Extracellular matrix organization


European Bioinformatics Institute, New York University Langone Medical Center, Ontario Institute for Cancer Research, Oregon Health and Science University.

The contents of this document may be freely copied and distributed in any media, provided the authors, plus the institutions, are credited, as stated under the terms of Creative Commons Attribution 4.0 International (CC BY 4.0) License. For more information see our license.

27/03/2022
Introduction

Reactome is open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. A system of evidence tracking ensures that all assertions are backed up by the primary literature. Reactome is used by clinicians, geneticists, genomics researchers, and molecular biologists to interpret the results of high-throughput experimental studies, by bioinformaticians seeking to develop novel algorithms for mining knowledge from genomic studies, and by systems biologists building predictive models of normal and disease variant pathways.

The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751), University of Toronto (CFREF Medicine by Design), European Union (EU STRP, EMI-CD), and the European Molecular Biology Laboratory (EBI Industry program).

Literature references


Reactome database release: 79

This document contains 10 pathways and 1 reaction (see Table of Contents)
The extracellular matrix (ECM) is a component of all mammalian tissues, a network consisting largely of the fibrous proteins collagen, elastin and associated-microfibrils, fibronectin and laminins embedded in a viscoelastic gel of anionic proteoglycan polymers. It performs many functions in addition to its structural role; as a major component of the cellular microenvironment it influences cell behaviours such as proliferation, adhesion and migration, and regulates cell differentiation and death (Hynes 2009).

ECM composition is highly heterogeneous and dynamic, being constantly remodeled (Frantz et al. 2010) and modulated, largely by matrix metalloproteinases (MMPs) and growth factors that bind to the ECM influencing the synthesis, crosslinking and degradation of ECM components (Hynes 2009). ECM remodeling is involved in the regulation of cell differentiation processes such as the establishment and maintenance of stem cell niches, branching morphogenesis, angiogenesis, bone remodeling, and wound repair. Redundant mechanisms modulate the expression and function of ECM modifying enzymes. Abnormal ECM dynamics can lead to deregulated cell proliferation and invasion, failure of cell death, and loss of cell differentiation, resulting in congenital defects and pathological processes including tissue fibrosis and cancer.

Collagen is the most abundant fibrous protein within the ECM constituting up to 30% of total protein in multicellular animals. Collagen provides tensile strength. It associates with elastic fibres, composed of elastin and fibrillin microfibrils, which give tissues the ability to recover after stretching. Other ECM proteins such as fibronectin, laminins, and matricellular proteins participate as connectors or linking proteins (Daley et al. 2008).

Chondroitin sulfate, dermatan sulfate and keratan sulfate proteoglycans are structural components associated with collagen fibrils (Scott & Haigh 1985; Scott & Orford 1981), serving to tether the fibril to the surrounding matrix. Decorin belongs to the small leucine-rich repeat proteoglycan family (SLRPs) which also includes biglycan, fibromodulin, lumican and asporin. All appear to be involved in collagen fibril formation and matrix assembly (Ameye & Young 2002).

ECM proteins such as osteonectin (SPARC), osteopontin and thrombospondins -1 and -2, collectively referred to as matricellular proteins (reviewed in Mosher & Adams 2012) appear to modulate cell-matrix in-
teractions. In general they induce de-adhesion, characterized by disruption of focal adhesions and a reorganization of actin stress fibers (Bornstein 2009). Thrombospondin (TS)-1 and -2 bind MMP2. The resulting complex is endocytosed by the low-density lipoprotein receptor-related protein (LRP), clearing MMP2 from the ECM (Yang et al. 2001).

Osteopontin (SPP1, bone sialoprotein-1) interacts with collagen and fibronectin (Mukherjee et al. 1995). It also contains several cell adhesive domains that interact with integrins and CD44.

Aggrecan is the predominant ECM proteoglycan in cartilage (Hardingham & Fosang 1992). Its relatives include versican, neurocan and brevican (Iozzo 1998). In articular cartilage the major non-fibrous macromolecules are aggrecan, hyaluronan and hyaluronan and proteoglycan link protein 1 (HAPLN1). The high negative charge density of these molecules leads to the binding of large amounts of water (Bruckner 2006). Hyaluronan is bound by several large proteoglycans proteoglycans belonging to the hyalectan family that form high-molecular weight aggregates (Roughley 2006), accounting for the turgid nature of cartilage.

