Continuous heterogeneity analysis of CryoEM images through Zernike polynomials and spherical harmonics

David Herreros Calero[1], Roy Lederman[2], James Krieger[3], David Myška[4], David Strelak[5], Jiri Filipovic[6], Ivet Bahar[7], Jose Maria Carazo[5] and Carlos Oscar Sorzano[5]

[1]CSIC, Madrid, Spain, [2]Yale University, United States, [3]Department of Computational and Systems Biology at the University of Pittsburgh, United States, [4]Institute of Computer Science, Masaryk University, Jihomoravsky kraj, Czech Republic, [5]CSIC, United States, [6]Institute of Computer Science, Masaryk University, United States, [7]University of Pittsburgh, United States

The study of molecular flexibility is greatly evolving thanks to the improvements introduced in fields like Electron Microscopy and to the development of new and specialized software focused on the analysis of the movements macro-molecules undergo. The combination of all these tools and fields is allowing us to understand how these motions take place, which is essential to reveal how macro-molecules' chemistry and structure meet to perform a specialized function in their environment.

When studying conformational variability, it is possible to follow two different approaches depending on the scope of the analysis. In some cases, it is useful to analyze just a set of stable conformations recovered after a reconstruction process (discrete heterogeneity) while other studies focus on the analysis of the continuous motion a macro-molecule undergoes when it transitions among different transient and stable conformations (continuous heterogeneity). In both cases, there is the need of developing new algorithms [1,2,3] capable of analyzing these data in a simpler, unsupervised manner, but without compromising the quality of the results and the information extracted from the study.

In this context, we present in this work the development of a new algorithm to study the continuous heterogeneity of macro-molecules where no user interaction is required. Since the user interaction is minimized, this new method introduces a simplification in the analysis of the continuous heterogeneity of macro-molecules and increases the reproducibility of the results. The analysis consists of the computation of a deformation field among different structures represented by a set of two-dimensional images using a mathematical basis defined on the sphere. These deformation fields can be afterward used to map all the structures simultaneously in the same coordinate framework to study their structural relationships. Also, the information stored in the deformation field can be decomposed into different components which provide more insights into how a specific transition develops among different states.
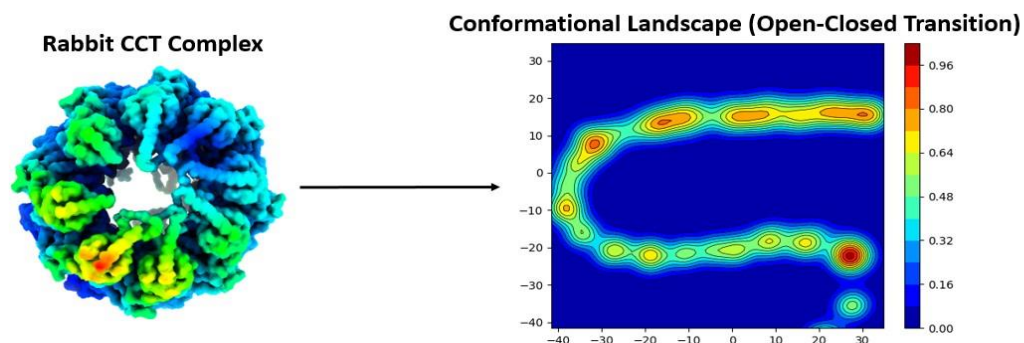
Figure 1. Conformational landscape recovered from an Open-Closed transition of a CCT complex from rabbit simulated using Molecular Dynamics.
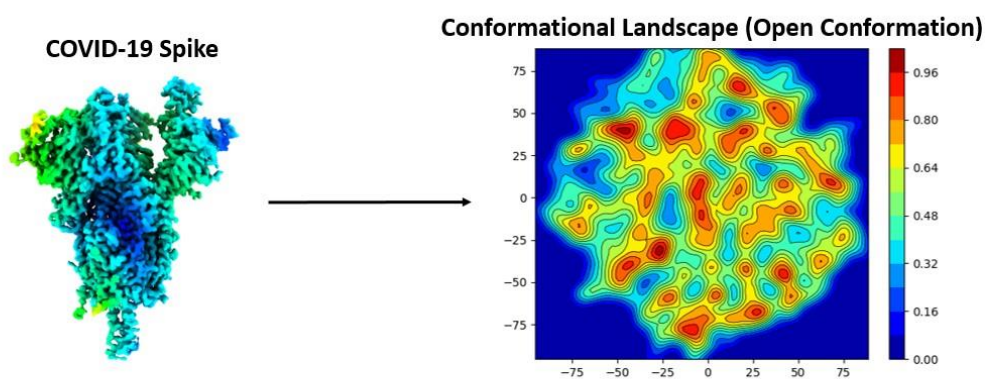

Figure 2. Conformational landscape recovered from the particles of the COVID-19 spike corresponding to the open state of the protein.

References

[1] C.O.S. Sorzano, A Jiménez, J Mota, J.L. Vilas, D. Maluenda, M Martínez, E. Ramírez-Aportela, T. Majtner, J. Segura, R Sánchez-García, Y. Rancel, L. Del Caño, P. Conesa, R. Melero, S. Jonic, J. Vargas, F. Cazals,Z. Freyberg, J. Krieger, I. Bahar, R. Marabini, and J.M. Carazo. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. Acta Crystallogr., 75:19–32, jan 2019.

[2] J. Frank, A. Ourmazd. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. Methods, 100:61-67, may 2016.

[3] M.Chen, M.L., Baker and S.J. Ludtke. Application of Deep Learning to CryoEM Heterogeneity Analysis. Biophy. Journal, 110:158a, feb 2016.