

Collagens at a glance

Karl E. Kadler, Clair Baldock, Jordi Bella and Raymond P. Boot-Handford

Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK

Author for correspondence (e-mail: karl.kadler@manchester.ac.uk)

Journal of Cell Science 120, 1955-1958
Published by The Company of Biologists 2007
doi:10.1242/jcs.03453

Collagens are a large family of triple helical proteins that are widespread throughout the body and are important for a broad range of functions, including tissue scaffolding, cell adhesion, cell migration, cancer, angiogenesis, tissue morphogenesis and tissue repair. Collagen is best known as the principal tensile element of vertebrate tissues such

as tendon, cartilage, bone and skin, where it occurs in the extracellular matrix as elongated fibrils. Collagen is also well known for its location in basement membranes – for example, in the kidney glomerulus, where it functions in the molecular filtration. However, the identification of transmembrane collagens on the surfaces of a wide variety of cells and collagens that are precursors of bioactive peptides that have paracrine functions has resulted in a revival of interest in collagen. Moreover, new developments in 3D reconstruction electron microscopy have led to new opportunities for studying intracellular trafficking of collagen. Newcomers to the field face the daunting task of sifting through 100,000 research papers that span 40 years. Here, we provide ‘the collagen basics’. Several excellent reviews are cited that are sources of more detailed descriptions and discussions.

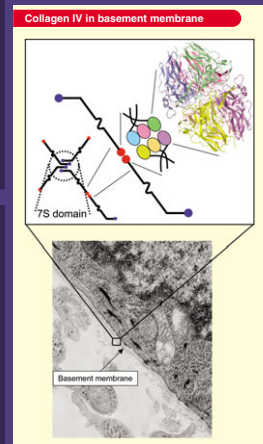
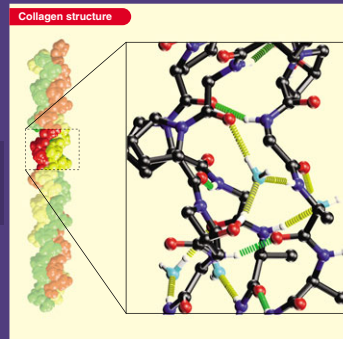
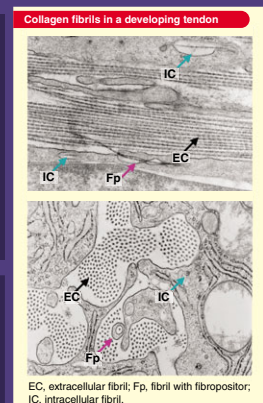
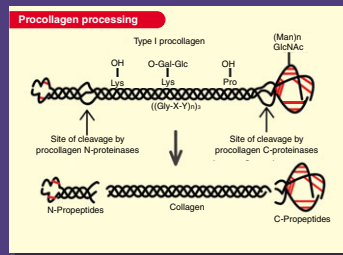
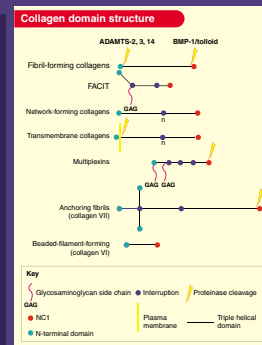
Structure and composition

Collagens contain three polypeptide (α) chains, displaying an extended polyproline-II conformation, a right-handed supercoil and a one-residue stagger between adjacent chains (Brodsky and Persikov, 2005). Each polypeptide chain has a repeating Gly-X-Y triplet in which glycyl residues occupy every third position and the X and Y positions are frequently occupied by proline and 4-hydroxyproline, respectively. The three α chains are held together by interchain hydrogen bonds. Highly ordered hydration networks surround the triple helices. The significance of these interactions to collagen stability remains a matter of debate. Some collagens have interruptions (containing numerous residues) and imperfections (one to three residues) in the triple helix. The conformational changes derived from

