Structural Biology of GAG–Protein Interactions

Martin Safo, Ph.D.

Department of Medicinal Chemistry

Email: msafo@vcu.edu

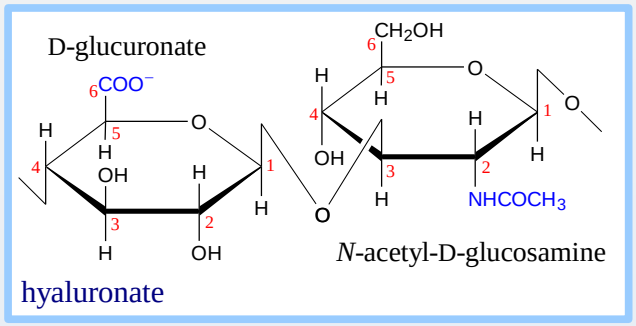
Office: Biotech One, Suite 212

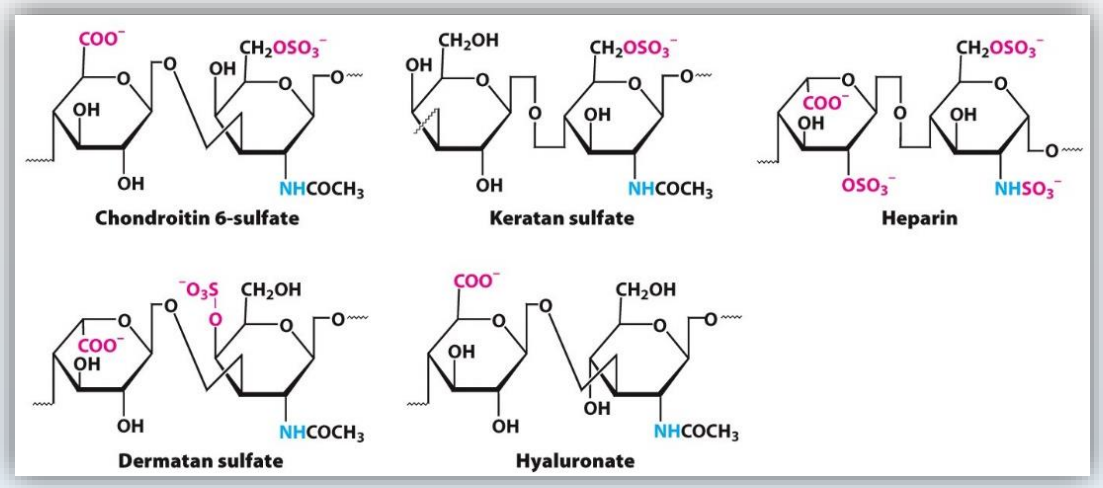
Glycosaminoglycans (GAGs)

- GAGs have diverse functional activities by interacting with numerous GAG binding proteins (GAGBP)
- Almost all GAGs contain charged sulfate/carboxylate groups that are sitting duck to interact with any protein with basic complementary enticing surface
- Yes there is some order in GAGs interaction – in terms of specificity, affinity and recognition
- The question is how does this come about – The answer lies in the current data of GAG interactions with GAG binding proteins

Glycosaminoglycans (GAGs)

- Repeating units of hexose/hexuronic acid (acidic sugar) and hexoxamine (amino sugar)





- The different GAGs differ according to the type of uronic acid and glycosamine unit they contain, as well as the geometry of the glycosidic bond between the units.
- The monosaccharides of the repeating disaccharides tend to be modified, with acidic groups, sulfated hydroxyl and amino groups, etc.
- Although each GAG has a predominant disaccharide component, heterogeneity exist in the sugars present in any given class of GAG

Many functions of proteoglycans are mediated by proteins which bind GAGs

Class	Example	Physiological/pathophysiological effect
Enzymes	glycosaminoglycan biosynthetic enzymes, thrombin and coagulation factors (proteases), complement proteins (esterases), extracellular superoxide dismutase, topoisomerase	multiple
Enzyme inhibitors	antithrombin III, heparin cofactor II, secretory leukocyte proteinase inhibitor, C1-esterase inhibitor	coagulation, inflammation, complement regulation
Cell adhesion proteins	P-selectin, L-selectin, some integrins	cell adhesion, inflammation, metastasis
Extracellular matrix proteins	laminin, fibronectin, collagens, thrombospondin, vitronectin, tenascin	cell adhesion, matrix organization
Chemokines	platelet factor IV, γ-interferon, interleukins	chemotaxis, signaling, inflammation
Growth factors	fibroblast growth factors, hepatocyte growth factor, vascular endothelial growth factor, insulin-like growth factor–binding proteins, TGF-β-binding proteins	mitogenesis, cell migration
Morphogens	hedgehogs, TGF-β family members	cell specification, tissue differentiation, development
Tyrosine-kinase growth factor receptors	fibroblast growth factor receptors, vascular endothelium growth factor receptor	mitogenesis
Lipid-binding proteins	apolipoproteins E and B, lipoprotein lipase, hepatic lipase, annexins	lipid metabolism, cell membrane functions

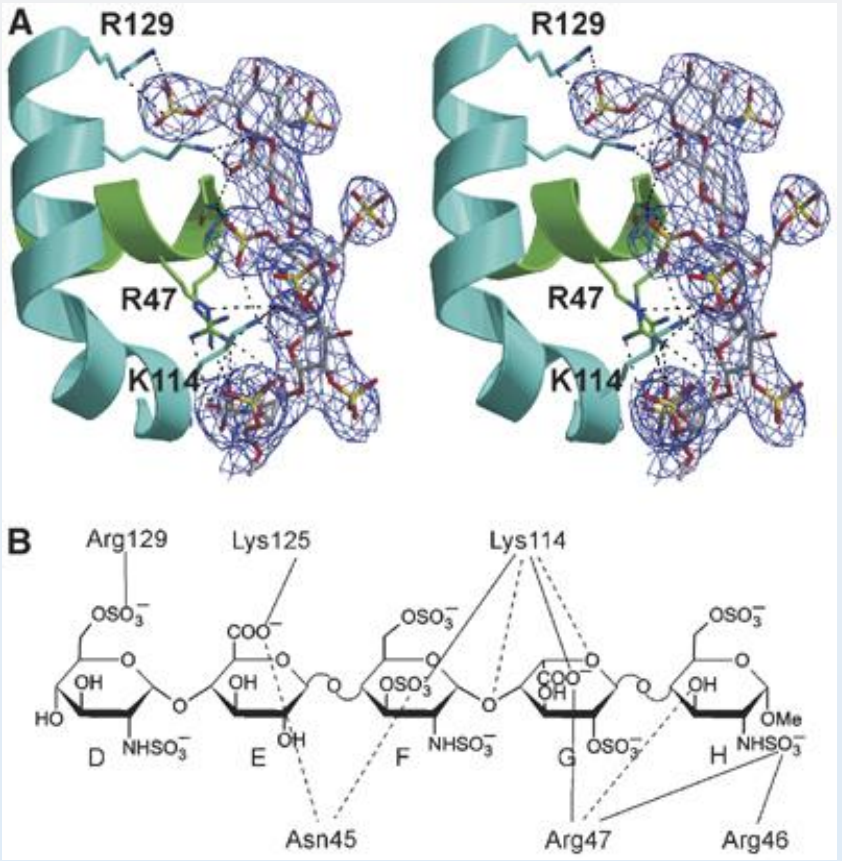
Adaptation, Table 35.1 Essentials in Glycobiology