The most significant enzymes in ECM remodeling are the Matrix Metalloproteinase (MMP) and A disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) families (Cawston & Young 2010). Other notable ECM degrading enzymes include plasmin and cathepsin G. Many ECM proteinases are initially present as precursors, activated by proteolytic processing. MMP precursors include an amino pro-domain which masks the catalytic Zn-binding motif (Page-McCaw et al. 2007). This can be removed by other proteinases, often other MMPs. ECM proteinases can be inactivated by degradation, or blocked by inhibitors. Some of these inhibitors, including alpha2-macroglobulin, alpha1-proteinase inhibitor, and alpha1-chymotrypsin can inhibit a large variety of proteinases (Woessner & Nagase 2000). The tissue inhibitors of metalloproteinases (TIMPs) are potent MMP inhibitors (Brew & Nagase 2010).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Edition</th>
<th>Author/Reviewer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-09-09</td>
<td>Authored</td>
</tr>
<tr>
<td>2012-02-21</td>
<td>Edited</td>
</tr>
<tr>
<td>2012-02-28</td>
<td>Reviewed</td>
</tr>
<tr>
<td>2013-05-21</td>
<td>Reviewed</td>
</tr>
<tr>
<td>2013-05-22</td>
<td>Reviewed</td>
</tr>
</tbody>
</table>

https://reactome.org
Collagen is a family of at least 29 structural proteins derived from over 40 human genes (Myllyharju & Kivirikko 2004). It is the main component of connective tissue, and the most abundant protein in mammals making up about 25% to 35% of whole-body protein content. A defining feature of collagens is the formation of trimeric left-handed polyproline II-type helical collagenous regions. The packing within these regions is made possible by the presence of the smallest amino acid, glycine, at every third residue, resulting in a repeating motif Gly-X-Y where X is often proline (Pro) and Y often 4-hydroxyproline (4Hyp). Gly-Pro-Hyp is the most common triplet in collagen (Ramshaw et al. 1998). Collagen peptide chains also have non-collagenous domains, with collagen subclasses having common chain structures. Collagen fibrils are mostly found in fibrous tissues such as tendon, ligament and skin. Other forms of collagen are abundant in cornea, cartilage, bone, blood vessels, the gut, and intervertebral disc. In muscle tissue, collagen is a major component of the endomysium, constituting up to 6% of muscle mass. Gelatin, used in food and industry, is collagen that has been irreversibly hydrolyzed.

On the basis of their fibre architecture in tissues, the genetically distinct collagens have been divided into subgroups. Group 1 collagens have uninterrupted triple-helical domains of about 300 nm, forming large extracellular fibrils. They are referred to as the fibril-forming collagens, consisting of collagens types I, II, III, V, XI, XXIV and XXVII. Group 2 collagens are types IV and VII, which have extended triple helices (>350 nm) with imperfections in the Gly-X-Y repeat sequences. Group 3 are the short-chain collagens. These have two subgroups. Group 3A have continuous triple-helical domains (type VI, VIII and X). Group 3B have interrupted triple-helical domains, referred to as the fibril-associated collagens with interrupted triple helices (FACIT collagens, Shaw & Olsen 1991). FACITs include collagen IX, XII, XIV, XVI, XIX, XX, XXI, XXII and XXVI plus the transmembrane collagens (XIII, XVII, XXIII and XXV) and the multiple triple helix domains and interruptions (Multiplexin) collagens XV and XVIII (Myllyharju & Kivirikko 2004). The non-collagenous domains of collagens have regulatory functions; several are biologically active when cleaved from the main peptide chain. Fibrillar collagen peptides all have a large triple helical domain (COL1) bordered by N and C terminal extensions, called the N- and C-propeptides, which are
cleaved prior to formation of the collagen fibril. The intact form is referred to as a collagen propeptide, not procollagen, which is used to refer to the trimeric triple-helical precursor of collagen before the propeptides are removed. The C-propeptide, also called the NC1 domain, directs chain association during assembly of the procollagen molecule from its three constituent alpha chains (Hulmes 2002).

Fibril forming collagens are the most familiar and best studied subgroup. Collagen fibres are aggregates or bundles of collagen fibrils, which are themselves polymers of tropocollagen complexes, each consisting of three polypeptide chains known as alpha chains. Tropocollagens are considered the subunit of larger collagen structures. They are approximately 300 nm long and 1.5 nm in diameter, with a left-handed triple-helical structure, which becomes twisted into a right-handed coiled-coil 'super helix' in the collagen fibril. Tropocollagens in the extracellular space polymerize spontaneously with regularly staggered ends (Hulmes 2002). In fibrillar collagens the molecules are staggered by about 67 nm, a unit known as D that changes depending upon the hydration state. Each D-period contains slightly more than four collagen molecules so that every D-period repeat of the microfibril has a region containing five molecules in cross-section, called the 'overlap', and a region containing only four molecules, called the 'gap'. The triple-helices are arranged in a hexagonal or quasi-hexagonal array in cross-section, in both the gap and overlap regions (Orgel et al. 2006). Collagen molecules cross-link covalently to each other via lysine and hydroxylsine side chains. These cross-links are unusual, occurring only in collagen and elastin, a related protein.

