

EQUIBIND: Geometric Deep Learning for Drug Binding Structure Prediction

Hannes Stärk, Octavian-Eugen Ganea, Lagnajit Pattanaik, Regina Barzilay, Tommi Jaakkola

摘要

预测药物样分子如何与特定蛋白质靶点结合是药物发现中最核心的问题。发现一种非常快速的计算绑定方法将有利于实现诸如快速虚拟筛选或药物工程等重要应用。但是现有的方法在需要耗费大量的计算资源，因为它们依赖于大量的候选采样，以及评分、排名和微调步骤。因此，作者使用 EQUIBIND 来解决这个问题，EQUIBIND 是一个三维等变几何深度学习模型，它能够对 i) 受体结合位置（盲对接）和 ii) 配体的结合姿势和方向等两个方面进行直接预测。与最近的传统方式的基线相比，EquiBind 实现了显著的速度提升和质量提升。此外，当以增加运行时间为代价将其与现有的微调技术相结合时，我们显示了额外的改进。最后，我们提出了一种新的快速微调模型，该模型基于到给定输入原子点云的冯·米塞斯角距离的封闭形式全局最小值来调整配体可旋转键的扭转角，避免了以前最小化差分进化策略昂贵的计算资源消耗。

关键词：深度学习；药物设计；配体结合预测；3 维等变图匹配网络

1 引言

在目前的药物发现领域中，每发现一个新药都需要大量的时间、金钱和人力成本，目前已知的具有潜在药物功能的化合物分子有 10^{60} 种，而人体内的蛋白质类型又有高于 100,000 种。在如此庞大的配体和受体数量下，基于传统的方式，即采用生物和化学实验的方式，需要耗费大量成本并且成功率低下，因此通过计算机技术对潜在的药物分子进行筛选能够极大的减少药物发现的时间和成本，从而达到加速新药发现和研制过程，降低病人用药成本以及掌握药物与蛋白质结合位点并发现其作用机制等方面具有重要意义。而在药物发现的过程中，其中一个核心问题是知道药物分子如何与靶蛋白进行相互作用形成复合物，这个问题涵盖了许多方面如结合动力学、构象变化、原子间相互作用等。目前用于预测药物的三维结合结构的商用软件通常是对药物的可能结合姿势进行穷举，然后经过打分函数进行评价和选择，这种方式需要很多的计算资源和时间，效率比较低下。因此提出了一个新模型 EquiBind，一个新的用于药物结合结构预测的集合与图深度学习模型，通过修改图匹配网络 (graph matching network, GMN) 和 E(3)-等变图神经网络 (E(3)-GNN) 得到，该模型能够直接预测配体——受体结合的复合物结构并且不需要通过大量的采样作为预处理（与传统方式相比），从而能够加快模型的预测时间。EquiBind 利用深度学习的方法，在各种指标上均取得改进，并且能够与现有的基于能量的方式结合，获得更优的结果^[1]。

2 相关工作

目前的计算机辅助药物设计主要可以分为两大类，一类是传统计算机辅助药物方法：基于配体（药效基团）的设计，基于结构（结合靶点）的设计和基于定量构效关系的设计。另一类则是近年兴起的利用机器学习模型对药物进行特征筛选并预测配体的结合位置和分子构象。

2.1 传统计算机辅助药物设计方法

传统的虚拟筛选方式是利用配体和受体之间的对接进行筛选，根据分子间的范德华力、氢键、疏水性等相互作用进行计算，最后得到配体和受体最有可能的结合位点，目前最常用的就是基于结合靶点的分子对接算法，例如 AutoDock Vina 及一系列基于 Vina 的优化算法。通过计算进行虚拟对接获得配体和受体的结合位置和姿势，并根据评分函数进行打分和筛选，这类方法采用的是基于大量候选位置信息进行采样筛选。但这种基于对接的方法有一个很大的弊端在于需要消耗大量的计算资源和时间，基于对接的软件需要首先知道配体相对于受体的结合位置，再通过随机生成大量配体构象信息，才能够进行对接、打分和筛选。

