

Original article

# A fibrillin-1-fragment containing the elastin-binding-protein GxxPG consensus sequence upregulates matrix metalloproteinase-1: biochemical and computational analysis

Patrick Booms<sup>a,1</sup>, Andreas Ney<sup>a,1</sup>, Frank Barthel<sup>b</sup>, Gautier Moroy<sup>c</sup>, Damian Counsell<sup>d</sup>,  
Christoph Gille<sup>e</sup>, Gao Guo<sup>a</sup>, Reinhard Pregla<sup>b</sup>, Stefan Mundlos<sup>a,f</sup>,  
Alain J.P. Alix<sup>c</sup>, Peter N. Robinson<sup>a,\*</sup>

<sup>a</sup> Institute of Medical Genetics, Charité University Hospital, Humboldt University, Augustenburger Platz, 13353 Berlin, Germany

<sup>b</sup> German Heart Institute, Berlin, Germany

<sup>c</sup> Laboratoire de Spectroscopies et Structures BioMoléculaires (LSSBM), IFR 53 Biomolécules, UFR Sciences Exactes et Naturelles, Université de Reims Champagne-Ardenne, Reims cedex 2, France

<sup>d</sup> Rosalind Franklin Center for Genomic Research, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

<sup>e</sup> Institute for Biochemistry, Charité University Hospital, Humboldt University, Berlin, Germany

<sup>f</sup> Max Planck Institute for Molecular Genetics, Berlin, Germany

Received 13 July 2005; received in revised form 23 November 2005; accepted 29 November 2005

## Abstract

Mutations in the gene for fibrillin-1 cause Marfan syndrome (MFS), a common hereditary disorder of connective tissue. Recent findings suggest that proteolysis, increased matrix metalloproteinase activity, and fragmentation of fibrillin-rich microfibrils in tissues of persons with MFS contribute to the complex pathogenesis of this disorder. In this study we show that a fibrillin-1 fragment containing a EGFEFG sequence that conforms to a putative GxxPG elastin-binding protein (EBP) consensus sequence upregulates the expression and production of matrix metalloproteinase (MMP)-1 by up to ninefold in a cell culture system. A mutation of the GxxPG consensus sequence site abrogated the effects. This is the first demonstration of such an effect for ligands other than elastin fragments. Molecular dynamics analysis of oligopeptides with the wildtype and mutant sequence support our biochemical results by predicting significant alterations of structural characteristics such as the potential for forming a type VIII  $\beta$ -turn that are thought to be important for binding to the EBP. These results suggest that fibrillin-1 fragments may regulate MMP-1 expression, and that the dysregulation of MMPs related to fragmentation of fibrillin might contribute to the development of MFS. Our Gene Ontology (GO) analysis of the human proteome shows that proteins with multiple GxxPG motifs are highly enriched for GO terms related to the extracellular matrix. Matrix proteins with multiple GxxPG sites include fibrillin-1, -2, and -3, elastin, fibronectin, laminin, and several tenascins and collagens. Some of these proteins have been associated with disorders involving alterations in MMP regulation, and the results of the present study suggest a potential mechanism for these observations.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Extracellular matrix; Fibrillin; Marfan syndrome; Matrix metalloproteinase; Elastin-binding protein; Molecular dynamics; Gene Ontology

## 1. Introduction

The Marfan syndrome (MFS) is a relatively common hereditary disorder of connective tissue with prominent manifestations in the cardiovascular, ocular, and skeletal systems [1,2].

The major cause of death in untreated persons with MFS is aortic dissection, which generally occurs following gradual dilatation of the aortic root over a period of many years. Mutations in the gene for fibrillin-1, *FBNI*, were shown to be the cause of MFS in 1991 [3]. However, the pathogenetic mechanisms leading from a *FBNI* mutation to clinical disease have not been completely elucidated. Several mechanisms have been proposed, ranging from a dominant negative effect of mutant monomers on the highly polymeric fibrillin-rich microfibrils [4], disturbances of tissue homeostasis [5,6], and alterations

\* Corresponding author.

E-mail address: [peter.robinson@charite.de](mailto:peter.robinson@charite.de) (P.N. Robinson).

<sup>1</sup> The first two authors contributed equally to this work.

in TGF $\beta$  signaling [7,8]. We and others have shown that *FBNI* mutations identified in MFS patients can destabilize recombinant fibrillin-1 fragments with respect to proteolysis in vitro [9–15]. In light of these results together with observations of fibrillin fragmentation in tissues of persons with MFS [16], and reports of increased levels of matrix metalloproteinases in tissues of persons with MFS [17,18], it is plausible that proteolytic degradation of microfibrils containing mutated fibrillin-1 monomers may be one component of the pathogenesis of MFS.

Extracellular matrix (ECM) proteins often have several functional roles in addition to being structure proteins. In particular, many ECM proteins can couple to cell membrane receptors and thereby trigger intracellular signaling mechanisms. The integrin family is the best-known class of cell surface molecules that mediate such signals [19]. In the present work, we have concentrated on the elastin–laminin binding protein (EBP), which is one of three protein subunits making up the elastin–laminin receptor (ELR) [20]. The EBP is a 67-kDa, enzymatically inactive alternative splice variant of  $\beta$ -galactosidase [21]. The EBP binds to tropoelastin intracellularly, blocking exposed hydrophobic VGVAPG-like domains<sup>2</sup> and thereby acting as a molecular chaperone by preventing premature self-aggregation and resultant proteolytic degradation of the highly hydrophobic tropoelastin monomers during transit through the secretory pathway [22]. The EBP–tropoelastin complex reaches the cell surface where tropoelastin is released from the EBP after the latter is engaged by galactosugars protruding from microfibrillar glycoproteins associated with the elastic-fiber scaffold [20,23]. EBP can act as a shuttle protein and is recycled intracellularly by endocytosis [24].

In addition to these functions, the EBP has been implicated in signaling processes mediated by EBP-binding of elastin's VGVAPG motif and related GxxPG motifs (where x is any amino acid). For instance, the ELR can function as a mechanotransducer in vascular smooth muscle [25,26]. EBP-transduced signals can cause proliferation of arterial smooth muscle cells [27]. VGVAPG and related hexapeptides in tropoelastin act as chemoattractants for neutrophils [28–30]. Finally, elastin-derived fragments and VGVAPG are able to upregulate MMP-2, MT1-MMP, and TIMP-2 production in HT-1080 cells [31], and xGxxPG hexapeptides including VGVAPG can upregulate MMP-1 and MMP-3 in dermal fibroblasts [32].

In the present study, we present evidence that a fibrillin-1 fragment containing a putative GxxPG EBP recognition se-

quence increases MMP-1 expression. Together with our recent finding that fibrillin-1 RGD-containing fragments can increase MMP-1 and MMP-3 production [33], our present findings add fibrillin-1 to a growing list of matrix proteins with multiple motifs that are capable of “out-to-in” signaling to the cell, such as elastin, laminin, and fibronectin. They provide a plausible contributory mechanism for microfibrillar fragmentation and increased MMP concentrations in tissues of Marfan patients. Our Gene-Ontology analysis identified multiple GxxPG motifs in numerous other ECM proteins and suggests a testable hypothesis for previously published observations of increased MMP activity in a number of disorders involving these proteins.

## 2. Material and methods

### 2.1. Fibrillin-1 constructs and in vitro mutagenesis

A recombinant construct, rFib47<sup>wt</sup>, was generated to correspond to amino acids I<sup>1929</sup> to E<sup>2205</sup> (nt 5785–6615, encoded by *FBNI* exons 47–53). This region encompasses cbEGF-like motifs 29, 30, 31, TGF $\beta$ 1bp-like (8-cys) motif 6, followed by cbEGF-like motifs 32 and 33. rFib47<sup>wt</sup> contains a putative EBP recognition motif EGFEFG at amino acids 2194–2199 (Fig. 1). Wild-type *FBNI* cDNA fragments were generated by reverse transcriptase-polymerase chain reaction (PCR) with primers designed such that the fragments contained a unique 5' *Sfi*I site and a unique 3' *Apa*I site. The primers for rFib47<sup>wt</sup> were gtatggcccagccggccATAGATGTTGATGAATGTG and gtatggcccTTCACATGTCATCATTGG (the *FBNI*-specific sequence is capitalized). RT-PCR and cloning into pSecTag2A were performed as previously described in [33].

In vitro mutagenesis was performed using the GeneTailor Site-Directed Mutagenesis system (Invitrogen Life Technologies) according to the manufacturer's instructions using the forward primer CCTGCGAGGAGGGATTTGAGTCCGGTCAATG (pos. 6572–6603), and the reverse primer CTCAAATCCCCTCTCGCAGGTGCATTCAA (pos. 6562–6591, rc). The forward (mutagenic) primer has a “T” instead of a “C” at position 6592, which changes the conserved proline residue at position 2198 to a serine residue (EGFEFG  $\rightarrow$  EGFE<sup>S</sup>FG), thereby abolishing the EBP consensus sequence. The recombinant polypeptides rFib47<sup>wt</sup> and rFib47<sup>mt</sup> were produced as previously described in [9].

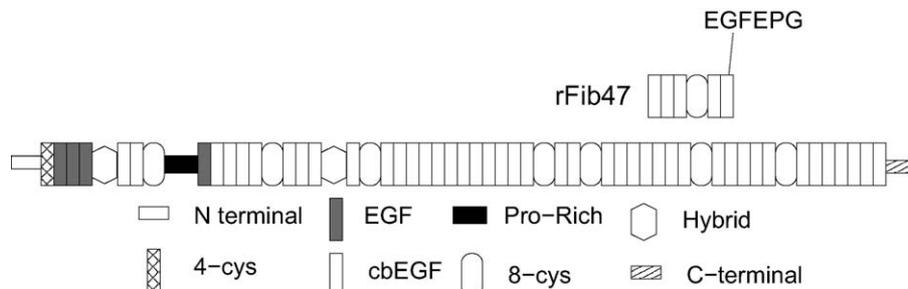


Fig. 1. Fibrillin-1 construct and location of putative EBP binding motifs in fibrillin-1. The recombinant fibrillin-1 construct used in this study rFib47<sup>wt</sup>. rFib47<sup>wt</sup> contains one of three putative EBP binding motifs that conform to the consensus sequence xGxxPG at position 2194–2199.

