2 Introduction

2.1 The sulfatase family

The sulfatases are a large class of enzymes found in both pro- and eukaryotic organisms which catalyze the hydrolysis of sulfate esters (Hanson *et al*., 2004). Currently 17 sulfatase enzymes have been characterized in humans, ten of which have been described to execute distinct catabolic or regulatory functions in the cell (Table 2.1) (Diez-Roux and Ballabio, 2005). This enzyme class can be divided into two categories based on sub-cellular localization and pH optimum (Diez-Roux and Ballabio, 2005). The acidic pH sulfatases represent the majority of the human sulfatases and are classified by their strict localization to the lysosome where they catabolically degrade sulfated glycolipids and glycosaminoglycans. In contrast, the neutral pH sulfatases represent a small sub-group of sulfatases with alternative localizations in the ER, Golgi or at the cell surface where they are involved in regulating the activity of sulfated substrates such as steroid-sulfates and glycosaminogylcans.

Table 2.1 **The human sulfatases, their sub-cellular localization, substrate specificity and associated genetic disorders**

Sulfatase	Physiological substrate	Localization	Associated genetic disorder
arylsulfatase A	*cerebroside-3-sulfate*	*lysosomal*	*Metachromatic leukodystrophy*
arylsulfatase B	*CS, DS*	*lysosomal*	*Maroteaux-Lamy syndrome (MPS VI)*
arylsulfatase C	**3β-hydroxysteriod-sulfate**	**ER**	**X-linked ichthyosis**
arylsulfatase D	**unknown**	**ER**	
arylsulfatase E	**unknown**	**Golgi**	**Chondrodysplasia punctata**
arylsulfatase F	unknown	ER	
arylsulfatase G	*unknown*	*lysosomal*	*unknown*
arylsulfatase H	unknown	unknown	
arylsulfatase I	unknown	unknown	
arylsulfatase J	unknown	unknown	
arylsulfatase K	unknown	unknown	
galactosamine-6-sulfatase	*CS, KS*	*lysosomal*	*Morquio A syndrome (MPS IVA)*
glucosamine-6-sulfatase	*HS, KS*	*lysosomal*	*Sanfilippo D (MPS IIID)*
sulfamidase	*HS*	*lysosomal*	*Sanfilippo A (MPS IIIA)*
iduronate-2-sulfatase	*HS, DS*	*lysosomal*	*Hunter's disease (MPS II)*
sulfatase 1 (Sulf1)	**HS**	**cell surface**	**unknown**
sulfatase 2 (Sulf2)	**HS**	**cell surface**	**unknown**

acidic pH sulfatases (cursive); neutral pH sulfatases (bold); HS: heparan sulfate; DS: dermatan sulfate; CS: chondroitin sulfate; KS: keratan sulfate. (Table adapted from Diez-Roux and Ballabio, 2005)

The biological importance of the sulfatases is underlined by the presence of at least eight monogenic diseases associated with specific sulfatase deficiencies (Hanson *et al*., 2004). Indeed, due to the catabolic function of many of these enzymes, sulfatase deficiencies often result in lysosomal storage disorders, a disease class characterized by the lysosomal accumulation of undegraded metabolites resulting in severe pathological symptoms and often death (Futerman and van Meer, 2004). As seen in table 2.1, many sulfatases still remain to be characterized or associated with human disease, indicating that much remains to be learned about the functionality of this enzyme class and the involvement of sulfation in mammalian development and disease.

The sulfatases represent a conserved family of enzymes which, despite sharing a common reaction mechanism, show distinct substrate specificities for a number of different sulfo-substrates including steroid sulfates, sulfolipids and glycosaminoglycans (Diez-Roux and Ballabio, 2005). Crystal structure comparison shows a high degree of structural similarity between the sulfatases, especially within the catalytic domain (Sardiello *et al*., 2005). Importantly, all sulfatases contain a unique post-translationally generated amino acid residue within their catalytic domain, C_α-formylglycine (FGly), which is essential for their catalytic activity (Schmidt *et al*., 1995). The importance of FGly for sulfatase activity is evident in the molecular mechanism of sulfate ester hydrolysis (von Bülow *et al*., 2001; Boltes *et al*., 2001). As seen in figure 2.1, the geminal diol of FGly-hydrate mediates a nucleophilic attack on the sulfate ester substrate, followed by a trans-sulfation/elimination reaction to release the covalently bound sulfate group. FGly is regenerated to FGly-hydrate by the addition of water.

