

Candidate Cell and Matrix Interaction Domains on the Collagen Fibril, the Predominant Protein of Vertebrates^{*[S]}

Received for publication, November 13, 2007, and in revised form, April 11, 2008 Published, JBC Papers in Press, May 15, 2008, DOI 10.1074/jbc.M709319200

Shawn M. Sweeney,^a Joseph P. Orgel,^b Andrzej Fertala,^c Jon D. McAuliffe,^d Kevin R. Turner,^e Gloria A. Di Lullo,^e Steven Chen,^f Olga Antipova,^b Shiamalee Perumal,^b Leena Ala-Kokko,^{g,h} Antonella Forlino,^{i,j} Wayne A. Cabral,^j Aileen M. Barnes,^j Joan C. Marini,^j and James D. San Antonio^{e1}

From the ^aCardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, the ^bCenter for Synchrotron Radiation Research and Instrumentation, Department of Biological, Chemical, and Physical Sciences, Illinois Institute of Technology, Chicago, Illinois 60616, the ^cDepartment of Dermatology and Cutaneous Biology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, the ^dDepartment of Statistics, Wharton School, University of Pennsylvania, Philadelphia 19104, Pennsylvania, the ^eChicago Medical School, North Chicago, Illinois 60064, the ^gCollagen Research Unit, Biocenter and Department of Medical Biochemistry and Molecular Biology, University of Oulu, Oulu, Finland, ^hConnective Tissue Gene Tests, Allentown, Pennsylvania 18103, the ⁱDepartment of Biochemistry A. Castellani, University of Pavia, Pavia, Italy, the ^jBone and Extracellular Matrix Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, and the ^eCardeza Foundation for Hematologic Research and Department of Medicine, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Type I collagen, the predominant protein of vertebrates, polymerizes with type III and V collagens and non-collagenous molecules into large cable-like fibrils, yet how the fibril interacts with cells and other binding partners remains poorly understood. To help reveal insights into the collagen structure-function relationship, a data base was assembled including hundreds of type I collagen ligand binding sites and mutations on a two-dimensional model of the fibril. Visual examination of the distribution of functional sites, and statistical analysis of mutation distributions on the fibril suggest it is organized into two domains. The “cell interaction domain” is proposed to regulate dynamic aspects of collagen biology, including integrin-mediated cell interactions and fibril remodeling. The “matrix interaction domain” may assume a structural role, mediating collagen cross-linking, proteoglycan interactions, and tissue mineralization. Molecular modeling was used to superimpose the positions of functional sites and mutations from the two-dimensional fibril map onto a three-dimensional x-ray diffraction structure of the collagen microfibril *in situ*, indicating the existence of domains in the native fibril. Sequence searches revealed that major fibril domain elements are conserved in type I collagens through evolution and in the type II/XI collagen fibril predominant in cartilage. Moreover, the fibril domain model provides potential insights into the genotype-phenotype relationship for several classes of human connective tissue diseases, mechanisms of integrin clustering by fibrils, the polarity of fibril assembly, heterotypic fibril function, and connective tissue pathology in diabetes and aging.

^{*} This work was supported, in whole or in part, by National Institutes of Health Grants AR048544 (to A. F.), AHA 0435339Z, NIH RR08630, and NSF 0644015 (to J. P. R. O. O.), and NIH AR049604 and HL053590 (to J. S. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^[S] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental text, equations, and Fig. S1.

¹ To whom correspondence should be addressed: Thomas Jefferson University, Cardeza Foundation for Hematologic Research and Dept. of Medicine, Rm. 809 Curtis, 1015 Walnut St., Philadelphia, PA 19107-5099. Tel.: 215-955-6121; Fax: 215-923-3836; E-mail: james.sanantonio@jefferson.edu.

Type I collagen is the most abundant protein in humans and other vertebrates, comprising much of the fibrous extracellular matrix scaffold of bones, tendons, skin, and many other tissues (1–4). In general, type I collagen and its binding partners are proposed to provide mechanical strength and form to tissues. Collagenous scaffolds are laid down and remodeled by cells and are also a predominant substrate for cell interactions, migration, and differentiation. Consequently, various debilitating human diseases are associated with type I collagen mutations, including osteogenesis imperfecta (OI,² brittle bone disease), Ehlers Danlos syndrome, vascular disorders, and others (3, 5). Type I collagen is also employed in human medicine as hemostatic sponges and implants to repair wounds and in tissue engineering applications as scaffolds (6).

Type I collagen is synthesized in the endoplasmic reticulum as $\alpha 1$ and $\alpha 2$ procollagen chains, each encoded by separate genes that are translated into proteins somewhat longer than 1000 amino acid residues (3, 7). Nucleation domains on the C-terminal propeptide promote the polymerization of two $\alpha 1$ and one $\alpha 2$ chains into the procollagen triple helical monomer (Fig. 1, A and B). The triple helical domain of procollagen is composed of contiguous glycine-X-Y tripeptide repeats, with the obligate glycine in the first position as its side chain is the only one small enough to fit within the coiled-coil of the triple helix. Extracellularly, N- and C-proteinases remove the globular termini of procollagen, and every ~ 67 nm along the fiber axis, five monomers assemble in a quarter-staggered fashion to form part of the supramolecular “helix,” the microfibril. Each microfibril, the proposed subunit of the collagen fibril (Fig. 1, C–E), and its immediate microfibrillar neighbors are connected by N- and C-terminal intermolecular cross-links. The basic repeating

² The abbreviations used are: OI, osteogenesis imperfecta; PG, proteoglycan; THP, triple helical peptide; MLBR, major ligand binding region; MMP, matrix metalloproteinase; SPARC, secreted protein, acidic, and rich in cysteine; FACIT, fibril-associated collagens with interrupted triple helices.

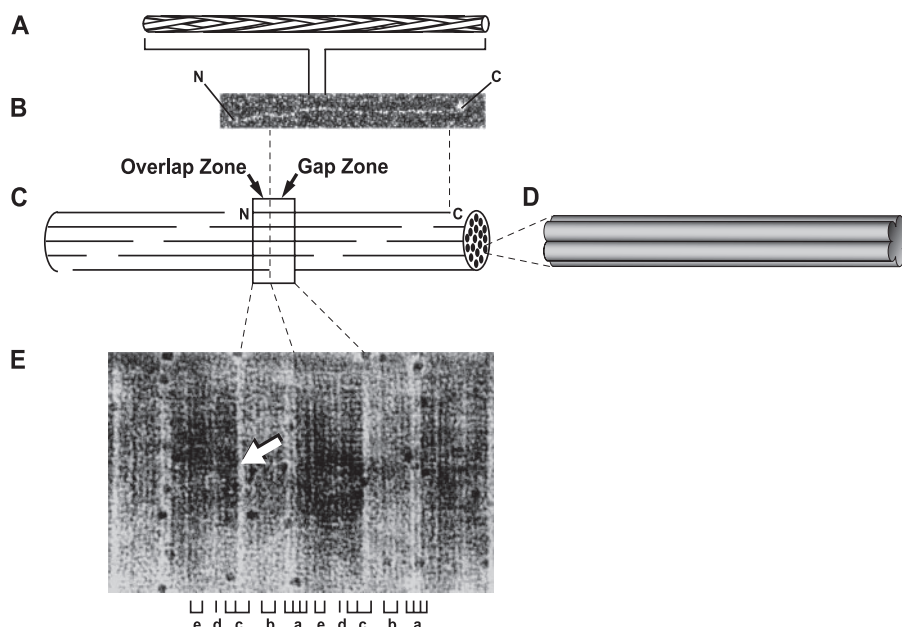


FIGURE 1. Assembly and structure of the type I collagen fibril. A fragment of a single type I collagen triple helix (monomer) is depicted (A). Type I collagen is secreted as a procollagen monomer ≈ 300 nm long, extracellularly, the propeptides (N and C) are cleaved by proteinases (dashed vertical lines, B) and the tropocollagen monomers assemble in a staggered fashion into collagen fibrils (C) where one D-period repeat (expanded two-dimensional view of 67 nm segment of microfibril, box) contains the complete collagen sequence from elements of the five monomers and includes an overlap and gap zone. The subunit structure of the fibril can be considered to be the D-periodic 5-mer referred to as the microfibril (D). Collagen fibrils appear as periodic banded structures by electron microscopy; arrow, left border of overlap zone; particles are heparin-albumin gold conjugates used to map the heparin-binding site (E). The collagen map in Fig. 2 represents an expanded view of one D-period. Modified from a figure reproduced from *J. Cell Biol.* by copyright permission of The Rockefeller University Press (53).

morphological structure of the fibril is the D-period, ~ 67 nm long, and composed of one overlap and one gap zone. Each D-period contains the complete monomer sequence derived from overlapping consecutive elements of five monomers (Fig. 1C, box). Other collagens, proteoglycans (PGs), and matrix macromolecules may further assemble with the fibril to impart tissue-specific properties to the heterotypic polymer (2, 4, 8). However, it is unclear how cells, matrix molecules, and other factors interact with the heterotypic collagen fibril. Therefore, a comprehensive structure-function model of the protein has been lacking.

