Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins

Ronald E. Hileman,[1] Jonathan R. Fromm,[1] John M. Weiler,[3,4] and Robert J. Linhardt[1,2]*

Summary

Although interactions of proteins with glycosaminoglycans (GAGs), such as heparin and heparan sulphate, are of great biological importance, structural requirements for protein-GAG binding have not been well-characterised. Ionic interactions are important in promoting protein-GAG binding. Polyelectrolyte theory suggests that much of the free energy of binding comes from entropically favourable release of cations from GAG chains. Despite their identical charges, arginine residues bind more tightly to GAGs than lysine residues. The spacing of these residues may determine protein-GAG affinity and specificity. Consensus sequences such as XBBBXXBX, XBBXBX and a critical 20 Å spacing of basic residues are found in some protein sites that bind GAG. A new consensus sequence TXXBXXTBXXXTBB is described, where turns bring basic interacting amino acid residues into proximity. Clearly, protein-GAG interactions play a prominent role in cell-cell interaction and cell growth. Pathogens including virus particles might target GAG-binding sites in envelope proteins leading to infection.
BioEssays 20:156–167, 1998. © 1998 John Wiley & Sons, Inc.

GLYCOSAMINOGLYCAN (GAG) STRUCTURE

Proteoglycans (PGs)

Heparin is synthesized in mast cells and basophils as a proteoglycan (PG) having a molecular weight of approximately 10^6 (Fig. 1). The protein core has multiple glycosaminoglycan (GAG) chains attached to serine residues through a neutral trisaccharide linkage region.[1] When mast cells or basophils degranulate, PG heparin is enzymatically degraded by proteases and β-endoglucuronidase to GAG heparin.[2] Heparan sulphate is also synthesized as a PG but these PGs contain a smaller number of GAG chains and are localised on cell surfaces and in the extracellular matrix.[3,4] The unique extracellular location of heparan sulphate PGs suggests an important role for these molecules in modifying activity of proteins involved in cell-cell interaction.

[1]Division of Medicinal and Natural Products Chemistry, University of Iowa, Iowa City.
[2]Department of Chemical and Biochemical Engineering, University of Iowa, Iowa City.
[3]University of Iowa and Department of Internal Medicine, Iowa City.
[4]VA Medical Center, University of Iowa, Iowa City.
*Correspondence to: Robert J. Linhardt, University of Iowa, PHAR-S328, Iowa City, IA 52242.
E-mail: robert-linhardt@uiowa.edu

Heparin

Heparin, the most studied of the GAGs, is a repeating linear copolymer of 1→ 4 linked uronic acid and glucosamine residues.[1,5] The uronic acid typically consists of 90% iduronic acid and 10% glucuronic acid. Heparin has a high negative charge density, the result of sulphate and carboxylate residues that are present in its structure. Indeed, the

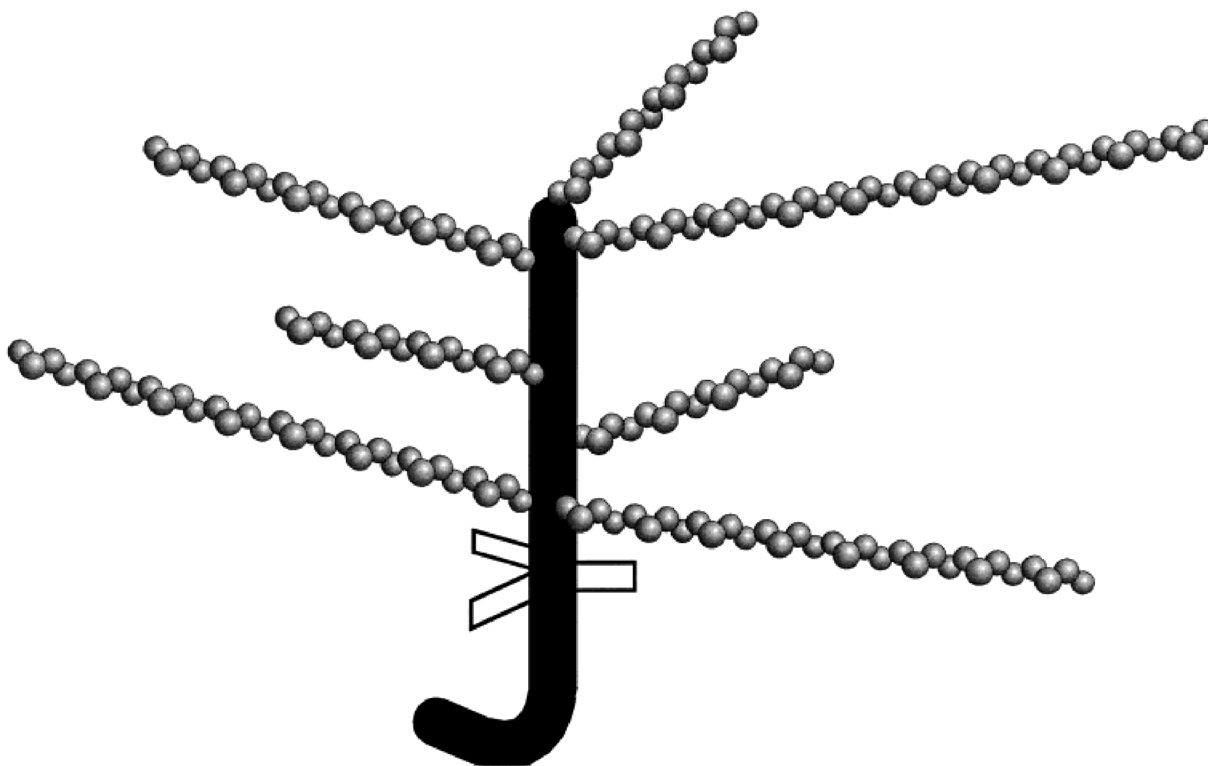

Figure 1. A representation of the structure of a proteoglycan. The structure is often likened to a bristle brush. The protein core is shown as a thick black line, saccharide residues of the GAG chains as spheres and *N*-linked glycans as open boxes.

average disaccharide in heparin contains 2.7 sulphate groups. Whereas the most common structure in heparin is the trisulphated disaccharide[6] (Fig. 2A), a number of structural variants exist making it microheterogeneous.[7] The two-position of uronic acid can be sulphated or unsulphated. Furthermore, the amino functionality of the two-position of the glucosamine residue may be sulphated, acetylated, or unsubstituted. The three- and six-positions of the glucosamine residue can either be sulphated or unsulphated. The presence (or absence) of these functional groups, as well as the presence of two different uronic acid residues, defines the sequence of the GAG chain. The length of heparin GAG chains can also vary. Heparin is polydisperse (can have a variety of chain lengths), the average chain in a commercial heparin has a molecular weight of 13,000 and chains range from 5,000 to 40,000.[8] The combination of heparin's polydispersity and microheterogeneity makes it structurally complex.

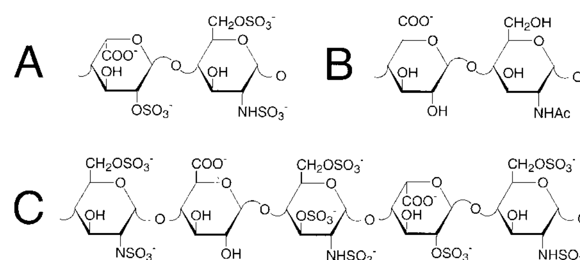

Figure 2. Structure of heparin and heparan sulphate. **A**: Major repeating disaccharide unit in heparin (iduronic acid-2-sulphate → glucosamine-2,6-disulphate; 75–90% of disaccharide sequences). **B**: Major disaccharide repeating unit of heparan sulphate (glucuronic acid → *N*-acetylglucosamine; 10–50% of disaccharide sequences). **C**: The ATIII pentasaccharide binding site sequence (4% of pentasaccharide sequences in heparin and $<$4% in heparan sulphate).

