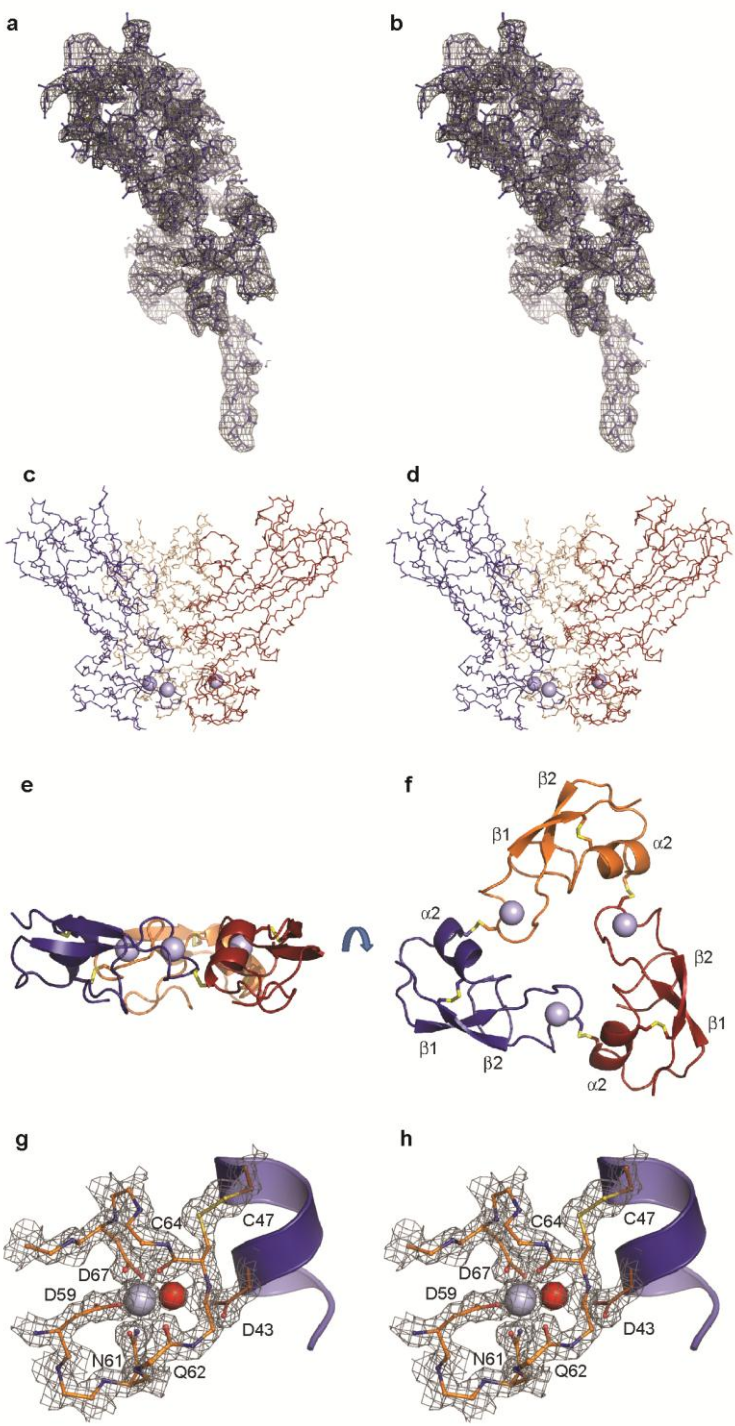
Supplementary Information

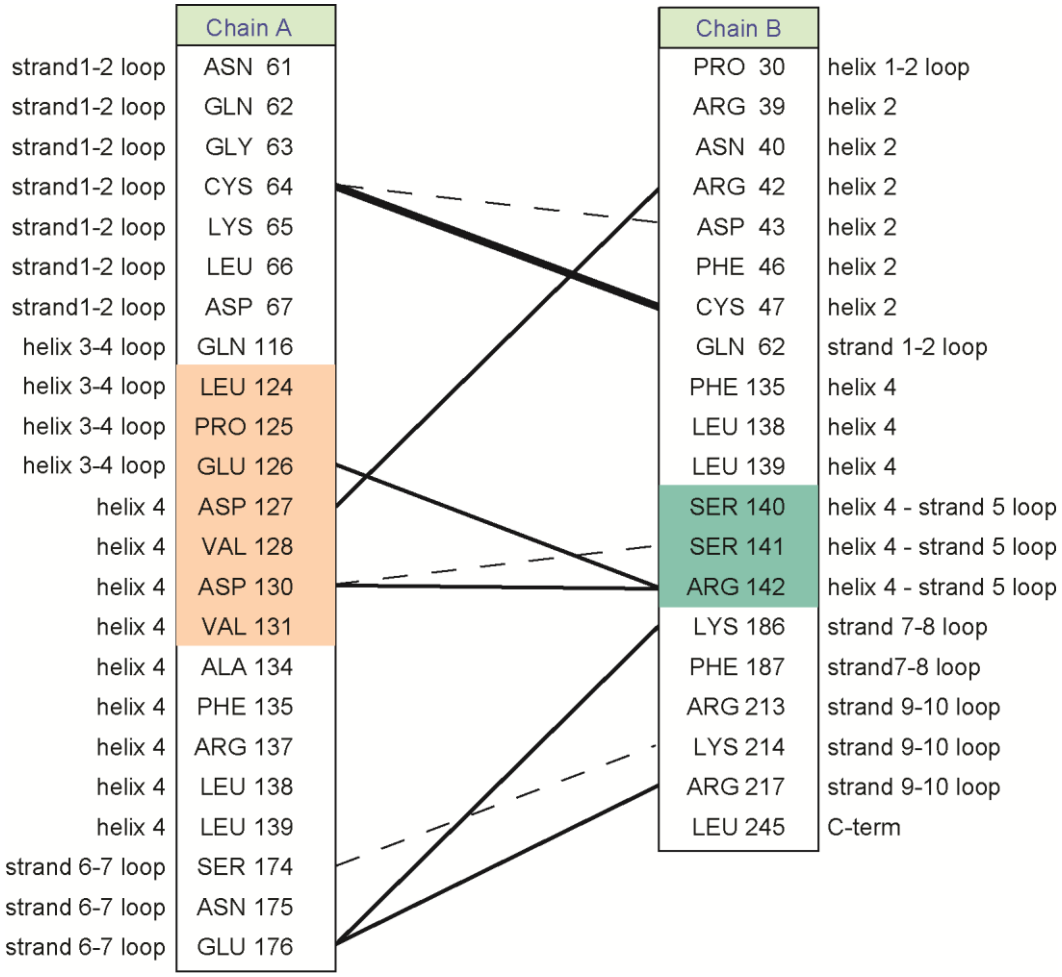
Structural Basis of Fibrillar Collagen Trimerization and Related Genetic Disorders

Jean-Marie Bourhis[1,2], Natacha Mariano[1], Yuguang Zhao[3], Karl Harlos[3], Jean-Yves Exposito[1], E.Yvonne Jones[3], Catherine Moali[1], Nushin Aghajari[4], David J.S. Hulmes[1]

