

Targeting the glycosaminoglycan-binding sites on proteins



Sulfated glycosaminoglycans bind to a wide variety of proteins, and in so doing have important roles in diverse biological processes. Selective mimics or inhibitors of protein–glycosaminoglycan interactions could have broad application in biology and medicine

Chemistry & Biology October 1994, 1:73–78

Interaction with glycosaminoglycans — especially heparin and heparan sulfate — appears to be important for the proper function of diverse cell-surface and secreted proteins. Compounds that selectively mimic or inhibit the binding of glycosaminoglycans to these proteins could be enormously useful in scientific and clinical areas as diverse as blood coagulation, cell growth control, atherosclerosis, inflammation, wound healing, tumor metastasis, and degenerative nervous system disease. Recent studies suggest that glycosaminoglycan-binding sites are typically extended clusters or bands of positively charged amino acids distributed along the surface of a protein, or in a shallow groove. Synthesizing or isolating compounds that both interact with these sites and exhibit protein selectivity could present some technical challenges, but the challenges would be well worth taking.

The importance of protein–glycosaminoglycan interactions

Sulfated glycosaminoglycans (GAGs) are linear polyanionic polysaccharides that are produced by most animal cells. They are constructed by the stepwise polymerization of simple disaccharide homopolymer backbones, which are then chemically modified in a variety of ways (for example, C-5 epimerization, N- and O-sulfation; see Fig. 1). The first steps of GAG polymerization typically occur on short oligosaccharides attached to serine hydroxyl groups of proteins. Generally, GAGs remain attached to the protein 'cores' on which they were synthesized; the resulting GAG-bearing glycoproteins are known as proteoglycans. GAGs occur abundantly (as proteoglycans) on the surfaces of animal cells and in the extracellular matrices that lie between cells. Depending on backbone structure, sulfated GAGs are said to belong to the heparan sulfate, chondroitin sulfate/dermatan sulfate, or keratan sulfate family. As a result of variations in chain polymerization and degree of modification, GAGs within any particular class can vary considerably in size, charge density, and fine structure (in other words, how dense the modifications are along the chain and the order in which they are arranged).

GAGs, especially heparan sulfates, are able to bind to a wide variety of physiologically and developmentally important cell-surface, secreted, and extracellular matrix proteins (Table 1). With only a few exceptions, the biological functions of GAGs are thought to stem from the effects they have on the proteins to which they bind.

For example, binding to GAGs very effectively protects some proteins from proteolytic degradation or denaturation. Other proteins are induced to change their conformation, and thus alter their biological activity, by binding to GAGs. By using GAGs to capture ligands, proteoglycans can sequester important proteins or, alternatively, deliver captured proteins efficiently to cells.

Perhaps the most fascinating functions of GAGs occur when a single GAG chain binds two protein ligands, an event that is favored by the large size of GAG chains. If the two proteins themselves interact, then the presence of a GAG chain can both catalyze that interaction (by effectively rendering the protein–protein interaction intramolecular, rather than bimolecular) and stabilize it (by forming a ternary GAG–protein–protein complex). The catalytic effect appears to explain much of the ability of the GAG heparin (a form of heparan sulfate) to increase by several orders of magnitude the rate of inactivation of the GAG-binding protease thrombin by the GAG-binding protease inhibitor antithrombin III [1]. On the other hand, it appears to be the stabilization effect that is the main factor in the ability of cell-surface heparan sulfate to mediate a large increase in the apparent affinity of the GAG-binding growth factor FGF-2 (basic fibroblast growth factor) for its GAG-binding cell surface receptor [2,3].

From these examples, it appears that GAGs are highly significant modulators of the functions of physiologically important proteins. The fact that GAGs are ubiquitously expressed, however, sometimes leads to the misapprehension that there must be little or no cell- or tissue-specificity in their actions. In fact, dramatic tissue-specific and cell-specific differences exist in the sizes and fine structures of GAGs, and these differences can cause the same type of GAG (for example, heparan sulfate) isolated from two different cell types to bind with very different affinities to a single protein ligand [4]. Indeed, for at least one protein, it is clear that high-affinity GAG-binding is critically dependent on a precise sequence of carbohydrate modifications [5]. For other proteins, however, it is likely that considerable variability in GAG structure can be tolerated.