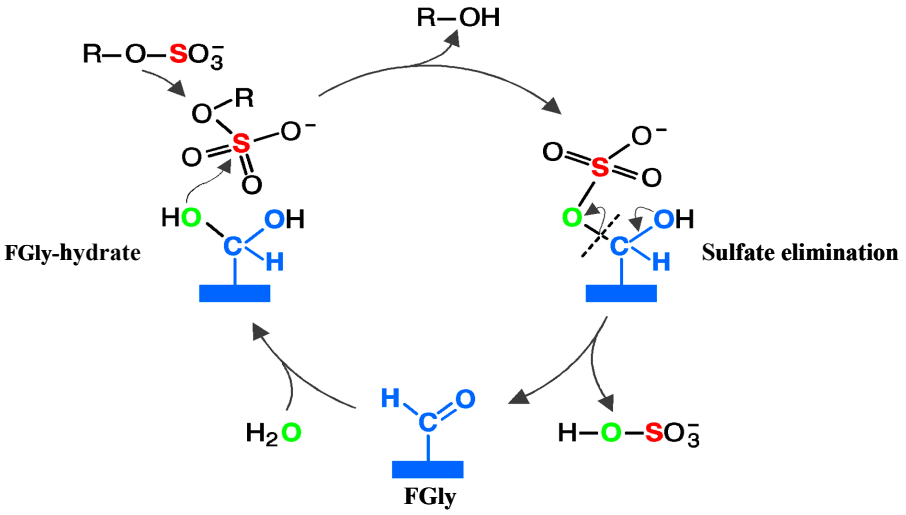

Figure 2.1 **The Sulfatase trans-sulfation/elimination reaction**

The geminal diol of the sulfatase FGly-hydrate carries out a nucleophilic attack on the sulfate ester resulting in the dissociation of the sulfate from its substrate side group and transient covalent association of the sulfate moiety with the enzyme. The sulfate group is then eliminated by the reconstitution of the FGly carbonyl. Finally, FGly is regenerated to FGly-hydrate by the addition of a water molecule.

In mammals, FGly is generated by the formylglycine generating enzyme (FGE) via post translational oxidation of a critical cysteine residue in the sulfatase active site (von

Figura *et al*., 1998; Dierks *et al*., 2003; Dierks *et al*., 2005). *In vitro* analysis has demonstrated that FGly is generated prior to protein folding in the ER and is dependent on recognition of the minimal sulfatase consensus sequence CXPXR by FGE (Dierks *et al*., 1997; Dierks *et al*., 1999). Deficiency of FGE in humans is the molecular basis of multiple sulfatase deficiency (MSD), a rare but fatal disorder in which all 17 sulfatases are inactive due to a lack of FGly (Dierks *et al*., 2003).

2.2 Heparan sulfate proteoglycans

Due to their high sulfate content at multiple sugar positions, glycosaminoglycans (GAGs) compose a major class of substrates for a wide variety of catabolic sulfatases in the lysosome as well as for the regulatory cell surface sulfatases Sulf1 and Sulf2 discussed in this study. GAGs are linear polysaccharides made up of repeating *N*-acetylhexosamine - hexose / hexuronic acid disaccharide units which can be sulfated at different positions and which can extend up to 200 disaccharides in length. Based on differences in sugar composition, GAGs can be classified into four primary families: keratan sulfate (KS), chondroitin sulfate (CS), dermatan sulfate (DS), and heparan sulfate (HS) (Bülow and Hobert, 2006) (Fig. 2.2). At the cell surface and within the extracellular matrix (ECM), these sulfated glycopolymers are found covalently attached to different proteins, forming a functionally diverse family of glycoproteins known as proteoglycans (PGs).