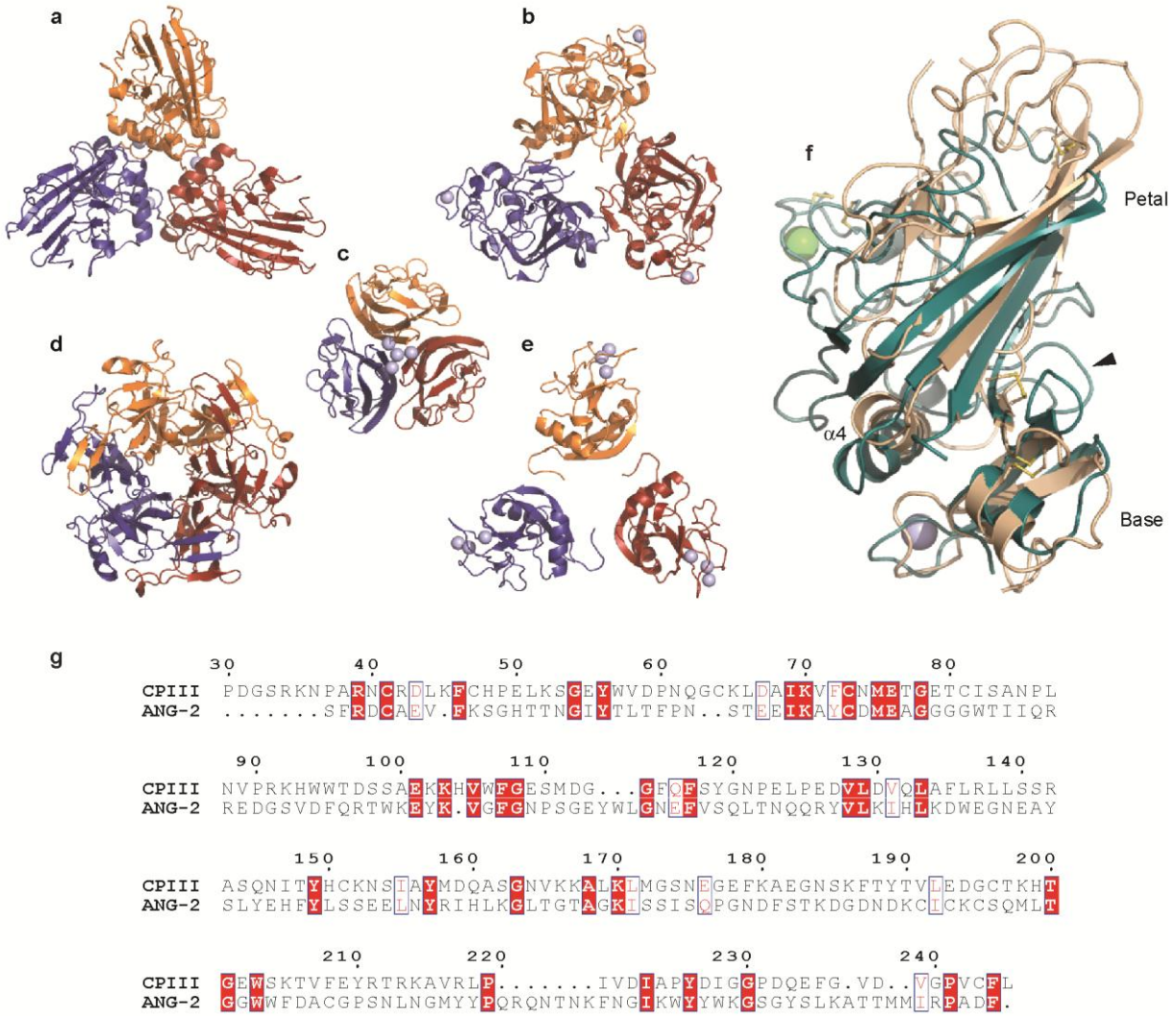
[1]Formation de Recherche en Evolution 3310, Institut de Biologie et Chimie des Protéines, Centre National de la Recherche Scientifique, Université Lyon 1, 69367 Lyon cedex 7, France; [2]Unit for Virus Host Cell Interactions, Unité Mixte Internationale 3265, Centre National de la Recherche Scientifique, Université Joseph Fourier, European Molecular Biology Laboratory, 38042 Grenoble cedex 9, France; [3]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford 0X3 7BN, UK; [4]Unité Mixte de Recherche 5086, Institut de Biologie et Chimie des Protéines, Centre National de la Recherche Scientifique, Université Lyon 1, 69367 Lyon cedex 7, France.



