of γ-irradiated embryonic fibroblasts (EF). Plates for ES-cell culture were coated with 0.2% gelatin (in water) before use by covering plates with the solution for at least 5min at r.t.

ES-cells were grown to 70% confluence in a T75 flask. Cells were trypsinized, counted, and resuspended in electroporation buffer [20mM HEPES (pH 7.0), 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO4, 6mM glucose and 0.1 mM β–mercaptoethanol] at a density of $2x10^7$ cells/ml. Linearized DNA (BEnT with SalI, and mgΔ[tgf2] with XhoI) was added to 0.8 ml of the cell suspension at 50 μg/ml. Electroporation was performed using a Gene Pulser (Bio-Rad), and by subjecting cells to one pulse at 400V and 25μF. After 10min incubation at r.t., electroporated cells were plated on ten 10-cm dishes covered with γ-irradiated EF cells (approximately $2x10^6$ cells/plate). The following day, cells were selected in ES-media containing 350μg/ml of a neomycin analog, G418 (GIBCO-BRL), and 0.2μM of the gancyclovir analog FIAU [1-(2-deoxy-2-fluoro-β-D-arabinofuranosil)-5-iodouracil] (Bristol Myers). One plate was selected with G418 alone in order to access the enrichment for homologous recombination by the tk-gene. Colonies were picked after 7-10 days of double selection.

Colony-picking and processing

After 7-10 days in selection media, colonies were picked with a yellow tip and dissociated in 20μl of trypsin/EDTA. After dissociation, 75μl of ES-media containing selection drugs was added to each colony. Each cell suspension was divided in the following manner: 35 μl were placed in wells of a 96-well plate previously seeded with γ-irradiated EF cells. These were grown for 2-4 days and then frozen at -70°C in ES-media

containing 10% DMSO. The remaining was transferred to 24-well plates, cultured until almost confluent, and used for DNA preparation.

DNA from colonies was prepared according to Laird et. al. (1991). Briefly, after cells reached confluency in the 24-well plates, the medium was replaced by 0.5 ml of lysis buffer (100mM Tris-HCl, pH 8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl and 100µg/ml Proteinase K). Plates were kept in cell-culture incubator at 37°C/o.n. The next day, plates were agitated for 15 min at r.t. This was followed by the addition of 0.5 ml of isopropanol to each well and continued agitation of plates until complete DNA precipitation. DNA precipitates were transferred into Eppendorf tubes, dried and resuspended in 50µl of TE buffer (10mM Tris-HCl, 0.1 mM EDTA, pH 7.5) at 55°C for several hours with gentle agitation.

Genomic Southern blot analysis

Approximately 5-10 µg of total genomic DNA was digested with BamHI for at least 5 hrs. at 37°C. Samples were loaded on a 0.7% agarose gel and run at either 100V/4 hrs. or 25V/o.n. The gel was treated with 0.1M HCl for 10min, denatured with 1.5M NaCl/0.5M NaOH for 1 hr, neutralized with 1.5M NaCl, 0.5M Tris (pH 7.0) for 1 hr and transferred in 10XSSC to nylon membrane (N^+-Hybond, Amersham) o.n. After transfer, DNA was UV-crosslinked to the membrane with the UV Stratalinker (Stratagene). The membrane was washed with 2XSSC/0.1%SDS for 30min/r.t. and air-dried on Whatman paper. Pre-hybridization was performed at 42°C for at least 1 hr in 5XSSC, 10X Denhardt's solution, 0.05 sodium phosphate (pH 6.5), 100µg/ml SS-DNA, 1%SDS, 5% dextran sulfate and 50% formamide. The same solution was used for hybridization overnight at 42°C with random prime-labeled cDNA probes. Filters were washed with 2XSSC/0.1%SDS for 15min/r.t.; 1XSSC/0.1%SDS for 15min/r.t.; 0.5XSSC/0.1%SDS for 30min/65°C and a final wash with 0.1XSSC/0.1%SDS for 15min/65°C. Bands were visualized by autoradiography.

Blastocysts injection of ES-cells

Blastocysts were flushed from the uterus of C57BL/6 mice at day three post coitus with injection media (DMEM supplemented with 10% FCS, 10mM HEPES buffer, and 5U of penicillin and streptomycin), and kept in media covered by mineral oil in a $37^{o}C/5\%CO_2$ incubator for a few hours before injection. ES-cells were plated on a EF layer two days before the injection day. Cells were trypsinized, replated in a gelatin coated dish, and left in the $37^{o}C/5\%CO_2$ incubator for 1 hr. During this incubation, the EF-cells attach more tightly to the dish then healthy ES-cells. Right before injection, the media was aspirated and ES-cells are detached and collected by gentle blowing of injection media. A few microliters of the ES-cell suspension was placed in a drop of injection media containing the blastocysts. Injection of ES-cells into the blastocysts was performed at $4^{o}C$, and approximately 10-15 ES-cells were injected into each blastocyst. An average of 20 injected blastocysts were transferred into the uteri of pseudopregnant recipients (Swiss) at day two post coitus. The surgical procedures involved in this section have been thoroughly described elsewhere (Mann, 1993).

In vitro differentiation of ES-cells

Cells grown in ES-media until 80% confluent were trypsinized and replated in the absence of EF-cells in ES-media with only 10% FCS. Two days later, cells were trypsinized again so that clumps of the order of 10^2 cells stayed intact, and plated in a bacterial dish. Media was changed every day for four days by letting clumps settle to the bottom of 15 ml conical tubes and adding new media. During this time clumps compacted and developed into the so-called embryonic bodies. These were transfered to gelatinized tissue culture plates. Two days later, fibroblasts and epithelial cells could be detected extending out of the embryonic bodies.

RESULTS

The results of this work are divided into two parts. The first one focuses on the human fibrillin gene - the complete FBN1 coding sequence, its genomic organization, and novel intragenic polymorphic markers are reported in this part. Results of this work constitute the topics of one published report and a manuscript in press (Pereira et al., 1993; Pereira et al., 1994). The second part describes the work performed in the mouse, starting with the isolation of the murine fibrillin gene and concluding with the gene targeting experiments. Portions of this work have been published, are about to be submitted for publication, or are yet to be completed (Li, et al. 1994; Yin et al., 1994).

Human fibrillin cDNA cloning

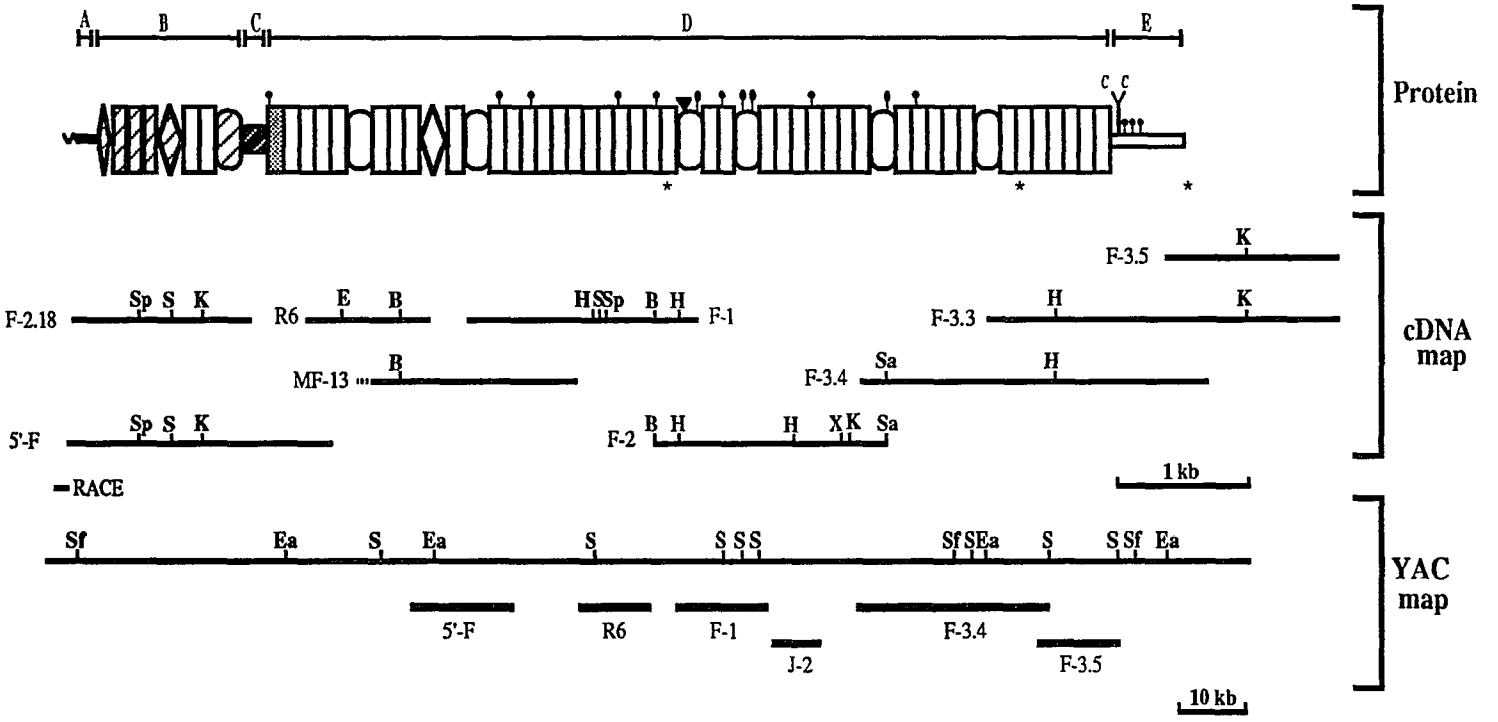
Northern blot analysis predicted that the size of the fibrillin mRNA is approximately 10-kb (Lee et al., 19991). Previous work characterized several overlapping cDNAs covering 7,142-nt of the FBN1 transcript (Lee et al., 1991; Maslen et al., 1991). The published sequence included an open reading frame (ORF) of 6226-nt and 916-nt of 3' untranslated sequence (3'UT) (see appendix).

Based on this data, six oligonucleotide probes were synthesized and used to screen a human placental cDNA library. This led to the isolation of five overlapping cDNA clones whose characterization confirmed the sequence reported by Maslen et al (1991) (fig. 8). In order to determine the sequence of the full-length fibrillin transcript, oligonucleotides corresponding to different 5' regions of the original clone MF-13 were synthesized and used to screen two human placental cDNA libraries. This approach however failed to detect novel 5' cDNA clones, suggesting a low representation of this portion of the FBN1 transcript in these libraries. In order to circumvent this problem, an alternative PCR-based approach was undertaken. Briefly, aliquots of the library were PCR amplified with primers specific for each one of the phage arms and with another primer (mf-42)

corresponding to the 5' region of clone MF-13. In this way, clones containing additional sequence 5' of mf-42 could potentially be amplified, regardless of their orientation within the phage vector. In addition, this method permitted rapid screening of a 100 fold greater number of independent recombinant phages than by the traditional library screening method.

PCR products were Southern hybridized to an oligonucleotide 5' of mf-42. One positive band was subcloned and sequenced. This product overlaps with part of MF-13, confirming that it is part of the FBN1 transcript. In order to avoid sequence errors introduced by the Taq DNA polymerase, the PCR clone was used to re-screen the same human placental cDNA libraries. The resulting positive clone (R6) was found to overlap with MF-13, extending the sequence of the fibrillin message 570-bp upstream of that clone. Sequence comparison of these two cDNAs identified a discordance in the very 5' sequence of MF-13. Analysis of a clone containing the corresponding genomic region showed identity to the R6 sequence. Thus, these results demonstrated that the first 144-bp of the original cDNA clone MF-13 are a cloning artifact (fig. 8). Clone R6 was then used to isolate cDNAs containing further 5' sequences. The PCR-based method described above generated probes which were subsequently used to screen the two placental cDNA libraries. This led to the isolation of two overlapping cDNA clones (5'-F and F-2.18) (fig. 8). Sequencing of clone 5'-F revealed that it overlaps with R6 for 110-nt, and contains 1.9-kb of additional 5' sequence.

Altogether, the overlapping fibrillin cDNA clones comprise approximately 9.6-kb of the fibrillin message. This value is very close to the predicted 10kb size of the mature fibrillin transcript. Analysis of the deduced amino acid sequence encoded by these cDNAs revealed a stretch of consecutive hydrophobic residues preceded by a methionine at the amino terminus. The ATG codon corresponding to this methionine is positioned in the context of a Kozak consensus sequence (Kozak, 1991), rendering it a potential start site of translation (fig. 9). Consistent with this observation, the stretch of hydrophobic residues is



41 aa $D X \frac{N}{D} E C X_6 C X_4 C X N X_2 G X \frac{Y}{F} X C X C X_2 G X_8 C X$

73 aa $D X R X_3 C \frac{Y}{F} X_8 C X_{12} C C C X_{10} C E X C P X_3 L C P X_{14}$

Fig 8 (previous page): Fibrillin protein, cDNAs and YAC-15. The cDNAs encoding fibrillin are depicted bellow a schematic representation of the protein. Symbols represent the following structural elements: signal peptide (wavy line); EGF-CB repeat (white rectangle); EGF repeat (hatched rectangle); TGFb repeat (white ovals); Fib-motif (white diamond); putative glycosylation sites (black dots); potential cell-attachment signal (black triagle). Hatched symbols indicate structures similar to the ones represented by the corresponding white symbols. Shaded rectangle indicates the EGF-CB motif missing the first three residues of the calcium binding consensus (see text). Letters above indicate the different structural regions of fibrillin. Asterisks indicate the position of the oligonucleotide probes used for the initial screening of cDNA libraries. The dotted line on clone MF-13 represents the cloning artifact found in the 5' end of this cDNA. cDNAs used for Southern blot hybridizations are shown below the YAC-15 map. B, BamHI; E, EcoRI; Ea, EagI; H, HindIII; K, KpnI; S, SmaI; Sa, SacI; Sf, SfiI; Sp, SphI. Below, the consensus sequence of the EGF-CB and TGFb motifs of fibrillin is shown: X indicates any residue different from a cysteine; subscripted numbers indicate average number of residues.

likely to constitute a signal peptide, which is usually present at the amino terminus of secreted proteins (Alberts et al., 1989).

Exhaustive library screening for further 5' clones failed to yield new sequences, and therefore an alternative approach was undertaken. The Rapid Amplification of cDNA Ends (RACE) protocol (Frohman et al., 1988), designed to isolate the 5' end of transcripts, was employed. Utilization of this method identified 38-nt additional 5' sequence. Analysis of the deduced amino acid sequence encoded by the 5' cDNA clones indicated that the ORF of the fibrillin transcript extends upstream of the putative initiator ATG.

Start site of transcription of FBN1

In order to determine the start site of transcription of FBN1, primer extension and RNase protection assays were employed (fig. 9). Two cell lines were utilized for these

studies: MG-63, an osteosarcoma cell line known to express FBN1, and HT-1080, a fibrosarcoma cell line unreactive with anti-fibrillin antibodies (Sakai et al., 1986). Primer mf-122 generated one single extension product of approximately 260-nt from MG-63 RNA, placing the start site of transcription approximately 134 nt upstream from the potential initiator ATG (fig. 9). No product was detected from RNA from HT-1080. To confirm this result, RNase protection assay was performed. A riboprobe extending from position -14 to position -358 from the potential initiator ATG was generated from a FBN1 genomic clone. Consistent with the results from the primer extension experiments, MG-63 RNA was able to protect a riboprobe fragment of approximately 118-nt (fig. 9). The protected fragment was not detected with RNA from HT-1080. Although another larger protected fragment was detected in the MG-63 sample, this product was unespecific, judged by its presence also in the HT-1080 sample. Thus, these two methods concurred in placing the start site of transcription of FBN1 134-nt upstream of the initiator ATG (fig. 9). Interestingly, the genomic sequence upstream of this potential transcription initiation site is extremely GC-rich and does not contain any TATA or CCAAT boxes.

Concurrent with our studies, Corson et al. described another three putative FBN1 exons (A, B and C) which are alternatively spliced 5' of exon 1 (Corson et al., 1993). Northern blot analysis using probes derived from these exonic sequences suggest that A is the most utilized exon among the three, since probes corresponding to exons B and C yielded very weak hybridization signals. The strongest signal, however, was obtained from a probe corresponding to exon 1. Moreover, the intensity of that signal could not be accounted by the sum of the signals from exons A, B and C. These experiments suggested that the most abundant form of the FBN1 transcript starts with exon 1 (or possibly another yet unidentified 5' exon). The other alternatively spliced forms, with exons A, B or C upstream of exon 1, are probably minor transcripts. We believe that they have no functional relevance, since they do not contain any in-frame ATG upstream of the initiating

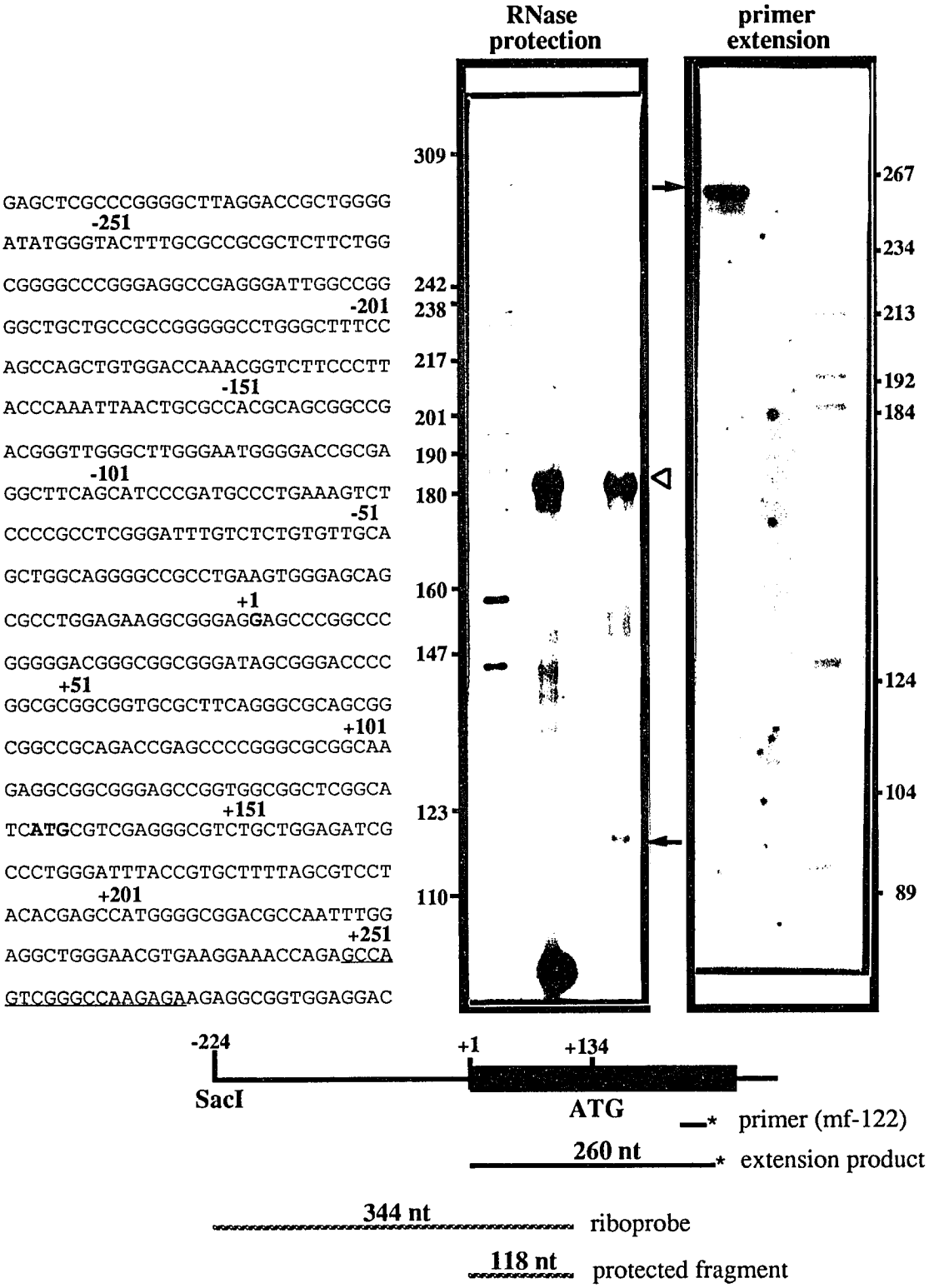

RNase protection
primer extension
GAGCTCGCCCGGGGCTTAGGACCGCTGGGG
-251
ATATGGGTACTTTGCGCCGCGCTCTTCTGG
CGGGGCCCGGGAGGCCGAGGGATTGGCCGG
-201
GGCTGCTGCCGCCGGGGGCCTGGGCTTTCC
AGCCAGCTGTGGACCAAACGGTCTTCCCTT
-151
ACCCAAATTAACTGCGCCACGCAGCGGCCG
ACGGGTTGGGCTTGGGAATGGGGACCGCGA
-101
GGCTTCAGCATCCCGATGCCCTGAAAGTCT
-51
CCCCGCCTCGGGATTTGTCTCTGTGTTGCA
GCTGGCAGGGGCCGCCTGAAGTGGGAGCAG
+1
CGCCTGGAGAAGGCGGGAGGAGCCCGGCCC
GGGGGACGGGCGGCGGGATAGCGGGACCCC
+51
GGCGCGGCGGTGCGCTTCAGGGCGCAGCGG
+101
CGGCCGCAGACCGAGCCCCGGGCGCGGCAA
GAGGCGGCGGGAGCCGGTGGCGGCTCGGCA
+151
TCATGCGTCGAGGGCGTCTGCTGGAGATCG
CCCTGGGATTTACCGTGCTTTTAGCGTCCT
+201
ACACGAGCCATGGGGCGGACGCCAATTTGG
+251
AGGCTGGGAACGTGAAGGAAACCAGAGCCA
GTCGGGCCAAGAGAAGAGGCGGTGGAGGAC
309
242
238
217
201
190
180
160
147
123
110
267
234
213
192
184
124
104
89
-224
+1
+134
SacI
ATG
primer (mf-122)
260 nt
extension product
344 nt
riboprobe
118 nt
protected fragment

Figure 9 (previous page): FBN1 start site of transcription. Nucleotide sequence of the 5' region of FBN1. Putative initiation codon is shown in bold letters, and primer mf-122 sequence is underlined. Numbering starts at the start site of transcription. RNase protection: RNA from cell lines HT-1080 (lane 1), and MG-63 (lane 2). Arrow indicates protected fragment, whereas arrow head shows inespecific product present in both samples. Primer extension: RNA from cell lines MG-63 (lane 1), and HT-1080 (lane 2).

codon in exon 1. The issue of FBN1 start site of transcription will be further discussed later, together with some additional evidence from the murine fibrillin gene.