Important Characteristic Features of GAG-Protein Interactions

- Main driving force for GAG-protein interactions is ionic interactions between the sulfate/carboxylate groups of GAGs and basic amino acids (Arg, Lys and to some extent His) of GAG binding protein's (GAGBP)
- The guanidino groups of Arg form more stable hydrogen bond/stronger electrostatic interactions with the sulfate/carboxylate groups of GAG
- Polar residues, usually Asn, Gln, and His play critical roles in GAG–protein interactions that involve hydrogen-bond
- Interactions of GAGs also involve van der Waals forces, hydrogen bonding and hydrophobic interactions with the sugar backbone
- Affinity of GAGs is also enhanced through hydrogen-bond interactions with smaller side-chain residues, e.g. Serine, Glycine – provides minimal steric constraint and flexibility
- GAGs recognize certain binding motifs in GAGBP
- Topology and distribution of the basic amino acids on the protein influences its affinity, specificity and recognition of GAG sequences
- Topology, sequence, monosaccharide modifications, and type/geometry of glycosidic bond of GAGs influence recognition of GAGBP

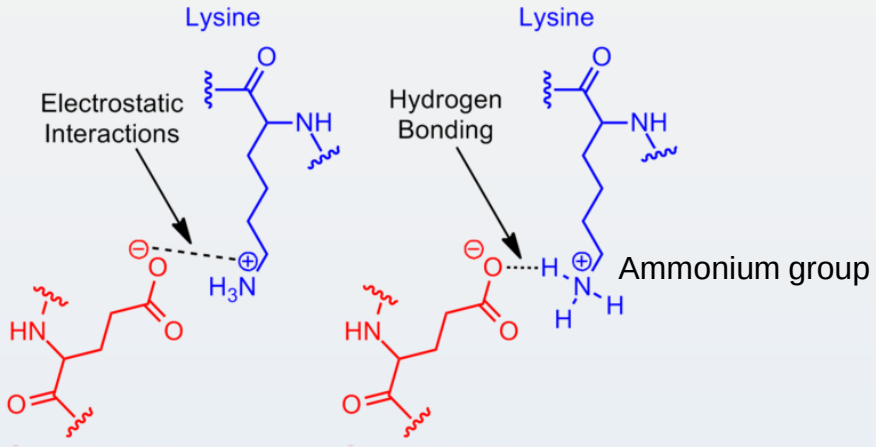
Electrostatic/Ionic Interactions are Central to GAG binding to GAG Binding Protein

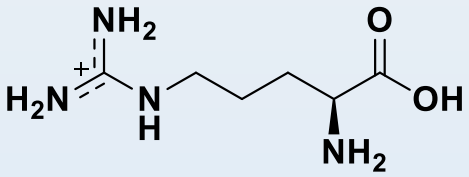
- Electrostatic interactions involving Lys and Arg dominate the interaction of most GAGBP and GAGs
- For example, the vast majority of characterized HA-binding sites contain 4 to 7 Lys and Arg residues. Typical of other GAGBPs interactions
- The basic amino acid in HA-ATIII interactions participates in at least 5 ionic interactions and contributes 40% of the binding energy. Remaining 60% by non-ionic interactions



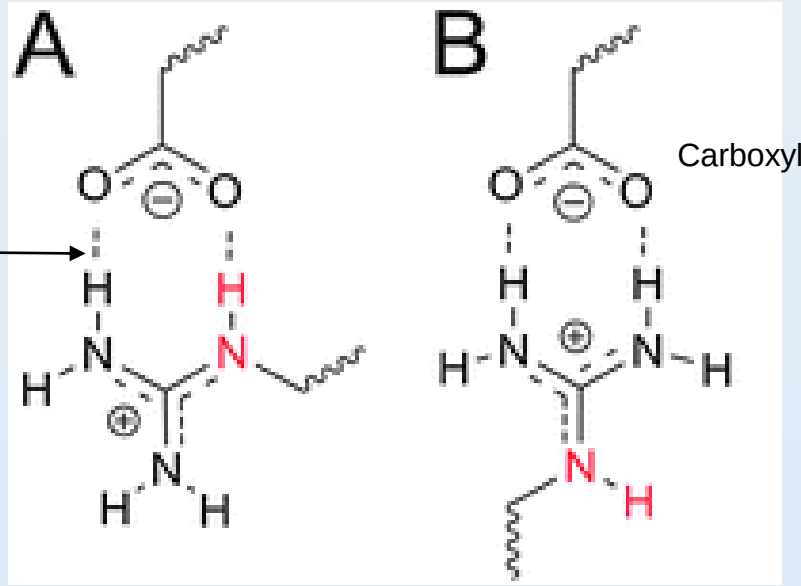
Antithrombin (AT-III)–pentasaccharide (heparin) interactions.

Arginine forms more stable hydrogen-bond/stronger electrostatic interactions than lysine





Bidentate interaction



Binding Affinity does not Depend only on the Number of GAGBP Basic Residues

- Number of basic residues does not usually correlate with the affinity of GAGBPs, but also includes the contribution of hydrogen bonding, Van der Waals forces, and hydrophobic interactions

- For example HA-binding sites of Fibroblast Growth Factor Receptor1 (FGFR1), thrombin, RAGE (receptor for advance glycation end products), contain seven basic residues, yet their affinities for HA range from 0.3 nM to 7 μM

- In contrast, HA-binding sites containing only four basic residues can have high binding affinity, e.g. basic vascular endothelial growth factor (VEGF) and fibroblast growth factors (FGF)

Important Characteristic Features of GAG-Protein Interactions

- Main driving force for GAG-protein interactions is ionic interactions between the sulfate/carboxylate groups of GAGs and basic amino acids (Arg, Lys and to some extent His) of GAG binding protein's (GAGBP)
- The guanidino groups of Arg form more stable hydrogen bond/stronger electrostatic interactions with the sulfate/carboxylate groups of GAG
- Polar residues, usually Asn, Gln, and His play critical roles in GAG–protein interactions that involve hydrogen-bond
- Interactions of GAGs also involve van der Waals forces, hydrogen bonding and hydrophobic interactions with the sugar backbone
- Affinity of GAGs is also enhanced through hydrogen-bond interactions with smaller side-chain residues, e.g. Serine, Glycine – provides minimal steric constraint and flexibility
- GAGs recognize certain binding motifs in GAGBP
- Topology and distribution of the basic amino acids on the protein influences its affinity, specificity and recognition of GAG sequences
- Topology, sequence, monosaccharide modifications, and type/geometry of glycosidic bond of GAGs influence recognition of GAGBP

