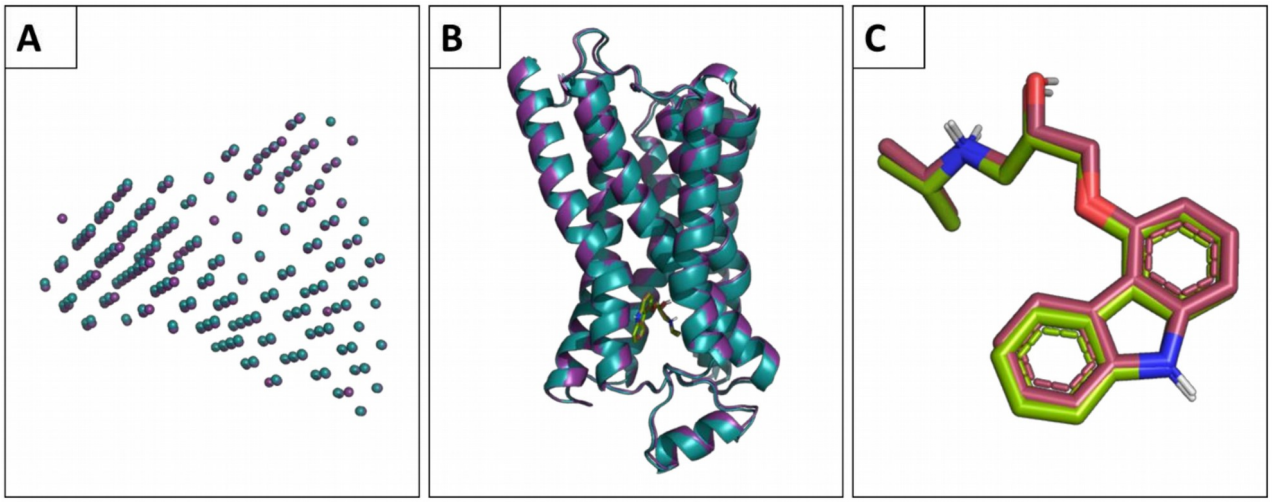
Master internship proposal

A Branch & Bound approach to align protein cavities



Alignment of two cavities represented as 3D point clouds (A). Alignment of the respective proteins (B) and co-crystallized ligands (C). Images taken from [1].

Advisors

- *Franck Hétroy-Wheeler* (hetroywheeler@unistra.fr), Engineering science, Computer science and Imaging Laboratory (ICube), Computer Graphics & Geometry team
- *Didier Rognan* (rognan@unistra.fr), Laboratory for Therapeutic Innovation (LIT), Structural Chemogenomics group

Keywords

3D point cloud registration; branch and bound; protein-ligand binding sites similarity.

Context

Comparing protein-ligand binding sites is of great importance in drug discovery. In particular, characterizing their local 3D similarities is a good starting point for drug design and repurposing. To this end, Desaphy *et al.* [2] have proposed to define a protein binding pocket as a negative image of the cavity in which a set of physicochemical 3D points mimic an ideal ligand (see Figure A). In order to assess the similarity between two binding sites, the goal is then to rigidly align the corresponding cavity point clouds and to compute some alignment score.

Many approaches can be applied for this purpose [3]. A first work has recently been carried in the Structural Chemogenomics group to assess if the combined use of the traditional RANSAC and ICP algorithms could be applied for protein cavity alignment [1]. Though results are promising, this approach does not guarantee that the best possible alignment is found since RANSAC is a stochastic approach and ICP is known to converge to a local, not necessarily global, optimum.

To overcome this issue, we will focus during this internship on a deterministic approach known to reach a global alignment optimum: the Branch and Bound approach [4,5].

Objectives

During this project, the student will first have to test Branch and Bound approaches [4,5] to rigidly align 3D point clouds of protein cavities. A detailed study on the performance of these approaches will have to be carried. This study should include quantitative results w.r.t. various alignment quality scores, a parameter sensitivity analysis, and an evaluation of the computing time and memory footprint. Test data will be provided by the Laboratory for Therapeutic Innovation.

The student will then have to investigate various branching strategies (depth- or breadth-first search, best evaluation first, etc.) as well as various bound evaluations to speed up the computation, since a low computing time is critical for our application.

Finally, the approach will be extended to handle the physicochemical properties of the points. This could be done for example by changing the evaluation in order to take these properties into account, or by penalizing mismatched properties in the branching strategy. This work will be carried in collaboration with chemoinformatics experts from the Laboratory for Therapeutic Innovation.

Student profile

- Master student, preferably in computer science or applied mathematics
- Creative and highly motivated
- Solid programming skills, especially in C++ and Matlab
- Good knowledge of advanced algorithms, computer vision and 3D geometry is a plus
- Fluent English or French spoken, fluent written English

Bibliography

[1] K.I.M. Eguida. Development of a new method to align and compare protein cavities with point clouds modeling. *Master thesis, University of Strasbourg* (2019).
[2] J. Desaphy, K. Azdimousa, E. Kellenberger & D. Rognan. Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *Journal of Chemical Information and Modeling* **52** (2012).
[3] G.K.L. Tam, Z.Q. Cheng, Y.K. Lai, F.C. Langbein, Y. Liu, D. Marshall, R.R. Martin, X.F. Sun & P.L. Roisin. Registration of 3D point clouds and meshes: a survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics* **19** (2013).
[4] J. Yang, H. Li, D. Campbell & Y. Jia. Go-ICP: a globally optimal solution to 3D ICP point-set registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38** (2016).
https://github.com/yangjiaolong/Go-ICP
[5] N. Dym & S.Z. Kovalsky. Linearly converging quasi branch and bound algorithms for global rigid registration. *Proceedings of the IEEE International Conference on Computer Vision* (2019).
https://github.com/nadavdym/quasiBnB