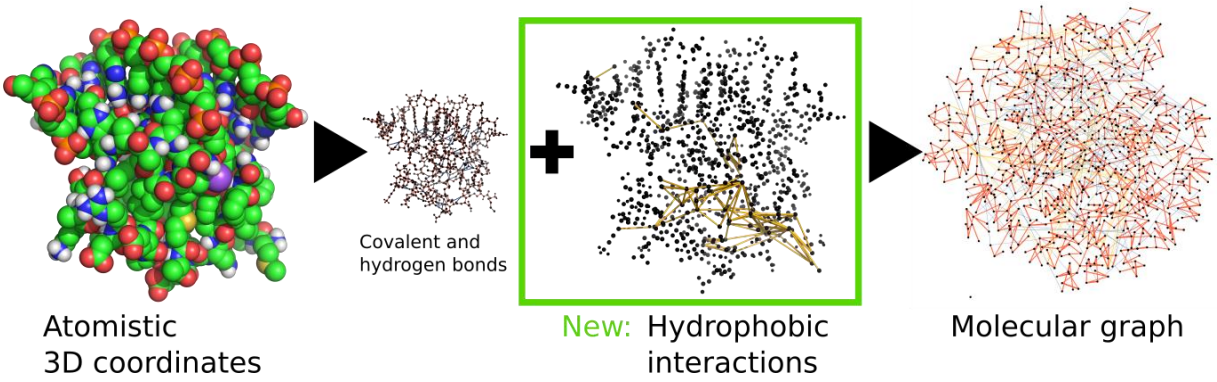
A novel methodology for encoding hydrophobic interactions in atomistic graphs constructed from biomolecular structures

Florian Song[1,3], Mauricio Barahona[2,3], Sophia Yaliraki[1,3]

[1] *Department of Chemistry, Imperial College London, London SW7 2AZ, UK*

[2] *Department of Mathematics, Imperial College London, London SW7 2AZ, UK*

[3] *Institute of Chemical Biology, Imperial College London, London SW7 2AZ, UK*



Studying protein structures at the atomistic level through the lens of mathematical graph theory has been successful in modelling, exploring and explaining a wide range of protein properties, such as flexibility [1], multi-scale organisation [2] and allostery [3], whilst remaining computationally feasible even for large system sizes.

Stronger types of bonding, particularly covalent bonds and hydrogen bonds can be clearly defined as the interaction between two atoms and thus fit well into the paradigm of molecular graphs. However, hydrophobic interactions are equally as crucial to protein shape and stability [4]. Yet, expressing them in computational models of protein structures is poorly understood, in parts due to their many-body property [5]. In this work, we introduce a novel methodology for incorporating both local and global aspects of hydrophobicity in biomolecular graphs. From a set of "candidate" interactions found through geometric and other constraints, a Relaxed Minimum Spanning Tree (RMST) [6] is computed to produce a sparsified set of accepted hydrophobic interactions. The universality of this procedure allows the consideration of both standard and non-standard residues as well as hydrophobic interactions between protein and ligand.

Despite no *a priori* information about the hydrophobicity of certain residues, we found that this approach successfully identifies hydrophobic connectivity in the buried interior of the protein, whilst capturing the many-body effect by creating regions of high connectivity consisting of many individual links. Together with existing methodology for constructing atomistic, biomolecular graphs from covalent and hydrogen bonds, we are able to encode detailed physico-chemical structural information in a concise, yet more realistic mathematical representation of protein structure.

[1] D.J. Jacobs, A.J. Rader, L.A. Kuhn, M.F. Thorpe, *Proteins: Struct., Funct., Genet.*, **2001**, *44*, 150-165.

[2] A. Delmotte, E.W. Tate, S.N. Yaliraki, M. Barahona, *Phys. Biol.*, **2011**, *8*, 055010.

[3] B.R.C. Amor, M.T. Schaub, S.N. Yaliraki, M. Barahona, *Nat. Commun.*, **2016**, *7*, 12477.

[4] K.A. Dill, *Biochem. J.*, **1990**, *29*, 7133-7155.

[5] M.S. Lin, N.L. Fawzi, T. Head-Gordon, *Structure*, **2007**, *15*, 727-740.

[6] M. Beguerisse-Diaz, B. Vangelov, M. Barahona, *2013 IEEE Global Conference on Signal and Information Processing*, **2013**, 937-940.