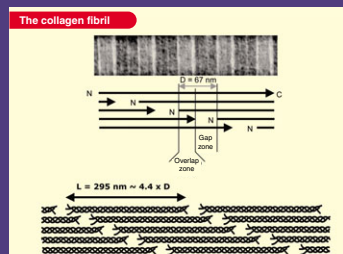
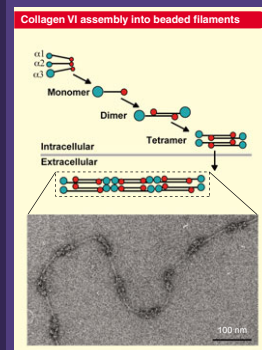
Collagens at a Glance

Karl E. Kadler, Clair Baldock, Jordi Bella and Raymond P. Boot-Handford

Collagen type	Classification	Distribution/tissues
I	Fibril-forming	Non-cartilaginous connective tissues – e.g. tendon, ligament, cornea, bone, mammalian fibrous skin
II	Fibril-forming	Cartilage, vitreous humour and nucleus pulposus
III	Fibril-forming	Co-distributes with collagen I, especially in embryonic skin and hollow organs
IV	Network-forming	Basement membranes
V	Fibril-forming	Co-distributes with collagen I, especially in embryonic tissues and in cornea
VI	Beaded-filament-forming	Widespread, especially muscle
VII	Anchoring fibrils	Dermal-epidermal junction
VIII	Network-forming	Descemet's membrane
IX	FACT	Co-distributes with collagen II, especially in cartilage and vitreous humour
X	Network-forming	Hypertrophic cartilage
XI	Fibril-forming	Co-distributes with collagen II
XII	FACT	Found with collagen I
XIII	Transmembrane	Neuronal junctions, skin
XIV	FACT	Found with collagen I
XV	Endostatin	Located between collagen fibrils that are close to basement membranes; found in the eye, muscle and microvessels; a close structural homologue of collagen XVIII
XVI	FACT	Integrated into collagen fibrils and fibrillar microfibrils
XVII	Transmembrane	Also known as the bellous pempigoid antigen ZBP180; localised to epithelia; an epithelial adhesion molecule; ectodomain cleaved by ADAM proteases
XVIII	Endostatin	Associated with basement membranes; endostatin is proteolytically released from the C-terminus of collagen XVIII; important for retinal vasculogenesis
XIX	FACT	Rare; localised to basement membrane zones; contributes to muscle physiology and differentiation
XX	FACT	Widespread distribution, most prevalent in corneal epithelium
XXI	FACT	Widespread distribution
XXII	FACT	Localised at tissue junctions – e.g. myotendinous junction, cartilage-synovial fluid, hair follicle-dermis
XXIII	Transmembrane	Limited tissue distribution; exists as a transmembrane and shed form
XXIV	Fibril-forming	Shares sequence homology with the fibril-forming collagens; has minor interruptions in the triple helix; selective expression in developing cornea and bone
XXV	Transmembrane	CLAC-P – precursor protein for CLAC (collagenous Alzheimer amyloid plaque component)
XXVI	Beaded-filament-forming	Also known as EMI domain-containing protein 2, protein Ema2, Emilin and multimerin domain-containing protein 2
XXVII	Fibril-forming	Shares sequence homology with the fibril-forming collagens; has minor interruptions in the triple helix; found in embryonic cartilage, developing dermis, cornea, inner limiting membrane of the retina and major arteries of the heart; restricted to cartilage in adults; found in fibrillar-like assemblies
XXVIII	Beaded-filament-forming	A component of the basement membrane around Schwann cells; a von Willebrand factor A domain-containing protein with numerous interruptions in the triple helical domain
Ectodysplasin A	Transmembrane	
Gliomedin	Transmembrane	Myelinating Schwann cells



Journal of Cell Science
jcs.biologists.org



(See poster insert)

some simple imperfections have been visualised in crystal structures of model peptides (Bella et al., 2006). Integrin adhesion sequences [e.g. Arg-Gly-Asp (RGD) and Gly-Phe-O-Gly-Arg, where O is hydroxyproline] occur in the triple helical domain of several collagens and contribute to integrin-ligand-binding specificity (Bella and Humphries, 2005).

Nomenclature and classification

There is no agreed definition for a collagen; there are triple helical proteins that are called collagens and there are proteins that have triple helical domains that are not regarded as collagens. In general, collagens are regarded as triple helical proteins that have functions in tissue assembly or maintenance. Inevitably, the line between 'collagens' and 'collagen-like' proteins is blurred. Vertebrate collagens are given a Roman numeral.

