INTRINSIC DISORDER OF THE EXTRACELLULAR MATRIX

Franck Peysselon[1], Bin Xue[2], Vladimir N. Uversky[2,3*], Sylvie Ricard-Blum[1*]

SUPPLEMENTARY TABLES

Supplementary Table 1S: Intrinsic disorder predicted in the protein domains found in the collagens. The signal peptide has been removed from the amino acid sequences before prediction. (C1q: C1q domain, C-propeptide: domain released during the maturation of procollagen, FNIII: fibronectin type III domain, Kunitz/BPTI: Kunitz/Bovine pancreatic trypsin inhibitor domain, N-propeptide: domain released during the maturation of procollagen, TSPN Thrombospondin N-terminal-like domains, VWFA: von Willebrand factor, type A, VWC: von Willebrand factor, type C). The analysis is based on UniProtKB sequence annotations. Intrinsic disorder has been predicted from protein sequences with IUPred1[6,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Domain name	Amount of predicted disorder (mean, %)	Range of disorder (%)	Number of domains in collagens
C1q	1.0	0.0 - 3.0	3
C-propeptide	21.0	10.5 - 37.1	9
Fibronectin III	31.0	1.1 - 60.9	42
Kunitz/BPTI	9.8	0.0 - 27.4	3
Matricryptins	7.5	0.0 - 22.0	7
N-propeptide	67.6	56.1 - 100.0	6
VWFA	2.5	0.0 - 21.6	50
VWFC	13.8	6.78 - 23.8	4
TSPN	4.6	0.0 - 16.1	16

Supplementary Table 2S: Intrinsic disorder predicted in collagen-like proteins containing a triple-helical domain but not referred to as collagens. The signal peptide has been removed from the sequence before prediction. (WD repeat-containing protein 33, Q9C0J8, a nuclear protein containing a triple-helical domain and neurogranin, Q92686, involved in signaling, were not included into the table because they are not secreted). Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Protein name	UniProtKB AC	Number of disordered amino acids/Number of all amino acids	Amount of predicted disorder (%)	Domain Collagen-like (%)
Acetylcholinesterase collagenic tail peptide	Q9Y215	248/433	57.3	43.7
Adiponectin	Q15848	79/226	35.0	29.2
Complement C1q subcomponent subunit A	P02745	105/223	47.1	35.4
Complement C1q subcomponent subunit B	P02746	107/226	47.4	34.5
Complement C1q subcomponent subunit C	P02747	114/217	52.5	37.8
Complement C1q-like protein 2	Q7Z5L3	121/266	45.5	16.2
Complement C1q-like protein 3	Q5VWW1	81/235	34.5	21.7
Complement C1q-like protein 4	Q86Z23	84/223	37.7	19.7
C1q-related factor	O75973	106/242	43.8	20.3
Complement C1q tumor necrosis factor-related protein 1	Q9BXJ1	91/256	35.6	16.4
Complement C1q tumor necrosis factor-related protein 2	Q9BXJ5	135/270	50.0	37.8
Complement C1q tumor necrosis factor-related protein 3	Q9BXJ4	84/224	37.5	28.1
Complement C1q tumor necrosis factor-related protein 5	Q9BXJ0	113/228	49.6	29.0
Complement C1q tumor necrosis factor-related protein 6	Q9BXI9	53/232	22.8	18.1
Complement C1q tumor necrosis factor-related protein 7	Q9BXJ2	118/273	43.2	37.4