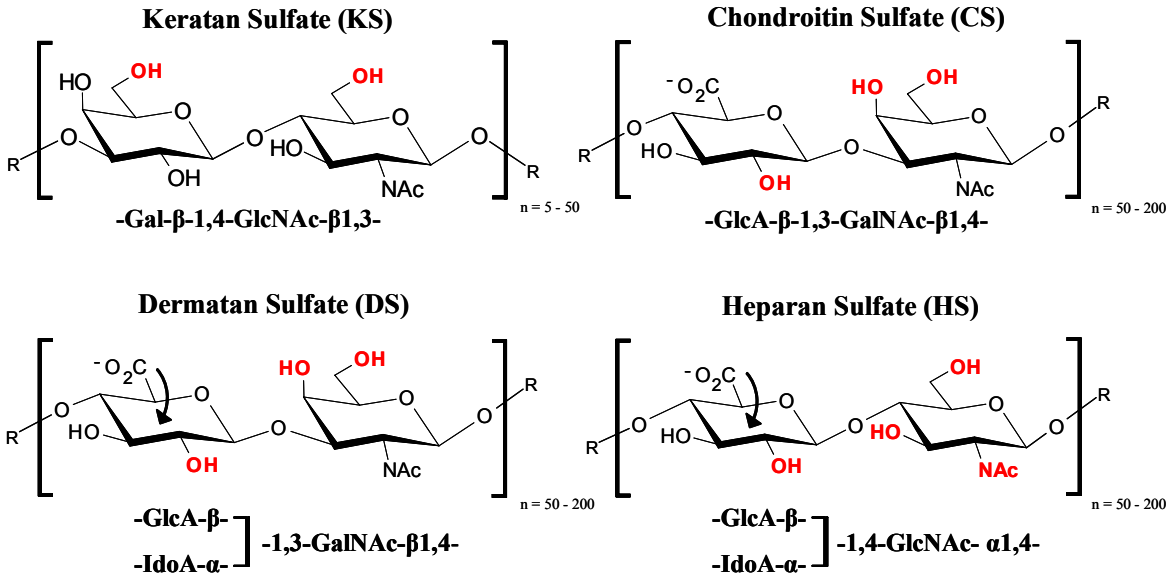

Figure 2.2 **The disaccharide units of glycosaminoglycans**

Chemical structures of the disaccharide repeat units that make up the different GAG chains keratan sulfate (KS), chondroitin sulfate (CS), dermatan sulfate (DS) and heparan sulfate (HS). Red functional groups indicate moieties which can be modified by the addition of sulfate. Arrow indicates that two C5 epimers are possible, generating GlcA or IdoA depending on the orientation.

The ability of HS to interact with a vast array of growth factors and signaling molecules sets this GAG sub-type apart from the other GAG family members and conveys heparan sulfate proteoglycans (HSPGs) with unique functional properties which are critical for regulating mammalian development (Bernfield *et al*., 1999). HSPGs can be divided into three primary categories based on protein back bone type and sub-cellular localization: the transmembrane syndecans (four isoforms), the glycosyl phosphatidyl inositol (GPI) anchored glypicans (six isoforms) and the ECM associated HSPGs agrin, collagen XVIII and perlecan (Bülow and Hobert, 2006) (Fig. 2.3). These HSPG families play distinct roles in regulating cell signaling, cellular adhesion, morphogen gradients, endocytosis and membrane trafficking (Kirkpatrick and Selleck, 2007).

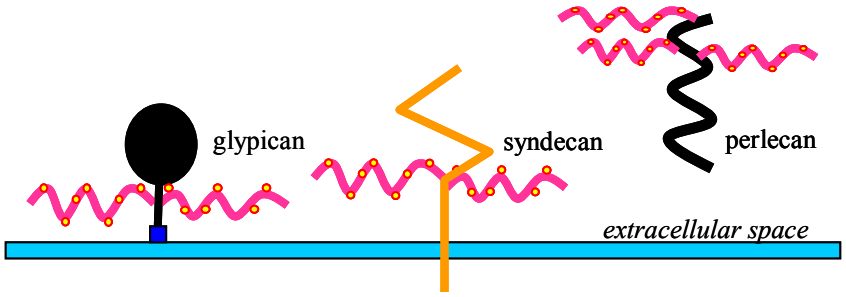

Figure 2.3 **The heparan sulfate proteoglycan family**
Representation of the three primary heparan sulfate proteoglycan family members: glypican, syndecan and perlecan. Glypicans are attached to the cell membrane by a GPI-anchor while syndecans are transmembrane HSPGs. The ECM is composed of multiple HSPGs including collagen XVIII, agrin and perlecan. For simplicity, only perlecan is shown due to its high abundance and near ubiquitous nature.