2.2 基于机器学习的计算机辅助药物设计

除了传统的已知结合位置进行对接筛选的方法之外，目前也出现了许多基于机器学习的方法，例如利用图神经网络和卷积神经网络对配体和受体结合亲和力进行预测；利用 3D 卷积神经网络识别蛋白质的可结合位点；利用差分进化算法来改进配体到活性位置及其结合姿势的微调。除上述的部分相关机器学习在药物设计方面的应用以外，与配体分子和蛋白质分子对接相似的问题是，蛋白质和蛋白质之间的相互对接，如利用两个输入的氨基酸序列，预测出两个蛋白质形成的复合物结构，或利用几何深度学习的方式对蛋白质进行刚体建模并对接等方式。以上工作均对计算机辅助药物设计的发展有所帮助

3 本文方法

3.1 本文方法概述

EquiBind 模型的输入和输出如图 1 所示，先利用 RDKit 工具库将药物——蛋白质复合物中的药物配体分子部分取出并进行三维空间中的随机建模，使之失去结合位置信息，再将配体和受体的三维结构图作为输入，并预测配体结合位置和结构。由于 PDBbind 数据集中只有 19K 的复合物数据，因此需要将正确的物理化学知识加入深度学习模型中，否则模型可能不足以从较小的数据量中获得很好的训练效果。为了实现这个目标，EquiBind 实现了：i) 保证了两个分子初始 3D 位置的独立性，确保配体和受体不存在相互之间的信息，实现结合位置姿势的预测。ii) 通过仅仅改变分子中的可旋转键角度，在实现配体灵活性的同时保持局部结构（二面角和键长）的固定，确保了构象的合理。iii) 利用非交集损失来防止空间冲突或不切实际的相互作用。

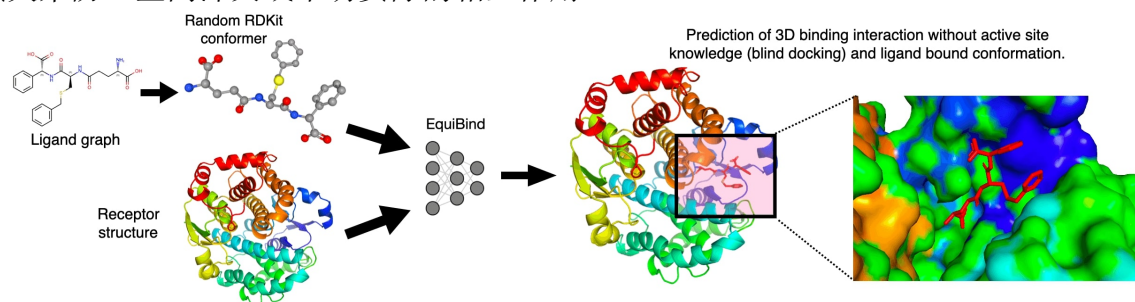


图 1: EquiBind 的输入以及预测的输出

3.2 模型框架

EquiBind 的具体模型结构如图 2 所示，EquiBind 主要可以分为多个独立等变图匹配网络层，多头注意力筛选以及配体分子结构拟合三个部分。

在等变图匹配网络层中，根据配体和受体的初始结构图信息，对配体和受体的三维坐标以及特征进行变换，并对该变化施加几何距离限制，确保构象的稳定性。EquiBind 中使用了 8 层的等变图匹配网络层，将获得的输出经过多头注意力。根据受体的均值向量来获得配体的重要位点，对于受体的重要位点同理。得到主要位点后，将配体中的位点拟合到受体上，就可以得到配体和受体相互作用的位置信息，获得结合复合物信息。为了避免构象紊乱，将初始结构图进行快速点云配体拟合，并在拟合过程中增加二面角限制，确保构象的稳定性，得到最终结果。