To help elucidate type I collagen structure-function relationships, a data base of functional domains, ligand binding sites, and human disease-associated mutations mapping to type I collagen was previously assembled and presented as a two-dimensional map of the collagen fibril (9). It was reported that ligand-binding hot spots appear on collagen, and the positions of certain classes of mutations may correlate with disease phenotype. Subsequently the molecular, microfibrillar, and fibrillar structures of type I collagen have been determined through x-ray diffraction (10).

This study is a theoretical analysis of the interrelationships between hundreds of new functional domains, ligand-binding sites, and mutations on the collagen map. Moreover, the map is analyzed alongside of a three-dimensional model of the collagen microfibrillar structure that was recently determined (11). This analysis reveals novel insights into how the type I collagen fibril, and perhaps collagen fibrils in general, are organized to fulfill their crucial role as scaffolds for cell

interactions and as the predominant structural elements of vertebrate tissues.

EXPERIMENTAL PROCEDURES

Ligand Binding Sites and Functional Domains—Positions of binding sites and functional domains were obtained from the literature and are indicated by labeled boxes placed next to the relevant sequences. Primary literature references for sites indicated on an earlier version of the map appear in the supplemental materials. Recently reported sites are referenced in legend to Fig. 2. Zones of interactions between ligands and broad regions of collagen fibrils, as observed by electron microscopy, are indicated by colored overlays. Many binding site locations were approximated based on low resolution approaches, such as electron microscopy of macromolecular complexes. On the other hand, the use of collagen model triple helical peptides (THPs) (12, 13), in some cases, including novel peptide Toolkits composed of

overlapping, 27-residue segments of collagen sequences (14), has allowed high resolution mapping of the positions of various ligand binding sites and functional domains on the collagen monomer.

In vivo, type I collagen-ligand interactions may depend upon the tissue source. Moreover, collagen fibrils may be heterotypic, i.e. contain other collagens such as types III and V (15), whose arrangements in the fibril are incompletely understood, and that could impact fibril-ligand binding. However, despite the degree of uncertainty implicit in the map, and that some discrepancies exist in the data base (9), the approach followed here is to identify fundamental features of collagen structure and function based on numerous lines of evidence and by focusing on the best characterized functional elements of collagen.

Identifying Interrelationships between Sites—Relevant relationships between binding sites and functional domains were identified in four ways. First, as sequences on the collagen monomer to which more than one ligand has been shown to bind, or that are near neighbors; second, as sequences that fall within the borders of a fibril region shown to bind a particular ligand; and third, as neighboring binding sites on adjacent monomers within the D-periodic packing scheme. In this last case it is proposed that interactions between ligands on adjacent monomers may be possible if their binding sites align vertically within the D-period, and if their two-dimensional structure allows them to simultaneously bind more than one triple helix and reach another ligand or ligand-binding site on neighboring monomer(s). Here, sites on two or more monomers are considered as aligned vertically if they overlap with or fall close to any

axis drawn perpendicular to the long axes of the monomers, and joining monomers 1 and 5. Fourth, select data from the two-dimensional map were correlated with the known three-dimensional packing structure of collagen within the fibrillar context (10, 16).

Correlating Two-dimensional Collagen Map Data with a Three-dimensional Collagen Microfibril Model—The collagen microfibril model used in this study was composed from the packing structure of rat tendon type I collagen molecules observed *in situ* as described (10). Fiber diffraction data collected from native and derivatized (heavy atom labeled) rat tail tendons was used to solve the *in situ* structure of the collagen molecules and microfibril using multiple isomorphous replacement (16). A molecular model was constructed based on the primary sequences of the $\alpha 1$ and $\alpha 2$ chains of rat type I collagen, and the superhelical parameters were determined from collagen-like peptide crystallographic structure determinations. The electron density map representing the experimentally determined microfibril has a resolution of 0.516 nm in the direction of the fiber (collagen molecule) axis and 1.11 nm in the direction perpendicular to this. To determine the position of functional sites from the two-dimensional collagen map on the three-dimensional microfibril model, solvent-accessible surface calculation and rendering was performed using SPOCK (17) with the default probe size of 0.14 nm. Distances between surface sites were determined by line-of-sight measurement of atoms central to each site and located at the microfibril surface, assembled by using the instructions and coordinates provided by RCSB structure file 1Y0F (10). To construct images from these analyses, the $C\alpha$ "worm" traces of relevant portions of individual triple helices were marked using a semi-transparent surface rendering. The semi-transparent surface was then rendered in the appropriate colors to represent the positions of relevant functional sequences and binding sites. That rat and human collagen protein sequences are highly homologous justifies the approach of localizing functional domains of human type I collagen on the rat type I collagen microfibril.

Three-dimensional Modeling of Integrin-Collagen Interactions—Molecular modeling of $\alpha 2\beta 1$ integrin I-domain binding to the collagen triple helix was performed on a Silicon Graphics (Octane) computer system using the SYBYL software package, version 7.2 (Tripos Inc.). A model of the collagen I triple helix was created (18) with a 36-amino acid peptide spanning the $\alpha 1$ -glycine 475 to $\alpha 1$ -asparagine 510 region and a corresponding $\alpha 2$ -glycine 475 to $\alpha 2$ -alanine 510 region of human collagen IA, in which a fructopyranose residue was linked to $\alpha 2$ -lysine 479. To minimize end group effects, the amino terminus was substituted with an *N*-acetyl group and the C terminus with NHCH_3 . The model was energy-minimized using a conjugate gradient method and subject to repeating cycles of molecular dynamics using Kollman force fields and united atoms (19). To illustrate the interaction of the $\alpha 1(\text{I})$ GFPGER^{502–507} sequence with the integrin I-domain a crystal structure of a complex between the I-domain and a triple helical collagen peptide was used (Protein Data Bank, ID code 1dzi). The intermolecular energy of the interaction was analyzed with the SYBYL/Dock module to identify possible binding confor-

mations. Surface calculations of the molecules were analyzed using SYBYL/Molcad module.

Human Mutations—Mutations were obtained from the Database of Human Type I and Type III Collagen Mutations (www.le.ac.uk/genetics/collagen/), the OI consortium mutation data base (5, 20), or published elsewhere (20–27). Unpublished mutations were detected by DNA sequencing (28) and were identified in patients referred to a clinical DNA diagnostics laboratory (Reading, PA) for mutation analysis of the *COL1A1* and *COL1A2* genes. Clinical diagnosis of OI was made by the referring medical personnel.

Collagen Sequence Analysis—Sequence homologies between human fibrillar collagens were calculated using the mature collagen primary sequences and the ClustalW function of MacVector 9.0 (Accelrys). In homology determinations between two proteins the number of homologous residues divided by the number of residues in the larger protein yielded the percent homology value. The find feature of the program was used to search for functional domain sequences of interest. Accession numbers for collagen sequences examined in this study include: *Homo sapiens* *COL1A1*, NM_000088; *H. sapiens* *COL1A2*, NM_000089; *H. sapiens* *COL2A1*, BC116449; *H. sapiens* *COL3A1*, NM_000090; *H. sapiens* *COL5A1*, NM_000093; *H. sapiens* *COL5A2*, NM_0000393; *H. sapiens* *COL5A3*, NM_015719; *H. sapiens* *COL11A1*, J04177, J05407, NM_080630, NM_08062, NM_001854; *H. sapiens* *COL11A2*, NM_080681, NM_080680, NM_080679; *Bos taurus* *COL1A1*, BC105184; *B. taurus* *COL1A2*, NM_174520; *Canis familiaris* *COL1A1*, NM_001003090; *C. familiaris* *COL1A2*, NM_001003187; *Xenopus laevis* *COL1A1*, BC049829; *X. laevis* *COL1A2*, BC049287; *Mus musculus* *COL1A1*, NM_007742; *M. musculus* *COL1A2*, NM_007743; *Rattus norvegicus* *COL1A1*, XM_001081230; *R. norvegicus* *COL1A2*, NM_053356; *Danio rerio* *COL1A1*, NM_199214, and *D. rerio* *COL1A2*, NM_182968.

Statistics—Statistical analyses of mutation distributions on the collagen fibril appear under "Results" and in the supplemental materials.

RESULTS AND DISCUSSION

The collagen map (Fig. 2) consists of the protein sequences of the human $\alpha 1$ and $\alpha 2$ chains of type I collagen, arranged as they are proposed to occur in the fibril, upon which positions of structural landmarks, ligand binding sites, and disease-associated mutations are superimposed.