Heparan Sulphate

Heparan sulphate is similar to heparin, but also has features that make it structurally unique. Like heparin, heparan sulphate is a repeating linear copolymer of variably sulphated uronic acid and glucosamine.[1] Heparan sulphate contains only an average of less than 1 sulphate per disaccharide. In addition, heparan sulphate is predominantly composed of glucuronic acid 1→4 linked to glu-

cosamine (Fig. 2B). The disaccharide composition of heparan sulphate suggests that it contains greater structural variation than is present in heparin.[7] Heparan sulphate is primarily composed of glucuronic acid, but can also contain substantial levels of iduronic acid. Although heparan sulphate contains all of the disaccharide sequences found in heparin, the amount of these minor sequences is greater in heparan sulphate making its structure and its overall sequence much more complex. In addition, heparan sulphate has domain structures made up of extended sequences having either low or high levels of sulphation.[3] The highly sulphated domains in heparan sulphate closely resembles heparin and may be involved in its binding to heparin binding proteins. Heparan sulphate chains are also polydisperse and are generally longer than heparin with an average molecular weight of 29,000 ranging[7] from 5,000 to 50,000.

Other GAGs

Chondroitin and dermatan sulphate are composed of *N*-acetylgalactosamine linked through alternating $1 \rightarrow 4$, $1 \rightarrow 3$ linkages to either glucuronic or iduronic acid, respectively. These GAGs can be sulphated at the four- or six-positions of *N*-acetyl galactosamine or the two-position of the uronic acid. Hyaluronan is an unsulphated $1 \rightarrow 4$, $1 \rightarrow 3$ linked copolymer of *N*-acetylglucosamine and glucuronic acid. Keratan sulphate is a sulphated $1 \rightarrow 4$ linked polymer of glucose and *N*-acetylglucosamine. These GAGs also bind proteins[9] but these interactions are generally less well-documented than those of heparin and heparan sulphate.

IMPORTANCE OF GAG-PROTEIN INTERACTIONS

GAGs are generally believed to exert their biological activities through the localisation, stabilisation, activation or inactivation of interacting proteins (Table 1).[10] These interactions play important roles in the normal physiology of animals[11] and are also involved in pathologic processes.[12]

Biology

Heparan sulphate PGs are localised on the external surface of cell membranes and in the extracellular matrix (ECM) and play a major role in cell-cell and cell-protein interaction.[10] These PGs promote adhesion of cells to ECM, through molecules such as fibronectin and laminin, and maintain boundaries between tissues of different types.[13] Heparan sulphate PGs are believed to be endogenous receptors for circulating growth factors and chemokines that regulate cell growth and migration.[10] Heparin is unique in its intracellular location in mast cell and basophil granules and is released as a GAG, together with vasoactive amines upon degranulation. Exogenous heparin is used as an anticoagulant/antithrombotic agent to maintain blood flow in the vasculature through the binding and activation of antithrombin III (ATIII, a coagulation protease inhibitor). In vivo, heparan sulphate PG localised on the endothelium is probably responsible for this activity.[14] It is noteworthy that GAGs are not generally antigenic. This may be due to the similarity of these polyanions to DNA. Indeed, antibodies produced against DNA in systemic lupus also cross-react with heparan sulphate.[15] In conclusion, while the biological roles of PGs are still not completely understood, clearly many of their activities are mediated through their interaction with proteins.

Therapeutic Potential of PGs and GAGs

PGs appear to be important in the control of tumor cell replication, migration, invasion, and vascularisation.[11] In vivo, PGs may also be involved in targeting, attachment and infection by bacteria, protozoa and viruses, of host-cells.[16–20] Heparin and related GAGs are used primarily as anticoagulant and antithrombotic agents to augment activity of ATIII resulting in the inactivation of coagulation proteases and blood anticoagulation. Recently, both a heparin decasaccharide and the polysulphonated heparin analog, suramin, have been reported to inhibit dengue virus infectivity of host cells.[19] GAGs or their analogs may also be administered to: 1) activate protein-based receptors (agonists); 2) inactivate protein-based receptors (antagonists); 3) compete with endogenous GAGs; and 4) inhibit GAG synthesising or metabolising enzymes. Thus, studies are ongoing to examine the potential for GAG-binding proteins, peptides, peptidomimetics, and analogs to prevent and treat a wide variety of disease processes.

INTERACTION THEORIES

Consensus Sequences

The first study to look at the general structural requirements for GAG-protein interaction was undertaken by Cardin and Weintraub and was reported in 1989.[21] They examined heparin binding domains of four proteins, human apo B, apo E, vitronectin, and platelet factor 4, and found sequence similarities. Their report describes defined arrays of basic amino acid clusters (lysine, arginine, and histidine) characterised by two consensus sequences, XBBXBX and XBBBXXBX, "where B is the probability of a basic residue and X is a hydropathic residue." These consensus patterns were shown to occur in other proteins known to bind heparin. Modeling suggested that if the XBBXBX pattern was in a β-strand conformation, the basic amino acids would be aligned on one face of the β-strand and the hydropathic amino acids would point back into the protein core of the protein. Likewise, if the XBBBXXBX pattern is folded in an α-helix, then the basic amino acids would be arrayed on one side of the helix with the hydropathic residues pointing back into the protein core (Fig. 3A).

TABLE 1. Selected Heparin Binding Proteins With Predicted Secondary Structure[53]

