Glycosaminoglycan–Protein Interactions: Molecular Modeling and Simulation Methods

Philip D. Mosier
Senior Scientist and Co-Director, HSPG PEG Computational Chemistry and Biology Core
Department of Medicinal Chemistry
Virginia Commonwealth University
7 January 2013
Biological Roles of GAGs

- **Heparin / HS Dermatan sulfate**
  - Antithrombin; HCII anticoagulation
  - Annexin V anticoagulation
  - C1 inhibitor complement activation
  - gB, gC, gD; gp120 viral invasion
  - Growth Factors: cell proliferation, etc.
  - Cytokines: PF-4; IL-8; Inflammation

Cytokines: PF-4; IL-8; Inflammation

- Growth Factors: cell proliferation, etc.
- Biological Functions:
  - Antithrombin; HCII anticoagulation
  - Annexin V anticoagulation
  - C1 inhibitor complement activation
  - gB, gC, gD; gp120 viral invasion
  - Growth Factors: cell proliferation, etc.
  - Cytokines: PF-4; IL-8; Inflammation

**Biological Roles of GAGs**
Glycosaminoglycans (GAGs)

- Linear (i.e. unbranched) chains consisting of repeating disaccharide units variously substituted with acetyl and sulfate groups.

- Free in solution or covalently linked to proteins (proteoglycans).

- Are phenomenally diverse with regard to saccharide ring conformation and substitution pattern.

- Unlike other biopolymers, there is no ‘genetic code’ that determines the precise GAG.
Monosaccharides

- Aldehydes and react with alcohols to give hemiacetals

\[
\begin{align*}
R\text{CHO} + R'\text{OH} & \rightleftharpoons R\text{OH} + R'\text{OR'}
\end{align*}
\]

Glucose: an aldose

\[
\begin{align*}
\text{FISCHER Projection} & \quad \text{HAWORTH Projections}
\end{align*}
\]

- Anomer: a stereoisomer that differs in configuration only at the anomeric carbon
- What is an epimer?
Pyranose Ring Conformations

- Based on the Cremer–Pople ring puckering indices $Q$, $\theta$ and $\psi$.

- The low-energy chair (C), boat (B) and skew-boat (S) conformations are designated by two numbers corresponding to ring atoms that lie above and below a plane defined by the remaining four ring atoms.

- The atom above the plane is superscripted; the one below is subscripted.

- The energetically most favorable conformations for GAGs are $^1C_4$, $^4C_1$ and $^2S_0$.

Pyranose Ring Conformations

\[ \alpha-\text{D-glucose} \]

\[ \begin{align*}
\text{H} & \quad \text{CH}_2\text{OH} \\
\text{HO} & \quad \text{H} \\
\text{HO} & \quad \text{HO} \\
\text{HO} & \quad \text{HO} \\
\text{HO} & \quad \text{H}
\end{align*} \]

\[ \xrightarrow{\text{C}_1} \]

\[ \begin{align*}
\text{H} & \quad \text{CH}_2\text{OH} \\
\text{HO} & \quad \text{H} \\
\text{HO} & \quad \text{HO} \\
\text{HO} & \quad \text{HO} \\
\text{HO} & \quad \text{H}
\end{align*} \]

\[ \text{Current Opinion in Microbiology} \]

\[ \alpha-\text{L-iduronate} \]

\[ \begin{align*}
\text{OOOC} & \quad \text{H} \\
\text{HO} & \quad \text{OH} \\
\text{HO} & \quad \text{OH} \\
\text{HO} & \quad \text{OH} \\
\text{HO} & \quad \text{H}
\end{align*} \]

\[ \xrightarrow{\text{C}_4} \]

\[ \begin{align*}
\text{H} & \quad \text{CH}_2\text{OH} \\
\text{HO} & \quad \text{H} \\
\text{HO} & \quad \text{HO} \\
\text{HO} & \quad \text{HO} \\
\text{HO} & \quad \text{H}
\end{align*} \]

\[ \text{2S}_0 \]

GAGs are heterogeneous, polydisperse, acidic polysaccharides

- **Hyaluronate**
  - d-Glucuronate
  - N-Acetyl-d-glucosamine

- **Dermatan sulfate**
  - l-Iduronate
  - N-Acetyl-d-glucosamine-4-sulfate

- **Chondroitin-4-sulfate**
  - d-Glucuronate
  - N-Acetyl-d-galactosamine-4-sulfate

- **Keratan sulfate**
  - d-Galactose
  - N-Acetyl-d-glucosamine-6-sulfate

- **Chondroitin-6-sulfate**
  - d-Glucuronate
  - N-Acetyl-d-galactosamine-6-sulfate

- **Heparin**
  - d-Glucuronate-2-sulfate
  - N-Sulfo-d-glucosamine-6-sulfate
Glycosidic Linkage

- The glycosidic linkage consists of two rotatable bonds designated $\phi$ and $\psi$.

- Consistent and regular values of $\phi$ and $\psi$ give rise to helical structure for some GAGs.