How do proteins bind GAGs?

In recent years, considerable progress has been made in elucidating the features of protein structure that are important

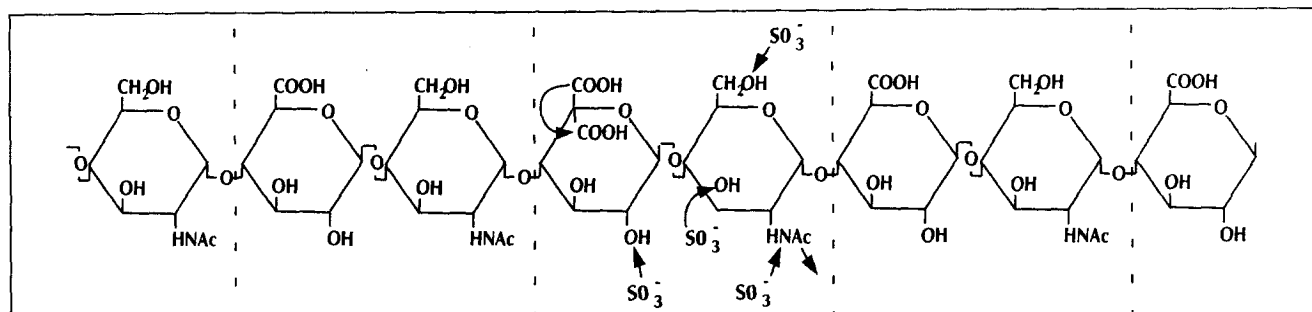


Fig. 1. Structure and biosynthesis of heparan sulfate, a protein-binding GAG. Eight contiguous sugars within a heparan sulfate polymer are shown (typical polymer lengths are 100–300 monosaccharides). The disaccharide repeat unit (D-glucuronic acid β 1 \rightarrow 4 D-N-acetylglucosamine α 1 \rightarrow 4) is demarcated by dotted lines. For the central disaccharide, all five known modification reactions are shown (three O-sulfations, one N-deacetylation/N-sulfation, and the epimerization of D-glucuronic acid to L-iduronic acid). In a mature heparan sulfate chain, each disaccharide may have undergone anywhere from zero to all five modifications, although not all combinations are possible, as some modifications seem to be prerequisites for others. Heparan sulfates also exhibit larger-scale structural features, in which blocks of largely unmodified sugars alternate with blocks of heavily modified ones. The GAG heparin is a form of heparan sulfate that is fairly uniformly and heavily modified; it is usually isolated in the form of GAG fragments of only 20–80 monosaccharides in length.

for binding GAGs. Such progress has relied upon the analysis of naturally occurring and site-directed mutations in GAG-binding proteins, as well as determinations of protein structure by X-ray crystallography and NMR.

Given that GAGs are essentially hydrophilic polyanions, it is not surprising that all studies have pointed to an essential role for the side chains of cationic amino acids (Lys, Arg) in GAG binding. Indeed, most GAG–protein interactions are predominantly electrostatic in character (see [6], for example), although at least one exception is

known (FGF-2, which has an unusually high affinity for heparin, appears to derive up to 70% of its free energy of binding from hydrogen bonds or van der Waals packing interactions [7]). Initial examinations of the sequences of known GAG-binding proteins suggested that certain short polybasic amino-acid motifs were characteristic of GAG-binding sites [8]. Recently, however, it has become clear from new three-dimensional structural information that GAG-binding sites can also be formed by the juxtaposition of amino-acid side chains that are discontinuous in the primary amino-acid sequence. Indeed, for

Table 1. Glycosaminoglycan-binding proteins.