Protein	GAG-binding site sequence with predicted secondary structure
Antistatin	^{93}PNGLKRDKLGCEYCECRPKRKLIPRLS^{120} ttt---tttttttt--ttttβββ----
Apo E	^{114}LRKRLLRD^{151} αααααααt ^{211}GERLRARM^{218} αααααααα
LPL	^{281}RKNRCNNLGYEINKVRAKR^{299} t-tttttttt---αααααt
EC-SOD	^{205}REHSERKKRRRESECKAA^{222} ααααααααααααtttt--
vWF	^{565}YIGLKDRKRPSELRRIASQVKYA^{587} αα------ttttββββββββββtt
NCAM	^{131}IWKHKGRDVILKKDVRFI^{148} βββtttt-----------
Fibronectin	^{6}RRARVTDATETTITISWRTKTETITGFQVDAIPANG^{41} ttβββββ-ββββββββββββββββββββββ--tttt ^{1906}YEKPGSPPREVVPRPRPGV^{1924} --tttttttt-tttttttt ^{1946}KNNQKSEPLIGRKKT^{1960} ttt----βββ-----
Laminin	^{641}RYVVLPRPVCFEKGMNYTVR^{660} ββββttttααααtttttβββ ^{202}RIQNLLKITNLRIKFVK^{218} βββββββββββββββββ ^{3010}KQNCLSSRASFRGCVRNLRLSR^{3032} tttt-tttt-ttttβββββ-tt
Vitronectin	^{347}AKKQRFRHRNRKGYR^{361} αααααααtttt----
ATIII	^{124}AKLNCRLYRKANKSSKLVSANR^{145} βααααααααααttttβββtttt
Protein C Inhibitor	^{264}SEKTLRKWLKMFKKRQLELY^{283} αααααααααααααααααααα
Thrombospondin	^{23}RKGSGRRLVK^{30} ααtttt--tt ^{77}RQMKKTR^{83} ααααttt
Acidic FGF	^{110}GLKKNGSCKRGPRTHYGQKAI^{130} α-tttt-tttt---ttttααα
Basic FGF	^{119}LKRTGQYKLGSKTGPGQK^{136} ββtttt---ttttttttα
TGF-β1	^{23}DFRKDLGWKWIHEPKGYHA^{41} β---ttttββββ-tttt--
Apo B100	^{3144}LSVKAQYKKNKHRHSI^{3159} αααααααααtttt--- ^{3352}YKLEGTTRLTRKRGLKLATA^{3371} αααααβββββttttαααααα
PDGF-A	^{194}GRPRESGKKRKRKRLKPT^{211} tt--ttttαααααααα--
Xanthine oxidase	^{781}LGVPANRIVVRVKRM^{795} αα-ttttββββββ-t ^{1106}KKKNPSGSWEDWVTAAY^{1122} t--tttt----αααααβ
Glia Derived Nexin	^{63}RYNVNGVGKVLKKINKAIVSKKNK^{86} ββtttt-αααααααααααααααtt
TFPI	^{212}GKCRPFKYSGCGGNENNFTSKQECLRACKKGF^{243} ---tttttttt--tttt-ttttααααtttttβ
AAMP	^{14}RRLRRMESESES^{25} αααααααααttt
IGFBP-5	^{221}RKGFYKRKQCKPSRGRKR^{238} ttt--------ttttttt
IGFBP-3	^{215}KKGFYKKKQCRPSKGRKR^{232} ttt--------ttttttt
HB-EGF	^{21}KRKKKGKGLGKKRDPCLRKYK^{41} ttt--tttt----tttt---t

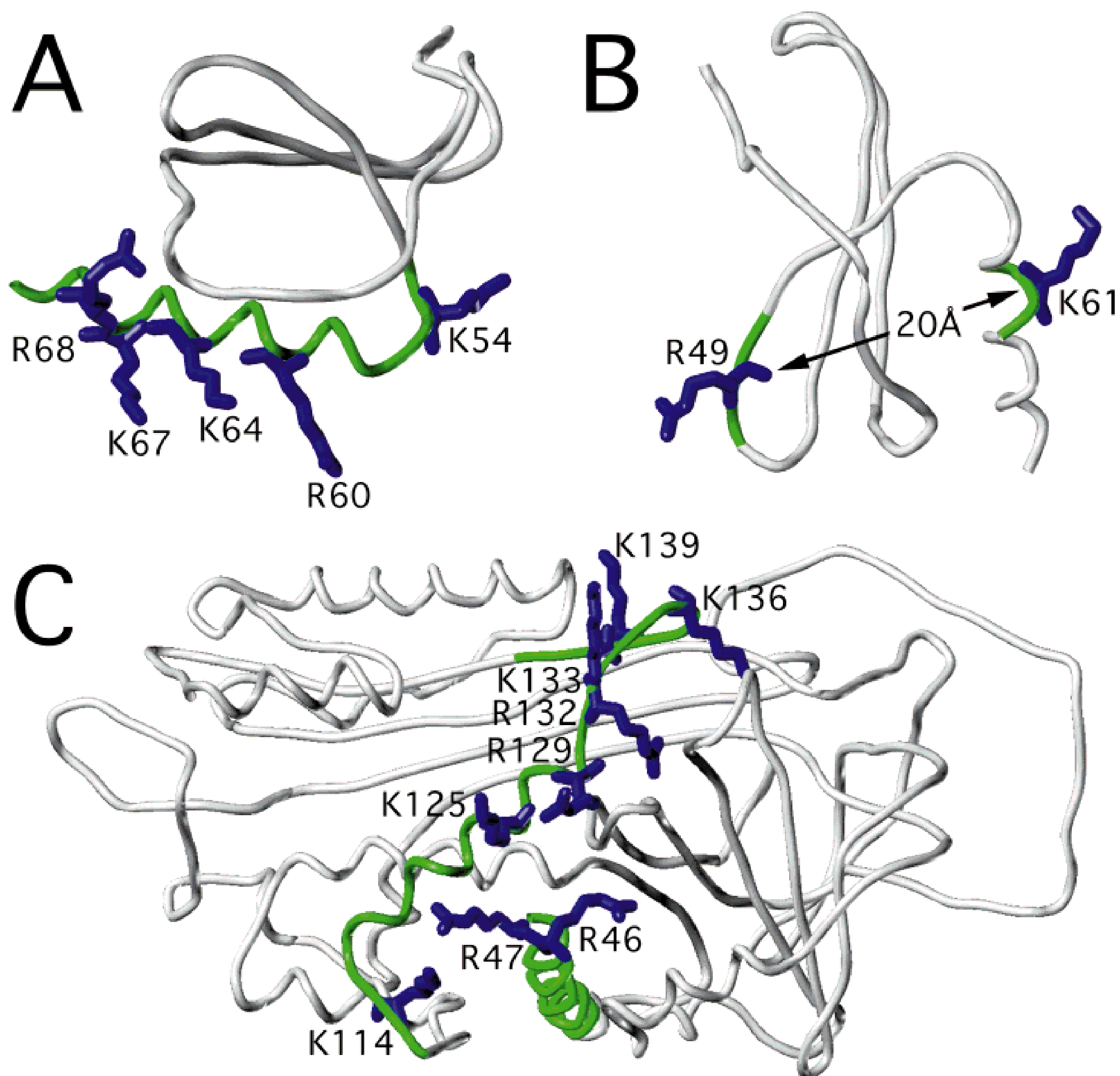

Figure 3. Types of GAG-binding sites present in proteins. **A**: An example of a linear XBBBXXBX motif with basic arginine and lysine residues (blue) oriented on one surface of the helix (green, residues 53–72). The structure is based on interleukin-8 X-ray coordinates (Brookhaven Protein Database 3IL8) **B**: A linear motif, having basic arginine and lysine residues (blue) spaced at a 20 Å linear distance, located on opposite surfaces (green, residues 48–50 and 60–62). The structure is based on a monomer of platelet factor-4 X-ray coordinates (Brookhaven Protein Database HPF4) **C**: An example of a higher order GAG-binding motif. The structure was constructed from antithrombin III X-ray coordinates (Brookhaven Protein Database ATH3 and ANT3) using INSIGHT. The binding domain is comprised of two linearly contiguous domains (green, residues 46–65 and 113–140) brought spatially close through folding of the protein (white) with basic arginine and lysine residues (blue) shown.

Sobel and coworkers[22] followed the approach of Cardin and Weintraub and proposed a third consensus sequence in the heparin binding protein von Willebrand factor: XBBBXX-BBBXXBBX, where "B represents a cationic residue." This consensus sequence may also occur in the heparin binding sites of other proteins.

Spacing of Basic Amino Acids

Margalit and coworkers[23] examined linearly contiguous heparin binding sites that were shown by others to interact with heparin. They proposed that a distance of approximately 20 Å between basic amino acids is important for interaction with heparin, irrespective of whether the heparin binding site

folds into an α-helix or a β-strand (Fig. 3B). In contrast to the work of Cardin and Weintraub,[21] the basic amino acids tended to be aligned on opposite sides of the secondary structural unit. The authors proposed that heparin must coil around the heparin binding domain to interact with the positively charged amino acids on opposite faces of the binding domain. This coiling may induce conformational changes in the heparin binding protein.