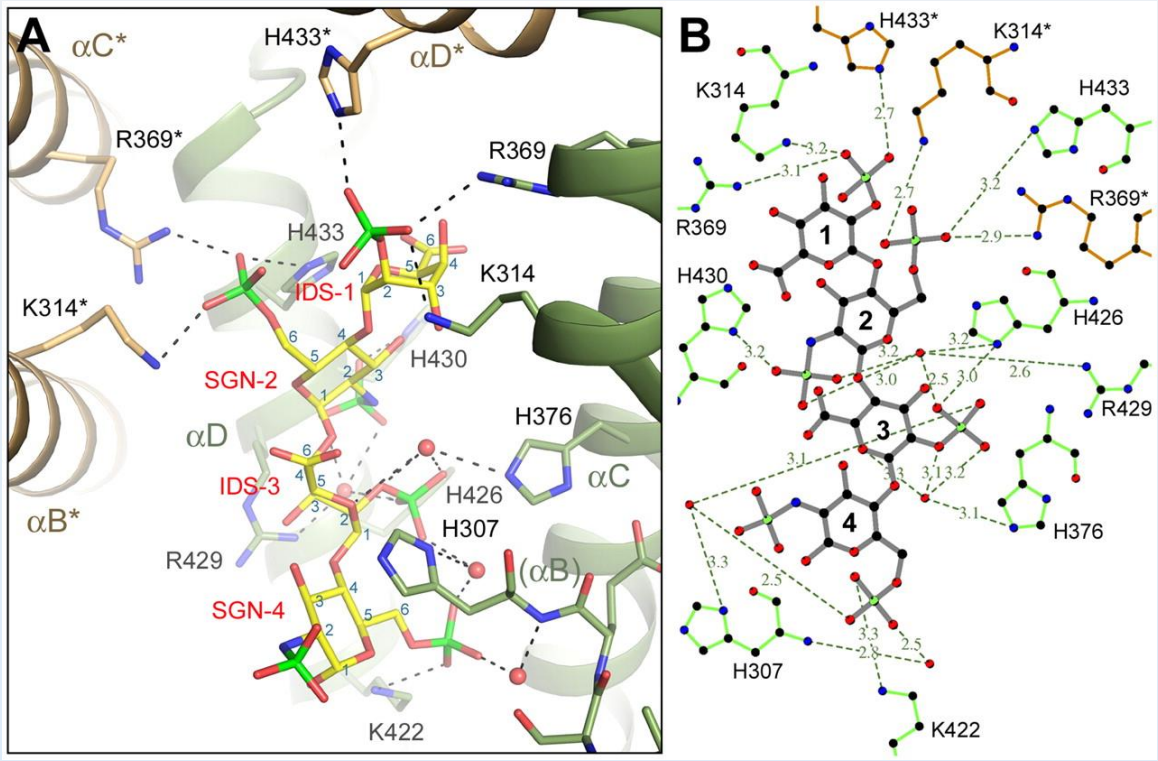
His Plays Important Role in GAG-binding

- His is not commonly found in GAG-binding sites but has been demonstrated by mutagenesis and/or X-ray crystallography in some GAGBPs, e.g. APLP-1, thrombin and annexin A2

- APLP-1 (E2 domain of amyloid precursor-like protein 1) plays important functions in synapse organization

- The HA-binding site of APLP-1 contains five His that make either direct or water-mediated hydrogen bonds with sulfate groups

- Mutation of each of the five His suggest that the histidines substantially contribute to binding

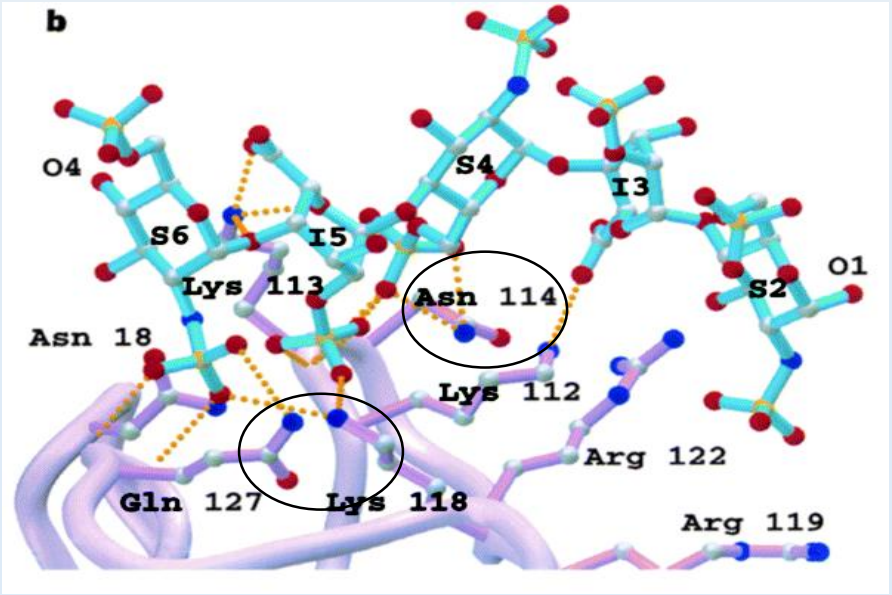


Asn and Gln are Important in GAG Interactions

- Other polar residues, especially Asn and Gln often make hydrogen bonds with GAG sulfate groups, e.g. in fibroblast growth factors (**FGF)**

- Contribution of Asn/Gln to binding is underestimated in most GAG binding sites because these residues are rarely targeted for study.

- One way to identify potential polar residues in GAG-binding sites would be to first define the boundary of GAG-binding sites by mutation of lysine and arginine residues, and then to target conserved polar residues within this region by mutagenesis.



Important Characteristic Features of GAG-Protein Interactions

- Main driving force for GAG-protein interactions is ionic interactions between the sulfate/carboxylate groups of GAGs and basic amino acids (Arg, Lys and to some extent His) of GAG binding protein's (GAGBP)
- The guanidino groups of Arg form more stable hydrogen bond/stronger electrostatic interactions with the sulfate/carboxylate groups of GAG
- Polar residues, usually Asn, Gln, and His play critical roles in GAG–protein interactions that involve hydrogen-bond
- Interactions of GAGs also involve van der Waals forces, hydrogen bonding and hydrophobic interactions with the sugar backbone
- Affinity of GAGs is also enhanced through hydrogen-bond interactions with smaller side-chain residues, e.g. Serine, Glycine – provides minimal steric constraint and flexibility
- GAGs recognize certain binding motifs in GAGBP
- Topology and distribution of the basic amino acids on the protein influences its affinity, specificity and recognition of GAG sequences
- Topology, sequence, monosaccharide modifications, and type/geometry of glycosidic bond of GAGs influence recognition of GAGBP

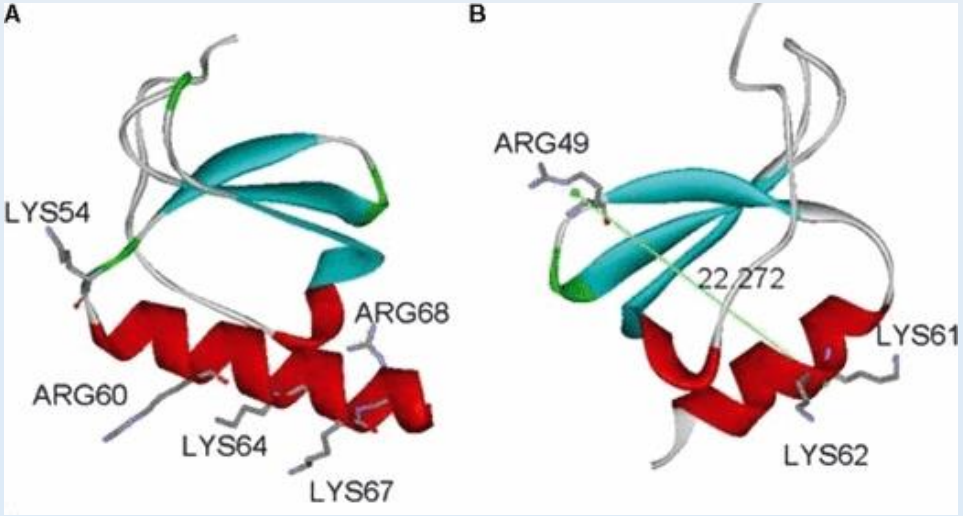
GAGs Recognize Certain Binding Motifs

- For example, several HS-protein complex structures suggest typical heparin-binding sites have the sequence XBBXBX or XBBBXXBX where B is usually Lys or Arg and X hydropathic residue - Asn, Ser, Ala, Gly, Ile, Leu and Tyr

- Other less occurring consensus sequences, e.g. XBBBXXBBBXXBBX

- Not all consensus sequences, including basic residues actually participate in the binding interaction – Specificity

- Basic residues that participate in the binding are usually far apart, and mostly located 20 Å apart



(A) Linear XBBBXXBX motif with basic arginine and lysine residues (grey) oriented on one surface of a helix (PDB code 3IL8).