## 2.2. Synthetic peptides, antibodies, and reagents

Synthetic hexapeptides corresponding to the EBP binding motif of tropoelastin (VGVAPG) as well as two putative EBP binding sites in fibrillin-1, EGFEPG (positions 2194–2199) and PGFPPG (422–427), were obtained from Ansynth Service (Roosendaal, The Netherlands). Anti-MMP-1 antibody (clone41-1E5) was obtained from Oncogene Research Products. This antibody recognizes both latent and active forms of MMP-1 [34].

## 2.3. Primary cell culture and treatment of cells

Primary human skin fibroblast were cultivated from explants of human skin biopsies in Dulbecco's minimal essential medium (DMEM) supplemented with 10% fetal bovine serum and in the presence of 5% CO<sub>2</sub> at 37 °C. Cells at passage 5–10 were used for subsequent experiments. Stimulation of cells with recombinant fibrillin-1 fragments and harvesting of cell culture medium and cellular fraction for RNA extraction were done as previously described in [33].

### 2.3.1. Quantitative reverse transcriptase PCR

Total RNA isolated from fibroblasts was reverse transcribed using Superscript II with standard protocols (Invitrogen Life Technologies). For MMP-1 and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), the following primer sequences were designed using the Primer Express software (Applied Biosystems): MMP-1f (TTACACGCCAGATTTGCCAAG), MMP-1r (GTTGTCCCGATGATCTCCCCT), GAPDHf (GAAGGTGAAGGTCCGAGTC) and GAPDhr (GAA GATGGTGATGGGATTC).

The reaction and measurements were performed in real time with the GeneAmp5700 Sequence Detection system (Applied Biosystems) using the SYBR-Green Core Reagent Kit (Applied Biosystems) with the PCR thermoprofile 50 °C for 2 min, 95 °C for 10 min, then 40 cycles of 95 °C for 15 s followed by 1 min at 60 °C. For each sample, six or more reactions were analyzed in parallel. To avoid contamination, the AmpErase UNG reaction was included (while using nucleotides with dUTP). The specificity of the reaction was evaluated by performing a melting reaction (GenAmp5700). The target gene expression was normalized to the expression of GAPDH using the standard-curve method.

### 2.3.2. Western blot analysis

Typically, the cell culture medium of six parallel samples was harvested following 48 hours incubation. The conditioned medium from each well was combined with 5 × SDS-PAGE sample buffer and 80 µl was subjected to gel electrophoresis using a Hoefer SE 600 apparatus (Amersham) followed by semi-dry Western blotting. Following incubation with primary anti-MMP-1 antibody at a final concentration of 1 µg ml<sup>-1</sup> overnight at 4 °C, blots were incubated with the secondary

antibody (alkaline phosphatase-conjugated anti mouse-IgG antibodies) for 2 h at room temperature (1 mg ml<sup>-1</sup> at a dilution of 1:3000), and developed using the ECF Western blotting kit according to the manufacturer's instructions (Amersham). Imaging of the blots was performed by scanning densitometry with a FluorImager SI Densitometer (Amersham). Background-corrected quantification was performed with ImageQuant software (Amersham).

## 2.4. Statistical evaluation

Experiments were repeated six times. The data are shown as mean ± S.E.M. Differences between control means and treated groups were assessed using an unpaired one-sided Student's *t*-test.

## 2.5. Molecular dynamics simulations

For EGFEPG and EGFESG peptides, two different starting structures were used for performing molecular dynamics. The first starting structure ("T1") was a fully extended model with dihedral angle values  $\varphi$ ,  $\psi$  and  $\omega = 180^\circ$ , except for the proline for which  $\varphi = -76.1^\circ$  and  $\psi = 180^\circ$ . The second starting structure ("T2") was built by homology modeling using the NMR structure of the 32nd and 33rd calcium-binding epidermal-growth factor (cbEGF) like module of fibrillin-1 (PDB code: 1emn) [35], which contains the EGFEPG motif. All heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms were explicitly considered. Charged N-terminus and C-terminus were used. An implicit solvent model based on the solvent accessible surface area [36,37], combined with CHARMM force field [38] was employed using the parameter set PARAM19. The simulations were done with the CHARMM program version c27b3. Shift functions were used for both electrostatic and van der Waals terms, without cutoff. All simulations were performed with an integration time step of 1 fs. The starting structure was energy minimized using consecutively 500 steps of steepest descent and gradient conjugate algorithms. A heating stage of 50 ps and an equilibration stage of 80 ps were used. During the 45 ns production phase, the 300 K temperature was kept constant by weak coupling to external bath [39] with a coupling constant of 5 ps. Numerical calculations were carried on our own cluster of 10 PCs, Athlon biprocessors (16 × 1.8 GHz + 4 × 2.8 GHz, eight nodes) under Linux.

The molecular dynamics trajectories were analyzed in terms of secondary structures ( $\beta$ -turn, helix,...) as previously described in [40]. The free energy of the type VIII  $\beta$ -turn was calculated for each trajectory using the Gibbs formula:

$$\Delta G_{nontypeVIII \rightarrow typeVIII} = -k_b T \ln \left( \frac{P_{typeVIII}}{P_{nontypeVIII}} \right)$$

## 2.6. Detection and classification of $\beta$ -turn in PDB entries

A Java program ("BetaTurn") was written to extract information from PDB files with which to detect and classify  $\beta$ -turns. A  $\beta$ -turn is a type of reverse turn in the protein backbone

<sup>2</sup> Human tropoelastin has six VGVAPG sequences and a total of 28 sequences conforming to the consensus sequence GxxPG.

that allows a significant change in the direction of the polypeptide chain. The basic definition of a  $\beta$ -turn states that the distance between  $C_{\alpha}(i)$  and  $C_{\alpha}(i+3)$  is  $\leq 7 \text{ \AA}$  and that the central residues are not helical according to the secondary structure assignments in PDB.

Eight conventional  $\beta$ -turn types (I, I', II, II', VIa, VIb, VIII, and "other") are defined according to  $\psi$  and  $\varphi$  limits for residues  $i+1$  and  $i+2$ , whereby a turn is assigned to a given class if all four angles lie within  $30^\circ$  of the ideal angles [41]. A substantial proportion of  $\beta$ -turns cannot be classified by this scheme; however, since most unclassified turns represent distortions of the classic  $\beta$ -turn types, allowing one of the angles to deviate by as much as  $45^\circ$  greatly increases the number of turns that can be classified [42]. This is the nomenclature used in the present work. Type VIII turns are defined according to the ideal angles  $\psi(i+1)$ :  $-60^\circ$ ,  $\varphi(i+1)$ :  $-30^\circ$ ,  $\psi(i+2)$ :  $-120^\circ$ ,  $\varphi(i+2)$ :  $120^\circ$ .

The program BetaTurn analyses protein structures from PDB files or clean coordinate files (CCF) produced by the EMBOSS program `pdmparse` [43]. BetaTurn uses routines from STRAP [44] for PDB or CCF parsing and organizing protein structure data. Routines for calculating and angles were adapted from the EMBOSS program `psiphi` [43].

BetaTurn was used to analyze a comprehensive list of 7795 nonredundant protein chains with at most 90% sequence identity from PDB (`cullpdb_pc90_res3.0_R1.0_d050129_chains7795`) obtained from PISCES [45]. We removed all chains for which consistent parsing of the PDB file was not possible, leaving 7044 chains. This dataset is termed N-7044. Information was collected for all tetrapeptides in these chains and results were entered into a MySQL database for further analysis. Results were summarized for all possible combinations of  $AxxB$ , where A and B refer to a specific amino acid residue and x means any amino acid residue. The average proportion of tetrapeptides in each class fulfilling the definition of a  $\beta$ -turn and of a type VIII  $\beta$ -turn were recorded and a  $\chi^2$  statistic was calculated with expected frequencies defined as the corresponding frequencies over all tetrapeptides.

The distribution of  $\beta$ -turns and GxxP  $\beta$ -turns was also analyzed in a second dataset with 1332 protein chains with at most 25% sequence identity according to the methods described in [46]. This dataset is termed N-1332.

## 2.7. Ramachandran plots

The Ramachandran plot displays the  $\psi$  and  $\varphi$  backbone conformational angles for each residue in a protein. The core or allowed regions are the areas in the plot that show the preferred regions for  $\psi/\varphi$  angle pairs for residues in a protein. Presumably, if the determination of protein structure is reliable, most pairs will be in the favored regions of the plot and only a few will be in "disallowed" regions [47]. A Ramachandran plot was used to display the distribution of and angle pairs for  $\beta$ -turns conforming to the consensus sequence "GxxP" that is part of the putative EBP consensus sequence GxxPG. A Java program was developed that creates a Ramachandran plot from an arbitrary

list of  $\psi$  and  $\varphi$  angle pairs and displays them together with the preferred regions [47,48] as well as the regions for the different types of  $\beta$ -turn.

## 2.8. Gene Ontology analysis of a comprehensive nonredundant protein set for proteins containing the GxxPG consensus sequence

A comprehensive set of FASTA sequences representing the human proteome was obtained from NCBI's genome resource [49]. Perl scripts were developed to identify sequences with and without GxxPG subsequences, and the NCBI RefSeq accession numbers were mapped to UniProt identifiers [50], in order to use the Gene Ontology (GO) annotations from the GOA database [51]. A total of 26,875 proteins were identified, and GO annotations were found for 15,283 of them.