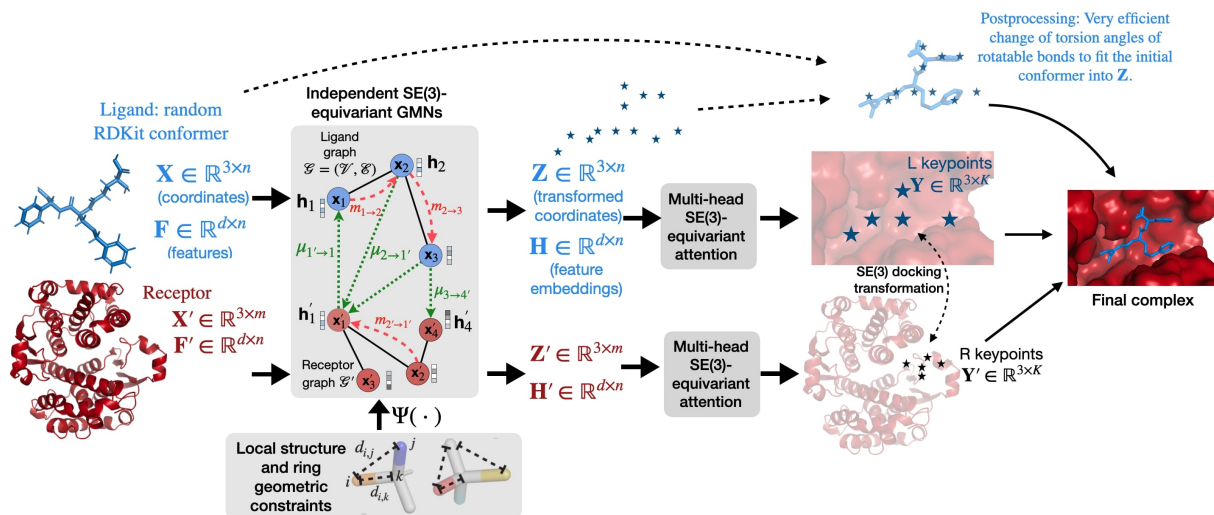


图 2: EquiBind 模型的整体框架

3.3 损失函数定义

4 复现细节

4.1 与已有开源代码对比

作者提供的 environment.yml 文件只包含了所使用的库名，不包括其版本号，在使用作者提供的 environment 文件创建的环境中无法正常处理配体和受体分子，也无法运行代码，因此修改了 environment 并提供了对应的版本号。

本次复现使用了原作者开源代码的 EquiBind 模型，在 EquiBind 的运行基础上，实现了相关代码，将模型的输出进一步处理并输入 AutoDock Vina 中进行调整。获得 EquiBind 预测的配体结构位置中心以及结合范围，并增加一定单位的可移动范围，确保当预测位置发生在蛋白质内部时能够通过微调使得构象更合理。实现了论文中的将 EquiBind 与传统方式结合的部分，并且与原论文所采用的方式不同，在实验中采用了 Vina 进行调整^[2]。

5 实验结果分析

实验得到配体和受体结合后的三维结构图信息，将实验结果可视化，取两个例子（如图 3 所示）可以发现 EquiBind 能够找到配体和受体实际结合的结合位点，但其构象不一定准确（图 3 中的黄色构象），在使用 Vina 后可以获得更接近真实构象的结果（图 3 中的蓝色构象）。除此以外，EquiBind 的结果有可能会由于在主要位点选择时在受体的内部，导致构象不合理（如图 4 所示），图中灰色为蛋白质受体，红色为配体实际结合位置，黄色为 EquiBind 输出结果，蓝色为 Vina 调整后的配体构象，可以观察到 EquiBind 预测的配体构象与蛋白质受体发生冲突，使用 Vina 能够避免这种现象的产生。