(B) Linear motif with basic arginine and lysine residues (blue) spaced at a 20 Å for interaction with GAG

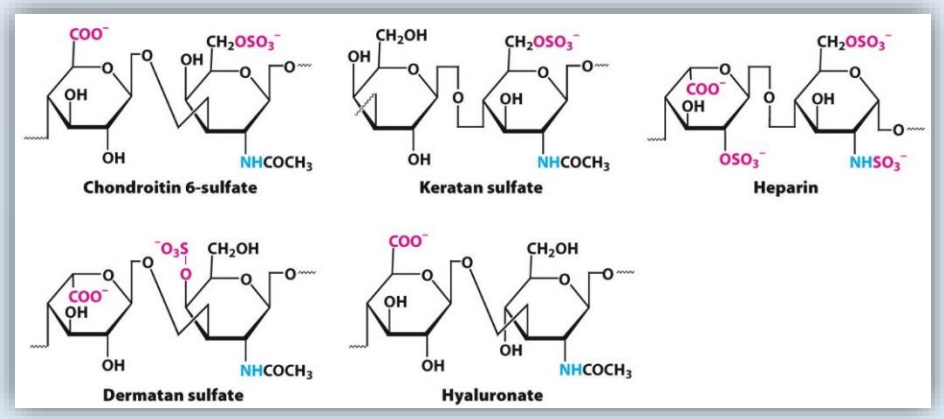
Characteristic Features of GAGs Affecting Recognition and Specificity

- The relative proportion of N- and O-linked sulfate groups and N-linked acetyl groups especially in HA/HS affect their interactions with protein

- The length and/or sequence of GAG influences its specificity and recognition of protein - provides optimal charge (orientation of sulfate groups) and surface (van der Waal contacts) complementarity with the protein

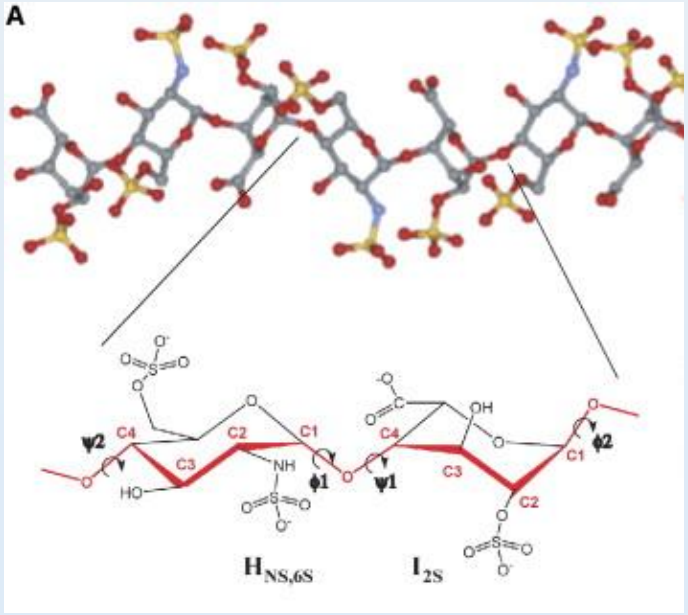
- Usually, Asparagine, Aspartate, Glutamine, Glutamate, Histidine and Tryptophan make the binding site for non-sulfated carbohydrates

- Binding leads to conformational change of the GAG around the glycosidic linkage (kinking) – to maximize binding and affinity



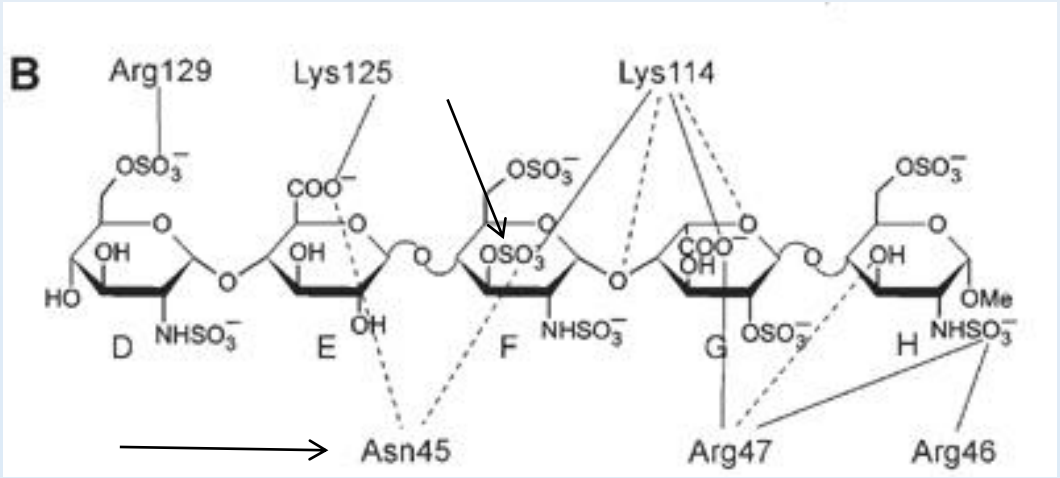
Kinking is Important for Optimal Binding and Recognition

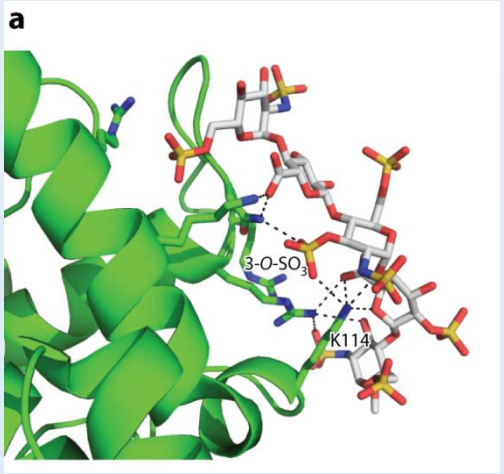
- Protein binding induces local distortions in the uniform helical structure of GAGs, manifested as changes in the glycosidic torsion angles
- The conformational changes enable an optimal fit in terms of both ionic and van der Waals contact between the oligosaccharide motif and the protein - provides specificity
- The kink-type depends on GAG sequence and protein binding site residues



Heparin–Antithrombin III Interactions

- The high-affinity interaction between AT and HA requires the 3-sulfo-*N*-sulfoglucosamine residue (GlcNS3S) in a very specific context.
- The 3-*O*-sulfate group, along with its amino acid partner in AT (Lys114), occupies a pivotal position in organizing both ionic and nonionic interaction networks that are responsible for the AT–HA interaction
- Binding studies suggest that the 3-*O*-sulfate group contributes ~60% of the binding free energy and that removal of this modification causes a 10^5-fold reduction in binding affinity to AT.

