REVIEW

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Mapping Structural Landmarks, Ligand Binding Sites, and Missense Mutations to the Collagen IV Heterotrimers Predicts Major Functional Domains, Novel Interactions, and Variation in Phenotypes in Inherited Diseases Affecting Basement Membranes

J. Des Parkin,[1] James D. San Antonio,[2] Vadim Pedchenko,[3] Billy Hudson,[3] Shane T. Jensen,[4] and Judy Savige[1]*

[1]*Department of Medicine (Northern Health), The University of Melbourne, Northern Health, Epping VIC 3076, Australia;* [2]*Operations, Orthovita Inc, Malvern, Pennsylvania;* [3]*Center for Matrix Biology and Department of Medicine, Vanderbilt University Medical Center, Tennessee;* [4]*Wharton Business School, University of Pennsylvania, Philadelphia, Pennsylvania*

Communicated by David N. Cooper

Received 8 May 2010; accepted revised manuscript 13 October 2010.
Published online 18 November 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/humu.21401

ABSTRACT: Collagen IV is the major protein found in basement membranes. It comprises three heterotrimers (α1α1α2, α3α4α5, and α5α5α6) that form distinct networks, and are responsible for membrane strength and integrity. We constructed linear maps of the collagen IV heterotrimers ("interactomes") that indicated major structural landmarks, known and predicted ligand-binding sites, and missense mutations, in order to identify functional and disease-associated domains, potential interactions between ligands, and genotype–phenotype relationships. The maps documented more than 30 known ligand-binding sites as well as motifs for integrins, heparin, von Willebrand factor (VWF), decorin, and bone morphogenetic protein (BMP). They predicted functional domains for angiogenesis and haemostasis, and disease domains for autoimmunity, tumor growth and inhibition, infection, and glycation. Cooperative ligand interactions were indicated by binding site proximity, for example, between integrins, matrix metalloproteinases, and heparin. The maps indicated that mutations affecting major ligand-binding sites, for example, for Von Hippel Lindau (VHL) protein in the α1 chain or integrins in the α5 chain, resulted in distinctive phenotypes (Hereditary Angiopathy, Nephropathy, Aneurysms, and muscle Cramps [HANAC] syndrome, and early-onset Alport syndrome, respectively). These maps further our understanding of basement membrane biology and disease, and suggest novel membrane interactions, functions, and therapeutic targets.
Hum Mutat 32:127–143, 2011. © 2011 Wiley-Liss, Inc.

KEY WORDS: interactome; genotype–phenotype correlation; collagen IV; Alport syndrome

Introduction

The collagens represent the major proteins of the extracellular matrix and 29 types (I–XXIX) assembled from at least 44 distinct α-chains have been identified [Myllyharju and Kivirikko, 2004; Soderhall et al., 2007]. Each molecule is a homo- or heterotrimer of three α-chains with the characteristic Gly-Xaa-Yaa repeat sequence where Xaa and Yaa are often proline and hydroxyproline. Collagens serve as scaffolds for the attachment of cells and matrix proteins, but are increasingly recognized to have many other ligands and be highly biologically active [Di Lullo et al., 2002; Sweeney et al., 2008; Timpl, 1989].

Collagen I

Collagen I is the most abundant protein in the body and contributes to the structural integrity of many tissues. It is a fibrillar molecule that comprises a heterotrimer of two α1 and one α2 chains encoded by the *COL1A1* and *COL1A2* genes. Collagen I has more than 100 different ligands, as diverse as bone morphogenetic protein (BMP), von Willebrand factor (VWF), and interleukin 2 [Di Lullo et al., 2002; Myllyharju and Kivirikko, 2004; Sweeney et al., 2008]. It is affected by mutations resulting in osteogenesis imperfecta and other connective tissue disorders, and also by glycation in diabetes and aging.

Collagen I Interactome

Linear protein maps ("interactomes") of the collagen I α1α1α2 heterotrimer have documented novel structural features and ligand-binding sites, predicted new interactions and functions, and summarized the molecule's diverse biological functions [Sweeney et al., 2008]. The maps demonstrated major ligand-binding regions, a "cell interaction" domain that regulates integrin-mediated cell binding and fibril remodeling, and a "matrix interaction" domain that determines crosslinking, proteoglycan interactions, and tissue mineralization. These maps suggested critical functional sites colocalize within such domains and that domain-specific, ligand-mediated functions were likely to be cooperative. For example, the proximity of sites for integrin-binding and collagenase cleavage predicted fibril remodeling

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Judy Savige, The University of Melbourne, Department of Medicine (Northern Health), The Northern Hospital, Epping VIC 3076, Australia.
E-mail: jasavige@unimelb.edu.au

Contract grant sponsor: NIH; Contract grant numbers: DK18381; DK065123 (to B.G.H.).

disrupts cell–fibril interaction; and the colocalization of binding sites for fibronectin, fibrillogenesis, and collagenase cleavage suggested a role for fibronectin in collagen assembly and degradation. Importantly, the collagen I map also correlated mutations in the α1 and α2 chains and clinical phenotypes in osteogenesis imperfecta (MIM 166200) [Marini et al., 2007; Sweeney et al., 2008]. Hundreds of missense mutations have been described, and the corresponding phenotypes vary from mild and asymptomatic, to severe with multiple, frequent fractures. Some of this variation is explained by mutation location and the nature of the substituting residues. Mutations closer to the carboxyl terminus generally result in more severe disease because disrupted helix propagation temporarily exposes residues amino-terminal to the site on all three chains to excessive hydroxylation and glycosylation [Engel and Prockop, 1991]. Severe disease also results from mutations where glycine is substituted with larger residues, such as valine, or more highly charged residues, such as aspartic acid [Byers et al., 1991; Marini et al., 2007]. Even single-point mutations influence the mechanical behavior of these tissues. Mutations associated with the most severe phenotypes correlate with weakened intermolecular adhesion, increased intermolecular spacing, reduced stiffness, and reduced strength of collagen fibrils [Gautieri et al., 2009]. However, the linear collagen I map provided evidence for a third mechanism for genotype–phenotype correlations: namely, that severe disease was more likely when missense mutations affected major structural or ligand-binding sites [Marini et al., 2007; Scott and Tenni, 1997; Sweeney et al., 2008].

Collagen IV

In contrast to collagen I, collagen IV forms networks, and is widely expressed in vascular and other basement membranes. The collagen IV family comprises six homologous α-chains, α1–α6, encoded by the *COL4A1–COL4A6* genes. These have arisen by reduplication from the ancestral *COL4A1* gene and are divided into two families—*COL4A1*-like (the α1, α3, and α5 chains), and *COL4A2*-like (the α2, α4, and α6 chains), where the corresponding genes share exon–intron organization, exon size, sequence homology, and the proteins have common structural features. Each collagen IV chain consists of the typical helical intermediate sequence as well as noncollagenous (NC) domains at the amino and carboxyl termini and multiple short noncollagenous interruptions [Khoshnoodi et al., 2008; Netzer et al., 1998]. The heterotrimers assemble intracellularly beginning with disulfide bond formation at the carboxyl terminal NC1 and progressing towards the 7S domain. They are then secreted to form a supramolecular network through dimerization at the carboxyl terminus and tetramerization at the 7S domain [Siebold et al., 1988], and the networks are further stabilized by lateral associations [Yurchenco and Ruben, 1987].

Collagen IV is found as three distinct heterotrimers in separate networks. The α1α1α2 network is ubiquitous in embryonic life and persists in vascular and other membranes (including brain, proximal renal tubule, muscle) in adulthood, but in specialized membranes in the glomerulus, lung, cochlea and retina is replaced in infancy by the α3α4α5 network, and by the α5α5α6 network in the epidermis, testis, and Bowman's capsule. The α1α1α2 and α3α4α5 networks are critical in embryogenesis, angiogenesis, and haemostasis, tumor growth and invasion, and microbial infection, and the α3α4α5 network, in particular, is responsible for the integrity of fluid–membrane barriers. The role of the α5α5α6 network is less clear.

The collagen IV networks are also affected in inherited and other diseases. Inherited diseases are most often due to missense mutations and associated with vascular or renal abnormalities. Mutations in the α1 chain result in stroke, porencephaly, (MIM 175780) and the **H**ereditary **A**ngiopathy, **N**ephropathy, **A**neurysms, and muscle **C**ramps syndrome (HANAC; MIM 611773) syndrome [Gould et al., 2005; Plaisier et al., 2007; Sibon et al., 2007]. Heterozygous mutations in the α3 or α4 chains produce Thin Basement Membrane Nephropathy (TBMN) with isolated hematuria, or rarely, autosomal dominant Alport syndrome (MIM 104200) with renal failure and hearing loss. Homozygous or compound heterozygous mutations in the α3 or α4 chains result in autosomal recessive Alport disease (MIM 203780) with renal failure, hearing loss, lenticonus, and retinopathy. Hemizygous mutations in the α5 chain cause X-linked Alport syndrome (MIM 301050). No disease-producing missense mutations have been described in the α2 or α6 chains.

The most clinically significant of these diseases is X-linked Alport syndrome. It affects one in 5,000 individuals, and more than 200 missense mutations have been described to date. Again, missense mutations affecting the carboxyl terminal residues of the α5 chain or where glycine is replaced by larger or more highly charged residues result in a severe phenotype with end-stage renal failure before the age of 30 [Gross et al., 2002; Jais et al., 2000; Persikov et al., 2004]. However it is still not always possible to predict the clinical course from the nature of the underlying mutations.

The α3α4α5 network is also affected by autoantibody-mediated rapidly progressive glomerulonephritis (antiGBM disease or "Goodpasture syndrome") [Saus et al., 1988]. Sometimes alloantibodies to components of the α3α4α5 network develop in X-linked Alport syndrome after renal transplantation leading to graft failure. In addition, the collagen IV networks are affected by glycation in diabetes and aging, and this alters matrix flexibility, proteolytic susceptibility, and subsequent function [Mott et al., 1997; Reigle et al., 2008; Tarsio et al., 1987].

Construction of the Collagen IV Interactomes

We have constructed linear maps of the three collagen IV heterotrimers indicating major structural landmarks, known and predicted ligand-binding sites, and missense mutations, in order to demonstrate potential functional domains and ligand interactions, and explain genotype–phenotype variation in inherited disease.

The human reference sequences (NP_001836, α1 isoform 1; NP_001837, α2; NP_000082, α3 isoform 1; NP_000083, α4; NP_000486, α5 isoform 1; and NP_001838, α6 isoform 1) were aligned as the α1α1α2, α3α4α5 and α5α5α6 heterotrimers in Microsoft Word according to the carboxyl terminal NC1 sequences with their 12 conserved cysteine residues, the triple helix NC interruptions, and the 7S domains using the Clustal W function of MacVector 9.0 (Accelrys). The collagen IV α1, α3, α5, and α6 chains also undergo alternative splicing. Isoform 1 represents the canonical sequence (Supp. Table S1), and other isoforms differ by small insertions to large deletions.

Structural domains and sites related to collagen IV assembly and turnover were identified from the literature and open access bioinformatics web sites (Uniprot, UCSC, etc.). Binding sites for integrins, cells, extracellular matrix molecules, and other ligands were identified from the literature, Web sites (Uniprot, UCSC, BioGrid, MINT, STRING, etc.), and by reference to the collagen I maps [Di Lullo et al., 2002; Matthews et al., 2009; Sweeney et al., 2008]. Some sites were predicted from binding motifs. Others

were derived from rotary shadowing electron microscopy measurements using the assumption that the average spacing of residues on the triple helix was 0.238 nm [Pietz and Reddi, 1984]. Still others were derived from experiments demonstrating binding to collagen IV proteolytic fragments or mimetic peptides. These were considered relevant because some ligands bind only to denatured collagen in vivo. In studies using collagen IV from the Engelbreth-Holm-Swarm (EHS) tumor, which comprises only the α1α1α2 heterotrimer, and where the chain was not identified, binding was presumed to occur to the more abundant α1 chain.

Sites involved in the major functions of the collagen IV networks (endothelial and epithelial cell binding domains, angiogenesis, hemostasis) or in disease (tumor growth and invasion, antitumor, microbial infection, glycation, and autoimmune disease) were indicated on the maps.

The consequences of missense mutations affecting major structural sites, ligand-binding sites, and functional domains of the collagen IV networks were then investigated. Missense variants were identified from the literature and open access Web databases (UniProt; Embl; HMGD database/Biobase, NCBI, etc.). Variants were classified as "pathogenic" or "nonpathogenic" by their contributors. Pathogenic missense variants in the α1–α5 chains were examined to determine whether those affecting a major structural domain or ligand-binding site were more likely to produce a distinctive clinical phenotype. In particular, whether mutations causing HANAC (α1 chain), autosomal dominant Alport syndrome rather than thin basement membrane nephropathy (α3 and α4 chains), or X-linked Alport syndrome with juvenile- (before the age of 30) or adult-onset end-stage renal failure were due to mutations affecting a major structural domain or ligand-binding site.

In addition, the distribution of α5 sequence variants was examined for randomness. Briefly, mutational densities for each exon were calculated and compared with simulated mutational maps consistent with the "null hypothesis" of no spatial variation. Exons with an unusually high or low density relative to the null distribution were then analyzed in more detail, and contiguous exons were further studied to increase the power of testing.