Complement C1q tumor necrosis factor-related protein 8	P60827	33/235	14.0	18.3
Complement C1q tumor necrosis factor-related protein 9A	P0C862	173/314	55.1	49.7
Complement C1q tumor necrosis factor-related protein 9B	B2RNN3	171/314	54.5	49.7
Collagen and calcium-binding EGF domain-containing protein 1	Q6UXH8	184/372	49.5	21.5
Otolin-1	A6NHN0	304/454	67.0	39.7
Mannose-binding protein C (collectin-1)	P11226	95/228	41.7	25.4
Collectin-10	Q9Y6Z7	83/250	33.2	24.0
Collectin-11	Q9BWP8	102/246	41.5	19.5
Collectin-12	Q5KU26	283/742	38.1	19.8
Ficolin-1	O00602	65/297	21.9	13.1
Ficolin-2	Q15485	64/288	22.2	14.6
Ficolin-3	O75636	57/276	20.7	12.0
Ectodysplasin A	Q92838	212/391	54.2	12.8
Macrophage receptor with collagenous structure (MARCO)	Q9UEW3	301/520	57.9	52.5
Macrophage scavenger receptor types I and II	P21757	167/451	37.0	15.3
Gliomedin	Q6ZMI3	246/551	44.7	15.6
Scavenger receptor class A member 3	Q6AZY7	176/606	29.0	19.6
Scavenger receptor class A member 5	Q6ZMJ2	115/495	23.2	12.1
Pulmonary surfactant A2	Q8IWL1	135/228	59.2	32.0
Pulmonary surfactant D	P35247	230/355	64.8	49.9
Pulmonary surfactant A1	Q8IWL2	131/228	57.5	32.0
EMI domain Containing protein 1	Q96A84	282/419	67.3	45.4
Emilin-1	Q9Y6C2	493/995	49.6	5.1
Emilin-2	Q9BXX0	458/1023	44.8	5.2
Collagen triple helix repeat-containing protein 1	Q96CG8	31/213	14.6	16.0

Supplementary Table 3S: Intrinsic disorder of elastin and microfibrillar-associated proteins. The signal peptide has been removed from the sequence before prediction. Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Protein name	UniProtKB AC	Number of disordered amino acids/Number of all amino acids	Amount of predicted disorder (%)
Elastin	P15502	39/760	5.1
Fibrillin-1	P35555	76/2844	2.7
Fibrillin-2	P35556	93/2884	3.2
Fibrillin-3	Q75N90	51/2778	1.8
EMILIN-1	Q9Y6C2	493/995	49.6
EMILIN-2	Q9BXX0	458/1023	44.8
EMILIN-3	Q9NT22	207/744	27.8
Multimerin-1	Q13201	226/1209	18.7
Multimerin2	Q9H8L6	230/927	24.8
EMI domain Containing protein 1 (Emu-1)	Q96A84	282/419	67.3
Microfibrillar-associated protein 1	P55081	305/439	69.5
Microfibrillar-associated protein 2 (Microfibril-associated glycoprotein 1)	p55001	71/166	42.8
Microfibril-associated glycoprotein 3	P55082	81/344	23.6
Microfibrillar-associated protein 3-like	O75121	148/381	38.9
Microfibril-associated glycoprotein 4	P55083	0/234	0.0
Microfibrillar-associated protein 5 (Microfibril-associated glycoprotein 2)	Q13361	27/152	17.8
Fibulin-1	P23142	10/674	1.5
Fibulin-2	P98095	318/1157	27.5
Fibulin-3	Q12805	53/476	11.1
Fibulin-4	O95967	34/418	8.1
Fibulin-5	Q9UBX5	1/425	0.2
Fibulin-6	Q96RW7	295/5614	5.3

Fibulin-7	Q53RD9	14/415	3.4
Lysyl oxidase (LOX)	P28300	153/396	38.6
Lysyl oxidase-like 1 (LOXL1)	Q08397	305/549	55.6
Lysyl oxidase-like 2 (LOXL2)	Q9Y4K0	5/749	0.7
Lysyl oxidase-like (LOXL3)	P58215	50/728	6.9
Lysyl oxidase-like (LOXL4)	Q96JB6	35/732	4.8