Type of protein	Examples	Physiological significance
Polypeptide growth factors	Fibroblast growth factors (FGFs 1–9); platelet-derived growth factor (PDGF); vascular endothelial growth factor (VEGF); hepatocyte growth factor (HGF); transforming growth factor β (TGF β); interferon γ ; interleukin 3 (IL-3); granulocyte–macrophage colony-stimulating factor (GM-CSF); heregulins; chemokines (including interleukin-8, MIP-1 β , and others)	Cell proliferation; blood vessel formation (angiogenesis); cell motility; immune response; inflammation; cancer and metastasis
Extracellular matrix components	Laminins; fibronectins; thrombospondins; tenascins; collagens; vitronectin; von Willebrand's factor, amyloid fibrils	Cell attachment, motility and invasion; cell differentiation, wound repair; metastasis, amyloid diseases (for example, Alzheimer's disease)
Proteases and anti-proteases	Thrombin, tissue plasminogen activator; urokinase plasminogen activator; clotting factors IX and XI; various complement components; mast cell proteases; antithrombin III; heparin cofactor II; protease nexin I; plasminogen activator inhibitor-1, amyloid β -protein precursor	Blood coagulation; fibrinolysis (dissolution of blood clots); immune response; inflammation; cell invasion; Alzheimer's disease
Cell–cell adhesion molecules	N-CAM; L1; myelin-associated glycoprotein; PECAM-1; selectins	Cell attachment; nervous system development; inflammation
Proteins involved in lipoprotein metabolism	Apolipoprotein B, apolipoprotein E, lipoprotein lipase, other lipases	Lipid uptake; atherosclerosis; Alzheimer's disease
Other	Angiogenin; lactoferrin; various viral proteins (for example, proteins from HIV, herpes viruses)	Blood vessel formation (angiogenesis); inflammation; infectious disease