SPECIFICITY OF GAG RECOGNITION

Perhaps the most studied group of heparin binding proteins are the serine protease inhibitors (serpins). These inhibitors form covalent complexes with proteases that they inhibit[24] and often their activity is altered by GAGs. This family includes ATIII, the most intensively studied heparin binding protein. When bound to a specific pentasaccharide sequence (Fig. 2C) in heparin,[25–27] ATIII undergoes a conformational change and becomes an activated inhibitor of thrombin and factors IX_a, X_a, XI_a, and XII_a. Not surprisingly, serine proteases acting on ATIII, such as coagulation proteins, thrombin, and Factor XI_a, also bind heparin.[28] In the case of ATIII and thrombin, the GAG chain may act as a template to bring the two interacting proteins together in a ternary complex.[29]

Using site directed mutagenesis, workers have proposed specific sites within proteins that bind GAGs.[30] Similarly, peptides containing GAG binding domains have also been prepared through the chemical and enzymatic fragmentation of heparin-binding proteins[31] as well as by chemical synthesis. A small linear sequence (17 amino acids) in a synthetic peptide, based on ATIII, was shown to specifically recognise ATIII's pentasaccharide binding site (Fig. 2C).[32]

Other GAGs have been studied extensively and specific sequences within them have been proposed as protein-binding sites. Heparin cofactor II (a serpin) binds to disulphated disaccharide sequences found in dermatan sulphate.[33,34] The dermatan and chondroitin sulphate chains of decorin and biglycan may play an important role in regulating hemostasis in the subluminal spaces.[34] Specific highly sulphated sequences in heparan sulphate and heparin have been shown to bind growth factors including fibroblast growth factors (FGF)[35–37] and hepatocyte growth factor.[38] Interestingly, all the specific GAG sequences described to date contain at least one iduronate residue. The conformational flexibility of this residue may play an important role in binding to proteins.[39] The X-ray co-crystal structure of FGF bound to a heparin hexasaccharide shows that a tightly interacting iduronic acid-2-sulphate residue adopts the unusual $^2S_0$ skew boat conformation whereas the other weakly interacting residue adopts the standard $^4C_1$ conformation.[35]

Clearly, proteins and peptides interact specifically with GAGs and with oligosaccharides. It is likely that new therapeutic agents maybe developed to regulate these important interactions. Due to the complex nature of GAG structure, synthetic approaches to produce GAG analogs, are currently limited. The majority of this review is focused on understanding the protein requirements for GAG interaction.

POLYELECTROLYTE EFFECT

Thermodynamics of Protein-Polyelectrolyte Interactions

Heparin resembles DNA; both are linear polymers that carry a high degree of negative charge density and are often referred to as polyelectrolytes.[40] Due the high repulsive energy that multiple negative charge groups incur in polyelectrolytes, cations (such as Na^+) bind to minimise these repulsive forces.[41] This has been demonstrated in NMR studies with DNA and hyaluronic acid.[42–44] The binding of Na^+ to the polyelectrolyte's anionic groups is entropically unfavourable[42] and has an obvious consequence for protein-GAG interactions. On binding, the positively charged amino acid residues (Arg^+ and Lys^+) of a GAG-binding protein must displace the bound Na^+, thus affecting the thermodynamics of the protein-polyelectrolyte interaction.[40,42]

Record and coworkers described the thermodynamics for DNA-protein interactions.[40] The binding of protein to polyelectrolyte was regarded as a simple ion exchange reaction. Rigorous mathematical analysis has demonstrated that a small fraction (f) of anion on the polyelectrolyte (GAG or DNA) is not charge-neutralized by Na^+.[45] Thus, the binding of a protein to GAG can be described by the equilibria

$$\text{Protein} + \text{GAG (m sites)} \rightleftharpoons \text{Protein} - \text{GAG} + m\,(1\text{-}f)\,Na^+$$

where m is the number of Na^+ released on protein-GAG binding. This equation suggests that ligand binding will decrease with increasing NaCl concentration as is commonly observed in both heparin[42,46–48] and DNA[44] binding proteins.

Much of the free energy (ΔG) of the interaction of heparin binding protein with GAG is derived from the entropically favourable release of Na^+ from the GAG. The energetically favourable release of bound Na^+ is called the "polyelectrolyte effect".[40,42] Specifically, the observed dissociation constant, K_d, is related to both ionic and nonionic contributions as follows

$$\log K_d = \log K_{d,nonionic} + m\,(1\text{-}f)\log [Na^+]$$

where $K_{d,nonionic}$ is the dissociation constant in the absence of the polyelectrolyte effect.[46–48] A plot of the log K_d vs. log $[Na^+]$ is linear with a slope of m (1-f). Thus, if (1-f) is known, the number of counterions released on the binding of protein to GAG (m) can be determined. Furthermore, a concentration of 1 M Na^+ abrogates the polyelectrolyte effect because the $\log K_d = \log K_{d,nonionic}$ (that is, m [1-f] log 1 = 0).

Consequently, binding measurements at a 1 M concentration of NaCl can be used to quantify the nonpolyelectrolyte mediated interaction of protein to GAG. The second equation, above, also suggests that the free energy (ΔG) of interaction is made up of a component due to the polyelectrolyte effect (the entropically favourable release of bound Na^+) as well as a component due to hydrogen bonding and hydrophobic interactions.[42,46–48]

Application of the Polyelectrolyte Effect to Binding Specificity

The polyelectrolyte effect on GAG binding to proteins at physiologic salt concentrations is significant.[49] This effect has been quantified for heparin interaction with thrombin,[46] mucus proteinase inhibitor,[47] and basic FGF.[48] Analysis of the interaction of thrombin with heparin at varying salt concentrations suggests that almost all of the free energy of interaction at physiologic ionic strength is due to the polyelectrolyte effect.[47] Thus, this is an entirely ionic-based interaction, without specific hydrophobic or hydrogen bonding interactions. Heparin binding to thrombin most closely resembles the simple interaction of heparin with an anion-exchange resin. A similar analysis of the interaction of mucus proteinase inhibitor with heparin shows that 85% of the binding energy is derived from the polyelectrolyte effect;[47] clearly some specific nonionic interactions play a role in this binding. Thompson and coworkers studied the interaction of basic FGF with heparin[48] as a function of NaCl concentration demonstrating the release of 2 to 3 Na^+ on binding. Only 30% of the free energy of binding was due to the polyelectrolyte effect suggesting a significant level of specificity in this interaction. X-ray crystallography of basic FGF bound to heparin hexasaccharide demonstrates a significant number of nonionic interactions.[35] As these examples suggest, the polyelectrolyte effect is important in understanding the specificity of interaction between heparin and proteins. Determination of the number of ionic interactions may also help to define the size of the GAG-binding domain within the protein.

CHARACTERISATION OF GAG-PROTEIN INTERACTIONS

Studies performed in our laboratories have characterised GAG-protein interaction.[19,35,37,50–57] We have focused on peptides as models for GAG binding proteins and have quantified affinity, using equilibrium dialysis, affinity chromatography, affinity electrophoresis, two-dimensional affinity resolution electrophoresis, fluorescence, NMR and Plasmon Resonance Spectrometry, and isothermal titration calorimetry (ITC).