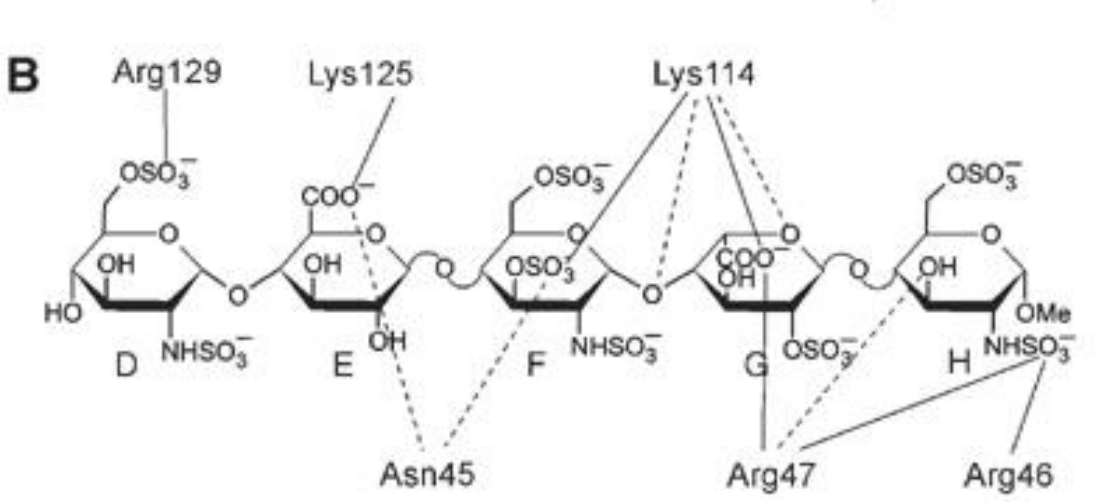




Specificity of HA–Antithrombin III Interactions

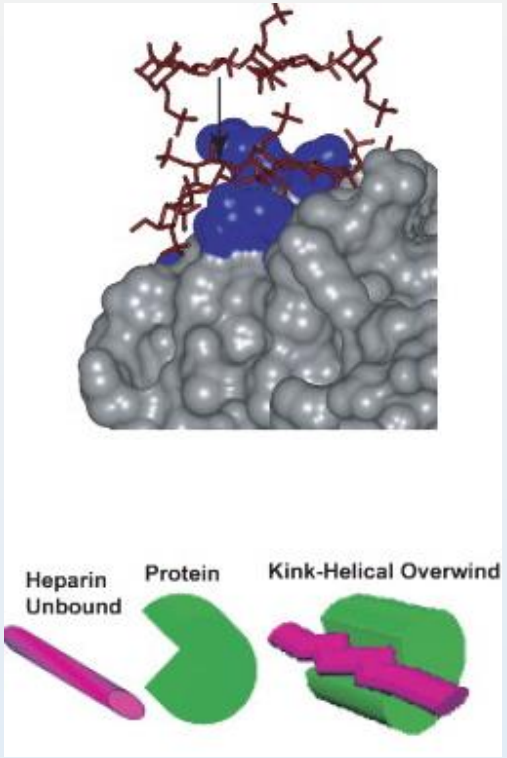
- In addition to the 3-*O*-sulfate/Lys114 interaction the surrounding sugar and protein residues also play important roles in promoting this interaction

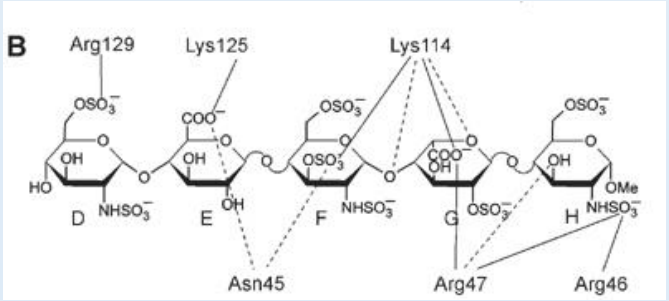
- For example, there must be a nonsulfated glucuronic acid at the nonreducing side of the GlcNS3S unit. If this residue is replaced with IdoA or 2-sulfo-IdoA (IdoA2S), then the oligosaccharide shows greatly reduced binding to AT

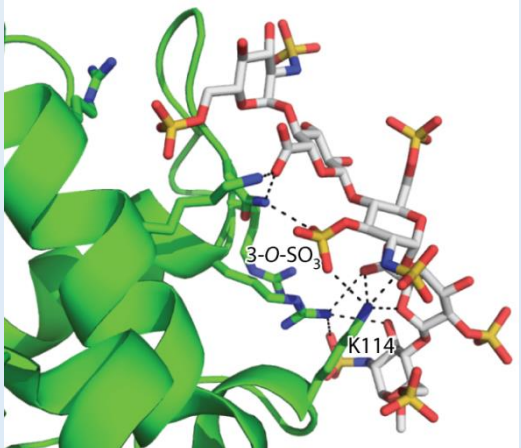


Structural Basis of the Specificity of HA–ATIII Interactions

- Contact between the HA sequence and the AT binding site residues induce a kink in the pentasaccharide motif
- The kink allows a critical interaction between the 3-O sulfate and Lys114 required for AT-III binding and recognition with HA
- Note that the kink is dependent on both HA sequence and the AT binding site residues





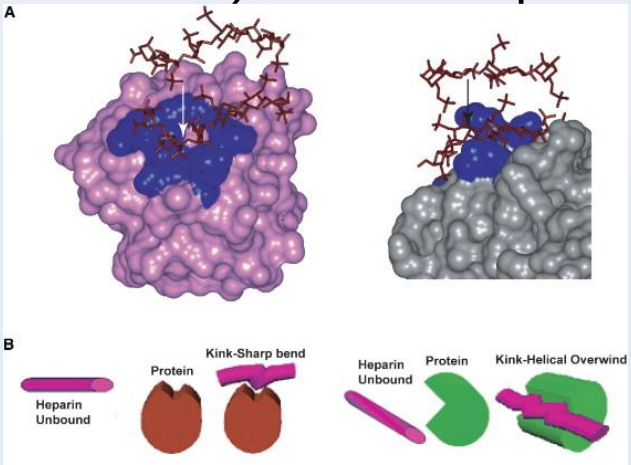


Recognition of ATIII and FGF2 by HA as a Result of Different Kinks

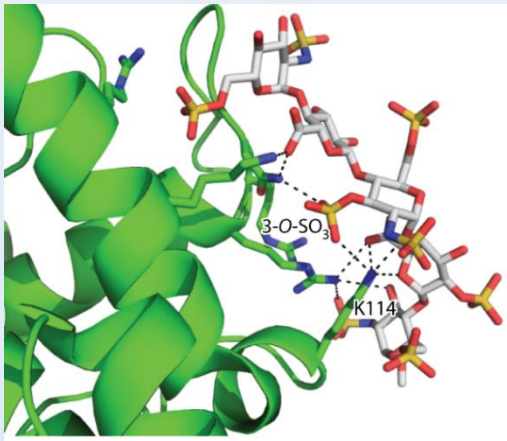
- The different protein binding sites in AT III and FGF2 induce different kinks in HA that promote recognition and binding

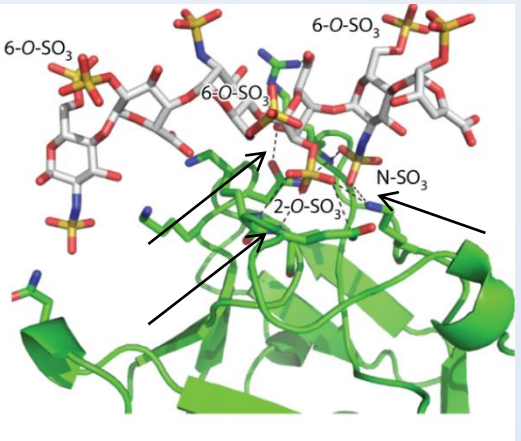
- HA-AT III binding stabilized by ionic interaction between Lys114 and the kink containing 3-*O*-sulfate

- For HA-FGF2 the trisaccharide kink contained the critical NS and 6-O sulfate groups on the hexosamine and 2-O sulfate groups on IdoA (IdoA2S-GlcNS6S) that are important for binding and recognition



HA–AT III HA h-FGF2



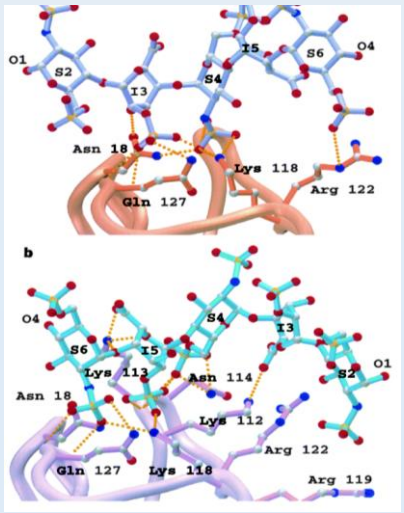


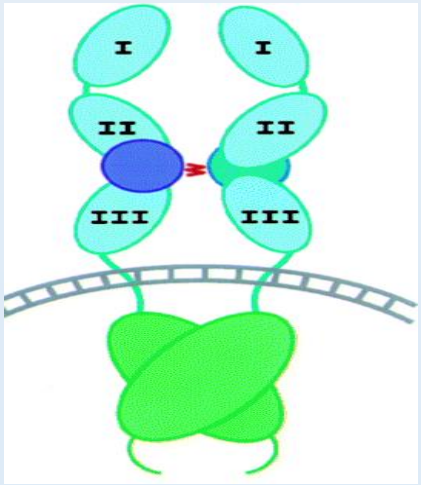
HA pentasaccharide–AT III HA hexasaccharide-FGF2