Ligand Binding Site Distribution—When the distribution of ligand-binding sites was examined on the map, more were found on the C-terminal half of the collagen monomer (Fig. 2), as noted previously (9). Moreover, three concentrations or "hot spots" of ligand binding are again evident, designated as major ligand binding regions (MLBRs) (*dashed boxes 1–3*, Fig. 2). The potential functional implications of some of the overlapping ligand binding sites were discussed before (9). A previously unobserved and less conspicuous hot spot appears for the first time on the revised map on and around sequence glycine-phenylalanine-hydroxyproline-glycine-glutamic acid-arginine (GFPGER)^{502–507} (Fig. 2, *orange box*), previously discovered as the predominant $\alpha 1\beta 1/\alpha 2\beta 1/\alpha 11\beta 1$ integrin and cell binding

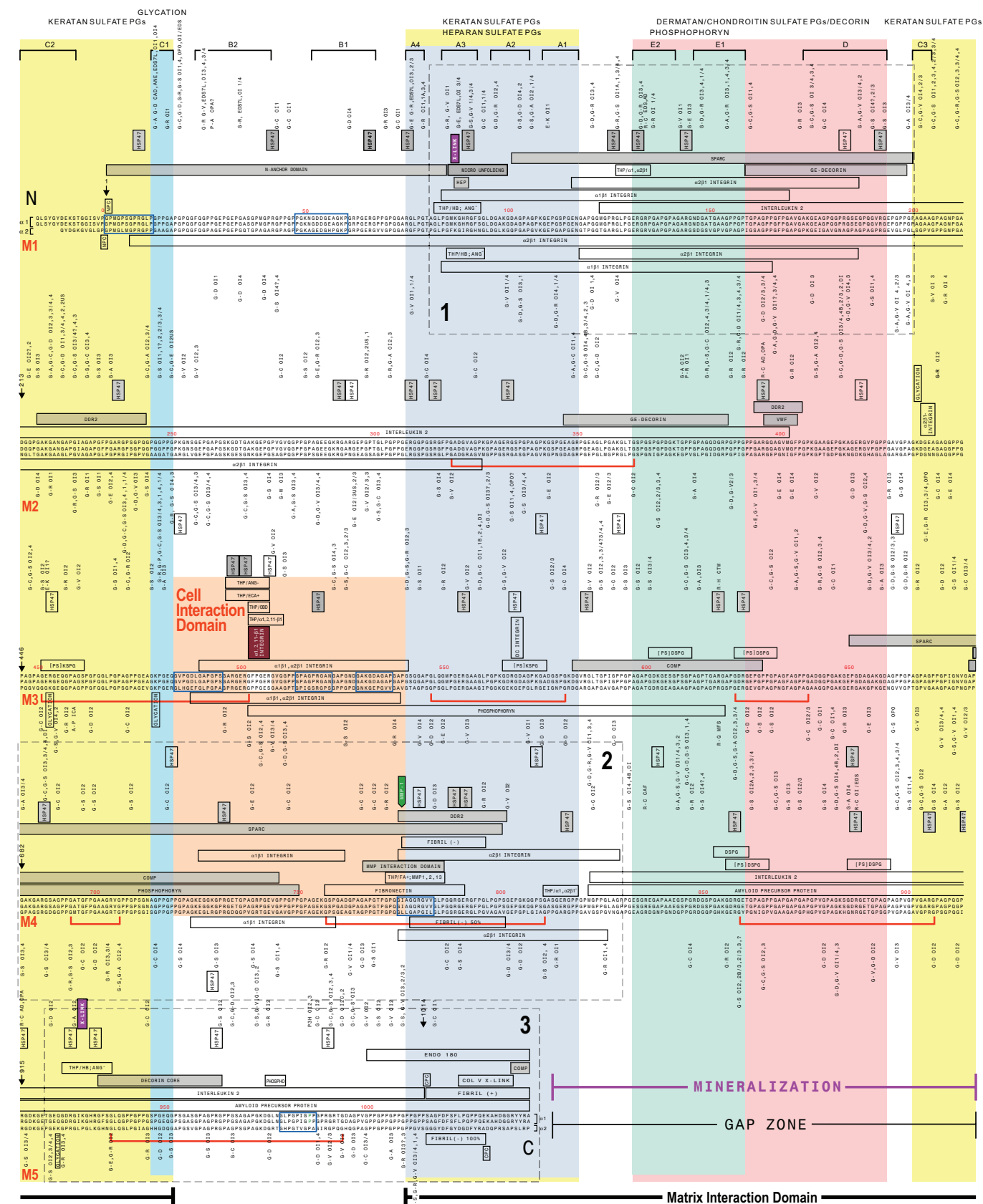


FIGURE 2. Ligand binding site and mutation map of the human type I collagen fibril. Protein sequences of the triple helix are shown (GenBankTM, $\alpha 1(I)$ accession #NP000079.2 and $\alpha 2(I)$ NP_000080.2; proline and hydroxyproline are designated as P, the latter being the third position in glycine-X-Y sequences. Note that Ala-459, $\alpha 2(I)$ results from a single nucleotide polymorphism observed in some patient samples in the non-redundant protein data base, whereas $\alpha 2(I)$ Pro-459 appears in the NP_000080.2 sequence. Yet, the association of Ala to Pro-459 substitution mutations with intracranial aneurysms (26) justifies its

$\alpha 1\beta 1$ integrin	$\alpha 1\beta 1$ integrin binding sites
$\alpha 2\beta 1$ integrin	$\alpha 2\beta 1$ integrin binding sites
$\alpha 11\beta 1$ integrin	$\alpha 11\beta 1$ integrin binding sites
AD	Mutation associated with arterial dissections
ANE	Mutation associated with aneurysm
CAD	Mutation associated with cervical artery dissections
CAF	Mutation associated with Caffey Disease (Infantile cortical hyperostosis)
Col V X-link	Collagen V cross-link site
COMP	Cartilage oligomeric matrix protein binding sites
CPC	C-Proteinase cleavage site
CTW	Mutation associated with connective tissue weakness
DC integrin	Integrin binding site on denatured collagen
DDR2	Proposed site of discoidin domain receptor 2 binding
Decorin core	Decorin core protein binding site
DI	Mutations associated with Dentinogenesis Imperfecta
[PS] DS PG	Proposed site of dermatan sulfate proteoglycan binding
EDS	Mutations associated with Ehlers-Danlos Syndrome
EDS7L	Mutations associated with Ehlers-Danlos VII-Like Syndrome
ENDO180	Collagen binding factor
Fibril (+)	C-telopeptide fibrillogenesis nucleation domain
Fibril (-)	Fibrillogenesis inhibition site
Fibril (-) 50%, 100%	50 or 100% inhibition of fibrillogenesis by peptides
G \rightarrow X	Substitution mutation
GE-decorin	Guanidine-extracted decorin binding sites
Glycation	Glycation sites
HEP	Heparin, heparan sulfate proteoglycan binding site
HSP47	Potential heat shock protein 47(chaperone) binding sites
ICA	Mutation associated with intracranial aneurysms
[PS] KS PG	Proposed sites of keratan sulfate proteoglycan binding
MFS	Mutation associated with Marfan Syndrome
MMP-1	Matrix Metalloproteinase-1 cleavage site
MMP 1, 2, 13	1, 2, 13
MMP Interaction Domain	Matrix Metalloproteinase Interaction Domain
N-ANCHOR DOMAIN	Proposed N-anchor Domain
NPC	N-Proteinase cleavage site
OI	Mutations associated with Osteogenesis imperfecta types 1, 2, 3, 4; ?/US denotes difficult diagnosis/unusual phenotype
OPA	Mutation associated with Osteopaenia; ? denotes difficult diagnosis/unusual phenotype
OPO	Mutations associated with Osteoporosis; ? denotes difficult diagnosis/unusual phenotype
P-3-H	Site of Prolyl-3-hydroxylation
PHOSPH	Phosphophoryn binding site
PHOSPHOPHORYN	
SPARC	Secreted protein acidic and rich in cysteine binding site
THP/ $\alpha 1, 2, 11\beta 1$ integrin	Triple helical peptide: binds integrin receptors
THP/ANG-	Triple helical peptide: inhibits angiogenesis
THP/ECA+	Triple helical peptide: promotes endothelial cell activation
THP/FA+	Triple helical peptide: promotes fibroblast adhesion, substrate for MMPs
THP/HB; ANG-	Triple helical peptide: binds heparin, inhibits angiogenesis
THP/OBD	Triple helical peptide: supports osteoblastic differentiation
μ Unfolding	Proposed Microunfolded domain
VWF	Proposed Site of von Willebrand Factor binding
X-LINK	Intermolecular cross-link site
Orange	Cell interaction domain
Yellow	Keratan sulfate proteoglycans binding region on the fibril
Blue	Glycation region on the fibril
Green	Phosphophoryn binding region on the fibril
Pink	Dermatan/chondroitin sulfate proteoglycans/decorin binding region on the fibril
Black Brackets	Matrix interaction domain
Red Brackets	Delineate clusters of lethal OI mutations on the $\alpha 2(I)$ chain
Blue Boxes	Regions containing 3 or more first position glycines on both chains where no human mutations have been reported.

sequence of type I collagen (29–31). Integrins are heterodimeric cell surface receptors that bind extracellular matrix molecules and are proposed to play roles in tissue morphogenesis, matrix assembly, and cell signaling (32, 33). Studies using THPs and molecular modeling approaches indicate that regions of the triple helix with high frequencies of ligand binding sites and functions do not have obvious correlations with local (at the scale of few triplets) characteristics in the conformation or stabilities of the triple helix (34). On the other hand, analysis of regional variations in the triple helix stability based on collagens with OI mutations identified two large flexible triple helical regions that aligned with zones proposed important for fibrillogenesis and ligand binding, which overlap with MLBR2 and perhaps MLBR1 (35). Putative binding sites for many of the major ligands exist on multiple monomers across the two-dimensional fibril map (Fig. 2); however, for the native collagen fibril it is not yet clear which of these sites are available for cell and ligand interactions.

