Exploring the 3D genome structure of viruses by means of multiscale simulations and polymer theory

Simón Poblete[1*],Fernando Mora[2], Pablo Huijse [2] , Cristian Huijse [2] and Sebastián Etchegaray[2†]
[1]Instituto de Ciencias Físicas y Matemáticas, Universidad Austral de Chile, Valdivia 509000, Chile.
[2]Instituto de Informática, Universidad Austral de Chile, Valdivia 509000, Chile
*simon.poblete@uach.cl, †sebastian@motivus.cl

Abstract

Understanding RNA viruses' assembly is vital for controlling their replication and propagation[1]. The structure of these agents consists of a single RNA strand containing the genetic information enclosed by a proteinaceous capsid of high symmetry. The assembly process involves an active interplay between capsid and genome, and, as usually found in many soft-matter and biological systems, it is a mechanism that entangles several time and length scales. This interplay between levels of resolution often demands incorporating different computational, theoretical, and experimental techniques to have a complete picture of the phenomenon. In particular, little is known about the structure of the RNA inside the virus, which has been, in some cases, partially accessed by cryo-electron microscopy and X-ray scattering.

Mean-field polymer theory, treating RNA as a negatively charged polyelectrolyte, has pointed out the relevance of electrostatics, persistence length, and secondary structure in the arrangement of the virus' interior [2]. In addition, other methodologies have indicated the importance of sequence-specific sites in the capsid proteins, which guide the folding process [3]. In the present work, we complement these efforts by proposing a multiscale strategy for obtaining realistic geometrical configurations of the genome, incorporating experimental data, and treating the RNA at different resolutions, which is needed to obtain a concrete picture, including finite-size and steric effects. We will discuss the parametrization of the polymer models used in the computational description at atomistic, coarse-grained [4,5], and mesoscopic levels of resolution. In addition, we will address some details about how to bridge these representations and the topological rules which must be obeyed [6].

Our methodology requires a massive sampling of small computer-generated structures, which we have performed using the Motivus [7] distributed computing platform, also introduced here, as a convenient and versatile tool that can be of use for other scientific purposes.

Finally, we discuss how this approach could be applied to other viruses using the experimental information available.

Acknowledgements to Project Fondecyt Iniciación 11181334 for financial support.

References

[1] R. Zandi, B. Dragnea, A. Travesset and R. Podgornik, Phys. Rep. 847, 1 (2020).
[2] G. Erdemci-Tardogan, H. Orland And R. Zandi, Phys. Rev. Lett. 119, 188102 (2017).
[3] R. Twarock, R. J. Bingham, E. C. Dykeman and P. G. Stockley, Curr. Opin. Vir. 30, 1 (2018)
[4] S. Poblete, S. Bottaro and G. Bussi, Nucleic Acids Res. 46, 1674 (2018).
[5] S. Poblete and H. V. Guzman, Viruses 13, 1555 (2021).
[6] B. C. Thiel, G. Bussi, S. Poblete and I. L. Hofacker, bioRxiv 2022.07.02.498583; doi: https://doi.org/10.1101/2022.07.02.498583
[7] https://motivus.cl/