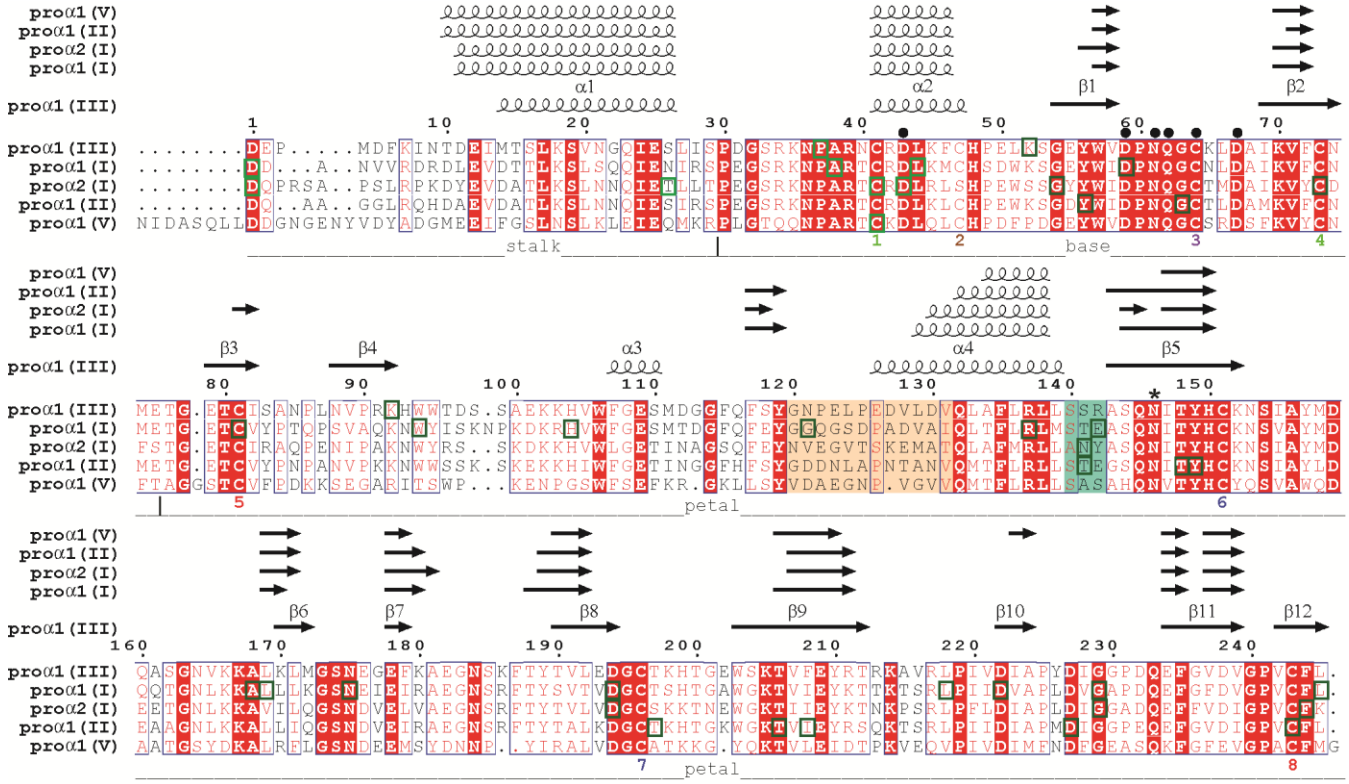
Supplementary Figure 1. Structural features of the human procollagen III C-propeptide trimer. (**a,b**) Stereo view of chain A from the 3.5 Å structure (form III) in the corresponding 2Fo-Fc electron density map contoured at 1.5σ. (**c,d**) Stereo view of the 1.7Å structure (form II; backbone trace) showing the three polypeptide chains as well as bound Ca^{2+} ions (light blue). Side (**e**) and top (**f**) views of the base region (form II; residues 30-76; Fig. 1b) showing bound Ca^{2+} ions, intra- and inter-chain disulfide bonds (yellow) and secondary structural elements. (**g,h**) Stereo view of the Ca^{2+} binding site (form II) in the 2Fo-Fc electron density map contoured at 1.5σ. Octahedral coordination involves Asp59, Asn61, Gln62, Cys64 and Asp67 in one chain as well as a water molecule (red) hydrogen bonded to Asp43 in a neighboring chain. Also shown is the inter-chain disulfide bond involving Cys47 and Cys64. All structures drawn using PyMOL, Version 1.4.1, Schrödinger, LLC.