Primary structure of human fibrillin

Analysis of the deduced primary structure of fibrillin revealed many interesting features of this 2,871 amino acid glycoprotein. Consistent with early protein studies, fibrillin has a predicted molecular mass of 347-kDa (Sakai et al., 1986). According to its primary sequence, fibrillin can be divided into five structurally distinct regions A, B, C, D and E (figs. 8 and 10).

Two of these regions, B and D, are composed of a series of consecutive cysteine-rich modules. Region D is the largest region of fibrillin with its 2,240 amino acids organized in a series of 49 consecutive cysteine-rich repeats (fig. 10). These repeats can be divided into three groups according to structural homologies. The largest group is composed of 41 repeats with close homology to the aforementioned subclass of EGF motifs involved in calcium binding (EGF-CB) (Handford et al., 1990). Alignment of these 41 EGF-CB yielded the following consensus sequence:

$$\mathbf{D}X(\mathbf{N/D})EC{X_{(6)}}C{X_{(4)}}CXN{X_2}G\mathbf{X}(\mathbf{Y/F})XCXC{X_2}G{X_{(9)}}C\mathbf{X}$$

(where numbers in parentheses indicate average number of residues and bold letters denote residues of the calcium binding consensus). One additional EGF-CB-like repeat lacking

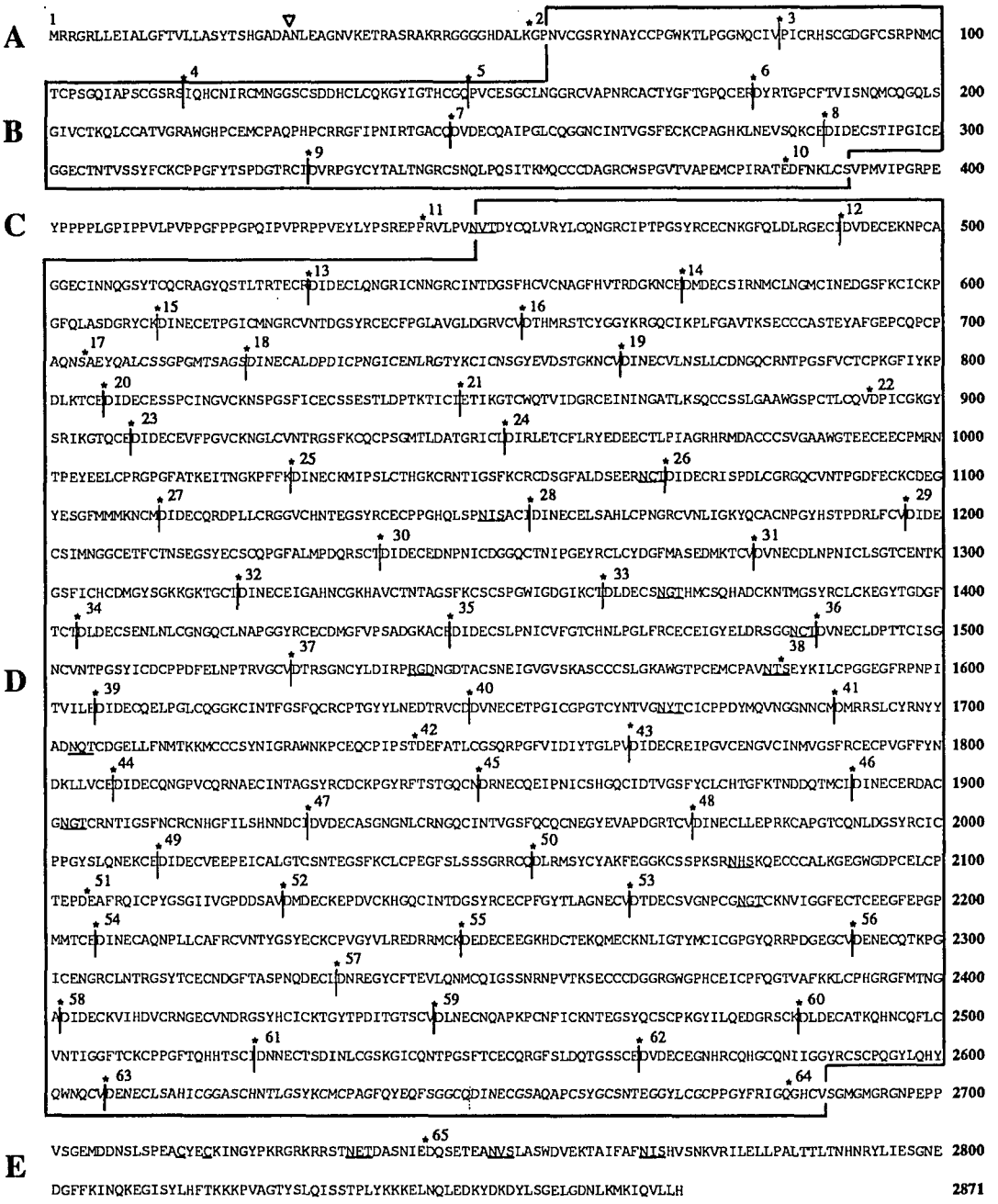

A
MRRGRLLEIALGFTVLLASYTSHGADANLEAGNVKETRASRAKRRGGGGHDALKGPNVCGSRYNAYCCPGWKTLPGGNQCIVPICRHSCGDGFCSRPNMC 100
B
TCPSGQIAPSCGSRSIQHCNIRCMNGGSCSDDHCLCQKGYIGTHCGQPVCESGCLNGGRCVAPNRCACTYGFTGPQCERDYRTGPCFTVISNQMCQGQLS 200
GIVCTKQLCCATVGRAWGHPCEMCPAQPHPCRRGFIPNIRTGACQDVDECQAIPGLCQGGNCINTVGSFECKCPAGHKLNEVSQKCEDIDECSTIPGICE 300
GGECTNTVSSYFCKCPPGFYTSPDGTRCIDVRPGYCYTALTNGRCSNQLPQSITKMQCCCDAGRCWSPGVTVAPEMCPIRATEDFNKLCSVPMVIPGRPE 400
C
YPPPPLGPIPPVLPVPPGFPPGPQIPVPRPPVEYLYPSREPPRVLPVNVTDYCQLVRYLCQNGRCIPTPGSYRCECNKGFQLDLRGECIDVDECEKNPCA 500
GGECINNQGSYTCQCRAGYQSTLTRTECRDIDECLQNGRICNNGRCINTDGSFHCVCNAGFHVTRDGKNCEDMDECSIRNMCLNGMCINEDGSFKCICKP 600
GFQLASDGRYCKDINECETPGICMNGRCVNTDGSYRCECFPGLAVGLDGRVCVDTHMRSTCYGGYKRGQCIKPLFGAVTKSECCCASTEYAFGEPCQPCP 700
AQNSAEYQALCSSGPGMTSAGSDINECALDPDICPNGICENLRGTYKCICNSGYEVDSTGKNCVDINECVLNSLLCDNGQCRNTPGSFVCTCPKGFIYKP 800
DLKTCEDIDECESSPCINGVCKNSPGSFICECSSESTLDPTKTICIETIKGTCWQTVIDGRCEININGATLKSQCCSSLGAAWGSPCTLCQVDPICGKGY 900
SRIKGTQCEDIDECEVFPGVCKNGLCVNTRGSFKCQCPSGMTLDATGRICLDIRLETCFLRYEDEECTLPIAGRHRMDACCCSVGAAWGTEECEECPMRN 1000
TPEYEELCPRGPGFATKEITNGKPFFKDINECKMIPSLCTHGKCRNTIGSFKCRCDSGFALDSEERNCTDIDECRISPDLCGRGQCVNTPGDFECKCDEG 1100
YESGFMMMKNCMDIDECQRDPLLCRGGVCHNTEGSYRCECPPGHQLSPNISACIDINECELSAHLCPNGRCVNLIGKYQCACNPGYHSTPDRLFCVDIDE 1200
CSIMNGGCETFCTNSEGSYECSCQPGFALMPDQRSCTDIDECEDNPNICDGGQCTNIPGEYRCLCYDGFMASEDMKTCVDVNECDLNPNICLSGTCENTK 1300
GSFICHCDMGYSGKKGKTGCIDINECEIGAHNCGKHAVCTNTAGSFKCSCSPGWIGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKEGYTGDGF 1400
TCTDLDECSENLNLCGNGQCLNAPGGYRCECDMGFVPSADGKACHDIDECSLPNICVFGTCHNLPGLFRCECEIGYELDRSGGNCTDVNECLDPTTCISG 1500
NCVNTPGSYICDCPPDFELNPTRVGCVDTRSGNCYLDIRPRGDNGDTACSNEIGVGVSKASCCCSLGKAWGTPCEMCPAVNTSEYKILCPGGEGFRPNPI 1600
TVILEDIDECQELPGLCQGGKCINTFGSFQCRCPTGYYLNEDTRVCDDVNECETPGICGPGTCYNTVGNYTCICPPDYMQVNGGNNCMDMRRSLCYRNYY 1700
ADNQTCDGELLFNMTKKMCCCSYNIGRAWNKPCEQCPIPSTDEFATLCGSQRPGFVIDIYTGLPVDIDECREIPGVCENGVCINMVGSFRCECPVGFFYN 1800
DKLLVCEDIDECQNGPVCQRNAECINTAGSYRCDCKPGYRFTSTGQCNDRNECQEIPNICSHGQCIDTVGSFYCLCHTGFKTNDDQTMCIDINECERDAC 1900
GNGTCRNTIGSFNCRCNHGFILSHNNDCIDVDECASGNGNLCRNGQCINTVGSFQCQCNEGYEVAPDGRTCVDINECLLEPRKCAPGTCQNLDGSYRCIC 2000
PPGYSLQNEKCEDIDECVEEPEICALGTCSNTEGSFKCLCPEGFSLSSSGRRCQDLRMSYCYAKFEGGKCSSPKSRNHSKQECCCALKGEGWGDPCELCP 2100
TEPDEAFRQICPYGSGIIVGPDDSAVDMDECKEPDVCKHGQCINTDGSYRCECPFGYTLAGNECVDTDECSVGNPCGNGTCKNVIGGFECTCEEGFEPGP 2200
MMTCEDINECAQNPLLCAFRCVNTYGSYECKCPVGYVLREDRRMCKDEDECEEGKHDCTEKQMECKNLIGTYMCICGPGYQRRPDGEGCVDENECQTKPG 2300
ICENGRCLNTRGSYTCECNDGFTASPNQDECLDNREGYCFTEVLQNMCQIGSSNRNPVTKSECCCDGGRGWGPHCEICPFQGTVAFKKLCPHGRGFMTNG 2400
ADIDECKVIHDVCRNGECVNDRGSYHCICKTGYTPDITGTSCVDLNECNQAPKPCNFICKNTEGSYQCSCPKGYILQEDGRSCKDLDECATKQHNCQFLC 2500
VNTIGGFTCKCPPGFTQHHTSCIDNNECTSDINLCGSKGICQNTPGSFTCECQRGFSLDQTGSSCEDVDECEGNHRCQHGCQNIIGGYRCSCPQGYLQHY 2600
QWNQCVDENECLSAHICGGASCHNTLGSYKCMCPAGFQYEQFSGGCQDINECGSAQAPCSYGCSNTEGGYLCGCPPGYFRIGQGHCVSGMGMGRGNPEPP 2700
E
VSGEMDDNSLSPEACYECKINGYPKRGRKRRSTNETDASNIEDQSETEANVSLASWDVEKTAIFAFNISHVSNKVRILELLPALTTLTNHNRYLIESGNE 2800
DGFFKINQKEGISYLHFTKKKPVAGTYSLQISSTPLYKKKELNQLEDKYDKDYLSGELGDNLKMKIQVLLH 2871

Figure 10 (previous page): Translation and organization of FBN1 coding sequence. Amino acids are numbered on the right from the start site of translation; a putative signal peptidase cleavage site is indicated by an arrow head; potential glycosylation sites and cell attachment sites are underlined. The five regions of fibrillin are designated by letters on the left. Cysteine rich regions are boxed. Asterisks indicate the exon boudaries (numbered above), whereas vertical lines indicate structural boundaries of the cysteine rich repeats. The dotted line in exon 63 indicates the absence of an expected intro/exon junction.

the two first amino acids of the calcium binding consensus was identified at the most amino terminus of region D (fig. 10).

The second group of cysteine-rich repeats of region D consists of 6 repeats scattered among the EGF-CB repeats (figs. 8 and 10). Motifs in this group show homology to the 8-cysteine motif first described in the transforming growth factor-β1 binding protein (TGFbp) (Kanzaki et al., 1990). The hallmark of the TGFbp repeats is a cluster of three consecutive cysteine residues. Finally, the third kind of cysteine-rich repeat in region D is a novel 8 cysteine-module characterized by two consecutive cysteines. This motif is apparently unique to the fibrillin molecule and thus was named Fib-motif.

The other cysteine-rich region of fibrillin, region B, extends for 336 residues (fig. 10). According to structural similarities, we identified in this region three EGF-like repeats, one 9-cysteine motif similar to the Fib-motif, two EGF-CB motifs and a derivative of the TGFbp motif. As discussed more extensively later, we defined a novel cysteine rich module reminiscent of the Fib-like motif, present at the very amino-terminus of region B.

The three non-cysteine-rich regions of fibrillin (A, C and E) comprise a total of 295 residues, and are structurally distinct from each other. Region C separates the two cysteine-rich regions B and D (fig. 10). It comprises 58 amino acids, nearly half of which (24 out of 58) are prolines. Region A is located at the amino terminus of fibrillin, and includes the putative signal peptide (fig. 10). Region E corresponds to the carboxy

terminus of the protein, and contains two cysteines and three potential N-linked glycosylation sites (fig. 10). Another twelve putative glycosylation sites, along with one potential cell-attachment signal (arginine-glycine-aspartate), can be identified in region D.

Genomic structure and exon organization

In addition to giving some insights on the evolution of the gene, the determination of the organization of FBN1 provided information relevant to the screening of fibrillin mutations in MFS patients. The work was divided into two parts. The first part was conducted by our collaborator Dr. B. Sykes. It involved the isolation of the entire FBN1 gene from a YAC library, and its preliminary characterization using the cDNAs isolated in our laboratory. The second part was performed in our laboratory and used a PCR-based approach to establish the intron/exon organization of the fibrillin gene.

Briefly, Dr. Sykes' group screened a YAC-library by PCR amplification of DNA pools and isolated a YAC clone (YAC-15) of approximately 230-kb. Southern analysis of YAC-15 DNA digested with rare cutter restriction enzymes and probed with the cDNA clones corresponding to different regions of the fibrillin gene indicated that FBN1 is approximately 110-kb long. This implies a ratio of coding to non-coding sequences of 1:10, suggesting a relatively high amount of interspersed intronic sequence in this gene. Expecting that this feature would hamper the determination of the FBN1 gene structure using traditional mapping/sequencing methods, we devised an alternative approach employing PCR amplification.

As previously mentioned, the original fibrillin cloning work identified three FBN1 exons. Each one of them codes for a single EGF-CB repeat (Lee et al., 1991), and begins and ends with complementing split codons. We therefore reasoned that the repeated structure of fibrillin might be reflected in the genomic organization of FBN1, so that each cysteine-rich module would be encoded by a single exon. Accordingly, we designed 59 pairs of PCR primers flanking the ends of the postulated FBN1 exons and used them to

amplify human genomic DNA. Electrophoretic analysis of the PCR reactions identified 47 amplified products which were sequenced and shown to be part of the FBN1 gene (fig. 11). The results confirmed the preliminary evidence for a modular organization of the fibrillin gene, since most of the exons identified were found to code for single cysteine-rich repeats. Six exceptions to this rule were noted. Four of the six TFGb repeats are encoded by two exons (exons number 16-17, 37-38, 41-42, and 50-51); two exons encode the Fib motif (exons number 21-22); and one exon encodes the last two EGF-CB repeats of region D (exon number 63) (fig. 10 and fig. 12).

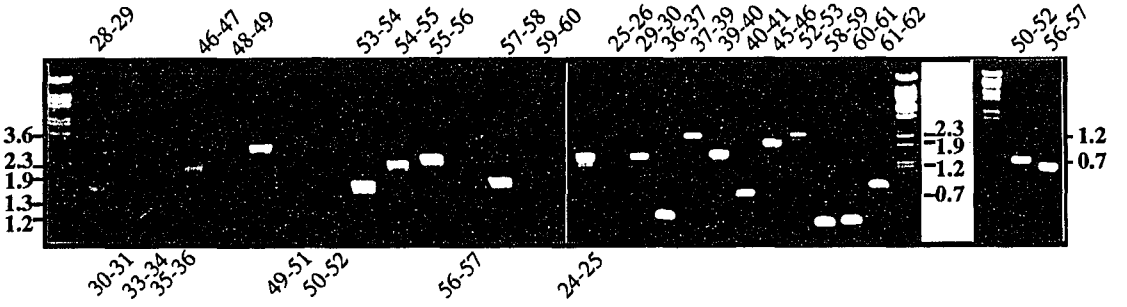


Figure 11: Agarose gels with products from PCR amplification of genomic DNA. Numbers indicate exons flanking the region amplified. Products confirmed to be from FBN1 by sequencing are numbered above. Size marker is λ/BstEI. Bands sizes in kb are indicated on the side of each gel.

The size of the DNA fragment to be amplified is one of the known limiting factors of the PCR approach (Saiki, 1989). This problem becomes more pronounced when one uses genomic DNA as opposed to cloned DNA as the substrate for amplification. To circumvent this problem and complete the characterization of FBN1, we isolated several phage clones using the primers that had failed to yield PCR products. Nine phage clones (HG-3, 5'-HG1, HG-1, HG-9, HG-4, λ4, λ3 and HG-6) were isolated from a placenta genomic library. Clones containing regions that were not detected in that library (HGY-30 and HGY-6.8) were isolated from the YAC-15-derived phage library. As for the genomic

DNA amplification, we could not obtain PCR products from DNA of these phage clones, indicating that the introns flanking exons 1 through 16, 30, 31, 40, 41, 43, 62, 63 and 65 are of substantial length. Thus, the characterization of those exons was performed by more traditional cloning approaches. DNA from the recombinant phages was subcloned into plasmids, and clones hybridizing to primers corresponding to these exons were sequenced.

Altogether, the experiments demonstrated that the FBN1 gene is composed of 65 exons (fig. 12). They also determined the sequence of the boundaries of all these exons, as well as the approximate size of 47 introns (fig. 13). In addition to confirming the original hypothesis of how the gene is structured, the results provided some evidence for how it might have evolved. Both of these points will be discussed more extensively in a later section.