Thus a GAG even though may have complementary surface to these proteins but if it does not induce a recognition kink will not bind

Heparin Modulated Activation of Fibroblast Growth factor (FGFR)

- Fibroblast growth factor (FGF) family (12 known, e.g. FGF-1 and FGF-2) are involved in cell signaling during growth and development by binding to and activating FGF tyrosine kinase receptors (FGFRs).
- FGFR stimulated biological response require HA binding to FGF
- How does HA recognize the different FGFs ??????
- Crystallized as dimer which is bridged by one molecule of heparin without any protein-protein contact. Each protomer binds sulfate groups of 5-6 monosaccharides on the opposite sides of the heparin





Specificity of HA Binding to Different FGFs

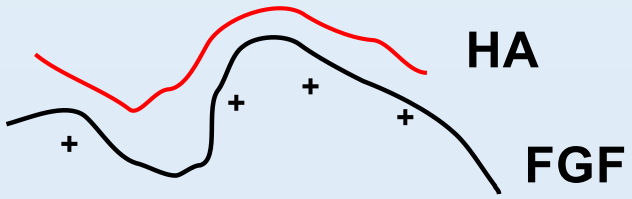
- Structures of HA-FGFs provide a molecular understanding for the specificity of HA oligosaccharide sequence to different FGFs

- The distribution of the basic residues within the HA binding loops imposes unique constraints on the orientation and spatial arrangement of the sulfate groups.

- The narrow binding pocket formed by the HA binding loops on each FGF imposes structural constraints on the HA oligosaccharide that induce a kink for optimal surface contact with the protein

- Together, these structural constraints determine the specificity of oligosaccharide sequence to bind to different FGFs.

- Requires the right GAG sequence to induce FGF recognition kink



Calcium-Dependent GAG-Protein Interaction

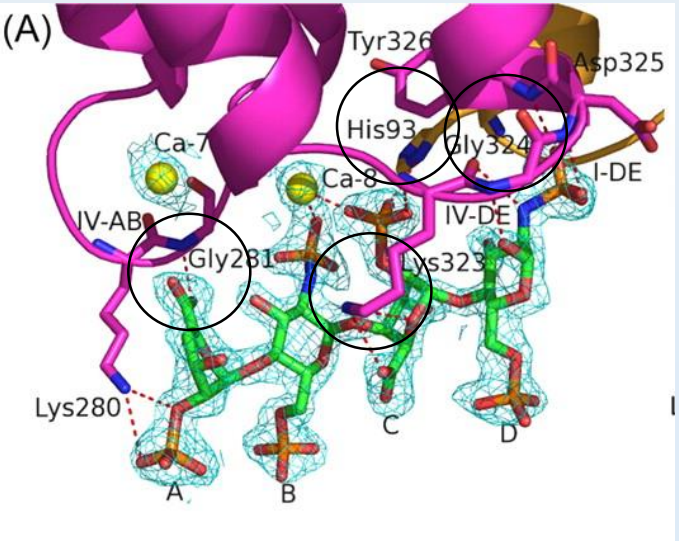
- Some GAG-protein interactions are mediated by Ca^{2+}
- Example is the family of Ca^{2+}-dependent membrane-binding protein, Annexin which are implicated in membrane trafficking, cell signaling, apoptosis, hemostasis, fibrinolytic pathway.
- Other examples of calcium-dependent HA-BPs include L-selectin, P-selectin, and type V collagen
- Annexin represents a notable exception to the general rule of four to seven basic residues per HA-binding site.

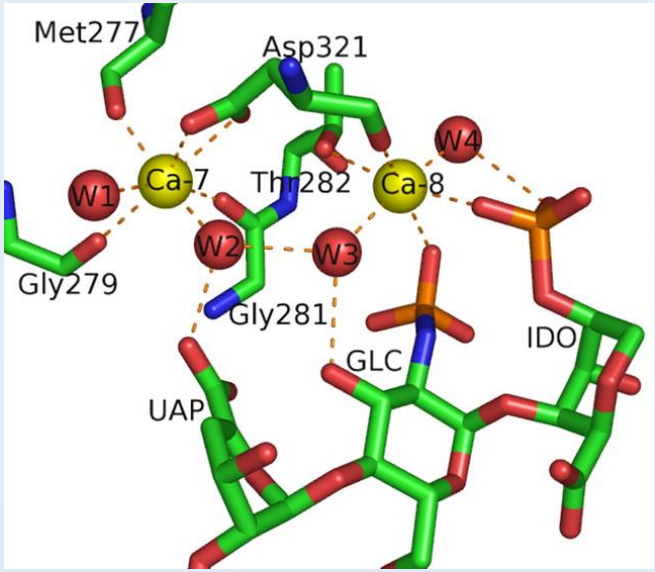
Calcium-Dependent HA-Annexin Interactions

- The crystal structures of Annexin A2 and heparin-derived oligosaccharides show only three basic residues participate in HA binding.

- Two calcium ions occupy two negatively charged pockets in close proximity to the two basic residues and interact either directly or indirectly with the sulfate and carboxylate groups of the oligosaccharide

- Carboxylate groups of the iduronate residue, and the N-sulfate and 6-O-sulfate of GlcNS6S have been shown to be essential for this binding.
-





Size of GAG Binding Site and Number of Basic Residues

- Size of the GAG-binding site does not serve as a good indicator of the binding affinity.
- High-affinity interactions do happen at small binding sites (e.g., FGF2, HGF, annexin A2) or large HA-binding sites, e.g., platelet factor 4 (PF4)
- Conversely, some GAGBPs with low affinity for HA have either small (e.g., thrombin) or large (e.g., FGFR1) HA-binding sites.
- Density of basic residues in the HA-binding site does not necessarily predict affinity as well.
- The HA-binding sites of thrombin and AT have surface areas of ~300 $Å^2$. Thrombin has a higher density of basic residues, but the affinity of AT for HA is three orders of magnitude greater than the affinity of thrombin for HA.

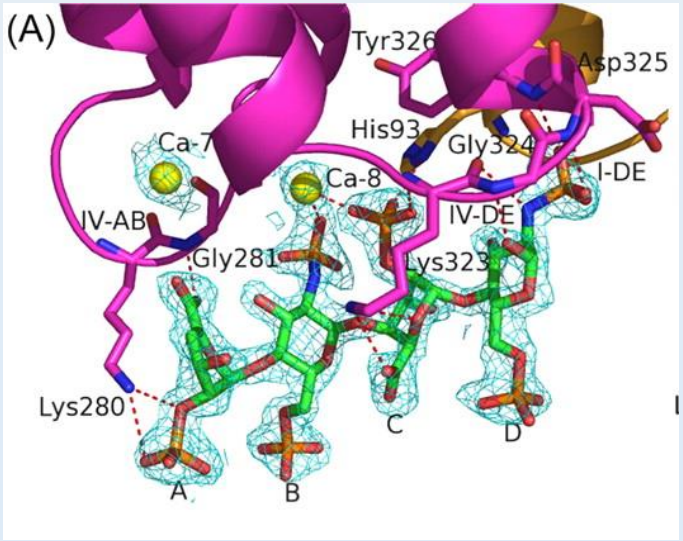
Secondary Structural Elements in GAG-Protein Binding

Generally, HA-binding sites consist of residues contributed by two to four spatially separated structural elements

Loops appear to be the most common secondary structural element, which may reflect their flexibility.

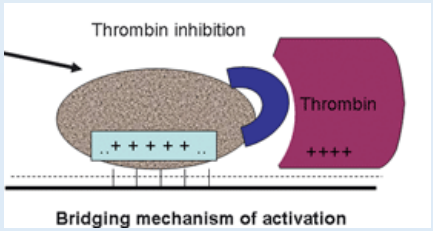
α-Helices and β-strands occur less frequently than loops, but dominant in some HA-binding (e.g., APLP-1, thrombospondin).