Supplementary Figure 2. Residues involved in stabilizing the interface between adjacent chains in the procollagen III C-propeptide trimer (excluding the stalk region). The inter-chain disulfide bond is indicated by the thick line, salt bridges by thin lines and hydrogen bonds by dotted lines. The remainder are interfacing residues. Corresponding secondary structure elements are also indicated. Residues in the chain recognition sequence are highlighted in wheat color (long stretch) and deep teal (short stretch). Interactions calculated using the PDBePISA[44] server at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html).



Supplementary Figure 3. Globular domains in fibrous protein trimerization and structural similarity with angiopoietin-2. (**a-e**) Different types of globular domain including (**a**) fibrillar procollagen C-propeptides (procollagen III; buried surface area ~ 3800 Å^2), (**b**) fibrinogen domains (L-ficolin[45]; PDB code 2J3G; buried surface area ~ 2900 Å^2) (**c**) C1q family[46] (collagen X NC1 trimer[47]; PDB code 1GR3; buried surface area ~ 5600 Å^2) (**d**) collagen IV NC1 region[48,49] (PDB code 1LI1; buried surface area ~ 13400 Å^2) and (**e**) collectins[50,51] (mannan binding lectin; PDB code P3OB; buried surface area ~ 900 Å^2). In all cases, views are from the C-terminal end of the molecule looking down the principal axis, with only the globular regions shown (i.e. omitting underlying collagen-like or α-helical coiled-coil regions). Bound Ca^{2+} ions are shown in light blue. Among these structures, only the procollagen C-propeptide trimer is stabilized by inter-chain disulfide bonds. Buried surface areas calculated using the PDBePISA[44] server at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html). (**f**) Structural alignment of chain A of the proα1(III) C-propeptide trimer (1.7Å structure; wheat color; Ca^{2+} ion in light blue) with the receptor binding domain of angiopoietin-2[52] (PDB code 1Z3S; deep teal color; Ca^{2+} ion in green), obtained using DALI[23]. Note that the equivalent of helix 4 in the C-propeptide structure is interrupted by a large loop (arrowhead) in angiopoietin-2. Structural similarity in the base region includes a conserved intra-chain disulfide bond. (**g**) Sequence alignment of the C-propeptide of human procollagen III (CPIII) with the receptor binding region of angiopoietin-2 (ANG-2). Sequence identity is less than 15 %. Sequence alignment and rendering done using CLUSTALW[34] and ESPript[35], respectively. All structures drawn using PyMOL, Version 1.4.1, Schrödinger, LLC.



Supplementary Figure 4. Alignment of the proα1(I), proα2(I), proα1(II), proα1(III) and proα1(V) C-propeptides showing the locations of all known naturally occurring missense mutations (green boxes; see Supplementary Table 1). Different regions and secondary structure elements found in the procollagen III C-propeptide are also indicated, as are predicted secondary structures (obtained using PsiPred[53]) for the other C-propeptides. Also shown are the positions of Cys residues (numbered according to the sequence and also as Cys 1 to 8) with intra-chain disulfide bonds identified as color-matched pairs. Residues involved in Ca^{2+} coordination are indicated by ● and the single N-linked glycosylation site by * (note Asn146 was mutated to Gln in the structure presented here). The long (12 residue) and short (3 residue) stretches of the discontinuous 15 residue chain recognition sequence are highlighted in wheat and deep teal color, respectively. Numbering refers to the C-propeptides of the proα1(III) chain. Sequence alignments and rendering were done using CLUSTALW[34] and ESPript[35], respectively.

SupplementaryTable 1. Positions and consequences of known missense mutations in the C-propeptides of the proα1(I), proα2(I), proα1(II), proα1(III) and proα1(V) chains. Numbering starts from the methionine at the translation initiation site, with corresponding positions in the procollagen III C-propeptide (CPIII) numbered from the BMP-1 cleavage site. Note that the disorders caused by the PLSD-T/SPD mutations in the proα1(II) chain are specific to the C-propeptide region[28].

