Correspondence

DOI: 10.5582/ddt.2023.01063

EQUIBIND: A geometric deep learning-based protein-ligand binding prediction method

Yuze Li[1], Li Li[1], Shuang Wang[1,2,*], Xiaowen Tang[1,*]

[1] Department of Medical Chemistry, School of Pharmacy, Qingdao University, Qingdao, Shandong, China;
[2] Department of Stomatology, Huangdao District Central Hospital, Qingdao, Shandong, China.

SUMMARY Structure-based virtual screening plays a critical role in drug discovery. However, numerous docking programs, such as AutoDock Vina and Glide, are time-consuming due to the necessity of generating numerous molecular conformations and executing steps like scoring, ranking, and refinement for the ligand-receptor complexes. Consequently, achieving rapid and reliable virtual screening remains a noteworthy challenge. Recently, a team of researchers from Massachusetts Institute of Technology, led by Stärk et al., developed an SE(3)-equivariant geometric deep learning based protein-ligand binding prediction approach, EQUIBIND. In comparison to conventional docking methods, EQUIBIND has the capacity to predict the binding modes of small molecules with target proteins rapidly and precisely. It presents an innovative resolution for high-throughput screening of drug-like compounds.

Keywords EQUIBIND, deep learning, virtual screening, protein-ligand binding prediction

Drug discovery is a highly costly and time-intensive process, requiring several years and billions of dollars to discover new drugs that target either newly identified or well-established receptors. Throughout this protracted timeline, issues like drug-induced side effects and adverse reactions can lead to the failure of drug development endeavors (*1*). The central challenge in drug discovery revolves around deciphering the complex binding interactions between drug-like compounds and proteins. As a result, the accurate and efficient identification of binding sites within proteins for compounds, the characterization of binding conformations, and the evaluation of ligand-protein interactions have become critical aspects of the drug discovery process. In recent decades, computational methodologies have gained significant traction in predicting drug binding sites and conducting virtual screenings to identify potential therapeutic agents (*2-4*).

In the field of drug design and development, widely used molecular docking approaches such as AutoDock Vina and Glide usually employ strategies based on geometric and energy-based matching principles (*5,6*). The protein-ligand binding complex undergoes a series of conformational sampling, score ranking and energy-driven refinement (*7*). Nowadays, driven by the advancement of computational technology, artificial intelligence-driven drug design has emerged as a prominent approach in drug development. Artificial intelligence approach like machine learning, deep learning, and related methodologies have been integrated in traditional virtual screening area. Recently, EQUIBIND, an SE(3)-equivariant geometric deep learning model was reported, which integrates Graph Matching Networks (GMN) (*8*) and E(3)-equivariant graph neural networks (E(3)-GNN) (*9*) to predict the binding conformations of ligand-receptor complexes (*10*). EQUIBIND significantly enhances predictive efficiency by bypassing the need for extensive sampling procedures. It provides direct predictions of binding sites on receptors (blind docking) and ligand binding conformations. A benchmark test demonstrates that EQUIBIND exhibits a processing efficiency of about nine times surpasses than currently prevalent commercial binding predicting programs (*10,11*).

Moreover, EQUIBIND capitalizes on a K-NN graph (k-nearest neighbor graph) to delineate three-dimensional structures. The input parameters predominantly encompass the rigid three-dimensional arrangement of the protein and the adaptable, optimizable conformations of small molecules, which are generated at random using RDKit. Subsequently, EQUIBIND employs Invariant Equivariant Graph Matching Networks (IEGMN) to transform the three-dimensional coordinates and predict the binding

conformation of the ligand. To ensure accuracy, EQUIBIND employs the SE(3)-equivariant mechanism to calculate the binding sites of receptors, in which the predictive sites are trained to match the actual binding sites. However, EQUIBIND does not guarantee the absolute correctness of complex structures, often resulting in bond lengths and bond angles that deviate from reasonable ranges. To address this issue, EQUIBIND aligns the initial conformations generated by RDKit with those generated by IEGMN. As this process involves only rotatable single bonds, it ensures the local structural integrity. The algorithm primarily focuses on aligning atomic positions around rotatable single bonds, each of which can be manipulated independently. In comparison to traditional docking methods, the EQUIBIND model demonstrates commendable predictive accuracy, enabling precise forecasts of ligand-target binding modes (*10*).

