Protein-protein interfaces as *de novo* drug design targets: insight from unsupervised machine learning

Bouchouireb, Z.[1,2*], and Bouvier, B.[2]

[*] Lead presenter

[1] zakaria.bouchouireb@u-picardie.fr

[2] *Laboratoire de Glycochimie, des Antimicrobiens et des Agroressources (LG2A), UMR-CNRS 7378, Université de Picardie Jules Verne, Amiens Cedex France*

Keywords: PPI, Machine learning, Drug design

Abstract:

Protein-protein interfaces are corner stone actors in many essential biological processes, including cell regulation and metabolic pathway signaling. They originate from a network of protein-protein interactions (PPI's) leading to the assembly of multiprotein complexes. Long forsaken, they have become an emerging class of molecular targets for designing therapeutic drugs. However, major challenges need to be addressed in order to identify PPI's druggability and design drugable compounds accordingly[1]. These challenges arise from the lack of a proper and accurate depiction of features that define PPI's chemical space. Therefore, rationalizing the structural and biochemical feature space underpinning PPI's chemical space is a key step to improve our understanding of disease and optimize accordingly the drug design process.

In this work, in order to gain more insight into the relative importance and classification power of the multiple features of PPI's, a 11k PPI's dataset spanning frequently occurring biological classes was derived from HippDB, SippDB and DippDB[2,3], and explored with unsupervised machine learning methodologies. A specific attention was paid to explore the relationship between hotspots, structural domains and half-interfaces as they represent the classical hierarchical decomposition of a PPI.

The early results derived from data analysis tend to comfort the consensus that PPI chemical space is hard to define[4]. However, hotspots amino acids neighborhood composition seems to play little to no role in their classification. This contradicts the trend sometimes observed at the single-hotspot level[5] and justifies the need to consider hotspot motifs rather than the hotspots themselves.

Bibliography:

[1] Macalino J. Y. S.; Basith S. *et al.*, Molecules, **2018**, 23(8):1963

[2] Bergey C. M.; Watkins A. M. *et al.*, Bioinformatics, **2013**, 29(21):2806-2807

[3] Watkins A. M.; Arora P. S., ACS Chem. Bio, **2014**, 9(8):1747-1754

[4] Bruno O. Villoutreix; Melaine A. Kuenemann *et al*, Mol Inform, **2014**, 33(6-7):414-437

[5] Wang L.; Liu Z. P.*et al.*, PEDS, **2012**, 25(3):119-126