

プロジェクト名(タイトル):

Data generation for machine learning for improving drug discovery processes

利用者氏名: ○大田 雅照 (1)

理研における所属研究室名:

(1) 計算科学研究センター HPC/AI 駆動型医薬プラットフォーム部門 AI 創薬連携基盤ユニット

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

サイクリックペプチドは、特異的で高い結合能を有し、従来は「undruggable」とされてきたタンパク質表面をターゲットにできることなどから、新規創薬モダリティとして注目されている。サイクリックペプチドは主鎖および側鎖に数多くの自由度を持っていることから、そのコンフォメーションは多様であり、水中、生体膜中、標的タンパク質との結合時などで周囲の環境に応じたコンフォメーション変化を起こすことも知られている。サイクリックペプチドの立体構造は、そのエネルギーにより規定されており、エネルギーは分子力場法などを用いて簡易的に評価することができるが、CH- π 相互作用、ハロゲン結合、原子近接による電荷移動など分子力場法では取り扱えない項が多く存在する。本方法では、まず非天然型アミノ酸を含む多様なサイクリックペプチドの多数の立体構造について、そのエネルギーを精密な量子化学手法である Fragment Molecular Orbital (FMO) 法を用いて算出する。次に、FMO 法で得られた多数のエネルギー値を学習させ、サイクリックペプチド立体構造から瞬時に精度よく量子化学レベルのエネルギーを推定する人工知能(Artificial Intelligence, AI)力場を構築することを試みている。

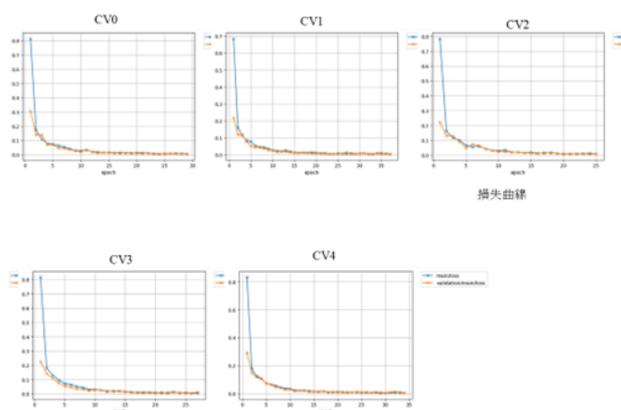
2. 具体的な利用内容、計算方法

サイクリックペプチドの立体構造構築: 非天然型アミノ酸も含めアミノ酸配列から初期立体構造を構築するソフトウェアを開発した。1サイクリックペプチドについて、構築した初期立体構造を基に、OMEGA を利用して最大 200 コンフォマーを生成した。さらに 200 コンフォマーそれぞれについて、MOE を用いて温度 200K, 300K, 400K, 500K の 200ps の分子動力学(MD)計算を実施し、最終時点でのコンフォマーを取得した(4 温度×200 コンフォマー)。1サイクリックペプチドについて、OMEGA (200 コンフォマー)と MD (800 コンフォマー)を合わせた 1,000 コンフォマーの FMO 計算を行い、FMO エネルギーを算出した。サイクリックペプチドに

ついては、アミノ酸残基数 7-13 の 16 種類のサイクリックペプチドの初期構造を生成し、10 サイクリックペプチドについては、それぞれ 1,000 コンフォマーの FMO 計算を終了させた。

3. 結果

サイクリックペプチド FMO AI 力場構築: 以下に 5 fold cross validation の結果(損失曲線)を示す。



4. まとめ

本方法で得られた AI 力場を用いることにより、FMO エネルギー値の絶対値を予測することには成功した。しかしながら、同一のサイクリックペプチドのコンフォマー(立体構造)間のエネルギー差を正確に推定することのできる精度ではなかった。

なお、我々が実施したエネルギー値のみを使用した AI 力場の構築に加え、我々が計算した FMO 法に基づくエネルギー微分データも含めて共同研究者にデータを提供し、エネルギー微分値とエネルギー値の両者を利用した AI 力場構築も実施している。

5. 今後の計画・展望

同一のサイクリックペプチドのコンフォマー(立体構造)間のエネルギー差を正確に推定することのできる AI 力場の構築をめざし、さらなるデータの追加(新たなサイクリックペプチ

ド構造の追加、既存サイクリックペプチドの変異体構造の追加など)、および、新たな AI モデル構築法(FMO エネルギー値のみではなく、FMO エネルギー微分値を用いた転移学習など)を検討していく。

6. 利用がなかった場合の理由