Collagen IV α1α1α2, α3α4α5, and α5α5α6 Interactomes (Figs. 1A–C)

Structural Landmarks and Sites Related to Assembly and Remodeling

The signal peptide at the amino terminal 7S domain directs posttranslational transport but is subsequently cleaved. The cysteine and lysine residues beyond the signal peptide form crosslinks through disulfide and lysine–hydroxylysine bonds respectively to produce the tetramer [Khoshnoodi et al., 2008]. Each α-chain has a collagenous Gly-Xaa-Yaa sequence, where Xaa and Yaa are often proline and hydroxyproline, as well as a number of short (1–24 residue) noncollagenous interruptions ranging from 21 in the α1 chain to 26 in the α4 chain. These confer flexibility and possibly have a role in connecting with supramolecular partners. The 7S kink is located 60 nm (about 250 residues) from the amino terminus on rotary shadowing [Pietz and Reddi, 1984]. Glycosylation is critical in protein folding and stability. N-glycosylation requires the Asp-Xaa-Ser/Thr sequence [Spiro, 2002]. O-glycosylation is more common and occurs at serine, threonine, hydroxylysine, and hydroxyproline residues within the collagenous domain without requiring specific sequences. Hydroxylation is a prerequisite for glycosylation, and there are about 50 hydroxylysine-linked disaccharides in each collagen chain [Hudson et al., 1993]. Prolines and lysines in the Yaa position are hydroxylated. Hydroxyprolines are underlined in the α1 chain in Figure 1A but have not been described for the other chains. We have used the term O for hydroxyproline rather than P where this was used in the original report, but O and P are generally interchangeable for collagen IV. There are also some–X-4Hyp-Ala sequences in the chains. The interaction of prolyl 4-hydroxylase clearly depends on the amino acid in the Xaa position and proline is particularly favorable, but alanine, leucine, arginine, valine, and glutamate are too. Hydroxylation is catalyzed mainly by prolyl 4-hydroxylase, and less often by prolyl 3-hydroxylase or lysyl 5- hydroxylase. If hydroxylation does not occur, the unfolded chain remains bound to the enzyme within the endoplasmic reticulum. Peptide-linked lysine is hydroxylated to form 5-hydroxylysine that then attaches glucosyl and galactosyl residues. Hydroxylysyl residues are also modified to form crosslinks and failure of lysine hydroxylation prevents tetramer formation.

The NC1 domain comprises the carboxyl terminal ~230 residues that fold to form the globular NC1 domain. Crosslinking results in a hexamer that can be dissociated into monomer and dimer subunits. The dimers are held together by covalent S-hydroxylysine-methionine crosslinks between methionine and hydroxylysine residues in opposite chains [Vanacore et al., 2005, 2009]. The 12 cysteines in each NC1 domain form intramolecular disulfide bonds. The domain swapping residues in the 13 residue donor β-hairpin motif and the 15-residue acceptor docking site with genetic hypervariability result in selective formation of heterotrimers [Khoshnoodi et al., 2006a]. The hypervariable regions of the α2 and α5 chains are critical in the formation of the α1α1α2 and α3α4α5 heterotrimers, respectively [Kang et al., 2007; Khoshnoodi et al., 2006b].

The highly complex folding and assembly of the collagen IV triple helix requires the coordination of many endoplasmic reticulum-based enzymes and molecular chaperones including HSP47 [Koide et al., 2006] and probably Secreted Protein, Acidic, and Rich in Cysteine (SPARC) [Martinek et al., 2007]. The collagen molecule moves from the endoplasmic reticulum to the Golgi body in partnership with HSP47 [Canty and Kadler, 2005]. HSP47 recognizes GXR where R is critical, and all potential binding sites occur in the triple helix. The GXR sequence is found at multiple locations in each chain ranging from 12 in the α5 to 26 in the α2 chain. HSP47 may compete for binding with prolyl 4-hydroxylase [Asada et al., 1999]. The $(GPP)_4$ sequence near the α3α4α5 carboxyl terminus may function in triple helix nucleation [Hyde et al., 2006] as well as platelet binding as discussed later.

Collagen IV is remodeled by enzymatic cleavage in embryogenesis and angiogenesis, as a result of normal turnover, as well as in tumor invasion and spread. It is degraded by a specific group of matrix metalloproteinases (MMP-2, -3, -9, -10,-13, -19, and -26; [Somerville et al., 2003] and by serine proteinases. MMP-2 and -9 are the major collagen IV collagenases. They have a common cleavage site in the α1 (G/I at residue 446) and α2 (G/L at residue 463) chains [Hostikka and Tryggvason, 1988]. These overlap with sites for integrin binding, and integrins appear important in MMP activation [Eble et al., 1993]. These motifs are conserved in other collagen IV chains. MMP-3 and -9 cleave asymmetrically between G/F and G/L on adjacent α1 and α2 chains leaving the NC1 domains intact [Gioia et al., 2009; Mott et al., 1997]. Predicted MMP-13 cleavage sites are at GPVGMK (near residue 990) and

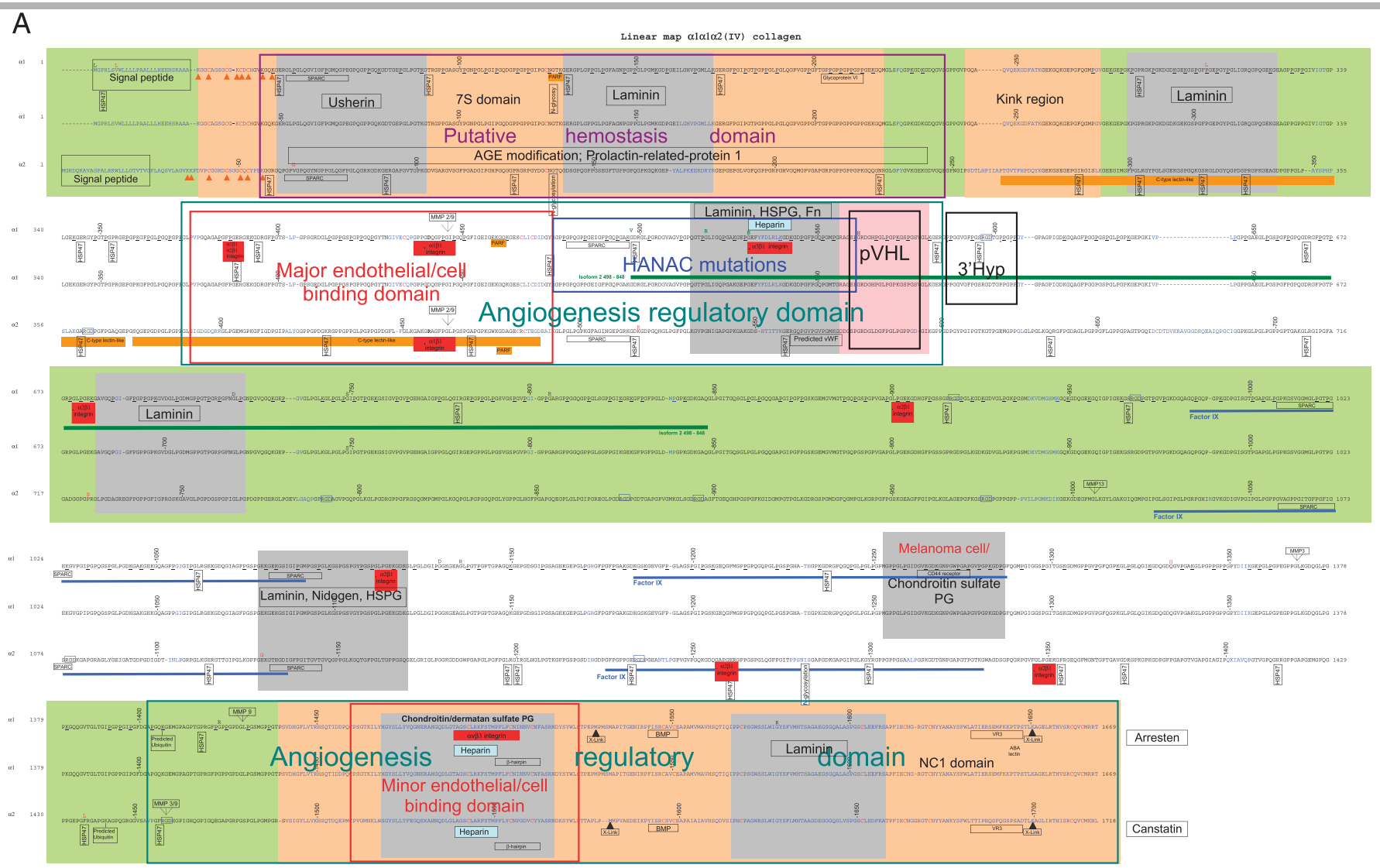

Figure 1. Linear protein maps of the **A:** α1α1α2; **B:** α3α4α5; and **C:** α5α5α6 heterotrimers of collagen IV. The protein sequences were derived and aligned as indicated in the text. The sequence is linear from left to right, with alternating bands shown in green or white only to distinguish between the continuous heterotrimer sequence. Noncollagenous interruptions in the sequence are the residues in blue. Cysteines are shown in orange and indicated with orange arrowheads. Binding sites are indicated on only one of the 2 α1 or α5 chains. Binding sites for laminin, nidogen, heparan sulfate proteoglycan, and fibronectin were derived from rotary shadowing studies in EHS-derived collagen and are shown here on both the α1 and α2 chains but the locations are approximate. Otherwise, binding sites were identified from binding motifs. Predicted sites have been identified from homology with known motifs. Underlined residues in the α1 chain are hydroxylated. Residues at the same location in the α2 chain are generally also hydroxylated. Asterisks indicate 3′ hydroxylation sites. Sequence variants are indicated above the wild type. Nonpathogenic changes are shown in red and pathogenic variants in black (where phenotype is not characterised or is: for α1 chain—vascular stroke or porencephaly; for α3 and α4: TBMN; and for α5: X-linked Alport syndrome with adult onset renal failure) or green (for α1 chain—HANAC; α3 and α4: autosomal dominant Alport syndrome; and for α5—X-linked Alport syndrome with juvenile onset renal failure. References are provided in the text.

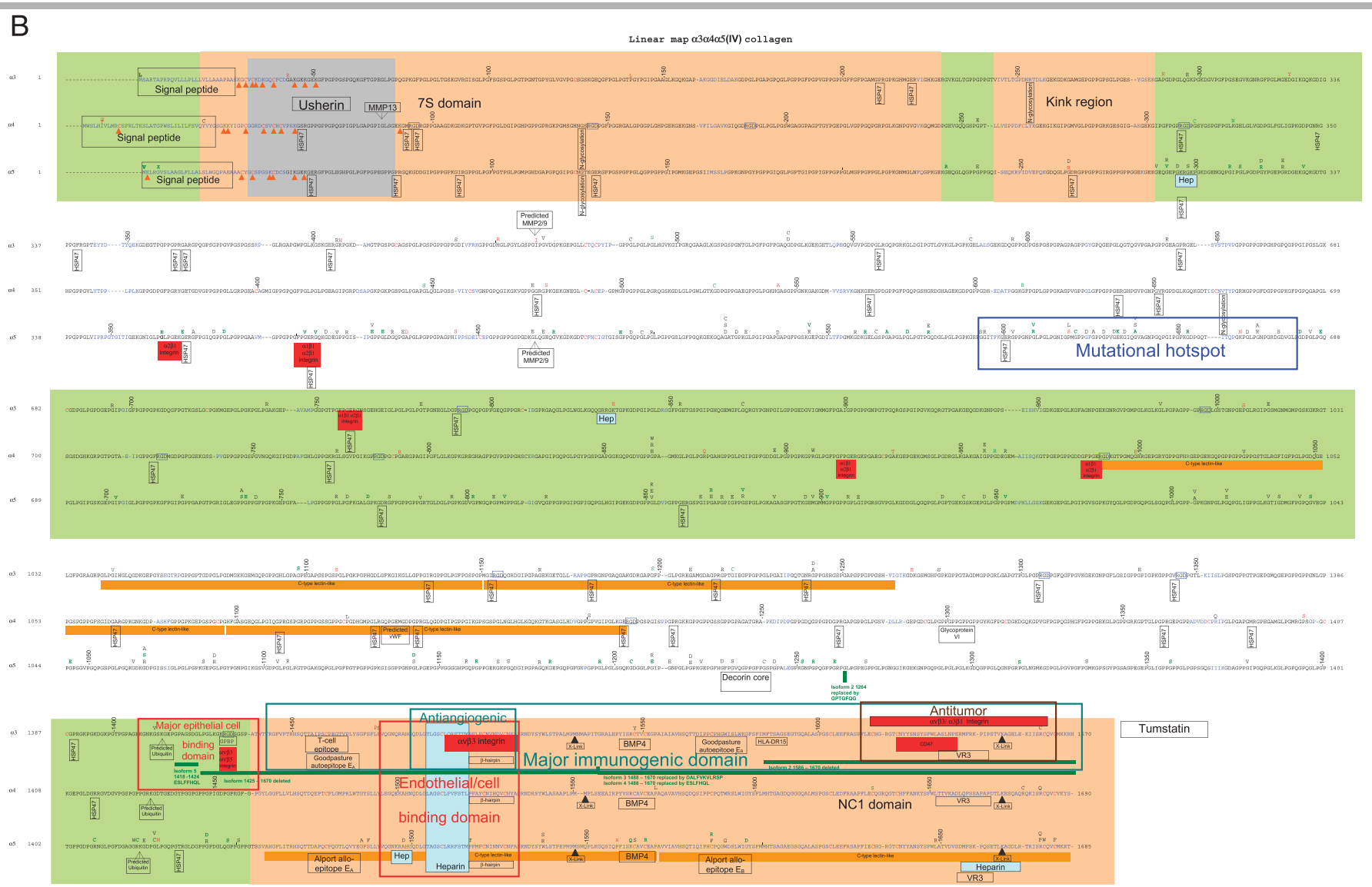

Figure 1. Continued.