- This can be used as a constraint to limit the number of rotatable bonds during the docking procedure.

DS Helical Allomorphs

Fig. 3. Mutually perpendicular views of $2_1$, $3_2$, and $8_3$ dermatan sulfate polyanions. All the allomorphs are characterized by O3I-O5N hydrogen bonds across the $\beta(1 \rightarrow 4)$ linkage.

GAG Binding Sites

- A typical GAG binding site will have the following characteristics:
  - Surface-exposed
  - Shallow
  - Highly positively charged (usually with many positively-charged amino acid residues that interact with the negatively-charged acidic GAG carboxylate and sulfate groups).

Thrombin (PDB ID = 1TB6)
Now we know something about the protein and ligand...so what drives the interaction between the two?
The affinity of a ligand for a receptor is determined by the free energy of binding $\Delta G$:

$$\Delta G = \Delta H - T\Delta S = -RT\ln K_{eq}$$
Thermodynamics: Enthalpy ($\Delta H$)

- The change in enthalpy is determined by the formation and breakage of inter- and intramolecular chemical bonds. For the most part, we need only be concerned with the weak non-covalent interactions. However, there are exceptions (affinity labels and irreversible “suicide” inhibitors, for example).

- ‘Bonds’ that are formed when the ligand binds to the receptor will stabilize (and thus favor) the ligand–receptor complex; ‘bonds’ that are broken when the ligand binds will disfavor the formation of the complex.
Thermodynamics: Entropy (\(\Delta S\))

Entropy represents the amount of uncertainty or degrees of freedom in a closed thermodynamic system. In terms of biological molecular interactions, there are two main factors that account for this:

- **Loss of translational and rotational degrees of freedom:** The position of the ligand, as well as freely-rotating single bonds within the ligand, are more or less fixed in place by the receptor upon binding, resulting in a decrease in entropy that *disfavors* the formation of the receptor–ligand complex.

- **Hydrophobic effect:** Organized water molecules surrounding the hydrophobic portions of the ligand and possibly inside the binding site are displaced during the docking event, resulting in an increase in entropy that *favors* the formation of the receptor–ligand complex.
So how do we translate these ideas into something practically useful?
The very basic idea is to treat molecules as physical entities that are governed by the conventional (non-quantum) “Newtonian” Laws of Physics.

\[ E_{\text{TOTAL}} = \sum E_{\text{stretch}} + \sum E_{\text{bend}} + \sum E_{\text{torsion}} + \sum E_{\text{o-o-p}} + \sum E_{\text{vdW}} + \sum E_{\text{Coulombic}} \]

The total internal energy of a molecule is a sum of various energy terms representing various stresses and strains on it.
Molecular Mechanics: $E_{\text{stretch}}$

$$E_{\text{stretch}} = \sum_{i=1}^{\text{All bonds}} k_i (\Delta r_i)^2 / 2$$

(Hooke’s Law)
Molecular Mechanics: $E_{\text{bend}}$

$$E_{\text{bend}} = \sum_{j=1}^{\text{All angles}} k_j (\Delta \theta_j)^2 / 2$$
Molecular Mechanics: $E_{\text{torsion}}$

$$E_{\text{torsion}} = \sum_{k=1}^{\text{All torsions}} V_k \left[ 1 + S_k \cos (|n_k| - \phi_k) \right] / 2$$
Molecular Mechanics: $E_{o-o-p}$

$$E_{o-o-p} = \sum_{k=1}^{\text{All out-of-planes}} k_1 (\Delta d_i)^2 / 2$$
Molecular Mechanics: $E_{vdW}$

No attraction + and − centered

Electron cloud gets pushed. Dipole forms

Electron cloud is closer so has more influence

Atoms attract each other

\[ E_{vdW} = \sum_{i=1}^{\text{All nonbonded}} \sum_{j>1}^{\phi_{ij}} \left[ \frac{1.0}{a_{ij}^{12}} - \frac{2.0}{a_{ij}^{6}} \right] \]
Molecular Mechanics: $E_{\text{Coulombic}}$

Where do the charges on atoms come from?

Charles Augustin de Coulomb
Molecular Mechanics: Atomic Charge

- Start with electronegativity…

---

**Periodic Table of the Elements**

Electronegativity

---

***Elements > 104 exist only for very short half-lives and the data is unknown.***

---

Lanthanides

Actinides
Molecular Mechanics: Atomic Charge

- calculate bonding (and atomic polarizability)…
partial charge on each atom is estimated based on this polarization:

In benzene every C has the same electronegativity; thus, all Cs have same partial charge. The Hs have same charge with opposite sign. The C partial charge reaches a maximum at the meta position relative to other Cs due to the e⁻-withdrawing methoxycarbonyl group. Amines are e⁻-releasing and C charge reaches a maximum (relative to the other Cs) at ortho and para.
We have the basic equations... Now we can bring in a computer to do the calculations.
Molecular Mechanics: Atomic Charge

• Together these equations comprise a “force field” that describes how to calculate the potential energy of a molecule.
• Bond lengths, bond angles and atomic radii are taken from high-resolution X-ray structures of small molecules and biomacromolecules.
• The force constants for Hooke’s law (stretches and bends) are taken from IR and similar experiments.
Molecular Mechanics: Minimization

• Energy Minimization ⇔ Structure Optimization
  – Simultaneous Process
  – *By definition* the most optimal structure is the one of lowest energy.
  – In vacuum, at 0 K.