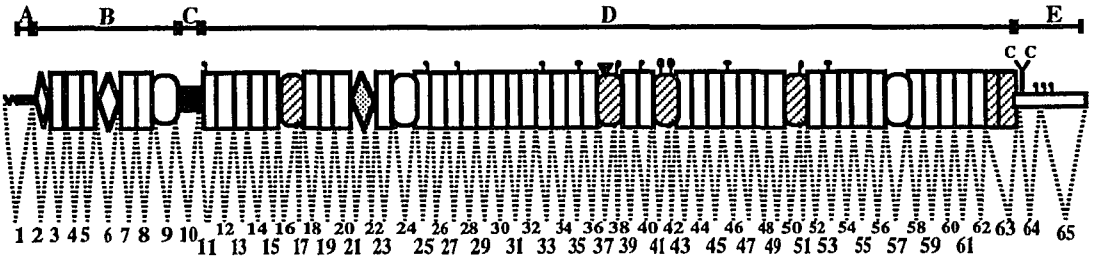
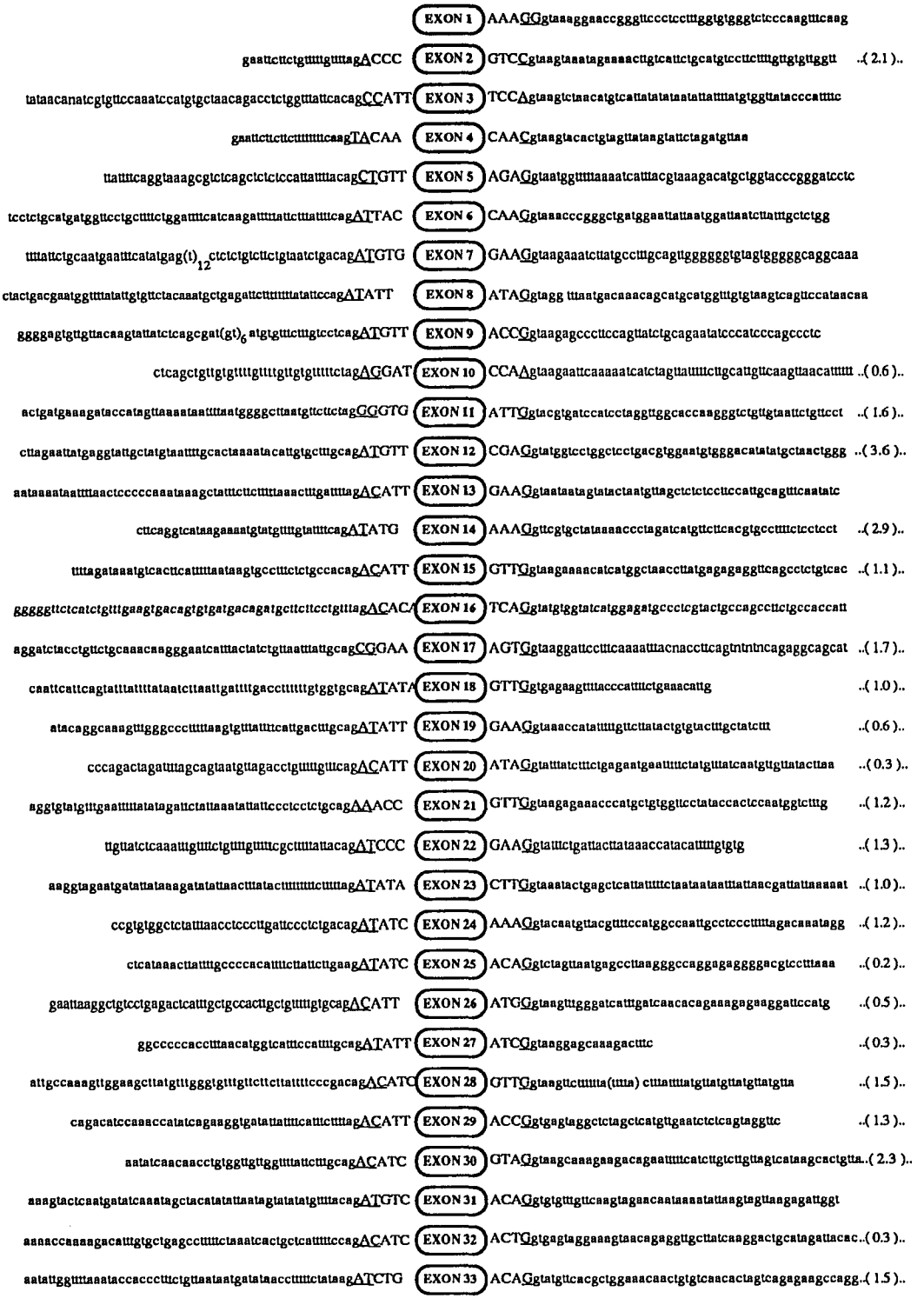


Figure 12: Genomic organization of FBN1. The number of the exons encoding each structure of fibrillin is indicated below. Shaded symbols represent the cysteine-rich motifs which are not encoded by single exons.


EXON 1 AAAGGgtaaaggaaccggggttccctccttggtgtgggtctcccaagttcaag
gaattcttctgttttgttttagACCC EXON 2 GTCCgtaagtaaatagaaaaacttgtcattctgcatgtccttcttttgttgtgttggtt ..(2.1)..
tataacanatcgtgttccaaatccatgtgctaacagacctctggttattcacagCCATT EXON 3 TCCAgtaagtctaacatgtcattatatatataatattatttttatgtggttatacccatttttc
gaattcttcttcttttttttcaagTACAA EXON 4 CAACgtaagtacactgtagttataagtattctagatgttaa
ttatttcaggtaaagcgtctcagctctctccattatttttacagCTGTT EXON 5 AGAGgtaatggttttaaaaatcatttacgtaaagacatgctggtacccgggatcctc
tcctctgcatgatggttcctgcttttctggatttcatcaagatttttattcttatttttcagATTAC EXON 6 CAAGgtaaacccgggctgatggaattattaatggattaatcttattgctctgg
ttttattctgcaatgaatttcatatgag(t)12ctctctgtctttctgtaatctgacagATGTG EXON 7 GAAGgtaagaaatcttatgccttgcagtgggggggtgtagtgggggcaggcaaa
ctactgacgaatggttttatattgtgttctacaaatgctgagattcttttttatatttccagATATT EXON 8 ATAGgtagg ttaaatgacaaacagcatgcatggttgtgtaagtcagttccataacaa
ggggagtgttgttacaagtattatctcagcgat(gt)6atgtgttcttgtcctcagATGTT EXON 9 ACCGgtaagagcccttccagttatctgcagaatatcccatcccagccctc
ctcagctgttgtgttttgttttgttgtgttttctagAGGAT EXON 10 CCAAgtaagaattcaaaaatcatctagttatttttcttgcattgttcaagttaacatttttt ..(0.6)..
actgatgaaagataccatagttaaaaataatttaatggggcttaatgttcttctagGGGTG EXON 11 ATTGgtacgtgatccatcctaggttggcaccaagggtctgttgtaattctgttcct ..(1.6)..
cttagaattatgaggtattgctatgtaatttgcactaaaatacattgtgcttgcagATGTT EXON 12 CGAGgtatggtcctggctcctgacgtggaatgtgggacatatatgctaactggg ..(3.6)..
aataaaataatttaactccccccaaataaagctatttcttcttttaaacttgatttagACATT EXON 13 GAAGgtaataatagtatactaatgttagctctctcctttccattgcagtttcaatatc
cttcaggtcataagaaaatgtatgtttgtatttcagATATG EXON 14 AAAGgttcgtgctatataaaacccctagatcatgttcttcacgtgccttttctcctcct ..(2.9)..
ttttagataaatgtcacttcatttttaataagtgccttttctctgccacagACATT EXON 15 GTTGgtaagaaaacatcatggctaaccttatgagagaggttcagcctctgtcac ..(1.1)..
ggggggttctcatctgtttgaagtgacagtgtgatgacagatgcttcttcctgtttagACACA EXON 16 TCAGgtatgtggtatcatggagatgccctcgtactgccagccttctgccaccatt
aggatctacctgttctgcaaacaagggaatcatttactatctgttaatttattgcagCGGAA EXON 17 AGTGgtaaggattcctttcaaaatttacnaccttcagtntntntncagaggcagcat ..(1.7)..
caattcattcagtatttatttttataatcttaattgatttttgaccttttttgtggtgcagATATA EXON 18 GTTGgtgagaagttttacccatttttctgaaacattg ..(1.0)..
atacaggcaaagttgggccctttttaagtgttatttcattgacttgcagATATT EXON 19 GAAGgtaaaccatatttttgttcttatactgtgtactttgctatctt ..(0.6)..
cccagactagatttagcagtaatgttagacctgttttgttcagACATT EXON 20 ATAGgtatttatctttctgagaatgaatttttctatgtttatcaatgttgttatacttaa ..(0.3)..
aggtgtatgtttgaatttttatatagatttctatttaaatattattccctcctctgcagAAACC EXON 21 GTTGgtaagagaaacccatgctgtggttcctataccactccaatggtcttttg ..(1.2)..
ttgttatctcaaatttgttttctgtttgttttcgctttttattacagATCCC EXON 22 GAAGgtatttctgatttacttataaaccatacatttttgtgtg ..(1.3)..
aaggtagaatgatattataaagatatattaacttttatactttttttctttttagATATA EXON 23 CTTGgtaaatactgagctcattatttttctaatataatttattaacgatttatttaaaaat ..(1.0)..
ccgtgtggctctatttaacctcccttgatttcccctctgacagATATC EXON 24 AAAGgtacaatgttacgttttccatggccaattgcctccctttttagacaaatagg ..(1.2)..
ctcataaactttattttgccccacatttctttatttctgaagATATC EXON 25 ACAGgtctagttaatgagccttaagggccaggagaggggacgtcctttaaa ..(0.2)..
gaattaaggctgtcctgagactcatttgctgccactgctgttttgtgcagACATT EXON 26 ATGGgtaagttgggatcatttgatcaacacagaaagagaaggatttccatg ..(0.5)..
ggccccccaccttaacatggtcatttccatttgcagATATT EXON 27 ATCGgtaaggagcaaagacttc ..(0.3)..
attgccaaagttggaagcttatgtttgggtgtttgttcttcttatttttcccgacagACATC EXON 28 GTTGgtaagttcttttta(ttta) cttatttttatgttatgttatgttatgtta ..(1.5)..
cagacatccaaaccatatcagaaggtgatattatttttcatttcttttagACATT EXON 29 ACCGgtgagtaggctctagctcatgttgaatctctcagtaggttc ..(1.3)..
aatatcaacaacctgtggttgttggttttattcttgcagACATC EXON 30 GTAGgtaagcaaagaagacagaatttttcatcttgtcttgttagtcataagcactgtta..(2.3)..
aaagtactcaatgatatcaaatagctacatatattaatagtatatatgttttacagATGTC EXON 31 ACAGgtgtgtttgttcaagtagaacaataaaatatttaagtagttaagagattggt
aaaaccaaaagacatttgtgctgagccttttttctaaatcactgctcatttttccagACATC EXON 32 ACTGgtgagtaggaaagtaacagaggttgcttatcaaggactgcatagattacac..(0.3)..
aatattggtttaaatacccacccttttctgttaataatgatatatacctttttctataagATCTG EXON 33 ACAGgtatgttcacgctggaaacaactgtgtcaacactagtcagagaagccagg..(1.5)..

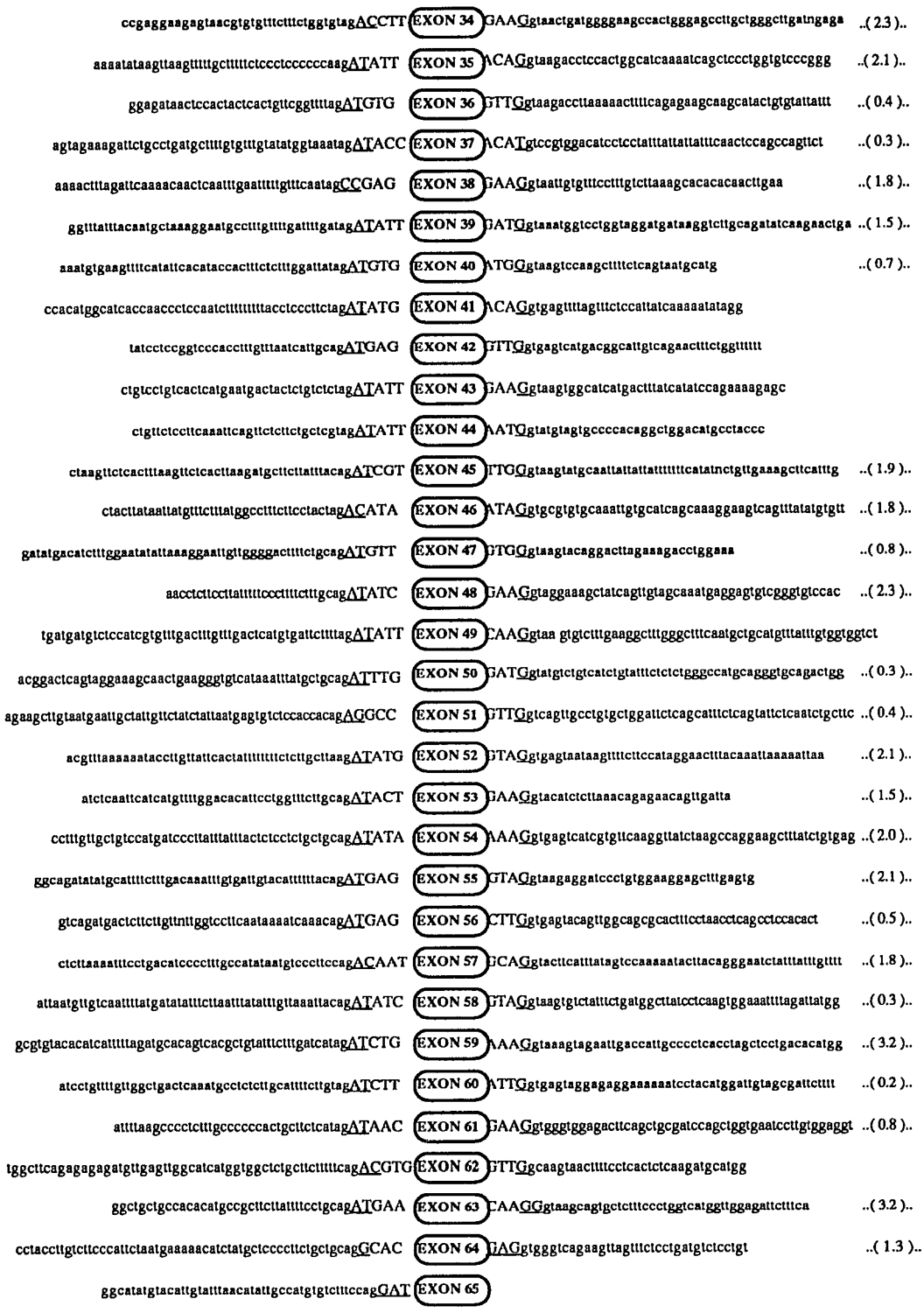

ccgaggaagagtaacgtgtgtttctttctggtgtagACCTT EXON 34 GAAGgtaactgatggggaagccactgggagccttgctgggcttgatngaga ..(2.3)..
aaaatataagttaagttttgctttttctccctccccccaagATATT EXON 35 ACAGgtaagacctccactggcatcaaaatcagctccctggtgtcccggg ..(2.1)..
ggagataactccactactcactgttcggttttagATGTG EXON 36 GTTGgtaagaccttaaaaactttcagagaagcaagcatactgtgtattatt ..(0.4)..
agtagaaagattctgcctgatgctttgtgtttgtatatggtaaatagATACC EXON 37 ACATgtccgtggacatcctcctatttattattatttcaactccagccagttct ..(0.3)..
aaaactttagattcaaaacaactcaatttgaatttttgtttcaatagCCGAG EXON 38 GAAGgtaattgtgtttcctttgtcttaaagcacacacaacttgaa ..(1.8)..
ggtttatttacaatgctaaaggaatgccttgttttgatttgatagATATT EXON 39 GATGgtaaatggtcctggtaggatgataaggtcttgcagatatcaagaactga ..(1.5)..
aaatgtgaagttttcatattcacataccacttttctcttggattatagATGTG EXON 40 ATGGgtaagtccaagcttttctcagtaatgcatg ..(0.7)..
ccacatggcatcaccaaccctccaatctttttttttacctcccttctagATATG EXON 41 ACAGgtgagttttagtttctccattatcaaaaatatagg
tatcctccggtcccaccttgtttaatcattgcagATGAG EXON 42 GTTGgtgagtcatgacggcattgtcagaactttctggttttt
ctgtcctgtcactcatgaatgactactctgtctctagATATT EXON 43 GAAGgtaagtggcatcatgacttttatcatatccagaaaagagc
ctgttctccttcaaattcagttctcttctgctcgtagATATT EXON 44 AATGgtatgtagtgccccacaggctggacatgcctaccc
ctaagttctcactttaagttctcacttaagatgcttcttatttacagATCGT EXON 45 TTGGgtaagtatgcaattattattatttttttcatatnctgttgaaagcttcatttg ..(1.9)..
ctacttataattatgtttcttttatggccttttcttcctactagACATA EXON 46 ATAGgtgcgtgtgcaaattgtgcatcagcaaaggaagtcagtttatatgtgtt ..(1.8)..
gatatgacatctttggaatatattaaaggaattgttggggacttttctgcagATGTT EXON 47 GTGGgtaagtacaggacttagaaagacctggaaa ..(0.8)..
aacctcttccttatttttccctttttcttgcagATATC EXON 48 GAAGgtaggaaagctatcagttgtagcaaatgaggagtgtcgggtgtccac ..(2.3)..
tgatgatgtctccatcgtgtttgacttgtttgactcatgtgattcttttagATATT EXON 49 CAAGgtaa gtgtctttgaaggctttgggctttcaatgctgcatgtttatttgtggtggtct
acggactcagtaggaaagcaactgaagggtgtcataaatttatgctgcagATTTG EXON 50 GATGgtatgtctgtcatctgtatttctctctgggccatgcagggtgcagactgg ..(0.3)..
agaagcttgtaatgaattgctattgttctatctattaatgagtgtctccaccacagAGGCC EXON 51 GTTGgtcagttgcctgtgctggattctcagcatttctcagtattctcaatctgcttc ..(0.4)..
acgtttaaaaaatacccttgttattcactatttttttctcttgcttaagATATG EXON 52 GTAGgtgagtaataagttttcttccataggaactttacaaattaaaaattaa ..(2.1)..
atctcaattcatcatgttttggacacattcctggtttcttgcagATACT EXON 53 GAAGgtacatctcttaaacagagaacagttgatta ..(1.5)..
cctttgttgctgtccatgatcccttatttatttactctcctctgctgcagATATA EXON 54 AAAGgtgagtcatcgtgttcaaggttatctaagccaggaagctttatctgtgag ..(2.0)..
ggcagatatatgcatttttctttgacaaaatttgtgattgtacatttttttacagATGAG EXON 55 GTAGgtaagaggatcccctgtggaaggagctttgagtg ..(2.1)..
gtcagatgactcttcttgttnttggtccttcaataaaatcaaacagATGAG EXON 56 CTTGgtgagtacagttggcagcgcacttttcctaacctcagcctccacact ..(0.5)..
ctcttaaaaattttcctgacatccccttttgccatataatgtcccttccagACAAT EXON 57 GCAGgtacttcatttatagtccaaaaatacttacagggaatctatttatttgtttt ..(1.8)..
attaatgttgtcaatttttatgatatatttcttaatttatatttgttaaattacagATATC EXON 58 GTAGgtaagtgtctatttctgatggcttatcctcaagtggaaatttttagattatgg ..(0.3)..
gcgtgtacacatcatttttagatgcacagtcacgctgtatttctttgatcatagATCTG EXON 59 AAAGgtaaagtagaattgaccattgcccctcacctagctcctgacacatgg ..(3.2)..
atcctgttttgttggctgactcaaatgcctctcttgcatttttcttgtagATCTT EXON 60 ATTGgtgagtaggagaggaaaaaatcctacatggattgtagcgattctttt ..(0.2)..
attttaagccccctctttgccccccactgcttctcatagATAAC EXON 61 GAAGgtgggtggagacttcagctgcgatccagctggtgaatccttgtggaggt ..(0.8)..
tggcttcagagagagatgttgagttggcatcatggtggctctgcttctttttcagACGTG EXON 62 GTTGgcaagtaacttttcctcactctcaagatgcatgg
ggctgctgccacacatgccgcttcttatttttcctgcagATGAA EXON 63 CAAGGgtaagcagtgctcttccctggtcatggttggagattctttca ..(3.2)..
cctaccttgtcttcccattctaatgaaaaacatctatgctccccttctgctgcagGCAC EXON 64 GAGgtgggtcagaagttagtttctcctgatgtctcctgt ..(1.3)..
ggcatatgtacattgtatttaacatattgccatgtgtctttccagGAT EXON 65

Figure 13 (two previous pages): Nucleotide sequence of the intron/exon junctions. Capital letters indicate coding sequence, whereas small letters indicate intronic sequences. Boundary codons are underlined. Numbers in parenthesis indicate the approximate size (in kb) of some introns.

Identification of polymorphic VNTRs in the FBN1 locus

A second aspect of the genomic analysis dealt with the identification of FBN1 polymorphic DNA markers. As already mentioned, these are useful diagnostic tools for the pre-symptomatic and pre-natal diagnosis of familial MFS. Previous work had identified two RFLPs and one VNTR in the FBN1 gene (Clark et al., 1991; Hewett et al., 1991; Lee et al., 1991). These markers however turned out to be of low informative value.

To isolate additional VNTRs, we screened the available collection of FBN1 phage and plasmid clones with a poly(dA-dC).poly(dT-dG) probe (Pharmacia). This analysis detected 3 VNTRs in phage clones HG-3, HG-2 and λ4. Sequence analysis showed that the new VNTRs (mts-1, mts-2 and mts-4) are composed of repeated CA-dinucleotides (fig. 14). This analysis also determined that the previously described VNTR (mts-3) is composed of TAAAA-pentanucleotide repeats (Lee et al., 1991) .

A preliminary study was then performed to estimate the degree of polymorphism of these new markers. PCR primers flanking the repeated sequences were used to amplify genomic DNA from 10 individuals of two distinct MFS families. Electrophoretic analysis of the PCR products detected at least two different alleles for each one of the markers, demonstrating polymorphisms of the latter (fig. 15). The segregation of these markers allowed the determination of the haplotypes for each one of the family members. The markers showed cosegregation of FBN1 and the MFS phenotype, and assigned an unaffected status to the individual III-2 in family B with previously ambiguous diagnosis (fig. 15).

mts-1

gatgaaacaatgatgatgaagaaaaagaagaaaaggaaaagatggcaacaaagaaggagaaacagaaccggaaa

aggaagaagaagaaaccaaaaaaaaaaaaaaaaaac **[CA]$_{21}$** aaatgaaaccctgcctctggaatacattgtcttt

attaataaagatattaccatacttcagttcaaaaccatgaaacttgacagcagacaatgtgtctttagggcaga

mts-2

ctttcagataagtagttgttatcttgcagagtctccaattaatcatggcacacgagtactcttcattaaaggaaaagagaaag

aaggaaaaaaaata **[CA]$_{18}$** tcttagagtcctagagggcagaaataatcttggggagcttttgcaagacgtatatgcc

tgggggggatcctctagagctgacctgcaggcatgcaag

mts-3

aaaacaagaaagataagtaaaagtagcgatgaaaacaaaatctcagagtacatagagtgttttagggagagatgaaataa

aataaaataaaataacataacataacataacataaaataaag **[TAAAA]$_{6}$** aagaacttaccaacacaaaatagccta

tcgggagttgaatggtagccagggttgcaggcacactgatacttccctatgagg

mts-4

ggaaagtggtttagttatctcaatttgctattcgttgatgtccctattgccatcaccaccaccccaacactgcacg **[CA]$_{16}$**

cgcacgcactttccatcttgtcttaccctgcacagggatccgcccgccttagcttcccaaagtgctaggattacaggtgt

gagcaaccg

Figure 14: Nucleotide sequence of the FBN1 VNTRs. PCR primer sequences are underlined. Dotted lines indicate primers used by Dr. Dietz in Pereira et al., 1994 (see appendix).

