ENHANCING AI-DRIVEN BIOMOLECULAR STRUCTURE PREDICTION WITH VOROMQA-AA: A NEW APPROACH TO ACCURACY ESTIMATION

Rita Banciul[1], Kliment Olechnovič[1,2], Justas Dapkūnas[1]

[1] Institute of Biotechnology, Life Sciences Center, Vilnius University
[2] CNRS Laboratoire Jean Kuntzmann, Grenoble
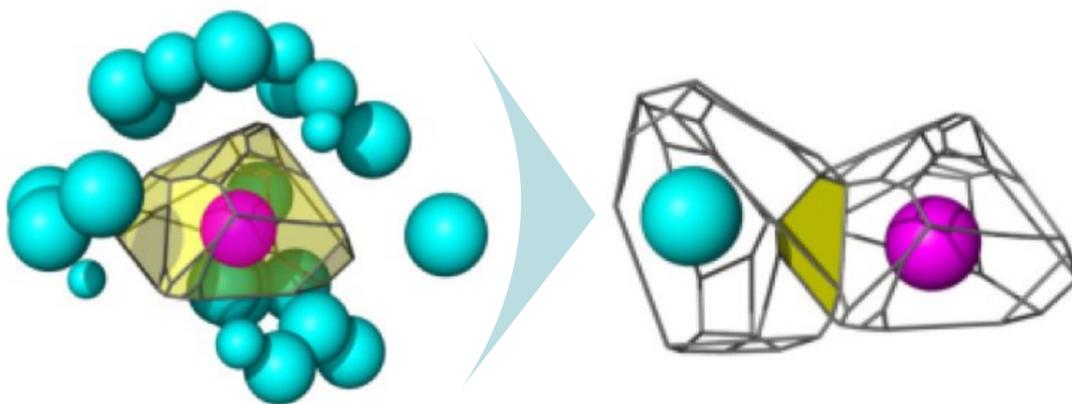rita.banciul@gmc.vu.lt

All cellular processes from energy production to cell division and signal transduction are based on interactions between proteins, nucleic acids, and other molecules. The structures of the resulting biomolecular complexes directly affect how they function. Therefore, knowing the three-dimensional arrangement is crucial for understanding fundamental biological processes, as well as designing biomolecules for therapeutic and industrial applications.

Recent advances in deep learning-based modeling have significantly improved the accuracy and accessibility of protein structure prediction. However, predictions for nucleic acids, ligands, and their complexes with proteins remain less reliable, having lower accuracy and less certain confidence metrics. While methods like AlphaFold and RoseTTAFold represent major progress, they function as black-box models, offering little transparency into their decision-making [1]. This lack of interpretability makes it difficult to assess reliability, identify errors, and fully trust their predictions, posing challenges for applications requiring high structural accuracy. Furthermore, multiple biomolecular modeling methods are available, each generating several structure models of varying accuracy, thus selecting the most reliable prediction becomes a major challenge.

To address both issues, structure quality assessment tools are essential. Previously, we developed VoroMQA [2], a quality assessment method for protein structures. However, with the release of methods capable of predicting biomolecular complexes including proteins, nucleic acids, ligands, and ions [3], there is an increasing need for accuracy estimation methods of broader scope.

Here, we present the development of VoroMQA-aa (VoroMQA all-atom), an advanced quality assessment tool capable of evaluating a broad range of macromolecular structures, including proteins, nucleic acids, and their complexes with ligands and ions. VoroMQA-aa is a knowledge-based statistical potential that utilizes contact areas derived from Voronoi tessellation to assess structure quality. To train this potential, we utilized a large dataset of experimentally determined structures. Due to the limited availability of structures containing nucleic acids, ligands, and ions, we adopted a novel training approach. Instead of relying on a clustered subset of reference structures, we incorporated all available structures, using clustering data to account for similarity of sequences and interaction interfaces. This strategy enhances the robustness and generalizability of VoroMQA-aa across diverse biomolecular systems.

VoroMQA-aa bridges the gap between AI-driven structure prediction and reliable model selection as well as interpretation, providing deeper insights into biomolecular interactions and functions.

Fig. 1. Constructing contact area of two atoms by Voronoi tessellation.

[1] von Eschenbach, W.J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. Philos. Technol. 34, 1607–1622.
[2] Olechnovič, K., Venclovas, Č. (2017). VoroMQA: Assessment of protein structure quality using interatomic contact areas. Proteins, 85(6), 1131–1145.
[3] Abramson, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630, 493–500.