Mutation Distribution—The phenotypic consequences of the most prevalent class of mutations on the map, those associated with OI, are in general proposed to arise from their disruption of collagen monomer folding or stability, not ligand-fibril interactions. OI is generally divided into four clinical types: type I (mild), II (lethal), III (severe), and IV (moderately severe) (5). The gradient model suggests that, because the helix folds in the C- to N-terminal direction, mutations near the C terminus affect collagen modification and assembly more significantly, overall resulting in a more severe phenotype (36). However, several exceptions to this pattern were observed (5); most relevant to the present study was the discovery that high concentrations of lethal OI mutations co-localize with cell and ligand-binding sites. For example, on the $\alpha 2(I)$ chain clusters of lethal OI mutations tend to coincide with putative zones of PG-fibril interactions (Fig. 2, *red brackets*) (5). New observations from the collagen map extend these findings. Thus, the N-terminal site for intermolecular cross-linking is associated with several lethal OI mutations but is bracketed by mild mutations, and sites for $\alpha 1\beta 1/\alpha 2\beta 1$ ligation at GFPGER^{502–507} and MMP-1 cleavage co-localize with mutation “silent zones” that may reflect embryonic lethality (see below) (Fig. 2). Finally, non-OI mutations also fail to follow an N- to C-terminal gradient of severity; yet cluster to several zones of the fibril D-period (supplemental Fig. S1).

Domain Model of the Collagen Fibril—Gross examination of the collagen map brings home a point previously reported, that much of the fibril is covered by PGs (2, 8), structural macromolecules considered to be full-time binding partners of collagen fibrils in many tissues (Fig. 2, *yellow, pink, and purple overlays*). PGs are composed of core proteins to which are covalently linked one or more high molecular weight, anionic glycosam-

inclusion here. Ligand binding regions are indicated by *rectangular boxes* adjacent to relevant sequences; ligand binding to type I procollagen or tropocollagen (*gray*), or to either $\alpha 1(I)$ or $\alpha 2(I)$ chains (*unshaded*). Ligand binding hot spots 1, 2, and 3 are enclosed by *dashed-lined black boxes*. *Highlighted sequences* include: the $\alpha 1\beta 1/\alpha 2\beta 1$ integrin binding site GFPGER^{502–507} (*maroon rectangle*); MMP-1 cleavage site (*green arrow*); N- and C-terminal intermolecular cross-links (*purple rectangles*). Studies responsible for mapping many of the functional domains on collagen are cited in the first map publication (9) and in the supplemental materials. New sites on the revised map include endo180 (57), HSP47 (58), micro-unfolding and N-anchor domains (59, 60), MMP-interaction domain (11), prolyl-3-hydroxylase substrate P986 (61), SPARC (62), phosphophoryn (63, 64), von Willebrand Factor (65), discoidin domain receptor 2 (66), and activities of THPs on angiogenesis (13), endothelial cell activation (67), and osteoblast differentiation (68). Mutations were from sources cited under the “Experimental Procedures.” This figure was modified from a previously published illustration (9).

inoglycan chains. PGs bound to the fibril may thus be expected to strongly influence fibril ligation of other factors competing for overlapping or neighboring binding sites. On the other hand, GFPGER^{502–507}, potentially the most critical cell interaction sequence of the fibril (see below), along with the matrix metalloproteinase (MMP) 1, 2, and 13 cleavage sequence and several other key ligand binding sites, localizes to the “b1/b2” bands region where PGs are absent (Fig. 2, *wide vertical non-shaded region*). These initial observations suggested the collagen fibril to be organized into two domains, one mediating dynamic cell-collagen interactions, and the other carrying out structural duties. Analysis of the map and microfibril structure model provided further support for this hypothesis, as detailed below.

Cell Interaction Domain—Cell-associated molecules proposed to bind type I collagen include the integrins, PGs, discoidin domain receptors, and molecules bridging cell surfaces and the extracellular matrix, such as fibronectin and SPARC (secreted protein, acidic, and rich in cysteine) (Fig. 2). However, evidence suggests integrin binding to be the most crucial determinant in cell-collagen interactions. Integrins that bind type I collagen include the $\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 11\beta 1$, and $\alpha V\beta 3$ receptors. Because the $\alpha V\beta 3$ integrin is thought to only interact with degraded collagen, it will not be further considered. Candidate $\alpha 1\beta 1/\alpha 2\beta 1/\alpha 11\beta 1$ integrin binding sequences on type I collagen have been mapped to residues 127–132, 502–507, and 811–816 (Fig. 2).

GFPGER^{502–507}: Ground Zero of Type I Collagen—Analysis of published data in the context of the map suggests a critical role for the central integrin binding site, GFPGER^{502–507} (Figs. 2 (*orange box*) and 3A), according to four lines of evidence. First, this sequence binds $\alpha 1\beta 1/\alpha 2\beta 1/\alpha 11\beta 1$ integrins and supports angiogenesis, endothelial activation, osteoblast differentiation, and cell adhesion. Second, it resides in the center of the largest “PG-clear zone” (Figs. 2 and 3A, *wide vertical non-shaded region*), ensuring its availability, and consistent with its dedicated function. Third, GFPGER^{502–507} lacks reported mutations, and also coincides with zones containing contiguous stretches of three or more first position glycines that are silent for mutations on the $\alpha 1$ and $\alpha 2$ chains (Figs. 2 and 3A, *blue boxes*), and in the fibril on M1 and M5 (Fig. 3B, *zones 2 and 7*). Such zones may be so critical to collagen function that mutations occurring there are embryonically lethal (37). These data imply that, during embryogenesis, cell-collagen interactions using GFPGER^{502–507} are crucial for fibril assembly, remodeling, or morphogenic processes such as angiogenesis. Moreover, considering the mutation density on the map, statistical analyses suggest that, although there is a high probability that the seven mutation silent zones may exist by random chance, it is very improbable they would align within such a narrow region of the D-period (see “Statistics” below), implying its special functional significance. Fourth, GFPGER^{502–507} is also a near neighbor of MLBR2, the highest concentration of ligand-binding sites of the fibril, including the site for cleavage by MMP-1, -2, and -13 required for collagen remodeling (Fig. 4A) (10). This constellation of sites also falls largely within the PG-clear zone. Analysis of the x-ray diffraction structure of the microfibril (16) confirmed these key sequences to be near neighbors, only

~3–11 nm apart on the 67 nm wide D-period (Fig. 4B). Thus, on the collagen microfibril, elements of M3 and M4 within the b1/b2 bands region, including GFPGER^{502–507}, the MMP cleavage site, and MLBR2 are proposed to comprise a cell interaction domain where cell-collagen interactions and fibril remodeling are controlled by cell surfaces (Figs. 2 (*orange box*) 4).

Two assumptions regarding the importance of GFPGER^{502–507} to fibril function deserve further comment. First, GLPGER^{127–132} and GASGER^{811–816} in type I procollagen and in THP form also bind $\alpha 1\beta 1/\alpha 2\beta 1$ integrins (38). At the level of the fibril, these sequences may overlap with PG binding zones, and at the monomer and fibril levels do not coincide with large mutation silent zones on both the $\alpha 1$ and $\alpha 2$ chains, distinguishing them from GFPGER^{502–507}. Yet, the relative functions of GFPGER^{502–507}, GLPGER^{127–132}, and GASGER^{811–816} in cell-collagen interactions must be resolved through further experimental investigation. A key consideration is whether these sites prove to be available for cell interactions in the native fibril. Second, a murine knockout of the $\alpha 2\beta 1$ integrin receptor yielded mice with a largely normal phenotype (39), which seems at odds with the proposal made here that GFPGER^{502–507} is crucial for cell-fibril interactions and embryogenesis. However, in the integrin null mouse, compensation for $\alpha 2\beta 1$ function could potentially arise from the expression of other collagen-interactive receptors, including the $\alpha 1\beta 1$ or $\alpha 11\beta 1$ integrins.

Statistics—Statistics was used to determine whether the localization of the seven mutation silent zones to a relatively narrow fibril region (Figs. 2, 3A, and 3B) is due to random chance. Thus, several properties related to the spatial distribution of first-position glycine (G_1 site) mutations in collagen were investigated. To compute the p values of interest required the derivation of some novel combinatorial quantities and algorithms. A complete presentation of these analyses appears in the supplemental materials; the following is a summary of the findings. G_1 site mutations do not occur in a spatially uniform way along the collagen monomer (p values 3.6×10^{-27} and 4.3×10^{-17}) under two reasonable models of spatial homogeneity. The adjacency patterns of mutation-free G_1 sites, however, are quite consistent with the hypothesis that they have no tendency to cluster together or remain apart (p values all ~0.5). On the other hand, several long sequences of adjacent mutation-free G_1 sites are arranged within a narrow vertical region of the D-period. The probability of their falling in this region, under a uniform random placement model, is 4.6×10^{-4} . The probability of their falling in *any* such region throughout the entire D-period is no more than 0.0025. Thus, the apparent clustering of mutation-free zones on the fibril likely did not arise randomly.