The macromolecular structures of collagen are diverse. Several group 3 collagens associate with larger collagen fibers, serving as molecular bridges which stabilize the organization of the extracellular matrix. Type IV collagen is arranged in an interlacing network within the dermal-epidermal junction and vascular basement membranes. Type VI collagen forms distinct microfibrils called beaded filaments. Type VII collagen forms anchoring fibrils. Type VIII and X collagens form hexagonal networks. Type XVII collagen is a component of hemidesmosomes where it is complexed with alpha6Beta4 integrin, plectin, and laminin-332 (de Pereda et al. 2009). Type XXIX collagen has been recently reported to be a putative epidermal collagen with highest expression in suprabasal layers (Soderhall et al. 2007). Collagen fibrils/aggregates arranged in varying combinations and concentrations in different tissues provide specific tissue properties. In bone, collagen triple helices lie in a parallel, staggered array with 40 nm gaps between the ends of the tropocollagen subunits, which probably serve as nucleation sites for the deposition of crystals of the mineral component, hydroxyapatite (Ca10(PO4)6(OH)2) with some phosphate. Collagen structure affects cell-cell and cell-matrix communication, tissue construction in growth and repair, and is changed in development and disease (Sweeney et al. 2006, Twardowski et al. 2007). A single collagen fibril can be heterogeneous along its axis, with significantly different mechanical properties in the gap and overlap regions, correlating with the different molecular organizations in these regions (Minard-Jolandan & Yu 2009).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-08-05</td>
<td>Authored</td>
<td>Jupe, S.</td>
</tr>
<tr>
<td>2012-04-11</td>
<td>Edited</td>
<td>Jupe, S.</td>
</tr>
<tr>
<td>2012-05-24</td>
<td>Reviewed</td>
<td>Canty-Laird, EG.</td>
</tr>
</tbody>
</table>
Fibronectin (FN1) is found in the extracellular matrix (ECM) of all cells as linear and branched networks that surround and connect neighbouring cells (Singh et al. 2010). Prior to matrix formation FN1 exists as a protein dimer. Often the two peptide chains represent differentially-spliced variants. The chains are linked by a pair of C-terminal disulfide bonds which are essential for subsequent multimerization (Schwarzbaur 1991). FN1 monomers have a molecular weight of 230-270 kDa depending on alternative splicing and contain three types of repeating unit, I, II, and III. I and II are stabilized by intra-chain disulfide bonds. The absence of disulfide bonds in type III modules allows them to partially unfold under applied force (Erickson 2002). Three regions of variable splicing occur along the length of the FN1 monomer (Mao & Schwarzbauer 2005). One or both of the 'extra' type III modules EIIIA and EIIIB may be present in cellular FN1, but never in plasma FN1. A variable (V) region exists between the 14th and 15th type III module. This contains the binding site for alpha4 beta1 and alpha4beta7 integrins. It is present in most cellular FN1, occasionally in plasma FN1. The modules are arranged into several functional and protein-binding domains. There are four FN1-binding domains (Mao & Schwarzbauer 2005). One of these domains (I1-5), referred to as the 'assembly domain', is required for the initiation of FN1 matrix assembly. Modules III9-10 correspond to the 'cell-binding domain' of FN1. The Arg-Gly-Asp (RGD) integrin binding sequence located in III10 is the primary site of FN1 to cell attachment, mediated predominantly by alpha5 beta1 and alphaV beta3 integrins. The 'synergy site' in III9 modulates FN1's association with alpha5 beta1 integrins. FN1 also contains interaction domains for fibrin (I1-5, I10-12), collagen (I6-9, III-2), fibulin-1 (III13-14), heparin, syndecan (III12-14) and fibrillin-1 (I6-9) (Mao & Schwartzbauer 2005, Sabatier et al. 2009).

FN1 dimer binding to alpha5beta1 integrin stimulates self-association. Binding is thought to lead to a conformational change in FN1 that triggers the addition of further FN1 dimers (Singh et al. 2010). I1-5 functions as a unit that is the primary FN1 matrix assembly domain (Sottile et al. 1991) but other units are likely to be involved (Singh et al. 2010), the process is not fully understood.
Several ECM proteins appear to require the FN1 matrix for their own assembly. Fibrillin-1 containing microfibrils are formed when fibrillin-1 multimers bind to the FN1 matrix (Sabatier et al. 2009). FN1 polymerization promotes the deposition of type I and type III collagen (Sottile and Hocking 2002, Velling et al. 2002). Inhibition of FN1 polymerization increases its turnover and a concomitant loss of collagen types I and III from the ECM (Sottile and Hocking 2002, Sottile et al. 2007). FN1 is regulated by matrix metalloproteinases, particularly MMP14 (Shi & Sottile 2011).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Role</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-07-12</td>
<td>Authored, Edited</td>
<td>Jupe, S.</td>
</tr>
<tr>
<td>2013-02-08</td>
<td>Reviewed</td>
<td>Reinhardt, DP.</td>
</tr>
</tbody>
</table>
FN1 binds Collagen types I-V, VII

Location: Extracellular matrix organization

Stable identifier: R-HSA-2327733