Supplementary Table 4S: Intrinsic disorder of the protein core of proteoglycans. The signal peptide has been removed from the sequence before prediction. Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Protein name	UniProtKB AC	Number of disordered amino acids /Number of all amino acids	Amount of predicted disorder (%)
Proteoglycans Hyalectans			
Aggrecan	P16112	1574/2399	65.6
Brevican	Q96GW7	307/889	34.5
Neurocan	O14594	718/1299	55.3
Versican	P13611	2121/3376	62.8
Small Leucine Rich Proteoglycans			
Biglycan (class I)	P21810	12/349	3.4
Decorin (class I)	P07585	23/343	6.7
Asporin (class I)	Q9BXN1	28/366	7.7
Extracellular matrix protein 2 (class I)	O94769	215/679	31.7
Fibromodulin (class II	Q06828	20/358	5.6
Keratocan (class II)	O60938	4/332	1.2
Lumican (class II)	P51884	0/320	0.0
Prolargin (PRELP) (class II)	P51888	47/362	13.0
Osteomodulin (osteoadherin) (class II)	Q99983	53/401	13.2
Epiphycan (class III)	Q99645	75/303	24.8
Opticin (class III)*	Q9UBM4	76/313	24.3
Mimecan (osteoglycin) (class III)	P20774	20/278	7.2
Chondroadherin (class IV)	O15335	6/337	1.8
Chondroadherin-like protein	Q6NUI6	77/732	10.5
Nyctalopin (class IV)	Q9GZU5	23/458	5.0
Tsukushi (class IV)	Q8WUA8	18/337	5.3
Podocan (class V)	Q7Z5L7	84/594	14.1
Podocan-like protein-1 (class V)	Q6PEZ8	5/486	1.0
Matrix-remodeling-associated protein 5 (Adlican, Adhesion	Q9NR99	1163/2802	41.5

protein with leucine-rich repeats and immunoglobulin domains related to perlecan)			
Basement membrane proteoglycans			
Agrin	O00468	308/2016	15.3
Bamacan	Q9UQE7	368/1217	30.2
Perlecan	P98160	934/4370	21.4
Leprecan-1	Q32P28	94/714	13.2
Leprecan like-1	Q8IVL5	78/684	11.4
Leprecan like-2	Q8IVL6	198/716	27.7
Leprecan like-4	Q92791	65/437	14.9
GPI-anchored proteoglycans			
Glypican-1	P35052	120/535	22.4
Glypican-2	Q8N158	161/556	29.0
Glypican-3	P51654	41/556	7.4
Glypican-4	O75487	67/538	12.5
Glypican-5	P78333	64/548	11.7
Glypican-6	Q9Y625	103/532	19.4
Membrane proteoglycans			
Syndecan-1	P18827	247/288	85.8
Syndecan-2	P34741	113/183	61.8
Syndecan-3	O75056	352/442	79.6
Syndecan-4	P31431	99/180	55.0
Neuropilin-1	O14786	92/902	10.2
Neuropilin-2	O60462	108/911	11.9
Neuroglycan	O95196	341/536	63.6
Betaglycan	Q03167	164/831	19.7
NG2	Q6UVK1	422/2293	18.4
CD44	P16070	497/722	68.8
Interphotoreceptor matrix proteoglycan2SPACRCAN	Q17R60	218/777	28.1
Endocan	Q3V4E3	23/115	20.0
Serglycin	P10124	62/131	47.3
Lubricin (proteoglycan 4)	Q92954	1021/1380	74.0
Testican-1	Q08629	62/418	14.8
Testican-2	Q92563	76/402	18.9
Testican-3	Q9BQ16	55/415	13.3

Supplementary Table 5S: Intrinsic disorder predicted in laminin human chains. The signal peptide has been removed from the sequence before prediction. Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Laminin chains	UniProtKB AC	Number of disordered amino acids /Number of all amino acids	Amount of predicted disorder (%)
α1	P25391	109/3058	3.6
α2	P24043	286/3100	9.2
α3	Q16787	358/3298	10.9
α4	Q16363	257/1799	14.3
α5	O15230	336/3660	9.2
β1	P07942	244/1765	13.8
β2	P55268	266/1766	15.1
β3	Q13751	150/1155	13.0
β4	A4D0S4	80/1742	4.6
γ1	P11047	385/1576	24.4
γ2	Q13753	261/1172	22.3
γ3	Q9Y6N6	230/1556	14.8