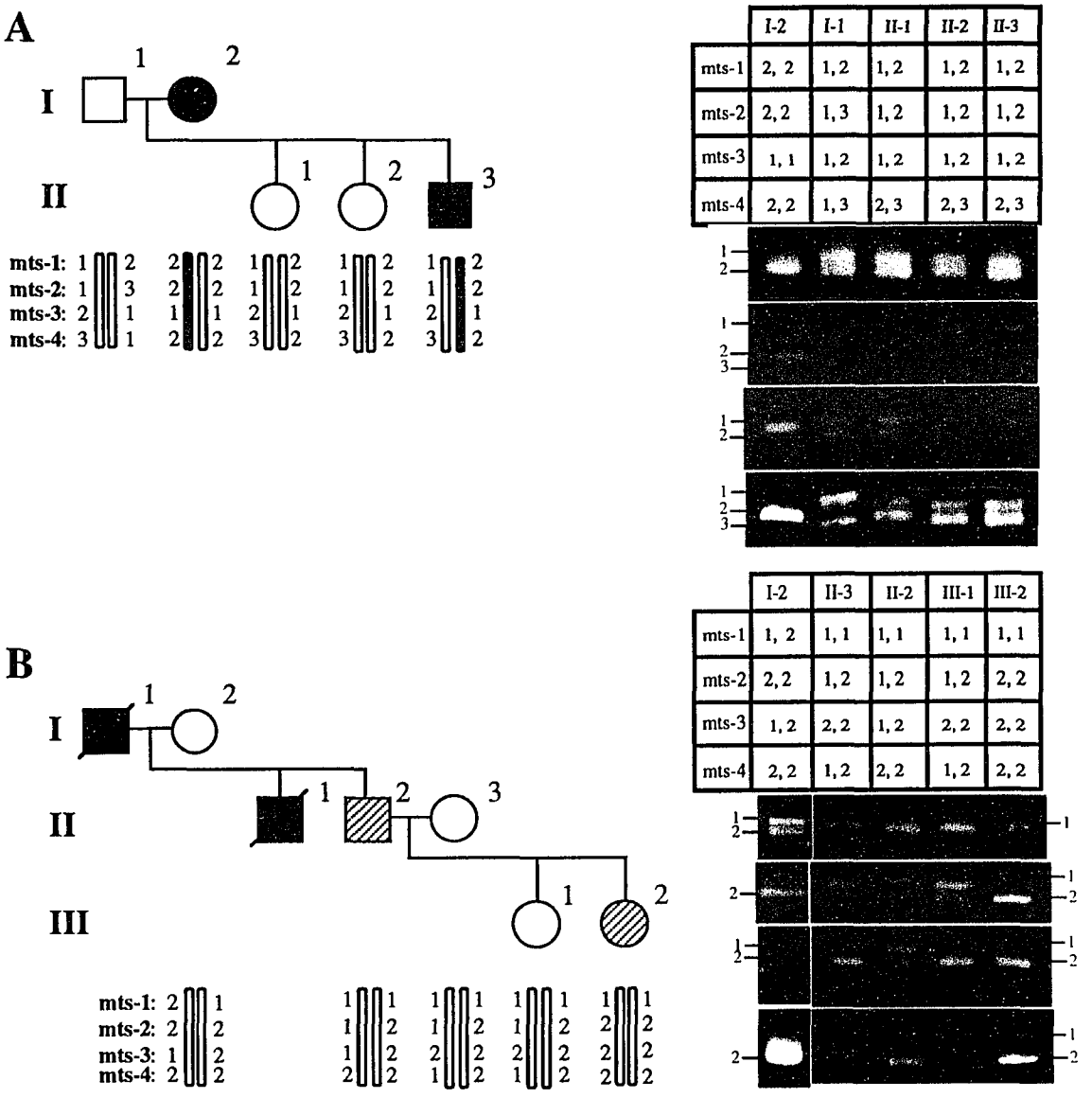


Figure 15: Use of VNTRs for diagnosis of MFS. Black symbols represent affected individuals, whereas gray symbols indicate individuals with an ambiguous diagnosis. On the right, agarose gels with PCR products from amplification of each VNTR. The compilation of the haplotypes is shown on the left.

In order to estimate the number and size of alleles for each one of the VNTRs, a more detailed analysis was performed by Dr. Hal Dietz (see appendix). The heterozygosity of each marker was determined by examination of DNA from 50 unrelated individuals. An

overall haplotype heterozygosity of 86% makes this group of markers highly informative. Finally, the position of each one of the VNTRs within the FBN1 gene was determined (fig. 16). Mapping of the markers within each respective phage clone demonstrated that these polymorphic sequences are spread throughout this locus: mts-1 in intron 1, mts-2 in intron 5, mts-3 in intron 28, and mts-4 in intron 43.

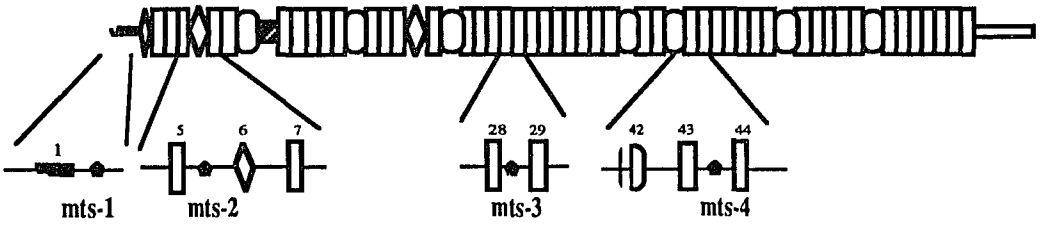


Figure 16: Location of the VNTRs within FBN1. VNTRs are indicated by black diamonds. Exons are represented by the structure they encode, and numbered above.

Murine fibrillin cDNA cloning

The second part of this work focused on the mouse gene and its utilization to study fibrillin pathogenesis. The different aspects of this study were performed with the collaboration of other investigators. Drs. Francke and Bonadio were involved on the characterization of the murine fibrillin gene, whereas Dr. Stuhlman provided invaluable guidance with the experiments involving ES-cells.

In order to initiate the studies of fibrillin in the mouse, we isolated the entire murine fibrillin transcript in collaboration with Dr. Jeff Bonadio's group. These investigators used the human cDNA probes characterized in our laboratory to screen an NIH-3T3 murine cDNA library, and isolated eight overlapping clones covering the entire murine fibrillin message. Sequence comparison between the human and murine fibrillin transcripts showed a high degree of homology at both nucleotide and amino acid levels (fig. 17). At the amino acid level, the two proteins are 90% homologous and 95% identical when one

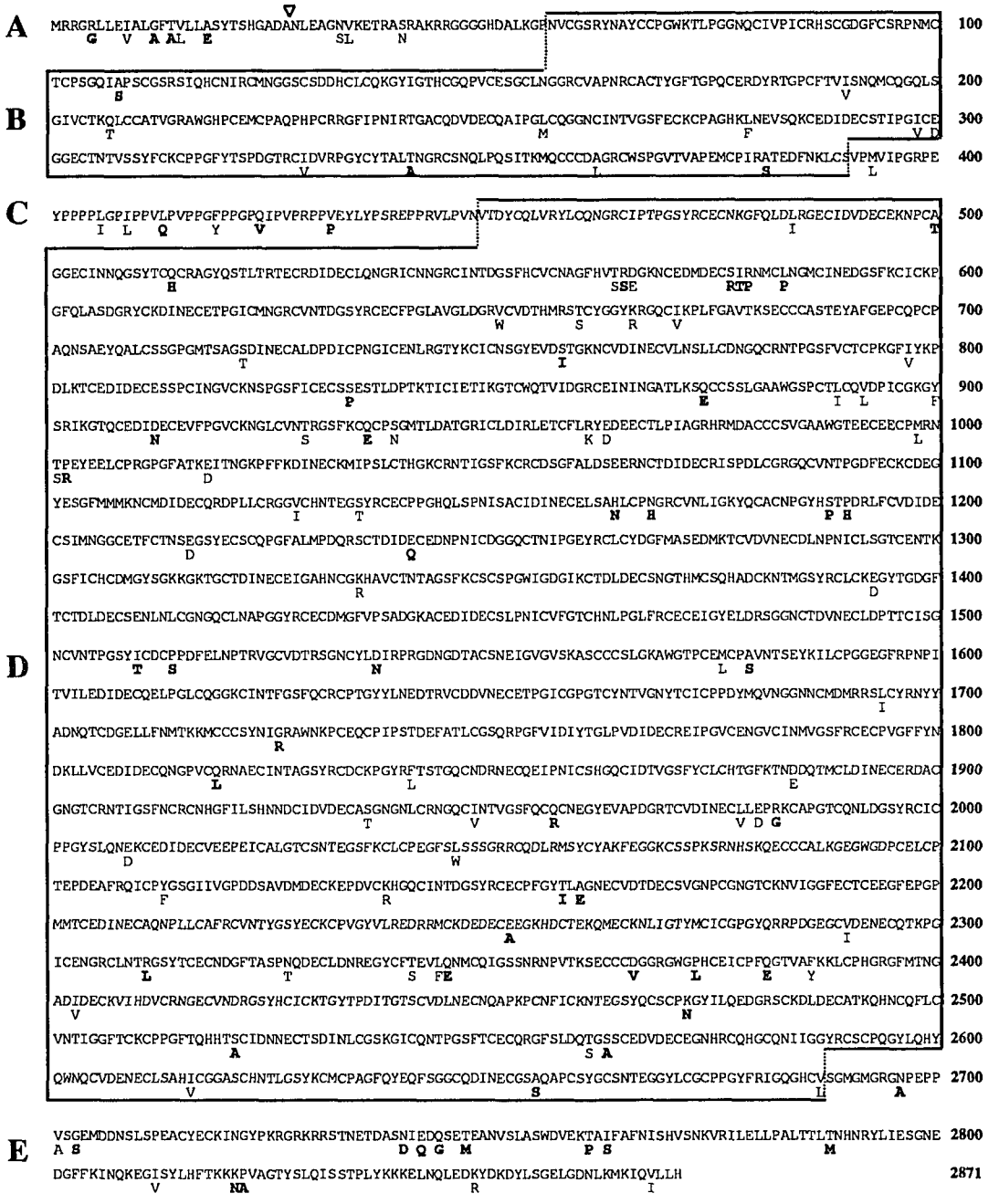

A
B
C
D
E

Figure 17 (previous page): Comparison between human and mouse fibrillin sequences. Letters below indicate mis-matched amino acids. Bolded letters show non-conserved changes.

considers the conservative amino acid changes between the human and murine fibrillins (fig. 17). This comparative analysis also provided evidence in support of the postulated start site of translation of the human transcript. In fact, an in-frame ATG can be found in the 5' region of the murine message at the same relative position as the potential initiator ATG in the human transcript. Upstream of this ATG-codon the homology between the two transcripts drops from 80% to 42%.

At the same time, and in collaboration with Dr. U. Francke, the murine fibrillin gene was mapped to mouse chromosomal region 2F. Assuming that the genomic organization of *Fbn1* would be the same as FBN1, PCR primers were synthesized to amplify sequences between exons 25 and 27 of *Fbn1*. As a result, we obtained a 1.2kb amplified fragment from mouse genomic DNA (fig. 18) which was confirmed to be from *Fbn1* by sequence analysis. The primers were then used for PCR amplification of DNA from a mapping panel consisting of 14 mouse X rodent somatic cell hybrids (Li, et al. 1994). The expected 1.2kb PCR product from *Fbn1* was consistently amplified only from DNA from hybrid cells containing mouse chromosome 2. All other mouse chromosomes were excluded by this panel. This result was confirmed by fluorescence *in situ* hybridization (FISH) employing the mouse genomic clone ΦB3 (see below), which contains exon 1. The FISH technique, performed in Dr. Francke's laboratory, detected fluorescent signals in both chromatins of chromosome 2, refining the localization of the *Fbn1* locus to band F of this chromosome (2F) (Li et al., 1994).

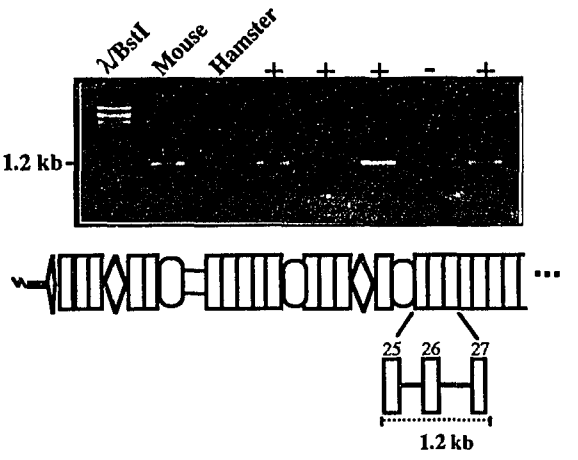


Figure 18: Example of specific PCR amplification of *Fbn1* sequences from mouse genomic DNA. (+) and (-) indicate DNA from hybrids including and excluding mouse chromosome 2, respectively. Diagram below indicates the amplified region of *Fbn1*.

This result is consistent with previous studies that have shown this murine chromosomal region to exhibit conserved synteny with the long arm of human chromosome 15.

Gene targeting

The high degree of conservation between the human and murine proteins reiterated the importance of maintaining the fibrillin structure. More generally, this finding supports the concept that mutations in the mouse fibrillin gene may produce manifestations similar to those of MFS.

Using the homologous recombination technique, we planed to generate reagents that will be used in the future to compare the effects of quantitative and qualitative changes in fibrillin on the development of transgenic mice. In order to derive the respective homologous recombination targeting constructs, genomic sequences from *Fbn1* were isolated and characterized. Since the use of isogenic DNA significantly increases the rate of homologous recombination in ES-cells, clones from specific regions of the *Fbn1* gene were isolated from a genomic library from the same mouse strain 129/Sv as the ES-cells utilized later (Deng and Capecchi, 1992). A clone containing exon 1 (ΦB3), and another

containing exons 24 through 28 (Φmg) were used for the construction of the targeting vectors BEnT and mgΔ[tgf2], respectively.

DNA from ΦB3 was used to engineer the targeting vector BEnT, designed to create an *Fbn1* null allele. Approximately 700 bp of the 5' of the *Fbn1* gene, including the initiator ATG and the sequence encoding the signal peptide, were substituted by the neo^r gene (fig. 19). In addition, a total of 9 kb of homologous sequences was included in BEnT in order to increase the frequency of homologous recombination events (fig 19). In order to study the effect of gross structural changes in fibrillin, we constructed the targeting vector mgΔ[tgf2] to delete one exon of the *Fbn1* gene. This vector consists of sequences 5' and 3' of exon 24, excluding the latter, flanking the neo^r gene (fig. 20). After the homologous recombination event, the resulting *Fbn1* allele should have an in-frame deletion of that exon, which encodes the second TGFbp motif of region D.

The BEnT and mgΔ[tgf2] constructs were electroporated separately into ES-cells. After selection with neomycin and FIAU, surviving colonies were picked and screened for homologous recombination events by Southern-blot analysis. In order to identify legitimate homologous recombination events, probes corresponding to sequences flanking the regions contained in the constructs were used for the Southern hybridization. In that way, regardless of the site of integration of the transgene, one band corresponding to the wild type allele was always detected. The appearance of a second band was solely due to the homologous recombination of the construct with one of the wild type alleles.

Southern-blot analysis of 160 neo-resistant/FIAU-sensitive ES-colonies electroporated with BEnT identified four positive clones for homologous recombination (B3EnT-23, -29, -60, and -156) (fig. 19). As estimated from the number of surviving cells selected with G418 alone, there was a two-fold enrichment for homologous recombination events by the negative selection with FIAU. One positive clone (mgΔ[tgf2]-66) was identified by Southern-blot analysis of 250 neo-resistant/FIAU-sensitive ES-colonies transfected with mgΔ[tgf2] (fig. 20). Similar to the first construct, a two-fold

enrichment by FIAU selection was observed. The positive clones were expanded and frozen in several aliquots at -70°C for future use.

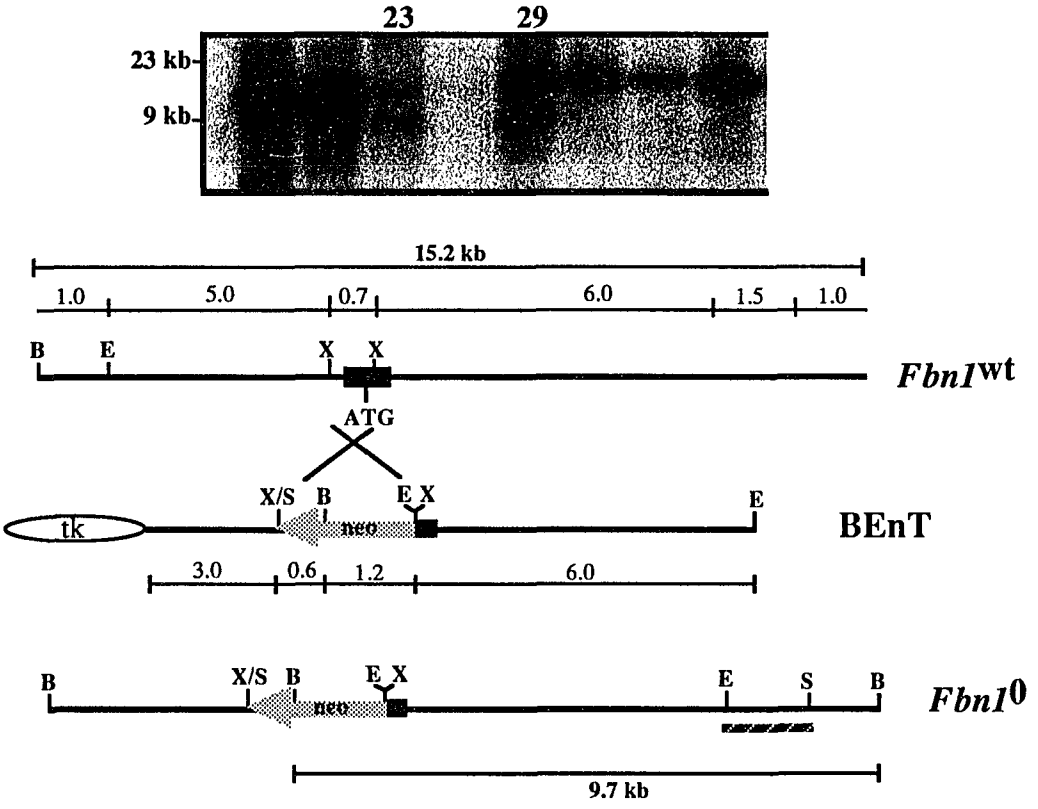


Figure 19: Null allele. Homologous recombination between the targeting vector BEnT and Fbn1. Size of DNA fragments are indicated in kb. Black rectangle represents exon 1 (initiator ATG is shown). Above, Southern blot analysis of ES-clones, including the two positive clones BEnT-23 and -29. Striped line indicates probe used for Southern analysis. Restriction sites are as follows: EcoRI (E), BamHI (B), XhoI (X), and SalI (S).

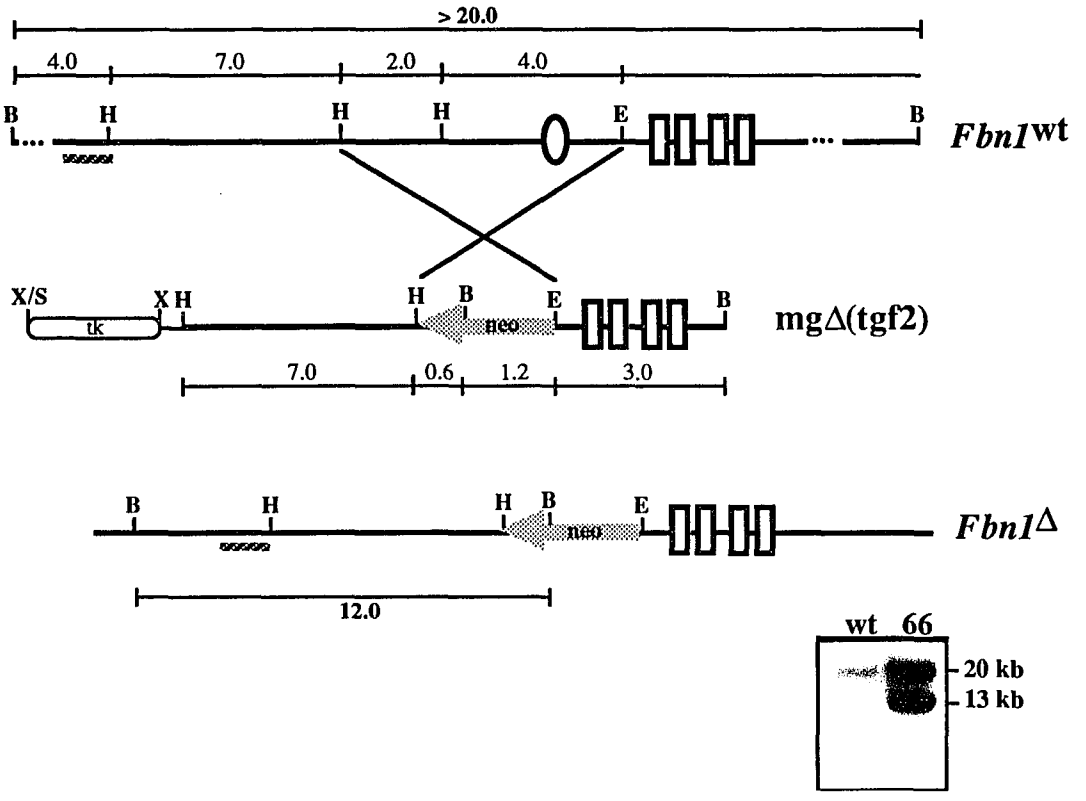


Figure 20: Exon 24 deletion. The wild type allele (above) undergoes homologous recombination with the targeting vector mgΔ(tgf2), generating a deletion of exon 24 (below). Exons are represented by rectangles (EGF-CB coding) and ovals (TGFbp coding). Striped line indicates probe used for Southern hybridization. Sizes of DNA fragments are indicated in kb. At the bottom, Southern-blot of the positive clone mgΔ(tgf2)-66. Restriction sites are as follows: EcoRI (E), BamHI (B), HindIII (H), and XhoI (X).

