

AI予測とMD計算を用いたタンパク質複合体予測の高精度スクリーニング法

大阪大学大学院生命機能研究科 河口真一

目的：タンパク質は、生体内の反応を促進する働きがあり、多くの場合、タンパク質間の相互作用によって制御されている。タンパク質は、そのアミノ酸配列に依存した立体構造を形成しており、表面構造が相互作用に重要である。タンパク質の立体構造を予測するAIプログラム、AlphaFoldが開発・公開されている。単量体のタンパク質については、予測精度が高いことが報告されているが、複合体の構造予測・結合予測では、信頼性スコアが高くても、実際には結合が検出できない場合も多くみられる。偽陽性を排除する二次スクリーニングを行い、結合実験の負担を軽減することが求められる。

内容：タンパク質複合体の予測精度を向上させるため、AlphaFoldが出力した複合体の立体構造に基づいて、その自由エネルギー、及び結合強度を、計算機科学の手法によって見積もる。見積もりを正確に行うためには、複合体の立体構造の精度が重要である。そのため、AlphaFoldが出力した構造を最適化し、自由エネルギーが低い状態にすることが必要である。この目的のために、D3センターのスーパーコンピュータで、分子動力的シミュレーション（MD）を行った。

結果：SRCタンパク質は、癌原遺伝子であり、多くの癌細胞で発現の上昇が見られる。異常な疾患状態では、通常とは異なるタンパク質間相互作用が形成され、疾患の一因となることが考えられる。AlphaFold2を用いて、約15,000種類の二量体構造予測を行った結果、信頼性スコアが比較的高い（0.7以上）タンパク質ペアは、SRCでは325個ほど見つかった（図1、2）。AlphaFold2が出力する複合体構造は、アミノ酸残基間の相互作用を考慮して予測されており、必ずしもエネルギー的に安定ではなく、結合の自由エネルギーを評価する目的には、構造精度が不十分である。そこで、1ナノ秒程度の分子動力学計算（MD）を行い、各原子を安定な位置に収束させ、現実の構造に近づけた後に、結合強度（解離定数）の評価を行なった（図3）。

利用した計算機：SQUID 汎用CPUノード群

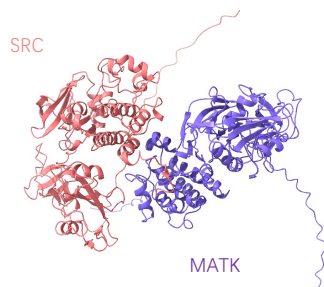


図1 AlphaFold2が予測したSRCとMATKとの複合体構造

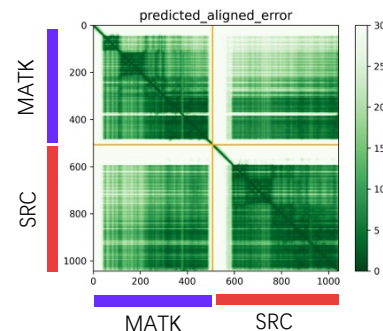


図2 予測した複合体構造の信頼性（緑が濃いほど確からしい）

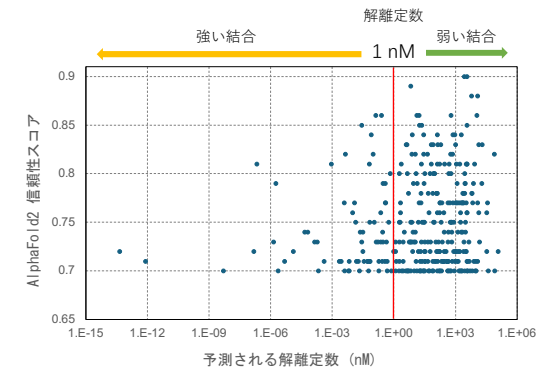


図3 AlphaFold2の予測スコアと見積もり解離定数との比較