J Bone Miner Metab (2000) 18:51–56

Journal of
Bone and
Mineral
Metabolism

Review article

Aggrecan from start to finish

Wei Luo[1], Chunxia Guo[1], Jing Zheng[1], Tung-Ling Chen[2], Pei Yin Wang[2], Barbara M. Vertel[2], and Marvin L. Tanzer[1]

[1] Department of BioStructure & Function, School of Dental Medicine, University of Connecticut Health Center, Farmington, CT 06030-3705, USA
[2] Department of Cell Biology and Anatomy, Finch University of Health Sciences/The Chicago Medical School, North Chicago, IL, USA

Key words: domains, chaperones, trafficking, misdirection, posttranslational modifications

Introduction

Aggrecan, the large aggregating chondroitin sulfate proteoglycan, first discovered in articular cartilage, is a prototypical molecule in the proteoglycan repertoire. As a consequence of the ability of aggrecan to trap large volumes of water, articular cartilage is able to resist physical compression. That ability is endowed by the molecular structure of aggrecan, which is dominated by negatively charged glycosaminoglycan (GAG) chains. These chains, covalently linked to the aggrecan core protein, are added to the core protein during its biosynthetic journey through the cell. At the start of this journey, the core protein is nascent and unfolded, but at journey's end each of its different motifs has folded properly and some domains have become heavily decorated with GAG chains. The finished aggrecan molecule is exported to the extracellular matrix (ECM) where it aggregates with hyaluronan (HA) and links proteins into unique macromolecular arrays. Acquisition of native aggrecan structure is undoubtedly monitored by intracellular regulatory systems that ensure that (1) the core protein has all its requisite domains; (2) all globular core protein domains fold correctly; and (3) GAG chains achieve their proper composition and length.

Aggrecan, its close relatives, and their multiple motifs

The four core proteins of aggrecan, versican, neurocan, and brevican are closely related [1]. Each has an N-terminal globular (G)1 domain and a C-terminal G3 domain that are joined by intervening nonglobular sequences of lengths characteristic for each core protein. In addition, the aggrecan core protein has a G2 domain. The G1, G2, and G3 domains each contain distinctive motifs, with each motif having its own folded structure (Fig. 1). The G1 domain contains two proteoglycan tandem repeats (PTR) and one Ig motif [2]. The folded PTRs in G1 define a binding site for HA [3]; this site engages HA after mature proteoglycans are secreted into the ECM. G3 motifs bear homologies to epidermal growth factor (EGF), C-type lectin, and sushi or complement reactive protein (CRP); each has a unique conformation [4,5]. The lectin and sushi motifs are always present whereas the presence of the EGF motif is variable.

How does the agrecan core protein fold up?

Each of the globular motifs must achieve its final conformation by folding from a nascent sequence in the core protein by a process that occurs quickly enough to accommodate the rate of translation and to support continuous protein flow through the secretory pathway. Folding includes the formation of transient intermediates [6]. Inevitably, completely folded globular motifs have a hydrophobic interior and a hydrophilic exterior [7]. Because spontaneous folding rates within the cell are too slow to achieve a timely proper folding, nascent chain folding is assisted by molecular chaperones that are found in virtually every cellular compartment [8]. Molecular chaperones do not catalyze folding but protect nascent chain hydrophobic groups when they are on the surface of partially folded intermediates, thereby hindering potential aggregation between the intermediates.

Because the aggrecan core protein has three globular domains, the folding behavior of individual motifs

Offprint requests to: M.L. Tanzer
Received: July 2, 1999

within and between its domains must somehow be orchestrated and coordinated. Presumably, mechanisms exist to determine when the motifs are properly folded and to allow the protein to progress through cellular compartments. Intracellular quality control systems, involving molecular chaperones, provide the necessary surveillance for monitoring nascent chain folding. Molecular chaperones also act as "gatekeepers," preventing incompletely or imperfectly folded nascent proteins from advancing to distal compartments, which is particularly important for movement from the endoplasmic reticulum (ER) to the Golgi complex [9]. Rejected proteins may be recycled for additional attempts to fold, or they may be diverted from the ER to cytosolic proteasomes for destruction [10,11].

What is the intracellular itinerary of aggrecan core protein and what events highlight its intracellular passage?

Aggrecan core protein mRNA is translated on polyribosomes that become associated with the cytoplasmic surface of the endoplasmic reticulum (ER) membrane. The nascent core protein has an N-terminal signal sequence that directs it to enter a translocon (Fig. 2), which is a unique protein complex that provides a dynamic channel through the ER membrane, allowing a temporary connection between the cytosol and the ER lumen [12,13].