Collagen I is the archetypal collagen in that it is trimeric, it is triple helical, its triple helix has no imperfections, it assembles into fibrils, and it has a predominately structural role in the tissue. However, most collagens differ from collagen I in one or more respects. For example, other collagens can have interruptions in the triple helix and do not necessarily assemble (in their own right) into fibrils. Furthermore, transmembrane collagens have numerous interruptions in the triple helix, do not self-assemble into fibrils, and have roles in cell adhesion and signaling.

At least 28 different collagens occur in vertebrates (numbered I-XXVIII; some with common names), together with a large group of collagen-like proteins (e.g. acetyl cholinesterase, adiponectin, C1q, ficolin, macrophage receptor and surfactant protein) (for a review, see Myllyharju and Kivirikko, 2001). In general, invertebrates have far fewer collagen genes but most have examples of fibrillar and basement membrane collagen (Huxley-Jones et al., 2007). Most collagens have evolved from orthologues present in invertebrates but little is known about the molecular composition of these molecules and no systematic nomenclature exists. Notably, ~200 cuticle-forming collagens are found in *Caenorhabditis elegans* (Page and Winter, 2003).

Collagens can be heterotrimeric – for example, type I collagen, which contains two identical α chains and a third chain that differs, $[\alpha 1(I)]_2\alpha 2(I)$. However, the majority of collagens are homotrimers – for example, collagen II, which contains three identical α chains $[\alpha 1(II)]_3$. Note that the $\alpha 1$ chain of one type of collagen (e.g. collagen I) has a primary structure different from that of the $\alpha 1$ chain of another type of collagen (e.g. collagen II). Collagens have non-triple helical domains at their N- and C-termini. These domains are called 'non collagenous' (NC) domains and are numbered from the C-terminus (NC1, NC2, etc.).

Vertebrate collagens are classified by function and domain homology

Fibril-forming collagens

Fibril-forming collagens occur as 67-nm *D*-periodic fibrils that are the principal source of tensile strength in animal tissues (Kadler et al., 1996). The fibrils are indeterminate in length and range in diameter from 12 nm to >500 nm, depending on the stage of development and tissue. The periodic structure of the fibril is due to regular staggering of triple helical collagen molecules. Mammals have 11 fibrillar collagen genes, which cluster phylogenetically into three distinct subclasses (Huxley-Jones et al., 2007). The Gly-X-Y domain of fibril-forming collagens contains ~1000 residues and is uninterrupted, with the notable exception of collagen XXIV and collagen XXVII.

Fibril-forming collagens are synthesised as procollagens containing N- and C-propeptides at each end of the triple helical domain. Cleavage of the C-propeptides is required for fibrillogenesis. The C-propeptides are cleaved by procollagen C-proteinases, which are identical to the BMP-1/tolloid proteinases (Greenspan, 2005). In the case of collagen V, the pro $\alpha 1(V)$ chain is cleaved by furin to release the C-propeptide. The N-propeptides are cleaved by procollagen N-proteinases, which are identical to the ADAMTS 2, ADAMTS 3 and ADAMTS 14 proteinases (Colige et al., 2005). The pro $\alpha 1(V)$ chain of collagen V is the exception: it is cleaved by BMP-1 (for details, see Greenspan, 2005). Cleavage of the propeptides exposes telopeptide

sequences that are short non-triple helical extensions of the polypeptide chains. The telopeptides contain binding sites for fibrillogenesis (Prockop and Fertala, 1998). The fibrillar collagens are stabilised by non-reducible covalent crosslinks that involve residues in the triple helix and in telopeptides (Eyre et al., 1984). The crosslinks are essential for the normal mechanical properties of collagen-containing tissues.

Fibril-associated collagens with interrupted triple helices (FACITs)

FACITs are relatively short collagens, have interruptions in the triple helical domain and can be found at the surfaces of collagen fibrils. Collagen IX is the archetypal FACIT; it is covalently crosslinked to collagen II (Wu et al., 1992), and is post-translationally modified to carry a glycosaminoglycan side chain.