We are currently attempting to investigate the effect of the *Fbn1* mutations *in vivo* by injecting the recombinant ES-cells into mouse blastocysts. So far, approximately 10% of the injected and reimplanted blastocysts developed into embryos. However, due to husbandry problems in the mouse colony, none of these survived for more than two days after birth. New hosts will be tested in order to persue this line of investigation.

As an alternative, we tested the ability of the ES-cells to form microfibrils after differentiation in cell culture. Repeated efforts by others to differentiate the J-1 ES-cells *in vitro* have failed (K. Andrikopolous, personal communication). However, the use of a

novel protocol (provided by Dr. Stuhlman) allowed us to differentiate these cells into a mixture of fibroblasts and epithelial cell lines which synthesize fibrillin. Work is currently being performed to access the ability of the mutant ES-cell lines to form a microfibrillar network in culture. This will be accessed in two ways: first, at the RNA level, where the amount of mutant transcript, as well as the sequence of the deleted transcript, will be tested. Second, at the protein level, microfibril assembly will be detected by the same immunofluorescent assay used to distinguish normal from MFS individuals using anti-fibrillin antibodies (Hollister et al., 1990).

DISCUSSION

Fibrillin transcript:

The cloning experiments reported here, together with previously published data, indicate that the size of the full-length fibrillin mRNA is 9,663-nt (Corson et al., 1993; Lee et al., 1991; Maslen et al., 1991; Pereira et al., 1993). This is consistent with Northern-blot analysis which estimated the size of this transcript to be approximately 10-kb, and the protein data which shows that fibrillin is a 350kDa protein (Lee et al., 1991; Sakai et al., 1986). Since the sequence of the amino terminus of fibrillin is unknown, we relied on the cDNA data to determine the start site of translation of the fibrillin message. One potential initiation codon was identified at the 5' of the FBN1 transcript. Although the ORF extends upstream of this putative initiator ATG, there are several indications that the latter is the real site of translation initiation. This codon is placed within the context of a Kozak consensus sequence, and the corresponding methionine is immediately upstream of a hydrophobic sequence likely to be a signal peptide. Moreover, there is a significant decrease in phylogenetic conservation of the nucleotide and amino acid sequences upstream of the ATG/methionine. Accordingly, we beleive that the corresponding 8,613-nt ORF of the fibrillin transcript translates into a 2,871 amino acid protein with a predicted molecular weight of 347-kDa.

As mentioned earlier, concurrent with our studies Corson et al. (1993) described another three putative FBN1 exons (A, B and C) which are alternatively spliced 5' of exon 1. Their results suggested that the most abundant form of the FBN1 transcript is the one starting with exon 1. The other forms, containing exons A, B or C upstream of exon 1, are probably minor transcripts of no physiological relevance, since they do not contain an in-frame ATG upstream of the initiating codon in exon 1. The data from the murine fibrillin cDNA cloning supports this hypothesis. Sequence comparison between the 5' region of

human and mouse fibrillin transcripts revealed an abrupt divergence between the two upstream of the putative initiator ATG, supporting the concept of this being the start-site of translation (Yin et al., 1994). Primer extension indicates that *Fbn1* mRNA has a short 5' UT of 34-nt, and no sequence significantly homologous to the presumed human exons A, B or C has been identified in the *Fbn1* mRNA (Yin et al., 1994). Sequence analysis of approximately 2.0kb of mouse genomic sequence upstream of exon 1 identified one region of strong homology to human exon A (Yin et al., 1994). However, the homology starts at the 5' region of that exon, extending into the upstream human intronic sequence. Although the data may just reflect differences between the FBN1 and *Fbn1* mRNAs, they are surprising in view of the high conservation of the coding sequence of both genes. Therefore, the physiological relevance of exons A, B and C awaits further analysis.

Regardless of the precise position of the start site of transcription, the genomic sequence at the 5' of the FBN1 gene is extremely GC-rich and does not contain any TATA or CCAAT boxes. These characteristics are usually associated with promoters of genes that present a broad spatial and temporal pattern of expression (Dynan, 1986). In fact, several GC-rich promoters lacking a TATA box have been described for genes encoding housekeeping enzymes, such as hypoxanthine phosphoribosyl transferase (HPRT) and dihydrofolate reductase (DHFR) (Patel et al., 1986; Yang, 1984). This type of promoter sequence can also be found in genes involved in growth control, such as the human Ha-ras gene and human EGF-receptor gene (Ishii et al., 1985a; Ishii et al., 1985b). Moreover, the promoter of the human elastin gene also belongs to this category - it is GC-rich, does not have a TATA-box, and, although it presents two CAAT boxes, their location suggests that they are not functional (Bashir et al., 1989). This might be an important observation in view of the process of elastogenesis (Serafini-Fracassini, 1984). In other words, it is conceivable to argue that fibrillin and elastin gene expression are coordinated through similar transcriptional control mechanisms during the biogenesis of the elastic fiber.

Fibrillin protein structure:

Fibrillin is mostly composed of cysteine-rich modules, particularly of EGF-like motifs - 47 of these modules are found in this protein. The stabilizing cysteine interactions within these repeats may account for the earlier observation that approximately two thirds of the cysteines in fibrillin are involved in intrachain disulfide bonds (Keene et al., 1991; Maddox et al., 1989). In addition, the spacing between the last cysteine of one EGF-like motif and the first cysteine of the next one is constant throughout the fibrillin molecule. This conserved spacing may allow consecutive EGF-like motifs to form a longer common β-sheet which may account for the rod-shaped appearance of fibrillin under EM analysis (Maddox et al., 1989) (fig. 21).

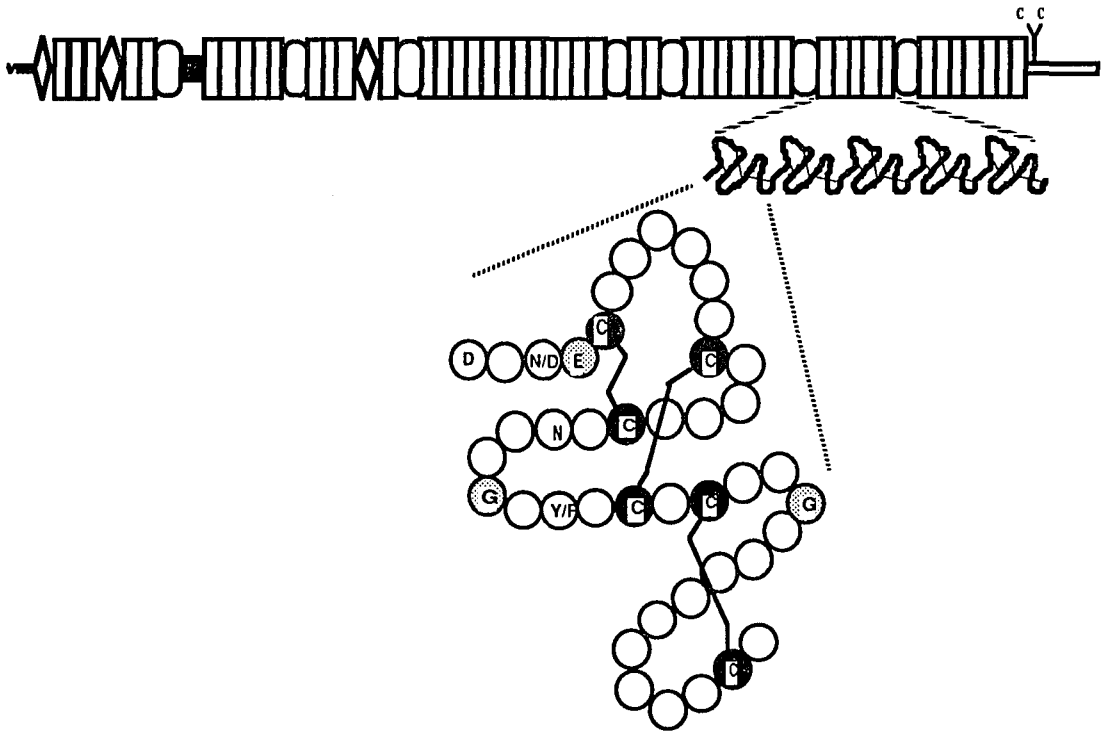


Figure 21: Folding of consecutive EGF-CB motifs in fibrillin. Below is the hypothetical folding of an EGF-CB motif. Cysteines (black circles), residues of the calcium binding consensus (white circles with letters), and residues invariant in fibrillin (stippled circles) are shown. Solid lines indicate disulfide bonds betweeen cysteines. Above that is the theoretical structure form by the interaction of consective EGF-CB.

EGF-like motifs are present in a variety of proteins as diverse as the mammalian LDL-receptor and the homeotic gene product *Notch* in *Drosophila* (Davis, 1990). Interestingly, there are also several ECM proteins which are partially composed of EGF-like motifs, including fibronectin, thrombospondin, and laminin (Davis, 1990; Engel, 1989). The function of the EGF-motif containing domains of these proteins is unknown. EGF-like domains have been proposed to have some mitogenic activity - for instance, fragments of laminin containing about 25 of these motifs stimulate cell growth in culture (Panayotou et al., 1989). However, it may be worth noting that it is merely a historical reason why these motifs are called EGF-like motifs - they were first described in that protein. This can mislead us into trying to correlate the function of these motifs in other proteins with the known mitogenic function of EGF. On the other hand, others have suggested that in some matrix proteins EGF-like motifs may serve merely as spacers, allowing different protein domains to interact with elements at more distant sites (Davis, 1990; Engel, 1989).

Most of the EGF-like motifs found in fibrillin (43 out of 47) contain an additional consensus of D, (D/N), N* and Y/F (where the asterisk denotes a potential hydroxylation site), identifying them with the calcium binding EGF-motif (Handford et al., 1990) (fig. 21). Studies of these EGF-CB motifs of the *Drosophila Notch* protein as well as of mutation in the same motifs of factor IX in hemophilia B suggest that they are involved in mediating protein-protein interactions (Handford et al., 1990; Rebay et al., 1991). Whether EGF-CB motifs play such role in fibrillin remains to be investigated. Several mutations altering one of the residues of the EGF-CB consensus have been identified in familial and sporadic cases of MFS (fig 3) (Dietz et al., 1992a; Dietz et al., 1992b; Dietz et al., 1993a; Hewett et al., 1993). Although in the *Drosophila Notch* gene mutations in different EGF-CB motifs are associated with different phenotypes (Kelley et al., 1987), so far no genotype/phenotype correlation has been established in MFS (Sykes, 1993).

Another previously described class of cysteine-rich motifs found in fibrillin is the TGFbp motif, with its characteristic cluster of three consecutive cysteine residues (Kanzaki et al., 1990). Region D contains six of these repeats, interspersed among the EGF-CB motifs. Since the cysteines in the latter are all involved in intrachain disulfide bonding, one can hypothesize that the strong interchain interactions observed among fibrillin molecules may occur via the cysteines in the TGFbp motifs. Moreover, one could also envision that the TGFbp motifs of adjacent fibrillin molecules must be aligned with each other in order to be able to form disulfide bonds. This organization in turn correlates well with the results of early immunolabeling studies of microfibrils with anti-fibrillin antibodies (Maddox et al., 1989). In addition to proposing a head-to-tail assembly of fibrillin into microfibrils, these experiments indicated that there is only one fibrillin-length between every two beaded structures of the microfibrils. For that to occur, the fibrillin molecules stretched between the beads must be aligned (fig. 22). One interesting inference from this model is that mutations in one fibrillin allele altering the number of repeats in region D would prevent the correct alignment of the normal and mutant molecules. As a result, this would disrupt the overall organization of fibrillin in the microfibrils. Deletions causing such structural abnormalities have been reported in the FBN1 gene of MFS patients (fig. 3) (Dietz et al., 1993a; Dietz et al., 1993b; Kainulainen et al., 1992). Further biochemical studies of these patients cells may help elucidate the effects of such structural changes in the assembly of fibrillin into microfibrils.

Finally, the third type of cysteine-rich module found in fibrillin is the Fib-motif, which has been first described in this protein. Closer examination of the Fib-motif revealed that it is a chimeric structure (fig. 23A). Its amino terminus is homologous to the amino terminus of the TGFbp motif - the main difference between them is the loss of one of the three clustered cysteine residues of the TGFbp motif. On the other hand, the carboxy-end of the Fib-motif resembles the carboxy-end of the EGF-CB repeat. In addition to presenting the same cysteine spacing, it also contains a glycine positioned two

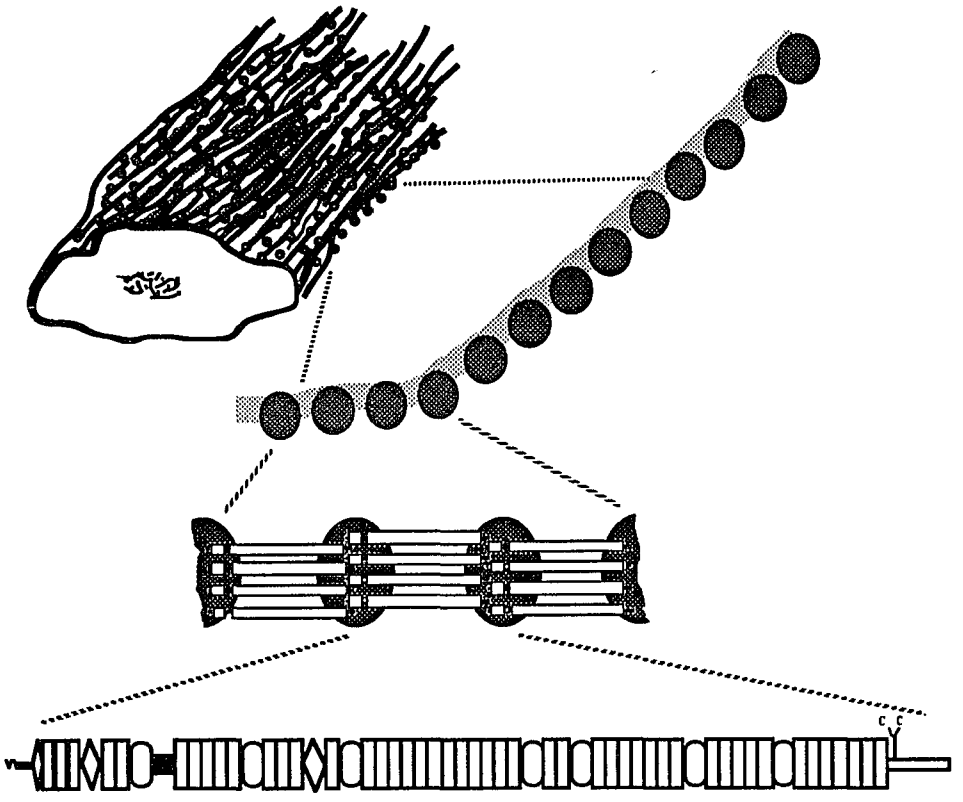


Figure 22: Assembly of fibrillin into elastic fiber. Aligned fibrillin molecules (bottom) assemble in a head-to-tail configuration, forming the beaded-string microfibrils. These will form a scaffold around the amorphous elastin core in the mature elastic fiber.

residues carboxy to the fifth cysteine, thus matching precisely the consensus sequence of the EGF-CB motifs of fibrillin. This observation suggests that the Fib-motif may have arisen from an unequal crossing-over between TGFbp and EGF-CB coding exons.

A combination of all of these different cysteine-rich motifs can be found in region B. This region contains slight variations of the TGFbp and Fib motifs (fig. 23). These differences in the primary sequence may allow alternative folding of these motifs, potentially rendering some of its many cysteine residues also available for interchain disulfide bonding. This in turn may account for some of the interactions required for a head-to-tail assembly of fibrillin molecules (Sakai et al., 1991).

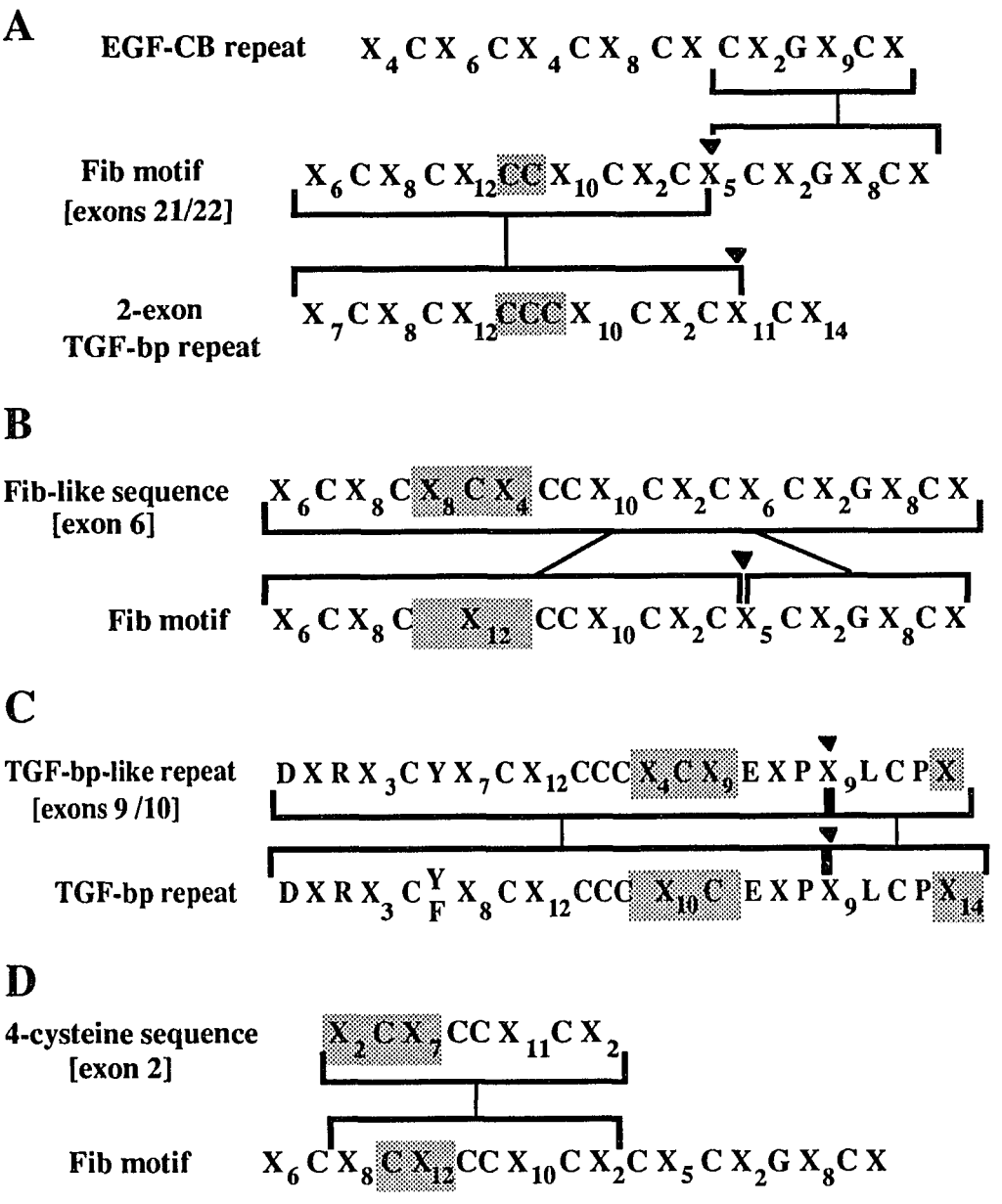


Figure 23: Postulated origin of cysteine-rich motifs of fibrillin. (X) represents any amino acid different from a cysteine. Black triagles indicate exon boundaries. Sequences hypothesized to have rearranged are shaded.

The high proline content of region C is unusual, and its structural significance unknown. Dystrophin presents a similar extended structure interrupted by a proline-rich

domain. This domain has been proposed to introduce hinges in dystrophin, conferring flexibility to this protein (Koening and Kunkel, 1990). By analogy, region C may allow fibrillin to bend. This could act in two ways; either by facilitating the presumed interactions of the residues of region B (head) with residues at the carboxy-end (tail) of a consecutive fibrillin molecule (Sakai et al., 1991) (fig. 22); or, perhaps, by accounting for the extension-contraction mechanism observed in microfibers. In fact, under EM, the periodicity of the beaded structures of microfibrils may range from 10 to 94 nm in stretched tissue preparations (Fleischmajer et al., 1991; Keene et al., 1991).

Fibrillin gene

This part of the work has generated both basic and applicable information about the FBN1 gene. Determination of the complete genomic organization provided insights into the evolution of the fibrillin gene family. Moreover, the same data have already been applied to the screening of MFS mutations at the genomic level. Finally, identification of polymorphic markers within the FBN1 gene has improved the pre-natal and pre-symptomatic diagnosis of familial MFS.