Fibrillin-1 has three occurrences of GxxPG, and by inspection, several well-known matrix proteins such as elastin and fibronectin were found to multiple copies of GxxPG sequences. Therefore, statistical analysis for overrepresentation of GO terms among proteins with at least three occurrences of GxxPG was performed according to the hypergeometric distribution [52] using version 2.0 of the Ontologizer [53]. A Bonferroni correction for multiple testing was performed by setting the significance threshold to 0.05 divided by the number of tests performed (4016 individual GO terms with more than one annotated protein in the population dataset were tested), corresponding to  $1.25 \times 10^{-5}$ .

## 2.9. Source code

The source code for BetaTurn, Ramachandran, and version 2.0 of the Ontologizer is available under the GNU Public License from <http://www.charite.de/ch/medgen/ontologizer>, <http://www.charite.de/ch/medgen/compngen/ramachandran>, and <http://www.charite.de/ch/medgen/compngen/betatum> or upon request to the authors.

## 3. Results

### 3.1. A fibrillin-1 fragment containing the EBP consensus sequence GxxPG upregulates MMP-1 expression and production

In this work we investigated the influence of a recombinant fibrillin-1 fragment containing a putative EBP binding site on the expression and production of MMP-1. Semiconfluent dermal fibroblasts from normal controls were incubated with 0, 0.1, and 0.2  $\mu\text{M}$  recombinant rFib47<sup>wt</sup>, which contains the amino acid sequence EGFEFG that conforms to the GxxPG consensus sequence for binding to the EBP [32]. After 48 h incubation, total RNA was isolated for RT-PCR and medium was isolated for determination of MMP-1 production. Quantitative RT-PCR showed a dose-dependent increase of MMP-1 expression as compared to that of GAPDH by a factor of up to 9.4

(Fig. 2A, left panel). We investigated conditioned cell culture medium by anti-MMP-1 Western blotting, demonstrating a dose-dependent increase in MMP-1 production by a factor of up to 7.0 (Fig. 2A, right panel). In separate experiments, we investigated the effect of synthetic hexapeptides corresponding to two putative EBP binding sites of fibrillin and the motif VGVAPG from tropoelastin; there was a clear increase in MMP-1 production following stimulation with each of the hexapeptides (Fig. 2B).

### 3.2. A mutation in the GxxPG site abrogates the effect of stimulation on MMP-1 expression and significantly reduces MMP-1 production

In order to demonstrate specificity of the stimulatory effect for an interaction with the EBP, we used *in vitro* mutagenesis to alter the EBP consensus sequence of fibrillin from GFEPG to GFESG. A recombinant construct was constructed to be identical to rFib47<sup>wt</sup> but to harbor this mutation (rFib47<sup>mt</sup>).

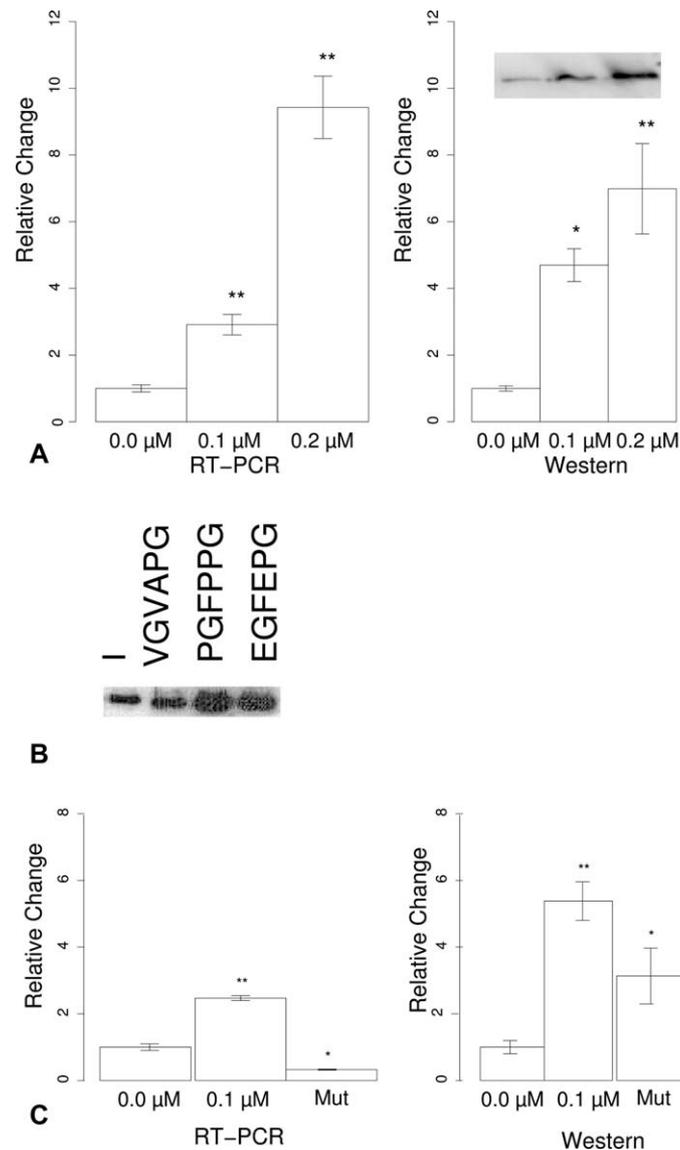


Fig. 2. Influence of a recombinant fibrillin-1 construct and synthetic hexapeptides on MMP-1 expression and production.

**A)** Influence of recombinant fibrillin-1 construct rFib47<sup>wt</sup> on expression and production of MMP-1 by human skin fibroblasts in culture. Following incubation of cultured human dermal fibroblasts for 48 h with recombinant fibrillin-1 fragment rFib47<sup>wt</sup>, culture medium was isolated for Western blotting with anti-MMP-1 antibody, and RNA was isolated for quantitative MMP-1 RT-PCR. Six parallel experiments at a concentration of 0–0.1–0.2 μM rFib47<sup>wt</sup> were performed. Left panel: Quantitative RT-PCR of MMP-1 normalized to GAPDH expression. Right panel: Quantitative Western blot analysis of MMP-1 production. The inset shows a typical blot. The error bars show 1.0 standard error.

**B)** Representative MMP-1 Western Blot following stimulation with synthetic hexapeptides at a concentration of 200 g ml<sup>-1</sup>. The first lane (“-”) shows a negative control (phosphate-buffered saline), the second VGVAPG (tropoelastin, six occurrences), the third PGFPPG (fibrillin proline-rich region, position 422–427), and the fourth EGFEPG (fibrillin, cbEGF33, position 2194–2199).

**C)** Influence of a mutation of the putative EBP binding site (EGFEPG → EGFESG). The mutant construct rFib47<sup>mt</sup> (EGFESG) was compared to the wild-type construct rFib47<sup>wt</sup> (EGFEPG) at a concentration of 0.1 μM. Left panel: Quantitative RT-PCR of MMP-1 normalized to GAPDH expression. Right panel: Quantitative Western blot analysis of MMP-1 production.

Stimulation with 0.1  $\mu$ M mutant construct was performed analogously to the above experiments. The mutated construct abolished the stimulatory effect on MMP-1 expression, and significantly reduced MMP-1 production (Fig. 2C).

### 3.3. Molecular dynamics of the EGFEPG and EGFESG Peptides

The hexapeptide Val-Gly-Val-Ala-Pro-Gly (VGVAPG) occurs multiply in elastin and is known for its chemotactic activity and ability to upregulate MMPs. Theoretical and experimental results concerning VGVAPG and related xGxxPG peptides suggested that these peptides contain a type VIII  $\beta$ -turn spanning the GxxP sequence [32,40,54,55]. No protein structure data are available for the entire tropoelastin molecule, and it is uncertain what the conformation of the VGVAPG sequences is in native elastin. We therefore performed molecular dynamics analysis of the biologically active EGFEPG and the inactive mutant peptide EGFESG to form hypotheses about potential structural correlates of our experimental observations.

Two starting structures for the molecular dynamics trajectories were used, “extended” (T1) and “homology” (T2) models (see Section 2). Table 1 shows the calculated global percentages for each type of  $\beta$ -turn defined for the three xGxx, Gxxx and xxxG sequences of EGFEPG (A) and EGFESG (B). The distribution along T1 and T2 of the different  $\beta$ -turn types are shown in Fig. 3A for EGFEPG and Fig. 3B for EGFESG.

Interestingly, the EGFEPG peptide structure is composed almost exclusively by two consecutive  $\beta$ -turns: a type II' on EGFE (T1: 73.8% and T2: 78.5%) and a type VIII on GFEP (T1: 35.1% and T2: 38.8%). According to the dihedral-angle values of the type II' and VIII, these two  $\beta$ -turns can be present among five consecutive residues [46,56]. The type II'  $\beta$ -turn on

Table 1

Global percentages of the  $\beta$ -turns observed for the peptides EGFEPG (A) and EGFESG (B). T1 and T2 correspond to the molecular dynamics starting with a fully extended model or with the homology model, respectively

A)	EGFEPG					
	Extended (T1)			Homology (T2)		
	EGFE	GFEP	FEPG	EGFE	GFEP	FEPG
I	2.7	–	–	2.9	–	–
I'	–	–	–	–	–	–
II	–	–	–	–	–	–
II'	73.8	–	–	78.5	–	–
VIII	1.9	35.1	–	1.1	38.8	–
Total	78.4	35.1	–	82.5	38.8	–

B)	EGFESG					
	Extended (T1)			Homology (T2)		
	EGFE	GFES	FESG	EGFE	GFES	FESG
I	7.9	25.8	20.8	6.6	19.4	23.3
I'	–	–	–	–	–	–
II	–	0.9	0.9	–	0.9	1.3
II'	64.7	–	2.3	45.7	2.1	2.3
VIII	1.4	22.7	6.5	2.3	15.3	8.1
Total	74.0	49.4	30.5	54.6	37.7	35.0

EGFE is stabilized by a hydrogen bond between the CO<sub>E1</sub> and the NH<sub>E4</sub> with an occurrence of 44.2% for T1 and 32.4% for T2. No  $\beta$ -turn is observed on FEPG.