Amino Acids Common in GAG-Binding Proteins

We assessed the affinity of a randomly synthesised 7-mer peptide library for both heparin and heparan sulphate. Peptides enriched in arginine and lysine, but not histidine, bound with greatest affinity. Peptides with high affinity for heparan sulphate were also enriched in other polar amino acids including serine.[50]

In addition, known heparin binding domains contain amino acids capable of hydrogen bonding. These hydrogen bonding interactions also play an important role in specificity (see Polyelectrolyte Effect). Asparagine residues are common in known heparin binding sites and glutamine residues are enriched in peptides that bind heparan sulphate, perhaps because of their hydrogen bonding amide-containing side chains. Moreover, a glutamine residue may be important for both acidic and basic FGF to interact with heparin (Fig. 4A,B).[35,59] Tyrosine residues, also enriched in known heparin binding sites, may hydrogen bond to hydroxyls of the GAG via their phenolic hydroxyl group. Bae and coworkers[32] suggested the importance of a tyrosine residue in a synthetic ATIII peptide in which an apparently specific, hydrophobic interaction with the *N*-acetyl group in heparin's ATIII pentasaccharide sequence (Fig. 2C) was observed.

Differences Between Arginine and Lysine in GAG Binding

Arginine-containing synthetic peptides bind tightest to heparin, whereas, lysine containing peptides bind less tightly to heparin, and peptides containing both arginine and lysine residues exhibit intermediate affinity for heparin.[51] Blocked arginine binds 2.5-times more tightly to heparin than blocked lysine and releases 2.5-times more heat on binding.[51] These data on blocked amino acids demonstrate that differences in binding can only be due to the ammonium side chain of lysine and the guanidino side chain of arginine. The tighter interaction observed for arginine may be the result of stronger hydrogen bond formation between the guanidino group of arginine and the sulphate of heparin. Alternatively, the softer (more diffuse) guanidinium cation of arginine may interact better with the soft sulphate anions of GAGs.

Spacing of Basic Amino Acids in GAG Binding

A single, isolated basic amino acid is most commonly found in heparin binding sites, followed by clusters of two and three basic amino acids.[52,53] A single nonbasic amino acid between basic amino acid clusters is the most commonly observed spacing pattern in heparin binding sites. Affinity studies demonstrate the importance of the number of basic residues within a cluster. Using ITC, synthetic peptides having the general sequence R_nW (where n = 9, 7, 5, 4, and 3) were titrated with heparin. We observed that affinity for heparin was not significantly increased for peptides having a cluster of greater than four arginines.[53]

We tested the affect of spacing on relative affinity for GAGs using a series of synthetic peptides.[52,53] Peptides

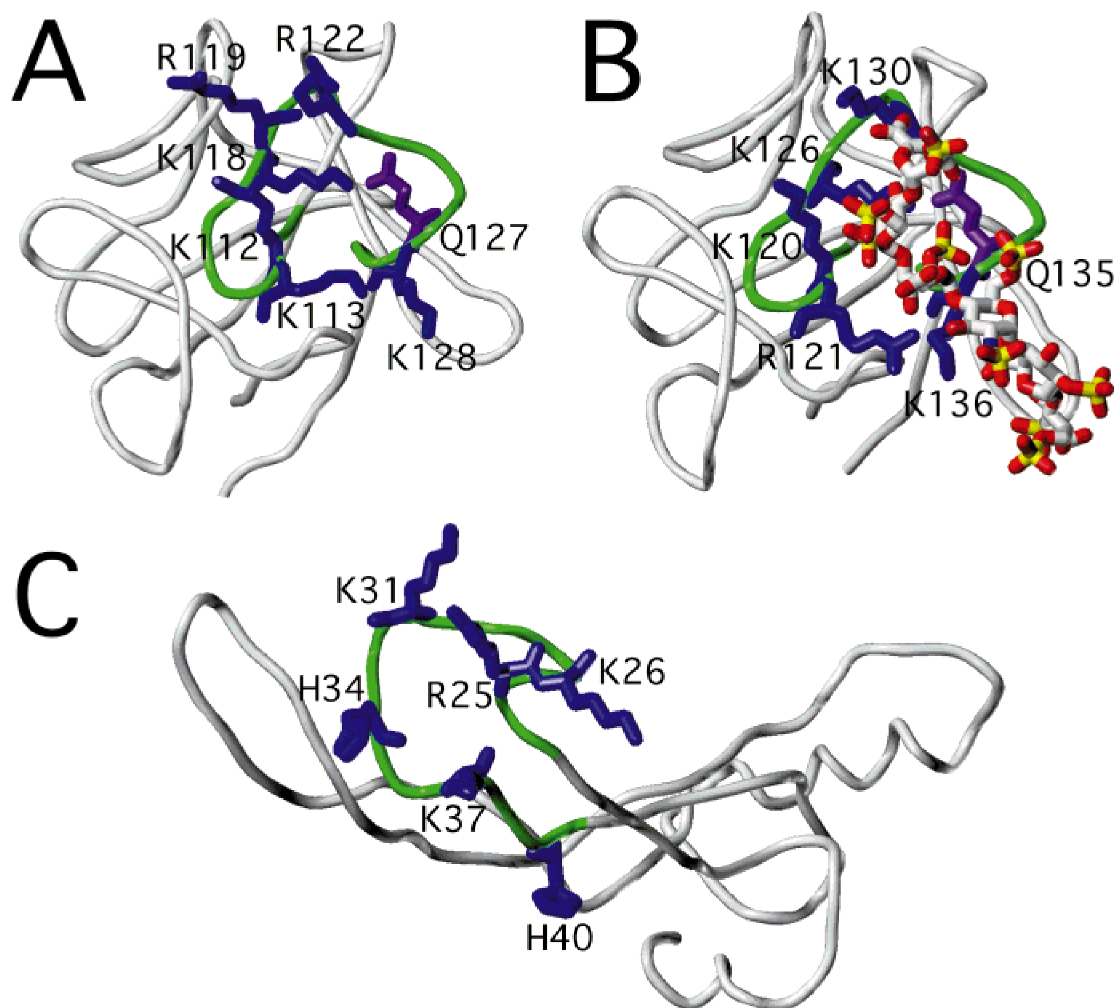

Figure 4. Linearly contiguous GAG-binding domains with the consensus TXXBXXTBXXXTBB. **A**: Acidic FGF (white) with the GAG-binding domain (green, residues 110–130) showing the potentially interacting basic residues (arginine and lysine, blue, and glutamine, purple). The structure is based on coordinates from Brookhaven Protein Database, 2AFG. **B**: Basic FGF (white) bound to a heparin-derived hexasaccharide showing the consensus motif (green, residues 119–136) and basic residues residues (arginine and lysine, blue, and glutamine, purple). The sulphate groups of heparin are shown with red and yellow (oxygen and sulpher atoms, respectively). The structure is based on co-crystal coordinates from Brookhaven Protein Database, 1BFG. **C**: The structure of TGFβ-1 (white) with the consensus sequence (green, residues 23–41) showing the potentially interacting basic residues (blue). The structure is based on coordinates from Brookhaven Protein database, 1KLC.

having the general sequence acetyl-GRRG$_n$RRG$_{5\text{-}n}$-amide showed that peptide with n = 0 and 3 bind most tightly to heparin-agarose. However, for peptides with the general sequence of acetyl-GRRRG$_n$RG$_{5\text{-}n}$-amide, the peptide with n = 0 bind most tightly to heparin-agarose, with the affinity of the peptides falling off rapidly as n approaches 6. These differences suggest that the pattern of the cluster (RRG$_n$RR vs. RRRG$_n$R) is important in defining affinity for GAG. Furthermore, binding sites containing patterns of four basic amino acids in a row (i.e., XBBBBX) or symmetrical clusters of basic amino acids spaced three amino acids apart (i.e., BBXXXBB) are predicted to recognise heparin or highly sulphated regions of heparan sulphate. In contrast, binding sites containing BBX$_n$BB sequences with n = 0 or 1 are predicted to bind most tightly to heparan sulphate. This suggests the importance of basic residue spacing on the GAG selectivity of binding proteins. Documented heparin binding sites most commonly contain patterns with one nonbasic residue between clusters of basic amino acids (that is, n = 1). In addition, the extracellular localisation of heparan sulphate (in

contrast to intracellular heparin) suggests it is the endogenous ligand for "heparin" binding proteins, most of which are extracellular (Table 1). Considering these data it is reasonable to suggest that binding sites containing BBXBB sequences optimally interact with heparan sulphate.