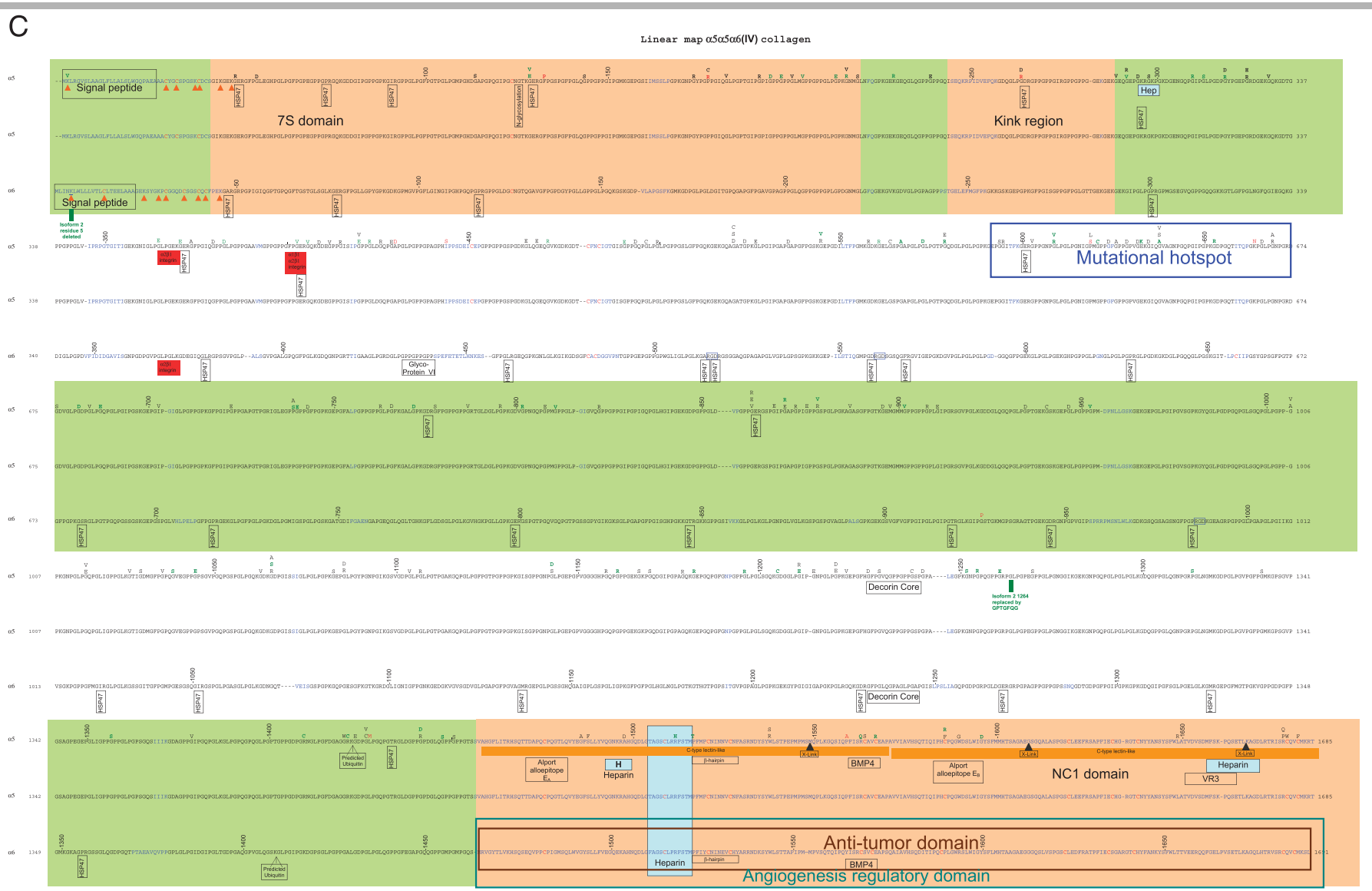

Figure 1. Continued.

GPMGLK (residue 1003) in the α2 chain, and at GPIGLS (residue 85) in the α4 chain [Deng et al., 2000].

Collagen IV is also cleaved by neutrophil proteinase 3, elastase (MMP-12), and cathepsins K, B, S, and possibly L. Neutrophil proteinase 3 cleaves at V/E, S/V, S/L, and Q/L [Rao et al., 1991], and there are many potential cleavage sites for these motifs except Q/L in each chain. Cathepsin K cleaves at G/K and is particularly important in chain turnover [Garnero et al., 1998; Nosaka et al., 1999].

Collagen IV also undergoes intracellular proteasomal degradation. Ubiquitin covalently attaches to a KG sequence, and the binding of multiple ubiquitins results in degradation. The KG site has only been described for the α3 chain in the triple helix near the NC1 domain (Uniprot), but this motif is conserved in all the chains.

Integrin-Binding Sites (Fig. 2)

Integrins mediate cell adhesion to all basement membrane proteins including collagen IV. The collagen IV integrin receptors belong to the β1 subgroup, namely, α1β1 and α2β1 [Leitinger and Hohenester, 2007; White et al., 2004]. Binding triggers pathways involved in cell migration and invasion, including phosphorylation of FAK, paxilin, activation of small G-proteins, PKC, and PI3 kinase as well as changes in intracellular calcium levels.

Collagen IV has 3 major integrin-binding motifs: GFOGER, which is the commonest and also occurs in fibrillar collagen; the classical RGD site; and other non-RGD binding motifs (Supp. Table S2). Integrin-binding sites are distributed throughout each heterotrimer, and the location of sites is important because receptor clustering appears to be necessary for activation. Some sites are cryptic and only accessible after denaturation, proteolysis etc. For example, cleavage of collagen IV during angiogenesis results in the loss of α1β1 but gain of αvβ3 binding [Xu et al., 2000]. RGD sites are present at multiple locations in the collagenous domains [Kim et al., 1994], but are generally inaccessible to cells in the native molecule [Herbst et al., 1988; Kim et al., 1994].

Integrins α1β1/α2β1: a major site for binding of α1β1 and α2β1 integrins has been identified within the triple-helical cyanogen bromide-derived fragment, CB3, located 100 nm from the amino-terminus of collagen IV [Vandenberg et al., 1991]. Antibodies to this fragment block cell binding by 80%. Subsequently, a single α1β1 and two α2β1 integrin binding sites were predicted on this fragment [Kern et al., 1993]. Further refinement identified a conformational-dependent site formed by the unique whole collagen heterotrimer spatial arrangement of the three residues, two Asp461 on the α1 chains and Arg 461 on the α2 chain, as critical for α1β1 integrin binding [Eble et al., 1993]. More recently, functional activity of this α1β1 binding site was confirmed using synthetic triple-helical peptides corresponding to residues 457–468 of the α1 and α2 chains stabilized with an artificial cysteine knot [Renner et al., 2004].

The precise identity and structure of the α2β1 binding site(s) in collagen IV remains unknown. A potential candidate is the GFOGER sequence identified as an integrin binding site in collagen I [Knight et al., 1998]. Interestingly, both the α1β1 and α2β1 integrins recognize GFOGER as the minimal binding motif on collagen [Knight et al., 2000; Siljander et al., 2004]. However, the α1β1 integrin binds with higher affinity to collagen IV, and α2β1 to collagen I [Kern et al., 1993; Tulla et al., 2001; Zhang et al., 2003]. The α2β1 integrin recognizes GXO/SGER, and then a hierarchy of GFPGER > GLPGER > GMPGER > GAPGER, and GLOGER and GASGER [Siljander et al., 2004]. The F is not critical for binding. There are no GLPGER, GASGER, GMPGER, GQRGER, GASGQR, or GFPGEK sequences in collagen IV. The most amino terminal GFOGER site on the α1 chain may represent the principal site for endothelial cell binding and activation [Knight et al., 2000; Xu et al., 2000], and for angiogenesis [Sweeney et al., 2008]. The α2β1 site on the α1 chain appears to facilitate lung cancer cell adhesion [Khoshnoodi et al., 2008].

Integrins α10β1 and α11β1: chondrocytes and fetal muscle cells adhere to collagen IV through these integrins [Tiger et al., 2001; Zhang et al., 2003], but the sites are unknown [Tulla et al., 2001; Zhang et al., 2003].

Integrin αvβ3: there are three binding sites at the carboxyl terminus of the α3 chain, two within the NC1 domain. The carboxyl terminal KRGDS site within the triple helix represents the only functional RGD cell-binding site in collagen IV [Pedchenko et al., 2004]. It mediates adhesion of podocytes [Borza et al., 2008], and overlaps with the binding site for the Goodpasture protein-binding protein (GPBP) [Raya et al., 1999]. A second non-RGD αvβ3 site located in the amino terminal part of the α3NC1 domain has anti-angiogenic activity [Maeshima et al., 2002; Sudhakar et al., 2003]. The third site in the carboxyl terminal part of the α3NC1 (amino acids 185–203) has antitumor activity, and colocalizes with the CD47/ IAP (integrin-associated protein) [Han et al., 1997; Shahan et al., 1999]. It also inhibits the activation of human neutrophils [Monboisse et al., 1994], inhibits the proliferation, and induces apoptosis of, capillary endothelial cells, and reduces tumor growth in vivo [Maeshima et al., 2000]. Both αvβ3 sites within the NC1 domain of the α3 chain are brought into close proximity by the β-hairpin binding to VR3, sufficient for activation. Both also overlap with heparin-binding sites which may enhance cell binding to the membrane through cell-surface proteoglycans.

The NC1 domains of the α1, α2, α3, and α6 chains all have anti-angiogenic properties [Colorado et al., 2000; Kamphaus et al., 2000; Petitclerc et al., 2000] that are attributed to integrin binding sites in at least the α1 and α3 chains [Sudhakar and Boosani, 2008; Sudhakar et al., 2005]. In the α1 chain, anti-angiogenic activity is mediated by the α1β1 integrin binding within the carboxyl terminal half of the NC1 domain [Nyberg et al., 2008].

Integrin α3β1: one site for α3β1 binding has been identified in the triple helical domain using a synthetic peptide corresponding to residues 531–543 of the α1 chain [Miles et al., 1995]. Interestingly, this peptide promoted adhesion of melanoma and ovarian carcinoma cell lines in single-stranded conformation, thus providing the first evidence for existence of triple-helix-independent integrin binding sites within the collagenous domain. Another α3β1 site is located at the carboxyl α3 NC1 domain and overlaps with the non-RGD αvβ3 binding site, suggesting that the α3β1 integrin trans-dominantly inhibits αvβ3 function [Borza et al., 2006; Hodivala-Dilke et al., 1998]. The antitumor activity of this region has been confirmed [Sudhakar and Boosani, 2008].

Cell-Binding Sites

The collagen IV networks in basement membranes bind all cells except erythrocytes. The α1α1α2 heterotrimer is usually anchored in vascular membranes to endothelial cells, but interacts also with neutrophils, lymphocytes, and platelets, as well as lung, breast, kidney, and colon tumor cells, and bacteria. In the kidney glomerulus, the α3α4α5 heterotrimer interacts specifically with glomerular epithelial and endothelial cells. Binding occurs through integrin and nonintegrin-mediated mechanisms. Tumor cells bind using the same integrins as endothelial cells. Integrin-mediated cell adhesion is promoted by the heparan sulfate side chains of perlecan, glypican, and syndecans, as well as glycoprotein VI and VWF.

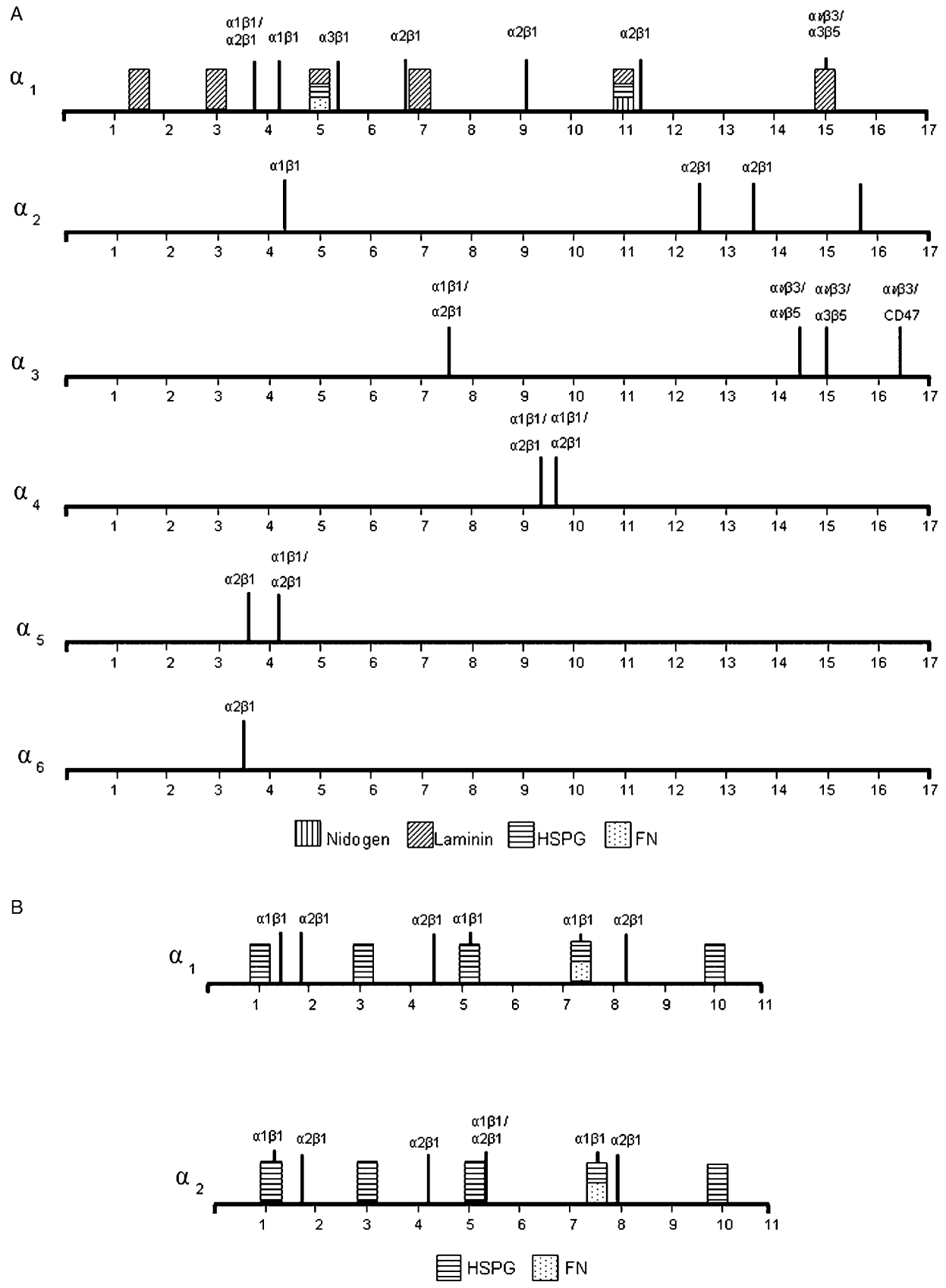

Figure 2. A comparison of the integrin and extracellular matrix binding sites for collagen IV (**A**) and collagen I (**B**). These diagrams demonstrate the periodicity of integrin and extracellular structural protein binding (laminin, nidogen, HSPG-heparan sulfate proteoglycan, FN-fibronectin) to the collagen chains. On the collagen IV heterotrimers they demonstrate integrin binding sites throughout the α1α1α2, α3α4α5 and α5α5α6 heterotrimers, and the periodicity of structural protein binding in the α1α1α2 heterotrimer. On the collagen I heterotrimer they demonstrate the periodicity of integrin and structural protein binding sites.