The EGFESG peptide shows a quite different dynamic behavior. The type II'  $\beta$ -turn is still the major one on the EGFE sequence (T1: 64.7% and T2: 45.7%) and is also stabilized by a hydrogen bond between the CO<sub>E1</sub> and the NH<sub>E4</sub> (H-bond rate of presence: 27.9% for T1 and 34.1% for T2). But, the conformation of GFES can adopt two types of  $\beta$ -turn: a type I (T1: 25.8% and T2: 19.4%) or a type VIII (T1: 22.7% and T2: 15.3%). Surprisingly, no hydrogen bond stabilizes these  $\beta$ -turns. The FESG sequence folds also in  $\beta$ -turn (T1: 30.5% and T2: 35.3%), the highest percentage corresponds to the type I with 20.8% for T1 and 23.3% for T2. In this peptide, in contrast to the EGFEPG peptide,  $\alpha$ -helix (T1: 2.2% and T2: 1.9%) and  $3_{10}$  helix (T1: 3.9% and T2: 3.4%) are also detected. Fig. 4 displays some typical conformations of the wild-type and mutant peptides.

The free energy,  $\Delta G$ , of the type VIII  $\beta$ -turn is plotted versus time for the four trajectories in Fig. 5. The  $\Delta G$  convergence, obtained after 30 ns, indicates that one reaches a conformational equilibrium between non-type VIII  $\beta$ -turn and type VIII  $\beta$ -turn. The mean  $\Delta G$  convergence values for the four trajectories are reported in Table 2.

It is also possible to predict the main secondary structural states for the residues in both hexapeptides. The major differences are seen in the C-terminal region (EPG vs. ESG). The final predicted assignments are (where  $\alpha$  refers to an  $\alpha$ -helix, I, II', and VIII refer to  $\beta$ -turn types, and c refers to random, un-ordered coil):

Final predicted assignments					
Native peptide					
E	G	F	E	P	G
$\alpha$	II'	II' + VIII	VIII	c	c
Mutant peptide					
E	G	F	E	S	G
$\alpha$	II'	II'	I	I	c

Notably, the mutant peptide does not contain any predicted type VIII  $\beta$ -turn.

### 3.4. Structural analysis of GxxPG sequences in the protein database (PDB)

These observations prompted us to investigate whether or not GxxP sequences within protein structures deposited in the PDB have a propensity to form  $\beta$ -turns and if so, what type of  $\beta$ -turn is the most common in such sequences. We therefore collected data from a set of 7044 nonredundant protein chains from the PDB, and analyzed the distribution of turns for all possible tetrapeptides. To perform a statistical test for overrepresentation of  $\beta$ -turns in GxxP peptides, we compared all possible AxxB peptides, where A and B are fixed to one given amino acid residue at a time and x means any amino acid.

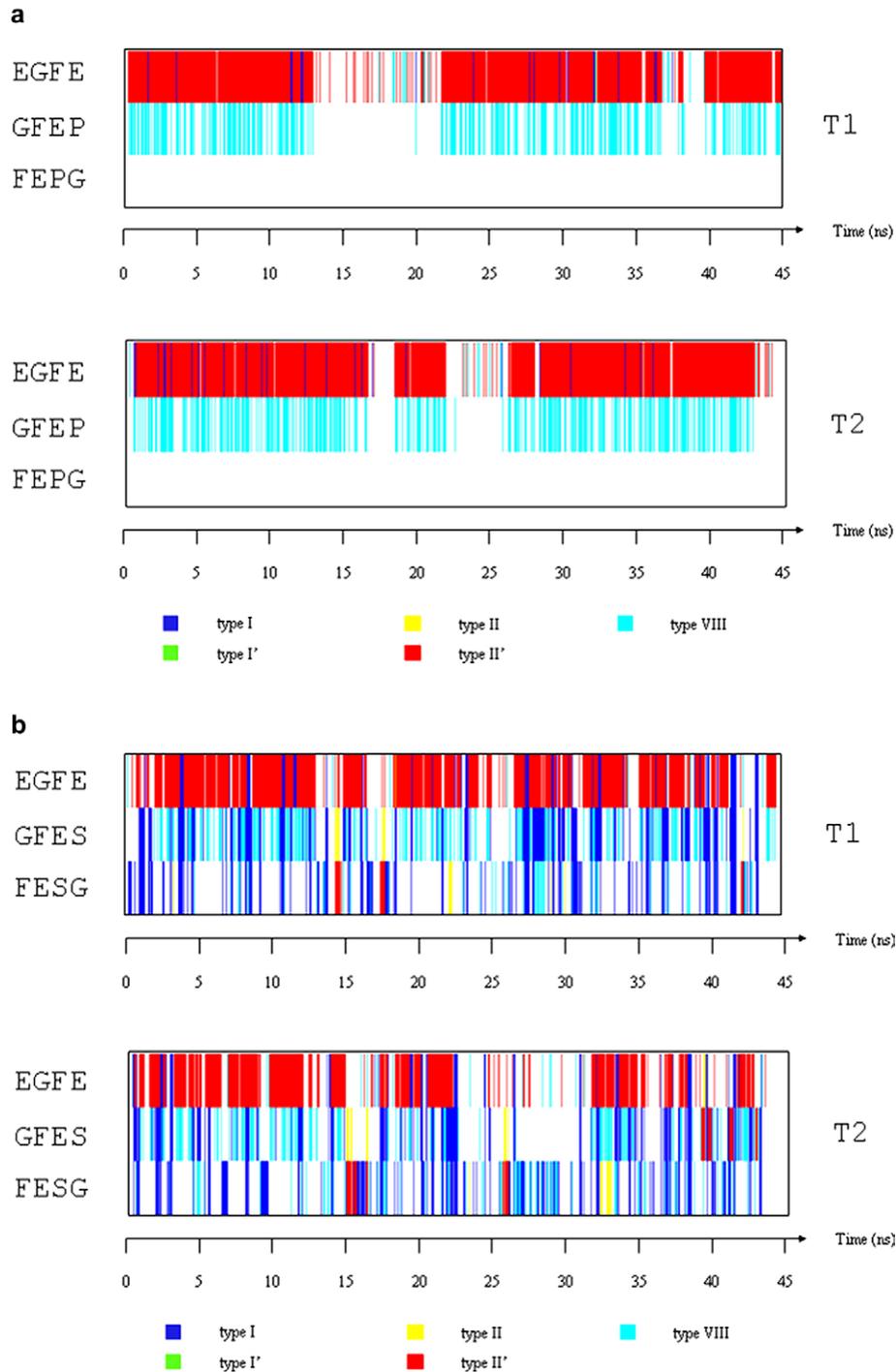


Fig. 3. Distribution of the different type  $\beta$ -turns along the molecular dynamics trajectory of the EGFEPG (A) and EGFESG (B) peptides. The starting structures are, respectively, the fully extended model (T1) and the model built by homology (T2).

A comprehensive nonredundant set of 7044 protein chains with structures deposited in the PDB with maximally 90% sequence identity was analyzed in order to characterize the distribution of  $\beta$ -turns among all possible tetrapeptides contained in the chains. Sequence and structural information could be extracted for a total of 1,795,010 residues corresponding to 1,686,526 tetrapeptides. 7.8% of all tetrapeptides fulfilled the criteria for a  $\beta$ -turn, corresponding to 23% of all residues being located within  $\beta$ -turn. 35.6% of the  $\beta$ -turns were type I, 10.5%

were type II, 10.4% were type VIII, and the remaining turns were distributed over the other types (Table 3). Individual tetrapeptides were grouped together according to AxxB consensus sequences, where A and B represent specific amino acids and x is any amino acid. There are thus  $20 \times 20$  possible AxxB motifs. Statistical analysis for overrepresentation of particular  $\beta$ -turn types among the AxxB motifs was performed using a  $\chi^2$  statistic, whereby the expected frequency of a turn type was taken to be its frequency over the entire dataset.

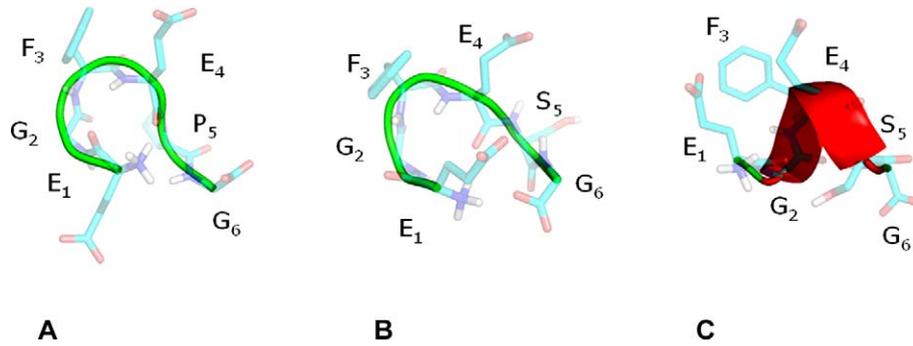


Fig. 4. Snapshots of some typical conformations of A) EGFEPG folded peptide where EGFE = type II' and GFEP = type VIII, B) EGFEFG folded peptide where EGFE = type II' and GFES = type I; a salt bridge links E<sub>1</sub> to G<sub>6</sub>; C) an EGFEFG peptide in  $3_{10}$  helical conformation.