Collagen Glycation and Ligand-Collagen Interactions—Glycation is the non-enzymatic addition of sugar to protein that occurs during aging and in an accelerated fashion in diabetes (40, 41). Thus, glucose reacts with the ϵ -amino group of a free lysine residue to generate a Schiff-base intermediate, followed by its rearrangement to the more stable Amadori product. Through subsequent reactions and modifications, fructosyl-lysines may then form cross-links with other proteins, forming advanced glycation end products. Collagen thus modified may become less flexible and exhibit altered cell- and ligand-colla-

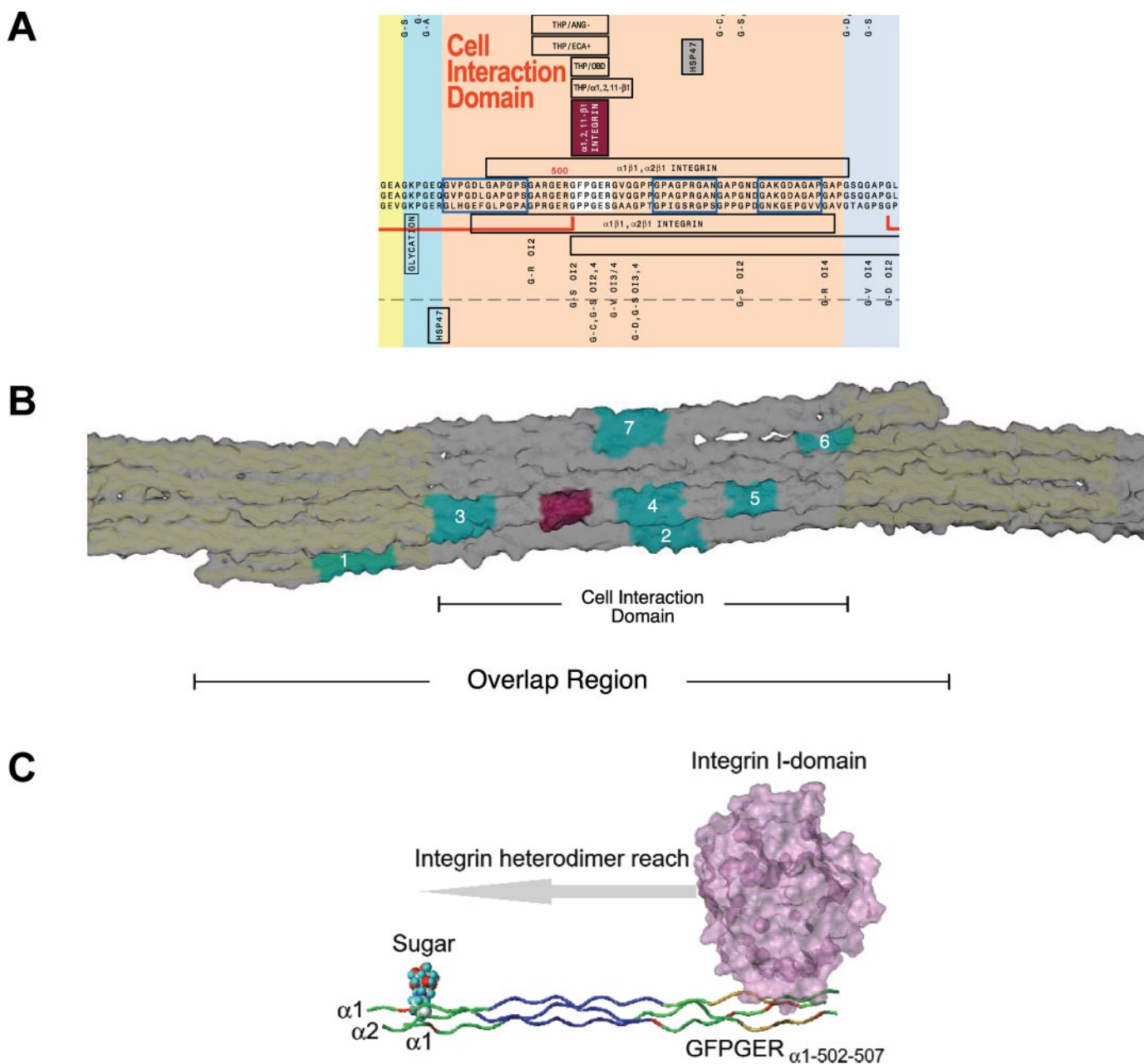


FIGURE 3. Integrin binding site of type I collagen: function and fibrillar environment. A, the $\alpha 1 \beta 1 / \alpha 2 \beta 1$ integrin binding site. GFPGER⁵⁰²⁻⁵⁰⁷ functions in angiogenesis, endothelial cell activation, and osteoblast differentiation, occupies a PG-clear zone, and coincides with sequences silent for mutations on the $\alpha 1$ and $\alpha 2$ chains of M3 and on other monomers in the fibril (blue boxes, Figs. 2 and 3B). GFPGER⁵⁰²⁻⁵⁰⁷ neighbors the major region of non-enzymatic glycation of the fibril (blue stripe) and the glycation substrate, lysine 479 (Fig. 3C). B, GFPGER⁵⁰²⁻⁵⁰⁷ and mutation silent zones localize to a narrow region of the fibril surface. GFPGER⁵⁰²⁻⁵⁰⁷ (maroon); mutation silent zones numbered sequentially from N to C termini (blue). C, glycation may influence integrin-collagen interactions. Model of $\alpha 2$ integrin subunit I-domain (purple) binding to GFPGER⁵⁰²⁻⁵⁰⁷ adjacent to $\alpha 2(I)$ fructosyl-lysine 479 (sugar). Note that glycation may potentially affect collagen ligation by the integrin heterodimer (gray arrow), but likely not the I-domain. On collagen: $\alpha 1$ GFPGER⁵⁰²⁻⁵⁰⁷ (yellow); mutation silent zone (blue); glycine residues associated with lethal OI mutations (red). Gray arrow, 10 nm is the maximal estimated reach of the integrin heterodimer.

gen interactions and fibrillar structure (41–45). Glycation occurs on numerous collagen residues but preferentially on hydroxylysines $\alpha 1(I)$ 434, and $\alpha 2(I)$ 453, 479, and 924 (Fig. 2), and the “c1” band fibril region (Figs. 2 and 3A, blue stripe). Because glycated collagen is a poor substrate for cell interactions, it is notable that lysine 479 and GFPGER⁵⁰²⁻⁵⁰⁷ are near neighbors on the two-dimensional fibril map (Fig. 2). To examine this relationship, a three-dimensional model was built of the triple helix spanning residues 475–510, with GFPGER⁵⁰²⁻⁵⁰⁷ ligated to the $\alpha 2 \beta 1$ integrin-I-domain (Fig. 3C). The lysine 479

fructosyl adduct was found to project outwards from the triple helix, poised to affect collagen-ligand interactions. Further, whereas glycation would likely not directly interfere with integrin-I-domain binding, interactions of the integrin heterodimer may potentially be affected. Examining cell and integrin interactions with native and glycated THPs, including sequences spanning the integrin binding site and lysine 479, could test such hypotheses. Lastly, it was reported that the fibril binding zone for keratan sulfate PGs, including the “c” bands, overlaps with all of the predominant glycation sites, in contrast

