

## **OT-RMSD: A GENERALIZATION OF THE ROOT-MEAN-SQUARE DEVIATION FOR ALIGNING UNEQUAL-LENGTH PROTEIN STRUCTURES**

**Yue Hu<sup>1</sup>, Zanxia Cao<sup>2</sup>, Yingchao Liu<sup>3</sup>**

<sup>1</sup>*Qilu University of Technology (Shandong Academy of Sciences), China*

<sup>2</sup>*Institute of Biophysics, Dezhou University, China*

<sup>3</sup>*Shandong Provincial Hospital, Shandong First Medical University, China*

[huyue@qlu.edu.cn](mailto:huyue@qlu.edu.cn)

The Root-Mean-Square Deviation (RMSD) metric, coupled with the Kabsch algorithm for optimal superposition, represents a cornerstone of quantitative structural biology. However, its application is fundamentally limited to the comparison of structures with a pre-defined, one-to-one correspondence of equal length, precluding its direct use for the vast majority of biologically relevant comparisons involving proteins of different lengths or unknown equivalences. To overcome this limitation, we introduce OT-RMSD, a novel, parameter-light method that generalizes the RMSD for comparing unequal-length protein structures without a pre-defined alignment. Our approach recasts the correspondence problem as an Optimal Transport (OT) task, utilizing a fundamental, distance-based cost function to identify the most economical mapping between two protein structures. This transport plan is then used as a soft-weighting scheme within an iterative Sinkhorn-Kabsch framework to simultaneously solve for correspondence and optimal superposition. We test our method on a benchmark dataset of 40 challenging protein pairs and demonstrate its superiority over both the widely-used PyMOL ‘align’ command and a more complex, heuristic-driven variant of our own framework. On average, OT-RMSD identifies larger and more accurate alignment cores, achieving a significantly lower RMSD. We conclude that OT-RMSD provides a mathematically principled, robust, and effective extension of the classic RMSD concept, offering a powerful tool for modern structural analysis.

This work is available as a preprint on bioRxiv:  
<https://www.biorxiv.org/content/10.1101/2025.09.23.678064v1.full>