Secondary Structure in GAG-Binding Domains

Topology of the heparin binding site may also be important. We examined the FGF heparin binding site using both experimental data[35,54–56,59] and molecular modeling.[52,54] The X-ray crystal structures of both acidic and basic FGF, with their binding sites occupied and unoccupied,[35,59] clearly demonstrate a well-conserved topology despite significant differences in primary sequence (Fig. 4A,B). The FGF heparin binding site contains a peptide backbone that loops back upon itself; three turns are present in the loop (Fig. 4). The domain can be described as a triangle, with each turn defining an apex.

X-ray and NMR structural data for acidic FGF, basic FGF and transforming growth factor β-1 (TGFβ-1) implicate a motif described as TXXBXXTBXXXTBB where T defines a turn, B a basic amino acid (arginine, lysine or occasionally glutamine), and X a hydropathic residue (Fig. 4). The distance between the first and second turn ranges from approximately 12 Å (TGFβ-1) to 17 Å (acidic and basic FGF), calculated from the C_α position of each residue. The distance between the second and third turns is approximately 14 Å. The distance between the first and third turns ranges from approximately 13 Å (TGFβ-1) to 18 Å (acidic and basic FGF). The structural similarity between these three GAG-binding domains characterizes a new type of GAG-binding domain.

Affinity studies demonstrated the importance of the shape of the binding site in FGF. A synthetic peptide with a sequence analogous to the native heparin binding site bound more tightly, as measured by both heparan sulphate-agarose affinity chromatography and ITC, than a peptide of identical sequence except containing an unnatural D-proline in the second turn.[52,54] A cyclic peptide, designed to resemble the structure of the heparin binding site in FGF, bound more tightly to heparan sulphate than the linear peptide of the same sequence. Furthermore, ITC demonstrated that the cyclic peptide bound to fewer sites in heparan sulphate when compared to the linear peptide, suggesting that it may be more specific in its interaction with GAG.

Prediction of GAG-Binding Domains in Proteins

Cardin and Weintraub[21] have suggested that linear arrays of basic residues segregate on a face of an α-helix or β-strand imparting heparin binding capacity. However, predicted turns[60] are common in these sites, and a number of heparin binding sites are predicted to be made up primarily of amino acids in turns (Table 1). Consequently, these sites should not be linear and should not segregate basic residues on the face of an α-helix or β-strand. We searched databases for sequences that correspond to a 17-residue consensus sequence (containing three turns) from the heparin binding sites of acidic and basic FGF.[52,54] As a control, we randomised the structural elements of this sequence and repeated the search. The FGF consensus sequence matched many more heparin binding sites in heparin binding proteins than the randomised sequence. This suggests that FGF type sites may be commonly found in heparin binding proteins other than FGF.

Higher Order Structures in GAG-Binding Domains

Not all heparin binding proteins contain linearly contiguous heparin binding sites. ATIII is known to contain a linearly contiguous heparin binding domain believed to be responsible for specificity[32] of interaction with heparin's pentasaccharide binding site (Fig. 2C). In addition, linearly remote interacting basic residues are brought spatially close through the folding of ATIII[61] (Fig. 3C). Indeed, it is well known that ATIII undergoes a significant conformational change on binding to the heparin pentasaccharide.[62] Thus, tertiary structure clearly plays a role in heparin binding to ATIII. Cooperative binding of multiple proteins to heparin might also take place suggesting the relevance of the quaternary structure of multi-protein complexes in promoting high affinity, high specificity, and productive interactions.

IMPLICATIONS AND CONCLUSIONS

These studies suggest how specificity for different GAG sequences may be incorporated in heparin binding sites. Asparagine and glutamine residues may align to hydrogen bond with hydroxyls of the GAG. Because the glutamine side chain is one methylene group longer than the asparagine side chain, in a particular heparin binding site asparagine residues may interact with more exposed hydroxyls whereas glutamine residues may "stretch or reach" less exposed hydroxyls in the GAG. The size of the heparin binding site (either the linear dimensions or the spacing of basic residues and number of turns present) may define the length of the GAG chain with which it interacts. Furthermore, heparin binding sites with widely spaced basic amino acid clusters would preferentially recognize GAG sequences in which the sulphate groups were more widely spaced.

Affinity of a heparin binding protein for a GAG sequence will depend on the ratio of arginine to lysine residues in positions that ion pair with anions of the GAG. A higher ratio would promote higher affinity interactions. Similarly, the proper alignment of amino acids capable of hydrogen bonding would also increase affinity of the GAG sequence for its heparin binding site in the protein.

A number of issues concerning specificity and affinity are raised by these studies. Neither heparin nor heparan sulphate contains functional groups that are usually considered hydrophobic. Heparan sulphate, however does have a significantly higher percentage of an *N*-acetyl functionality on the amino group of glucosamine than does heparin (Fig. 2A,B). It is possible that hydrophobic amino acids of proteins form hydrophobic interactions with the *N*-acetyl groups of heparan sulphate promoting specificity of this interaction. Leucine and tyrosine residues occur more commonly in known heparin binding sites than statistically expected, perhaps fulfilling this role.[50] For example, a solvent exposed valine residue is found in the predicted heparin binding site of keratinocyte growth factor.[35] Bae and coworkers[32] demonstrated a close spatial proximity between a hydrophobic tyrosine residue of a peptide and the *N*-acetyl group using nuclear Overhauser effect spectroscopy. Furthermore, it has been reported that some glucosamine residues in heparan sulphate are unsubstituted (having a free amino group).[63,64] Because these amines would be protonated at physiologic pH, proteins that recognize free amino containing GAG sequences would bind weakly if a basic amino acid were also present. Binding sites in proteins that recognize free amino containing GAG sequences may have neutral residues capable of hydrogen bonding, anionic residues (aspartic and glutamic acid) to form ionic interactions, or metals that interact with the lone pair electrons of this free amino group.

Other factors influencing GAG-protein interactions include the on- and off-rate kinetics of binding and enthalpically- vs. entropically-driven binding. Most of the known binding sites are surface exposed shallow sites, suggesting both a high on- and off-rate. Binding entropy in the interaction of polyanions and polycations is directly influenced by the displacement of counterions (for example, Na^+ on the GAGs and Cl^- on the proteins) during binding. The presence of internal ion pairing in place of counterions (i.e., a heparin binding site containing both basic and acidic residues) or a high energy (unstable) binding site in which counterions are excluded might decrease the role of entropy and increase the role of enthalpy in the binding process. For the design of a pharmaceutically active compound (either a GAG or a peptide analog), factors affecting avidity, specificity of interaction, as well as off-rate kinetics of binding will be important.