Network-forming collagens

Collagen IV is the prototypical network-forming collagen. It forms an interlaced network in basement membranes, where it has an important molecular filtration function. The network is generated by head-to-head interactions of two trimeric NC1 domains. The resultant hexamer is stabilised by covalent Met-Lys crosslinks (Than et al., 2002). N-to-N interactions between four collagen IV molecules establish the crosslinked '7S domain', which is an important interaction node in the extended network. Collagen VIII is a major component of Descemet's membrane and vascular subendothelial matrices, where it occurs as polygonal superstructures. The related collagen X occurs in the hypertrophic zone of growth plate cartilage and is thought to form a network similar to that of collagen VIII (Stephan et al., 2004).

Transmembrane collagens

These collagens are type II transmembrane proteins that have a short cytosolic N-terminal domain and long interrupted triple helical extracellular (ecto) domains. They include collagens XIII and XXV, which have cell adhesive properties and occur on numerous cell types, including malignant cells. The ectodomains can be proteolytically shed

by furin-like proprotein convertases. Collagen XVII is cleaved by ADAM family proteinases. A growing number of collagen-like transmembrane proteins that have triple-helical ecto domains are being identified in vertebrates and invertebrates. These have not been assigned to a specific class but have important roles in neural function and neural tube dorsalisation, eye development, modulation of growth factor activity, and have cell adhesive functions. These un-adopted collagens include ectodysplasin, gliomedin and other members of the colmedin sub-family of transmembrane collagens. The ectodomain of gliomedin is shed by BMP-1/tolloid proteinases (Maertens et al., 2007).

Endostatin-producing collagens

Collagen XV is found bridging adjacent collagen fibrils near basement membranes and can form a variety of oligomeric assemblies (Myers et al., 2007). Collagen XVIII is found in some basement membranes. Cleavage of part of the NC1 domains of collagens XV and XVIII releases endostatins, which are inhibitors of endothelial cell migration and angiogenesis, reduce tumour growth in animals, and control neuronal guidance in *C. elegans* (Marnaros and Olsen, 2005).

Anchoring fibrils

Collagen VII is the major component of the anchoring fibrils beneath the lamina densa of epithelia. The NC1 domain of collagen VII is cleaved by BMP-1/tolloid proteinases.

Beaded-filament-forming collagen

Collagen VI is the archetypal beaded-filament-forming collagen. It is found in most tissues where it forms structural links with cells. Collagen VI monomers crosslink into tetramers that assemble into long molecular chains known as microfibrils, which have a beaded repeat of 105 nm.

Collagens and disease

The importance of collagen is exemplified in a wide spectrum of diseases, which are caused by >1000 mutations; see OMIM (Online

Mendelian Inheritance of Man) for a comprehensive listing (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>). These diseases include Alport syndrome (collagen IV), certain arterial aneurysms (collagen III), Bethlem myopathy and Ullrich muscular dystrophy (collagen VI) (Baker et al., 2005), certain chondrodysplasias (collagen IX and XI), some subtypes of the Ehlers-Danlos syndrome (collagen I and V), specific subtypes of epidermolysis bullosa (collagen VII), Kniest dysplasia (collagen II), Knobloch syndrome (collagen XVIII), osteogenesis imperfecta (collagen I) (Marini et al., 2007), some instances of osteoporosis and osteoarthritis, and Stickler syndrome (collagen II, IX and XI) (Van Camp et al., 2006). Mice with genetically engineered collagen mutations have been produced. Collagen accumulates or is ectopically expressed in adhesions, fibrosis, cirrhosis, cardiovascular disease and scars.

Collagen biosynthesis

Collagens undergo extensive post-translational modification in the endoplasmic reticulum prior to triple helix formation. A number of enzymes and molecular chaperones assist in their correct folding and trimerisation. These include several hydroxylases, two collagen glycosyltransferases and peptidyl cis-trans isomerase, in addition to protein disulphide isomerase (PDI) (for a review, see Myllyharju and Kivirikko, 2004). HSP47 is a collagen-specific molecular chaperone that is essential for the normal synthesis of collagen (Nagata, 2003). Fibril-forming collagen molecules fold in a C- to N-terminal direction. The correct trimerisation of the NC1 domains is crucial for collagen assembly and has precluded the use of antigenic tags or fluorescent proteins at the C-terminus of the chains. Folding of the trimeric NC1 domain involves the formation of intra- and inter-chain disulphide bonds (Lees and Bulleid, 1994).

