

Evolution of Protein Structure and Function: Conserved Domains

Dr. Stephen R. Campion¹ & David B. Pierce¹

1. Alvernia University, 400 Saint Bernadine St. Reading, PA 19607

This study is part of an ongoing exploration into the evolution of protein structure and function. Individual proteins that share a common element or domain of protein structure have been compiled and compared in order to characterize which of the available chemical functionalities inherent in the twenty common amino acids, are most heavily preserved throughout evolution, as proteins diverge over time to generate diverse protein functions from a common primordial ancestor. This study focuses specifically on a compiled family of protein structures containing one or more LINK domains within their overall protein architecture. These domains are unique to extracellular matrix (ECM) proteins that comprise the "core" of a variety of proteoglycan structures that are critical to cellular organization and communication within biological tissues and organs. The conservation of both polar and non-polar amino acids have been identified at strategic locations with the LINK domain consensus sequence. Mapping of these amino acids within the LNK 3D structure is currently limited by the unavailability of diverse LINK structural data.

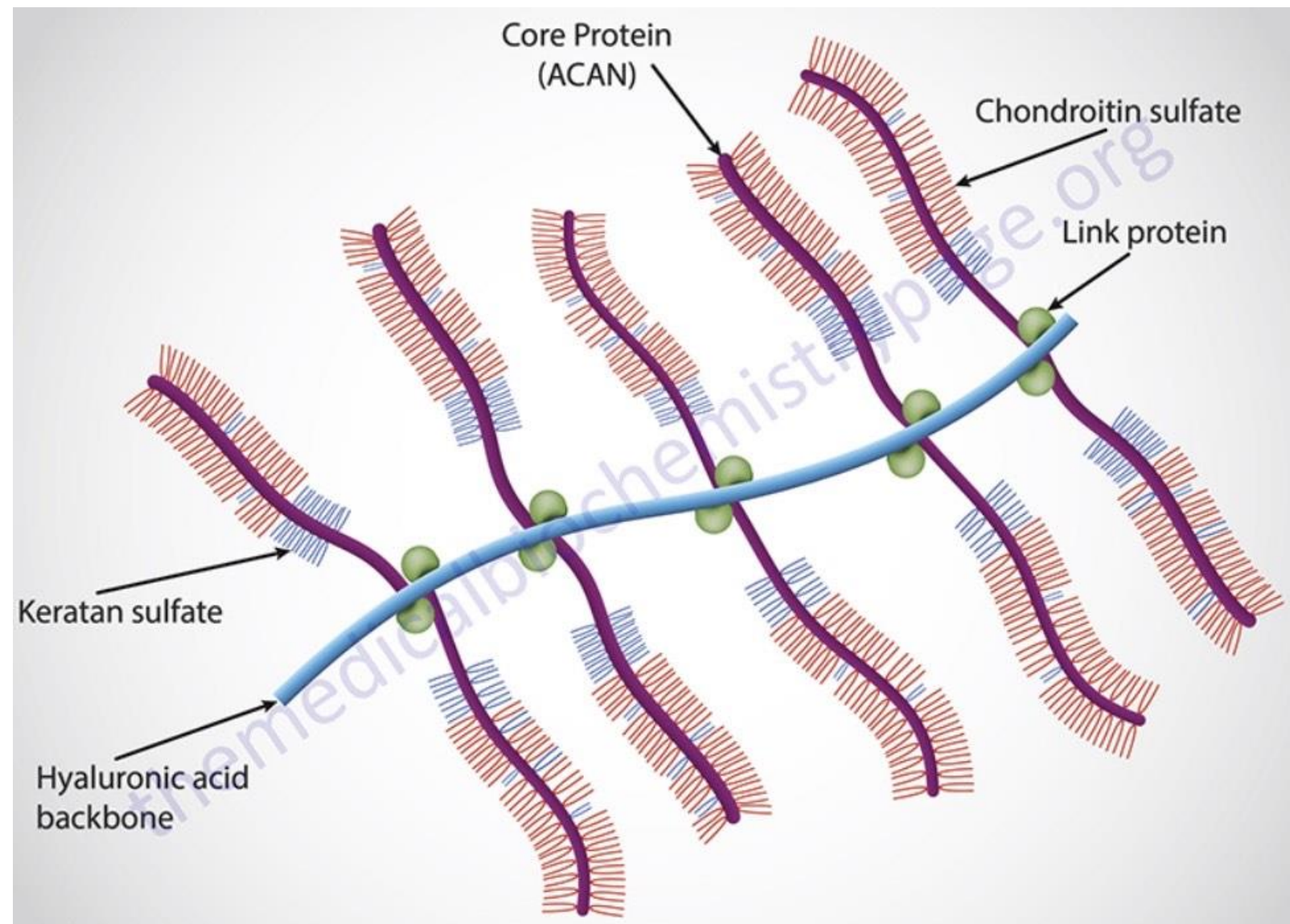


Figure 1: An image depicting the interactions between glycosaminoglycans, core proteins such as that of aggrecan, link proteins, and the hyaluronic acid backbone. <https://images.app.goo.gl/14GVjtHriMob2Exv5>

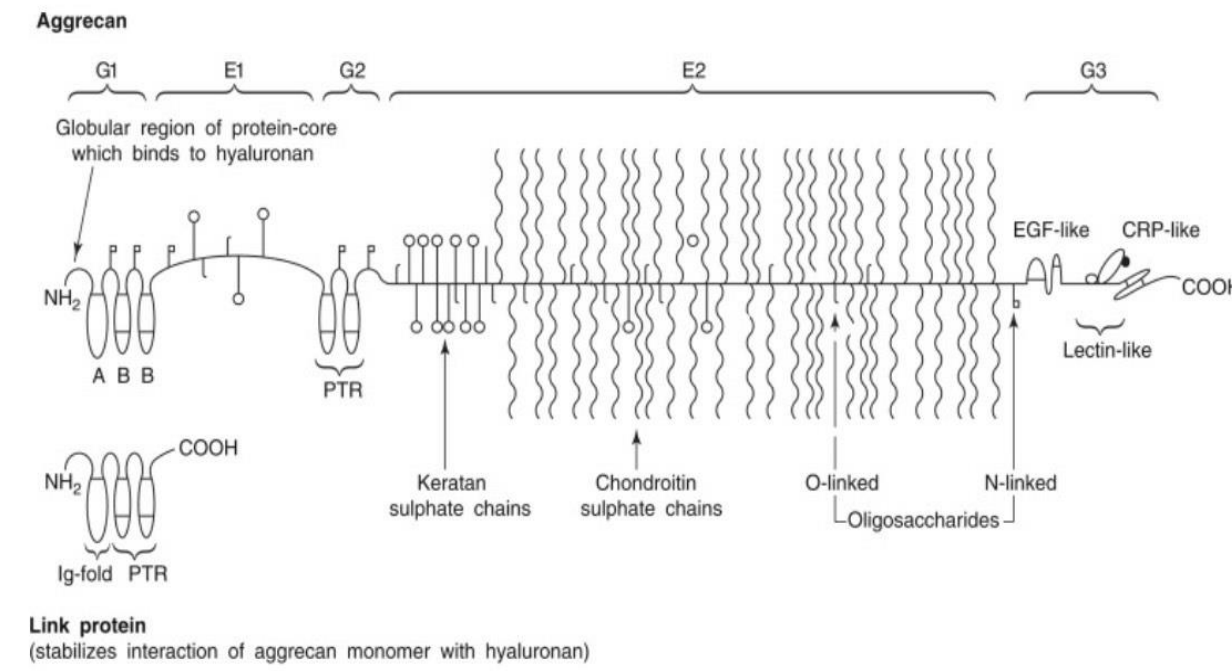


Figure 3: An image highlighting the regions and interactions of GAGs and saccharides within aggrecan. <https://images.app.goo.gl/GltpT72VGv2FhVd8>

