

TSUBAME 共同利用 令和 4 年度 学術利用 成果報告書

利用課題名 深層学習を用いた分子動力学法ベースの水和熱力学量計算法の高速化
英文: A Fast Computation Method for Solvation Thermodynamic Quantities Using Deep Learning

吉留 崇
Takashi Yoshidome

東北大学 大学院工学研究科 応用物理学専攻
Department of Applied Physics, Graduate School of Engineering, Tohoku University
URL

Grid Inhomogeneous Solvation Theory (GIST)は、水和エネルギーと水和エントロピーの空間的な分布を計算することで、水和を詳細に表すことが出来るものの、分子動力学法を用いるため大きな計算コストがかかることが問題である。そこで、本研究では水和自由エネルギー分布を予測する深層学習モデルの開発を行った。我々の深層学習モデルは、水和自由エネルギー分布を約 1 分程度で計算でき、かつ GIST の計算結果を定量的に再現することに成功した。

Grid Inhomogeneous Solvation Theory (GIST) can describe hydrations around proteins using the spatial distribution of the hydration energy and hydration entropy. However, huge computation is required for the GIST because the GIST requires the trajectory obtained using molecular dynamics simulations. In the present study, a deep learning (DL) model to predict the hydration free energy distribution was proposed. It was found that our DL model can quantitatively predict the hydration free energy distribution in less than 1 minute.

Keywords: タンパク質、水和、深層学習、グリッド不均一溶液理論

背景と目的

タンパク質における水和は、折り畳みやリガンドとの結合などの様々な過程において重要な役割を果たす。Grid Inhomogeneous Solvation Theory (GIST)[1]は、水和エネルギーと水和エントロピーの空間的な分布を計算することで、水和を詳細に表すことに成功した。しかし、分子動力学法を用いるため大きな計算コストがかかる。このため、GIST のドッキングアルゴリズムなどへの応用例は未だに少ない。

本プロジェクトでは、GIST 計算を高速化することを目的とし、タンパク質立体構造からその周りの水和自由エネルギーの分布を予測する深層学習モデルの開発を行った。2-fold cross validation を用いて深層学習モデルの最適なハイパーパラメータを決定し、深層学習モデルを構築した。リガンド結合に伴い、タンパク質表面から移動する水分子の自由エネルギーを計算した所、GIST で計算したものと良く一致した。

概要

本研究では、以前我々が提案した、深層学習に基づ

く水和分布予測法[2]と同じ深層学習モデルを採用した。深層学習の学習データとテストデータとして、以前、分子動力学法と GIST を用いて計算した 76 のタンパク質の水和自由エネルギー分布を使用した。ハイパーパラメータを最適化するために、128 通りのハイパーパラメータを用意し、それぞれに対し 2-fold cross validation を実行した。実行には 68 個のタンパク質を用いた。誤差関数を最小とするハイパーパラメータを選び、残り 10 個でテスト計算を行った。さらに、誤差関数が小さいものから 40 個を選び、水和エントロピーと水和エネルギーを予測する深層学習モデルの構築と 2-fold cross validation を実行した。

結果および考察

最適なハイパーパラメータを決定し、深層学習モデルを構築した後、そのパフォーマンスを調べるために 10 個のタンパク質に適用した。具体的には、リガンドが結合する際にタンパク質から移動する水分子の自由エネルギーを計算した。構築した深層学習モデルと GIST の両方で行い、結果を比較した。その結果、深層学習

モデルで得た水分子の自由エネルギー (ΔG_{Deep}) の GIST 計算結果 (ΔG_{GIST}) からの誤差率 ($|\Delta G_{\text{Deep}} - \Delta G_{\text{GIST}}| / \Delta G_{\text{GIST}}$) は、約 4% であった。この結果は、我々の深層学習が高精度で GIST の計算結果を再現できることを意味する。また、我々の深層学習モデルは、約 1 分程度で水和自由エネルギーの分布を計算でき、この計算時間は GIST 計算 (2 日程度) の約 1/1000 であった。よって、我々の深層学習モデルは、高速にかつ正確に水和自由エネルギー分布を計算できることが分かった。

まとめ、今後の課題

本プロジェクトでは、高速にかつ正確に水和自由エネルギー分布を計算できる深層学習モデルを開発した。今後は、開発した深層学習とドッキングアルゴリズムとの融合を行う。

[1] C.N. Nguyen, T.K. Young, and M.K. Gilson, *J. Chem. Phys.*, **137**, 044101 (2012).

[2] K. Kawama, Y. Fukushima, M. Ikeguchi, M. Ohta, and T. Yoshidome, *J. Chem. Inf. Model.*, **62**, 4460 (2022).