Endothelial cells

Endothelial cells typically bind collagen IV through the α1β1 and α2β1 integrins but also via αvβ3 and other integrins [Marneros and Olsen, 2001; Pedchenko et al., 2004, 2005; Tsilibary et al., 1990]. The major endothelial cell binding sites in the α1α1α2 hetrotrimer are the GFOGER sequences in the more amino terminal triple helix, and the TAGSCLRKFSTM peptide derived from the α1 NC1 domain promotes adhesion and spreading of bovine endothelial cells [Tsilibary et al., 1990]. There are similar motifs to this in all the other collagen IV NC1 domains.

Epithelial cells

Glomerular, retinal, and probably other epithelial, as well as endothelial, cells bind to the KRGDS αvβ3 integrin-binding site in the α3 chain triple helix adjacent to the NC1 domain [Borza et al., 2008; Pedchenko et al., 2004]. No other epithelial-specific binding sites have been identified.

Neutrophils

Neutrophils bind to the αvβ3 integrin binding site in the α3 NC1 domain, and binding downregulates neutrophil activation and, probably decreases tissue damage as the cells traverse the capillary wall [Monboisse et al., 1994].

Platelets

Platelets adhere to collagens I and III through the α1β1 and α2β1 integrin receptors, and adhesion is enhanced by binding to the glycoprotein VI and VWF receptors. Collagen IV has binding sites for these integrins and glycoprotein VI as well as predicted sites for VWF.

Molecules That Enhance Platelet and Cell Binding

Cell surface proteoglycans

Heparin and heparan sulfate proteoglycan (HSPG) binding sites potentially support cell–collagen IV interactions through binding to cell surface HSPGs such as syndecan and glypican. Some sites have been demonstrated experimentally and others predicted from Cardin and Weintraub [1989] consensus sequences. However, the predicted sequences are probably only active in an α-helix, which is not found in the triple helical regions of collagen IV.

SPARC (osteonectin or BM-40)

This small glycoprotein modulates cell–matrix interactions and collagen assembly [Mayer et al., 1991]. It is essential for embryonic development and may also function as a chaperone. In collagen I, the GVMGFO motif where F is critical for binding [Hohenester et al., 2008] is a common binding site for SPARC, VWF, and the discoidin domain receptor 2 (DDR2), but this site is not found in collagen IV. In collagen IV, SPARC recognizes GFP or GLP [Hohenester et al., 2008] but it is unclear whether VWF and DDR1 (collagen IV binds DDR1 not DDR2) also bind to this motif.

Von Willebrand factor (VWF)

vWF is a large, multimeric molecule that mediates platelet adhesion to collagen, and is a carrier for coagulation factor VIII. The binding motif on collagen III is RGQPGVMGF [Lisman et al., 2006] and in collagen IV similar motifs occur on the α2 (RGQPGVPGVPGMKGD), α1, α4 (RGQPGEMGD), and, possibly, the α3 (RGQPGRKGL) chains. This presumes the homotrimeric structure found in collagen III is not necessary for binding. These motifs do not include the GFP or GLP sequences needed for SPARC binding and, if confirmed, must represent an independent binding mechanism.

Glycoprotein VI

The binding of glycoprotein VI to collagen I tethers and activates platelets prior to the platelet release reaction [Dubois et al., 2006]. The $(GPP)_4$ sequence simultaneously binds and activates two glycoprotein VI molecules [Smethurst et al., 2007]. The α1, α4, and α6 chains each have a single glycoprotein VI binding site but at different locations in the amino, midpoint, or carboxyl terminus, of the triple helix. Only the α1α1α2 heterotrimer has two glycoprotein VI binding sites, and these are at the amino terminus of the α1 triple helix, between binding sites for SPARC and α1β1/α2β1 integrin. This represents a potential platelet binding site. Sites in the other heterotrimers may have other functions such as triple helix nucleation or stabilization.

Binding to Extracellular Matrix Structural Proteins

Collagen IV interacts with laminin, nidogen, and HSPG (mainly perlecan, but also chondroitin and dermatan sulfate, and agrin). Molecules bind at multiple sites sometimes by different mechanisms. The following locations have been determined mainly from rotary shadowing electron microscopy, and some are unconfirmed.

Laminin

Laminin is the major noncollagenous protein found in basement membranes. It forms a distinct network that binds to the collagen networks directly [McKee et al., 2007] or through a nidogen bridge.

Laminin binding to collagen IV has been studied by rotary shadowing in the EHS tumor and there are up to six sites throughout the α1α1α2 heterotrimer (Supp. Table S3) [Aumailley et al., 1989; Charonis et al., 1985; Laurie et al., 1986; Ohno et al., 1991; Rao et al., 1985]. The sites 251–291, 174–178, and 75–87 nm from the NC1 have been confirmed in at least two studies, and potentially overlap with sites for nidogen, HSPG, and fibronectin.

Nidogen ("entactin")

Nidogen is ubiquitous in basement membranes and links the collagen IV and laminin networks [Aumailley et al., 1989]. Only one binding site, 80 nm from the NC1 domain, which is potentially shared with HSPG, has been identified [Aumailley et al., 1989].

HSPG sites

There are two major binding sites for HSPG in the collagen IV triple helix. These are 200–300 nm and 100 nm from the NC1

[Koliakos et al., 1989; Laurie et al., 1986]. A further site in the NC1 domain has the highest affinity and binds preferentially to chondroitin or dermatan sulphate.

Heparin is a glycosaminoglycan with repeating disaccharide subunits of glucosamine and sulfated iduronic or glucuronic acids that represents a structural analog of HSPG. Three potential heparin-binding motifs have been identified in collagen IV [Koliakos et al., 1989]. These are termed Hep- I in the α1 chain (TAGSCLRKFSTM), Hep-II in α2 (LAGSCLARFSTM), and Hep-III in α1 (GEFYFDLRLKGDK). The Hep-III site overlaps with a laminin/HSPG/fibronectin site identified on rotary shadowing. The following sequences are analogous to the Hep-I and Hep-II sites: TLGSCLQRFTTM in α3; LAGSCLPVFSTL in α4; and TAGSCLRRFSTM in α5. These are located in the NC1 domains close to, or overlapping with, integrin-binding sites.

Further potential heparin-binding sites have the sequences XBBXBX and XBBBXXBX, where B are basic and X are hydropathic residues [Cardin and Weintraub, 1989]. They include GRRGKT (residues 830–835) in the α3 chain, and GKRGKP and NKRAHG (residues 296–300 and 1,489–1,495, respectively) in the α5 chain.

Melanoma cell/CD44 receptor

CD44 is a chondroitin sulfate PG that is a receptor on the surface of melanoma cells, and binds to the α1 chain (GVKGDKGDPGYPGAP) [Lauer-Fields et al., 2003].

Other Molecules That Bind to Type IV Collagen

Bone morphogenetic protein 4 (BMP4)

This cytokine is a member of the TGF-β superfamily and regulates vascular endothelium proliferation, differentiation, and survival. It is critical in embryogenesis and vascular remodeling, and in macrophage and T cell responses [Wang et al., 2008]. Its binding motif (Y/FI/VSRCXVCE) appears at the same location within the NC1 domain in all collagen IV chains [Wang et al., 2008]. BMP4 binds heparin [Paralkar et al., 1990] and the sites for BMP are near binding sites for heparin on all the collagen IV chains.

Fibronectin

Fibronectin binding is controversial but rotary shadowing studies suggest a site 205 nm from the NC1 domain on the α1α1α2 heterotrimer at about residue 580 [Laurie et al., 1986]. Fibronectin typically binds via an RGD motif and is enhanced by HSPG [Tarsio et al., 1987]. The proposed location overlaps with a possible HSPG site.

Usherin

Binding to collagen IV occurs at the hinge region between the 7S domain and the triple helix [Bhattacharya et al., 2004] where there are multiple disulfide bonds. It is not clear whether usherin binds to one or all collagen IV chains, but it is found in the same membranes as the α3α4α5 heterotrimer (the stria vascularis of the cochlea and Bruch's membrane in the retina) and has been added her to both the α1α1α2 and α3α4α5 maps. Usher's syndrome results from mutations in the corresponding gene and causes retinitis pigmentosa and hearing loss but not renal disease.

Von Hippel Lindau (VHL) protein

VHL protein acts as a tumor suppressor in two major pathways: the hypoxia-inducible factor (HIF) α and an extracellular matrix pathway. The VHL-HIFα interaction requires HIFα hydroxylation by cytosolic prolyl hydroxylases. In cells with mutant VHL, HIFα accumulates and its targets, VEGF and TGF-α are activated [Kurban et al., 2008]. VHL protein binds to both the α1 and α2 chains of unassembled intracellular collagen IV [Grosfeld et al., 2007]. Binding to the α2 chain is specific and also depends on hydroxylation [Kurban et al., 2008]. The VHL protein interacts with the 70 kDa amino terminal fragment of the α2 chain protruding from the endoplasmic reticulum. This represents a domain at about residues 500–600 near the 3′ hydroxylated residues on the α2 chain. VHL also binds to fibronectin and the proposed location contains a fibronectin-binding site [Ohh et al., 1998]. The potential VHL binding domain corresponds to the region affected by mutations affecting the α1 chain and responsible for HANAC.

Factor IX

The active form of this serine protease hydrolyses and activates Factor X. Factor IX may have a role in coagulation during endothelial membrane rupture. It binds to the α1 chain at residues 985–1,092 and 1,182–1,288, and to the α2 chain at residues 1,030–1,137 and 1,227–1,333 [Cheung et al., 1996; Wolberg et al., 1997]. Mutations in this protein result in hemophilia B.

Prolactin-related protein 1

This glycoprotein is produced by the placenta, binds to the collagen IV 7S domain, and probably acts on cells traversing the placenta [Takahashi et al., 2008].

Many other proteins are also found in the basement membrane and bind to collagen IV but their binding motifs are not known (Table 1).

Functional and Disease-Associated Domains

The collagen IV heterotrimers play critical roles in both physiological and disease states. Digested or denatured collagen fragments may have different roles from the native molecule if functional sites have been destroyed and cryptic sites exposed and activated. The α1α1α2 heterotrimer is the most susceptible of the collagen molecules to proteolysis.

Angiogenesis regulatory domains

Angiogenesis is critical in embryogenesis, and in the adult, in tissue regeneration and wound healing. It depends on the interaction of endothelial cells with extracellular matrix proteins or their fragments, as well as with growth factors, such as the VHL protein.

A major putative angiogenesis regulatory site is present at the amino terminus of the triple helix of the α1 chain. In collagen I, endothelial cell ligation of the α1β1/α2β1 integrin-binding motif, GFOGER, in the triple helix induces angiogenesis [Sweeney et al., 2008] and this motif is also present in the collagen IV α1 chain. It is near binding sites for laminin/HSPG/fibronectin, SPARC, VHL protein, and predicted sites for heparin and VWF. Fragments of the triple helix containing the GFPGER motif inhibit angiogenesis by preventing endothelial cell binding to GFPGER in the native collagen IV.