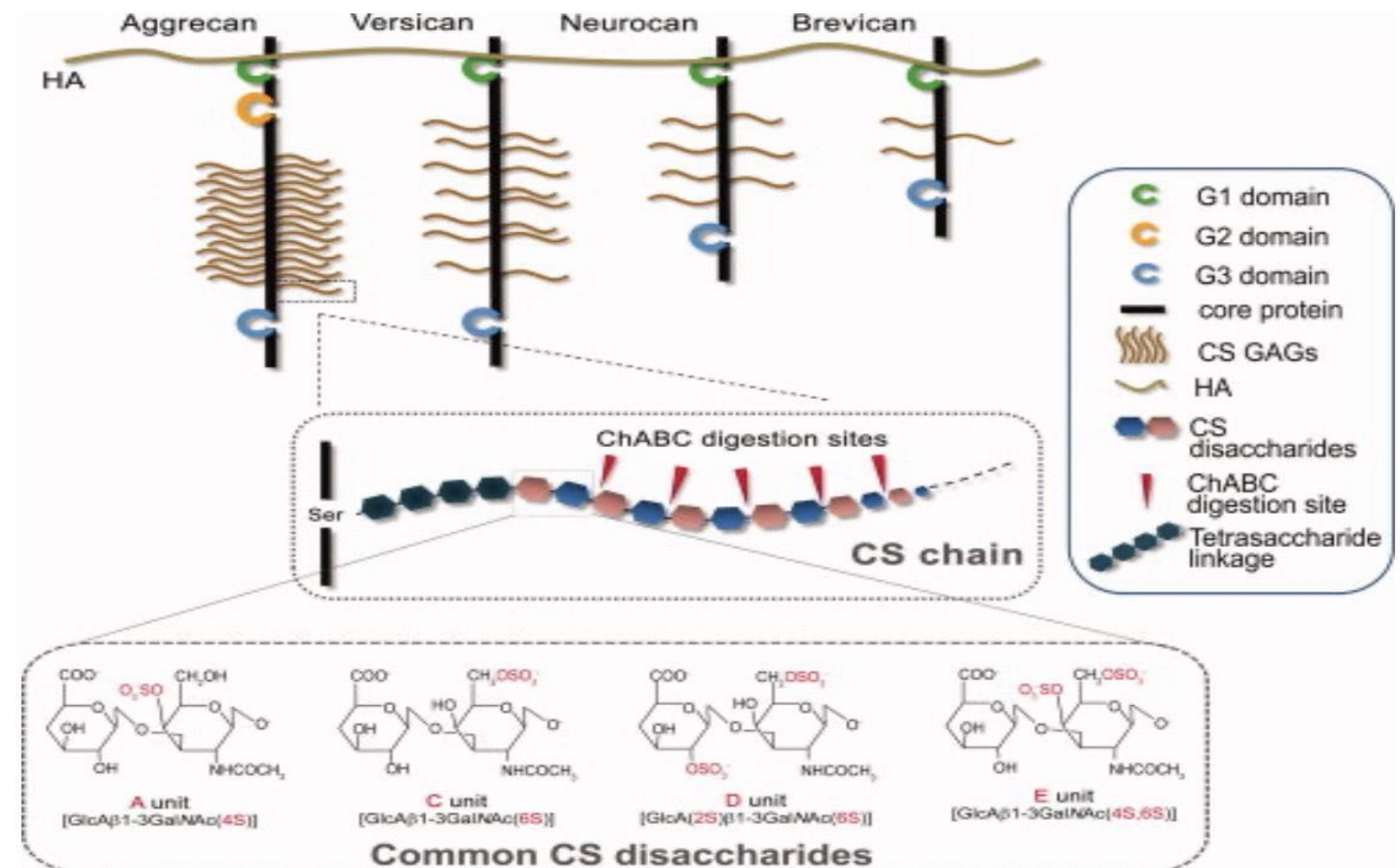


Figure 4: An image illustrating a more in-depth look at chondroitin sulfate and chondroitin sulfate disaccharides in relation to Aggrecan, Versican, Neurocan, and Brevican and their respective domains. <https://images.app.goo.gl/n1j9sd3uhZ2Sf7yx8>