This study presents a new type of heparin binding site (Fig. 4) having the following characteristics. 1) A peptide backbone that turns back upon itself to form a cup of positive charge to interact with the negatively charged GAG. The majority of interacting amino acids are contained in this contiguous region of the protein with a consensus sequence of TXXBXXTBXXXTBB. 2) Proper spacing and patterns of basic residues in part define affinity and specificity for GAG interaction. 3) Arginine residues promote tighter interaction of these sites for GAG than lysine residues. The correct ratio of these two residues, in part, defines affinity of a site for GAG. 4) Other polar (and perhaps hydrophobic) amino acids distributed throughout the site, in part, define specificity and affinity of these sites for GAG.

The definition of the consensus sites within GAG-binding proteins will increase our understanding of protein-GAG interactions and facilitate the search for potential GAG-binding domains in target proteins. GAG-protein interactions are clearly important in many biological processes. The design of therapeutics based on protein-GAG interaction, to modify cell-cell interaction, viral infection and cell growth represents an important goal for the future.

REFERENCES

1 Lane, D. A. and Lindahl, U. (1989) *Heparin, Chemical and Biological Properties, Clinical Applications.* CRC Press, Inc., Boca Raton, FL.

2 Jacobsson, K. G. and Lindahl, U. (1987) Degradation of heparin proteoglycan in cultured mouse mastocytoma cells. *Biochem J* **246**:409–415.

3 Gallagher, J. T., Turnbull, J. E. and Lyon, M. (1992) Heparan sulphate proteoglycans: Molecular organization of membrane-associated species and an approach to polysaccharide sequence. *Adv Exp Med Biol* **313**:49–57.

4 Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M. T., Spring, J., Gallo, R. L. and Lose, E. J. (1992) Biology of the syndecans [review]. *Ann Rev Cell Biol.* **8**:365–393.

5 Casu, B. (1985) Structure and biological activity of heparin. *Adv Carbohydr Chem Biochem* **43**:51–134.

6 Linhardt, R. J., Ampofo, S. A., Fareed, J., Hoppensteadt, D., Mulliken, J. B. and Folkman, J. (1992) Isolation and characterization of human heparin. *Biochemistry* **31**:12441–12445.

7 Griffin, C. C., Linhardt, R. J., Van Gorp, C. L., Toida, T., Hileman, R. E., Schubert, R. L. and Brown, S. E. (1995) Isolation and characterization of heparan sulfate from crude porcine intestinal mucosal peptidoglycan heparin. *Carbohydr Res* **276**:183–197.

8 Linhardt, R. J., Rice, K. G., Kim, Y. S., Engelken, J. and Weiler, J. M. (1988) Homogeneous, structurally defined heparin oligosaccharides with low anticoagulant activity inhibit the generation of the amplification pathway C3 convertase. *J Biol Chem* **263**:13090–13096.

9 Templeton, D. M. (1992) Proteoglycans in cell regulation [review]. *Crit Rev Clin Lab Sci* **29**:141–184.

10 Linhardt, R. J. and Toida, T. (1997) Heparin analogs—development and applications [review]. In *Carbohydrates as Drugs* (eds. Z. B. Witczak and K. A. Nieforth) Marcel Dekker, New York, Chapter 7, pp. 277–341.

11 Folkman, J. and Shing, Y. (1992) Control of angiogenesis by heparin and other sulfated polysaccharides [review]. *Adv Exp Med Biol* **313**:355–364.

12 Riegel, A. T. and Wellstein, A. (1994) The potential role of the heparin-binding growth factor pleiotrophin in breast cancer [review]. *Breast Ca Res Treatment* **31**:309–314.

13 Newman, S. A., Frenz, D. A., Hasegawa, E. and Akiyama, S. K. (1987) Matrix-derived translocation: Dependence on interaction of amino-terminal domain of fibronectin with heparin-like surface components of cells or particles. *Proc Natl Acad Sci USA* **84**:4791–4795.

14 Harpel, P. C., Zhang, X. and Borth, W. (1996) Homocysteine and hemostasis: pathogenic mechanisms predisposing to thrombosis [review]. *J Nutrition* **126**(4 Suppl)1285S–1289S.

15 Suenaga, R. and Abdou, N. I. (1995) Heparan sulphate-ELISA gives false positive results for anti-DNA-DNA/histone immune complexes in sera of patients with SLE. *Lupus* **4**:57–62.

16 van Putten, J. P. and Paul, S. M. (1995) Binding of syndecan-like cell surface proteoglycan receptors is required for Neisseria gonorrhoeae entry into human mucosal cells. *EMBO J* **14**:2144–2154.

17 Kardami, E., Pearson, T. W., Beecroft, R. P. and Fandrich, R. R. (1992) Identification of basic fibroblast growth factor-like proteins in African trypanosomes and Leishmania. *Mol Biochem Parasit* **51**:171–181.

18 Flynn, S. J. and Ryan, P. (1996) The receptor-binding domain of pseudorabies virus glycoprotein gC is composed of multiple discrete units that are functionally redundant. *J Virology* **70**:1355–1364.

19 Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J. and Marks, R. M. (1997) Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nature Med* **3**:866–871.

20 Sawitzky, D. (1996) Protein-glycosaminoglycan interactions: infectiological aspects [review]. *Med Microbiol Immunol* **184**:155–161.

21 Cardin, A. D. and Weintraub, H. J. R. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**:21–32.

22 Sobel, M., Soler, D. F., Kermode, J. C. and Harris, R. B. (1992) Localisation and characterization of a heparin binding domain peptide of human von Willebrand factor. *J Biol Chem* **267:**8857–8862.

23 Margalit, H., Fischer, N. and Ben-Sasson, S. A. (1993) Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues. *J Biol Chem* **268**:19228–19231.

24 Carrell, R. W., Christey, P. B. and Boswell, D. R. (1987) Haemostasis. In *Thrombosis*, (eds. M. Verstraete, J. Vermylen, H.R. Lijenen and J. Arnout), p 1-15. Leuven University Press, Leuven.

25 Petitou, M., Lormeau, J.-C. and Choay, J. (1988) Interaction of heparin and antithrombin III. The role of *O*-sulfate groups. *Eur J Biochem* **176**:637–640.

26 Sinaÿ, P. (1987) Recent advances in the study of the binding site of heparin to antithrombin. *Pure Appl Chem* **59**:445–448.

27 Sinaÿ, P., Jacquinet, J.-C., Petitou, M., Duchaussoy, P., Lederman, I., Choay, J. and Torri, G. (1984) Total synthesis of a heparin pentasaccharide fragment having high affinity for antithrombin III. *Carbohydr Res* **132**:C5–C9.

28 Jackson, R. L., Busch, S. J. and Cardin, A. D. (1991) Glycosaminoglycans: Molecular properties, protein interactions, and role in physiological processes [review]. *Physiol Rev* **71**:481–539.

29 Sheehan, J. P. and Sadler, J. E. (1994) Molecular mapping of the heparin-binding exosite of thrombin. *Proc Natl Acad Sci USA* **91**: 5518–5522.

30 Godavarti, R., Cooney, C. L., Langer, R. and Sasisekharan, R. (1996) Heparinase I from Flavobacterium heparinum. Identification of a critical histidine residue essential for catalysis as probed by chemical modification and site-directed mutagenesis. *Biochemistry* **35**:6846–6852.

31 Rosenfeld, L. and Danishefsky, I. (1986) A fragment of antithrombin that binds both heparin and thrombin. *Biochem J* **237**:639–646.