Table 1. Proteins That Bind to Collagen IV But Where the Binding Site is Unknown

Ligand	Role
Acetylcholinesterase	This molecule supports cell adhesion [Paraoanu and Layer, 2008]. Stress produces a splice variant, "acetylcholinesterase-related peptide," that binds collagen IV and laminin, and inhibits cell adhesion by competing with other forms of acetylcholinerase [Johnson and Moore, 2007]
C1q receptor 1	This molecule is widely expressed on cell surfaces and has a conserved sequence that is homologous with collagen IV but also binds to it [Ghebrehiwet et al., 1992]
Collagen type VII	The noncollagenous domain of collagen VII binds to collagen IV [Chen et al., 1997]. Mutations cause the blistering disease epidermolysis bullosa
Discoidin domain receptor (DDR)	DDR1 and 2 are receptor tyrosine kinases that function as collagen receptors. Collagen IV stimulates DDR1 in the absence of integrins [Vogel et al., 2000] but the relevant motif is not known [Khoshnoodi et al., 2008]. The binding motif on collagen I is common to DDR2, SPARC, and VWF
Disrupted in schizophrenia 1 (DISC1)	This is a multifunctional protein associated with the centrosome and spindle, that binds to many cytoskeletal and signalling receptors and also to collagen IV [Morris et al., 2003]
Extracellular matrix protein 1	This is a secreted glycoprotein that binds to collagen IV [Sercu et al., 2008]
Fibulin 2, 4	This is a family of 5 extracellular matrix proteins found in close association with microfibrils containing fibronectin or fibrillin. Both fibulin -2 and fibulin-4 bind to collagen IV [Kobayashi et al., 2007; Sasaki et al., 1995]
Insulin-like growth factor binding protein 7 (Igfbp7)	Also known as "angiomodulin," interacts with extracellular matrix proteins expressed in most blood vessels, including collagen IV [Nagakubo et al., 2003]
Lymphoid chemokines	The cytokines CCL21, CXCL13, and CXCL12 are secreted by high endothelial venules and play a critical role in lymphoid trafficking [Yang et al., 2007]
Mac-2 binding protein	This is a cell-adhesive protein found in the extracellular matrix [Sasaki et al., 1998]
Matrilins	This family of extracellular adaptor molecules binds to collagen I, fibronectin and the laminin-nidogen complex and possibly also to collagen IV [Mates et al., 2004]
Microfibrillar-associated protein 2	This is the major antigen of elastin-associated microfibrils. It may be affected in inherited connective tissues disease [Finnis and Gibson, 1997]
Myelin-associated glycoprotein	This multifunctional adhesion molecule is found in the central and peripheral nervous system. It binds to fibrillary collagens more avidly than collagen IV possibly through glycosaminoglycans [Fahrig et al., 1987]
Nucleosomes	These comprise nuclear chromatin and proteins, especially histones, and it is unclear why they bind to extracellular matrix proteins [Mjelle et al., 2007]
Oncostatin M	This is a cytokine in the IL6 family and binds to collagen I, III, IV, and VI [Somasundaram et al., 2002]
Plasminogen	This is the precursor of the serine protease plasmin. It binds to the α1 and α2 chains of collagen IV [Stack et al., 1992]
Platelet-derived growth factor	Some extracellular matrix components interact with growth factors and cytokines thus limiting the location of their biological activities [Somasundaram and Schuppan, 1996]
Serpins	These serine protease inhibitors inhibit thrombin, urokinase and plasmin. Some including C′ esterase inhibitor and nexin 1 bind to collagen IV [Donovan et al., 1994]
Serum amyloid A	This is an acute phase protein of unknown function that binds with high affinity to laminin and lower affinity to type IV collagen [Ancsin and Kisilevsky, 1997]
Transforming growth factor β1	This protein is critical in cell proliferation and differentiation and binds to collagen IV [Paralkar et al., 1991]
Thrombospondin 1	This is one of a family of thrombospondins released from platelets during aggregation. It is involved in many biological reactions and bind weakly to collagen IV [Galvin et al., 1987]

The collagen IV NC1 domains also represent major angiogenesis regulatory domains because they have binding sites for endothelial cell integrins, and HSPG/heparin. Although integrin binding sites in the NC1 are angiogenic, the same sites on the fragments produced by, for example, MMP cleavage during membrane turnover, are anti-angiogenic. Thus, the NC1 domains of the α1, α2, α3, and α6 chains of collagen IV that result from proteolysis (sometimes known as "arresten," "canstatin," and "tumstatin" for the α1–α3 chains, respectively) are all anti-angiogenic [Mundel and Kalluri, 2007; Mundel et al., 2008; Petitclerc et al., 2000]. The α1 NC1 domain disrupts angiogenesis through blocking growth factor-dependent endothelial cell growth possibly through effects on the α1β1 integrin and perlecan [Colorado et al., 2000]. The α2 NC1 domain inhibits endothelial cell growth and migration, and induces apoptosis [Kamphaus et al., 2000]. The α3 NC1 domain includes 2 αvβ3 integrin-binding sites, one with anti-angiogenic and one with antitumor properties [Maeshima et al., 2000; Shahan et al., 1999]. One of these NC1 fragments is currently in clinical trials for the treatment of human renal cell carcinoma [Eikesdal et al., 2008].

Hemostasis

The blood vessel wall stroma comprises mainly collagen I and III, and the endothelial basement membrane is predominantly collagen IV α1α1α2. Platelet adhesion under high shear stress depends on the binding of VWF to collagen, and, in turn, on binding to glycoprotein VI and the α2β1 integrin. Platelets bind to collagen I and III, but have only weak affinity for collagen IV. The two glycoprotein VI sites of the α1α1α2 heterotrimer may

contribute to platelet binding, and these sites are close to the putative integrin, SPARC, and VWF sites.

Infections

Adhesion of microbial pathogens to lectin-like sequences on collagen IV is the initial step in tissue colonization and infection. Many bacteria and fungi including *Staphylococcus aureus, Streptoccus pyogenes*, *Escherichia coli, Yersinia enterocolitica, Candida albicans,* and *Agaricus bisporus* bind to collagen IV through a variety of mechanisms including microbial glycoprotein adherence to lectin-binding domains [Alonso et al., 2001; Dinkla et al., 2009; Farfan et al., 2008; Flugel et al., 1994; Kajimura et al., 2004; Vercellotti et al., 1985].

The lectin-binding sites are widely dispersed in the different collagen IV chains. Agaricus bisporus agglutinin binds to the α1 NC1 domain [Kajimura et al., 2004]. *Escherichia coli* binds to the 7S domain of collagen IV in the urinary tract and thus to the α1α1α2 heterotrimer [Selvarangan et al., 2004; Westerlund et al., 1989]. The α2–α5 collagen chains each have a Ca-dependent C-lectin-like domain that overlap in the α3, α4, and α5 chains (Swiss protein Web site). The M3 serotype of *Streptococcus pyogenes* induces glomerulonephritis and rheumatic heart disease and a bacterial "peptide associated with rheumatic fever" ("PARF") binds to placental type IV collagen, 20 and 100 nm from the 7S domain in the α1α1α2 triple helix resulting in subsequent autoantibody production [Dinkla et al., 2009].

Tumor growth and spread

Tumor growth and spread depends on the development of an adequate blood supply and migration through the vascular endothelium. Basement membrane collagen is integral to these activities. Tumor cells adhere to collagen IV through integrins and induce angiogenesis. However, the upregulation of integrins also inhibits tumor cell migration [Bago et al., 2009]. The full-length α3 NC1 domain has no effect on tumor cell growth [Maeshima et al., 2000], but the corresponding synthetic peptide (residues 185–203) that binds to α3β1 and CD47/αvβ3 integrin complex inhibits the proliferation of various epithelial tumor and melanoma cell lines [Han et al., 1997; Maeshima et al., 2000; Shahan et al., 1999]. Surprisingly, recent studies show this peptide also possesses anti-angiogenic activity [Shahan et al., 2004]. Thus, proteolytic degradation of the α3 NC1 may release a cryptic fragment with antitumor activity. The NC1 of the α6 chain also has antitumor activity [Mundel et al., 2008].

Collagen glycation

Glycation is the nonenzymatic binding of glucose to the ε-amino group of lysine, and the subsequent cross-linking of frucosyl-lysine to produce advanced glycation end products. Glycation occurs on many residues but preferentially on hydroxylysine, and is normal in aging and accelerated in diabetes. Glycation of collagen I results in a molecule that is less flexible [Reiser et al., 1992], and has altered binding to cells and ligands including integrins, HSPG and fibronectin [Reigle et al., 2008; Tarsio et al., 1987].

The principal residues affected by glycation in collagen IV are not known except for locations in the 7S and NC1 domains of the α1 and α2 chains [Raabe et al., 1996]. These potentially interfere with hexamer formation, and the binding of laminin and usherin. Glycation interferes with collagen IV assembly in diabetes [Tsilibary et al., 1988], and with digestion by MMP-3 and MMP-9 [Mott et al., 1997], and hence, tissue remodeling. Glycation may also contribute to the delayed wound healing and the increased risk of tumor metastasis seen in diabetes and aging.

Elevated glucose levels in diabetes also produce reactive dicarbonyl species ("carbonyl stress"). One of the major products of glucose degradation, methylglyoxal, specifically reacts with arginine residues in proteins. Arginine is a key residue in most integrin binding sites (RGD, GFOGER, etc.), and modification of collagen IV and its fragments, including RGD-containing fragments of the α3 chain by methylglyoxal, disrupts integrin-mediated cell– matrix interactions [Pedchenko et al., 2005].

Immunoreactive determinants

Epitopes for the autoantibodies in Goodpasture disease, and alloantibodies in Alport posttransplant glomerulonephritis are located within the α3 and α5 NC1 domains [Pedchenko et al., 2010]. The Goodpasture autoepitopes E_A and E_B comprise α3 NC1 residues 17–31 and 127–141 [Kalluri et al., 1991; Netzer et al., 1999) and at the homologous residues to E_A, 17–31, on α5NC1 [Pedchenko et al., 2010]. The Goodpasture T cell epitope overlaps with the E_A epitope [Bolton et al., 2005]. The Goodpasture antigen-binding protein (GPBP) is a nonconventional serine/threonine kinase that phosphorylates the KRGDS motif of the α3 chain located just before the NC1 domain [Raya et al., 1999, 2000; Revert et al., 2008]. It occurs in two spliced forms, the more active of which is present in tissues affected by Goodpasture disease.

About 5% of patients with X-linked Alport syndrome who receive a kidney transplant develop alloantibodies against Alport antigenic sites that they "recognize" immunologically because the α3α4α5 network is absent from their native kidneys. Three alloepitopes have been identified in the NC1 domain of the α5 chain, and two in the linear sequence [Kang et al., 2007]. Other epitopes of Alport alloantibodies with unknown motifs are present in the α3 and α4 chains [Kalluri et al., 2000]. The Goodpasture and Alport epitopes overlap, but the Goodpasture epitopes are sequestered ("cryptic") within the α3α4α5 NC1 hexamer, whereas the Alport epitopes are accessible to alloantibodies, suggesting different key residues [Hudson et al., 2003; Pedchenko et al., 2010].

Another disease, with severe subepidermal bullous eruptions and renal insufficiency, is associated with IgG autoantibodies directed against an unknown epitope in the NC1 domain of the α5 chain [Ghohestani et al., 2000].

HLA DR15 binds to FIMFTSAGS in the NC1 domain and is responsible for the specific B and T cell response [Phelps et al., 1998]. BMP4 binds nearby in the NC1 and also has potentially a role in the macrophage and T cell response in autoantibody-mediated disease.

T lymphocytes bind to a specific site on the α3 chain and this binding is enhanced by lectins [Rabinovich et al., 1999].

Missense Mutations and Clinical Phenotype

Missense mutations resulting in HANAC are limited to the collagen IV α1 chain binding site for VHL and other proteins (integrins, heparin, VWF) involved in angiogenesis. One known mutation affects an integrin-binding site.

VHL syndrome is characterized by hemangioblastoma of the cerebellum, retina, and spinal cord; renal cysts, and clear cell

cancer. Basement membranes in tissues affected by VHL syndrome, including the proximal renal tubule membrane-derived renal cysts and cancer, all comprise the α1α1α2 network. The absence of clinical features from tissues with collagen IV α3α4α5-containing basement membranes suggest the VHL protein does not bind to this network. Inheritance of VHL disease is autosomal dominant but the germline mutation predisposes to a "second hit" and loss of the functional protein that normally directs the α2 chain into the heterotrimer. Both HANAC and VHL are characterized by vascular abnormalities and renal cysts, and HANAC probably results from defective binding of the VHL protein to the α1 chain and subsequent loss of function. Mutations elsewhere in the collagen α1 chain result in vascular stroke and porencephaly.

More than 100 mutations have been described in the α3 and α4 chains in autosomal recessive Alport syndrome and thin basement membrane nephropathy. These occur throughout both chains but are too few to determine randomness or explain genotype–phenotype variation. Of the nine missense mutations resulting in autosomal dominant Alport syndrome, only a cysteine substitution in the α4 chain NC1 domain (C1634S) is likely to have major structural consequences through affecting disulfide bond and globular domain formation [Marcocci et al., 2009].

Two hundred thirty-three missense mutations have been described in the α5 chain in X-linked Alport syndrome. Clinical data are available for 105, 73 of which result in severe disease with "early-onset" renal failure. Glycine substitutions with glutamic acid, valine, or arginine trend to severe disease (22/33, 67% vs. 9/27, 33%, $P = 0.1755$ by Fisher's two-tailed test, Supp. Table S4a and b). Both mutations affecting an integrin-binding site also result in severe disease, but this may be attributed to the nature of the substituting residue (glutamic acid, valine).

The distribution of mutations in the α5 chain is not random. Mutations are more common in exons 25 and 26 ($P = 0.00348$) but do not result in more severe disease and no known ligands bind here. Mutations are underrepresented in exons 1–6 and 42–45 ($P = 0.0006$, 0.00078, respectively). The nonrandom distribution of α5 mutations is not due to a nonviable phenotype because the α3α4α5 network is not present in embryonic life and even α5 chain nonsense mutations in males are not lethal.

No missense mutations are known for the α2 and α6 chains. This may be because the phenotype is too severe to be viable, too mild to come to medical attention, or too rare to be detected.

Alternative Splicing Isoforms

Collagen IV isoforms are found in different tissues and it is interesting to consider the consequences of alternative splicing on the retention of major ligand-binding sites and functional domains. For example, the ?1 isoform is missing the major angiogenesis regulatory. The α1 isoform is missing the major angiogenesis regulatory domain and the site affected by mutations in HANAC syndrome. The α3 isoforms 2–4 lack various of the NC1 immunogenic, endothelial cell binding, antiangiogenic, and antitumor domains. Isoforms of the α5 and α6 chains retain all major ligand-binding sites.

Conclusions

Structural and Functional Domains

Collagen IV is the major constituent of basement membranes, but the predominant heterotrimer and its binding partners and functions depend on individual tissues. Each of the maps described here documented all known ligand-binding sites from tissues as diverse as blood vessels, placenta, muscle, kidney, small bowel, and skin. More ligands were identified for the α1α1α2 network because it is more widespread, more abundant where it occurs, and has been best studied.

Overall, the carboxyl terminal domains had the highest density of ligand-binding sites, and represented major ligand-binding and often functional domains. There were other lesser "hotspots." There were also regions with few or no ligand-binding sites, such as residues 750–900 in the α1 chain. These may have a structural "load-bearing" role or provide space between biologically active regions. Some ligands such as integrins, BMP, and heparin bound at similar locations on different collagen chains.