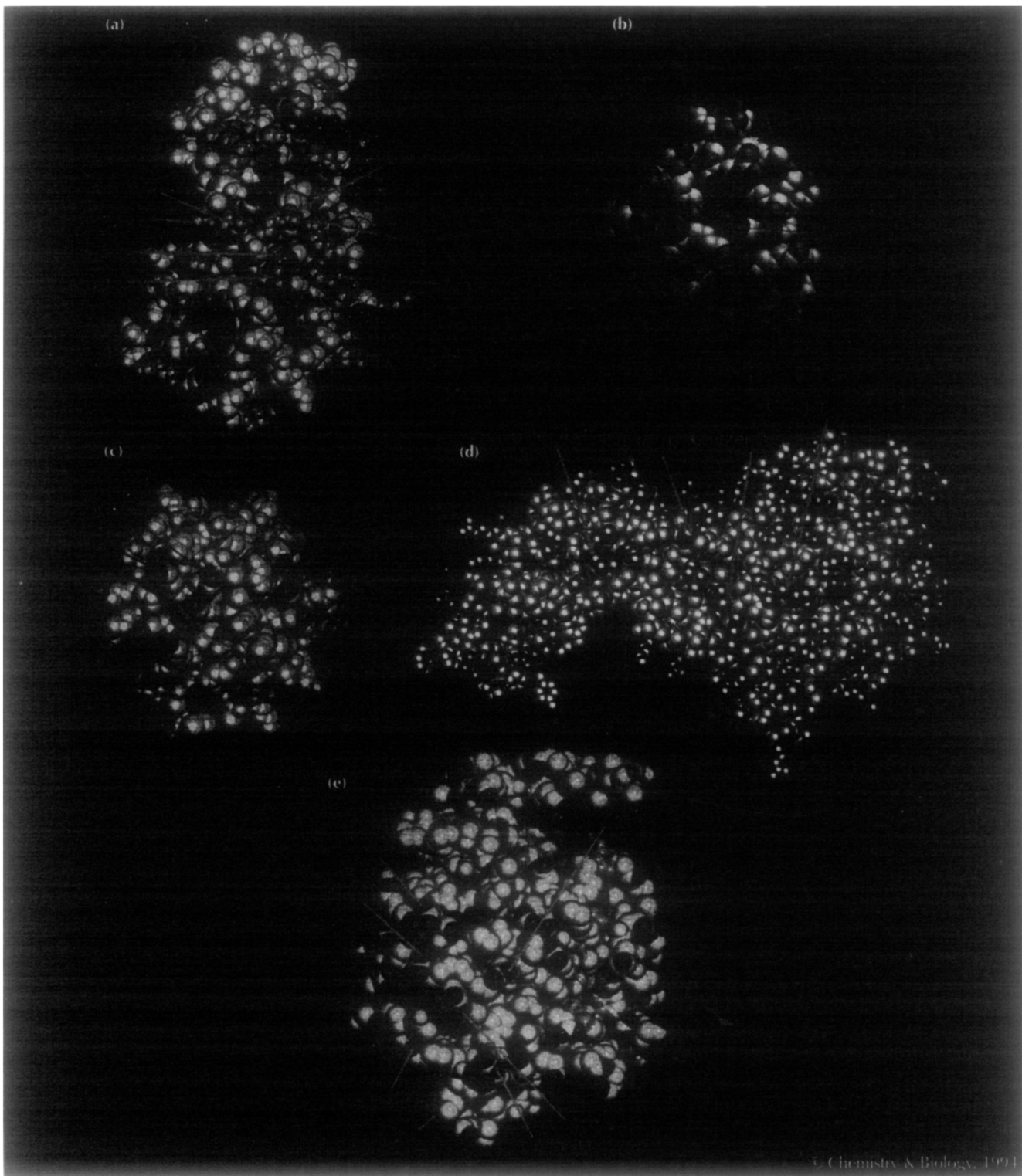


Fig. 2. Apparent GAG-binding sites of representative proteins. **(a)** Crystal structure of the active form of antithrombin III, a highly GAG-sequence-specific serine protease inhibitor [21]. Residues important for binding heparin are indicated. Basic residues are colored blue, acidic residues colored red. Image produced using Quanta 4.0. **(b)** Basic fibroblast growth factor. This growth factor has an unusually high affinity for heparin and heparan sulfate. Amino-acid side chains important for heparin binding are indicated [22]. Nitrogen atoms are colored blue, oxygen atoms red. **(c)** Platelet factor four. The AB dimer of the homotetramer is shown (in the tetramer, the CD dimer lies behind the structure shown) [23]. Four lysine residues known to be essential for heparin binding (labeled) are contributed by each monomer. These eight lysines define a linear array of positive charge that is continued around the sides of the dimer by additional arginine residues. The nitrogen atoms of Lys and Arg side chains are colored blue and the oxygen atoms of Asp and Glu side chains red. **(d)** Lipoprotein lipase. This enzyme remains associated with cell surfaces by binding to the GAGs of cell-surface proteoglycans. Four clusters of basic residues (labeled) form a highly extended patch of positive charge that appears to constitute the GAG-binding site [24]. The side chains of basic residues are colored blue; the side chains of acidic residues are red. Image produced using TurboFrodo, Bio-Graphics, Marseilles. **(e)** Lactoferrin. Two non-contiguous clusters of residues contribute six Arg and one Lys (labeled) to the apparent GAG-binding site of this inflammatory response protein [25]. Atoms are colored as in (c).

oligomeric proteins, GAG-binding sites can involve residues on multiple polypeptides (see [9] for example).

Some examples of apparent GAG-binding sites are shown in Fig. 2. Although the assignments of these sites have yet to be confirmed by direct structural analysis of protein-GAG complexes, they are supported by considerable data from protein fragments and mutants. The overall picture that has emerged is that GAG-binding sites are extended collections of positive charge, often lying along one exposed face of a protein, and sometimes wrapping around multiple faces. The end-to-end lengths of these extended clusters are comparable to the minimum GAG chain lengths that are required for binding (typically 6–12 monosaccharide units, ~25–50 Å long). GAG-binding sites may lie in shallow grooves, but so far have not been found in deep clefts or pockets. To date, the protein folding motifs that can be involved in forming GAG-binding sites include immunoglobulin domains [10], fibronectin type III repeats [11], kringle domains [12], four-helix bundles [13], laminin A chain globule domains [14], and collagen triple helices [15].

Can GAGs be used as drugs? — the good news and the bad

Heparin, a particularly highly modified (and therefore highly charged) type of heparan sulfate, is relatively easily and cheaply purified from various animal tissues, and is widely used both pharmacologically and as an affinity ligand for protein chromatography (indeed, for most of the proteins in Table 1, it was adsorption to heparin affinity columns that provided the first hints of GAG-binding). The main pharmacological use of heparin stems from its ability to catalyze the inactivation of thrombin by the blood protein antithrombin III (see above). As thrombin is the key enzyme that initiates clotting, heparin that finds its way into the blood is a potent anticoagulant. Because the affinity of heparin for antithrombin III is relatively strong, and because heparin's effects are essentially catalytic, it is potent at relatively low systemic doses. Indeed, heparin has long been the drug of choice for short-term anticoagulation therapy.

