

ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名： タンパク質構造を予測する技術の開発	実施時期：2024/12/20～2025/3/31 所属機関名：産業技術総合研究所 代表者氏名：富井 健太郎
---------------------------	--

成果概要：
深層学習を含む機械学習等の手法を利用し、タンパク質の立体構造を予測する新規手法を開発しました。
特に、単数のアミノ酸配列を入力とする予測手法の精度改善に挑戦しています。

成果のポイント：

(A) モデルアーキテクチャ

ESM 3B Prediction Our Prediction Ground Truth

(B) コンタクトマップ予測 on T1028 (CASP14)

(C) 3次元構造予測
T1025 (CASP14)
tm score 90%

- 大規模言語モデルはタンパク質の構造予測にも用いられているが、一般的なモデルはタンパク質構造予測の精度があまり高くありません。我々は、蛋白質構造の情報に基づいたタンパク質言語モデルを開発しました。
- 我々の手法（図A）は、8分の1のレイヤー数しか使用していないにもかかわらず、SOTAの一つであるESMfoldと同等の予測精度を達成しました。
- 残基コンタクト予測においては、我々のモデルは遥かに大きなESM-3B（5倍のサイズ）の精度を大きく上回っています。たとえば、ターゲットT1028（図B）では、ESM-3Bが見逃した多くのコンタクトを我々のモデルでは正確に予測されています。
- 現在、モデルの改善を続けています。

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：