Integrin-binding sites were most abundant in the α1α1α2 heterotrimer, and endothelial cell binding occurred via the α1β1, α2β1, and αvβ3 integrins throughout the triple helix and NC1 domains of the α1 and α2 chains. These sites enable the vascular endothelium to adhere to the underlying collagen network. Furthermore, the GFOGER motif ensures cell-directed collagen IV assembly into the basement membrane. The integrins also facilitate binding to other cells including tumor cells and platelets. The α3α4α5 heterotrimer has a distinctive epithelial cell binding site in the triple helix near the NC1 of the α3 chain that is not present in other chains. This is particularly relevant for glomerular, alveolar and retinal epithelial cells which rest on a membrane comprising the α3α4α5 network.

The proximity of binding sites for functionally related ligands on individual or nearby collagen chains suggested cooperative interactions. Integrin receptors span up to 10 nm and the glycosaminoglycan side chains of HSPGs extend 20 nm or more [Doyle et al., 1975]. The collagen heterotrimer itself is 1.0–1.5 nm wide [Trus and Piez, 1980] and ligands binding to vertically aligned sites may also interact [Di Lullo et al., 2002]. The collagen networks represent scaffolds for the clustering of ligands binding to the triple helix, and the NC1 domains form intermolecular aggregates that bring together many ligands. Cooperative binding potentially occurred: between integrin sites, especially in the NC1 domains; between integrins and HSPG side chains, or MMPs; between SPARC and glycoprotein VI; and between the Goodpasture antigen, T cells, HLA DR15 and BMP4.

Modeling the Collagen Type IV Scaffold

The collagen IV maps suggest an orientation with respect to the endothelial and epithelial membrane surfaces in vivo. The collagen IV molecule is 400 nm long with 57 nm between the 7S domain and kink, and 340 nm between the kink and NC1 domain, but basement membranes are typically only 50–300 nm wide. The collagen IV triple helical domains are too long to allow the monomer to lie perpendicular to the outer membrane margin but the noncollagenous interruptions and supercoiling confer some flexibility. Our observation that collagen IV heterotrimers demonstrate polarity with respect to cell and ligand interactions suggests a model for orientation. Thus, in the α1α1α2 heterotrimer, the GFOGER motif at the amino terminus represents the major binding site for vascular endothelium. The kink allows the N-terminal 57 nm of the triple helix to lie flat against the endothelial surface of the basement membrane, and the neighboring 7S domains to self-associate and covalently crosslink. The kink enables the molecule to span, at an acute angle, from the endothelium to the epithelium. In cross-section the molecule is "accordion-like" and potentially stabilized by interactions with other extracellular matrix molecules at different levels between the

membrane margins. The major binding site for epithelial cells is the αvβ3 motif in the triple helix near the α3 NC1 domain of the α3α4α5 heterotrimer. This location also allows the N-terminal triple helix to associate near the endothelial surface.

The collagen IV interactomes have been based on ligand interactions with native collagen, at various stages of denaturation, or with synthetic triple helices or peptides. Binding was detected with low resolution rotary shadowing or high-resolution methods. The maps' major limitations were that they did not indicate critical physicochemical characteristics that might be apparent on space filling and other multidimensional maps, and that they did not demonstrate the effect of ligand binding on collagen IV physicostructural properties such as flexibility and elasticity using techniques like "optical tweezers" and molecular dynamics measured by coarse grain simulation. Molecular level properties such as chemistry and nanomechanics require a more sophisticated approach, such as multiscale modeling, and more powerful layering than has been possible here. Nevertheless, there is already evidence that mutations in Alport syndrome alter not only the molecular structure but also nanomechanical properties [Srinivasan et al., 2009].

Future collagen IV interactomes are likely to be specific for tissues and developmental stage, and binding sites will be confirmed using high-resolution methods. These linear protein maps indicate the functional domains responsible for collagen IV's diverse biological activities, and potentially facilitate the development of antagonists for these activities through targeting the corresponding domains.

Acknowledgments

This work was supported in part by NIH Grants DK18381 and DK065123 (to B.G.H.) and by grants from the National Health and Medical Research Council of Australia and Kidney Health Australia (to J.S.). The authors have declared no conflict of interest.

References

Alonso R, Llopis I, Flores C, Murgui A, Timoneda J. 2001. Different adhesins for type IV collagen on *Candida albicans*: identification of a lectin-like adhesin recognizing the 7S(IV) domain. Microbiology 147:1971–1981.

Ancsin JB, Kisilevsky R. 1997. Characterization of high affinity binding between laminin and the acute-phase protein, serum amyloid A. J Biol Chem 272: 406–413.

Asada S, Koide T, Yasui H, Nagata K. 1999. Effect of HSP47 on prolyl 4-hydroxylation of collagen model peptides. Cell Struct Funct 24:187–196.

Aumailley M, Wiedemann H, Mann K, Timpl R. 1989. Binding of nidogen and the laminin-nidogen complex to basement-membrane collagen type-IV. Eur J Biochem 184:241–248.

Bago R, Pavelic J, Vlahovicek GM, Bosnar MH. 2009. Nm23-H1 promotes adhesion of CAL 27 cells in vitro. Mol Carcinogenesis 48:779–789.

Bhattacharya G, Kalluri R, Orten DJ, Kimberling WJ, Cosgrove D. 2004. A domain-specific usherin/collagen IV interaction may be required for stable integration into the basement membrane superstructure. J Cell Sci 117:233–242.

Bolton WK, Chen L, Hellmark T, Fox J, Wieslander J. 2005. Molecular mapping of the Goodpasture's epitope for glomerulonephritis. Trans Am Clin Climatol Assoc 116:229–236.

Borza CM, Borza DB, Pedchenko V, Saleem MA, Mathieson PW, Sado Y, Hudson HM, Pozzi A, Saus J, Abrahamson DR, Zent R, Hudson BG. 2008. Human podocytes adhere to the KRGDS motif of the α3α 4α5 collagen IV network. J Am Soc Nephrol 19:677–684.

Borza CM, Pozzi A, Borza DB, Pedchenko V, Hellmark T, Hudson BG, Zent R. 2006. Integrin α3β1, a novel receptor for α3(IV) noncollagenous domain and a trans-dominant inhibitor for integrin αvβ3. J Biol Chem 281:20932–20939.

Byers PH, Wallis GA, Willing MC. 1991. Osteogenesis imperfecta—translation of mutation to phenotype. J Med Genet 28:433–442.

Canty EG, Kadler KE. 2005. Procollagen trafficking, processing and fibrillogenesis. J Cell Sci 118:1341–1353.

Cardin AD, Weintraub HJR. 1989. Molecular modeling of protein–glycosaminoglycan interactions. Arteriosclerosis 9:21–32.

Charonis AS, Tsilibary EC, Yurchenco PD, Furthmayr H. 1985. Binding of laminin to type-IV collagen—a morphological study. J Cell Biol 100:1848–1853.

Chen M, Marinkovich MP, Veis A, Cai XY, Rao CN, Otoole EA, Woodley DT. 1997. Interactions of the amino-terminal noncollagenous (NC1) domain of type VII collagen with extracellular matrix components—a potential role in epidermal-dermal adherence in human skin. J Biol Chem 272:14516–14522.

Cheung WF, van Den Born J, Kuhn K, Kjellen L, Hudson BG, Stafford DW. 1996. Identification of the endothelial cell binding site for factor IX. Proc Nat Acad Sci USA 93:11068–11073.

Colorado PC, Torre A, Kamphaus G, Maeshima Y, Hopfer H, Takahashi K, Volk R, Zamborsky ED, Herman S, Sarkar PK, Ericksen MB, Dhanabal M, Simons M, Post M, Kufe DW, Weichselbaum RR, Sukhatme VP, Kalluri R. 2000. Anti-angiogenic cues from vascular basement membrane collagen. Cancer Res 60: 2520–2526.

Deng SJ, Bickett DM, Mitchell JL, Lambert MH, Blackburn RK, Carter HL, Neugebauer J, Pahel G, Weiner MP, Moss ML. 2000. Substrate specificity of human collagenase 3 assessed using a phage-displayed peptide library. J Biol Chem 275:31422–31427.

Di Lullo GA, Sweeney SM, Korkko J, Ala-Kokko L, San Antonio JD. 2002. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. J Biol Chem 277:4223–4231.

Dinkla K, Talay S, Morgelin M, Graham R, Rohde M, Nitsche-Schmitz D, Chhatwal G. 2009. Crucial role of the CB3-region of collagen IV in PARF-induced acute rheumatic fever. PLOS One 4:e4666.

Donovan FM, Vaughan PJ, Cunningham DD. 1994. Regulation of protease NEXIN-1 target protease specificity by collagen type-IV. J Biol Chem 269:17199–17205.

Doyle BB, Hukins DWL, Hulmes DJS, Miller A, Woodheadgallowa J. 1975. Collagen polymorphism—its origins in amino-acid sequence. J Mol Biol 91:79.

Dubois C, Panicot-Dubois L, Merrill-Skoloff G, Furie B, Furie BC. 2006. Glycoprotein VI-dependent and -independent pathways of thrombus formation in vivo. Blood 107:3902–3906.

Eble JA, Golbik R, Mann K, Kuhn K. 1993. The α1β1 integrin recognition site of the basement membrane collagen molecule $[\alpha1(IV)]_2\alpha2(IV)$. EMBO J 12:4795–4802.

Eikesdal HP, Sugimoto H, Birrane G, Maeshima Y, Cooke VG, Kieran M, Kalluri R. 2008. Identification of amino acids essential for the antiangiogenic activity of tumstatin and its use in combination antitumor activity. Proc Nat Acad Sci USA 105:15040–15045.

Engel J, Prockop DJ. 1991. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. Annu Rev Biophys Biophysical Chem 20:137–152.

Fahrig T, Landa C, Pesheva P, Kuhn K, Schachner M. 1987. Characterization of binding-properties of the myelin-associated glycoprotein to extracellular-matrix constituents. EMBO J 6:2875–2883.

Farfan MJ, Inman KG, Nataro JP. 2008. The major pilin subunit of the AAF/II fimbriae from enteroaggregative *Escherichia coli* mediates binding to extracellular matrix proteins. Infect Immunity 76:4378–4384.

Finnis ML, Gibson MA. 1997. Microfibril-associated glycoprotein-1 (MAGP-1) binds to the pepsin-resistant domain of the α3(VI) chain of type VI collagen. J Biol Chem 272:22817–22823.

Flugel A, Schulzekoops H, Heesemann J, Kuhn K, Sorokin L, Burkhardt H, Vondermark K, Emmrich F. 1994. Interaction of enteropathogenic *Yersinia enterocolitica* with complex basement-membranes and the extracellular-matrix proteins collagen type-IV, laminin-1 and laminin-2, and nidogen/entactin. J Biol Chem 269:29732–29738.

Galvin NJ, Vance PM, Dixit VM, Fink B, Frazier WA. 1987. Interaction of human thrombospondin with type IV collagen—direct binding and electron-microscopy. J Cell Biol 104:1413–1422.

Garnero P, Borel O, Byrjalsen I, Ferreras M, Drake FH, McQueney MS, Foged NT, Delmas PD, Delaisse JM. 1998. The collagenolytic activity of cathepsin K is unique among mammalian proteinases. J Biol Chem 273:32347–32352.

Gautieri A, Uzel S, Vesentini S, Redaelli A, Buehler MJ. 2009. Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils. Biophysical J 97:857–865.

Ghebrehiwet B, Peerschke EIB, Hong YQ, Munoz P, Gorevic PD. 1992. Short amino-acid-sequences derived from C1Q-Receptor (C1Q-R) show homology with the α-chains of fibronectin and vitronectin receptors and collagen type-IV. J Leukocyte Biol 51:546–556.

Ghohestani RF, Hudson BG, Claudy A, Uitto J. 2000. The α5 chain of type IV collagen is the target of IgG autoantibodies in a novel autoimmune disease with subepidermal blisters and renal insufficiency. J Biol Chem 275:16002–16006.

Gioia M, Monaco S, Van Den Steen PE, Sbardella D, Grasso G, Marini S, Overall CM, Opdenakker G, Coletta M. 2009. The collagen binding domain of gelatinase A modulates degradation of collagen IV by gelatinase B. J Mol Biol 386:419–434.

Gould DB, Phalan FC, Breedveld GJ, van Mil SE, Smith RS, Schimenti JC, Aguglia U, van der Knaap MS, Heutink P, John SWM. 2005. Mutations in *COL4A1* cause perinatal cerebral hemorrhage and porencephaly. Science 308:1167–1171.

Grosfeld A, Stolze IP, Cockman ME, Pugh CW, Edelmann M, Kessler B, Bullock AN, Ratcliffe PJ, Masson N. 2007. Interaction of hydroxylated collagen IV with the von Hippel-Lindau tumor suppressor. J Biol Chem 282:13264–13269.

Gross O, Netzer KO, Lambrecht R, Seibold S, Weber M. 2002. Meta-analysis of genotype–phenotype correlation in X-linked Alport syndrome: impact on clinical counselling. Nephrol Dial Transplant 17:1218–1227.

Han J, Ohno N, Pasco S, Monboisse JC, Borel JP, Kefalides NA. 1997. A cell binding domain from the α3 chain of type IV collagen inhibits proliferation of melanoma cells. J Biol Chem 272:20395–20401.

Herbst TJ, McCarthy JB, Tsilibary EC, Furcht LT. 1988. Differential-effects of laminin, intact type-IV collagen, and specific domains of type-IV collagen on endothelial-cell adhesion and migration. J Cell Biol 106:1365–1373.

Hodivala-Dilke KM, DiPersio CM, Kreidberg JA, Hynes RO. 1998. Novel roles for α3β1 integrin as a regulator of cytoskeletal assembly and as a trans-dominant inhibitor of integrin receptor function in mouse keratinocytes. J Cell Biol 142: 1357–1369.

Hohenester E, Sasaki T, Giudici C, Farndale RW, Bachinger HP. 2008. Structural basis of sequence-specific collagen recognition by SPARC. Proc Natl Acad Sci USA 105:18273–18277.

Hostikka SL, Tryggvason K. 1988. The complete primary structure of the α2 chain of human type-IV collagen chain and comaparison with the α1 (IV) chain. J Biol Chem 263:19488–19493.