Collagen trafficking

Collagens are relatively large proteins and are not accommodated in conventional small-diameter transport vesicles (Bonfanti et al., 1998; Trucco et al., 2004). Procollagen I can be cleaved

to collagen inside the cell and intracellular collagen fibrils can occur in plasma membrane protrusions called fibripositors (see review by Canty and Kadler, 2005) (see also Canty et al., 2004). In embryonic stages, collagen fibrils are closely associated with the plasma membrane of tendon and corneal fibroblasts (Birk and Trelstad, 1986). Collagen V is needed for the nucleation of collagen-I-containing fibrils in vivo (Wenstrup et al., 2004).

Degradation

The triple helix is resistant to proteolytic cleavage by pepsin, trypsin and papain. *Clostridium histolyticum* produces collagenases that cleave triple helices at numerous sites. The ability of collagens to resist cleavage by pepsin and trypsin, and their sensitivity to cleavage by bacterial collagenase, are used as research tools to identify and characterise collagens. Degradation of collagen and gelatin (unfolded or denatured collagen) in vivo can be mediated by MMPs, cysteine proteinases (e.g. cathepsins B, K and L), and serine proteinases (e.g. plasmin and plasminogen activator) (for reviews, see Everts et al., 1996; Sabeh et al., 2004).

References

- Baker, N. L., Morgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F. and Lemande, S. R. (2005). Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum. Mol. Genet.* **14**, 279-293.
- Bella, J. and Humphries, M. J. (2005). Alpha-H...O=C hydrogen bonds contribute to the specificity of RGD cell-adhesion interactions. *BMC Struct. Biol.* **5**, 4.
- Bella, J., Liu, J., Kramer, R., Brodsky, B. and Berman, H. M. (2006). Conformational effects of Gly-X-Gly interruptions in the collagen triple helix. *J. Mol. Biol.* **362**, 298-311.
- Birk, D. E. and Trelstad, R. L. (1986). Extracellular compartments in tendon morphogenesis: collagen fibril, bundle, and macroaggregate formation. *J. Cell Biol.* **103**, 231-240.
- Bonfanti, L., Mironov, A. A., Martinez-Menarguez, J. A., Martella, O., Fusella, A., Baldassarre, M., Buccione, R., Geuze, H. J. and Luini, A. (1998). Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. *Cell* **95**, 993-1003.
- Brodsky, B. and Persikov, A. V. (2005). Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **70**, 301-339.
- Canty, E. G. and Kadler, K. E. (2005). Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* **118**, 1341-1353.
- Canty, E. G., Lu, Y., Meadows, R. S., Shaw, M. K., Holmes, D. F. and Kadler, K. E. (2004). Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon. *J. Cell Biol.* **165**, 553-563.
- Colige, A., Ruggiero, F., Vandenberghe, I., Dubail, J., Kesteloot, F., Van Beeumen, J., Beschin, A., Brys, L., Lapiere, C. M. and Nusgens, B. (2005). Domains and