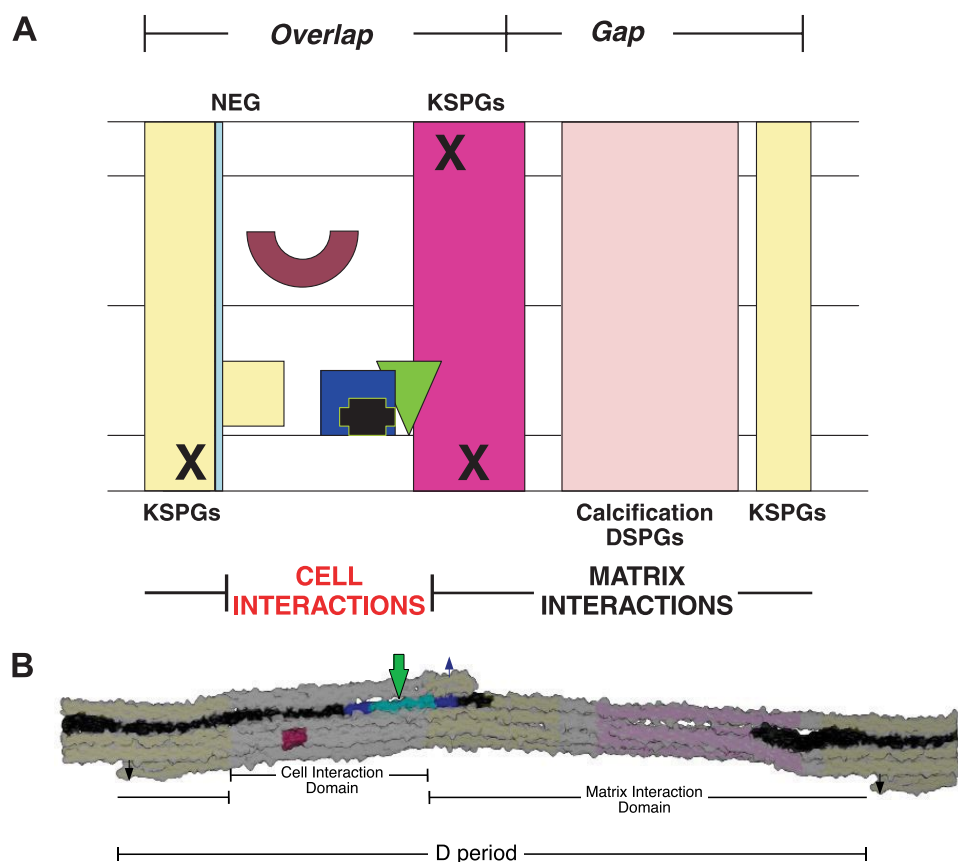


FIGURE 4. Domain model of the collagen fibril. A, schematic of domains on the collagen fibril D-period. The putative cell interaction domain, including the $\alpha 1 \beta 1 / \alpha 2 \beta 1$ integrin binding site GFPGER^{502–507} (maroon horse-shoe); MMP cleavage site (green arrow); and elements of MLBR2 (geometric figures); occupy the fibril overlap zone. The putative matrix interaction domain, including sites for intermolecular cross-linking (X); PG-binding (yellow and pink); and mineralization occupy the remainder of the fibril. B, physical mapping of the cell and matrix interaction domains on the collagen microfibril. Selected elements of the putative cell and matrix interaction domains were mapped, including GFPGER^{502–507} (maroon); MMP cleavage site (green arrowhead), MMP interaction domain (cyan); lateral edges of the fibronectin binding site (blue); keratan sulfate PG binding sites (yellow); dermatan sulfate PG binding sites (pink); N- and C-terminal intermolecular cross-links (black and blue arrows), respectively. The SPARC binding site (black) was rendered transparent where it overlaps the fibronectin and MMP binding/cleavage sites. The three-dimensional image was derived from an x-ray diffraction structure of the collagen microfibril *in situ*. The microfibrillar aspect that faces toward the outside of the fibril is at the top of the picture (the C terminus points outside of the fibril, while the N terminus faces inside the fibril). Distances between GFPGER^{502–507} and domains for MMP interaction, or the N-terminal end of the fibronectin-binding domain were 11.2 and 5.0 nm, respectively. A view of microfibril function in the context of the collagen fibril appears in Fig. 5.

to that of dermatan sulfate PGs that localize to the “d” and “e” bands (9). Consistent with this observation, it was reported that keratan sulfate PGs, but not dermatan sulfate PGs, exhibit reduced affinity for glycosaminoglycans (46). Together these observations justify further investigations into the consequences of glycation to cell- and ligand-collagen interactions.

Matrix Interaction Domain—The map suggests that, aside from the cell interaction domain, the remainder of the D-period comprises a substrate for the binding of structural molecules (Figs. 2 and 4), which is supported by various lines of evidence. Thus, dermatan sulfate PGs occupy the e and d bands regions, keratan sulfate PGs fall within the a and c bands regions, and heparin binds the a bands region. Moreover, intermolecular cross-links in type I collagen exist between residues 87 and 930, and between type I collagen at residues 1023–1036 and type V collagen in the a and c band regions. Notably, the N- and C-terminal cross-links closely bracket the cell interaction domain, potentially stabilizing its complement of key sites, or otherwise

influencing its function or availability. Next, the gap or “hole zone” and its associated proteins are proposed to contain the nucleation sites upon which hydroxyapatite crystal growth during bone mineralization occurs. Last, several domains important for collagen fibrillogenesis map to the a bands region. Thus, the overlap zone and lateral borders of the gap zone may comprise a “matrix interaction domain” where organic and inorganic matrix molecules bind to connect the fibril with other extracellular matrix elements, and to impart distinct structural properties to collagen fibrils and the tissue stroma in which they reside (Figs. 2 (black brackets) and 4).

Non-overlapping Fibril Domains—Integrins and MMPs (<5 nm wide) may be envisioned to interact with their binding sites in the cell interaction domain of ~20 nm wide and thus mainly influence ligands binding adjacent sequences. However, the impact of fibril-bound PGs to cell-collagen interactions may be more considerable. For example, the decorin core protein of ~40 kDa may contain a glycosaminoglycan chain averaging 40 kDa (47). The core protein is proposed to bind collagen monomer(s) in the d/e bands region. However, if its glycosaminoglycan chain is unrestrained, it could reach 80 nm or more, potentially disrupting cell- and ligand-collagen interactions at distant sites. Instead, it is proposed that the anionic

glycosaminoglycans are mostly constrained within the matrix interaction domain via their binding to the transverse electropositive bands on the fibril surface, adjacent to where the PG core protein binds (Figs. 1, 2, and 5). Thus, the fibril is viewed to consist of non-overlapping cell and matrix interaction domains (Figs. 4 and 5). Nonetheless, some cross-talk between domains may be likely, e.g. if glycation affects integrin-GFPGER^{502–507} ligation, or matrix proteins like SPARC with broad fibril-binding footprints influence the function of both domains simultaneously (Figs. 2 and 4).

That numerous functional sites of the type I collagen molecule fall into either of two fibril domains according to their function and that a preferred glycation substrate, lysine 479, coincides with the major fibril glycation zone support the accuracy of the overlap model of collagen fibril structure (48) upon which the map is based. Lastly, FACIT (fibril-associated collagens with interrupted triple helices) collagens (49) are proposed to interact with type I or II fibrils to influence inter-monomer

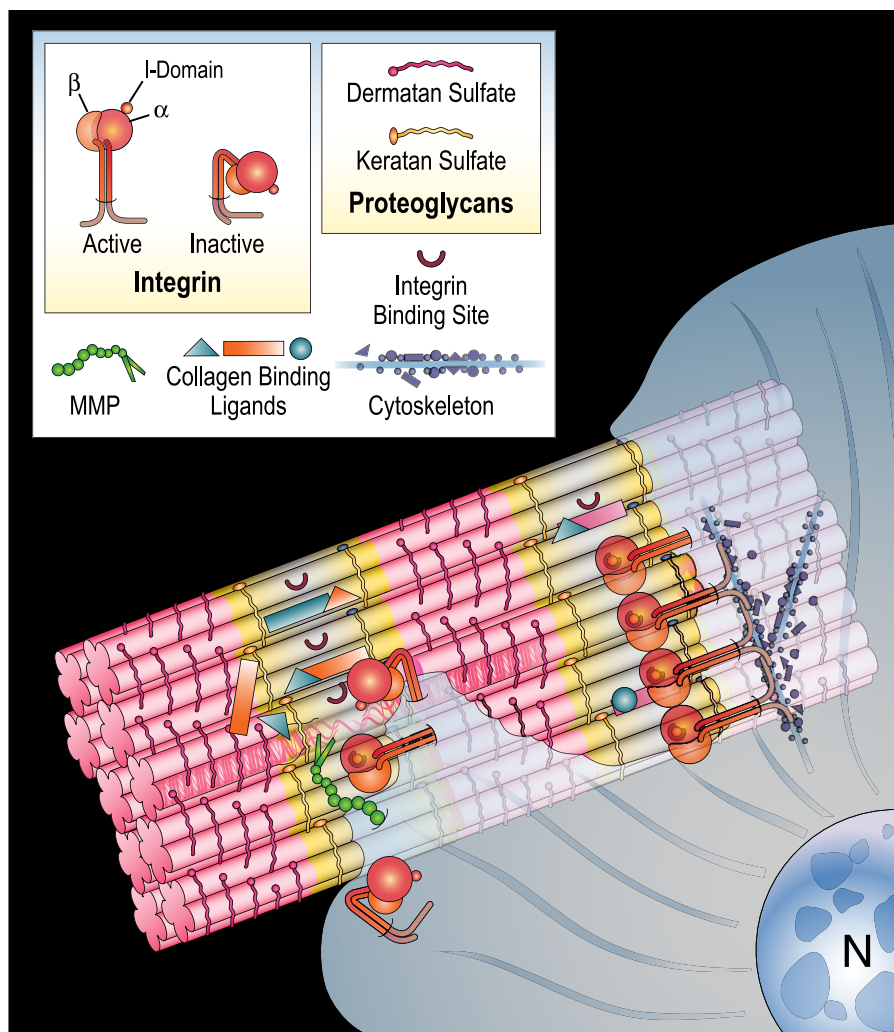


Illustration by Nanita Barchi

FIGURE 5. Domain regulation of collagen fibril function. The collagen fibril is proposed to be composed of cell interaction domains (unshaded) and matrix interaction domains (yellow and pink shading). In the cell interaction domain integrin binding sites promote cell adhesion and integrin receptor ligation, clustering, and activation. The confluence of crucial cell and ligand-binding sites on the fibril surface implies their coordinate regulation. Thus, fibril-bound integrins may modulate other ligand-collagen interactions, and MMP cleavage of one monomer may enhance or disrupt ligand-fibril interactions elsewhere. In the matrix interaction domain, PGs and other matrix molecules bind to impart key structural properties to the fibril. Note: the fibril shown in this schematic includes ten microfibrils and spans two and a half, 67 nm wide collagen D-periods; however, the average collagen fibril *in vivo* contains many more microfibrils than depicted here and may reach hundreds of microns in length (52).

associations within the fibril or inter-fibril dynamics (50). Whether FACIT collagens modulate the function of the cell and matrix interaction domains proposed here must await further understanding of the physical nature of FACIT-fibril interactions.

