**MSMS algorithms for building protein surfaces**

The area of computational biology has been growing exponentially for the last decade or maybe even more. Proteins are one of a few key classes of molecules crucial for any live organism, so studying them can be quite rewarding. The surface of a protein is especially critical for understanding its behaviour since it is what interacts with the outer world mostly and thus defines a protein’s functions. Moving closely to Mathematica, the bio-direction is a priority now as I understand it. It’s especially true because of the ongoing pandemic. I found that although Mathematica has a “Protein” entity and tools to work with it, it lacks such an important property as the surface of a protein. Nowadays surfaces of proteins are used [1-3] for studying protein-protein, protein-ligand and other interactions via advanced optimization techniques, particularly NNs. So I thought I could implement the first step on the way of bringing sophisticated protein analysis to Mathematica - implement a package with a few functions necessary for basic surface analysis. I am planning to write 3 main functions and a supporting one:

1) DownloadPDB[pdbID]. Takes ID of a protein in a Protein Data Bank and returns the whole file from the database in a single string.

2) ProtonatePDB[pdbStr]. Takes a string which must contain a full .pdb file. Such a string can be obtained by DownloadPDB or by Import[file.pdb, {“PDB”, “String”}]. Returns a string containing a full pdb file of a protonated protein. This function utilizes the Reduce program [4].

3) DrawProteinSAS[pdbStr, Options[]]. Takes a string which must contain a full .pdb file. Possible options are radius of a probe molecule and type of atom radiuses to use. Probe radius corresponds to a type of solvent one wants to draw a surface with. Given a solvent, a fair approximation of a proba radius would be the radius of a smallest atom of a solvent molecule which has a part of its surface exposed to the outer world. Default probeR=1.5 (A). 3 types of radiuses are supported: Atomic, Covalent, VanDerWaals (default). This function draws a solvent accessible surface [5] of a given protein (collection of atoms with known chemical types).

4) ConstructSESmesh[pdbStr, Options[]]. Takes a string which must contain a full .pdb file. Possible options are triangulation density (5 vert/A^2), probe molecule radius (1.5 (A)), verbose (T/\_F\_), path to put files of faces and vertices of the built mesh (default is the $TmpDir). This function builds a solvent excluded surface [5] of a given protein.

[1] Gainza, P., Sverrisson, F., Monti, F. *et al.* Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods* 17, 184–192 (2020).

doi.org/10.1038/s41592-019-0666-6

[2] Yoichi Murakami, Kenji Mizuguchi, Applying the Naïve Bayes classifier with kernel density estimation to the prediction of protein–protein interaction sites, *Bioinformatics*, Volume 26, Issue 15, 1 August 2010, Pages 1841–1848,

[doi.org/10.1093/bioinformatics/btq302](https://doi.org/10.1093/bioinformatics/btq302)

[3] Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins*. 2007;66(3):630-645.

doi:10.1002/prot.21248

[4] Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol*. 1999;285(4):1735-1747. doi:10.1006/jmbi.1998.2401

[5] Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*. 1996;38(3):305-320. doi:10.1002/(SICI)1097-0282(199603)38:3%3C305::AID-BIP4%3E3.0.CO;2-Y