The bad news is that the potent anticoagulant action of heparin limits its clinical utility to low systemic doses (excessive anticoagulation can, of course, be fatal). Unfortunately, the need to use low doses means that heparin has a less than optimal chance of influencing the functions of GAG-binding proteins that either have lower affinities for heparin than antithrombin III, or are poorly accessible to the circulation. Despite this limitation, studies have shown that physiologically tolerable doses of heparin can have dramatic effects on serum lipid metabolism in clinical trials [16], and on blood vessel growth [17] and vascular smooth muscle proliferation [18] in animal models.

Were anticoagulation the only problem associated with heparin as a drug, it could be (and in some cases has been) surmounted by prefractionating heparin to remove the chains that bind antithrombin III with high affinity

(only about 30 % of most commercial heparin does so). Unfortunately, there is a larger underlying problem that is not so easily solved. The fact that heparin binds to so many GAG-binding proteins (it binds to most of the molecules in Table 1) means that it cannot be used as a specific inhibitor of any one of them. Many approaches have been proposed, some of which have been tried, to make heparin a more specific inhibitor. Such approaches include fractionation of heparin by affinity for different proteins, fractionation of heparin by size, chemical modification of heparin, and chemical synthesis of defined heparin oligosaccharides. In addition, attempts have been made to develop other GAGs, or other polyanionic polysaccharides (for example, sulfated dextrans) as heparin mimetics. Overall, these approaches have met with limited success.

How chemists can help

To be fair, it is too early to know whether GAGs, modified GAGs, or other polysaccharides can ever be developed into specific drugs that are useful in altering the functions of individual GAG-binding proteins. Still, it is worth considering whether some altogether different approaches might stand a better chance of success. For example, one might argue that there are inherent limitations in the binding affinity that can be achieved by compounds that interact with proteins only in the way that GAGs usually do, that is, by forming multiple electrostatic contacts with cationic amino-acid side chains. Consistent with this view, most GAG-binding proteins bind GAGs with dissociation constants in the range of 2×10^{-8} to 10^{-5} M (as mentioned earlier, FGF-2 binds heparin with a tighter affinity ($K_d \sim 10^{-9}$ M), but apparently achieves much of this strength through non-ionic interactions). Although drugs with affinities similar to GAGs themselves might be useful in mimicking GAG actions, drugs that are meant to be competitive inhibitors of GAG binding would benefit from much tighter affinities (especially as GAGs are so abundant).

Similarly, one might argue that there are inherent limitations not just in the affinity, but also in the specificity that can be achieved by compounds that bind proteins predominantly through electrostatic interactions. For example, studies of protein-protein and protein-nucleic acid binding suggest that the source of high specificity usually lies in hydrophobic and hydrogen-bonding interactions (see [19] for example).

For these reasons, it is natural to ask whether compounds could be found that interact with the GAG-binding sites on proteins, yet make sufficient non-electrostatic contacts to achieve both high affinity and high specificity. A guess at what such a molecule might look like is shown schematically in Fig. 3. In this modular amphiphilic compound, a central highly anionic module is flanked on both sides by uncharged polar or apolar modules. The idea is that the central module would interact electrostatically with some or all of the cationic residues that define the protein's GAG-binding site, while the flanking