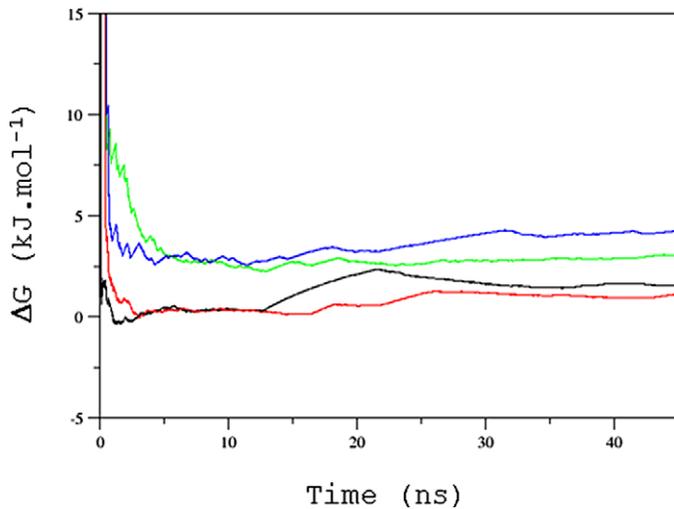


Fig. 5. Free energy versus the trajectory time: in black EGFEFG (T1), in red EGFEFG (T2), in green EGFEFG (T1) and in blue EGFEFG (T2).

Table 2  
Convergence value of  $\Delta G$  for the four trajectories of molecular dynamics

	EGFEFG		EGFEFG	
	Extended (T1)	Homology (T2)	Extended (T1)	Homology (T2)
$\Delta G$ convergence value ( $\text{kJ mol}^{-1}$ )	1.5	1.1	3.0	4.2

A total of 6357 GxxP tetrapeptides were identified in the dataset, of which 435 (6.84%) were  $\beta$ -turn. Among the GxxP tetrapeptides identified as  $\beta$ -turns, 43.2% fulfilled the definition of a type VIII  $\beta$ -turn, compared to an overall level of 10.4%

Table 3

Distribution of GxxP  $\beta$ -turns in PDB data. Results are shown for the N-7044 (90%) database, which comprises 1,686,529 residues, and N-1332 (25%), which comprises 224,369 residues

Databases $\beta$ -Turn type	N-7044 (90%)		GXXP		N-1332 (25%)		GXXP	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
I	46,738	35.6	7	1.6	6500	32.9	2	2.9
II	13,800	10.5	3	0.7	1,943	9.8	1	1.4
VIII	13,586	10.4	188	43.2	1,815	9.2	32	46.4
I'	4708	3.6	–	–	600	3.0	–	–
II'	2899	2.2	1	0.2	383	1.9	–	–
VI	852	0.6	–	–	266	1.3	–	–
Other	48,609	37.1	236	54.3	8249	41.8	34	49.3
Total	131,191	100.0	435	100.0	19,756	100.0	69	100.0

( $\chi^2_{df=2} = 506$ ,  $P \ll 0.001$ ). An additional 225 tetrapeptides conformed to the  $\psi$ - and  $\phi$ -dihedral angle criteria for a  $\beta$ -turn but have  $C_{(i)}-C_{(i+3)}$  distances in the range 7–7.5 Å. Additionally, 97 of the 236  $\beta$ -turns of type “other” were close to being type VIII turns (at least two of the angles were within  $30^\circ$  and an additional angle within  $45^\circ$  of the ideal angles for type VIII).

On the other hand, GxxP was significantly less likely than expected by chance to form a type I  $\beta$ -turn: only 1.6% of GxxP  $\beta$ -turns were type I compared to 35.6% overall ( $\chi^2_{df=2} = 219$ ,  $P \ll 0.001$ ). Table 3 shows the distribution of  $\beta$ -turn types in the N-7044 dataset as well as in the N-1332 dataset. Fig. 6 shows Ramachandran plots [47] for all GxxP  $\beta$ -turns in the N-7044 dataset.

Therefore, the motif GxxP is most often found in a non- $\beta$ -turn conformation, but when it does take on a  $\beta$ -turn conformation it is mainly in the canonical type VIII conformation or close to it. This tendency is highly statistically significant. Factors such as the neighboring residues and structures and the nature of the middle two residues may also play a role in deciding whether a given GxxP sequence takes on a  $\beta$ -turn conformation.

### 3.5. GO analysis of proteins containing multiple GxxPG sequences

A comprehensive set of 26,835 proteins from the human proteome was analyzed for the presence of GxxPG subsequences. Gene-Ontology analysis was performed to search for overrepre-

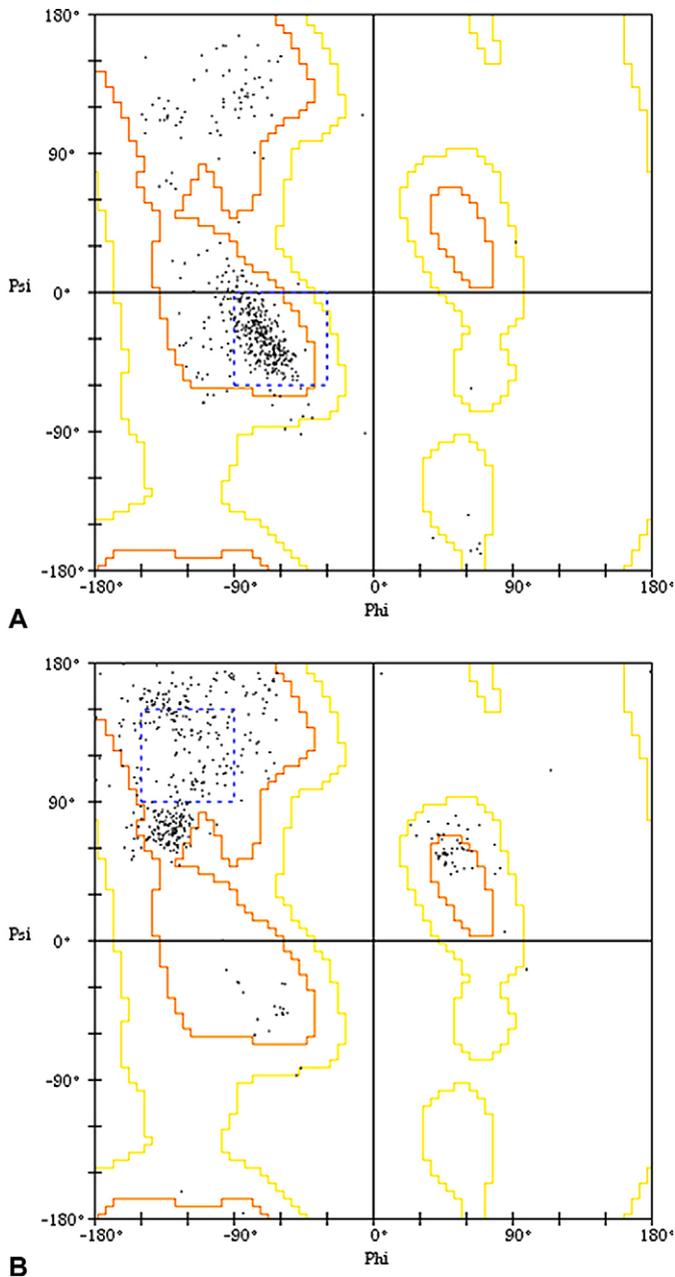


Fig. 6. Ramachandran plots of the dihedral angles  $\psi/\phi$  for the central residues  $i + 1$  and  $i + 2$  of the GxxP tetrapeptides from the N-7055 database.

A) Residue  $i + 1$ . B) Residue  $i + 2$ . The angles are seen to cluster around to core areas for a type VIII  $\beta$ -turn (dashed blue lines).

sented GO terms [52,53,57]. There was a remarkable overrepresentation of the terms *ECM (sensu Metazoa)*, *ECM*, and *ECM structural constituent*, together with the term *binding* (Table 4)

Table 4

GO analysis of proteins with  $\geq 3$  GxxPG Sequences. *ECM* and *ECM (sensu Metazoa)* are from the subontology cellular component, and *ECM structural constituent* and *binding* are from the subontology molecular function

GO term	Proteins with $\geq 3$ GxxPG	Population (%)
<i>ECM (sensu Metazoa)</i>	17 (5.63%; $1.2 \times 10^{-7}$ )	1.16%
<i>ECM</i>	17 (5.63%; $1.8 \times 10^{-7}$ )	1.20%
<i>ECM structural constituent</i>	12 (3.97%; $8.53 \times 10^{-11}$ )	0.30%
<i>Binding</i>	134 (44.4%; $4.13 \times 10^{-6}$ )	32.0%

proteins with three or more occurrences of the GxxPG consensus sequence. These results provide indirect but strong evidence of a potential biologically significant role of at least some GxxPG sequences in proteins in the ECM.

#### 4. Discussion

The molecular pathogenesis of MFS remains elusive despite over a decade of research following the initial discovery of mutations in the *FBNI* gene in 1991. Initially, it was hypothesized that mutations in fibrillin would result in a structural weakness in the fibrillin-rich microfibrils according to a dominant negative paradigm [58]. More recently, it has been shown that fibrillin-1 mutations can have other secondary effects above and beyond effects on the structural integrity of tissue. For instance, altered TGF $\beta$  activity was demonstrated to be associated with developmental abnormalities in the lung in a Marfan mouse model [7]. Recently, our laboratory has shown that fibrillin-1 fragments containing an RGD integrin binding motif upregulate expression and production of MMP-1 and MMP-3 in a cell culture system [33]. This suggested the possibility that altered matrix properties including fragmentation of tissue could be a cause of the increased MMP concentrations seen in a mouse model of MFS [5] as well as a small number of clinical studies [17,18]. In this study, we show that a GxxPG-containing fragment is able to increase MMP-1 expression and production in a cell culture system.

##### 4.1. Elastin-binding protein, VGVAPG and similar elastin-derived peptides, and the proposed GxxPG consensus sequence for EBP activity

The VGVAPG cell recognition domain(s) are accessible on the surface of growing elastic fibers [59]. In mature fibers, however, these hydrophobic sequences probably remain masked. However, after tissue injury and release of potent elastolytic enzymes by leukocytes, these domains could be unmasked or cleaved off the elastin polymer so that they could bind to the EBP of the adjacent cells [32].