The exon organization of FBN1 reflects the modular arrangement of the fibrillin protein, i.e., most of the cysteine-rich modules in regions B and D are encoded by single exons (fig. 12). This type of modular organization was first described in the repetitive structure of the fibrillar collagen genes, where multiples of the Gly-X-Y motif characteristic of this gene family are encoded by single exons (Vuorio and de Crombrugghe, 1990; Yamada et al., 1980). Analogous to the postulated evolution of the collagen gene family the modular arrangement of fibrillin's coding sequence suggests that this gene may have evolved by duplications of ancestral cysteine-rich units (Yamada et al., 1980; Pereira et al., 1993). One of these units, the EGF-CB motif, is repeated 43 times in fibrillin. These are all encoded by single exons, with the exception of the last two EGF-CB in region D, which are encoded by a fused exon. The TGFbp motifs are encoded by either two exons or one

fused exon, suggesting that the ancestor unit, composed of two exons, later evolved into a single exon that was then duplicated. The same is the case for the Fib-motif - two exons encode the Fib-motif in region D, whereas one fused exon encodes a homologous sequence in region B. Interestingly, the TGFbp and the Fib motifs encoded by exons 6 and 9, respectively, underwent some rearrangements (fig. 23). In addition, exon 2 encodes a novel cysteine module that could have been derived from the amino portion of the Fib-motif (fig. 23). Altogether, these observations may indicate an evolutionarily delayed assembly of the 5' portion of the FBN1 gene. Finally, and still following the modular arrangement of fibrillin, regions A and C, initially defined based on their primary sequences, are also encoded by single exons. Interestingly, each exon of regions B, C and D starts with a 2nt-split codon and ends with a 1nt-split codon (fig. 13). The 3'-nt of one exon complements the most 5'-2nt-split codon of the following exon. Thus, the deletion of any number of exons from these regions should keep intact the reading frame of the fibrillin transcript.

Finally, it is worth noting that the sequences of transition between different regions of fibrillin are encoded by mosaic exons, i.e., exons that span the structural boundaries (exons 1, 10, 11 and 64). Similar junction exons have been described in the collagen genes, where they code for transition sequences between the structurally and functionally distinct triple-helical domain and the amino and carboxy propeptides of the collagen molecules. Consistent with the idea of a functional organization of the gene, it has been proposed that these exons may encode functional domains. Alternatively, it has also been suggested that the junction exons may just reflect how the propeptide-genes were fused to the helical-gene (reviewed by Boedtker et al., 1985). The latter hypothesis raises a very interesting point regarding the evolution of the fibrillin gene family. The FBN2 gene product presents a domain structure similar to fibrillin. Regions A, B, D and E are extremely conserved, consisting of the same structural elements (Zhang et al., 1994). However, region C, which is proline-rich in fibrillin, is rich in glycine residues in fibrillin-2. Thus, one may speculate the existence of a family of "region C-genes" which were

independently joined to the conserved "region B-gene" and "region D-gene" by the adapter-like junction exons. Therefore, perhaps region C provides a distinct function to each fibrillin protein. In addition, preliminary data from the FBN2 gene indicates that it presents the same genomic organization as FBN1 (Lee, et al. 1991; Zhang, personal communication), implying that both genes originated from a common progenitor gene, which might have also generated other fibrillin-like genes.

The intron-exon boundary sequences for all 65 FBN1 exons were determined in this study (fig. 13). The availability of these sequences will facilitate the screening of individuals for mutations in the fibrillin gene, since its entire coding sequence can now be amplified from genomic DNA as opposed to reversely transcribed RNAs. In fact, a set of PCR primers is now being developed based on those sequences in order to amplify each one of the 65 exons of the fibrillin gene. These will be made available to different groups by the Marfan Syndrome and Related Disorders Consortium. However, it should be pointed out that, due to the reported inefficiency of FBN1 mutation detection by PCR-based methods, one may consider using alternative approaches. For instance, traditional Southern blot analysis could detect large FBN1 deletions or rearrangements which, when present in heterozygosity, may remain undetected by PCR analysis.

The heterogeneity of FBN1 mutations involved in MFS has precluded the development of molecular diagnosis for this disease. Still, linkage analysis can be employed in familial MFS. The use of the three new intragenic markers described in this study, with an overall heterozygosity of 86%, will significantly improve the pre-natal and pre-symptomatic diagnosis of familial MFS. This is particularly important for the management of the cardiovascular manifestations of MFS patients. It has been proposed that the earlier the treatment of individuals at risk with β-adrenergic blockers, which slows the rate of aortic root dilatation, the greater the efficacy of this therapy (Shores et al., in press). Thus, the early identification of MFS patients will improve their clinical

management. In addition, exclusion of the MFS diagnosis will prevent the unnecessary follow-up and diagnostic testing of ambiguous cases.

Mouse fibrillin gene

The study of the murine fibrillin gene was undertaken with the long-term goal of developing a system in which to investigate microfibril assembly and pathogenesis. The immediate results provided further evidence for the start site of translation and the controversial composition of the 5' end of the fibrillin transcript. In addition, the ES-cell lines generated here represent an important step towards the development of mice carrying mutations in the *Fbn1* gene.

As mentioned above, sequence analysis of the 5' end of the murine transcript reinforced the results obtained from the characterization of the human FBN1 message. A potential initiator ATG can be found in the *Fbn1* transcript at the same relative position as the one in the human message. Moreover, the nucleotide homology between the two transcripts drops from 80% throughout the coding region to 42% upstream of that ATG, indicating that the latter region is structurally less important. In addition, different from FBN1, no evidence was found for alternative splicing at the 5' end of *Fbn1*. Taken together, these results indicate that the start site of translation of both *Fbn1* and FBN1 is at the ATG in exon 1. Moreover, they suggest that the minor forms of the human FBN1 transcript, containing exons A, B or C, may not be functionally relevant.

Murine fibrillin follows the same modular arrangement as the human protein, with the same five structurally distinct domains (fig. 17). Based on the high homology between the two proteins, one can expect that mutations disrupting the function of human fibrillin should have the same effect in the murine protein. This hypothesis constituted the basis for the experiments of gene targeting in ES-cells.

Mapping of the murine fibrillin gene to the chromosomal region 2F reinforced the conserved mouse/human synteny on mouse chromosome 2. That chromosomal region has

been found to contain genes homologous to human genes mapped to the chromosomal region 15q13-q21 (Lyon and Kirby, 1993). Incidentally, the *Tight-skin* (*Tsk*) locus, corresponding to a spontaneous dominant mutation in mice (Green et al., 1976), has been also mapped to mouse chromosome 2 band F (Siracusa et al., 1993). Mice heterozygous for the *Tsk* mutation present excessive collagen accumulation in the dermis and subdermal tissues, leading to increase skin thickness (Jiminez et al., 1986). These animals are potential models for the study of human scleroderma. This acquired disease of the connective tissue is characterized by excessive ECM deposition in the skin, heart, kidney and lung, and by microvascular injury (Medsger, 1993). Mapping of *Fbn1* to the *Tsk* locus makes this gene a candidate for the molecular lesion in the *Tsk* mice, indicating that the synthesis of fibrillin should be investigated in the affected animals. Although the manifestations of the *Tsk* mutation do not precisely mimic those of MFS, it is possible that mutations in some specific region of fibrillin generate distinct phenotypes. Alternatively, linkage of *Fbn1* to the *Tsk* phenotype may provide an additional marker useful for the positional cloning of the true *Tsk* gene. Towards this end, we are currently screening the *Fbn1* gene for informative VNTRs.

Homologous recombination was used to produce four ES-cell lines containing a null *Fbn1* allele ($Fbn1^0$). A 695-bp fragment, including the initiator ATG and the sequence coding for the signal peptide, was replaced by the neo^r gene in the $Fbn1^0$ allele (fig. 24). Since the promoter of *Fbn1* has not been characterized, it is possible that enough of its sequence remains in the $Fbn1^0$ allele to drive gene expression. Nevertheless, sequence analysis showed that if any transcript were made from the $Fbn1^0$ allele, in addition to lacking sequence coding for a signal peptide, it would have five out-of-frame ATGs upstream of the first in-frame ATG on exon 3 (fig. 24). Thus, in order for $Fbn1^0$ to produce anything resembling fibrillin, the translation machinery would have to skip those five potential initiation codons before starting transcription from the in-frame ATG on exon

3. Even so, this protein would lack the targeting sequence required for it to be secreted from cells. Thus, it is very likely that *Fbn1*^{0} is in fact a null allele.

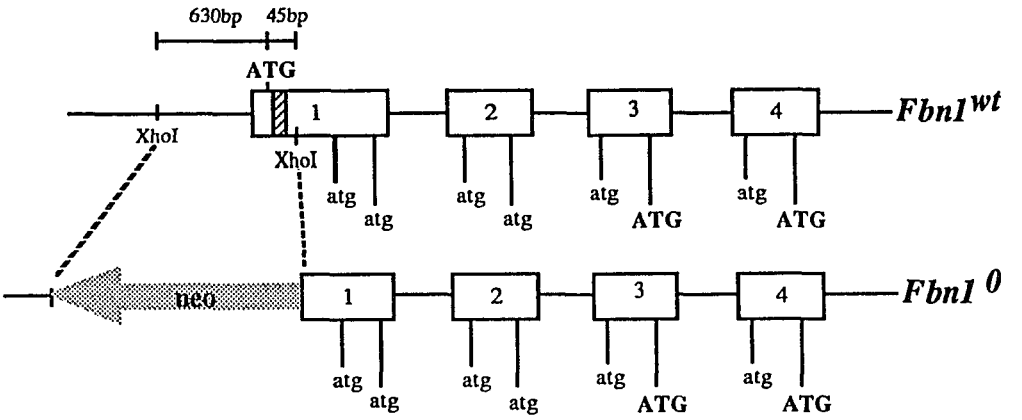


Figure 24: Null *Fbn1* allele. Exons are represented by rectangles, and the neo^{r} gene by an arrow. Signal peptide encoding sequence is depicted by a striped rectangle. Capital letters indicate in-frame codons, whereas small letters indicate out-of-frame codons.

Mice generated from the BenT ES-cells will allow the investigation of the effect of quantitative changes in fibrillin, an issue that has not been resolved in humans. According to the dominant negative model for fibrillin pathogenesis, a null-allele in heterozygosity should not disturb microfibril formation. No MFS patient carrying a null FBN1 allele has yet been described. However, one recent study described a patient with a mild variant of MFS carrying what might mimic a null allele (Dietz et al., 1993a). The level of expression of the mutant allele, consisting of a 4-bp insertion leading to an early stop codon at exon 41, was only 6% of that of the normal. Consistent with the dominant negative model, the authors proposed that the reduced amount of abnormal fibrillin monomers allows a preferential interaction among normal fibrillin molecules during microfibril assembly, justifying the patient's mild phenotype (Dietz et al., 1993a). However, in the same report, a 16% level of expression of a different mutant FBN1 transcript was detected in a patient

with severe MFS. The authors reconciled these two results by arguing that expression of the classic MFS phenotype may require a critical threshold ratio of 6-16% between mutant and normal transcripts (Dietz et al., 1993a). Although this explanation is consistent with the dominant negative mechanism of fibrillin pathogenesis, the effect of quantitative changes in fibrillin awaits further investigation.

A systematic analysis of fibrillin gene dosage may be performed in mice carrying the *Fbn1*^{0} allele. In fact, transgenic mice have been used to address the same issue in another dominantly inherited disease of the connective tissue, OI-type I. OI-type I is a very mild variant of the aforementioned neonatal lethal OI-type II (Byers, 1989b). The former is characterized by bone fragility, and contrasting OI-type II, patients with OI-type I live well into adulthood. Interestingly, both conditions are associated with mutations in the COL1A1 and COL1A2 genes (reviewed by Byers and Bonadio, 1989). The transgenic mouse strain *Mov-13* carries a proviral insert in the promoter region of that gene which was shown to cause inactivation of transcription from that COL1A1 allele. Studies of these mice, heterozygous for the null-COL1A1 allele, indicated that fiber structure and the integrity of the connective tissue are jeopardized by the reduced amount of collagen type I (Bonadio et al., 1990). Moreover, the mild phenotype of those mice resembled OI-type I, suggesting that a null COL1A1 allele present in heterozygosity is responsible for the mild phenotype. By analogy, the phenotype of mice heterozygous for the *Fbn1*^{0} allele may indicate a human condition caused by such mutation.

The second mutation generated in ES-cells was identified in clone mgΔ[tgf2]-66, and consists of an in-frame one-exon deletion in *Fbn1*. This exon, number 24, codes for the second TGFbp repeat in region D. Although this specific mutation has not been described in humans, other similar deletions have been found in the FBN1 transcript of MFS patients, including the skipping of exon 51, which encodes the carboxy half of the fifth TGFbp repeat in region D (Dietz et al., 1993a; Dietz et al., 1993b; Godfrey et al., 1993; Kainulainen et al., 1992). The current model predicts that such structural

modifications may have the same general effect, hampering the supposed alignment of fibrillin molecules required for microfibril assembly (fig. 22). It will be interesting to investigate whether this is the sole mode of action of such structural mutation. One would predict that mice homozygous for this mutation will have a homogeneous pool of truncated fibrillin molecules which may align with each other, and perhaps form microfibrils. On the other hand, the deletion of a TGFbp repeat, presumably involved in interchain disulfide bonds between fibrillin molecules, may weaken these interactions, rendering these microfibers fragile. Alternatively, if that repeat is a site of interaction between fibrillin and other microfibril components, the assembly of microfibrils will be disturbed in the homozygous mice despite of the interaction among the fibrillin molecules.

Interestingly, a recent report suggested the existence of a correlation between mutations around that exon of FBN1 and a neonatal severe form of MFS (n-MFS) (Kainulainen et al., 1994). Indeed, two missense mutations in exons 24 and 26, and one in-frame deletion of exon 25 were found in n-MFS individuals (fig. 3). However, the very first FBN1 mutation, a missense mutation in that same region, was described in a patient with classical MFS (Dietz et al., 1991a). Future studies on mice generated from the mgΔ[tgf2]-66 clone will contribute to clarify this emerging controversy.

The gene targeting approach can be further exploited in future studies aimed at determining structure/function correlations in fibrillin. For instance, an interesting domain to be investigated is the unique proline-rich region C of fibrillin. As hypothesized above, this region may provide the different functions of the members of the fibrillin family. One may test this hypothesis by swapping region C-encoding-exons between FBN1 and FBN2. Additionally, specific promoter elements of these two genes, once defined, may also be exchanged. Preliminary data indicate that the fibrillin genes are differentially expressed during development, suggesting a distinct role for each fibrillin protein (Zhang, personal communication). The gene targeting approach can help investigate the temporal/functional relationship between the two fibrillin proteins. Furthermore, one can also address the issue

of extensive clinical variability of MFS by studing the effects of a given mutation in different inbred strains of mice. These experiments may shed some light on the contribution of the genetic background for the MFS phenotype.

SUMMARY

In conclusion, the results of this study have substantially improved the diagnosis and identification of individuals at risk for MFS. Completion of the fibrillin transcript sequence will now enable the screening of MFS patients for mutations in its entire length. Moreover, the determination of the sequences of all FBN1 intron/exon junctions will greatly facilitate this process by extending the screening to the genomic level. In addition, the highly informative set of intragenic polymorphic markers identified in this work will be essential for the pre-natal and pre-symptomatic diagnosis of familial cases of MFS. Their usefulness has been exemplified in this report by the exclusion of the MFS phenotype in two individuals with ambiguous diagnosis.

In addition to contributing to the improvement of the diagnosis of MFS, this work has provided new knowledge of the fibrillin protein and gene. Completion of the primary sequence of fibrillin, along with elucidation of the gene organization, shed new light on the structure and function of this protein, as well as the evolutionary origin of the fibrillin gene family.

Finally, this study has initiated a broader body of work aimed at establishing the transgenic mouse as an animal model for the study of fibrillin biosynthesis and microfibril assembly. Isolation of the mouse fibrillin transcript has revealed a high degree of conservation between the human and murine fibrillin. In addition, it has provided additional evidence for the location of the putative start site of translation of the human protein. These data were utilized to design vectors for fibrillin gene targeting in mouse ES-cells. The successful generation of ES-cell lines bearing two different mutations in the fibrillin gene has laid the ground for future investigations aimed at characterizing the mechanisms and factors responsible for microfibril pathogenesis.

APPENDIX

(i) FBN1 coding sequence. Nucleotides are numbered starting at the ATG.