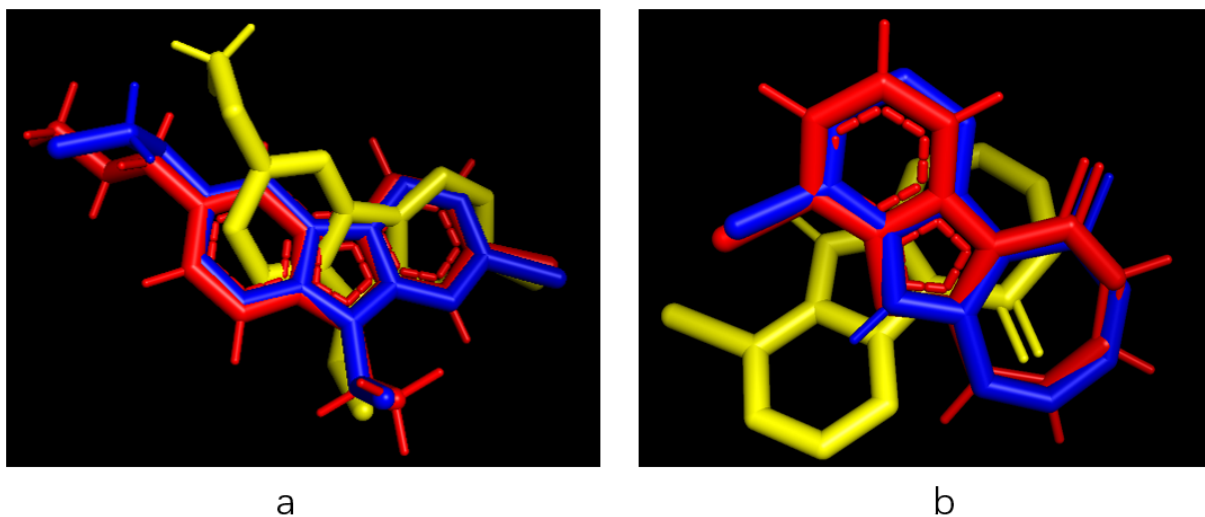


图 3: 配体位置和姿势实验结果, 分别是实际构象 (红色)、EquiBind 的输出构象 (黄色) 和 EquiBind+Vina 的输出构象 (蓝色)

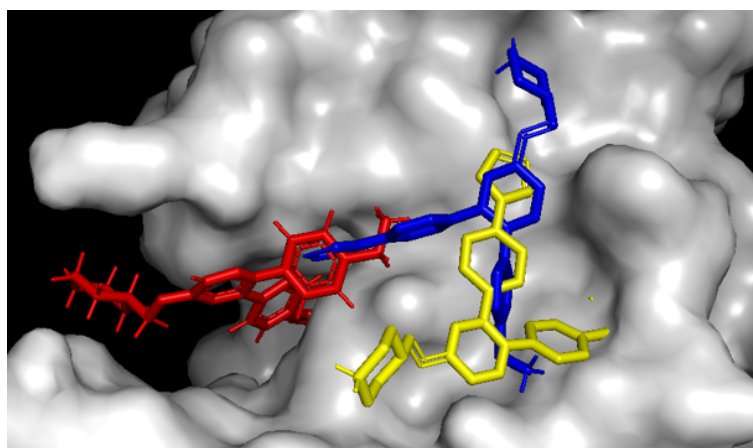


图 4: EquiBind 的结果可能出现不合理构象, 使用 Vina 能改善这种情况。其中红色为实际构象, 黄色为 EquiBind 的输出, 蓝色为 EquiBind+Vina 的输出。

6 总结与展望

EquiBind 实现了在不知道配体和受体之间信息的前提下, 预测出配体与受体结合的位置和姿势, 并且相比于具体的配体结合姿势, EquiBind 在预测结合位置中心方面具有更高的准确性, 因此可以利用这一点, 将 EquiBind 与传统的对接方式进行结合, 可以得到更加接近真实构象的预测结果并避免发生结构冲突。EquiBind 为计算机辅助药物设计提供了新的思路, 通过深度学习的方法预测药物分子与蛋白质的相互作用, 对于快速虚拟筛选具有重要意义。EquiBind 极大的缩短了筛选所需要的时间, 但其准确性还需要进一步的提升。

参考文献

- [1] STÄRK H, GANEA O, PATTANAIK L, et al. EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction[J]., 2022, 162: 20503-20521.
- [2] TROTT O, OLSON A J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading[J]. Journal of Computational Chemistry, 2010, 31(2): 455-461.