	Predicted disorder (%)				
Laminin chains	Laminin G-like 1	Laminin G-like 2	Laminin G-like 3	Laminin G-like 4	Laminin G-like 5
α1	11.1	7.3	0.0	0.0	5.5
α2	3.3	0.0	12.4	0.6	13.4
α3	13.4	12.3	0.0	1.2	10.9
α4	0.0	0.0	3.6	1.7	2.3
α5	10.3	0.0	15.4	25.2	0.0
Mean	7.7	3.9	6.3	5.7	6.4

Supplementary Table 6S: Intrinsic disorder of matricellular proteins. The signal peptide has been removed from the sequence before prediction. Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Protein name	UniProtKB AC	Number of disordered amino acids /Number of all amino acids	Amount of predicted disorder (%)
CCN1	O00622	3/357	0.8
CCN2	P29279	0/323	0.0
CCN3	P48745	4/326	1.2
CCN4	O95388	16/345	4.6
CCN5	O76076	13/227	5.7
CCN6	O95389	13/331	3.9
Osteopontin	P10451	294/298	98.7
Sparc	P09486	50/286	17.5
Smoc1	Q9H4F8	128/408	31.4
Smoc2	Q9H3U7	181/425	42.6
Hevin	Q14515	410/648	63.3
Tenascin C	P24821	281/2179	12.8
Tenascin X	P22105	2025/4266	47.5
Tsp1	P07996	237/1152	20.6
Tsp2	P35442	218/1154	18.9
Tsp3	P49746	222/934	23.8
Tsp4	P35443	196/935	21.0
Tsp5	P49747	238/737	32.3
Galectin 1	P09382	0/135	0.0
Galectin 2	P05162	16/132	12.1
Galectin 3	P17931	106/250	42.4
Galectin 4	P56470	34/323	10.5
Galectin 7	P47929	27/136	19.9
Galectin 8	O00214	6/317	1.9
Galectin 9	O00182	31/355	8.7
Galectin 10	Q05315	0/142	0.0
Galectin 12	Q96DT0	5/336	1.5
Galectin 13	Q9UHV8	0/139	0.0
Galectin 14	Q8TCE9	0/139	0.0
Plasminogen Activator Inhibitor type 1	P05121	0/379	0.0
Autotaxin	Q13822	36/836	4.3
Periostin	Q15063	49/815	6.0

Supplementary Table 7S: Intrinsic disorder of Small Integrin Binding N-Linked Glycoproteins (SIBLINGs). The signal peptide has been removed from the sequence before prediction. Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Protein name	UniProtKB AC	Number of disordered amino acids /Number of all amino acids	Amount of predicted disorder (%)
Bone sialoprotein	P21815	274/301	91.0
Dentin matrix protein 1	Q13316	497/497	100.0
Dentin sialophosphoprotein	Q9NZW4	1285/1286	99.9
Matrix extracellular phosphoglycoprotein	Q9NQ76	492/508	96.9
Osteopontin	P10451	294/298	98.7

Supplementary Table 8S: Intrinsic disorder of matrix metalloproteinases (MMPs).The signal peptide has been removed from the sequence before prediction. Intrinsic disorder has been predicted from protein sequences with IUPred[16,17] as detailed in the Material and Methods section. The percentage of intrinsic disorder was calculated by counting all the disordered amino acids in the protein, dividing it by the number of all amino acids in the protein and multiplying it by 100.[18]