2.2.1 Heparan sulfate proteoglycan biosynthesis

HSPG biosynthesis is a non-template driven process, relying on multi-enzyme regulation to generate a dynamically sulfated HS polymer attached to a protein core. As displayed in figure 2.4, HS biosynthesis begins in the Golgi with the sequential attachment of four monosaccharides (xylose, two galactose and glucuronate) to a target serine residue of the proteoglycan protein backbone. The addition of the initial *N*-acetylglucosamine (GlcNAc) to this tetrasaccharide linker by the enzyme EXTL3 initiates HS polymer synthesis. Subsequently, alternating glucuronate (GlcA) and GlcNAc residues are added by the co-polymerase enzyme complex EXT1 and EXT2, extending the HS chain up to 200 disaccharide units in length. Once synthesized, the diversity of HS structure is generated by a variety of HS biosynthetic enzymes in the Golgi (Esko and Selleck; 2002). The initial sulfate modification to the nascent HS chain is made by the enzyme *N*-deacetylase/*N*-sulfotransferase (NDST) which can modify up to 50% of the GlcNAc residues to *N*-sulfoglucosamine (GlcNS). Following *N*-sulfation, epimerization by the enzyme *C*-5 epimerase (C5epi) converts a portion of GlcA residues to iduronate (IdoA). Lastly, sulfation reactions at the 2*O*- position of uronic acid (UA) or the 3*O*- and/or 6*O*- position of GlcNAc/GlcNS are carried out by specific

HS sulfotransferases. Importantly, many of the HS biosynthetic enzymes consist of multiple isomers whose variable substrate specificity and expression patterns help modulate the composition and overall organization of the HS sulfation patterns.