Hudson BG, Reeders ST, Tryggvason K. 1993. Type-IV collagen—structure, gene organization, and role in human-diseases—molecular-basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. J Biol Chem 268:26033–26036.

Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG. 2003. Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med 348 2543–2556.

Hyde TJ, Bryan MA, Brodsky B, Baum J. 2006. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. J Biol Chem 281:36937–36943.

Jais JP, Knebelmann B, Giatras I, De Marchi M, Rizzoni G, Renieri A, Weber M, Gross O, Netzer KO, Flinter F, Pirson Y, Dahan K, Wieslander J, Persson U, Tryggvason K, Martin P, Hertz JM, Schröder C, Sanak M, Carvalho MF, Saus J, Antignac C, Smeets H, Gubler MC. 2000. X-linked Alport syndrome: natural history in 195 families and genotype–phenotype correlations in males. J Am Soc Nephrol 11:649–657.

Johnson G, Moore SW. 2007. Acetylcholinesterase readthrough peptide shares sequence similarity to the 28–53 peptide sequence of the acetylcholinesterase adhesion-mediating site and competes for ligand binding in vitro. J Mol Neurosci 31:113–125.

Kajimura D, Takahashi S, Yoshikawa K, Hattori S, Sado Y, Imamura Y, Hayashi T. 2004. Non-helical type IV collagen polypeptides in human placenta. Biochem Biophys Res Commun 314:11–16.

Kalluri R, Gunwar S, Reeders ST, Morrison KC, Mariyama M, Ebner KE, Noelken ME, Hudson BG. 1991. Goodpasture syndrome—localization of the epitope for the autoantibodies to the carboxyl-terminal region of the α3 (IV) chain of basement-membrane collagen. J Biol Chem 266:24018–24024.

Kalluri R, Torre A, Shield CF, Zamborsky ED, Werner MC, Suchin E, Wolf G, Helmchen UM, van den Heuvel L, Grossman R, Aradhye S, Neilson EG. 2000. Identification of α3, α4 and α5 chains of type IV collagen as alloantigens for Alport posttransplant antiglomerular basement membrane antibodies. Transplantation 69:679–683.

Kamphaus GD, Colorado PC, Panka DJ, Hopfer H, Ramchandran R, Torre A, Maeshima Y, Mier JW, Sukhatme VP, Kalluri R. 2000. Canstatin, a novel matrix-derived inhibitor of angiogenesis and tumor growth. J Biol Chem 275: 1209–1215.

Kang JS, Kashtan CE, Turner AN, Heidet L, Hudson BG, Borza DB. 2007. The alloantigenic sites of α3α4α5(IV) collagen. J Biol Chem 282:10670–10677.

Kern A, Eble J, Golbik R, Kuhn K. 1993. Interaction of type-IV collagen with the isolated integrin-α1β1 and integrin-α2β1. Eur J Biochem 215:151–159.

Khoshnoodi J, Cartailler JP, Alvares K, Veis A, Hudson BG. 2006a. Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. J Biol Chem 281:38117–38121.

Khoshnoodi J, Pedchenko V, Hudson BG. 2008. Mammalian collagen IV. Microsc Res Tech 71:357–370.

Khoshnoodi J, Sigmundsson K, Cartailler JP, Bondar O, Sundaramoorthy M, Hudson BG. 2006b. Mechanism of chain selection in the assembly of collagen IV—a prominent role for the α2 chain. J Biol Chem 281:6058–6069.

Kim JP, Chen JD, Wilke MS, Schall TJ, Woodley DT. 1994. Human keratinocyte migration on type-IV collagen-roles of heparin-binding site and α2β1 integrin. Lab Invest 71:401–408.

Knight CG, Morton LF, Onley DJ, Peachey AR, Messent AJ, Smethurst PA, Tuckwell DS, Farndale RW, Barnes MJ. 1998. Identification in collagen type I of an integrin α2β1-binding site containing an essential GER sequence. J Biol Chem 273:33287–33294.

Knight CG, Morton LF, Peachey AR, Tuckwell DS, Farndale RW, Barnes MJ. 2000. The collagen-binding A-domains of integrins α1β1 and α2β1 recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. J Biol Chem 275:35–40.

Kobayashi N, Kostka G, Garbe JHO, Keene DR, Bachinger HP, Hanisch FG, Markova D, Tsuda T, Timpl R, Chu ML, Sasaki T. 2007. A comparative analysis of the fibulin protein family—biochemical characterization, binding interactions, and tissue localization. J Biol Chem 282:11805–11816.

Koide T, Nishikawa Y, Asada S, Yamazaki CM, Takahara Y, Homma DL, Otaka A, Ohtani K, Wakamiya N, Nagata K, Kitagawa K. 2006. Specific recognition of the collagen triple helix by chaperone HSP47 II. The HSP47-binding structural motif in collagens and related proteins. J Biol Chem 281:11177–11185.

Koliakos GG, Kouzikoliakos K, Furcht LT, Reger LA, Tsilibary EC. 1989. The binding of heparin to type-IV collagen-domain specificity with identification of peptide sequences from the α1(IV) and α2(IV) which preferentially bind heparin. J Biol Chem 264:2313–2323.

Kurban G, Duplan E, Ramlal N, Hudon V, Sado Y, Ninomiya Y, Pause A. 2008. Collagen matrix assembly is driven by the interaction of von Hippel-Lindau tumor suppressor protein with hydroxylated collagen IV α2. Oncogene 27: 1004–1012.

Lauer-Fields JL, Malkar NB, Richet G, Drauz K, Fields GB. 2003. Melanoma cell CD44 interaction with the α1(IV)1263-1277 region from basement membrane collagen is modulated by ligand glycosylation. J Biol Chem 278: 14321–14330.

Laurie GW, Bing JT, Kleinman HK, Hassell JR, Aumailley M, Martin GR, Feldmann RJ. 1986. Localization of binding-sites for laminin, heparan-sulfate proteoglycan and fibronectin on basement-membrane (type-IV) collagen. J Mol Biol 189:205–216.

Leitinger B, Hohenester E. 2007. Mammalian collagen receptors. Matrix Biol 26: 146–155.

Lisman T, Raynal N, Groeneveld D, Maddox B, Peachey AR, Huizinga EG, de Groot PG, Farndale RW. 2006. A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. Blood 108:3753–3756.

Maeshima Y, Colorado PC, Torre A, Holthaus KA, Grunkemeyer JA, Ericksen MB, Hopfer H, Xiao YW, Stillman IE, Kalluri R. 2000. Distinct antitumor properties of a type IV collagen domain derived from basement membrane. J Biol Chem 275:21340–21348.

Maeshima Y, Sudhakar A, Lively JC, Ueki K, Kharbanda S, Kahn CR, Sonenberg N, Hynes RO, Kalluri R. 2002. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. Science 295:140–143.

Marcocci E, Uliana V, Bruttini M, Artuso R, Silengo MC, Zerial M, Bergesio F, Amoroso A, Savoldi S, Pennesi M, Giachino D, Rombolà G, Fogazzi GB, Rosatelli C, Martinhago CD, Carmellini M, Mancini R, Di Costanzo G, Longo I, Renieri A, Mari F. 2009. Autosomal dominant Alport syndrome: molecular analysis of the *COL4A4* gene and clinical outcome. Nephrol Dials Transplant 24: 1464–1471.

Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Korkko J, Prockop DJ, De Paepe A, Coucke P, Symoens S, Glorieux FH, Roughley PJ, Lund AM, Kuurila-Svahn K, Hartikka H, Cohn DH, Krakow D, Mottes M, Schwarze U, Chen D, Yang K, Kuslich C, Troendle J, Dalgleish R, Byers PH. 2007. Consortium for osteogeneslis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Hum Mutat 28: 209–221.

Marneros AG, Olsen BR. 2001. The role of collagen-derived proteolytic fragments in angiogenesis. Matrix Biol 20:337–345.

Martinek N, Shahab J, Sodek J, Ringuette M. 2007. Is SPARC an evolutionarily conserved collagen chaperone? J Dent Res 86:296–305.

Mates L, Nicolae C, Morgelin M, Deak F, Kiss I, Aszodi A. 2004. Mice lacking the extracellular matrix adaptor protein matrilin-2 develop without obvious abnormalities. Matrix Biol 23:195–204.

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P. 2009. Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res 37:D619–D622.

Mayer U, Aumailley M, Mann K, Timpl R, Engel J. 1991. Calcium-dependent binding of basement-membrane protein BM-40 (osteonectin, SPARC) to basement-membrane collagen type-IV. Eur J Biochem 198:141–150.

McKee KK, Harrison D, Capizzi S, Yurchenco PD. 2007. Role of laminin terminal globular domains in basement membrane assembly. J Biol Chem 282: 21437–21447.

Miles AJ, Knutson JB, Skubitz APN, Furcht LT, McCarthy JB, Fields GB. 1995. A peptide model of basement-membrane collagen α1(IV) 531–543 binds the α3β1 integrin. J Biol Chem 270:29047–29050.

Mjelle JE, Rekvig OP, Fenton KA. 2007. Nucleosomes possess a high affinity for glomerular laminin and collagen IV and bind nephritogenic antibodies in murine lupus-like nephritis. Ann Rheum Dis 66:1661–1668.

Monboisse JC, Garnotel R, Bellon G, Ohno N, Perreau C, Borel JP, Kefalides NA. 1994. The α3 chain of type IV collagen prevents activation of human polymorphonuclear leukocytes. J Biol Chem 269:25475–25482.

Morris JA, Kandpal G, Ma L, Austin CP. 2003. DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. Hum Mol Genet 12:1591–1608.

Mott JD, Khalifah RG, Nagase H, Shield CF, Hudson JK, Hudson BG. 1997. Nonenzymatic glycation of type IV collagen and matrix metalloproteinase susceptibility. Kidney Int 52:1302–1312.

Mundel TM, Kalluri R. 2007. Type IV collagen-derived angiogenesis inhibitors. Microvasc Res 74:85–89.

Mundel TM, Yliniemi AM, Maeshima Y, Sugimoto H, Kieran M, Kalluri R. 2008. Type IV collagen α6 chain-derived non-collagenous domain 1 (α6(IV) NC1) inhibits angiogenesis and tumor growth. Int J Cancer 122:1738–1744.

Myllyharju J, Kivirikko KI. 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20:33–43.

Nagakubo D, Murai T, Tanaka T, Usui T, Matsumoto M, Sekiguchi K, Miyasaka M. 2003. A high endothelial venule secretory protein mac25/angiomodulin, interacts with multiple high endothelial venule-associated molecules including chemokines. J Immunol 171:553–561.

Netzer KO, Leinonen A, Boutaud A, Borza DB, Todd P, Gunwar S, Langeveld JPM, Hudson BG. 1999. The Goodpasture autoantigen—mapping the major conformational epitope(s) of α3(IV) collagen to residues 17–31 and 127–141 of the NC1 domain. J Biol Chem 274:11267–11274.

Netzer KO, Suzuki K, Itoh Y, Hudson BG, Khalifah RG. 1998. Comparative analysis of the noncollagenous NC1 domain of type IV collagen: Identification of structural features important for assembly, function, and pathogenesis. Protein Sci 7: 1340–1351.

Nosaka AY, Kanaori K, Teno N, Togame H, Inaoka T, Takai M, Kokubo T. 1999. Conformational studies on the specific cleavage site of type I collagen (α1) fragment (157–192) by cathepsins K and L by proton NMR spectroscopy. Bioorg Med Chem 7:375–379.

Nyberg P, Xie L, Sugimoto H, Colorado P, Sund M, Holthaus K, Sudhakar A, Salo T, Kalluri R. 2008. Characterization of the anti-angiogenic properties of arresten, an α1β1 integrin-dependent collagen-derived tumor suppressor. Exp Cell Res 314:3292–3305.

Ohh M, Yauch RL, Lonergan KM, Whaley JM, Stemmer-Rachamimov AO, Louis DN, Gavin BJ, Kley N, Kaelin WG, Iliopoulos O. 1998. The von Hippel-Lindau tumor suppressor protein is required for proper assembly of an extracellular fibronectin matrix. Mol Cell 1:959–968.

Ohno M, Ohno N, Kefalides NA. 1991. Studies on human laminin and laminin–collagen complexes. Connect Tissue Res 25:251–263.

Paralkar VM, Nandedkar AKN, Pointer RH, Kleinman HK, Reddi AH. 1990. Interaction of osteogenin, a heparin binding bone morphogenetic protein, with type-IV collagen. J Biol Chem 265:17281–17284.

Paralkar VM, Vukicevic S, Reddi AH. 1991. Transforming growth-factor-β type-1 binds to collagen-IV of basement-membrane matrix—implications for development. Dev Biol 143:303–308.

Paraoanu LE, Layer PG. 2008. Acetylcholinesterase in cell adhesion, neurite growth and network formation. FEBS J 275:618–624.

Pedchenko V, Bondar O, Fogo AB, Vanacore R, Voziyan P, Kitching AR, Wieslander J, Kashtan CX, Borza DB, Neilson EG, Wilson CB, Hudson BG. 2010. Molecular architecture of the Goodpasture autoantigen in antiGBM nephritis. N Engl J Med 363:343–354.

Pedchenko V, Zent R, Hudson BG. 2004. αvβ3 and αvβ5 integrins bind both the proximal RGD site and non-RGD motifs within noncollagenous (NC1) domain of the α3 chain of type IV collagen—implication for the mechanism of endothelial cell adhesion. J Biol Chem 279:2772–2780.

Pedchenko VK, Chetyrkin SV, Chuang P, Ham AJL, Saleem MA, Mathieson PW, Hudson BG, Voziyan PA. 2005. Mechanism of perturbation of integrin-mediated cell–matrix interactions by reactive carbonyl compounds and its implication for pathogenesis of diabetic nephropathy. Diabetes 54:2952–2960.

Persikov AV, Pillitteri RJ, Amin P, Schwarze U, Byers PH, Brodsky B. 2004. Stability related bias in residues replacing glyclines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. Hum Mutat 24:330–337.