• Although there are a variety of permutations on the algorithms – mostly to increase speed – the basic idea is: if the energy is decreasing, then the structure is “improving”.
  – However, it is not so simple...
Molecular Mechanics: Minimization

![Graph showing energy vs. position, with minima and energy landscapes for different conformations: planar, boat, and chair.]
Molecular Mechanics: Local Minima

• Exhaustive Search
  – Define structures and calculate energies for “all” conformers.
  – Structure with lowest energy is at Global minimum.
  – The process is generally varying torsion angles for “rotatable” bonds. At a 15° increment, number of structures = 24^n, where n is number of these bonds.

• Molecular Dynamics
  – Simply, add “heat” and time to molecular mechanics and break out of local minima.
  – Use with larger molecules.
Molecular Mechanics: Rotatable Bonds

- 7 to 15 rotatable bonds – depending on how you count them.
- Becomes an issue when docking larger oligosaccharides.
- Can use the “rigid backbone” approximation to partially address this.
Molecular Dynamics

START

Initialize Positions and Velocities

Calculate forces for all molecules using Potential

Apply Thermostat and Volume Changes

Update Position and Velocities

Analyze the Data

Temperature (K)

Given Temperature
Calculated Temperature

MD steps
Docking

Involves finding an optimal complementary ‘fit’ between a ligand molecule and a suitable binding site in a receptor, and typically involves three operations:

- Generate plausible candidate ligand conformations

In general, the conformation(s) that the ligand adopts when bound to the receptor (i.e. the *binding mode*) is not known. The docking program must therefore generate a number of possible conformations to test.

- Place the ligand into the binding site

The docking program must be able to place the ligand into the binding site such that the distances and angles between complementary functional groups are chemically meaningful and that the atoms are not clashing with one another.

- Assign a score or fitness value to the docked conformation.

Ligand docking is treated computationally as a mathematical optimization problem. Thus, the docked solutions (or *poses*) that are generated must be evaluated by a *fitness* or *scoring function* to determine the quality of each pose.
GAG Virtual Libraries

- GAG virtual libraries capture the phenomenal diversity of GAGs by enumerating the sulfation and acetylation patterns of the GAG sequences, as well as the GAG conformations at the monosaccharide level.

- Example: H/HS hexasaccharide library:
  - 1 hexasaccharide sequence = 3 disaccharide units
  - $36 \times 36 \times 36 = 36^3 = 46,656$ virtual H/HS sequences
Virtual Screening

- Virtual screening (VS) provides a means to dock many GAG sequences to a given protein and select the ones that bind with high affinity and selectivity.

- For large libraries, a dual-filter strategy is used:

  1. Build Virtual Library of H/HS sequences (Use ‘Average Backbone’ Model)
  2. 1st GOLD Docking (10,000 iterations; one run per sequence)
  3. Affinity Filter (GOLD score in top ~1%)
     - NO
     - YES
  4. 2nd GOLD Docking (100,000 iterations; three runs per sequence)
  5. Specificity Filter (RMSD < 2.5 Å)
     - NO
     - YES
  6. ‘High Affinity & High Specificity’ Sequence(s)
  7. Discard

Modeling GAG–Protein Interactions

- Molecular modeling provides us with the tools to model **ANY** GAG–protein interaction.

- Powerful and easy and to do, but filled with caveats:
  - Flexibility of the GAG and/or binding site amino acid residues
  - Ionization state of charged groups
  - Involvement of water
T–AT Binding Site Analysis

- Identification of highly relevant water molecules
  - Displacement of these water molecules greatly enhances the affinity of the GAG for its binding site.

Case Study: HS–AT Interaction

- The H/HS pentasaccharide sequence (i.e. ‘DEFGH’) binds with high specificity to AT.

- Screening of a combinatorial virtual library of 6,859 heparin hexasaccharides using a dual-filter strategy, in which antithrombin affinity was the first filter and self-consistency of docking was the second, resulted in only 10 sequences.

- The results explain the biochemical observations with single substitution and ring truncation variants of the pentasaccharide.
Case Study: HS–AT Interaction

Case Study: HS–AT Interaction

Case Study: HS–AT Interaction

Conclusions

- Molecular modeling techniques provide powerful and general methods to address GAG–protein binding.

- These tools can be used to identify potential high-affinity, high specificity GAG sequences for ANY potential GAG binding site.
Acknowledgments

• VCU
  • Dr. Umesh Desai
  • Dr. Glen Kellogg

• All of you!

• NIH P01 HL107152