Although the core protein N-terminal signal sequence targets it to the translocon, targeting may be regulated in various ways. Recent studies indicate that there are specialized N-terminal signal sequences that modulate nascent chain entry into the ER lumen and that downstream sequences may also affect the process [14–16]. Although the mechanism underlying this observation has not been completely deciphered, it may involve reciprocal specialized signal recognition particles (SRPs) and specialized SRP receptors (SRP docks) on the ER membrane cytosolic surface [12]. SRPs and their receptors are critical elements in the targeting process (Fig. 2). Furthermore, targeting to the translocon may be delayed until the nascent chain is completed, leading to the physiological process of posttranslational translocation seen in bacteria and yeast [12] and also exemplified in mammalian cells by the posttranslational translocation into mitochondria. In certain instances, as yet undefined, nascent chain targeting to the translocon may not occur, with the nascent protein being released into the cytosol (see Fig. 2). The translocon cytosolic portal becomes sealed once ribosomal–translocon union occurs, while its ER lumenal portal is thought to be guarded by the molecular chaperone BiP, whose mechanical ability to act as a trapdoor is driven by ATP hydrolysis [17,18].

After successful targeting, nascent core protein becomes transiently anchored, via its signal sequence, in the translocon. After the entire nascent chain passes through the translocon and enters the ER lumen, the N-terminal signal sequence is often but not necessarily cleaved. Cleavage occurs within the translocon channel, catalyzed by a specific protease complex [19]. The enzymatic transfer of N-linked oligosaccharides from a lipid-linked complex in the ER membrane to consensus sites in the nascent chain [20] is coordinated with these events; the modified chain is then released into the ER lumen.

While in the ER lumen, the nascent chain's globular motifs must properly fold, protected by chaperones, and must pass through quality control systems that usually contain the same chaperones. For example, the ER intralumenal chaperones calnexin and calreticulin, which assist nascent glycoprotein folding, must recognize specific immature N-linked oligosaccharides on the nascent chain as essential features for quality control surveillance [21]. Aggrecan core protein contains N-linked carbohydrates, and thus interactions between aggrecan core protein and calreticulin/calnexin are anticipated. Other candidate chaperones in the ER lumen that might also interact with aggrecan core protein [22] include BiP (see Fig. 2) and proline disulfide isomerase, a chaperone that facilitates disulfide bond formation; aggrecan core protein folded globular domains contain abundant disulfide bonds.

Once a folded glycoprotein passes surveillance, its oligosaccharides are enzymatically trimmed to stubs within the ER lumen [23]. Subsequently, within the Golgi stacks further trimming occurs and specific monosaccharides are sequentially added to the stubs, creating new oligosaccharide structures. GAG chains are added to the same core proteins as the molecules are moved through the secretory pathway. Xylosyl transferase recognizes GAG chain consensus sites on the core protein, and GAG chain growth is initiated by the addition of xylose to serine residues in the late ER [24–26]. Then, core protein is transported to the *cis*-Golgi region where other linkage monosaccharides are added. Transport through the medial and *trans*-Golgi elements follows, coupled with the completion of the oligosaccharide linkage region and the subsequent elongation of GAG chains by the addition of repeating disaccharides. The sulfation of GAG chains completes the complex series of events involved in producing a mature aggrecan molecule [27,28].

Fig. 1. Schematic diagram of aggrecan core protein—the different globular motifs of Ig, PTR, EGF-like, C lectin-like, and sushi (CRP)—arranged within the G1, G2, and G3 domains. The folded appearance of each motif has been depicted by MOLMOL [39], using relevant files in the Brookhaven protein database. GAG chains are predominantly added to the core protein in the region from G2 to G3 by cellular biosynthetic processes

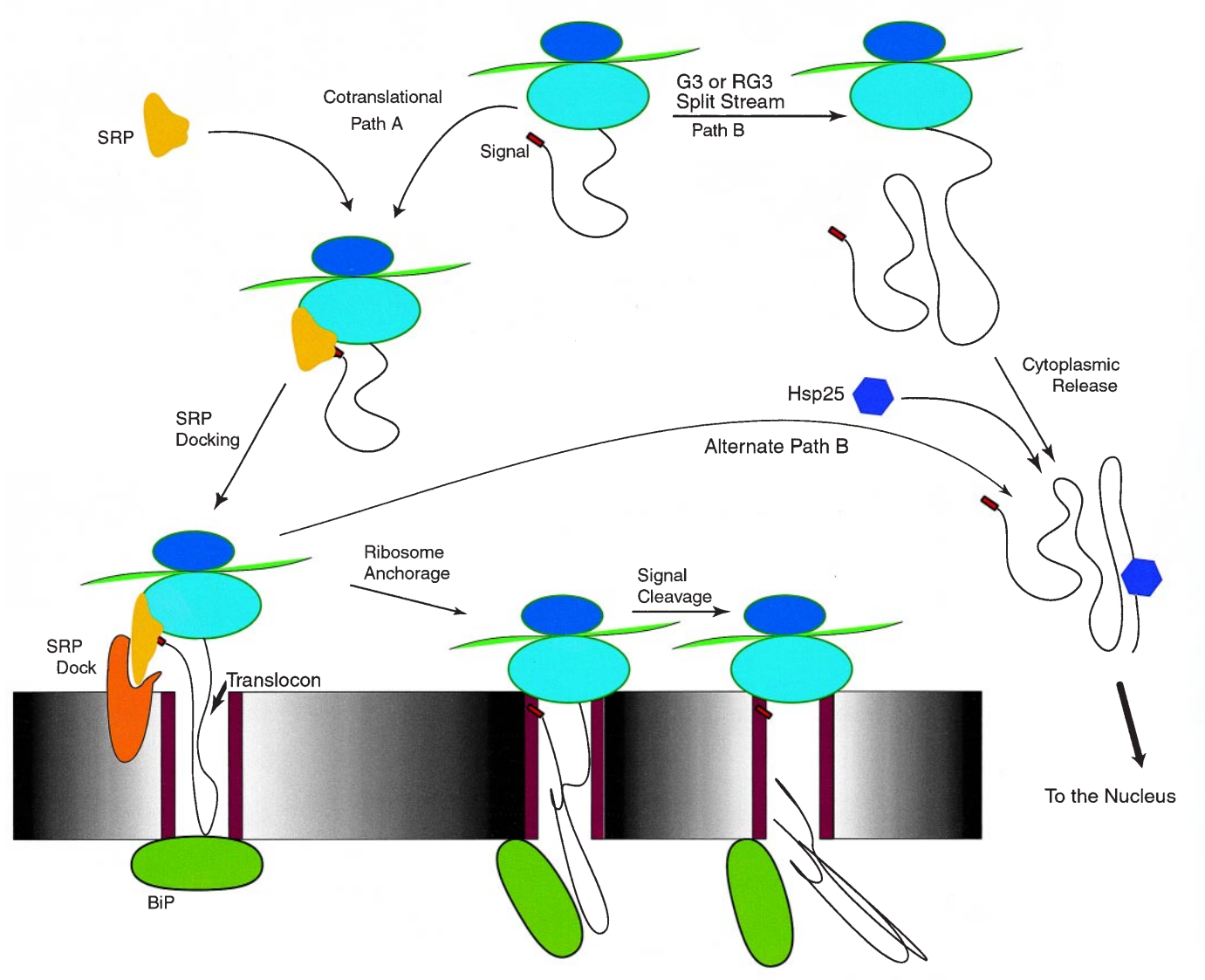

Fig. 2. Schematic illustration of the cotranslational routing of aggrecan core protein and its entry into a translocon of the ER membrane, via signal recognition particle (SRP) recognition and SRP docking, followed by ribosomal anchorage, signal peptide cleavage, and nascent chain release into the ER lumen (*path A*). The signal sequence is depicted with an extended downstream nascent chain, as suggested by Hegde and Lingappa [12]. When G3 domains alone are expressed, intracellular split streaming occurs, with a portion of G3 or RG3 protein entering path A and the remainder entering postulated *path B* or *alternative path B*, with progressive nuclear accumulation of G3 or RG3 proteins. Heat shock protein (Hsp25) is depicted as associating with G3 in path B; Hsp25 also associates with RG3, both in path B (not shown) and conceptually in path A during a posttranslational pause (not shown), as previously suggested [34]

Insights gained from expressing aggrecan globular motifs

In its characteristic progress through the default secretory pathway, nascent aggrecan core protein enters the ER lumen and travels from the ER lumen to the Golgi apparatus, through the various Golgi compartments, and then to the cell surface, where completed aggrecan is released into the ECM. In the lethal chondrodysplasia, nanomelia, a premature stop codon within the chondroitin sulfate (CS) domain underlies production of a truncated core protein that lacks G3 and part of the flanking CS domain [29]; the truncated protein fails to reach the Golgi elements and appears to be targeted for ER-associated degradation [30]. The effects of this particular mutation implicate G3 as being vital for the movement of aggrecan core protein from the ER to the Golgi.

An experimental paradigm

The role of globular domains in aggrecan trafficking was pursued by experiments employing molecular genetic techniques. Individual aggrecan core protein globular domains, G1 and G3, and GAG consensus sequences were arranged in separate genetic constructs to examine their expression and routing behavior in transfected cells. Each construct contained the aggrecan core protein natural N-terminal signal sequence and a C-terminal hexahistidine sequence, sometimes accompanied by a FLAG epitope. Each construct also contained the same GAG consensus sequences, positioned either at the natural flanking sites of G1 or G3, or at their opposite flanks. G3, when expressed as a rearranged protein with flanking GAG consensus sites located at its C-terminus instead of its N-terminus, was termed RG3.

In each construct, the GAG consensus sites served as Golgi reporter groups, indicating that an expressed protein had passed through the exocytic secretory pathway. These various constructs were expressed in suitable host cells, and the behavior of their expressed proteins was examined using biochemical, immunofluorescent, and immunoelectron microscopic techniques [(Chen et al. in manuscript); 31–34].

G3 may facilitate the progression of G1

We found that G3 by itself, or G3 in combination with G1 in a "miniaggrecan" construct, moves through the ER and the Golgi and is secreted as the corresponding neoproteoglycan. In contrast, G1 by itself resides primarily in the ER and is released from the cell 24 to 100 fold more slowly than is G3 [32,33]. Interestingly, miniaggrecan is released from the host cells at a rate intermediate to those of G1 and G3. Surprisingly, the secreted G1 protein lacks GAG chains. This absence of GAG chains may reflect the inaccessibility of GAG consensus sites to requisite enzymes during exocytic pathway trafficking or may reflect an alternate pathway out of the cell. (The notion of an alternative pathway has been suggested for certain proteins such as growth factors and cytokines and is discussed by Rapoport et al. [19]). By expressing core protein in mutant cells that have inactive xylosyl transferase, we showed in related experiments that GAG chain initiation is not necessary for core protein to continue its exocytic journey [32]. Collectively, these results indicate that G3 facilitates the progression of G1 beyond the ER lumen whereas G1 retards the progression of G3 through the cell.

The G3 lectin motif may aid many different proteins

Dissection of G3 into its constituent motifs by individual expression of those motifs showed that only the lectin motif is necessary and sufficient to facilitate G1 passage into the Golgi stacks [34]. Because the lectin motif is common to a superfamily of 131 different cell-surface and secretory proteins and often located at the C-terminus, it may provide a generic signal for intracellular routing. Within the lectin motif, the routing signal is encoded by a portion of the central of three exons; two segments of highly conserved amino acids within that central exon appear to be necessary and sufficient for neoproteoglycan appearance in the cell culture medium [35].

A path not usually taken

Unexpectedly, in certain instances, intracellular G3 follows a curious path: while a portion of G3 journeys through the secretory pathway, the remainder enters and accumulates in the nucleus [(Chen et al. in manuscript); 31]. When expressed, both G3 and RG3 form nuclear inclusions with unique patterns of intranuclear distribution, and the G3 inclusions appear to be irreversibly aggregated. The nuclear deposits are particularly surprising in light of the presence of the natural N-terminal signal peptide sequence in the constructs, which should encode information to direct the proteins to the ER lumen via the SRP–translocon mechanism described earlier.

The simplest explanation for our observations is that some G3 and RG3 nascent protein chains initially fail to enter the ER lumen, yet continue to elongate and are subsequently released into the cytosol (see Fig. 2). Alternatively, entry into the ER lumen may begin, but progress inefficiently, and further entry may be stymied, followed by retrograde translocation from the lumen

into the cytosol. (The efficiency of SRP-dependent docking to the ER membrane and translocation may both be regulated, as discussed earlier.) From there, the proteins relocate to the nucleus by mechanisms yet to be determined. When G3 and RG3 are expressed from constructs that lack the signal sequence, movement to the nucleus is rapid and extensive. On the other hand, when miniaggrecan and several related constructs are expressed, nuclear or cytosolic distributions are never seen, suggesting that G1 and other domains may foil diversion from the ER lumen to the cytosol (and then on to the nucleus).

The nuclear deposits we observed are reminiscent of the nuclear inclusions that characterize neurodegenerative disorders such as Huntington's disease and spinocerebellar ataxia [36]; thus, G3- and RG3-transfected cells may serve as useful models for studying nuclear inclusion diseases. Although our work was initially designed to learn more about aggrecan trafficking through the secretory pathway, our results raise new and fundamental questions about intracellular trafficking at the levels of the ER membrane and the nuclear envelope.

Unusual chaperone interactions

In another surprising outcome, our results showed that G3, when expressed in the absence of G1, could be chemically cross-linked in the host cell to the molecular chaperone Hsp25 [34]. In contrast, neither miniaggrecan nor G1 becomes cross-linked to Hsp25 or any other identifiable molecular chaperones within the cell. Even more surprisingly, RG3 is secreted as a neoproteoglycan along with Hsp25. Neither G3, G1, nor miniaggrecan are secreted in association with Hsp25. Hsp25 is primarily located in the cytoplasm where it is thought to have multiple functions [37]. Cellular heat shock induces the migration of Hsp25 from the cytoplasm to the nucleus, and upon cell recovery from heat shock, Hsp25 returns to the cytoplasm [38]. These phenomena may presage the ability of Hsp25 to move into the nucleus with proteins such as G3 if they have inadvertently been released into the cytosol. This behavior of Hsp25 is precisely what we have found. At normal incubation temperatures, Hsp25 is frequently colocalized in the nucleus with accumulated G3 or RG3, in cells that are expressing those proteins. In control cells at the same temperature that are not expressing G3 or RG3, only cytoplasmic Hsp25 is seen [(Chen et al. in manuscript); 31].

Taken together, our experimental results indicate that Hsp25 associates with unique G3-containing proteins in more than one intracellular compartment. At the same time, the results provide a distinguishing example of specific biological and molecular interactions in vivo for this small heat shock protein.

Conclusions

Overall, the expression of aggrecan domains has provided insights relevant to aggrecan folding, processing, and trafficking. On a more primary level, because the aggrecan domains are generic and their behavior reflects fundamental intracellular processes, the results challenge contemporary paradigms of intracellular trafficking and chaperone interactions.

Where do we go from here?

The following are among the questions that arise from these recent observations. Does aggrecan core protein have a *specialized* signal sequence for translocon targeting? What are the mechanisms by which G1 and G3 influence one another's intracellular trafficking? Might amino acid sequences within the G3 lectin motif provide targeting signals for Golgi entry? If so, what elements in the anterograde pathway to the Golgi recognize such signals? What mechanisms regulate nascent G3 split streaming between the ER and the cytosol? What is the G3 route to the nucleus and what role does its structure play in nuclear targeting? What determines the patterns of G3 protein deposition in the nucleus? What effects do such deposits have upon the cell? Are there functional relationships between Hsp25 and G3 proteins in various intracellular compartments?

Acknowledgments. Original research results referred to in this review have been supported by grants from the National Institutes of Health to MLT and BMV, and from the Arthritis Foundation to MLT and BMV.

References

1. Schwartz NB, Pirok EW III, Mensch JR Jr, Domowicz MS (1999) Domain organization, genomic structure, evolution, and regulation of expression of the aggrecan gene family. Prog Nucleic Acid Res Mol Biol 62:177–225
2. Perkins SJ, Nealis AS, Dudhia J, Hardingham TE (1989) Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein. J Mol Biol 206:737–753
3. Watanabe H, Cheung SC, Itano N, Kimata K, Yamada Y (1997) Identification of hyaluronan-binding domains of aggrecan. J Biol Chem 272:28057–28065
4. Brissett NC, Perkins SJ (1998) Conserved basic residues in the C-type lectin and short complement repeat domains of the G3 region of proteoglycans. Biochem J 329:415–424
5. Perkins SJ, Nealis AS, Dunham DG, Hardingham TE, Muir IH (1991) Molecular modeling of the multidomain structures of the proteoglycan binding region and the link protein of cartilage by neutron and synchrotron X-ray scattering. Biochemistry 30: 10708–10716
6. Creighton TE, Darby NJ, Kemmink J (1996) The roles of partly folded intermediates in protein folding. FASEB J 10:110–118

7. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4:4–6
8. Fink AL (1999) Chaperone-mediated protein folding. Physiol Rev 79:425–429
9. Hammond C, Helenius A (1995) Quality control in the secretory pathway. Curr Opin Cell Biol 7:523–529
10. Bonifacino JS, Weissman AM (1998) Ubiquitin and the control of protein fate in the secretory and endocytic pathways. Annu Rev Cell Dev Biol 14:19–57
11. Rivett AJ (1998) Intracellular distribution of proteasomes. Curr Opin Immunol 10:110–114
12. Hegde RS, Lingappa VR (1999) Regulation of protein biogenesis at the endoplasmic reticulum membrane. Trends Cell Biol 9:132–137
13. Matlack KE, Mothes W, Rapoport TA (1998) Protein translocation: tunnel vision. Cell 92:381–390
14. Al-Qahtani A, Teilhet M, Mensa-Wilmot K (1998) Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from *Trypanosoma brucei* and *Leishmania*. Biochem J 331:521–529
15. Martoglio B, Dobberstein B (1998) Signal sequences: more than just greasy peptides. Trends Cell Biol 8:410–415
16. Zheng N, Gierasch LM (1996) Signal sequences: the same yet different. Cell 86:849–852
17. Crowley KS, Liao S, Worrell VE, Reinhart GD, Johnson AE (1994) Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. Cell 78:461–471
18. Hamman BD, Hendershot LM, Johnson AE (1998) BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation. Cell 92:747–758
19. Rapoport TA, Jungnickel B, Kutay U (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. Annu Rev Biochem 65:271–303
20. Burda P, Aebi M (1999) The dolichol pathway of N-linked glycosylation. Biochim Biophys Acta 1426:239–257
21. Trombetta ES, Helenius A (1998) Lectins as chaperones in glycoprotein folding. Curr Opin Struct Biol 8:587–592
22. Herrmann JM, Malkus P, Schekman R (1999) Out of the ER—outfitters, escorts and guides. Trends Cell Biol 9:5–7
23. Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem 54:631–634
24. Kearns AE, Vertel BM, Schwartz NB (1993) Topography of glycosylation and UDP-xylose production. J Biol Chem 268:11097–11104
25. Vertel BM, Walters LM, Flay N, Kearns AE, Schwartz NB (1993) Xylosylation is an endoplasmic reticulum to Golgi event. J Biol Chem 268:11105–11112
26. Schwartz NB (1995) Xylosylation: the first step in synthesis of proteoglcyan. Trends Glycosci Glycotechnol 7:429–445
27. Silbert JE, Sugumaran G (1995) Intracellular membranes in the synthesis, transport, and metabolism of proteoglycans. Biochim Biophys Acta 1241:371–384
28. Vertel BM (1995) The ins and outs of aggrecan. Trends Cell Biol 5:458–464
29. Li H, Schwartz NB, Vertel BM (1993) cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia. J Biol Chem 268:23504–23511
30. Vertel BM, Walters LM, Grier B, Maine N, Goetinck PF (1993) Nanomelic chondrocytes synthesize, but fail to translocate, a truncated aggrecan precursor. J Cell Sci 104:939–948
31. Chen TL, Wang P, Gwon S, Luo W, Zheng J, Guo C, Tanzer ML, Vertel BM (1998) Nuclear misdirection of aggrecan domains expected to traffic through the exocytic pathway. Mol Biol Cell 9 (supp):9a
32. Luo W, Kuwada TS, Chandrasekaran L, Zheng J, Tanzer ML (1996) Divergent secretory behavior of the opposite ends of aggrecan. J Biol Chem 271:16447–16450
33. Wang P, Chen TL, Luo W, Zheng J, Qian R, Tanzer ML, Colley KJ, Vertel BM (1999). Immunolocalization of 6xhis-tagged proteins in CHO cells with QIAexpress anti-his antibodies. Qiagen News 1 (1999):3–6 (www.qiagen.com/literature/qiagennews)
34. Zheng J, Luo W, Tanzer ML (1998) Aggrecan synthesis and secretion. A paradigm for molecular and cellular coordination of multiglobular protein folding and intracellular trafficking. J Biol Chem 273:12999–13006
35. Luo W, Guo C, Tanzer ML (1998) Aggrecan trafficking and secretion is facilitated by hydrophobic sequences from domain G3. Mol Biol Cell 9 (suppl):54a
36. Lunkes A, Mandel JL (1997) Polyglutamines, nuclear inclusions and neurodegeneration. Nat Med 3:1201–1202
37. Welsh MJ, Gaestel M (1998) Small heat-shock protein family: function in health and disease. Ann N Y Acad Sci 851:28–35
38. Arrigo AP, Suhan JP, Welch WJ (1988) Dynamic changes in the structure and intracellular locale of the mammalian low-molecular-weight heat shock protein. Mol Cell Biol 8:5059–5071
39. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55, 29–32