Some elastin-derived peptides (EDPs) other than VGVAPG have also been shown to have biological effects related to EBP binding, including PGAIPG [30] and GVLPG and GVAPG [60]. A laminin-derived hexapeptide sequence, LGTIPG, has chemotactic effects comparable to those of VGVAPG [61]. Furthermore, evidence was presented from experiments on synthetic oligopeptides that GxxPG represents the core recognition sequence for the EBP [32]. Each of the three fibrillins contain multiple sequences conforming to the GxxPG that thus represent candidates for EBP interaction motifs.

In the present study, we have shown that one of these motifs, located in the 33rd cbEGF module of fibrillin-1, possesses stimulatory properties. An increase in MMP-1 production and expression by a factor of up to seven was demonstrated by quantitative RT-PCR and Western analysis. We demonstrated specificity by mutating one of the residues of the EBP consensus sequence (i.e. xGxxPG  $\rightarrow$  xGxxSG). We note that the mu-

tation  $xGxxPG \rightarrow xGxxSG$  was designed to alter the putative EBP consensus sequence and is not a mutation associated with MFS. This mutation abolished the stimulatory activity of the fragment as measured by RT-PCR and significantly reduced the production of MMP-1 as measured by Western blotting. To the best of our knowledge, no previously published experiments have directly shown a role for GxxPG sequences in larger polypeptides other than elastin.

Our findings provide evidence that distinct motifs in fibrillin-1 are capable of triggering increases in MMP expression and production, and suggest a plausible molecular explanation for the fragmentation of tissue and upregulation on MMPs observed in clinical studies and animal models [5,16–18]. Our results provide the basis for a testable hypothesis that mutations in fibrillin-1 may be associated with deleterious secondary effects that go beyond a mere structural weakness. This would have direct relevance for the development of novel therapeutic strategies for MFS. In addition to an influence on MMP activity, another important effect of increased levels of proteases might be the release of matrix-bound growth factors [62] and the cleavage and activation of growth factors such as TGF- $\beta$  [63].

A number of ECM proteins are known to have multiple binding motifs that can trigger signals to the cells causing a variety of responses such as increased production of MMPs or chemotaxis. For instance, distinct fibronectin fragments can increase MMP expression by binding interactions including but not limited to RGD-mediated interaction with integrins (reviewed in [64]). Distinct laminin fragments can increase MMP expression by syndecan binding [65] as well as by integrin binding [66]. There is evidence for both fibronectin and laminin that protein fragments may acquire novel signaling capabilities that the corresponding intact proteins do not possess [64,67]. It appears plausible that fragments of matrix proteins with MMP-stimulatory activity could trigger a sort of vicious cycle leading to increased MMP concentrations with increased matrix fragmentation [9,64]. In light of evidence that TGF- $\beta$  can itself increase MMP activity in some circumstances (see, for instance, [68]), it is plausible that increased protease-mediated TGF- $\beta$  release could represent an additional contributory factor to the vicious cycle.

#### 4.2. Molecular dynamics analysis supports a structural distortion due to the $GxxPG \rightarrow GxxSG$ mutation

We used a molecular dynamics approach to search for potential molecular reasons for the lack of biological activity of the EGFESG mutant. Our analysis showed significant differences between the two peptides. The observation that GxxP tetrapeptides tend to form  $\beta$ -turns has been pointed out in several publications [40,46,54,55]. The type VIII  $\beta$ -turn is the favored conformation for the GxxP motif. The presence of a type VIII  $\beta$ -turn appears to predict EDP biological activity through EBP binding [40]. Our analysis shows that the EGFEPG peptide folds only into type II'  $\beta$ -turn on EGFE and into type VIII  $\beta$ -turn on GFEP (Fig. 3A).

The EGFESG peptide folds also into type II'  $\beta$ -turn in the EGFE sequence. In contrast to the EGFEPG peptide, the GFES motif adopts two folded conformations: type I (Fig. 3B) and type VIII  $\beta$ -turn. The only difference between type I and VIII  $\beta$ -turns is the range for the dihedral angles of the second central residue (here the E residue)  $\varphi = -90^\circ$  and  $\psi = 0^\circ$  for type I and  $\varphi = -120^\circ$  and  $\psi = 120^\circ$  (the so-called  $\beta$ -sheet zone) for type VIII. It is well known that the steric constraints of the proline residue constrain the dihedral angle values of the preceding residue to be close to the  $\beta$ -sheet zone [69]. The lack of such a proline residue in the mutant hexapeptide relaxes the steric constraints on the preceding residue. The GFES conformation is thus not restrained to only one folded conformation.

Moreover, in contrast to the FEPG sequence, which does not adopt a  $\beta$ -turn conformation, the FESG motif folds mainly into a type I  $\beta$ -turn. The direct consequence is a salt bond formation between the N and C termini (Fig. 4B). In a previous study, it was demonstrated that without a glycine residue located after the GxxP motif, an EDP has no biological activity [32]. So, the  $G_6$  residue has no structural importance for the EDPs but is required for EBP binding [40]. In the EGFESG peptide, the  $G_6$  residue seems thus to be unable to interact with the EBP. The difference in the free energy of the type VIII  $\beta$ -turn folding between EGFEPG and EGFESG peptide is relatively significant (T1:  $1.5 \text{ kJ mol}^{-1}$  and T2:  $3.1 \text{ kJ mol}^{-1}$ ), and given that the mean difference of the percentages of type VIII  $\beta$ -turn between the two peptides is only around 18%, the proportion of type VIII  $\beta$ -turn does not seem to be the only explanation of the differences in the biological activity of the two constructs.

The difference in biochemical activity between the EGFEPG and EGFESG peptides can be explained by two complementary hypotheses. First, the EGFESG peptide folds into several different conformations whereas the EGFEPG folds into a single structure. The probability that the mutant EGFESG peptide will adopt a suitable conformation for binding to the EBP is thus predicted to be drastically decreased.

The second hypothesis involves the role of the  $G_6$  residue. In the EGFEPG peptide, no  $\beta$ -turn or salt bond involving the  $G_6$  residue can be observed, meaning that this residue is free to interact with the EBP. On the other hand, in the EGFESG peptide, the  $G_6$  residue is blocked by a salt bridge and thus will not be able to interact with the EBP.

#### 4.3. Analysis of GxxP structures in PDB

The PDB offers a resource of solved protein structures that can be used to search for general principles of protein structure. In this work, we have developed a Java program to search a comprehensive and nonredundant set of 7044 protein chains with maximally 90% sequence identity for the distribution of  $\beta$ -turn types across all GxxP  $\beta$ -turns in those chains and showed a high significant tendency for GxxP  $\beta$ -turns to adopt a type VIII conformation. We interpret this result as well as the results of the molecular dynamics analysis showing that the EGFEPG  $\rightarrow$  EGFESG mutation significantly alters the pre-

dicted structural behavior of the hexapeptide as providing strong but indirect support for the notion that the EGFEPG sequence of fibrillin is biologically active due to an ability to interact with EBP. Nonetheless, several uncertainties remain.

Protein structures are not available for many matrix proteins, and only partial structures have been solved for fibrillin-1. One NMR spectroscopy-derived structure (PDB code 1emn) is available that encompasses the EGFEPG sequence in the 33rd cbEGF module [35] that is contained in rFib47<sup>wt</sup>. In this structure, the EGFEPG sequence is located at a surface of the second cbEGF module. The structure of this sequence does not fulfill the criteria for a  $\beta$ -turn, since the distance between the  $C_{\alpha}(i)$  and  $C_{\alpha}(i + 3)$  is 9.7 Å. The EGFEPG sequence is not part of a specific secondary structure motif such as a helix or a sheet in 1emn, but is essentially part of a loop. Given that the EGFEPG sequence is at the extreme C-terminal end of the structure, it is uncertain if the structure of the EGFEPG sequence is identical in the intact native fibrillin-1 molecule.

A degree of flexibility has been shown for a number of proteins including elastin [70], in which methods such as circular dichroism, NMR spectroscopy, molecular dynamics simulations have been used to show that a polypeptide chain from one of the exons coding the hydrophobic sequences can oscillate between rather extended conformations, such as polyproline II structure, and folded ones, such as  $\beta$ -turn [71]. Although the present experiments were not designed to address this question, we speculate that the EGFEPG sequence in the fibrillin-1 fragment we investigated may display conformational flexibility, taking on a  $\beta$ -turn conformation upon binding to the EBP. The EGFEPG sequence is located near to the C-terminus of the recombinant polypeptide we investigated. It is possible that the corresponding sequence is either not accessible or is less flexible in intact microfibrils. Thus, fragmentation of fibrillin and fibrillin-rich microfibrils could lead to the exposure of motifs and to changes in cellular behavior owing to signaling through integrins and the EBP and perhaps other pathways.

#### 4.4. GO analysis shows overrepresentation of ECM proteins among proteins with multiple GxxPG sites

We performed computational analysis on comprehensive sets of protein structure and sequence data to further investigate the biological relevance of GxxPG sequences in matrix proteins. GO analysis demonstrated a highly statistically significant overrepresentation of annotations to terms related to the ECM among proteins with three or more GxxPG sequences (Table 4). Table 5 gives an overview of proteins with three or more GxxPG occurrences that are annotated to ECM or ECM structural constituent. Many of the proteins listed in Table have been associated with disorders characterized by increased MMP concentration. To the best of our knowledge, the role of EBP and EBP motifs has not been investigated for any of the proteins except elastin and fibrillin-1 (present work).