32 Bae, J., Desai, U. R., Pervin, A., Caldwell, E. E. O., Weiler, J. M. and Linhardt, R. J. (1994) Interaction of heparin with synthetic antithrombin III peptide analogues. *Biochem J* **301**:121–129.

33 Tollefsen, D. M., Maimone, M. M., McGuire, E. A. and Peacock, M. E. (1989) Heparin cofactor II activation by dermatan sulfate. *Ann NY Acad Sci* **556**:116–122.

34 Linhardt, R. J. and Hileman, R. E. (1995) Dermatan sulfate as a potential therapeutic agent. *Gen Pharmac* **26**:443–451.

35 Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J. and Rees, D. C. (1996) Heparin structure and interactions with basic fibroblast growth factor. *Science* **271**:1116–1120.

36 Walker, A., Turnbull, J. E. and Gallagher, J. T. (1994) Specific heparan sulfate saccharides mediate the activity of basic fibroblast growth factor. *J Biol Chem* **269**:931–935.

37 Wang, H., Bernfield, M., Toida, T., Kim, Y. S., Capila, I., Hileman, R. E. and Linhardt, R. J. (1997) Glycosaminoglycans can influence fibroblast growth factor-2 mitogenicity without significant growth factor binding. *Biochem Biophys Res Comm* **235**:369–373.

38 Ashikari, S., Habuchi, H. and Kimata, K. (1995) Characterization of heparin sulfate oligosaccharides that bind to hepatocyte growth factor. *J Biol Chem* **270**:29586–29593.

39 Mulloy, B., Forster, M. J., Jones, C. and Davies, D. B. (1993) NMR and molecular modeling of the solution conformation of heparin. *Biochem J* **293**:849–858.

40 Record, T. M., Lohman, T. M. and de Haseth, P. J. (1976) Ion effects on ligand-nucleic acid interactions. *Mol Biol* **107**:145–158.

41 Manning, G. S. (1972) On the application of polyelectrolyte "limiting laws" to the helix-coil transition of DNA. I. Excess univalent cations. *Biopolymers* **11**, 937-949.

42 Lohman, T. M. and Mascotti, D. P. (1992) Thermodynamics of ligand-nucleic acid interactions. *Methods Enzymol* **212**:400–424.

43 Sicinska, W., Adams, B. and Lerner, L. (1993) A detailed 1H and 13C NMR study of a repeating disaccharide of hyaluronan: the effects of temperature and counterion type. *Carbohydr Res* **242**:29–51.

44 Anderson, C. F., Record, M. T. and Hart, P. A. (1978) Sodium-23 NMR studies of cation-DNA interactions. *Biophys Chem* **7**:301–316.

45 Manning, G. S. (1972) On the application of polyelectrolyte "limiting laws" to the helix-coil transition of DNA. I. Excess univalent cations. *Biopolymers* **11**:937–949.

46 Olson, S. T., Halvorson, H. R. and Bjork, I. (1991) Quantitative characterization of the thrombin-heparin interaction. *J Biol Chem* **266**:6342–6352.

47 Faller, B., Mely, Y., Gerard, D. and Bieth, J. G. (1992) Heparin-induced conformational change and activation of mucus proteinase inhibitor. *Biochemistry* **31**:8285–8290.

48 Thompson, L. D., Pantoliano, M. W. and Springer, B. A. (1994) Energetic characterization of the basic fibroblast growth factor-heparin interaction: Identification of the heparin binding domain. *Biochemistry* **33**:3831–3840.

49 Mascotti, D. P. and Lohman, T. M. (1995) Thermodynamics of charged oligopeptide-heparin interactions. *Biochemistry* **34**:2908–2915.

50 Caldwell, E. E. O., Nadkarni, V. D., Fromm, J. R., Linhardt, R. J. and Weiler, J. M. (1996) Importance of specific amino acids in protein binding sites for heparin and heparan sulfate. *Intl J Biochem Cell Biol* **28**:203–216.

51 Fromm, J. R., Hileman, R. E., Caldwell, E. E. O., Weiler, J. M. and Linhardt, R. J. (1995) Differences in the interaction of heparin with arginine and lysine and the importance of these basic amino acids in the binding of heparin to acidic fibroblast growth factor. *Arch Biochem Biophys* **323**:279–287.

52 Fromm, J. R. (1996) The interaction of proteins with glycosaminoglycans: Studies on the interaction of acidic fibroblast growth factor with heparin and heparan sulfate. *Dissertation* University of Iowa, Iowa City, IA.

53 Fromm, J. R., Hileman, R. E., Caldwell, E. E. O., Weiler, J. M. and Linhardt, R. J. (1997) Pattern and spacing of basic amino acids in heparin binding sites. *Arch Biochem Biophys* **343**:92–100.

54 Fromm, J. R., Hileman, R. E., Weiler, J. M. and Linhardt, R. J. (1997) Interaction of fibroblast growth factor-1 and related peptides with heparan sulfate and its oligosaccharides. *Arch Biochem Biophys* **346**:252–262.

55 Volkin, D. B., Tsai, P. K., Dabora, J. M., Gress, J. O., Burke, C. J. and Linhardt, R. J. (1993) Physical stabilization of acidic fibroblast growth factor by polyanions. *Arch Biochem Biophys* **300**:30–41.

56 Mach, H., Volkin, D. B., Burke, C. J., Middaugh, C. R., Linhardt, R. J., Fromm, J., Loganathan, D. and Mattsson, L. (1993) Nature of the interaction of heparin with acidic fibroblast growth factor. *Biochemistry* **32**:5480–5489.

57 Edens, R. E., Fromm, J. R., Fromm, S. J., Linhardt, R. J. and Weiler, J. M. (1995) Two-dimension affinity resolution electrophoresis demonstrates that three distinct heparin populations interact with antithrombin III. *Biochemistry* **34**:2400–2407.

58 Nelson, R. M., Venot, A., Bevilacqua, M. P., Linhardt, R. J. and Stamenkovic, I. (1995) Carbohydrate-protein interactions in vascular biology [review]. *Ann Rev Cell Dev Biol* **11**:601–631.
59 Zhu, X., Hsu, B. T. and Rees, D. C. (1993) Structural studies of the binding of the anti-ulcer drug sucrose octasulfate to acidic fibroblast growth factor. *Structure* **1**:27–34.
60 Chou, P. Y. and Fasman, G. D. (1978) Empirical predictions of protein conformation. *Ann Rev Biochem* **47**:251–278.
61 Huntington, J. A., Olson, S. T., Fan, B. and Gettins, P. G. (1996) Mechanism of heparin activation of antithrombin. Evidence for reactive centerloop preinsertion with expulsion upon heparin binding. *Biochemistry* **35**:8495–8503.
62 Mille, B., Watton, J., Barrowcliffe, T. W., Mani, J. C. and Lane, D. A. (1994) Role of N- and C-terminal amino acids in antithrombin binding to pentasaccharide. *J Biol Chem* **269**:29435–29443.
63 van den Born, J., Gunnarsson, K., Bakker, M. A., Kjellen, L., Kusche-Gullberg, M., Maccarana, M., Berden, J. H. and Lindahl, U. (1995) Presence of *N*-unsubstituted glucosamine units in native heparan sulfate revealed by a monoclonal antibody. *J Biol Chem* **270**:31303–31309.
64 Toida, T., Yoshida, H., Toyoda, H., Koshiishi, I., Imanari, T., Hileman, R. E., Fromm, J. R. and Linhardt, R. J. (1997) Structural differences and the presence of unsubstituted amino groups in heparan sulphates from different tissues and species. *Biochem J* **322**:499–506.