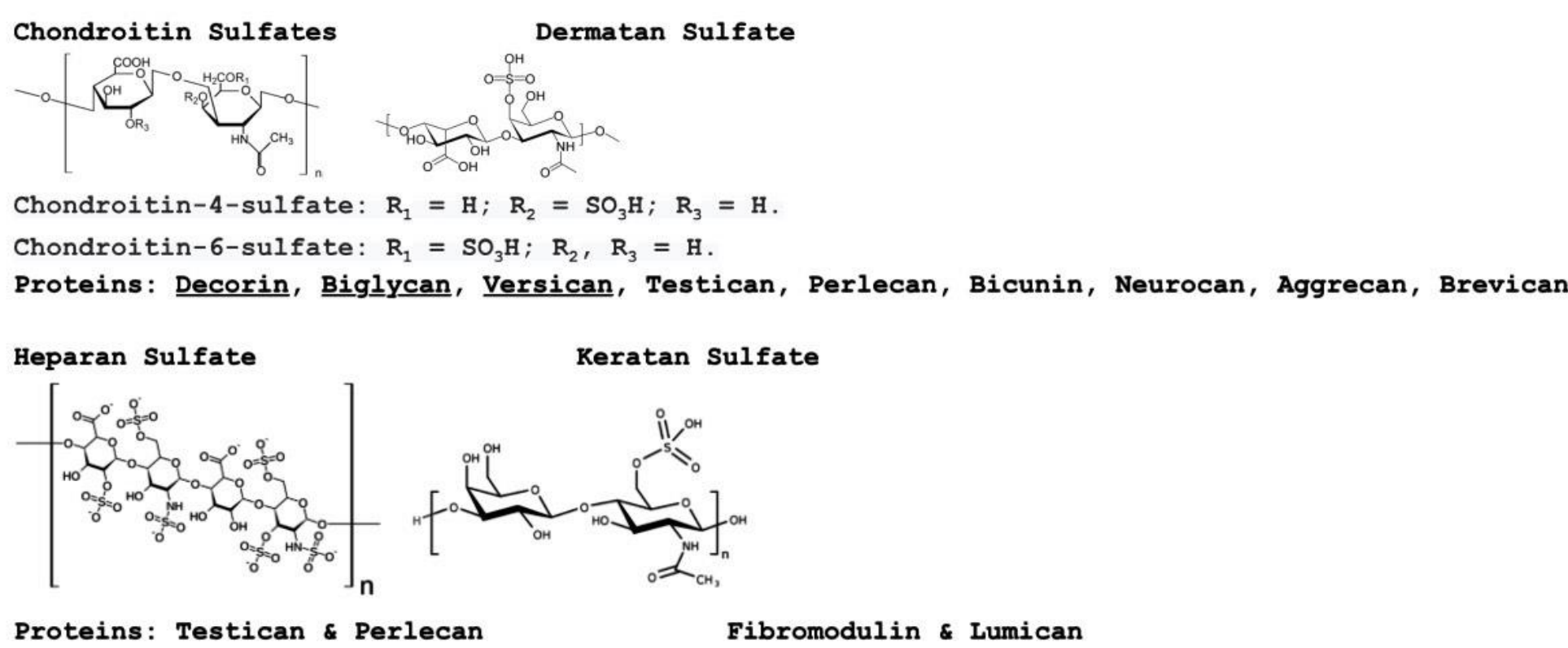


Figure 2: The structures of the four major Glycosaminoglycan groups relevant to this study and their proteins.