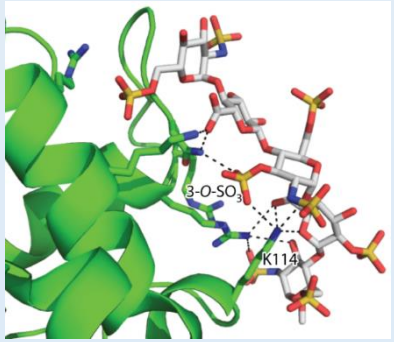
Specificity for HA does not seem to depend on particular secondary structural elements.



HA Catalysis of Thrombin Inhibition by AT

- Heparin forms a ternary complex with thrombin and AT-III, accelerating > 2000-fold the rate of thrombin inhibition by AT-III to prevent coagulation
- AT-III recognizes a specific pentasacharride sequence motif within heparin that allows heparin to bind with kd of 0.6 nM.
- Interaction between 3-O sulfate group and Lys114 as a result of the induced kink is very important for recognition between HA and AT-III.
- Pentasacharride binding and subsequent conformational change activates AT-III inhibition of thrombin
- Biochemical studies with other mimetics however, implicate a longer chain of at least 16-18 in addition to the pentasacharride sequence.



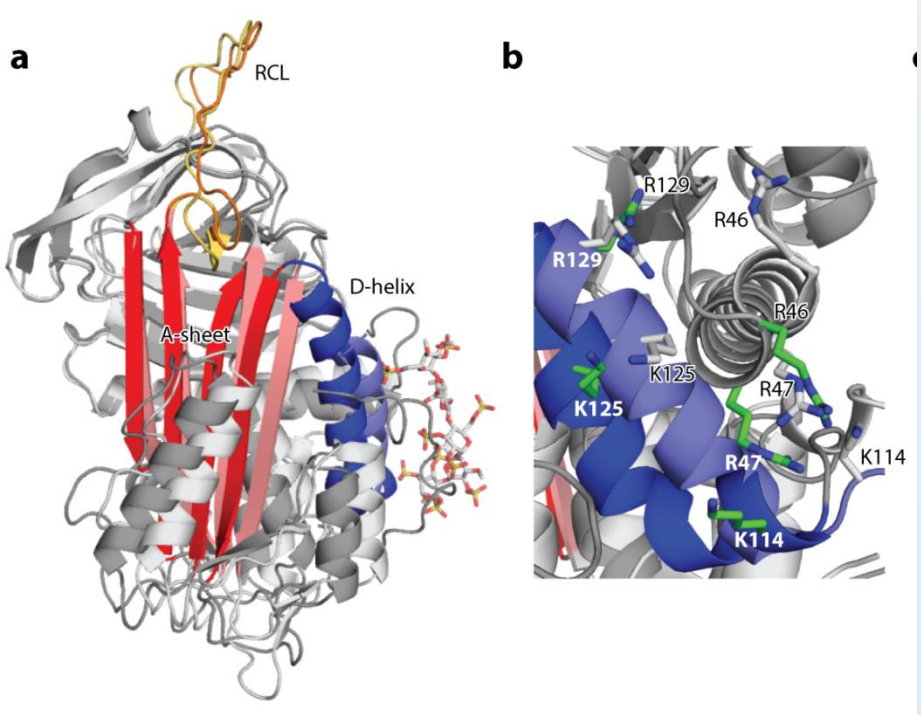


HA pentasaccharide–AT III

Conformational Changes in AT-III with Heparin Binding

- Key structural elements site undergo significant rearrangement with HA-binding

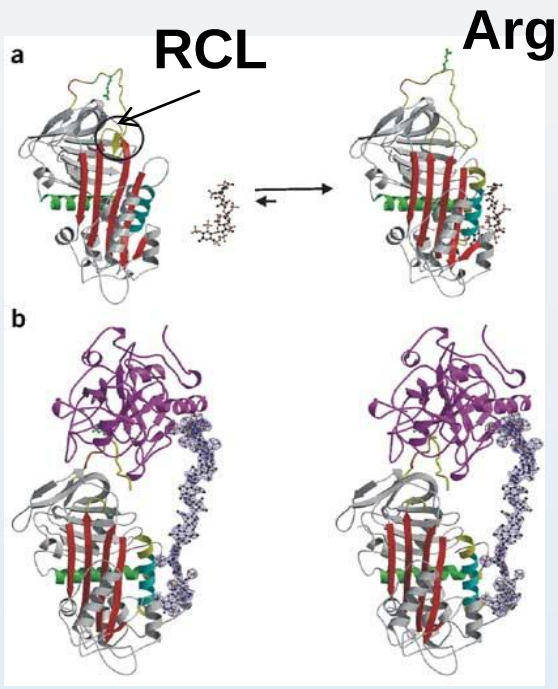
- There are also significant rearrangement of AT-binding residues (green) following HA binding, allowing the side chains of the HA-binding amino acid residues (green) to move by 5 to 17 Å to accommodate HA

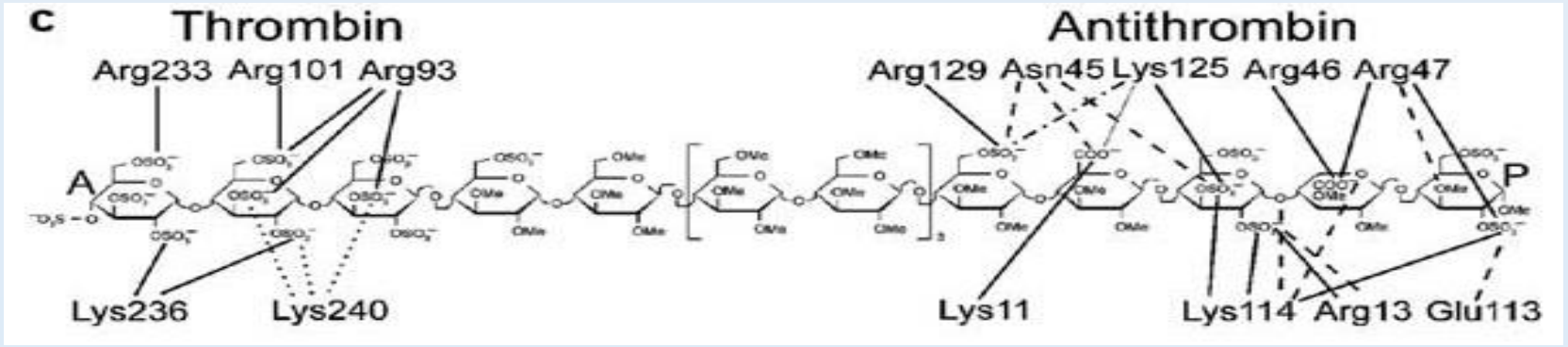


Overlay of native AT-III (*light gray*) and pentasaccharide-bound AT-III (*dark gray*). The important structural elements are the D-helix in purple (native) and blue (bound), the A-sheet in pink (native) and red (bound), and the reactive central loop (RCL) in yellow (native) and orange (bound).