Although a role of EBP-mediated signaling for matrix proteins other than elastin and fibrillin is purely speculative, we suggest that EBP may plausibly be involved in the pathogenesis

of at least some of these disorders and that research in this direction may be fruitful. For instance, Tenascin-C can increase MMP-9 expression in cell-culture [72], upregulate MMP-9 in breast-cancer cells [73], and is associated with increased MMP-2 in calcified aortic stenosis [74]. Increased MMP concentrations have been identified in several disorders associated with the collagens with multiple GxxPG sites, such as collagen VII $\alpha$ 1. Several types of epidermolysis bullosa are associated with collagen VII $\alpha$ 1 mutations. One form, severe recessive dystrophic epidermolysis bullosa, is associated with increased MMP-3 levels [75]. Since type VII $\alpha$ 1 collagen is susceptible to digestion with MMP-3, a vicious cycle similar to that proposed for fibronectin fragments [64] and fibrillin-1 fragments [9] appears possible. Aggrecan is a proteoglycan that is prominent in cartilage, and experimental [76] and clinical [77] studies have shown that levels of aggrecan in joint fluid increase in response to injury or in conditions such as rheumatoid arthritis. Since these conditions are characterized by increased levels of MMPs and other manifestations such as influx of polymorphonuclear leukocytes that could plausibly be related to signaling through the ELR, it is conceivable that aggrecan fragments bearing GxxPG sites might contribute to the pathophysiology of these disorders.

## 5. Conclusion

In summary, the present study has demonstrated that a polypeptide fragment of fibrillin-1 containing a GxxPG site increases MMP-1 expression and production in cell culture. Together with the integrin binding RGD site of fibrillin-1 [33], this is the second motif in fibrillin-1 that is capable of influencing MMP expression, placing fibrillin in a growing list of ECM proteins with multiple signaling interactions, including fibronectin and laminin. VGVAPG and related GxxPG peptides

Table 5  
ECM proteins with at least three GxxPG sequences

Name	GxxPG sequences	NCBI	UniProt
Elastin	28	NP_000492	P15502
Tenascin X	11	NP_061978	P22105
Tenascin N <sup>a</sup>	8	XP_040527	Q9UQP3
Fibronectin	7	NP_997641	P02751
KIAA0543 protein	7	XP_376720	–
Fibrillin 2	7	NP_001990	P35556
Collagen VII $\alpha$ 1	7	NP_000085	Q02388
Galectin-3	5	NP_002297	P17931
Fibrillin 3	4	NP_115823	Q75N90
Collagen III $\alpha$ 1	3	NP_000081	P02461
Collagen IV $\alpha$ 5	3	NP_000486	P29400
Collagen Va2	3	NP_000384	P05997
Collagen VIII $\alpha$ 2	3	NP_005193	P25067
PDZK3	3	NP_055837	O15018
Laminin 3	3	NP_000218	Q16787
Fibrillin 1	3	NP_000129	Q75N87
Tenascin C <sup>a</sup>	3	NP_002151	–
Aggrecan 1 <sup>a</sup>	3	NP_001126	–
Netrin-2-like	3	NP_006172	O00634
EMILIN-1	3	NP_008977	Q9Y6C2
EMILIN-2	3	NP_114437	Q9BXX0

<sup>a</sup> Not annotated in GOA, but known ECM localization.

in elastin have been shown to have a variety of biological activities mediated by the ELR, including stimulation of MMP production and increased chemotaxis. To our knowledge, this is the first study to demonstrate biological activity for GxxPG site in a large polypeptide other than for elastin, and our results add further weight to the notion that the GxxPG core sequence is required for biological activity. GO analysis demonstrated a highly significant overrepresentation of annotations to ECM among proteins with three or more GxxPG sites. This result, together with an analysis of the literature, suggests GxxPG motifs may have a broader role in the physiology and pathophysiology of a subset of matrix proteins and provides a concrete hypothesis for future studies.

### Acknowledgements

This work was supported in part by the Deutsche Forschungsgemeinschaft, Ro-2005/3. We are grateful to the Temerty Family Foundation and the Canadian Marfan Association for additional support.

### References

- [1] Pyeritz RE. The Marfan syndrome. *Annu Rev Med* 2000;51:481–510.
- [2] Robinson PN, Godfrey M. Marfan syndrome: a primer for clinicians and scientists. In: Landes Bioscience. Eurekah.com and New York: Kluwer Academic/Plenum Publishers; 2004.
- [3] Dietz HC, Cutting CR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, et al. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 1991;352:337–9.
- [4] Dietz HC, McIntosh I, Sakai LY, Corson GM, Chalberg SC, Pyeritz RE, et al. Four novel *FBN1* mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. *Genomics* 1993;17:468–75.
- [5] Pereira L, Lee SY, Gayraud B, Andrikopoulos K, Shapiro SD, Bunton T, et al. Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1. *Proc Natl Acad Sci USA* 1999;96:3819–23.
- [6] Pereira L, Andrikopoulos K, Tian J, Lee SY, Keene DR, Ono R, et al. Targeting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome. *Nat Genet* 1997;17:218–22.
- [7] Neptune ER, Frischmeyer PA, Arking DE, Myers L, Bunton TE, Gayraud B, et al. Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* 2003;33:407–11.
- [8] Ng CM, Cheng A, Myers LA, Martinez-Murillo F, Jie C, Bedja D, et al. TGF-beta-dependent pathogenesis of mitral valve prolapse in a mouse model of Marfan syndrome. *J Clin Invest* 2004;114:1586–92.
- [9] Booms P, Tietze F, Rosenberg T, Hagemeyer C, Robinson PN. Differential effect of *FBN1* mutations on *in vitro* proteolysis of recombinant fibrillin-1 fragments. *Hum Genet* 2000;107:216–24.
- [10] McGettrick AJ, Knott V, Willis A, Handford PA. Molecular effects of calcium binding mutations in Marfan syndrome depend on domain context. *Hum Mol Genet* 2000;9:1987–94.
- [11] Reinhardt DP, Ono RN, Notbohm H, Muller PK, Bachinger HP, Sakai LY. Mutations in calcium-binding epidermal growth factor modules render fibrillin-1 susceptible to proteolysis. A potential disease-causing mechanism in Marfan syndrome. *J Biol Chem* 2000;275:12339–45.
- [12] Robinson PN, Booms P. The molecular pathogenesis of the Marfan syndrome. *Cell Mol Life Sci* 2001;58:1698–707.
- [13] Whiteman P, Smallridge RS, Knott V, Cordle JJ, Downing AK, Handford PAA. G1127S change in calcium-binding epidermal growth factor-like domain 13 of human fibrillin-1 causes short range conformational effects. *J Biol Chem* 2001;276:17156–62.
- [14] Suk JY, Jensen S, McGettrick A, Willis AC, Whiteman P, Redfield C, et al. Structural consequences of cysteine substitutions C1977Y and C1977R in calcium-binding epidermal growth factor-like domain 30 of human fibrillin-1. *J Biol Chem* 2004;279:51258–65.
- [15] Vollbrandt T, Tiedemann K, El-Hallous E, Lin G, Brinckmann J, John H, et al. Consequences of cysteine mutations in calcium-binding epidermal growth factor modules of fibrillin-1. *J Biol Chem* 2004;279:32924–31.
- [16] Fleischer KJ, Nousari HC, Anhalt GJ, Stone CD, Laschinger JC. Immunohistochemical abnormalities of fibrillin in cardiovascular tissues in Marfan's syndrome. *Ann Thorac Surg* 1997;63:1012–7.
- [17] Sachdev NH, Di Girolamo N, McCluskey PJ, Jennings AV, McGuinness R, Wakefield D, et al. Lens dislocation in Marfan syndrome: potential role of matrix metalloproteinases in fibrillin degradation. *Arch Ophthalmol* 2002;120:833–5.
- [18] Segura AM, Luna RE, Horiba K, Stetler-Stevenson WG, McAllister Jr. HA, Willerson JT, et al. Immunohistochemistry of matrix metalloproteinases and their inhibitors in thoracic aortic aneurysms and aortic valves of patients with Marfan's syndrome. *Circulation* 1998;98:II331–II337 (discussion II337–8).
- [19] Pozzi A, Zent R. Integrins: sensors of extracellular matrix and modulators of cell function. *Nephron Exp Nephrol* 2003;94:e77–e84.
- [20] Mecham RP, Hinek A, Entwistle R, Wrenn DS, Griffin GL, Senior RM. Elastin binds to a multifunctional 67-kDa peripheral membrane protein. *Biochemistry* 1989;28:3716–22.
- [21] Privitera S, Prody CA, Callahan JW, Hinek A. The 67-kDa enzymatically inactive alternatively spliced variant of  $\beta$ -galactosidase is identical to the elastin/laminin-binding protein. *J Biol Chem* 1998;273:6319–26.
- [22] Hinek A, Rabinovitch M. 67-kD elastin-binding protein is a protective “companion” of extracellular insoluble elastin and intracellular tropoelastin. *J Cell Biol* 1994;126:563–74.
- [23] Hinek A, Mecham RP, Keeley F, Rabinovitch M. Impaired elastin fiber assembly related to reduced 67-kDa elastin-binding protein in fetal lamb ductus arteriosus and in cultured aortic smooth muscle cells treated with chondroitin sulfate. *J Clin Invest* 1991;88:2083–94.
- [24] Hinek A, Keeley FW, Callahan J. Recycling of the 67-kDa elastin binding protein in arterial myocytes is imperative for secretion of tropoelastin. *Exp Cell Res* 1995;220:312–24.
- [25] Spofford CM, Chilian WM. Mechanotransduction via the elastin–laminin receptor (ELR) in resistance arteries. *J Biomech* 2003;36:645–52.
- [26] Spofford CM, Chilian WM. The elastin–laminin receptor functions as a mechanotransducer in vascular smooth muscle. *Am J Physiol Heart Circ Physiol* 2001;280:H1354–H1360.
- [27] Mochizuki S, Brassart B, Hinek A. Signaling pathways transduced through the elastin receptor facilitate proliferation of arterial smooth muscle cells. *J Biol Chem* 2002;277:44854–63.
- [28] Senior RM, Griffin GL, Mecham RP, Wrenn DS, Prasad KU, Urry DW. Val-Gly-Val-Ala-Pro-Gly, a repeating peptide in elastin, is chemotactic for fibroblasts and monocytes. *J Cell Biol* 1984;99:870–4.
- [29] Blood CH, Sasse J, Brodt P, Zetter BR. Identification of a tumor cell receptor for VGVAPG, an elastin-derived chemotactic peptide. *J Cell Biol* 1988;107:1987–93.
- [30] Grosso LE, Scott M. PGAIPG, a repeated hexapeptide of bovine and human tropoelastin, is chemotactic for neutrophils and Lewis lung carcinoma cells. *Arch Biochem Biophys* 1993;305:401–4.
- [31] Brassart B, Randoux A, Homebeck W, Emonard H. Regulation of matrix metalloproteinase-2 (gelatinase A, MMP-2), membrane-type matrix metalloproteinase-1 (MT1-MMP) and tissue inhibitor of metalloproteinases-2 (TIMP-2) expression by elastin-derived peptides in human HT-1080 fibrosarcoma cell line. *Clin Exp Metastasis* 1998;16:489–500.
- [32] Brassart B, Fuchs P, Huet E, Alix AJP, Wallach J, Tamburro AM, et al. Conformational dependence of collagenase (matrix metalloproteinase-1) up-regulation by elastin peptides in cultured fibroblasts. *J Biol Chem* 2001;276:5222–7.
- [33] Booms P, Pregla R, Ney A, Barthel F, Reinhardt DP, Pletschacher A, et al. RGD-containing fibrillin-1 fragments upregulate matrix metalloproteinase expression in cell culture: a potential factor in the pathogenesis of the Marfan syndrome. *Hum Genet* 2005;116:51–61.