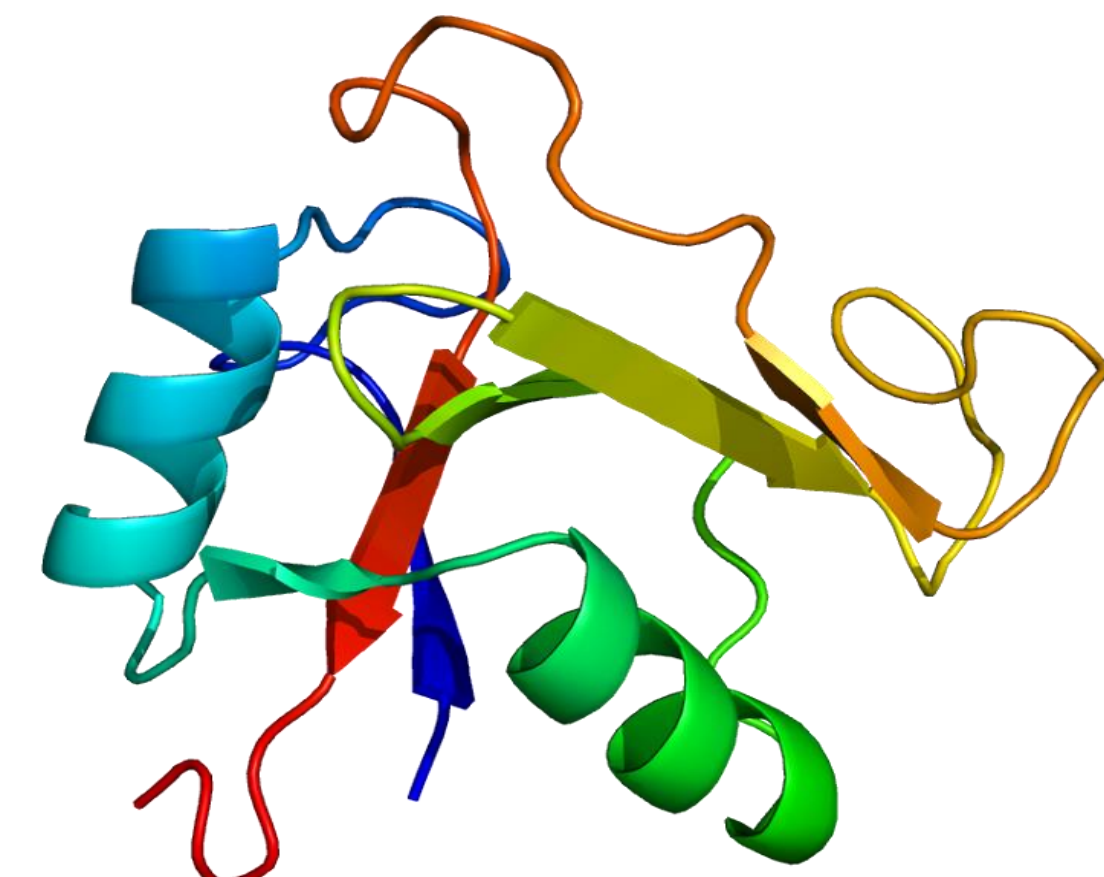


Figure 5: A 3-dimensional model of Tumor Necrosis Factor-Inducible Protein Tsg-6. <https://images.app.goo.gl/5URnhf4k8HDi2PLJ8>

Sequencing, Protein, and Organism¹

•XP_003261610 249 DGDVDFHVT---PSKFTFEEAAKE C E-NQDARLATVGELOAAWR-NGFDQC DY GW LSDASVRHPVTVARA--Q C GGGLLVRTLYRFENQTGFPPPPDSRFDAY C FK 347 Versican (Northern White-Cheeked Gibbon)
 •AAI34426 155 KGVVFLYREGSARYAFSFAQAQA C A-RIGARIATPEQLYAYL-GGYEQ C DA GW LSDQTVRYPIQTPRE--A C YGDMDFGFPVGRVNYGVVDPDDLYDVY C YA 252 Brevican (Cattle)
 •XP_003438010 256 EVFHSS-VPG---RLSLSSADRC V-SLGGQLATVGGQLYAWR-AGLDSC AP GW LSDGSVRYVPTWPR--D C GGTEPGVRTVTPNDTADNTTALYDAY C Y 348 Neurocan (Nile Tilapia)
 •XP_016854347 617 EVFFAT-QLE---QFTLLEAQEY C E-SQNTLASAQQLYAAWK-LGLDN C RA GW LADGSLRYPITNPR--G C GGDKPGVRTVLYHNTQGFDPDQSKHHAF C F 712 Aggrecan (Green Anole)
 •XP_005802294 32 GVFLVE-GESR-YILTFEMAMKV C E-QLESTMASPAQVQEAAYN-NSMET C RY GW INNGSSVILRHTQHE--N C AKNKTGFIVNSYTGDFHDAH C Y 120 CD44 Antigen (Southern Platyfish)
 •EDM09988 157 QGVVFPYFPRGLGRYLNLFHEARQA C L-DQDAVIASFQDLYDAWR-GGLDW C NA GW LSDGSVQYPIKPRE--P C GGQNTVPGVGRVNYGFWDKDKSRYDVF C FT 254 Hyaluronan and Proteoglycan Link Protein 1 (Norway Rat)
 •XP_003413832 263 RVYYLE-HPE---KLTLTEAKAA C Q-EDGAQIAKVGQLFAAWKFRGLDR C DA GW LADGSVRYVPAHPRP--N C GSAEPGVRSFGFPDPQSHKYGVY C Y 353 Hyaluronan and Proteoglycan Link Protein 3 (African Savanna Elephant)
 •XP_011277789 2209 GVFHLQ-APSGPYGLNFSEAKAA C G-AQGAVALASLPQLSAAQK-LGFHLC RV GW LANGSAAHPVVFPA--D C GGGQLGVISLGRKNHRSERWDTY C Y 2301 Stabilin-1 (Domestic Cat)
 •EGW09706 1449 GVFLHR-SPLGQYKLTDFDCAKEA C A-NEAASMATYNQLSYAQK-ASYHL C SA GW LENGVRGYPTTYASP--K C GSNIVGIVDYGIRPNKSEMWDV C Y 1541 Stabilin-2 (Chinese Hamster)
 •XP_007245828 31 RVFLLS-LRNGSNG-GYEEASQA C E-AQDARLASGMELRHAVVECSFSA C TQ GW LTGPSIGT-----T--V C RSMAGSLRAVNMRENVTETSERLAAF C V 121 Sushi Domain-Containing Protein 5 (Mexican Tetra)
 •2PF5_A 1 GVYHRE-ARSGKYKLTAEAKAV C E-FEGGHLATYKQLEAARK-IGFHV C AA GW MAKGRVGYPIVKPGP--N C GFGKGTIIDYGIIRLNRSERWDAY C Y 93 Tumor Necrosis Factor-Inducible Protein Tsg-6 (Human)
 •XP_002942315 40 VLLAES-SDHK-YKFNFTRAESV C H-VLGLQLASKNQVEKANK-YGFET C SF GW VSEQVVISRIQPNE--K C GQSKTGLVPMVMTVNAAKDFHAY C F 130 Lymphatic Vessel Endothelial Hyaluronan Acid Receptor 1 (Western Clawed Frog)