Protein name	UniProtKB AC	Number of disordered amino acids /Number of all amino acids	Amount of predicted disorder (%)
Collagenases			
MMP1	P03956	66/450	14.7
MMP8	P22894	62/447	13.9
MMP13	P45452	39/452	8.6
Gelatinases			
MMP2	P08253	13/631	2.1
MMP9	P14780	84/688	12.2
Stromelysins			
MMP3	P08254	48/460	10.4
MMP10	P09238	30/459	6.5
Matrilysins			
MMP7	P09237	25/250	10.0
MMP26	Q9NRE1	13/244	5.3
MMP11	P24347	83/476	17.4
Membrane-type MMPs			
Transmembrane type			
MMP14	P50281	72/562	12.8
MMP15	P51511	137/628	21.8
MMP16	P51512	94/576	16.3
MMP24	Q9Y5R2	86/593	14.5
GPI-anchored			
MMP17	Q9ULZ9	124/568	21.8
MMP25	Q9NPA2	178/541	32.9
Others			
MMP12	P39900	18/454	4.0
MMP19	Q99542	71/490	14.5
MMP20	O60882	17/461	3.7
MMP21	Q8N119	124/545	22.8
MMP23	O75900	19/390	4.9
MMP27	Q9H306	18/496	3.6
MMP28	Q9H239	22/498	4.4

Supplementary Table 9S: Human proteins annotated with the UniProtKB keyword "extracellular matrix", predicted to contain more than 50% of global disorder and included in the analysis using DAVID bioinformatics resource.[14,15]

UniProtKB AC	DAVID gene identifier	DAVID Gene Name
P16112	802704	aggrecan
Q9NP70	797689	ameloblastin (enamel matrix protein)
Q99217	804457	amelogenin (amelogenesis imperfecta 1, X-linked)
Q99218	810092	amelogenin, Y-linked
P27797	788614	calreticulin
Q6UXA7	772442	chromosome 6 open reading frame 15 (protein STG)
P02452	808600	collagen, type I, alpha 1
P08123	781103	collagen, type I, alpha 2
P02458	787944	collagen, type II, alpha 1
P02461	795569	collagen, type III, alpha 1
P02462	800275	collagen, type IV, alpha 1
P08572	804645	collagen, type IV, alpha 2
Q01955	824817	collagen, type IV, alpha 3 (Goodpasture antigen)
P53420	819591	collagen, type IV, alpha 4
P29400	784210	collagen, type IV, alpha 5
Q14031	812933	collagen, type IV, alpha 6
P20849	826638	collagen, type IX, alpha 1
Q14055	788090	collagen, type IX, alpha 2
Q14050	780333	collagen, type IX, alpha 3
P20908	780010	collagen, type V, alpha 1
P05997	814924	collagen, type V, alpha 2
P25940	810626	collagen, type V, alpha 3
Q02388	774727	collagen, type VII, alpha 1
P27658	787706	collagen, type VIII, alpha 1
P25067	772627	collagen, type VIII, alpha 2
Q03692	803916	collagen, type X, alpha 1
P12107	777404	collagen, type XI, alpha 1
P13942	803177	collagen, type XI, alpha 2
Q14993	774673	collagen, type XIX, alpha 1
P39059	826696	collagen, type XV, alpha 1
Q07092	823173	collagen, type XVI, alpha 1
Q9UMD9	789642	collagen, type XVII, alpha 1
P39060	777598	collagen, type XVIII, alpha 1
Q96P44	783035	collagen, type XXI, alpha 1
Q8NFW1	776338	collagen, type XXII, alpha 1
Q17RW2	782335	collagen, type XXIV, alpha 1
Q8IZC6	792574	collagen, type XXVII, alpha 1

Q2UY09	791505	collagen, type XXVIII, alpha 1
Q13316	772614	dentin matrix acidic phosphoprotein 1
Q9NZW4	773812	dentin sialophosphoprotein
Q96A84	793960	EMI domain containing 1
Q96A83	774209	EMI domain containing 2 (collagen, type XXVI, alpha 1)
Q9NRM1	817683	enamelin
O00515	814366	ladinin 1
Q9NQ76	810893	matrix extracellular phosphoglycoprotein
P55081	783592	microfibrillar-associated protein 1
Q6XPR3	778770	repetin
Q14515	792300	SPARC-like 1 (hevin)
Q8IWL2	781929	surfactant protein A1B; surfactant protein A1
Q8IWL1	808468	surfactant protein A2; surfactant protein A2B
P35247	783108	surfactant protein D
Q8WVF2	781416	upper zone of growth plate and cartilage matrix associated (UCMA)
P13611	800485	versican