ATG CGT CGA GGG CGT CTG CTG GAG ATC GCC CTG GGA TTT ACC GTG CTT TTA GCG TCC TAC **120**

ACG AGC CAT GGG GCG GAC GCC AAT TTG GAG GCT GGG AAC GTG AAG GAA ACC AGA GCC AGT **180**

CGG GCC AAG AGA AGA GGC GGT GGA GGA CAC GAC GCG CTT AAA GGA CCC AAT GTC TGT GGA **240**

TCA CGT TAT AAT GCT TAC TGT TGC CCT GGA TGG AAA ACC TTA CCT GGC GGA AAT CAG TGT **300**

ATT GTC CCC ATT TGC CGG CAT TCC TGT GGG GAT GGA TTT TGT TCG AGG CCA AAT ATG TGC **360**

ACT TGC CCA TCT GGT CAG ATA GCT CCT TCC TGT GGC TCC AGA TCC ATA CAA CAC TGC AAT **420**

ATT CGC TGT ATG AAT GGA GGT AGC TGC AGT GAC GAT CAC TGT CTA TGC CAG AAA GGA TAC **480**

ATA GGG ACT CAC TGT GGA CAA CCT GTT TGT GAA AGT GGC TGT CTC AAT GGA GGA AGG TGT **540**

GTG GCC CCA AAT CGA TGT GCA TGC ACT TAC GGA TTT ACT GGA CCC CAG TGT GAA AGA GAT **600**

TAC AGG ACA GGC CCA TGT TTT ACT GTG ATC AGC AAC CAG ATG TGC CAG GGA CAA CTC AGC **660**

GGG ATT GTC TGC ACA AAA CAG CTC TGC TGT GCC ACA GTC GGC CGA GCC TGG GGC CAC CCC **720**

TGT GAG ATG TGT CCT GCC CAG CCT CAC CCC TGC CGC CGT GGC TTC ATT CCA AAT ATC CGC **780**

ACG GGA GCT TGT CAA GAT GTG GAT GAA TGC CAG GCC ATC CCC GGG CTC TGT CAG GGA GGA **840**

AAT TGC ATT AAT ACT GTT GGG TCT TTT GAG TGC AAA TGC CCT GCT GGA CAC AAA CTT AAT **900**

GAA GTG TCA CAA AAA TGT GAA GAT ATT GAT GAA TGC AGC ACC ATT CCT GGA ATC TGT GAA **960**

GGG GGT GAA TGT ACA AAC ACA GTC AGC AGT TAC TTT TGC AAA TGT CCC CCT GGT TTT TAC **1020**

ACC TCT CCA GAT GGT ACC AGA TGC ATA GAT GTT CGC CCA GGA TAC TGT TAC ACA GCT CTG **1080**

ACA AAC GGG CGC TGC TCT AAC CAG CTG CCA CAG TCC ATA ACC AAA ATG CAG TGC TGC TGT **1140**

GAT GCC GGC CGA TGC TGG TCT CCA GGG GTC ACT GTC GCC CCT GAG ATG TGT CCC ATC AGA **1200**

GCA ACC GAG GAT TTC AAC AAG CTG TGC TCT GTT CCT ATG GTA ATT CCT GGG AGA CCA GAA **1260**

TAT CCT CCC CCA CCC CTT GGC CCC ATT CCT CCA GTT CTC CCT GTT CCT CCT GGC TTT CCT **1320**

CCT GGA CCT CAA ATT CCG GTC CCT CGA CCA CCA GTG GAA TAT CTG TAT CCA TCT CGG GAG **1380**

CCA CCA AGG GTG CTG CCA GTA AAC GTT ACT GAT TAC TGC CAG TTG GTC CGC TAT CTC TGT **1440**

CAA AAT GGA CGC TGC ATT CCA ACT CCT GGG AGT TAC CGG TGT GAG TGC AAC AAA GGG TTC **1500**

CAG CTG GAC CTC CGT GGG GAG TGT ATT GAT GTT GAT GAA TGT GAG AAA AAC CCC TGT GCT **1560**

GGT GGT GAG TGT ATT AAC AAC CAG GGT TCG TAC ACC TGT CAG TGC CGA GCT GGA TAT CAG **1620**

AGC ACA CTC ACG CGG ACA GAA TGC CGA GAC ATT GAT GAG TGT TTA CAG AAT GGC CGG ATC **1680**

TGC AAT AAT GGA CGC TGC ATC AAC ACA GAT GGC AGT TTT CAT TGC GTG TGT AAT GCG GGC **1740**

TTT CAT GTT ACA CGA GAT GGG AAG AAC TGT GAA GAT ATG GAT GAA TGC AGC ATA AGG AAC **1800**

ATG TGC CTT AAT GGA ATG TGT ATC AAT GAA GAT GGC AGT TTT AAA TGT ATT TGC AAA CCT **1860**

GGA TTC CAG CTG GCA TCA GAT GGA CGT TAT TGC AAA GAC ATT AAC GAG TGT GAA ACC CCT **1920**

GGG ATC TGC ATG AAT GGG CGT TGC GTC AAC ACT GAT GGC TCC TAC AGA TGT GAA TGC TTC **1980**

CCT GGA CTG GCT GTG GGT CTG GAT GGC CGT GTG TGT GTT GAC ACA CAC ATG CGG AGC ACA **1940**

TGC TAT GGT GGA TAC AAG AGA GGC CAG TGT ATC AAA CCT TTG TTT GGT GCT GTC ACT AAA **2100**

TCT GAA TGC TGT TGC GCC AGC ACT GAG TAT GCA TTT GGG GAA CCT TGC CAG CCG TGT CCT **2160**

GCA CAG AAT TCA GCG GAA TAT CAG GCA CTC TGC AGC AGT GGG CCA GGA ATG ACG TCA GCA **2220**

GGC AGT GAT ATA AAT GAA TGT GCA CTA GAT CCT GAT ATT TGC CCA AAT GGA ATC TGT GAA **2280**

AAC CTT CGT GGG ACC TAT AAA TGT ATA TGC AAT TCA GGA TAT GAA GTG GAT TCA ACT GGG **2340**

AAA AAC TGC GTT GAT ATT AAT GAA TGT GTA CTG AAC AGT CTC CTT TGT GAC AAT GGA CAA **2400**

TGT AGA AAT ACT CCT GGA AGT TTT GTC TGT ACC TGC CCC AAG GGA TTT ATC TAC AAA CCT **2460**

GAT CTA AAA ACA TGT GAA GAC ATT GAT GAA TGC GAA TCA AGT CCT TGC ATT AAT GGA GTC **2520**

TGC AAG AAC AGC CCA GGC TCT TTT ATT TGT GAA TGT TCT TCT GAA AGT ACT TTG GAT CCA **2580**

ACA AAA ACC ATC TGC ATA GAA ACC ATC AAG GGC ACT TGC TGG CAG ACT GTC ATT GAT GGG **2640**

CGA TGT GAG ATC AAC ATC AAT GGA GCC ACC TTA AAG TCC CAG TGC TGC TCC TCC CTC GGT **2700**

GCT GCG TGG GGA AGC CCG TGC ACC CTA TGC CAA GTT GAT CCC ATA TGT GGT AAA GGG TAC **2760**

TCA AGA ATT AAA GGA ACA CAA TGT GAA GAT ATA GAT GAA TGT GAA GTG TTC CCA GGA GTG **2820**

TGT AAA AAT GGC CTG TGT GTT AAC ACT AGG GGG TCA TTC AAG TGT CAG TGT CCC AGT GGA **2880**

ATG ACT TTG GAT GCC ACA GGA AGG ATC TGT CTT GAT ATC CGC CTG GAA ACC TGC TTC CTG **2940**

AGG TAC GAG GAC GAG GAG TGC ACC CTG CCT ATT GCT GGC CGC CAC CGC ATG GAC GCC TGC **3000**

TGC TGC TCC GTC GGG GCA GCC TGG GGT ACT GAG GAA TGC GAG GAG TGT CCC ATG AGA AAT **3060**

ACT CCT GAG TAC GAG GAG CTG TGT CCG AGA GGA CCC GGA TTT GCC ACA AAA GAA ATT ACA **3120**

AAT GGA AAG CCT TTC TTC AAA GAT ATC AAT GAG TGC AAG ATG ATA CCC AGC CTC TGC ACC **3180**

CAC GGC AAG TGC AGA AAC ACC ATT GGC AGC TTT AAG TGC AGG TGT GAC AGC GGC TTT GCT **3240**

CTT GAT TCT GAA GAA AGG AAC TGC ACA GAC ATT GAC GAA TGC CGC ATA TCT CCT GAC CTC **3300**

TGT GGC AGA GGC CAG TGT GTG AAC ACC CCT GGG GAC TTT GAA TGC AAG TGT GAC GAA GGC **3360**

TAT GAA AGT GGA TTC ATG ATG ATG AAG AAC TGC ATG GAT ATT GAT GAG TGT CAG AGA GAT **3420**

CCT CTC CTA TGC CGA GGT GGT GTT TGC CAT AAC ACA GAG GGA AGT TAC CGC TGT GAA TGC **3480**

CCG CCT GGC CAT CAG CTG TCC CCC AAC ATC TCC GCG TGT ATC GAC ATC AAT GAA TGT GAG **3540**

CTG AGT GCA CAC CTG TGC CCC AAT GGC CGT TGC GTG AAC CTC ATA GGG AAG TAT CAG TGT **3600**

GCC TGC AAC CCT GGC TAC CAT TCA ACT CCC GAT AGG CTA TTT TGT GTT GAC ATT GAT GAA **3660**

TGC AGC ATA ATG AAT GGT GGT TGT GAA ACC TTC TGC ACA AAC TCT GAA GGC AGC TAT GAA **3720**

TGT AGC TGT CAG CCG GGA TTT GCA CTA ATG CCT GAC CAG AGA TCA TGC ACC GAC ATC GAT **3780**

GAG TGT GAA GAT AAT CCC AAT ATC TGT GAT GGT GGT CAG TGC ACA AAT ATC CCT GGA GAG **3840**

TAC AGG TGC TTG TGT TAT GAT GGA TTC ATG GCA TCT GAA GAC ATG AAG ACT TGT GTA GAT **3900**

GTC AAT GAG TGT GAC CTG AAT CCA AAT ATC TGC CTA AGT GGG ACC TGT GAA AAC ACG AAA **3960**

GGC TCA TTT ATC TGC CAC TGT GAT ATG GGC TAC TCC GGC AAA AAA GGA AAA ACT GGC TGT **4000**

ACA GAC ATC AAT GAA TGT GAA ATT GGA GCA CAC AAC TGT GGC AAA CAT GCT GTA TGT ACC **4080**

AAT ACA GCA GGA AGC TTC AAA TGT AGC TGC AGT CCC GGG TGG ATT GGA GAT GGC ATT AAG **4140**

TGC ACT GAT CTG GAC GAA TGT TCC AAT GGA ACC CAT ATG TGC AGC CAG CAT GCA GAC TGC **4200**

AAG AAT ACC ATG GGA TCT TAC CGC TGT CTG TGC AAG GAA GGA TAC ACA GGT GAT GGC TTC **4260**

ACT TGT ACA GAC CTT GAT GAG TGC TCT GAG AAC CTG AAT CTC TGT GGC AAT GGC CAG TGC **4320**

CTC AAT GCA CCA GGA GGA TAC CGC TGT GAA TGC GAC ATG GGC TTC GTG CCC AGT GCT GAC **4380**

GGG AAA GCC TGT GAA GAT ATT GAT GAG TGC TCC CTT CCG AAC ATC TGT GTC TTT GGA ACT **4440**

TGC CAC AAC CTC CCT GGC CTG TTC CGC TGT GAG TGT GAG ATA GGC TAC GAA CTG GAC AGA **4500**

AGC GGC GGG AAC TGC ACA GAT GTG AAT GAA TGC CTG GAT CCA ACC ACG TGC ATC AGT GGG **4560**

AAC TGT GTC AAC ACT CCA GGC AGC TAT ATC TGT GAC TGC CCA CCT GAT TTT GAA CTG AAC **4620**

CCA ACT CGA GTT GGC TGT GTT GAT ACC CGC TCT GGA AAT TGC TAT TTG GAT ATT CGA CCT **4680**

CGA GGA GAC AAT GGA GAT ACA GCC TGC AGC AAT GAA ATT GGA GTT GGT GTT TCC AAA GCT **4740**

TCC TGC TGC TGT TCT CTG GGT AAA GCC TGG GGT ACT CCT TGT GAG ATG TGT CCT GCT GTG **4800**

AAC ACA TCC GAG TAC AAA ATT CTT TGT CCT GGA GGG GAA GGT TTC CGA CCA AAT CCT ATC **4860**

ACC GTT ATA TTG GAA GAT ATT GAT GAG TGC CAG GAG CTA CCA GGG CTG TGC CAA GGA GGA **4920**

AAA TGT ATC AAC ACC TTT GGG AGT TTC CAG TGC CGC TGT CCA ACC GGC TAC TAC CTG AAT **4980**

GAA GAT ACA CGA GTG TGT GAT GAT GTG AAT GAA TGT GAG ACT CCT GGA ATC TGT GGT CCA **5040**

GGG ACA TGT TAC AAC ACC GTT GGC AAC TAC ACC TGT ATC TGT CCT CCA GAC TAC ATG CAA **5100**

GTG AAT GGG GGA AAT AAT TGC ATG GAT ATG AGA AGA AGT TTG TGC TAC AGA AAC TAC TAT **5160**

GCT GAC AAC CAG ACC TGT GAT GGA GAA TTG TTA TTC AAC ATG ACC AAG AAG ATG TGC TGC **5220**

TGT TCC TAC AAC ATT GGC CGG GCG TGG AAC AAG CCC TGT GAA CAG TGT CCC ATC CCA AGT **5280**

ACA GAT GAG TTT GCT ACA CTC TGT GGA AGT CAA AGG CCA GGC TTT GTC ATC GAC ATT TAT **5340**

ACC GGT TTA CCC GTT GAT ATT GAT GAG TGC CGG GAG ATC CCA GGG GTC TGT GAA AAT GGA **5400**

GTG TGT ATC AAC ATG GTT GGC AGC TTC CGA TGT GAA TGT CCA GTG GGA TTC TTC TAT AAT **5460**

GAC AAG TTG TTG GTT TGT GAA GAT ATT GAC GAG TGT CAG AAC GGC CCA GTG TGC CAG CGC **5520**

AAC GCC GAA TGC ATC AAC ACT GCA GGC AGC TAC CGC TGT GAC TGT AAG CCC GGC TAC CGC **5580**

TTC ACC TCC ACA GGA CAG TGC AAT GAT CGT AAT GAA TGT CAA GAA ATC CCC AAT ATA TGC **5640**

AGT CAT GGG CAG TGC ATT GAC ACA GTT GGA AGC TTT TAT TGC CTT TGC CAC ACT GGT TTT **5700**

AAA ACA AAT GAT GAC CAA ACC ATG TGC TTG GAC ATA AAT GAA TGT GAA AGA GAT GCC TGT **5760**

GGG AAT GGA ACT TGC CGG AAC ACA ATT GGT TCC TTC AAC TGC CGC TGC AAT CAT GGT TTC **5820**

ATC CTT TCT CAC AAC AAT GAC TGT ATA GAT GTT GAT GAA TGT GCA AGT GGA AAT GGG AAT **5880**

CTT TGC AGA AAT GGC CAA TGC ATT AAT ACA GTG GGG TCT TTC CAG TGC CAG TGC AAT GAA **5940**

GGC TAT GAG GTG GCT CCA GAT GGG AGG ACC TGT GTG GAT ATC AAT GAA TGT CTT CTA GAA **6000**

CCC AGA AAA TGT GCA CCA GGT ACC TGT CAA AAC TTG GAT GGG TCC TAC AGA TGC ATT TGC **6060**

CCA CCT GGA TAC AGT CTT CAA AAT GAG AAG TGT GAA GAT ATT GAT GAG TGT GTC GAA GAG **6120**

CCA GAA ATT TGT GCC CTG GGC ACA TGC AGT AAC ACT GAA GGC AGC TTC AAA TGT CTG TGT **6180**

CCA GAA GGG TTT TCC TTG TCC TCC AGT GGA AGA AGG TGC CAA GAT TTG CGA ATG AGC TAC **6240**

TGT TAT GCG AAG TTT GAA GGA GGA AAG TGT TCA TCA CCC AAA TCC AGA AAT CAC TCC AAG **6300**

CAG GAA TGC TGC TGT GCC TTG AAG GGA GAA GGC TGG GGA GAC CCC TGC GAG CTC TGC CCC **6360**

ACG GAA CCT GAT GAG GCC TTC CGC CAG ATA TGT CCT TAT GGA AGT GGG ATC ATC GTG GGA **6420**

CCT GAT GAT TCA GCA GTT GAT ATG GAC GAA TGC AAA GAA CCC GAT GTC TGT AAA CAT GGA **6480**

CAG TGC ATC AAT ACA GAT GGT TCC TAT CGC TGC GAG TGT CCC TTT GGT TAT ACT CTA GCA **6540**

GGG AAT GAA TGT GTA GAT ACT GAT GAA TGT TCT GTT GGC AAT CCT TGT GGA AAT GGA ACC **6600**

TGC AAG AAT GTG ATT GGA GGT TTT GAA TGC ACC TGC GAG GAG GGA TTT GAG CCC GGT CCA **6660**

ATG ATG ACA TGT GAA GAT ATA AAT GAA TGT GCC CAG AAT CCT CTG CTC TGT GCC TTC CGA **6720**

TGT GTG AAC ACT TAT GGG TCA TAT GAA TGC AAA TGT CCC GTG GGA TAT GTG CTC AGA GAA **6780**

GAC CGT AGG ATG TGC AAA GAT GAG GAT GAG TGT GAA GAG GGA AAA CAT GAC TGT ACT GAA **6840**

AAA CAA ATG GAA TGC AAG AAC CTC ATT GGC ACA TAT ATG TGC ATC TGT GGA CCC GGG TAT **6900**

CAG CGG AGA CCT GAT GGA GAA GGC TGT GTA GAT GAG AAT GAA TGT CAG ACG AAG CCA GGG **6960**

ATC TGT GAG AAT GGG CGC TGC CTC AAC ACC CGT GGG AGC TAC ACC TGT GAG TGT AAT GAT **7020**

GGG TTT ACC GCC AGC CCC AAC CAG GAC GAG TGC CTT GAC AAT CGG GAA GGG TAC TGC TTC **7080**

ACA GAG GTG CTA CAA AAC ATG TGT CAG ATC GGC TCC AGC AAC AGG AAC CCC GTC ACC AAA **7140**

TCG GAA TGC TGC TGT GAC GGA GGG AGA GGC TGG GGT CCC CAC TGT GAG ATC TGC CCT TTC **7200**

CAG GGG ACT GTG GCT TTC AAG AAA CTC TGT CCC CAT GGC CGA GGA TTC ATG ACC AAT GGA **7260**

GCA GAT ATC GAT GAA TGC AAG GTT ATT CAC GAT GTT TGC CGA AAT GGG GAA TGT GTC AAT **7320**

GAC AGA GGA TCA TAT CAT TGC ATT TGT AAA ACT GGG TAC ACT CCA GAT ATA ACT GGG ACT **7380**

TCC TGT GTA GAT CTG AAC GAG TGC AAC CAG GCT CCC AAA CCC TGC AAT TTT ATC TGC AAA **7440**

AAC ACA GAA GGG AGT TAC CAG TGT TCA TGC CCG AAA GGC TAC ATT CTG CAA GAG GAT GGA **7500**

AGG AGC TGC AAA GAT CTT GAT GAG TGT GCA ACC AAG CAA CAC AAC TGC CAG TTC CTA TGT **7560**

GTT AAC ACC ATT GGC GGC TTC ACA TGC AAA TGT CCT CCC GGA TTT ACC CAA CAC CAT ACG **7620**

TCC TGC ATT GAT AAC AAT GAA TGC ACC TCT GAC ATC AAT CTG TGC GGG TCT AAG GGC ATT **7680**

TGC CAG AAC ACT CCT GGA AGC TTC ACC TGT GAA TGC CAG CGG GGA TTC TCA CTT GAT CAG **7740**

ACC GGC TCC AGC TGT GAA GAC GTG GAC GAG TGT GAG GGT AAC CAC CGC TGC CAG CAT GGC **7800**

TGC CAG AAC ATC ATT GGG GGC TAC AGG TGC AGC TGC CCC CAG GGC TAC CTC CAG CAC TAC **7860**

CAG TGG AAC CAG TGT GTT GAT GAA AAC GAA TGC CTC AGC GCT CAC ATC TGC GGA GGA GCC **7920**

TCC TGT CAC AAC ACC CTG GGG AGC TAC AAG TGC ATG TGT CCC GCC GGC TTC CAG TAT GAA **7980**

CAG TTC AGT GGA GGA TGC CAA GAC ATC AAT GAA TGT GGC TCT GCG CAG GCC CCC TGC AGC **8040**

TAT GGC TGT TCC AAT ACC GAG GGC GGT TAC CTG TGT GGC TGT CCA CCT GGT TAC TTC CGC **8100**

ATA GGC CAA GGG CAC TGT GTT TCT GGA ATG GGC ATG GGC CGA GGA AAC CCA GAG CCA CCT **8160**

GTC AGT GGT GAA ATG GAT GAC AAT TCA CTC TCC CCA GAG GCT TGT TAC GAG TGT AAG ATC **8220**

AAT GGC TAC CCC AAA CGG GGC AGG AAA CGG AGA AGC ACA AAC GAA ACT GAT GCC TCC AAT **8280**

ATC GAG GAT CAG TCT GAG ACA GAA GCC AAT GTG AGT CTT GCA AGT TGG GAT GTT GAG AAG **8340**

ACA GCC ATC TTT GCT TTC AAT ATT TCC CAC GTC AGT AAC AAG GTT CGA ATC CTA GAA CTC **8400**

CTT CCA GCT CTT ACA ACT CTG ACG AAT CAC AAC AGA TAC TTG ATC GAA TCT GGA AAT GAA **8460**

GAT GGC TTC TTT AAA ATC AAC CAA AAG GAA GGG ATC AGC TAC CTC CAC TTC ACA AAG AAG **8520**

AAG CCA GTG GCT GGA ACC TAT TCA TTA CAA ATC AGT AGT ACT CCA CTT TAT AAA AAG AAA **8580**

GAA CTT AAC CAA CTA GAA GAC AAA TAT GAC AAA GAC TAC CTC AGT GGT GAA CTG GGT GAT **8640**

AAT CTG AAG ATG AAA ATC CAG GTT TTG CTT CAT TAA TTC ACC ATC CAG AGA CCA AAT AAT **8700**

TAA AAG AAA AAC AAA TAT AGA TAG GTA GAA CTA TAT TTT CCC CCA ATC AGA ATC ATC ATA **8760**

TCA TAG GTA CAA TCT TTC ACC AAG TAA ATT TGT ATA AAT AAG CAC TAT TCT TTG TAT TAC **8820**

CAA AGC AAG GTA CAG GTG ACT ACC CTA GTT CAA AAC AAC CAC TTT CTC AGG CTT CTC ATG **8880**

TGT GTA GCT AAG CTA CCT TGT CAT ATG TGT TGA TTC TTG AAA ACT GGG ACG TGT ATT TCC **8940**

ATT GGG GGT TGG CCA TTT ATG CTG ACA TGC CAT CCT TCC AGC AAA CGT AYG GGA ATG TGC **9000**

TTT CAA TTG ATG GAC TAC TCT ATT TTT TGC AAA TTT GTA AAC TTT GCT TCT CCA AAT ACA **9060**

AGT ACT AGG TTG TCC ATT TAT GGT ACC TAT TTG GTG CTA GTA AAT TTT CAA ACT AGA TTT **9120**

ATA AAT GCA CTG TAA TAT GTA CAC AAC TTA GAA ACC AAA TTA CAA GTA TTC AGT TCC AAT **9180**

ACT TCA TTA ATT TCA ATC AAC CAA AGT TAG TTC AGT AGC TTA TCT CAG TTA TGA GTA TAA **9240**

TAC ATT ACA TGT AAA TTA AGT GTG TGT ATA CTG TAA TCG TGC TAT TTT TTA TCA TTG AAA **9300**

CAT TTA TAA ACT AGA ATA ATA ATG CCC TTA ATG TGA GGG TTT GTA ATG GTG CTT ATT AAG **9360**

ACC AAA GAC TTG TTA AAT GTA TAC ACC AAG TGG TAA TGA AAT TTC KGT GAC TGG CCC ACA **9420**

CGT GCA TAG AGG TCT GGG AGG ACC AGG AAA CAG CCT CAG TGG CCA GAG GAT CAC CAG TGC **9480**

ATC CTT CAT CAC AGC ATG TGC AAT ATG CCA AGA TTA CCC TCG GTC ATT CCT GTC AAC AAG **9540**

GGG TCA ATG TCA TAA ATG TCA CAA TAA AAC AAT CTC TTC TTT TTT TTA GTT TAA AAA AAA **9600**

AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA

(ii) Allele size and frequency for FBN1 intragenic VNTRs. (analysis performed by Dr. Dietz. Table adapted from Pereira et al., 1994)

	mts-1		mts-2		mts-3		mts-4	
Allele	**size**(bp)	**freq.**	**size**(bp)	**freq.**	**size**(bp)	**freq.**	**size**(bp)	**freq.**
1	146	0.04	165	0.00#	197	0.01	128	0.06
2	144	0.00*	163	0.07	192	0.78	126	0.02
3	142	0.05	161	0.05	187	0.08	124	0.07
4	140	0.01	159	0.07	182	0.13	122	0.00#
5	138	0.61	157	0.08			120	0.05
6	136	0.07	155	0.07			118	0.06
7	134	0.05	153	0.02			116	0.52
8	132	0.03	151	0.00*			114	0.11
9	130	0.13	149	0.01			112	0.11
10	128	0.01	147	0.02				
11			145	0.51				
12			143	0.01				
13			141	0.00*				
14			139	0.01				
15			137	0.08				
Het.		**0.50**		**0.58**		**0.30**		**0.70**

C.Het. 0.86

* allele not observed upon screening of 100 chromosomes, but predicted by allele sizing

allele not observed upon screening of 100 chromosomes, but seen during pedigree analysis

Het.: heterozygosity of each marker. C.Het.: combined heterozygosity of the four markers.

BIBLIOGRAPHY

Abraham, P. A., Perejda, A.J., Carnes, W.H., and Uitto, J. (1982). Marfan syndrome. Demonstration of abnormal elastin in aorta. J. Clin. Invest., 70, 1245.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1989). In Molecular Biology of the Cell (pp. 414). New York: Garland Publishing, Inc.

Appel, A., Horwitz, A.L., and Dorfman, A. (1979). Cell-free synthesis of hyaluronic acid in Marfan syndrome. J. Biol. Chem, 254, 12199.

Bashir, M. M., Indik, Z., Yeh, H., Ornstein-Goldstein, N., Rosenbloom, J.C., Abrams, W., Fazio, M., Uitto, J., and Rosenbloom, J. (1989). Characterization of the complete human elastin gene. Delineation of unusual features in the 5'-flanking region. J. Biol. Chem., 264, 8887.

Beals, R. K., and Hecht, F. (1971). Congenital contractural arachnodactyly: a heritable disorder of connective tissue. J. Bone Joint Surg., 53, 887.