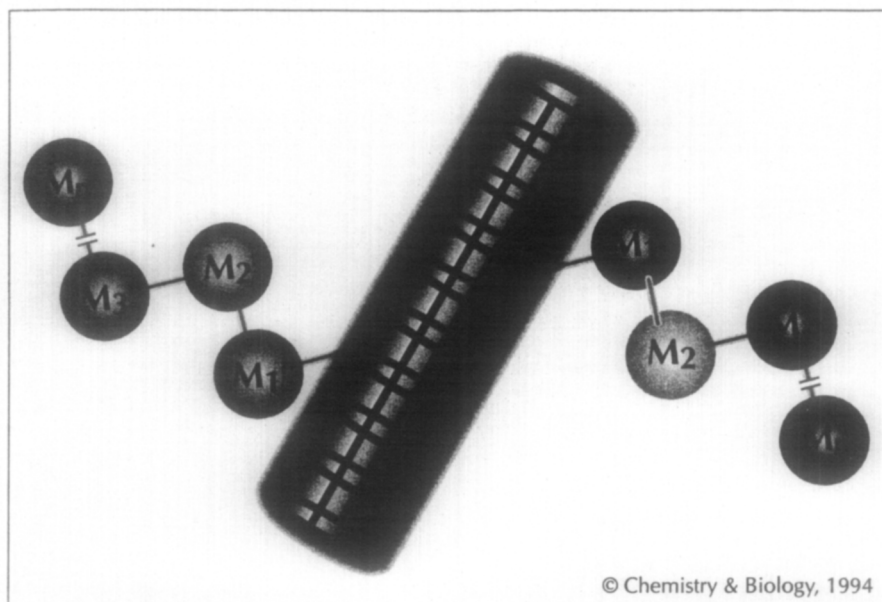


Fig. 3. How compounds that bind the GAG-binding sites on specific proteins might look. The red central domain is negatively charged, and might be constant in many different types of specific GAG mimics. The 'arms' are variable, and might be generated using combinatorial chemistry approaches.

modules form hydrogen bonds or hydrophobic interactions with residues on either side of the GAG-binding site. Although it might be possible to design molecules of this type 'rationally', based on known protein structures, it seems that such compounds would just as readily lend themselves to semi-random combinatorial approaches, in which active molecules are isolated by screening. For example, one might want to choose a fixed central module (such as a large polysulfonated aromatic structure with derivatizable groups at either end) and add the flanking modules using any of the random, modular syntheses that have recently been developed for the production of molecule 'libraries' (see [20] for review). One distinct advantage of a combinatorial approach is that screening for 'lead' compounds could be very straightforward. For example, one could adsorb mixtures of known GAG-binding proteins to heparin columns, and screen for compounds that elute single protein species.

Obviously, Fig. 3 illustrates only one of many kinds of approaches that could be used in synthesizing ligands for the GAG-binding sites on specific proteins. The hope is that this discussion will inspire chemists to give serious thought to GAG-binding proteins as both worthwhile and feasible targets for the synthesis of specific new ligands — ligands that should benefit both the scientific and medical communities.

Acknowledgements: The author gratefully acknowledges the generous assistance of the investigators who provided the images shown in Fig. 2. The structure of basic fibroblast growth factor was determined in the laboratory of Brian W. Matthews, University of Oregon; the image was kindly provided by Ingrid Vetter. Coordinates for platelet factor four were provided by Xiaohua Zhang, and the photograph prepared with the assistance of Imre Berger and David Smith. The photograph of the lactoferrin GAG-binding site was prepared by David Mann, who kindly provided it prior to publication. The image of lipoprotein lipase was kindly provided by Christian Cambillau, and that of antithrombin III by James Whisstock and Robin Carrell.