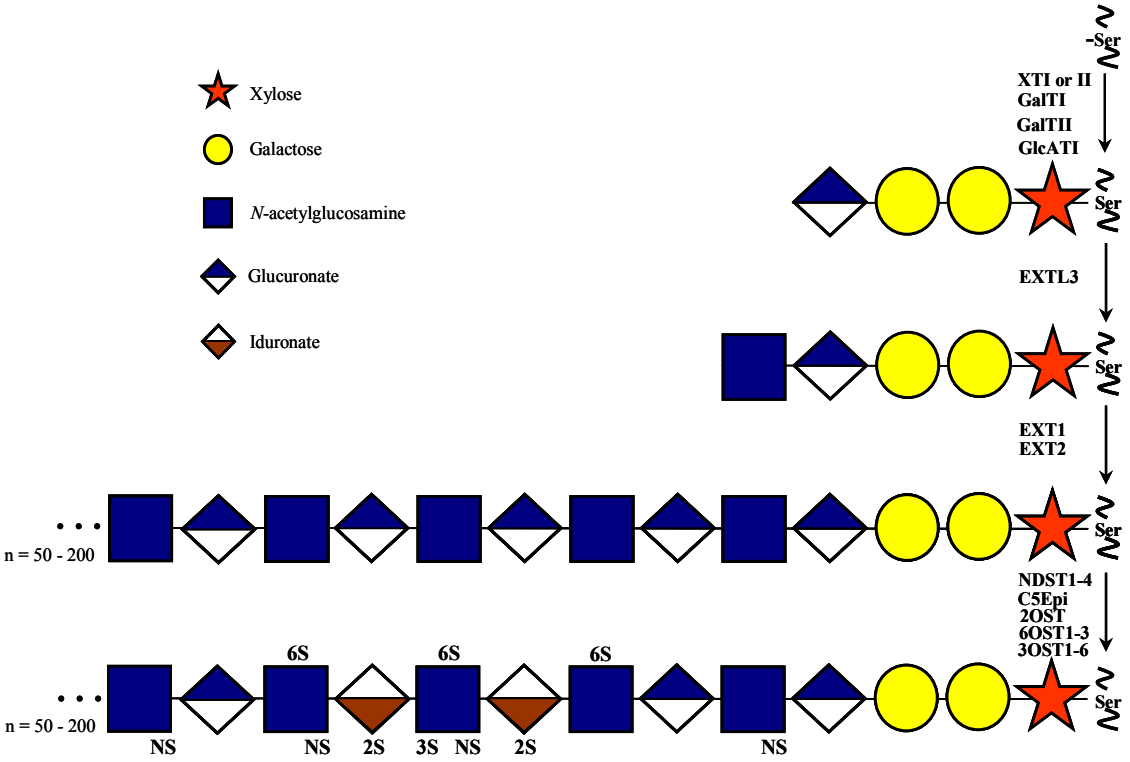

Figure 2.4 **Heparan sulfate biosynthesis**

HS biosynthesis begins with the stepwise addition of a tetrasaccharide linker to a serine residue of the HS proteoglycan backbone by the enzymes xylosyltransferase (XTI or II), galactosyltransferase (GalTI and II) and glucuronosyltransferase (GlcATI). The initial GlcNAc residue is added by the GlcNAc transferase EXTL3, followed by chain extension of alternating GlcA and GlcNAc residues by the EXTI/EXTII co-polymerases. *N*-deacetylase/*N*-sulfotransferase (NDST) enzymes add sulfate to the *N*-position of glucosamine and GlcA is partially epimerized to IdoA by uronosyl C5 epimerase (C5epi). *O*-sulfation is carried out by heparan sulfate 2*O*-sulfotransferase (2OST), 3*O*-sulfotranserase (3OST1-6) and 6*O*-sulfotransferase (6OST1-3) enzymes. The number (n) refers to the number of UA-GlcNAc disaccharide units in a typical HS chain.

As seen in figure 2.4 the cumulative action of the HS biosynthetic enzymes is incomplete, generating highly, partially and non-sulfated regions (Maccarana *et al*., 1996). These differentially sulfated regions create a distinct HS domain structure which is implicit to HS function (Fig. 2.5). Regions of consecutive *N*-sulfation vary from approximately 6 to 24 saccharides in length and are designated as S-domains. These highly sulfated domains are flanked by short (4-6 saccharides) less sulfated regions containing variable *N*-sulfation and are refered to as transition zones. S-domains and their associated transition zones are spaced throughout the chains by non-sulfated regions. Importantly, HS sulfation patterns and their domain organization along the HS chain are cell type and developmental stage specific (Ledin *et al*., 2004), serving as

dynamic templates to promote or inhibit specific cellular interactions and signaling events (Gallagher, 2006; Kreuger *et al*., 2006).

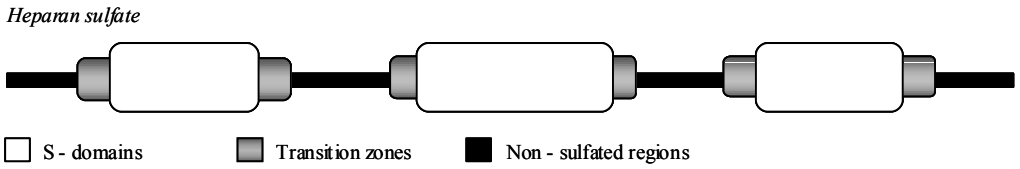

Figure 2.5 **Heparan sulfate domain structure**

Heparan sulfate diversity is generated by the biosynthetic addition of sulfate groups in specific regions throughout the chain resulting in variably sized domains of consecutive *N*-sulfation (S-domains) flanked by short, less sulfated, transition zones. S-domains and their associated transition zones are spaced throughout the chains by non-sulfated regions.

2.2.2 HSPG regulation of cell signaling

A wide array of growth factors, growth factor binding proteins, morphogens and cytokines have been described to require HSPGs for the induction of cellular response as displayed in table 2.2. HSPG modulation of cell signaling is primarily dependent on the HS polymers whose sulfate epitopes serve to present HS binding ligands to their respective receptors in an active conformation.

Table 2.2 **Heparan sulfate dependent signaling and regulatory ligands**

Growth factors	Growth factor binding proteins	Morphogens	Cytokines
FGF 1-15	Follistatin	BMP 2 and 4	CC Chemokines
TGFβ 1 and 2	IGFBP 3 and 5	GDNF	CXC Chemokines
VEGF165 and 189	Noggin	Chordin	IL-2, -3, -4, -5, -7, -8, -12
PDGF		Frizzled-related peptides	GM-CSF
Heparin-binding EGF		Sonic hedgehog	Interferon γ
Neuregulin		Wnts 1-13	TNF α
Betacellulin			Platelet factor 4
Amphiregulin			

BMP: bone morphogenic protein; EGF: epidermal growth factor; FGF: fibroblast growth factor; GDNF: glial cell line-derived neurotrophic factor; GM-CSF: granulocyte-monocyte colony stimulating factor; IGFBP: insulin–like growth factor binding protein; IL: interleukin; PDGF: platelet-derived growth factor; TGF: transforming growth factor; TNF: tumor necrosis factor; VEGF: vascular endothelial growth factor (Table adapted from Bernfield *et al*., 1999)

The molecular mechanism by which HS is able to mediate cell signaling response is best studied for the FGF1 and FGF2 members of the FGF growth factor family, both of which require *N*-, 2*O*- and 6*O*-sulfated HS oligomers to activate their cognate receptors (Yayon *et al*., 1991; Guimond *et al*., 1993). Examination of co-crystal structures of FGF1 growth factor dimers in association with heparin and the FGF-receptor-2

(FGFR2) binding domain reveals a stabilizing function for heparin, mediating FGF1/Heparin/FGFR2 ternary complex formation in a 2:1:2 assembly (Fig. 2.6A) (Pellegrini *et al.*, 2000). Indeed, structural analysis of the crystallized FGF1/Heparin/FGFR2 complex reveals electrostatic interactions between the growth factor dimer and *N*- and 2*O*- sulfate moieties in the heparin oligosaccharide, while 6*O*-sulfate mediates association with the FGF-receptor dimer (Fig. 2.6B and 2.6C).

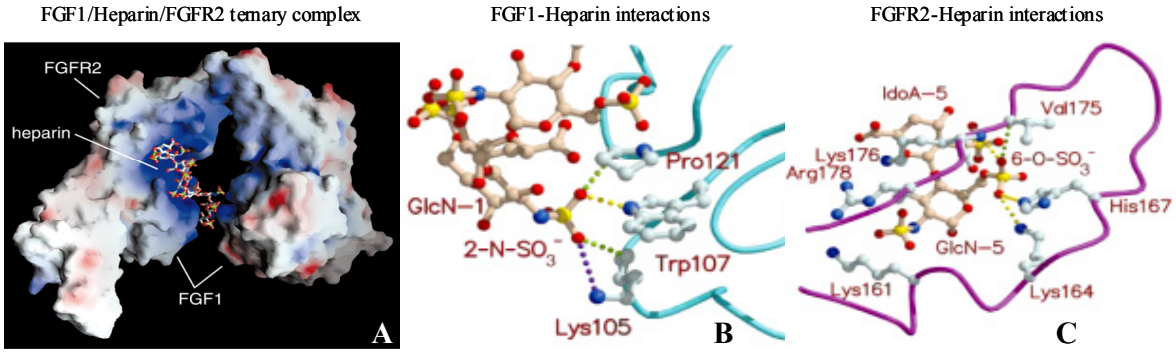

Figure 2.6 **FGF1/Heparin/FGFR2 ternary complex crystal structure**

The co-crystallization of FGF1 and the FGF receptor (FGFR) with heparin reveals a 2:1:2 ternary complex of FGF1:Heparin:FGFR2. (A) Overall architecture of the FGF1/Heparin/FGFR2 complex showing the electrostatic potential on the molecular surface of the growth factor and receptor complex. (B) Molecular analysis of the FGF1-heparin association showing electrostatic interactions of the growth factor with *N*-sulfate residues of the heparin oligosaccharide. (C) Molecular analysis of FGFR2-heparin association showing electrostatic interactions of the growth factor receptor with 6*O*-sulfate residues of the heparin oligosaccharide. Figure adapted from Pellegrini *et al.*, 2000.

Interestingly, while the presence of specific HS sulfate moieties is essential for some growth factors such as FGF, other HS binding ligands have been shown to rely not as much on sulfate composition but rather on HS charge/domain distribution for active HS conjugation. Exemplifying this concept, HS binding growth factors such as interferon-γ, interleukin-8, platelet factor-4 and VEGF-165 have been shown to require large, multi-domain regions on the HS molecule to elicit cell signaling response (Kreuger *et al.*, 2006). These binding regions are referred to as "composite sulfate regions" and usually consist of multiple S-domains and transition zones, spanning extensive areas of the HS polymer (Gallagher, 2006). Further, it is becoming increasingly evident that in addition to the HS chains, the HSPG protein core can also play an important role in modulating cell signaling response. Indeed, HSPG core proteins have been shown to undergo direct protein-protein interactions with certain growth factors such as FGF2, FGF10 and BMP4 independent of the HS chain (Kirkpatrick *et al.*, 2006). Additionally, proteolytic processing of the glypican core protein was shown to act in concert with HS to modulate Wnt signaling (de Cat *et al.*, 2003). These observations indicate that HSPG regulation of cell signaling involves a co-factor function mediated by HS sulfation patterns which, in some cases, can be further modified by the HSPG protein core to fine tune cellular response.