Fragment based modeling of protein-GAG complexes
Isaure Chauvot de Beauchêne, Sergey A Samsonov, Martin Zacharias

To cite this version:
Isaure Chauvot de Beauchêne, Sergey A Samsonov, Martin Zacharias. Fragment based modeling of protein-GAG complexes. GGMM 2017 - 20e congrès du Groupe de Graphisme et Modelisation Moleculaire, May 2017, Reims, France. pp.1. hal-01927283

HAL Id: hal-01927283
https://hal.archives-ouvertes.fr/hal-01927283
Submitted on 19 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
**INTRODUCTION**

Glycosaminoglycans (GAGs) are linear anionic periodic polysaccharides. They bind to their protein targets in the extracellular matrix, and so participate in many cell-signaling processes. As such, they are very promising targets for the design of novel functional biomaterials for regenerative medicine.

Experimental methods (X-ray, NMR) to solve the structure of GAG-protein complexes are impaired by GAG flexibility and high negative charge density that makes it difficult to obtain pure homogeneous samples. Therefore, it often requires computational docking methods. But GAGs are also too flexible for (semi-)rigid docking, and too large for fully flexible docking.

Fragment-based docking permits to deal with the high number of degrees of freedom of the ligand. However, most current methods are limited to small ligands and well delineated binding sites. We have recently extended such approach to flexible linear polymers docked on the whole protein surface. By rigid docking of fragment libraries and exhaustive assembly of the compatible poses, we could model protein-bound GAG, another highly flexible anionic linear polymer. We present here a new variation on the theme, using flexible docking of fragments by AutoDock, on a coarsely known binding site. We successfully applied it on protein–GAG complexes.

**METHODS**

**Inputs**

We chose 3mer as fragment length in order to concentrate binding specificity (enough contacts per frag.) and a low number of flexible bonds. GAG having a periodicity in [A-B]n, we dock two types of trimers : A-B-A and B-A-B.

**Flexible Docking**

- **AutoDock** (Morris et al. 2006)
- Charges : GLYCAM07 and literature (for SO4)
- All atoms representation, implicit solvent
- Grid centered on the COM of the bound ligand
- 1000 poses for each fragment type
- Rigid receptor (bound), fully flexible ligand

**Combinatorial assembly**

Pose compatibility depends on their overlapping RMSD. For time saving, only some atoms are considered (5/5/0/N). The cutoff is chosen so as to get at least 1000 chains for 5mers or 10.000 for 6/7mers respectively, up to a maximum of 3 Å.

**Refinement**

The assembled chains are clustered (0.5 Å) and the overlapping parts of the compatible poses are averaged. This results in incorrect monomer geometries, which is corrected by replacing each monomer by the closest one in a structural library, extracted from all initial docking poses.

**RESULTS**

Glycosaminoglycans (GAGs) are linear anionic periodic polysaccharides. They bind to their protein targets in the extracellular matrix, and so participate in many cell-signaling processes. As such, they are very promising targets for the design of novel functional biomaterials for regenerative medicine.

Experimental methods (X-ray, NMR) to solve the structure of GAG-protein complexes are impaired by GAG flexibility and high negative charge density that makes it difficult to obtain pure homogeneous samples. Therefore, it often requires computational docking methods. But GAGs are also too flexible for (semi-)rigid docking, and too large for fully flexible docking.

Fragment-based docking permits to deal with the high number of degrees of freedom of the ligand. However, most current methods are limited to small ligands and well delineated binding sites. We have recently extended such approach to flexible linear polymers docked on the whole protein surface. By rigid docking of fragment libraries and exhaustive assembly of the compatible poses, we could model protein-bound GAG, another highly flexible anionic linear polymer. We present here a new variation on the theme, using flexible docking of fragments by AutoDock, on a coarsely known binding site. We successfully applied it on protein–GAG complexes.

**Protein-GAG benchmark**

Here we consider the binding site as coarsely known, covering half of the surface.

**Hits enrichment by fragment assembly**

Filtering the poses by fragment assembly propensity (counting the possible chains they belong to) enriches the pool in native poses (<5Å) more effectively than AD scoring. However, this filtering did not permit to precise the binding site prediction by counting the number of contacts per amino-acid.

**Chain assembly for GAGs with different length (dp)**

For each GAG we tested to assemble all sub-chains with dp6 to 7.

We could sample near-native (<5Å) dp7 chains for 12/13 complexes, and quasi-native (<3Å) for 6/13 complexes. Those ratio diminished to 6/11 for dp6, and 2/11 for dp7 chains.

**CONCLUSION**

We developed a new method to model protein-bound GAGs with high accuracy, on a coarsely known binding site on the protein. The approach proved effective to select correct fragment poses at the surface of the protein, and to model GAG up to a degree of polymerisation dp7 with an accuracy of 3.0 Å RMSD, with more than 10% correct models.

**Perspectives**

- We considered the binding region as coarsely known. We will extend the method to cases were the binding region could not be predicted with sufficient confidence. We could either repeat the AD docking with a large number of grids, or use another docking engine for exhaustive sampling.
- For this second option, we will implement a saccharide coarse-grained representation in the ATTRACT docking program.
- We considered the protein as rigid. We intend to model also the protein flexibility, by representing parts of the surface in several conformations, that will have to match with neighbor conformations in the chain assembly.
- Finally, we will apply the method to the docking of intrinsically disordered proteins (IDPs).

**REFERENCES**


[N. Samsomov and M.T. Pisabarro (2016) Glycobio]

The authors thanks the Deutsche Forschungsgemeinschaft (DFG) for funding part of this work, and SjefJ. De Vries for useful discussions.