References

1. Evans, D.L., Marshall, C.J., Christey, P.B. & Carrell, R.W. (1992). Heparin-binding site, conformational change and activation of antithrombin. *Biochemistry* **31**, 12629–12642.
2. Nugent, M.A. & Edelman, E.R. (1992). Kinetics of basic fibroblast growth factor binding to its receptor and heparan sulfate proteoglycan: A mechanism for cooperativity. *Biochemistry* **31**, 8876–8883.
3. Kan, M., Wang, F., Xu, J., Crabb, J.W., Hon, J. & McKeehan, W.I. (1993). An essential heparin-binding domain in the fibroblast growth factor receptor kinase. *Science* **259**, 1918–1921.
4. Sanderson, R.D., Turnbull, J.E., Gallagher, J.T. & Lander, A.D. (1994). Fine structure of heparan sulfate regulates syndecan-1 function and cell behavior. *J. Biol. Chem.* **269**, 13100–13106.
5. Lindahl, U. (1989). Biosynthesis of heparin and related polysaccharides. In *Heparin* (Lane, D.A. and Lindahl, U. eds) pp. 159–189. CRC Press, Boca Raton, FL.
6. Olson, S.T., Halvorson, H.R. & Björk, I. (1991). Quantitative characterization of the thrombin–heparin interaction. Discrimination between specific and non-specific binding models. *J. Biol. Chem.* **266**, 6342–6352.
7. Thompson, L.D., Pantoliano, M.W. & Springer, B.A. (1994). Energetic characterization of the basic fibroblast growth factor–heparin interaction: Identification of the heparin-binding domain. *Biochemistry* **33**, 3831–3840.
8. Cardin, A.D. & Weintraub, H.J.R. (1989). Molecular modeling of protein–glycosaminoglycan interactions. *Arteriosclerosis* **9**, 21–32.
9. Stuckey, J.A., St. Charles, R. & Edwards, B.F.P. (1992). A model of the platelet factor 4 complex with heparin. *Proteins* **14**, 277–287.
10. Frelinger, A.L. III & Rutishauser, U. (1986). Topography of N-CAM structural and functional determinants. II. Placement of monoclonal antibody epitopes. *J. Cell Biol.* **103**, 1729–1737.
11. Ingham, K.C., Brew, S.A., Migliorini, M.M. & Busby, T.F. (1993). Binding of heparin by type III domains and peptides from the carboxy-terminal hep-2 region of fibronectin. *Biochemistry* **32**, 12548–12553.
12. Li, X., Bokman, A.M., Llinas, M., Smith, R.A. & Dobson, C.M. (1994). Solution structure of the kringle domain from urokinase-type plasminogen activator. *J. Mol. Biol.* **235**, 1548–1549.
13. Diederichs, K., Boone, T. & Karplus, P.A. (1991). Novel fold and putative receptor binding site of granulocyte–macrophage colony-stimulating factor. *Science* **254**, 1779–1782.
14. Sung, U., O'Rear, J.J. & Yurchenco, P.D. (1993). Cell and heparin binding sites in the distal long arm of laminin: identification of active and cryptic sites with recombinant and hybrid glycoprotein. *J. Cell Biol.* **123**, 1255–1268.
15. San Antonio, J.D., Lander, A.D., Karnovsky, M.J. & Slayter, I.I.S. (1994). Mapping the heparin-binding sites on type I collagen monomers and fibrils. *J. Cell Biol.* **125**, 1179–1188.
16. Weintraub, M., et al., & Rubinstein, A. (1994). Continuous intravenous heparin administration in humans causes a decrease in serum lipolytic activity and accumulation of chylomicrons in circulation. *J. Lipid Res.* **35**, 229–238.

17. Folkman, J. & Shing, Y. (1992). Control of angiogenesis by heparin and other sulfated polysaccharides. *Adv. Exp. Med. Biol.* **313**, 355–364.
18. Pukac, L.A., Hirsch, G.M., Lormeau, J.C., Petitou, M., Choay, J. & Karnovsky, M.J. (1991). Antiproliferative effects of novel, nonanticoagulant heparin derivative on vascular smooth muscle cells *in vitro* and *in vivo*. *Am. J. Pathol.* **139**, 1501–1509.
19. Janin, J. & Chothia, C. (1990). The structure of protein–protein recognition sites. *J. Biol. Chem.* **265**, 16027–16030.
20. Mitchison, T.J. (1994). Towards a pharmacological genetics. *Chemistry & Biology* **1**, 3–6.
21. Carrell, R.W., Stein, P.E., Fermi, G. & Wardell, M.R. (1994). Biological implication of a 3 Å structure of dimeric antithrombin. *Structure* **2**, 257–270.
22. Eriksson, A.E., Cousens, L.S. & Matthews, B.W. (1993). Refinement of the structure of human basic fibroblast growth factor at 1.6 Å resolution and analysis of presumed heparin binding site by selenate substitution. *Protein Science* **2**, 1274–1284.
23. Zhang, X., Chen, L., Bancroft, D.P., Lai, C.K. & Maione, T.E. (1994). Crystal structure of recombinant human platelet factor 4. *Biochemistry* **33**, 8361–8366.
24. van Tilbeurgh, H., Roussel, A., Lalouel, J.-M. & Cambillau, C. (1994). Lipoprotein lipase. Molecular model based on the pancreatic lipase X-ray structure: consequences for heparin-binding and catalysis. *J. Biol. Chem.* **269**, 4626–4633.
25. Mann, D.M., Romm, E. & Migliorini, M. (1994). Delineation of the glycosaminoglycan-binding site in the human inflammatory response protein lactoferrin. *J. Biol. Chem.* **269**, 23661–23667.

Arthur D Lander, Department of Brain and Cognitive Sciences and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA