Matrix Biology 34 (2014) 64–79

Contents lists available at ScienceDirect

Matrix Biology

journal homepage: www.elsevier.com/locate/matbio

# Border patrol: Insights into the unique role of perlecan/heparan sulfate proteoglycan 2 at cell and tissue borders

Mary C. Farach-Carson *, Curtis R. Warren, Daniel A. Harrington, Daniel D. Carson

*Department of Biochemistry and Cell Biology, Rice University W100 George R. Brown Hall P.O. Box 1892, MS-140, Houston, TX 77251–1892, United States*

ARTICLE INFO

*Article history:*
Received 8 July 2013
Received in revised form 16 August 2013
Accepted 17 August 2013
Available online 31 August 2013

*Keywords:*
Perlecan
HSPG2
Basement membrane
Basal lamina
Heparan sulfate
Proteoglycan
Tissue structure
Tissue borders

ABSTRACT

The extracellular matrix proteoglycan (ECM) perlecan, also known as heparan sulfate proteoglycan 2 or HSPG2, is one of the largest (>200 nm) and oldest (>550 M years) extracellular matrix molecules. In vertebrates, perlecan's five-domain structure contains numerous independently folding modules with sequence similarities to other ECM proteins, all connected like cars into one long, diverse complex train following a unique N-terminal domain I decorated with three long glycosaminoglycan chains, and an additional glycosaminoglycan attachment site in the C-terminal domain V. In lower invertebrates, perlecan is not typically a proteoglycan, possessing the majority of the core protein modules, but lacking domain I where the attachment sites for glycosaminoglycan chains are located. This suggests that uniting the heparan sulfate binding growth factor functions of domain I and the core protein functions of the rest of the molecule in domains II–V occurred later in evolution for a new functional purpose. In this review, we surveyed several decades of pertinent literature to ask a fundamental question: Why did nature design this protein uniquely as an extraordinarily long multifunctional proteoglycan with a single promoter regulating expression, rather than separating these functions into individual proteins that could be independently regulated? We arrived at the conclusion that the concentration of perlecan at functional borders separating tissues and tissue layers is an ancient key function of the core protein. The addition of the heparan sulfate chains in domain I likely occurred as an additional means of binding the core protein to other ECM proteins in territorial matrices and basement membranes, and as a means to reserve growth factors in an on-site depot to assist with rapid repair of those borders when compromised, such as would occur during wounding. We propose a function for perlecan that extends its role from that of an extracellular scaffold, as we previously suggested, to that of a critical agent for establishing and patrolling tissue borders in complex tissues in metazoans. We also propose that understanding these unique functions of the individual portions of the perlecan molecule can provide new insights and tools for engineering of complex multi-layered tissues including providing the necessary cues for establishing neotissue borders.



## 1. Introduction

Perlecan, or heparan sulfate proteoglycan 2 (HSPG2), is an exceptionally large, secreted extracellular matrix (ECM) proteoglycan with a modular structure that supports its many functions in complex tissues. Human perlecan consists of 4391 amino acids including a twenty one amino acid long signal sequence that is absent in the mature protein. The protein core of perlecan consists of a series of individual folding motifs or "modules" that include one SEA domain, three each of laminin G-like domains and laminin B domains, four each of LDL-receptor class A domains and EGF-like domains, eleven laminin EGF-like domains including two that are split (1st and 5th), and one that is incomplete (4th), and twenty two Ig-like C2 type immunoglobulin-like domains, all but one of which are in domain IV. The structural features of these modules and their organization into five domains have been well reviewed previously and the interested reader is referred to any of these excellent articles for more information (Iozzo et al., 1994; Gomes et al., 2002; Hassell et al., 2002; Melrose et al., 2002; Iozzo, 2005; Knox and Whitelock, 2006; Farach-Carson and Carson, 2007; Iozzo et al., 2009; Ida-Yonemochi et al., 2011; Ishijima et al., 2012; Kaneko et al., 2013).

One fascinating aspect of perlecan in vertebrates is its almost invariable existence as a long modular multi-functional protein. While a few variants of the human protein have been reported (i.e. miniperl GenBank AAL79552.1), vertebrate perlecan almost always is produced as a single chain core protein with the biological variability largely attributable to the nature and extent of its decoration with glycosaminoglycan chains, specifically some combination of heparan sulfate (HS) and chondroitin sulfate, added to the core protein at specific sites in modular domains I (see annotation in UniProt P98160) and sometimes V (Tapanadechopone et al., 1999). A single promoter drives and regulates perlecan expression. Then, like train cars linked behind one engine, the expression of the forty eight independently folding modules of human perlecan follows an encoded signal peptide locomotive that

* Corresponding author. Tel.: +1 713 348 5052; fax: +1 302 831 2281.
*E-mail address:* farachca@rice.edu (M.C. Farach-Carson).

0945-053X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.matbio.2013.08.004

directs the entire sequence into the secretory path for carbohydrate decoration. Beginning with the SEA module as the coal car and with the final laminin G-like 3 (LG3) module as the caboose, the diverse modules of perlecan are hitched together intracellularly by their assembly as a single polypeptide chain (Fig. 1). After modification during intracellular trafficking through the secretory route, the whole assembly then is deposited into the extracellular space where the individual modules can interact with a variety of other ECM proteins, heparan sulfate binding proteins, and other protein modifiers (Farach-Carson and Carson, 2007; Whitelock et al., 2008). While some cells, such as chondrocytes and endothelial cells have the capacity, chaperones, and sugar nucleotide pools to make large amounts of fully decorated perlecan, the size and complexity of the protein prevent it from being produced easily as a full length recombinant protein. Most laboratories that study perlecan thus rely on purification protocols using large volumes of medium from cultured cells or tissue extracts, and take advantage of its large size and charge to separate it from other ECM proteins (Castillo et al., 1996; Kaji et al., 2000; Whitelock, 2001). Other approaches include the recombinant production of individual perlecan domains or clusters of modules (Chakravarti et al., 1995; Hopf et al., 1999; French et al., 2002; Clarke et al., 2012; Decarlo et al., 2012), or synthesis of perlecan sequence-containing peptides (Farach-Carson et al., 2008). The human *HSPG2* gene is located on the reverse strand of chromosome 1 and covers 115,000 base pairs. There are 97 exons in the transcript, which is 14,327 base pairs in length.

The *HSPG2* gene promoter that drives and modulates perlecan expression in vertebrates has been defined as the 2.7 kilobase (kb) long region lying immediately upstream of the start site of transcription (Iozzo et al., 1997). The proximal promoter region has the features of a CpG island promoter, lacking a TATA box. There are five apparent transcriptional start sites, and a short 5′ UTR consisting of 80 base pairs. Putative *cis*-responsive elements (CREs) were identified previously in the *HSPG2* promoter (Iozzo et al., 1997). To better understand the evolutionary conservation of this promoter, we performed an in silico analysis using the PhastCons (Siepel et al., 2005) function of the UCSC genome browser, along with a series of complementary analyses using publically available programs. This analysis revealed evolutionary conservation in the noncoding region upstream of the *HSPG2* gene amongst human, chimpanzee, rhesus monkey and mouse genomes, with much less conservation between the upstream regions of the *HSPG2* gene in human and zebrafish genomes. The 20 kb upstream region of the *HSPG2* gene includes 78 putative *cis*-responsive elements (CREs) as identified by the presence of known conserved sequences that interact with known transcription factors. Nineteen of these conserved putative CREs reside in the proximal 2.7 kb. Upstream of this conserved region lie approximately 3 kb of non-conserved sequence, followed by the next highly conserved upstream region. This indicates that the majority of functional *cis* elements in the HSPG2 promoter likely are found within 2.7 kb upstream of the transcriptional start site. Although additional putative CREs can be identified in the promoter region, when investigating the human genome alone, we suggest that the nineteen conserved CREs proximal to the transcriptional start site of *HSPG2* are the most likely to function in primary gene regulation to produce the primary perlecan transcript (Fig. 2A).

Despite the complexity of the perlecan gene and the number of exons, the perlecan protein is surprisingly conserved (Fig. 2B). This retention of protein structure throughout millions of years of evolution suggests an accompanying retention of function. In a previous article (Farach-Carson and Carson, 2007), we proposed a function for perlecan as an extracellular scaffold capable of modulating local intracellular signaling. In this review, we will extend that idea and further present a role for perlecan as a key molecule designed by nature to patrol her cell and tissue boundaries, where it can not only regulate local signals, but also help physically separate signals and confine them to individual cell and tissue compartments: hence, a role in “border patrol”.

## 2. Unique modular structure creates unique functions

As described above, the perlecan core protein historically has been described in terms of five functional domains, each of which contains

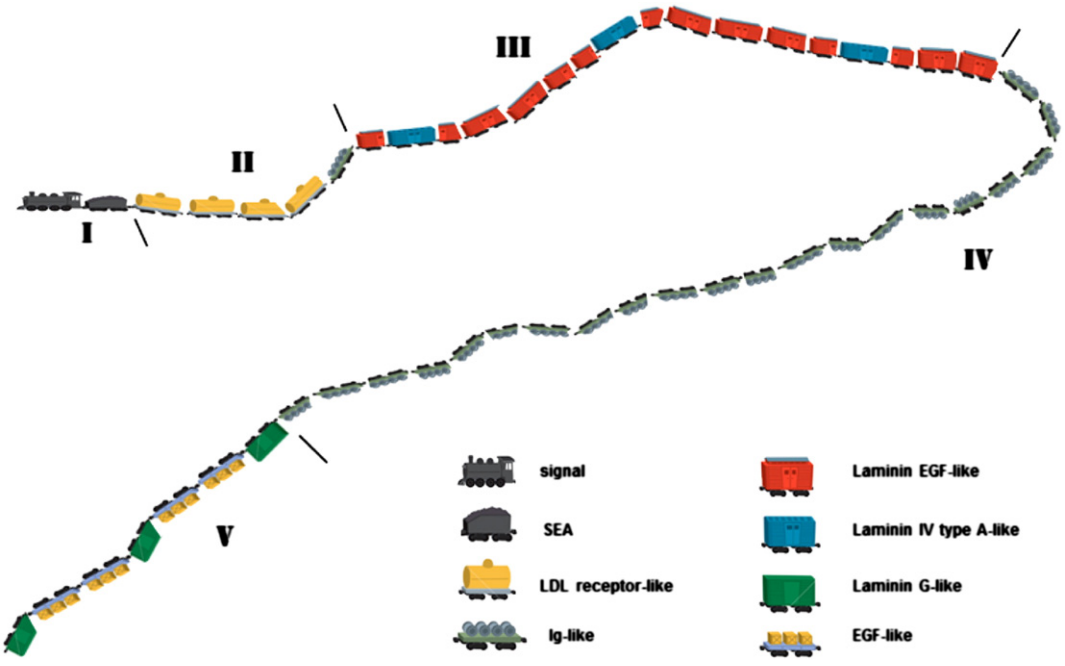


**Fig. 1.** Modular structure of perlecan chain depicted as train cars assembled in tandem behind a common “engine” that is unique to perlecan. The individual modules are self-folding, joined by linker sequences between. The locomotive depicts the signal sequence driving the train into the cellular secretory path for secretion into the extracellular space. The SEA module is envisioned as the coal car, and is near the region where the majority of the glycosylation occurs in domain I (sugars not shown). Features of note include the “split cars” such as the three laminin EGF-like modules split by a laminin IV type A-like module (red split by blue). Note the one Ig-like module at the end of domain II, with the rest in long domain IV. The curvature of the train is suggested by the visualization of the perlecan monomer by atomic force microscopy (see text).