Beighton, P., de Paepe, A., Danks, D., Finidori, G., et al. (1988). International nosology of heritable disorders of connective tissue, Berlin, 1986. Am. J. Med. Genet., 29, 581.

Besser, T.E., Potter, K.A., Bryan, G.M., and Knowlen, G.A. (1990). An animal model for Marfan syndrome. Am. J. Med. Genet., 37, 159.

Blanton, S. H., Sarfarazi, M., Eiberg, H., de Groote, J., Farndon, P.A., Kilpatrick, M.W., Child, A.H., Pope, F.M., Peltonen, L., Francomano, C.A., Boileau, C., Keston, M., Tsipouras, P. (1990). An exclusion map of Marfan syndrome. J. Med. Genet., 27, 73.

Boileau, C., Jondeau, G., Bonaiti, C., Coulon, M., Delorme, G., Dubourg, O., Bourdarias, J.P., and Junien, C. (1990). Linkage analysis of five fibrillar collagen loci in a large French Marfan syndrome family. J. Med. Genet., 27, 78.

Bonadio, J., Saunders, T.L., Tsai, E., Goldstein, S.A., Morris-Wiman, J., Brinkley, L., Dolan, D.F., Altschuler, R.A., Hawkins, J.E., Bateman, J.F., Mascara, T., and Jaenisch, R. (1990). Transgenic mouse model of the mild dominant form of osteogenesis imperfecta. Proc. Natl. Acad. Sci. U.S.A., 87, 7145.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980). Construction of a genetic linkage map using restriction length polymorphisms. Am. J. Hum. Genet., 32, 314.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimeras from embryo-derived teratocarcinoma cell lines. Nature, 309, 255.

Bressan, G. M., Daga-Gordini, D., Colombatti, A., Castellani, I., Marigo, V., and Volpin, D. (1993). Emilin, a component of elastic fibers preferentially located at the elastin-microfibrils interface. J. Cell Biol., 121, 201.

Byers, P. H. (1983). Inherited disorders of collagen biosynthesis: Ehlerst-Danlos syndrome, the Marfan syndrome, and osteogenesis imperfecta. In J. A. Spittel (Eds.), J. Clin. Med. (pp. 1). Philadelphia: Harper & Row, publishers.

Byers, P. H., and Bonadio, J.F. (1989a). In B. R. Olsen and Nimmi, M. (Eds.), Collagen: Molecular Biology. (pp. 125). Boca Raton: CRC.

Byers, P. H. (1989b). In C. R. Scriver Beaudet, A.L., Sly, W.S., and Valle, D. (Eds.), The Metabolic Bases for Inherited Diseases (pp. 2805). New York: McGraw-Hill.

Capecchi, M.P. (1989). The new mouse genetics: altering the genome by gene targeting. Trends Genet., 5, 70.

Clark, B. A., Maslen, C.L., Sakai, L.Y., Dhalimi, M.A., Litt, R., and Litt, M. (1991). A BamHI polymorphism at the fibrillin (FBN) locus. Nucl. Acids. Res., 19(15), 4309.

Cooke, R. M., Wilkinson, A.J., Baron, M., Pastore, A., Tappin, M.J., Campbell, I.D., Gregory, H., and Sheard, B. (1987). The solution structure of human epidermal growth factor. Nature, 327, 339.

Corson, G. M., Chalberg, S.C., Dietz, H.C., Charbonneau, N.L., and Sakai, L.Y. (1993). Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5'end. Genomics, 17, 476.

Dalgleish, R., Hawkins, J.R., and Keston, M. (1987). Exclusion of the a2(I) and a1(III) collagen genes as the mutant loci in a Marfan syndrome family. J. Med. Genet., 24, 148.

Davis, C. (1990). The many faces of epidermal growth factor repeats. New Biologist, 2, 410.

Deng, C., and Capecchi, M.R. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the targeting locus. Mol. Cell. Biol., 12, 3365.

Dietz, H. C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M., Stetten, G., Meyers, D.A., and Francomano, C.A. (1991a). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature, 352, 337.

Dietz, H. C., Pyeritz, R.E., Hall, B.D., Cadle, R.G., Hamosh, A., Schwartz, J., Meyers, D.A., and Francomano, C.A. (1991b). The Marfan syndrome locus: confirmation of assignment to chromosome 15 and identification of tightly linked markers at 15q15-q21.3. Genomics, 9, 355.

Dietz, H. C., Saraiva, J.M., Pyeritz, R.E., Cutting, G.R., and Francomano, C.A. (1992a). Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome patients at cysteine residues in the EGF-like domains. Hum. Mut., 1, 366.

Dietz, H. C., Pyeritz, R.E., Puffenberger, E.G., Kendzior, R.J., Corson, G.M., Maslen, C.J., Sakai, L.Y., Francomano, C.A., Cutting, G.R. (1992b). Marfan phenotype variability in a family segregating a missense mutation in the EGF-like motif of the fibrillin gene. J. Clin. Invest., 89, 1674.

Dietz, H. C., McIntosh, I., Sakai, L.Y., Corson, G.M., Chalberg, S.C., Pyeritz, R.E., and Francomano, C.A. (1993a). Four novel FBN1 mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. Genomics, 17.

Dietz, H. C., Valle, D., Francomano, C.A., Kendzior, F.J., Pieritz, R.E., and Cutting, G.R. (1993b). The skipping of constitutive exons in vivo induced by nonsense mutation. Science, 254, 680.

Dynan, S. W. (1986). Promoters for housekeeping genes. Trends in Genetics, 2, 196.

Engel, J. (1989). EGF-like domains in extracellular matrix proteins: localized signals for growth and differentiation. FEBS Letters, 251, 1.

Evans, M. J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature, 292, 154.

Fahrenbach, W. H., Sandberg, L.B., and Cleary, E.G. (1966). Ultrastructural studies on early elastogenesis. Anat. Rec., 155, 563.

Fleischmajer, R., Contard, P., Schwartz, E., MacDonald II, E.D.,Jacobs II, LLoydstone, and Sakay, L.Y. (1991). Elastin-associated microfibrils (10nm) in a three-dimensional fibroblast culture. J. of Invest. Dermat., 4, 638.

Francomano, C. A., Streeten, E.A., Meyers, D.A., and Pyeritz, R.E. (1988). Exclusion of fibrillar procollagens as causes of Marfan syndrome. Am. J. Med. Genet., 29, 457.

Frohman, M. A., Dush, M.K., and Martin, G.R. (1988). Rapid amplification of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc Natl. Acad. Sci. U.S.A., 85, 8998.

Gibson, M. A., Kumaratilake, J.S., and Cleary, E.G. (1989). The protein components of the 12-nanometer microfribils of elastic and non-elastic tissues. J. Biol. Chem., 264, 4590.

Gibson, M. A., Sandberg, L.B., Grosso, L.E., Cleary, E.G. (1991). Complementary DNA cloning establishes microfibril-associated glycoprotein (MAPG) to be a discrete component of the elastin-associated microfibrils. J. Biol. Chem., 266, 7596.

Glesby, M. J., and Pyeritz, R.E. (1989). Association of mitral valve prolapse and systemic abnormalities of connective tissue: A phenotypic continuum. J. Am. Med. Assoc., 262, 523.

Godfrey, M., Menashe, V., Weleber, R.G., Koler, R.D., Bigley, R.H., Lovrien, E., Zonana, J., and Hollister, D.W. (1990). Cosegregation of elastin-associated microfibrillar abnormalities with the Marfan phenotype in families. Am. J. Hum. Genet., 46, 652.

Godfrey, M., Vandemark, N., Wang, M., Velinov, M., Wargowsli, D., Tsipouras, P., Han, J., Becker, J., Robertson, W., Droste, S., and Rao, V.H. (1993). Prenatal diagnosis and a donor splice site mutation in fibrillin in a family with Marfan syndrome. Am. J. Hum.Genet., 53, 472.

Gordon, J. W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A., and Ruddle, F.H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. Proc. Natl. Acad. Sci. U.S.A., 77, 7380.

Green, M. C., Sweet, H.O., and Bunker, L.E. (1976). Tight-skin, a new mutation in the mouse causing excessive growth of connective tissue and skeleton. Am. J. Pathol., 82, 493.

Greenlee, T. K., Jr, Ross, R., and Hartman, J.L. (1966). The fine structure of elastic fibers. J. Cell Biol., 30, 59.

Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G.G., and Campbell, I.D. (1990). The first EGF-like domain from human factor IX contains a high affinity calcium binding site. EMBO J, 9, 475.

Hasty, P., Ramirez-Solis, R., Krumlauf, R., and Bradley, A. (1991). Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. Nature, 350, 243.

Hewett, D. R., Lynch, J.R., Godfrey, M., and Sykes, B.C. (1991). G/A polymorphism in an intron of the fibrillin gene FBN1. Nucl. Acids Res., 19(24), 6975.

Hewett, D. R., Lynch, J.R., Smith, R., and Sykes, B.C. (1993). A novel fibrillin mutation in the Marfan syndrome which could disrupt calcium binding of the epidermal growth factor-like module. Hum. Mol. Genet., 2, 475.

Hollister, D. W., Godfrey, M., Sakai, L.Y., and Pyeritz, R.E. (1990). Marfan syndrome:immunohistologic abnormalities of the elastin-associated microfibrillar fiber system. N. Engl. J. Med., 323, 152.

Horrigan, S. K., Rich, C. B., Streeten, B.W., Li, Z., and Foster, J.A. (1992). Characterization of an associated microfibril protein through recombinant DNA techniques. J. Biol. Chem., 267(14), 10087.

Ishii, S., Xuy, Y., Stratton, R.H., Roe, B.A., Merlino, G.T., and Pastan, I. (1985a). Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. U.S.A., 82, 4920.

Ishii, S., Merlino, G.T., and Pastan, I. (1985b). Promoter region of the human Harvey ras proto-oncogene: similarity to the EGF-receptor proto-oncogene promoter. Science, 230, 1378.

Jiminez, S. A., Williams, C.J., Myers, J.C., and Bashey, R.I. (1986). Increased collagen biosynthesis and increased expression of type I and type III procollagen genes in tight skin (TSK) mouse fibroblasts. J. Biol. Chem., 261, 657.

Kagan, H. M., Vaccaro, C.A., Bronson, R.E., Tang, S.S., and Brody, J.S. (1986). Ultrastructural immunolocalization of lysyl oxidase in vascular connective tissue. J. Cell Biol., 103, 1121.

Kainulainen, K., Pulkkinen, L., Savolainen, A., Kaitila, I., Peltonen, L. (1990a). Location of chromosome 15 of the gene defect causing Marfan syndrome. N. Engl. J. Med., 323, 935.

Kainulainen, K., Savolainen, A., Palotie, A., Kaitilia, I., Rosenbloom, J., and Peltonen, L. (1990b). Marfan syndrome: exclusion of genetic linkage of five genes coding for connective tissue components in the long arm of chromsome 2. Hum. Genet., 84, 233.

Kainulainen, K., Sakai, L.Y., Child, A., Pope, M.F., Puhakka, L., Ryhanen, L., Palotie, A., Kaitila, I., and Peltonen, L. (1992). Two mutations in Marfan syndrome resulting in truncated polypeptide chains of fibrillin. Proc. Natl. Acad. Sci. U.S.A., 89, 5917.

Kainulainen, K., Karttunen, L., Puhakka, L., Sakai, L., and Peltone, L. (1994). Mutations in the fibrillin gene responsible for dominant ectopia lentis and neonatal Marfan syndrome. Nature Genet., 6, 64.

Kanzaki, T., Olofsson, A., Moren, A., Wernstedt, C., Hellman, U., Miyazono, K., Claesson-Welsh, L., Heldin, C.H. (1990). TGF-b1 binding protein: a component of the large latent complex of TGF-b1 with multiple repeat sequences. Cell, 61, 1051.

Karrer, H. E., and Cox, J. (1961). Electron microscope study of developing chick embryo aorta. J. Ultrastruct.Res., 4, 420.

Keene, D. R., Maddox, B.K., Kuo, H.J., Sakai, L.Y., and Glanville, R.W. (1991). Extraction of extendable beaded structures and their identification as fibrillin-containing extracellular matrix microfibrils. J. Histochem. Cytochem., 39, 441.

Kelley, M. R., Kidd, S., Deutsch, W.A., and Young M.W. (1987). Mutations altering the structure of epidermal growth factor-like coding sequences at the Drosophila Notch locus. Cell, 51, 539.

Kivirikko, K. I. (1993). Collagens and their abnormalities in a wide spectrum of diseases. Ann. Med., 25, 113.

Koening, M., and Kunkel, L.M. (1990). Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. J. Biol. Chem., 265(8), 4560.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. of Biol. Chem., 266, 19867.

Krieg, T., and Müller, P.K. (1977). The Marfan's syndrome. In vitro study of collagen metabolism in tissue specimens of the aorta. Exp. Cell. Biol., 45, 207.

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M., Jaenisch, R., and Berns, A. (1991). Simplified mammalian DNA isolation procedure. Nucl. Acids. Res., 19(15), 4293.

Laitinen, O., Uitto, J., Iivanainen, M., Hannuksela, M., and Kivirikko, K.I. (1968). Collagen metabolism of the skin in Marfan's syndrome. Clin. Chim. Acta, 21, 321.

Lamberg, S. I. (1978). Stimulatory effect of exogenous hyaluronic acid distinguishes cultured fibroblasts of Marfan's disease from controls. J. Invest. Dermatol., 71, 391.

Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattei, M.G., Sarfarazi, M., Tsipouras, P., Ramirez, F., and Hollister, D.W. (1991). Linkage of Marfan syndrome and a phenotypically related disorder to two fibrillin genes. Nature, 352, 330.

Li, X., Pereira, L., Zhang, H., Sanguineti, C., Ramirez, F., Bonadio, J., and Francke, U. (1994). Fibrillin genes map to regions of conserved mouse/hamster synteny on mouse chromosomes 2 and 18. Genomics, 18, 667.

Low, F. N. (1962). Microfibrils: fine filamentous components of the tissue space. Anat. Rec., 142, 131.

Lyon, M. F., and Kirby, M.C. (1993). Mouse chromosome atlas. Mouse Genome, 91, 40.

Maddox, B. K., Sakai, L.Y,, Keene, D.R., and Glanville, R.W. (1989). Connective tissue microfibrils; isolation and characterization of three large pepsin-resistant domains of fibrillin. J. Biol .Chem., 264, 21381.

Mann, J. R. (1993). Guide to techniques in mouse development: Surgical techniques in production of transgenic mice. Methods in Enzymology, 225, 782.

Maslen, C. L., Corson, G.M., Maddox, B.K., Glanville, R.W., and Sakai, L.Y. (1991). Partial sequence of a candidate gene for the Marfan syndrome. Nature, 352, 334.

McGookey-Milewicz, D., Pyeritz R.E., Crawford E.S., and Byers P.H. (1992). Marfan syndrome:defective synthesis, secretion and extracellular matrix formation of fibrillin by cultured dermal fibroblasts. J. Clin. Invest., 89, 79.

McKusick, V. (1956). Heritable Disorders of Connective Tissue. 1st ed. St. Louis: CV Mosby.

Mecham, R. P., and Hauser, J.E. (1992). The Elastic Fiber. In E. D. Hay (Eds.), Cell Biology of the Extracellular Matrix, 2nd ed. (pp. 79). Plenum Publish Co.

Medsger, T. A., Jr. (1993). Systemic sclerosis (scleroderma), localized scleroderma, eosinophilic fasciitis, and calcinosis. In Arthritis and Allied Conditions (pp. 1253). Philadelphia: Lea & Febiger.

Nakashima, Y. (1986). Reduced activity of serum β-glucuronidase in Marfan syndrome. Angiolog., 37, 576.

Ogilvie, D. J., Wordsworth, B.P., Priestley, L.M., Dalgleish, R., Schmidtke, J., Zoll, B., and Sykes, B. (1987). Segregation of all four major fibrillar collagen genes in the Marfan syndrome. Am. J. Hum. Genet., 41, 1071.

Panayotou, G., Aumailley, M., Timpl, R., and Engel, J. (1989). Domains of laminin with growth-factor activity. Cell, 56, 93.

Patel, I., Framson, P.E., Caskey, C.T., and Chinault, A.C. (1986). Fine structure of the human hypoxanthine phosphoribosyl transferase gene. Mol. Cell. Biol., 6, 393.

Pereira, L., D'Alessio, M, Ramirez, F, Lynch, JR, Sykes, , Pangilinan, T, Bonadio, J (1993). Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome. Hum. Mol. Genet., 2, 961.

Pereira, L., Levran, O., Ramirez, F., Lynch, J.R., Sykes, B., Pyeritz, R.E., and Dietz, H.C. (1994). Diagnosis of Marfan syndrome: a molecular approach for stratification of cardiovascular risk within families. N. Engl. J. Med., in press.

Potter, K., Hoffman, Y., Sakai, L.Y., Byers, P.H., Besser, T.E. and Milewicz, D. (1993). Abnormal fibrillin metabolism in bovine Marfan syndrome. Am. J. Pathol., 142(3), 803.

Pyeritz, R. E., and McKusick, V.A. (1979). The Marfan syndrome:diagnosis and management. N. Engl. J. Med., 300, 772.

Rebay, I., Fleming, R.J., Fehon, R.G., Cherbas, L., Cherbas, P., and Artavanis-Tsakonas, S. (1991). Specific EGF repeats of Notch mediate interactions with Delta and Serrate:implications for Notch as a multifunctional receptor. Cell, 67, 687.

Ren, Z. X., Brewton, R.G., and Mayne, R. (1991). An analysis by rotary shadowing of the structure of the mammalian vitreous humor and zonular apparatus. J. Struct. Biol., 106, 57.

Royce, P. M., and Danks, D.M. (1982). Normal lysyl oxidase activity in skin fibroblasts from patients with Marfan's syndrome. IRCS Med. Sci., 10, 41.

Saiki, R.K. (1989). The design and optimization of the PCR. In H.A. Erlich (Ed.), PCR Technology. New York: Stockton Press.

Sakai, L. Y., Keene, D.R., and Engvall, E. (1986). Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. J. Cell. Biol., 103, 2499.

Sakai, L. Y., Keene, D.R., Glanville, R.W., and Bachinger, H.P. (1991). Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective tissue microfibrils. J. Biol. Chem., 266, 14763.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual., 2nd ed. Cold Spring Harbor Laboratory Press, U.S.A.

Serafini-Fracassini, A., Venterella, G., Field, M.J., Hinnie, J., Onyezili, and Griffiths, R. (1981). Characterization of a structural glycoprotein from bovine ligamentum nuchae exhibiting dual amine oxidase activity. Biochemistry, 20, 5424.

Serafini-Fracassini, A. (1984). Elastogenesis in embryonic and post-natal development. In A. Ruggeri and Motta, P.M. (Eds.), Ultrastructure of the Connective Tissue Matrix (pp. 140). Nijhoff: The Hague.

Shores, J., Berger, K.R., Murphy, E.A., and Pyeritz, R.E. Progression of aortic root dilatation and the benefit of long-term b-adrenergic blockade in the Marfan syndrome. N.Engl. J. Med., in press.

Siracusa, L. D., Christner, P., McGrath, R., Mowers, S.D., Nelson, K.K., and Jimenez, S.A. (1993). The tight skin (Tsk) mutation in the mouse, a model for human fibrotic disease, is tightly linked to the b2-microglobulin (B2m) gene on chromosome 2. Genomics, 17, 748.

Stacey, A., Bateman, J., Choi, T., Mascara, T., Cole, W., and Jaenisch, R. (1988). Perinatal lethal osteogenesis imperfecta in transgenic mice bearing an engineered mutant pro-a1(I) collagen gene. Nature, 322(10), 131.

Starcher, B. C., and Galione, M.J. (1976). Purification and comparison of elastins from different animal species. Anal. Biochem., 74, 441.

Sykes, B. (1993). Marfan gene dissected. Nature Genet., 3(2), 99.

Tsipouras, P., Borresen, A.L., Bamforth, S., Harper, P.S., and Berg, K. (1986). Marfan Syndrome: exclusion of genetic linkage to the COL1A2 gene. Clin. Genet., 30, 428.

Tsipouras, P., Sarfarazi, M., Devi, A., Weiffenbach, B., and Boxer, M. (1991). Marfan syndrome is closely linked to a marker on 15q1.5-q2.1. Proc. Natl. Acad. Sci.U.S.A., 88, 4486.

Tsipouras, P., Del Mastro, R., Sarfarazi, M., Lee, B., Vitale, E., Child, A., Godfrey, M., Devereux, R., Hewett, D., Steinmann, B., Viljoen, D., Sykes, B.C., Kilkpatrick, M., and Ramirez, F. (1992). Linkage of Marfan syndrome, dominant ectopia lentis and congenital contractural arachnodactyl to the fibrillin genes on chromsomes 15 and 5. N. Engl. J. Med., 326, 905.

Vuorio, E., and de Crombrugghe, B. (1990). The family of collagen genes. Ann. Rev. Biochem., 59, 837.

Wyman, A.R., and White, R. (1980). A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. U.S.A., 77, 6754.

Yamada, Y., Avvedimento, V.E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I., and de Crombrugghe, B. (1980). The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54bp. Cell, 22, 887.

Yang, J. K., Masters, N., and Attardi, G. (1984). Human dihydrofolate reductase gene organization. J. Mol. Biol., 176, 169.

Yin, W., Germiller, J., Sanguineti, C., Smiley, E., Pereira, L., Ramirez, F., and Bonadio, J. (1994). Primary structure and developmental expression of *Fbn1*, the mouse fibrillin gene. Manuscript submited.

Zhang, H., Apfelroth, S.D., Hu, W., Davis, E.C., Sanguineti, C. (1994). Structure and expression of fibrillin-2, a novel microfibrillar component preferentially located in elastic matrices. J. Cell. Biol., in press.