

2023 年度年次報告書
次世代 AI を築く数理・情報科学の革新
2023 年度採択研究代表者

飯田 慎仁

北里大学 未来工学部
助教

拡散モデルによる蛋白質の立体構造集団の生成

研究成果の概要

近年開発された、蛋白質の立体構造予測手法 AlphaFold により、たんぱく質の立体構造予測は格段に向上した。しかしながら、既存の予測手法らはたんぱく質の揺らぎを考慮した予測結果を与えない。そこで本課題では、分子動力学シミュレーションと生成モデルを用いた、ペプチドの立体構造集団の生成モデルを構築することを目的とする。

今年度は、多数のペプチドに対して全原子分子動力学シミュレーションを行うことで学習データを作成した。そして、この得られた学習データを学習させた拡散モデル(pepflow)を構築した。このモデルを用いて、ペプチドの立体構造(集団)の生成を行った。ペプチドの主鎖構造を再現できたり、またアミノ酸側鎖の構造および水素原子の位置も適切に生成できた。

これらの立体構造集団に対する、詳細な解析として、主鎖二面角の自由エネルギー地形を描写した。具体例として、Gly-Ala-Gly 配列の主鎖二面角に対する自由エネルギー地形(ラマチャンドランプロット)を計算したところ、分子動力学シミュレーションで得られた地形と拡散モデルから得られたものが類似していることが確認できた。このことは適切にモデルの学習が進んでいることを示している。また、他の配列でも同様の解析を行い、分子動力学シミュレーションと拡散モデルの結果の類似性を確認できた。

今後の方針として、まず(1)より長いペプチドに対するデータの生成を行えるようにするため、データの拡充を行う。(2)今回学習したモデルを用いて、より長いペプチドを生成できるか(外挿性)を評価する。