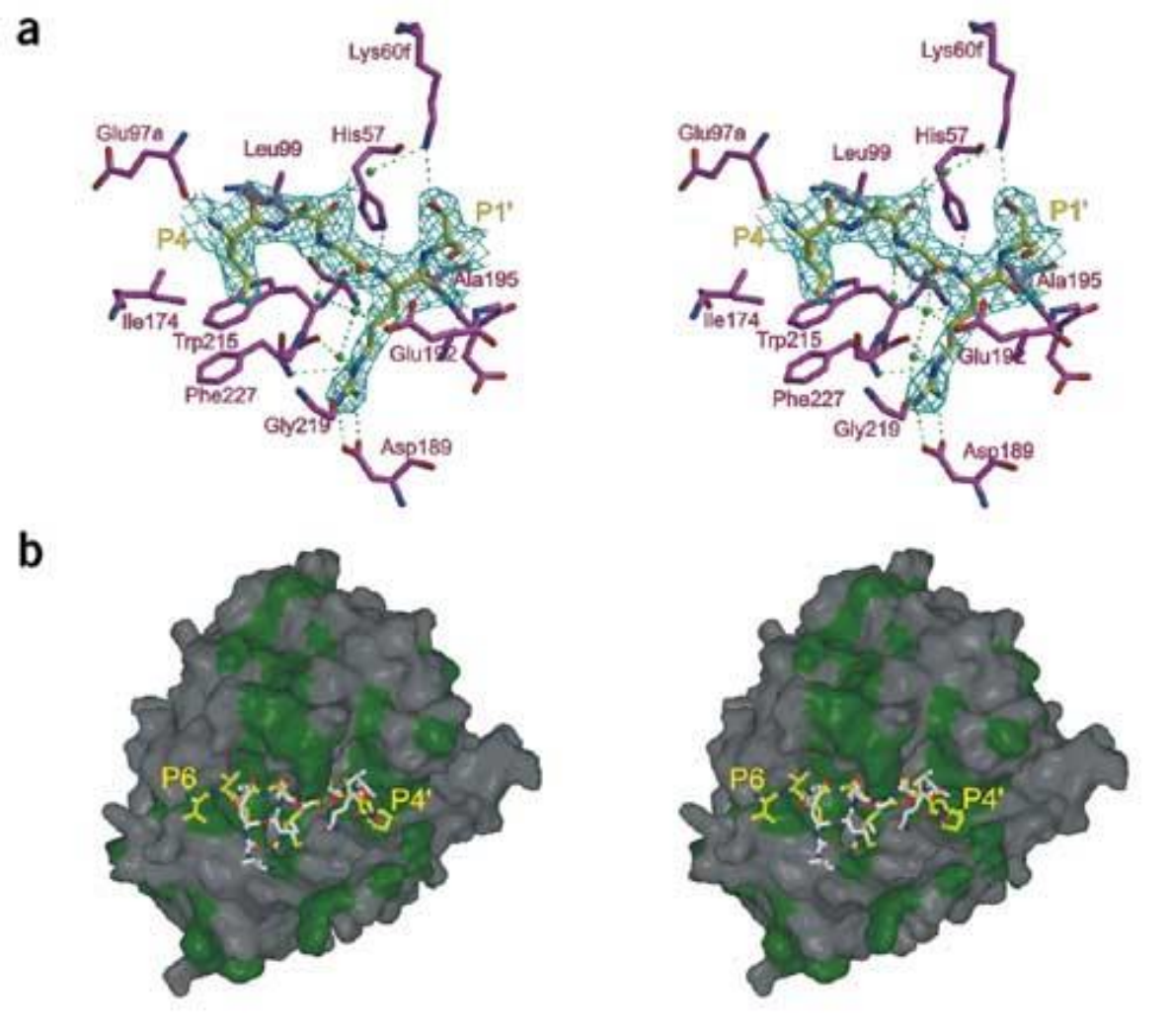
Heparin Catalysis of Thrombin Inhibition by Antithrombin

- The induced conformational change of ATIII with HA binding leads to expulsion of the hinge region of the reactive center loop (RCL)
- This increases the flexibility of the RCL and liberates the P1 Arg
- Thrombin then binds to antithrombin region, and makes several exosite interactions.





Specific Interactions between Thrombin and AT-III



The RCL of antithrombin (yellow) binds within the active site of thrombin (magenta) via extensive hydrogen bonding (green dashed lines) and by hydrophobic interactions.

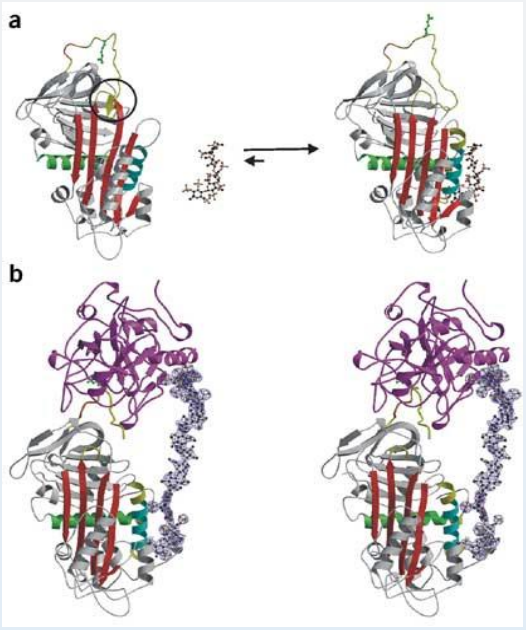
Heparin as an Allosteric Regulator

- AT is unique among serpins in that its RCL adopts a more constrained conformation, which makes it less accessible for certain proteases (Factors Xa and IXa) and is responsible for the poor inhibitory activity of native AT

- Only when there are conformational changes through the interaction with a specific heparin pentasaccharide sequence that exposes the RCL does AT-III become an efficient inhibitor of factor Xa or thrombin

- In contrast, other serpins have exposed RCL that acts as bait to achieve inhibition of target serine protease. Thus

- HA can be viewed as an allosteric modulator of AT-III



Hyaluronic acid Interactions

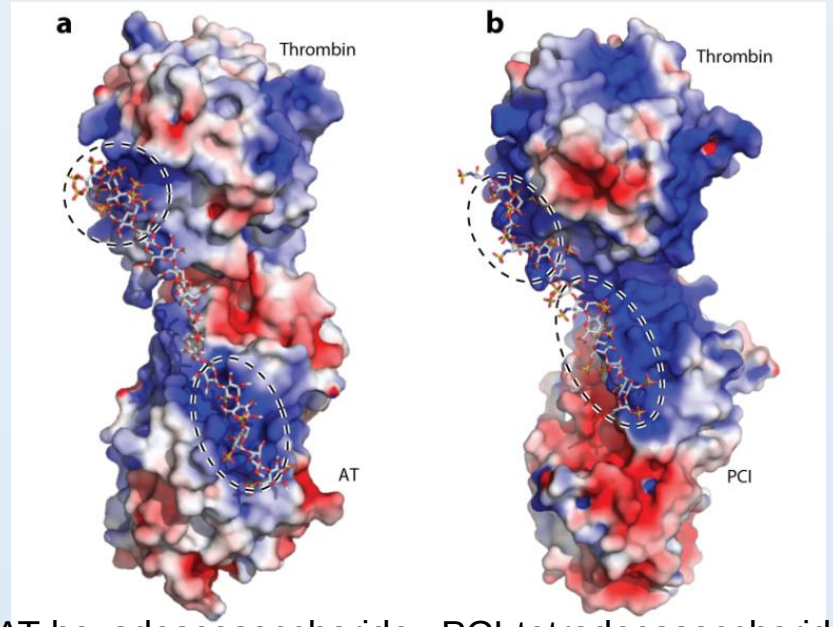
- Hyaluronic binding proteins are known as hyaladherins, and the best-studied HA-protein interaction is that with CD44, which plays a key role in cancer.

- NMR studies have implicated that the overall conformation of HA in solution is flexible and this flexibility of HA is implicated to assemble different hyaladherins where each protein captures distinct conformations of the different oligosaccharide binding motifs in HA.

- Availability of these crystal structures would provide better structural insights into the specificity of HA-protein interactions.

Protein C Inhibitor and Antithrombin – Chain Length Recognition

- Protein C inhibitor (PCI) is another serpin that acts on thrombin
- The HA-binding site of PCI is situated much closer to the PCI/thrombin interface than that found in AT/thrombin complexes, enabling the formation of a HA-binding site composed of domains from both PCI and thrombin
- The close proximity between the two HA-binding sites enables shorter oligosaccharides (≥14) to act as a scaffold, in contrast to the 18 oligosaccharide required for bridging AT and thrombin



AT-hexadescasaccharide PCI-tetradecasaccharide.