Gene	Mutation and position	CPIII equivalent	Phenotype	Ref	Region	Probable consequence
COL1A1	Asp1219Glu	1 (Asp)	OI type I (mild)	[54]	stalk	prevents BMP-1 cleavage
COL1A1	Asp1219Asn	1 (Asp)	OI type I (mild)	[16]	stalk	prevents BMP-1 cleavage
COL1A1	Ala1256Thr	38 (Ala)	OI type I (mild)	[55]	base	destabilizes base
COL1A1	Leu1262Arg	44 (Leu)	OI type II (lethal)	a	base	disrupts S-S bond
COL1A1	Asp1277His	59 (Asp)	OI type II (lethal)	[31]	base	disrupts Ca^{2+} binding
COL1A1	Asp1277Glu	59 (Asp)	OI type II (lethal)	[56]	base	disrupts Ca^{2+} binding
COL1A1	Cys1299Trp	81 (Cys)	OI type I (mild)	[25]	petal	disrupts S-S bond
COL1A1	Trp1312Cys	94 (Trp)	OI type II (lethal)	[29]	petal	disrupts S-S bond
COL1A1	His1323Tyr	104 (His)	OI type I (mild)	[57]	petal	disrupts surface interactions
COL1A1	Gly1340Ser	121 (Asn)	OI type I (mild)	a	petal	disrupts CRS
COL1A1	Arg1356His	137 (Arg)	OI type II/III (lethal)	a	petal	disrupts inter-chain interactions
COL1A1	Glu1361Lys	142 (Arg)	OI type IV (moderate)	a	petal	disrupts CRS
COL1A1	Ala1387Val	168 (Ala)	OI type IIC (lethal)	[58]	petal	disrupts packing
COL1A1	Leu1388Arg	169 (Leu)	OI type II (lethal)	[31]	petal	disrupts hydrophobic core
COL1A1	Asn1394Ser	175 (Asn)	OI type I (mild)	a	petal	disrupts inter-chain interactions
COL1A1	Asp1413Asn	194 (Asp)	OI type II (lethal)	[54,56,59]	petal	disrupts S-S bond
COL1A1	Leu1437Gln	218 (Leu)	OI type II (lethal)	[60]	petal	disrupts hydrophobic core
COL1A1	Asp1441His	222 (Asp)	OI type I (mild)	[61]	petal	disrupts surface interactions
COL1A1	Asp1441Tyr	222 (Asp)	OI type II (lethal)	[62]	petal	disrupts surface interactions
COL1A1	Gly1448Asp	229 (Gly)	OI type I/IV (moderate)	a	petal	disrupts surface interactions
COL1A1	Leu1464Pro	245 (Leu)	OI type III (severe)	[27]	petal	disrupts S-S bond
COL1A2	Asp1120Ala	1 (Asp)	OI type I (mild)	a	stalk	prevents BMP-1 cleavage
COL1A2	Thr1148Pro	26 (Ser)	OI type III (severe)	[27]	stalk	disrupts coiled coil
COL1A2	Cys1163Arg	41 (Cys)	OI type IV (moderate)	[63]	base	disrupts S-S bond
COL1A2	Asp1165Glu	43 (Asp)	OI type I (mild)	a	base	disrupts Ca^{2+} binding
COL1A2	Gly1176Val	54 (Gly)	OI type IV (moderate)	[63]	base	disrupts surface interactions
COL1A2	Cys1195Tyr	73 (Cys)	OI type I (mild)	[54]	base	disrupts S-S bond
COL1A2	Asn1262Ser	141 (Ser)	OI type I (mild)	a	petal	disrupts CRS
COL1A2	Asp1315Val	194 (Asp)	OI type IV (moderate)	[63]	petal	disrupts surface interactions
COL1A2	Gly1350Ser	229 (Gly)	OI type I (mild)	a	petal	disrupts surface interactions
COL1A2	Phe1365Leu	244 (Phe)	OI type I (mild)	[55]	petal	disrupts S-S bond
COL2A1	Tyr1298Asn	56 (Tyr)	SEDT (moderate)	[64]	base	destabilizes base

COL2A1	Gly1305Ala	63 (Gly)	VEPD (moderate)	[65]	base	disrupts Ca^{2+} binding loop
COL2A1	Thr1383Met	141 (Ser)	ANFH (mild)	[66]	petal	disrupts CRS
COL2A1	Thr1390Asn	148 (Thr)	ACGII-HCG (PLSD-T) (lethal)	[28,67]	petal	disrupts S-S bond
COL2A1	Tyr1391Cys	149 (Tyr)	PLSD-T/SPD (lethal/severe)	[30,68]	petal	disrupts S-S bond
COL2A1	Thr1439Met	197 (Thr)	SEDC/PSACH (severe ; double mutant COL2A1/COMP)	[69]	petal	disrupts S-S bond
COL2A1	Thr1448Pro	206 (Thr)	PLSD-T (lethal)	[28]	petal	disrupts secondary structure
COL2A1	Ile1450Thr	208 (Phe)	SEDC-M-like (mild)	[64]	petal	disrupts hydrophobic core
COL2A1	Ile1450Asn	208 (Phe)	SEDC-M-like (mild)	[64]	petal	disrupts hydrophobic core
COL2A1	Asp1469His	227 (Asp)	PLSD-T (moderate)	[28]	petal	disrupts surface interactions
COL2A1	Cys1485Gly	243 (Cys)	PLSD-T (lethal/severe; mother and child)	[28]	petal	disrupts S-S bond
COL3A1	Pro1258Ser	37 (Pro)	EDS type IV (vascular form, mild)	b	base	disrupts secondary structure
COL3A1	Lys1273Arg	52 (Lys)	EDS type IV (vascular form, mild)	b	base	disrupts surface interactions
COL3A1	Lys1313Arg	92 (Lys)	EDS type IV/TSAH (vascular form, lethal following traumatic injury)	[70]	petal	disrupts surface interactions
COL5A1	Cys1639Ser	41 (Cys)	EDS type I (hyperextensible skin, joint laxity)	[32]	base	disrupts S-S bond

[a] Unpublished, see https://oi.gene.le.ac.uk

[b] Personal communication (X. Jeunemaitre, Paris Cardiovascular Research Centre, INSERM U970, Paris, France).

Other sources : https://eds.gene.le.ac.uk, http://bioinf.umbc.edu/dmdm

Abbreviations: ACGII-HCG = achondrogenesis type II – hypochondrogenesis; ANFH = avascular necrosis of the femoral head; COMP = cartilage oligomeric matrix protein; CPIII = C-propeptide of procollagen III; EDS = Ehlers-Danlos syndrome; CRS = chain recognition sequence; OI = osteogenesis imperfecta; PSACH = pseudoachondroplasia; PLSD-T = platyspondylic lethal skeletal dysplasia Torrance type (N.B. despite its name, this syndrome is not always lethal); PSACH = pseudoachondrodysplasia; SEDC = spondyloepiphyseal dysplasia congenita; SEDC-M = spondyloepiphyseal dysplasia congenita mild; SEDT = spondyloepiphyseal dysplasia tarda; SPD = spondyloperipheral dysplasia; TSAH = traumatic sub-arachnoid haemorrhage; VPED = vitreoretinopathy and phalangeal epiphyseal dysplasia.