However, EQUIBIND is not devoid of its limitations. It indirectly models the atomic positions of side chains by encoding features within the receptor α-carbon graph utilizing localized frameworks, as a result, some predictions may manifest notable disparities. Stärk *et al*. have attempted surface atom refinements and fine-tuning approaches by incorporating atomic subgraphs of the receptor. Regrettably, these endeavors failed to yield significant enhancements to the performance and instead contributed to extended runtimes. Consequently, there remains efforts for the future advancement and refinement of EQUIBIND.

In conclusion, EQUIBIND serves as a novel deep learning-based protein-ligand binding prediction tool. By leveraging SE(3)-equivariant transformations and utilizing minimal input information, it can directly predict the binding sites on receptors and the binding modes of ligands. This capability ensures both high prediction accuracy and a substantial enhancement in prediction efficiency. In the future, whether employed in conjunction with existing virtual screening tools or utilized independently for protein-ligand binding prediction, EQUIBIND holds the potential to emerge as a crucial asset in the drug molecule screening and discovery process.

Funding:This work was supposed by the Qingdao Marine Science and Technology Innovation Special Project (No. 22-3-3-HYGG-25-HY).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Reymond JL, Awale M. Exploring chemical space for drug discovery using the chemical universe database. ACS Chem Neurosci. 2012; 3:649-657.
2. Evans R, O'Neill M, Pritzel A, *et al*. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2021; 2021.10.04.463034.
3. Jindal A, Kotelnikov S, Padhorny D, Kozakov D, Zhu Y, Chowdhury R, Vajda S. Side-chain packing using SE(3)-transformer. Pac Symp Biocomput. 2021; 27:46-55.
4. Ganea O-E, Huang X, Bunne C, Bian Y, Barzilay R, Jaakkola T, Krause A. Independent SE(3)-equivariant models for end-to-end rigid protein docking. arXiv. 2021; 2111.07786.
5. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: A new approach for rapid, accurate docking and scoring.2. enrichment factors in database screening. J Med Chem. 2004; 47:1750-1759.
6. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455-461.
7. Hassan NM, Alhossary AA, Mu Y, Kwoh CK. Protein-ligand blind docking using QuickVina-W with inter-process spatio-temporal integration. Sci Rep. 2017; 7:15451.
8. Li Y, Gu C, Dullien T, Vinyals O, Kohli P. Graph matching networks for learning the similarity of graph structured objects. ICML. 2019; 34:3835-3845.
9. Satorras VG, Hoogeboom E, Welling M. E(n)-equivariant graph neural networks. arXiv. 2021; 2102.09844.
10. Stärk H, Ganea O-E, Pattanaik L, Barzilay R, Jaakkola T. EQUIBIND: Geometric deep learning for drug binding structure prediction. arXiv. 2022; 2022.05146.
11. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, Sunseri J, Koes DR. GNINA 1.0: molecular docking with deep learning. J Cheminform. 2021; 13:43.

Received August 19, 2023; Revised September 15, 2023; Accepted September 19, 2023.

**Address correspondence to:*
Shuang Wang, Department of Medical Chemistry, School of Pharmacy, Qingdao University, Qingdao 266071, Shandong, China; Department of Stomatology, Huangdao District Central Hospital, Qingdao 266555, Shandong, China.
E-mail: ws1201youyou@163.com

Xiaowen Tang, Department of Medical Chemistry, School of Pharmacy, Qingdao University, Qingdao 266071, Shandong, China.
E-mail: xwtang1219@qdu.edu.cn

Released online in J-STAGE as advance publication September 26, 2023.