Domains in Other Fibrillar Collagens—The best characterized elements of the cell and matrix interaction domains of the human fibril as proposed here include sites for integrin binding, GFPGER^{502–507}; MMP-1 cleavage, GPQGIA; and N- and C-terminal intermolecular cross-linking, GMKGHR and GIKGHR, respectively (Fig. 2). These elements were searched for in other fibrillar collagens (data not shown). The sequences, and their positions in the triple helices relative to the cross-link sites, are identical or highly homologous in type I collagens of other vertebrates and in the human type II/XI collagen heterotypic fibril predominant in

cartilage. Thus, it may be speculated that vertebrate fibrillar collagens share the same domain structure.

Domains in Heterotypic Fibrillar Collagens—The type I collagen heterotypic fibril often contains type III and V collagens, in which some of the domain elements are altered in ways predicted to be functionally significant (51). It thus is proposed that the physical and cell-interactive properties of heterotypic fibrils may be modulated according to their contents of minor collagens carrying variants of one or more key sequences. For example, type V collagen, which lacks both an active integrin binding site at the appropriate position and an MMP-1 cleavage site, when incorporated in the type I fibril, may hinder cell interactions and fibril remodeling in tissues where low rates of collagen metabolism are desirable, such as the cornea.

Domains and the Polarity of Fibril Assembly—That type I and type II collagens and their fibrillar collagen-binding partners may share the same domain structure implies they assemble in parallel into the heterotypic fibril, where key functional domains fall into register. That atypical mutations in type I and III collagens appear at disparate regions of their triple helical domains, yet cluster to common zones of the fibril supports this contention (supplemental Fig. S1). However, although collagen monomer assembly into fibrils *in vivo* occurs most commonly in a head-to-

tail orientation, head-to-head, or tail-to-tail assembly is also seen and plays a role in tissue morphogenesis and repair (52). It is thus notable that GFPGER^{502–507}, proposed here to be the most critical sequence of type I collagen, is located midway between the N- and C-terminal cross-links. Therefore, if monomers assemble in a unipolar or bipolar direction into fibrils, GFPGER^{502–507} remains accessible for cell interactions.

Regulation of Cell-Fibril Interactions—The domain structure of the collagen microfibril was next considered in the context of the multivalent collagen fibril (Fig. 1). Although the fine structure of the collagen fibril surface remains incompletely understood, our following speculations assume that, at least in some cases, ligand-binding sites may align between microfibrils, as has been observed for some collagen-binding ligands (8, 53), although there are exceptions (54). Thus, in the fibril it is proposed that cell interaction domains appear every 5–6 nm across

the fibril's width, and at 67 nm intervals along its length (Fig. 5). Assuming integrin heterodimers to be ≤ 10 nm in diameter (55, 56) and collagen microfibril diameters ~ 5 –6 nm (10), GFPGER^{502–507}-bound integrins are proposed to cluster in stripes perpendicular to the long axis of the fibril, with one integrin bound per one or two microfibrils. According to this model, the fibril is an ideal substrate for integrin clustering, considered a key component of receptor activation and signaling (51, 56). Other collagen-binding ligands of ~ 5 –10 nm in diameter may potentially interact with one or more cell interaction domains, each occupying areas of $\sim 5 \times 20$ nm on the fibril. It follows that the cell interaction domain may accommodate one, or at most several ligands coincidentally, implying that cell- and ligand-binding interactions may be coordinately regulated. Therefore, it is speculated that integrins may positively or negatively modulate the interactions of other ligands with the fibril, or vice versa. Furthermore, MMP cleavage of one monomer may destabilize the fibril, displacing ligands, including integrins from neighboring monomers, or, conversely, could facilitate cell-fibril interactions (11) (not shown).

In summary, analysis of the distribution of functional sites and mutations on a two-dimensional model of the type I collagen fibril, and on an x-ray diffraction structure of the microfibril *in situ*, has identified candidate cell and matrix interaction domains. Moreover, such fibril domains may exist in heterotypic type I fibrils and other vertebrate fibrillar collagens. Defining the fine structure of the fibril domains and their functional inter-relationships may open new avenues for the therapeutic modulation of collagen function and metabolism in diseases, including fibrosis, atherosclerosis, and OI, where the protein plays a prominent role and for the engineering of fibrillar collagens and synthetic polymers for numerous applications in human medicine.

Acknowledgments—We thank Laila Huq for advice on phosphororin; Francisca Malfait for input regarding atypical mutations; Gregg Fields for discussion on MMPs; Richard Farndale for comments on von Willebrand Factor and discoidin domain receptor2-collagen interactions; Sergey Leikin, Barbara Brodsky, Anton Persikov, Dale Bodian, Warren Ewens, Olena Jacenko, Donald Gullberg, and Morris Karnovsky for helpful comments; and Andrew Likens and Nanita Barchi for artwork. Please note that an interactive data base containing a library of collagen ligand-binding sites and mutations integrated with physicochemical properties of the collagen proteins is under construction.³

REFERENCES

- Piez, K. A., and Reddi, A. H. (1984) *Extracellular Matrix Biochemistry*, Elsevier, New York
- Ayad, S., Boot-handford, R. P., Humpries, M. J., Kadler, K. E., and Shuttleworth, C. A. (1998) *The Extracellular Matrix Facts Book*, 2nd Ed., Academic Press, San Diego
- Prockop, D. J., and Kivirikko, K. I. (1995) *Annu. Rev. Biochem.* **64**, 403–434
- Kadler, K. E., Baldock, C., Bella, J., and Boot-Handford, R. P. (2007) *J. Cell Sci.* **120**, 1955–1958
- Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., Hyland, J. C., Korkko, J., Prockop, D. J., De Paepe, A., Coucke, P., Symoens, S., Glorieux, F. H., Roughley, P. J., Lund, A. M., Kuurila-