- maturation processes that regulate the activity of ADAMTS-2, a metalloproteinase cleaving the aminopropeptide of fibrillar procollagens types I-III and V. *J. Biol. Chem.* **280**, 34397-34408.
- Everts, V., van der Zee, E., Creemers, L. and Beertsen, W.** (1996). Phagocytosis and intracellular digestion of collagen, its role in turnover and remodelling. *Histochem. J.* **28**, 229-245.
- Eyre, D. R., Paz, M. A. and Gallop, P. M.** (1984). Cross-linking in collagen and elastin. *Annu. Rev. Biochem.* **53**, 717-748.
- Greenspan, D. S.** (2005). Biosynthetic processing of collagen molecules. *Top. Curr. Chem.* **247**, 149-183.
- Huxley-Jones, J., Robertson, D. L. and Boot-Handford, R. P.** (2007). On the origins of the extracellular matrix in vertebrates. *Matrix Biol.* **26**, 2-11.
- Kadler, K. E., Holmes, D. F., Trotter, J. A. and Chapman, J. A.** (1996). Collagen fibril formation. *Biochem. J.* **316** (Pt 1), 1-11.
- Lees, J. F. and Bulleid, N. J.** (1994). The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J. Biol. Chem.* **269**, 24354-24360.
- Maertens, B., Hopkins, D., Franzke, C. W., Keene, D. R., Bruckner-Tuderman, L., Greenspan, D. S. and Koch, M.** (2007). Cleavage and oligomerization of gliomedin, a transmembrane collagen required for node of Ranvier formation. *J. Biol. Chem.* **282**, 10647-10659.
- Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., Hyland, J. C., Korkko, J., Prockop, D. J., De Paep, A. et al.** (2007). Consortium for osteogenesis 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.
- Marners, A. G. and Olsen, B. R.** (2005). Physiological role of collagen XVIII and endostatin. *FASEB J.* **19**, 716-728.
- Myers, J. C., Amenta, P. S., Dion, A. S., Sciancalepore, J. P., Nagaswami, C., Weisel, J. W. and Yurchenco, P. D.** (2007). The molecular structure of human tissue type XV presents a unique conformation among the collagens. *Biochem. J.* PMID: 17355226 E-press ahead of print
- Myllyharju, J. and Kivirikko, K. I.** (2001). Collagens and collagen-related diseases. *Ann. Med.* **33**, 7-21.
- Myllyharju, J. and Kivirikko, K. I.** (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **20**, 33-43.
- Nagata, K.** (2003). HSP47 as a collagen-specific molecular chaperone: function and expression in normal mouse development. *Semin. Cell Dev. Biol.* **14**, 275-282.
- Page, A. P. and Winter, A. D.** (2003). Enzymes involved in the biogenesis of the nematode cuticle. *Adv. Parasitol.* **53**, 85-148.
- Prockop, D. J. and Fertala, A.** (1998). Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides. Demonstration that assembly is driven by specific binding sites on the monomers. *J. Biol. Chem.* **273**, 15598-15604.
- Sabeh, F., Ota, I., Holmbeck, K., Birkedal-Hansen, H., Soloway, P., Balbin, M., Lopez-Otin, C., Shapiro, S., Inada, M., Krane, S. et al.** (2004). Tumor cell traffic through the extracellular matrix is controlled by the membrane-anchored collagenase MT1-MMP. *J. Cell Biol.* **167**, 769-781.
- Stephan, S., Sherratt, M. J., Hodson, N., Shuttleworth, C. A. and Kielty, C. M.** (2004). Expression and supramolecular assembly of recombinant alpha1(viii) and alpha2(viii) collagen homotrimers. *J. Biol. Chem.* **279**, 21469-21477.
- Than, M. E., Henrich, S., Huber, R., Ries, A., Mann, K., Kuhn, K., Timpl, R., Bourenkov, G. P., Bartunik, H. D. and Bode, W.** (2002). The 1.9-A crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. *Proc. Natl. Acad. Sci. USA* **99**, 6607-6612.
- Trucco, A., Polishchuk, R. S., Martella, O., Di Pentima, A., Fusella, A., Di Giandomenico, D., San Pietro, E., Beznoussenko, G. V., Polishchuk, E. V., Baldassarre, M. et al.** (2004). Secretory traffic triggers the formation of tubular continuities across Golgi sub-compartments. *Nat. Cell Biol.* **6**, 1071-1081.
- Van Camp, G., Snoeckx, R. L., Hilgert, N., van den Ende, J., Fukuoka, H., Wagatsuma, M., Suzuki, H., Smets, R. M., Vanhoenacker, F., Declau, F. et al.** (2006). A new autosomal recessive form of Stickler syndrome is caused by a mutation in the COL9A1 gene. *Am. J. Hum. Genet.* **79**, 449-457.
- Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I. and Birk, D. E.** (2004). Type V collagen controls the initiation of collagen fibril assembly. *J. Biol. Chem.* **279**, 53331-53337.
- Wu, J. J., Woods, P. E. and Eyre, D. R.** (1992). Identification of cross-linking sites in bovine cartilage type IX collagen reveals an antiparallel type II-type IX molecular relationship and type IX to type IX bonding. *J. Biol. Chem.* **267**, 23007-23014.

Cell Science at a Glance on the Web
Electronic copies of the poster insert are available in the online version of this article at jcs.biologists.org. The JPEG images can be downloaded for printing or used as slides.