- [34] Zhang J, Fujimoto N, Iwata K, Sakai T, Okada Y, Hayakawa T. A one-step sandwich enzyme immunoassay for human matrix metalloproteinase 1 (interstitial collagenase) using monoclonal antibodies. *Clin Chim Acta* 1993;219:1–14.
- [35] Downing AK, Knott V, Werner JM, Cardy CM, Campbell ID, Handford PA. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* 1996;85:597–605.
- [36] Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999;35:133–52.
- [37] Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel beta-sheet peptide. *Proc Natl Acad Sci USA* 2000;97:10780–5.
- [38] Brooks BR, Brucoleri BD, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J Comp Chem* 1983;4:187–217.
- [39] Berendsen HJC, Postma JPM, van Gusteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–90.
- [40] Moroy G, Alix AJP, Hery-Huynh S. Structural characterization of human elastin derived peptides containing the GXXP sequence. *Biopolymers* 2005;78:206–20.
- [41] Wilmot CM, Thornton JM. Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol* 1988;203:221–32.
- [42] Wilmot CM, Thornton JM. Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng* 1990;3:479–93.
- [43] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;16:276–7.
- [44] Gille C, Frommel C. STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* 2001;17:377–8.
- [45] Wang G, Dunbrack Jr. RL. PISCES: a protein sequence culling server. *Bioinformatics* 2003;19:1589–91.
- [46] Fuchs PF, Alix AJP. High accuracy prediction of beta-turns and their types using propensities and multiple alignments. *Proteins* 2005;59:828–39.
- [47] Kleywegt GJ, Jones TA. Phi/psi-chology: Ramachandran revisited. *Structure* 1996;4:1395–400.
- [48] Lovell SC, Davis IW, Arendall 3rd WB, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Ca geometry:  $\phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins* 2003;50:437–50.
- [49] Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2005;33:D39–D45.
- [50] Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 2004;32:D115–D119.
- [51] Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, et al. The Gene Ontology Annotation (GOA) Database: sharing knowledge in UniProt with Gene Ontology. *Nucleic Acids Res* 2004;32:D262–D266.
- [52] Castillo-Davis CI, Hartl DL. GeneMerge—post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics* 2003;19:891–2.
- [53] Robinson PN, Wollstein A, Bohme U, Beattie B. Ontologizing gene-expression microarray data: characterizing clusters with Gene Ontology. *Bioinformatics* 2004;20:979–81.
- [54] Alix AJP. A turning point in the knowledge of the structure–function–activity relations of elastin. *J Soc Biol* 2001;195:181–93.
- [55] Floquet N, Hery-Huynh S, Dauchez M, Derreumaux P, Tamburro AM, Alix AJP. Structural characterization of VGVAPG, an elastin-derived peptide. *Biopolymers* 2004;76:266–80.
- [56] Hutchinson EG, Thornton JM. A revised set of potentials for  $\beta$ -turn formation in proteins. *Protein Sci* 1994;3:2207–16.
- [57] Robinson PN, Bohme U, Lopez R, Mundlos S, Nurnberg P. Gene-Ontology analysis reveals association of tissue-specific 5' CpG-island genes with development and embryogenesis. *Hum Mol Genet* 2004;13:1969–78.
- [58] Dietz HC, Valle D, Francomano CA, Kendzior Jr. RJ, Pyeritz RE, Cutting GR. The skipping of constitutive exons in vivo induced by non-sense mutations. *Science* 1993;259:680–3.
- [59] Wrenn DS, Griffin GL, Senior RM, Mechem RP. Characterization of biologically active domains on elastin: identification of a monoclonal antibody to a cell recognition site. *Biochemistry* 1986;25:5172–6.
- [60] Bisaccia F, Morelli MA, De Biasi M, Traniello S, Spisani S, Tamburro AM. Migration of monocytes in the presence of elastolytic fragments of elastin and in synthetic derivatives. Structure-activity relationships. *Int J Pept Protein Res* 1994;44:332–41.
- [61] Mechem RP, Hinek A, Griffin GL, Senior RM, Liotta LA. The elastin receptor shows structural and functional similarities to the 67-kDa tumor cell laminin receptor. *J Biol Chem* 1989;264:16652–7.
- [62] Imai K, Hiramatsu A, Fukushima D, Pierschbacher MD, Okada Y. Degradation of decorin by matrix metalloproteinases: identification of the cleavage sites, kinetic analyses and transforming growth factor-beta1 release. *Biochem J* 1997;322(Pt 3):809–14.
- [63] Yu Q, Stamenkovic I. Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF- $\beta$  and promotes tumor invasion and angiogenesis. *Genes Dev* 2000;14:163–76.
- [64] Homandberg GA. Potential regulation of cartilage metabolism in osteoarthritis by fibronectin fragments. *Front Biosci* 1999;4:D713–D730.
- [65] Utani A, Momota Y, Endo H, Kasuya Y, Beck K, Suzuki N, et al. Laminin alpha 3 LG4 module induces matrix metalloproteinase-1 through mitogen-activated protein kinase signaling. *J Biol Chem* 2003;278:34483–90.
- [66] Adair-Kirk TL, Atkinson JJ, Broekelmann TJ, Doi M, Tryggvason K, Miner JH, et al. A site on laminin  $\alpha 5$ , AQARSAASKVKVSMKF, induces inflammatory cell production of matrix metalloproteinase-9 and chemotaxis. *J Immunol* 2003;171:398–406.
- [67] Khan KM, Falcone DJ. Selective activation of MAPK(erk1/2) by laminin-1 peptide  $\alpha 1$ :Ser(2091)-Arg(2108) regulates macrophage degradative phenotype. *J Biol Chem* 2000;275:4492–8.
- [68] Munshi HG, Wu YI, Mukhopadhyay S, Ottaviano AJ, Sassano A, Koblinski JE, et al. Differential regulation of membrane type 1-matrix metalloproteinase activity by ERK 1/2- and p38 MAPK-modulated tissue inhibitor of metalloproteinases 2 expression controls transforming growth factor- $\beta 1$ -induced pericellular collagenolysis. *J Biol Chem* 2004;279:39042–50.
- [69] MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. *J Mol Biol* 1991;218:397–412.
- [70] Mithieux SM, Weiss AS. Elastin. *Adv Protein Chem* 2005;70:437–61.
- [71] Bochicchio B, Floquet N, Pepe A, Alix AJP, Tamburro AM. Dissection of human tropoelastin: solution structure, dynamics and self-assembly of the exon 5 peptide. *Chemistry (Easton)* 2004;10:3166–76.
- [72] Ilunga K, Nishiura R, Inada H, El-Karef A, Imanaka-Yoshida K, Sakakura T, et al. Co-stimulation of human breast cancer cells with transforming growth factor-beta and tenascin-C enhances matrix metalloproteinase-9 expression and cancer cell invasion. *Int J Exp Pathol* 2004;85:373–9.
- [73] Kalembei I, Inada H, Nishiura R, Imanaka-Yoshida K, Sakakura T, Yoshida T. Tenascin-C upregulates matrix metalloproteinase-9 in breast cancer cells: direct and synergistic effects with transforming growth factor  $\beta 1$ . *Int J Cancer* 2003;105:53–60.
- [74] Jian B, Jones PL, Li Q, Mohler 3rd ER, Schoen FJ, Levy RJ. Matrix metalloproteinase-2 is associated with tenascin-C in calcific aortic stenosis. *Am J Pathol* 2001;159:321–7.
- [75] Sawamura D, Sugawara T, Hashimoto I, Bruckner-Tuderman L, Fujimoto D, Okada Y, et al. Increased gene expression of matrix metalloproteinase-3 (stromelysin) in skin fibroblasts from patients with severe recessive dystrophic epidermolysis bullosa. *Biochem Biophys Res Commun* 1991;174:1003–8.
- [76] Matyas JR, Atley L, Ionescu M, Eyre DR, Poole AR. Analysis of cartilage biomarkers in the early phases of canine experimental osteoarthritis. *Arthritis Rheum* 2004;50:543–52.
- [77] Ishiguro N, Ito T, Oguchi T, Kojima T, Iwata H, Ionescu M, et al. Relationships of matrix metalloproteinases and their inhibitors to cartilage proteoglycan and collagen turnover and inflammation as revealed by analyses of synovial fluids from patients with rheumatoid arthritis. *Arthritis Rheum* 2001;44:2503–11.