- Svahn, K., Hartikka, H., Cohn, D. H., Krakow, D., Mottes, M., Schwarze, U., Chen, D., Yang, K., Kuslich, C., Troendle, J., Dagleish, R., and Byers, P. H. (2007) *Hum. Mutat.* **28**, 209–221
- Yang, C., Hillas, P. J., Baez, J. A., Nokelainen, M., Balan, J., Tang, J., Spiro, R., and Polarek, J. W. (2004) *BioDrugs* **18**, 103–119
- Prockop, D. J., and Kivirikko, K. I. (1984) *N. Engl. J. Med.* **311**, 376–386
- Scott, J. E. (1988) *Biochem. J.* **252**, 313–323
- Di Lullo, G. A., Sweeney, S. M., Korkko, J., Ala-Kokko, L., and San Antonio, J. D. (2002) *J. Biol. Chem.* **277**, 4223–4231
- Orgel, J. P., Irving, T. C., Miller, A., and Wess, T. J. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9001–9005
- Perumal, S., Antipova, O., and Orgel, J. P. (2008) *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2824–2829
- Lauer-Fields, J. L., Tuzinski, K. A., Shimokawa, K., Nagase, H., and Fields, G. B. (2000) *J. Biol. Chem.* **275**, 13282–13290
- Sweeney, S. M., DiLullo, G., Slater, S. J., Martinez, J., Iozzo, R. V., Lauer-Fields, J. L., Fields, G. B., and San Antonio, J. D. (2003) *J. Biol. Chem.* **278**, 30516–30524
- Farndale, R. W., Lisman, T., Bihan, D., Hamaia, S., Smerling, C. S., Pugh, N., Konitsiotis, A., Leitinger, B., de Groot, P. G., Jarvis, G. E., and Raynal, N. (2008) *Biochem. Soc. Trans.* **36**, 241–250
- Linsenmayer, T. F. (1991) in *Cell Biology of Extracellular Matrix* (Hay, E. D., ed) pp. 7–44, Plenum Press, New York
- Orgel, J. P., Wess, T. J., and Miller, A. (2000) *Structure* **8**, 137–142
- Christopher, J. A., Swanson, R., and Baldwin, T. O. (1996) *Comput. Chem.* **20**, 339–345
- Chen, J. M., Sheldon, A., and Pincus, M. R. (1995) *J. Biomol. Struct. Dyn.* **12**, 1129–1159
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784
- Lee, K. S., Song, H. R., Cho, T. J., Kim, H. J., Lee, T. M., Jin, H. S., Park, H. Y., Kang, S., Jung, S. C., and Koo, S. K. (2006) *Hum. Mutat.* **27**, 599
- Ward, L. M., Lalic, L., Roughley, P. J., and Glorieux, F. H. (2001) *Hum. Mutat.* **17**, 434
- Pallos, D., Hart, P. S., Cortelli, J. R., Vian, S., Wright, J. T., Korkko, J., Brunoni, D., and Hart, T. C. (2001) *Arch. Oral Biol.* **46**, 459–470
- Malfait, F., Symoens, S., De Backer, J., Hermanns-Le, T., Sakalihsan, N., Lapiere, C. M., Coucke, P., and De Paepe, A. (2007) *Hum. Mutat.* **28**, 387–395
- Gensure, R. C., Makitie, O., Barclay, C., Chan, C., Depalma, S. R., Bastepe, M., Abuzahra, H., Couper, R., Mundlos, S., Sillence, D., Ala Kokko, L., Seidman, J. G., Cole, W. G., and Juppner, H. (2005) *J. Clin. Invest.* **115**, 1250–1257
- Pollitt, R., McMahon, R., Nunn, J., Bamford, R., Afifi, A., Bishop, N., and Dalton, A. (2006) *Hum. Mutat.* **27**, 716
- Yoneyama, T., Kasuya, H., Onda, H., Akagawa, H., Hashiguchi, K., Nakajima, T., Hori, T., and Inoue, I. (2004) *Stroke* **35**, 443–448
- Cabral, W. A., Makareeva, E., Letocha, A. D., Scribanu, N., Fertala, A., Stepewski, A., Keene, D. R., Persikov, A. V., Leikin, S., and Marini, J. C. (2007) *Hum. Mutat.* **28**, 396–405
- Körkkö, J., Ala-Kokko, L., De Paepe, A., Nuytinck, L., Earley, J., and Prockop, D. J. (1998) *Am. J. Hum. Genet.* **62**, 98–110
- Knight, C. G., Morton, L. F., Onley, D. J., Peachey, A. R., Messent, A. J., Smethurst, P. A., Tuckwell, D. S., Farndale, R. W., and Barnes, M. J. (1998) *J. Biol. Chem.* **273**, 33287–33294
- Knight, C. G., Morton, L. F., Peachey, A. R., Tuckwell, D. S., Farndale, R. W., and Barnes, M. J. (2000) *J. Biol. Chem.* **275**, 35–40
- Zhang, W. M., Kapyla, J., Puranen, J. S., Knight, C. G., Tiger, C. F., Penttinen, O. T., Johnson, M. S., Farndale, R. W., Heino, J., and Gullberg, D. (2003) *J. Biol. Chem.* **278**, 7270–7277
- Giancotti, F. G., and Ruoslahti, E. (1999) *Science* **285**, 1028–1032
- Hay, E. D. (ed) (1991) *Cell Biology of Extracellular Matrix*, Plenum Press, New York and London
- Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2005) *J. Biol. Chem.* **280**, 19343–19349
- Makareeva, E., Mertz, E. L., Kuznetsova, N. V., Sutter, M. B., DeRidder, A. M., Cabral, W. A., Barnes, A. M., McBride, D. J., Marini, J. C., and Leikin, S. (2008) *J. Biol. Chem.* **283**, 4787–4798

³ D. L. Bodian and T. E. Klein, manuscript in preparation.

36. Byers, P. H., Wallis, G. A., and Willing, M. C. (1991) *J. Med. Genet.* **28**, 433–442
37. Scott, J. E., and Tenni, R. (1997) *Cell Biochem. Funct.* **15**, 283–286
38. Xu, Y., Gurusiddappa, S., Rich, R. L., Owens, R. T., Keene, D. R., Mayne, R., Hook, A., and Hook, M. (2000) *J. Biol. Chem.* **275**, 38981–38989
39. Holtkotter, O., Nieswandt, B., Smyth, N., Muller, W., Hafner, M., Schulte, V., Krieg, T., and Eckes, B. (2002) *J. Biol. Chem.* **277**, 10789–10794
40. Tsilibary, E. C. (2003) *J. Pathol.* **200**, 537–546
41. Paul, R. G., and Bailey, A. J. (1996) *Int. J. Biochem. Cell Biol.* **28**, 1297–1310
42. Brennan, M. (1989) *J. Biol. Chem.* **264**, 20947–20952
43. McCarthy, A. D., Etcheverry, S. B., Bruzzzone, L., Lettieri, G., Barrio, D. A., and Cortizo, A. M. (2001) *BMC Cell Biol.* **2**, 16
44. Paul, R. G., and Bailey, A. J. (1999) *Int. J. Biochem. Cell Biol.* **31**, 653–660
45. Chen, J., Brodsky, S., Li, H., Hampel, D. J., Miyata, T., Weinstein, T., Gafter, U., Norman, J. T., Fine, L. G., and Goligorsky, M. S. (2001) *Am. J. Physiol.* **281**, F71–F80
46. Reigle, K. L., Di Lullo, G., Turner, K. R., Last, J. A., Chervoneva, I., Birk, D. E., Funderburgh, J. L., Elrod, E., Germann, M. W., Surber, C., Sander-son, R. D., and San Antonio, J. D. (2008) *J. Cell. Biochem.*, in press
47. Vogel, K. G., and Heinegard, D. (1985) *J. Biol. Chem.* **260**, 9298–9306
48. Chapman, J. A. (1974) *Connect. Tiss. Res.* **2**, 137–150
49. Shaw, L. M., and Olsen, B. R. (1991) *Trends Biochem. Sci.* **16**, 191–194
50. Eyre, D. R., Pietka, T., Weis, M. A., and Wu, J. J. (2004) *J. Biol. Chem.* **279**, 2568–2574
51. Siljander, P. R., Hamaia, S., Peachey, A. R., Slatter, D. A., Smethurst, P. A., Ouwehand, W. H., Knight, C. G., and Farndale, R. W. (2004) *J. Biol. Chem.* **279**, 47763–47772
52. Kadler, K. E., Holmes, D. F., Trotter, J. A., and Chapman, J. A. (1996) *Biochem. J.* **316**, 1–11
53. San Antonio, J. D., Lander, A. D., Karnovsky, M. J., and Slayter, H. S. (1994) *J. Cell Biol.* **125**, 1179–1188
54. Holmes, D. F., Gilpin, C. J., Baldock, C., Ziese, U., Koster, A. J., and Kadler, K. E. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7307–7312
55. Nermut, M. V., Green, N. M., Eason, P., Yamada, S. S., and Yamada, K. M. (1988) *EMBO J.* **7**, 4093–4099
56. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000) *Cell* **101**, 47–56
57. Thomas, E. K., Nakamura, M., Wienke, D., Isacke, C. M., Pozzi, A., and Liang, P. (2005) *J. Biol. Chem.* **280**, 22596–22605
58. Koide, T., Takahara, Y., Asada, S., and Nagata, K. (2002) *J. Biol. Chem.* **277**, 6178–6182
59. Cabral, W. A., Makareeva, E., Colige, A., Letocha, A. D., Ty, J. M., Yeowell, H. N., Pals, G., Leikin, S., and Marini, J. C. (2005) *J. Biol. Chem.* **280**, 19259–19269
60. Makareeva, E., Cabral, W. A., Marini, J. C., and Leikin, S. (2006) *J. Biol. Chem.* **281**, 6463–6470
61. Morello, R., Bertin, T. K., Chen, Y., Hicks, J., Tonachini, L., Monticone, M., Castagnola, P., Rauch, F., Glorieux, F. H., Vranka, J., Bachinger, H. P., Pace, J. M., Schwarze, U., Byers, P. H., Weis, M., Fernandes, R. J., Eyre, D. R., Yao, Z., Boyce, B. F., and Lee, B. (2006) *Cell* **127**, 291–304
62. Wang, H., Fertala, A., Ratner, B. D., Sage, E. H., and Jiang, S. (2005) *Anal. Chem.* **77**, 6765–6771
63. Fujisawa, R., Zhou, H., and Kuboki, Y. (1994) *Connect. Tiss. Res.* **31**, 1–10
64. Huq, N. L., Loganathan, A., Cross, K. J., Chen, Y. Y., Johnson, N. I., Wil-letts, M., Veith, P. D., and Reynolds, E. C. (2005) *Arch. Oral Biol.* **50**, 807–819
65. Lisman, T., Raynal, N., Groeneveld, D., Maddox, B., Peachey, A. R., Huiz-inga, E. G., de Groot, P. G., and Farndale, R. W. (2006) *Blood* **108**, 3753–3756
66. Konitsiotis, A. D., Raynal, N., Bihan, D., Hohenester, E., Farndale, R. W., and Leitinger, B. (2008) *J. Biol. Chem.* **283**, 6861–6868
67. Baronas-Lowell, D., Lauer-Fields, J. L., and Fields, G. B. (2004) *J. Biol. Chem.* **279**, 952–962
68. Reyes, C. D., and Garcia, A. J. (2004) *J. Biomed. Mater. Res. A* **69**, 591–600