Petitclerc E, Boutaud A, Prestayko A, Xu JS, Sado Y, Ninomiya Y, Sarras MP, Hudson BG, Brooks PC. 2000. New functions for non-collagenous domains of human collagen type IV—novel integrin ligands inhibiting angiogenesis and tumor growth in vivo. J Biol Chem 275:8051–8061.

Phelps RG, Jones VL, Coughlan M, Turner AN, Rees AJ. 1998. Presentation of the Goodpasture autoantigen to CD4 T cells is influenced more by processing constraints than by HLA class II peptide binding preferences. J Biol Chem 273: 11440–11447.

Pietz K, Reddi AH. 1984. Extracellular matrix biochemisty. New York: Elsevier.

Plaisier E, Gribouval O, Alamowitch S, Mougenot B, Prost C, Verpont MC, Marro B, Desmettre T, Cohen SY, Roullet E, Dracon M, Fardeau M, Van Agtmael T, Kerjaschki D, Antignac C, Ronco P. 2007. *COL4A1* mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps. N Engl J Med 357: 2687–2695.

Raabe HM, Molsen H, Mlinaric SM, Acil Y, Sinnecker GHG, Notbohm H, Kruse K, Muller PK. 1996. Biochemical alterations in collagen IV induced by in vitro glycation. Biochem J 319:699–704.

Rabinovich GA, Ariel A, Hershkovitz R, Hirabayashi J, Kasai KI, Lider O. 1999. Specific inhibition of T-cell adhesion to extracellular matrix and proinflammatory cytokine secretion by human recombinant galectin-1. Immunol 97: 100–106.

Rao CN, Margulies IMK, Liotta LA. 1985. Binding domain for laminin on type-IV collagen. Biochem Biophys Res Commun 128:45–52.

Rao NV, Wehner NG, Marshall BC, Gray WR, Gray BH, Hoidal JR. 1991. Characterization of proteinase-3 (PR-3), a neutrophil serine proteinase—structural and functional-properties. J Biol Chem 266:9540–9548.

Raya A, Revert-Ros F, Martinez-Martinez P, Navarro S, Rosello E, Vieitas B, Granero F, Forteza J, Saus J. 2000. Goodpasture antigen-binding protein, the kinase that phosphorylates the Goodpasture antigen, is an alternatively spliced variant implicated in autoimmune pathogenesis. J Biol Chem 275:40392–40399.

Raya A, Revert F, Navarro S, Saus J. 1999. Characterization of a novel type of serine/threonine kinase that specifically phosphorylates the human Goodpasture antigen. J Biol Chem 274:12642–12649.

Reigle KL, Di Lullo G, Turner KR, Last JA, Chervoneva I, Birk DE, Funderburgh JL, Elrod E, Germann MW, Surber C, Sanderson RD, San Antonio JD. 2008. Non-enzymatic glycation of type I collagen diminishes collagen-proteoglycan binding and weakens cell adhesion. J Cell Biochem 104:1684–1698.

Reiser KM, Amigable MA, Last JA. 1992. Nonenzymatic glycation of type-1 collagen—the effects of aging on preferential glycation sites. J Biol Chem 267: 24207–24216.

Renner C, Sacca B, Moroder L. 2004. Synthetic heterotrimeric collagen peptides as mimics of cell adhesion sites of the basement membrane. Peptide Sci 76: 34–47.

Revert F, Ventura I, Martinez-Martinez P, Granero-Molto F, Revert-Ros F, Macias J, Saus J. 2008. Goodpasture antigen-binding protein is a soluble exportable protein that interacts with type IV collagen identification of novel membrane-bound isoforms. J Biol Chem 283:30246–30255.

Sasaki T, Brakebusch C, Engel J, Timpl R. 1998. Mac-2 binding protein is a cell-adhesive protein of the extracellular matrix which self-assembles into ring-like structures and binds β1 integrins, collagens and fibronectin. EMBO J 17: 1606–1613.

Sasaki T, Gohring W, Pan TC, Chu ML, Timpl R. 1995. Binding of mouse and human fibulin-2 to extracellular-matrix ligands. J Mol Biol 254:892–899.

Saus J, Wieslander J, Langeveld JPM, Quinones S, Hudson BG. 1988. Identification of the Goodpasture antigen as the α3(IV) chain of collagen-IV. J Biol Chem 263: 13374–13380.

Scott JE, Tenni R. 1997. Osteogenesis imperfecta mutations may probe vital functional domains (eg proteoglycan binding sites) of type 1 collagen fibrils. Cell Biochem Function 15:283–286.

Selvarangan R, Goluszko P, Singhal J, Carnoy C, Moseley S, Hudson B, Nowicki S, Nowicki B. 2004. Interaction of Dr adhesin with collagen type IV is a critical step in *Escherichia coli* renal persistence. Infect Immun 72:4827–4835.

Sercu S, Zhang M, Oyama N, Hansen U, Ghalbzouri AEL, Jun G, Geentjens K, Zhang LR, Merregaert JH. 2008. Interaction of extracellular matrix protein 1 with extracellular matrix components: ECM1 is a basement membrane protein of the skin. J Invest Dermatol 128:1397–1408.

Shahan T, Grant D, Tootell M, Ziaie Z, Ohno N, Mousa S, Mohamad S, Delisser H, Kefalides N. 2004. Oncothanin, a peptide from the α3 chain of type IV collagen, modifies endothelial cell function and inhibits angiogenesis. Connect Tissue Res 45:151–163.

Shahan TA, Ziaie Z, Pasco S, Fawzi A, Bellon G, Monboisse JC, Kefalides NA. 1999. Identification of CD47/integrin-associated protein and αvβ3 as two receptors for the α3(IV) chain of type IV collagen on tumor cells. Cancer Res 59: 4584–4590.

Sibon I, Coupry I, Menegon P, Boucher JP, Gorry P, Burgelin I, Calvas P, Orignac I, Dousset V, Lacombe D, Orgogozo J-M, Arveiler B, Goizet C. 2007. *COL4A1*

mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. Ann Neurol 62:177–184.

Siebold B, Deutzmann R, Kuhn K. 1988. The arrangement of intramolecular and intermolecular disulfide bonds in the carboxyterminal, non-collagenous aggregation and cross-linking domain of basement membrane type-IV collagen. Eur J Biochem 176:617–624.

Siljander PRM, Hamaia S, Peachey AR, Slatter DA, Smethurst PA, Ouwehand WH, Knight CG, Farndale RW. 2004. Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. J Biol Chem 279: 47763–47772.

Smethurst PA, Onley DJ, Jarvis GE, O'Connor MN, Knight CG, Herr AB, Ouwehand WH, Farndale RW. 2007. Structural basis for the platelet-collagen interaction—the smallest motif within collagen that recognizes and activates platelet glycoprotein VI contains two glycine-proline-hydroxyproline triplets. J Biol Chem 282:1296–1304.

Soderhall C, Marenholz I, Kerscher T, Ruschendorf F, Esparza-Gordillo J, Worm M, Gruber C, Mayr G, Albrecht M, Rohde K, Schulz H, Wahn U, Hubner N, Lee YA. 2007. Variants in a novel epidermal collagen gene (*COL29A1*) are associated with atopic dermatitis. PLOS Biol 5:1952–1961.

Somasundaram R, Ruehl M, Schaefer B, Schmid M, Ackermann R, Riecken EO, Zeitz M, Schuppan D. 2002. Interstitial collagens I, III, and VI sequester and modulate the multifunctional cytokine oncostatin M. J Biol Chem 277: 3242–3246.

Somasundaram R, Schuppan D. 1996. Type I, II, III, IV, V, and VI collagens serve as extracellular ligands for the isoforms of platelet-derived growth factor (AA, BB, and AB). J Biol Chem 271:26884–26891.

Somerville RPT, Oblander SA, Apte SS. 2003. Matrix metalloproteinases: old dogs with new tricks. Genome Biol 4:11.

Spiro RG. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12:43R–56R.

Srinivasan M, Uzel SGM, Gautieri A, Keten S, Buehler MJ. 2009. Alport syndrome mutations in type IV tropocollagen alter molecular structure and nanomechanical properties. J Struct Biol 168:503–510.

Stack MS, Moser TL, Pizzo SV. 1992. Binding of human plasminogen to basement-membrane (type-IV) collagen. Biochem J 284:103–108.

Sudhakar A, Boosani CS. 2008. Inhibition of tumor angiogenesis by tumstatin: insights into signaling mechanisms and implications in cancer regression. Pharmaceut Res 25:2731–2739.

Sudhakar A, Nyberg P, Keshamouni VG, Mannam AP, Li J, Sugimoto H, Cosgrove D, Kalluri R. 2005. Human α1 type IV collagen NC1 domain exhibits distinct antiangiogenic activity mediated by α1β1 integrin. J Clin Invest 115: 2801–2810.

Sudhakar A, Sugimoto H, Yang CQ, Lively J, Zeisberg M, Kalluri R. 2003. Human tumstatin and human endostatin exhibit distinct antiangiogenic activities mediated by αvβ3 and α5β1 integrins. Proc Natl Acad Sci USA 100: 4766–4771.

Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, Forlino A, Cabral WA, Barnes AM, Marini JC, San Antonio JD. 2008. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J Biol Chem 283:21187–21197.

Takahashi T, Yamada O, Soares MJ, Hashizume K. 2008. Bovine prolactin-related protein-I is anchored to the extracellular matrix through interactions with type IV collagen. J Endocrinol 196:225–234.

Tarsio JF, Reger LA, Furcht LT. 1987. Decreased interaction of fibronectin, type-IV collagen, and heparin due to nonenzymatic glycation—implications for diabetes-mellitus. Biochemistry 26:1014–1020.

Tiger CF, Fougerousse F, Grundstrom G, Velling T, Gullberg D. 2001. α1β1 integrin is a receptor for interstitial collagens involved in cell migration and collagen reorganization on mesenchymal nonmuscle cells. Dev Biol 237:116–129.

Timpl R. 1989. Structure and biological activity of basement membrane proteins. Eur J Biochem 180:487–502.

Trus BL, Piez KA. 1980. Compressed microfibril models of the native collagen fibril. Nature 286:300–301.

Tsilibary EC, Koliakos GG, Charonis AS, Vogel AM, Reger LA, Furcht LT. 1988. Heparin type IV collagen interactions: equilibrium binding and inhibition of type IV collagen self-assembly. J Biol Chem 263:19112–19118.

Tsilibary EC, Reger LA, Vogel AM, Koliakos GG, Anderson SS, Charonis AS, Alegre JN, Furcht LT. 1990. Identification of a multifunctional, cell-binding peptide sequence from the α1(NC1) of type-IV collagen. J Cell Biol 111: 1583–1591.

Tulla M, Pentikainen OT, Viitasalo T, Kapyla J, Impola U, Nykvist P, Nissinen L, Johnson MS, Heino J. 2001. Selective binding of collagen subtypes by integrin αI-1, αI-2 and αI-10 domains. J Biol Chem 276:48206–48212.

Vanacore RM, Friedman DB, Ham AJ, Sundaramoorthy M, Hudson BG. 2005. Identification of S-hydroxylysyl-methionine as the covalent cross-link of the noncollagenous (NC1) hexameric α1α1α2 collagen IV network: a role for the post-translational modification of lysine 211 to hydroxylysine 211 in hexamer assembly. J Biol Chem 280:29300–29310.

Vanacore R, Ham AJ, Voehler M, Sanders CR, Conrads TP, Veenstra TD, Sharpless KB, Dawson PE, Hudson BG. 2009. A sulfilimine bond identified in collagen IV. Science 325:1230–1234.

Vandenberg P, Kern A, Ries A, Luckenbilledds L, Mann K, Kuhn K. 1991. Characterization of a type-IV collagen major cell binding-site with affinity to the α1β1 and the α2β1 integrins. J Cell Biol 113:1475–1483.

Vercellotti GM, McCarthy JB, Lindholm P, Peterson PK, Jacob HS, Furcht LT. 1985. Extracellular-matrix proteins (fibronectin, laminin, and type-IV collagen) bind and aggregate bacteria. Am J Pathol 120:13–21.

Vogel W, Brakebusch C, Fassler R, Alves F, Ruggiero F, Pawson T. 2000. Discoidin domain receptor 1 is activated independently of β1 integrin. J Biol Chem 275: 5779–5784.

Wang XM, Harris RE, Bayston LJ, Ashe HL. 2008. Type IV collagens regulate BMP signalling in *Drosophila*. Nature 455:U72–U49.

Westerlund B, Kuusela P, Risteli J, Risteli L, Vartio T, Rauvala H, Virkola R, Korhonen TK. 1989. The O75X adhesin of uropathogenic *Escherichia-coli* is a type-4 collagen-binding protein. Mol Microbiol 3:329–337.

White DJ, Puranen S, Johnson MS, Heino J. 2004. The collagen receptor subfamily of the integrins. Int J Biochem Cell Biol 36:1405–1410.

Wolberg AS, Stafford DW, Erie DA. 1997. Human factor IX binds to specific sites on the collagenous domain of collagen IV. J Biol Chem 272:16717–16720.

Xu Y, Gurusiddappa S, Rich RL, Owens RT, Keene DR, Mayne R, Hook A, Hook M. 2000. Multiple binding sites in collagen type I for the integrins α1β1 and α2β1. J Biol Chem 275:38981–38989.

Yang BG, Tanaka T, Jang MH, Bai Z, Hayasaka H, Miyasaka M. 2007. Binding of lymphoid chemokines to collagen IV that accumulates in the basal lamina of high endothelial venules: its implications in lymphocyte trafficking. J Immunol 179:4376–4382.

Yurchenco PD, Ruben GC. 1987. Basement–membrane structure insitu—evidence for lateral associations in the type-IV collagen network. J Cell Biol 105: 2559–2568.

Zhang WM, Kapyla J, Puranen JS, Knight CG, Tiger CF, Pentikainen OT, Johnson MS, Farndale RW, Heino J, Gullberg D. 2003. α11β1 integrin recognizes the GFOGER sequence in interstitial collagens. J Biol Chem 278:7270–7277.