44. Krissinel,E. & Henrick,K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol* **372**, 774-797 (2007).
45. Garlatti,V. *et al.* Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J.* **26**, 623-633 (2007).
46. Gaboriaud,C. *et al.* The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J. Biol. Chem.* **278**, 46974-46982 (2003).
47. Bogin,O. *et al.* Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Struct. Fold. Des.* **10**, 165-173 (2002).
48. Than,M.E. *et al.* The 1.9-Å 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. Nat. Acad. Sci. USA* **99**, 6607-6612 (2002).
49. Sundaramoorthy,M., Meiyappan,M., Todd,P., & Hudson,B.G. Crystal structure of NC1 domains: Structural basis for type IV collagen assembly in basement membranes. *J. Biol. Chem.* **277**, 31142-31153 (2002).
50. Sheriff,S., Chang,C.Y., & Ezekowitz,R.A. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat. Struct. Biol.* **1**, 789-794 (1994).
51. Weis,W.I. & Drickamer,K. Trimeric structure of a C-type mannose-binding protein. *Structure.* **2**, 1227-1240 (1994).
52. Barton,W.A., Tzvetkova,D., & Nikolov,D.B. Structure of the angiopoietin-2 receptor binding domain and identification of surfaces involved in Tie2 recognition. *Structure.* **13**, 825-832 (2005).
53. Buchan,D.W. *et al.* Protein annotation and modelling servers at University College London. *Nucleic Acids Res.* **38**, W563-W568 (2010).
54. Pollitt,R. *et al.* Mutation analysis of COL1A1 and COL1A2 in patients diagnosed with osteogenesis imperfecta type I-IV. *Hum. Mutat.* **27**, 716 (2006).
55. Zhang,Z.L. *et al.* The identification of novel mutations in COL1A1, COL1A2 and LEPRE1 genes in Chinese patients with osteogenesis imperfecta. *J. Bone Miner. Metab.* **30**, 69-77 (2012).
56. Pyott,S.M. *et al.* Recurrence of perinatal lethal osteogenesis imperfecta in sibships: parsing the risk between parental mosaicism for dominant mutations and autosomal recessive inheritance. *Genet. Med.* **13**, 125-130 (2011).
57. Ries-Levavi,L. *et al.* Genetic and biochemical analyses of Israeli osteogenesis imperfecta patients. *Hum. Mutat.* **23**, 399-400 (2004).
58. Takagi,M. *et al.* Heterozygous C-propeptide mutations in COL1A1: osteogenesis imperfecta type IIC and dense bone variant. *Am. J. Med. Genet. A* **155A**, 2269-2273 (2011).
59. Bodian,D.L. *et al.* Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype-phenotype relationships. *Hum. Mol. Genet.* **18**, 463-471 (2009).
60. Rugolotto,S. *et al.* Pulmonary function tests in an infant with osteogenesis imperfecta and early biphosphonate treatment. *Acta Paediatrica* **96**, 1855-1862 (2007).
61. Ke,L.F. *et al.* Molecular diagnosis of a Chineses pedigree with osteogenesis imperfecta type I. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **26**, 50-53 (2009).
62. Pace,J.M. *et al.* A single amino acid substitution (D1441Y) in the carboxyl-terminal propeptide of the proalpha1(I) chain of type I collagen results in a lethal variant of osteogenesis imperfecta with features of dense bone diseases. *J. Med. Genet.* **39**, 23-29 (2002).
63. Pace,J.M. *et al.* Defective C-propeptides of the proalpha2(I) chain of type I procollagen impede molecular assembly and result in osteogenesis imperfecta. *J. Biol. Chem.* **283**, 16061-16067 (2008).
64. Nishimura,G. *et al.* The phenotypic spectrum of COL2A1 mutations. *Hum. Mutat.* **26**, 36-43 (2005).
65. Richards,A.J. *et al.* Vitreoretinopathy with phalangeal epiphyseal dysplasia, a type II collagenopathy resulting from a novel mutation in the C-propeptide region of the molecule. *J. Med. Genet.* **39**, 661-665 (2002).
66. Kannu,P., O'Rielly,D.D., Hyland,J.C., & Kokko,L.A. Avascular necrosis of the femoral head due to a novel C propeptide mutation in COL2A1. *Am. J. Med. Genet. A* **155A**, 1759-1762 (2011).
67. Mortier,G.R. *et al.* Report of five novel and one recurrent COL2A1 mutations with analysis of genotype-phenotype correlation in patients with a lethal type II collagen disorder. *J. Med. Genet.* **37**, 263-271 (2000).
68. Hoornaert,K.P. *et al.* Czech dysplasia metatarsal type: another type II collagen disorder. *Eur. J. Hum. Genet.* **15**, 1269-1275 (2007).
69. Unger,S. *et al.* Double heterozygosity for pseudoachondroplasia and spondyloepiphyseal dysplasia congenita. *Am. J. Med. Genet.* **104**, 140-146 (2001).
70. Pickup,M.J. & Pollanen,M.S. Traumatic subarachnoid hemorrhage and the COL3A1 gene: emergence of a potential causal link. *Forensic Sci. Med. Pathol.* **7**, 192-197 (2011).