KEY

I = Immunoglobulin-like domain
 L = Link domain
 HY = Hyaluronan-binding domain
 Ca = Calcium-binding domain
 E = EGF-like domain
 CL = C-Lectin domain
 FN = Fibronectin type domain
 SR = Sushi repeats
 CH = Carbohydrate-binding domain
 LG = Ligan-binding domain
 RL = Receptor-ligand interactions
 F = Fasciclin domain
 E3 = EGF-3 domain
 M = Midasin domain
 TM = Transmembrane domain
 HTD = Heterodimer interface
 HMD = Homodimer Interface
 S = Sheet
 H = Helix
 DS = Disulfide bond

Conserved Domains¹

•XP_003261610 I (36-151)--L (150-244)--HY (160-161)--L (251-346)--HY (258-259)--E (1336-1370)--E (1372-1408)--C (1372,1375,1389)--CL (1414-1537)--FN (1527-1529)--Ca (1475,1479,1502,1508,1509)--Ca (1479,1509)--CH (1499,1501,1508,1522,1523)--SR (1541-1597)--RL (1551-1571)
 •AAI34426 I (44-154)--I (49-158)--L (157-251)--HY (167-168)--L (258-353)--HY (255-266)
 •XP_003438010 I (45-156)--I (47-152)--L (155-249)--HY (165-166)--L (256-348)--HY (263-264)--E (1002-1032)--E (1036-1072)--Ca (1036,1039,1053)--CL (1078-1201)--LG (1172,1176,1178,1184,1186-1189,1192-1194)--SR (1205-1261)--RL (1215,1235)
 •XP_016854347 (Aggrecan core protein only)
 •XP_005802294 L (32-120)--HY (41-42)
 •EDM09988 I (53-160)--I (54-155)--L (159-253)--HY (169-170)--L (258-339)
 •XP_003413832 I (58-163)--I (59-158)--L (162-256)--HY (172-173)--L (263-353)--HY (270-271)
 •XP_011277789 F (394-491)--F (520-644)--E3 (832-860)--E3 (915-946)--E3 (952-988)--F (1000-1121)--F (1140-1242)--E3 (1460-1496)--E3 (1502-1539)--E3 (1545-1582)--F (16-7-1711)--F (1737-1867)--E3 (2096-2131)--E3 (2137-2174)--L (2209-2301)--HY (2219-2220)--(four repeated fasciclin 1 domains 2368-2463)
 •EGW09706 E3 (118-153)--F (174-296)--F (293-383)--E3 (517-545)--E3 (554-589)--E3 (599-642)--F (650-758)--F (758-865)--E3 (951-987)--E3 (993-1029)--F (1060-1160)--F (1186-1258)--E3 (1337-1372)--E3 (1378-1415)--L (1449-1541)--HY (1459-1460,1507,1518,1526)--(four repeated fasciclin 1 domains 1607-1699)
 •XP_007245828 L (31-121)--HY (41-42)--SR (128-186)--RL (138,156)--M (220-429)--TM (657-693)--HTD (658-659,661,664,667,670-672,674-675,678-679,682)--HMD (659-662,664-667,670-672,674-676,678-679,682)
 •2PF5_A L (1-93)--S (1-6)--HY (11,12,59,70,78)--H (16-26)--DS (23,92)--H (33-42)--S (49-53)--S (55-62)--S (73-81)--S (88-94)
 •XP_002942315 L (33-179)--HY (49-50)

Acknowledgements:

This research was funded by Alvernia University

References:

1. <https://www.ncbi.nlm.nih.gov/guide/proteins/>



David.Pierce@alvernia.edu