Multiscale Dynamics of Proteins revealed by Graph Partitioning

Antoine Delmotte[1], Renaud Lambiotte[2], Mauricio Barahona[1,2], Sophia N. Yaliraki[2,3], Edward W. Tate[3]

***Short Abstract* — Proteins are complex systems with structural and dynamical properties occurring at various time and spatial scales which the current experimental and computational methods can only partly access. This work presents a novel approach extracting structural and dynamical features of proteins at different scales using a graph partitioning algorithm. The method has been able to identify biochemically relevant partitions such as amino acids, secondary structures, and functional domains. The robustness of the partitions, probed by the variation of information, and compared with random graphs, is shown to be related to the biochemical significance of the detected communities.**

***Keywords* — Multiscale modeling of proteins, graph partitioning, robustness of partitions and variation of information**

I. Introduction

BECAUSE of their very high complexity, understanding the multiscale structural organization of proteins, and its relation to their dynamic behavior, is extremely difficult. Many experimental techniques, including Nuclear Magnetic Resonance (NMR) spectrometry, X-ray crystallography and Fluorescence Resonance Energy Transfer (FRET), and computational methods, including the detailed Molecular Dynamics simulation and coarse-grained models such as Normal Mode Analysis (NMA) or Elastic Network Models (ENM), have already been successfully applied. Nevertheless, each of them has only access to limited time and spatial scales, and none really allows relating the behavior at the femtoseconds atomistic level to the micro to milliseconds large domains motions.

II. Methods

Recently, M. Barahona and coworkers [2,3] defined a novel quality function for graph partitioning. As random walkers explore the graph, the community structure is extracted by identifying groups of nodes where the random walkers have a higher probability of being trapped than expected purely by chance. Depending on the time allowed to the random walkers, regions of different sizes can be explored, and, if there is any, the hierarchy of structures into communities can then be extracted.

Because of their intrinsic multiscale organization, the same quality function, associated with an algorithm generating the partitions, such as the very efficient Louvain algorithm [4], can be applied on proteins. The latter can indeed be represented in the form of a graph, where each node corresponds to a particular atom, and each edge, to an interaction. In this representation, it has been suggested that the dynamics of the protein can be related to the dynamics of the random process taking place on the graph [5].

In order to assess their relevance, the identified partitions are associated with a measure of their robustness to small perturbations. Namely, as small variations are introduced in the partitioning algorithm, an information theory measure called the variation of information [6] calculates the distance between the resulting partitions. By comparing the values obtained in proteins with different types of random graphs, it is shown that biochemically relevant partitions can be identified from their robustness.

III. Results and Conclusion

This method has been successfully applied to different proteins, including Adenylate Kinase, Cyclin-dependent Kinase 9 in complex with Cyclin T1, and the Myosin Tail Interacting Protein. It has allowed extracting biochemically relevant structural features of protein at different scales, and relating them with their corresponding dynamics.

Acknowledgements: This work was funded by the British Heart Foundation (BHF) centre of Research Excellence at Imperial College London.

[1]Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK, E-mail: antoine.delmotte09@imperial.ac.uk

[2]Institute for Mathematical Sciences, Imperial College London, London, SW7 2AZ, UK

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

References

[1] Henzler-Wildman, K. & Kern, D. (2007), 'Dynamic personalities of proteins', *Nature* **450**(7172), 964–972.

[2] Delvenne, J., Yaliraki, S. N. & Barahona, M. (2008), 'Stability of graph communities across time scales', ArXiv e-prints

[3] Lambiotte, R., Delvenne, J. C. & Barahona, M. (2009), 'Laplacian dynamics and multiscale modular structure in networks', ArXiv e-prints

[4] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008), 'Fast unfolding of communities in large networks', Journal of statistical Mechanics: Theory and Experiment 2008(10), P10008+.

[5] Meliga, S. (2009), Graph Clustering of Atomic Networks for Protein Dynamics, Master's thesis, Imperial College London, London, United Kingdom.

[6] Meila, M. (2007), 'Comparing clusterings–an information based distance', Journal of Multivariate Analysis **98**(5), 873 – 895.