

## タンパク質の長時間ダイナミクスを短時間 MD で作る

山口 孝太 (筑波大学 生物学類) 指導教員: 原田 隆平 (筑波大学 生命環境系)

### 【背景・目的】

近年、計算機の進歩に伴い生体分子の分子動力学 (MD) シミュレーション (MD) を実行することが可能となってきた。例えば MD 汎用計算機である "Anton" はミリ秒オーダーの計算を可能にした。しかし、Anton を利用してミリ秒オーダーのトラジェクトリ (原始座標の時系列データ) が取れたとしても、生体分子が取り得る全ての状態をサンプルできる保証はない。そこで、ディープラーニング (GAN) を利用することで、複数の短時間 MD シミュレーションのデータから長時間 MD のシミュレーションデータを予測するようなモデル<sup>1</sup>が最近提案されている。現状において予測モデルが適用されているのは高分子だけであり、タンパク質をはじめとする生体分子には適用されていない。適用が難しい理由は、タンパク質は高分子と比較してダイナミクスとして緩和が遅い (レイイベント) であるため、短時間 MD では取り得る状態を全てサンプルしきれないからである。そこで、レイイベントサンプリング法として開発された PaCS-MD<sup>2</sup>を適用し、状態 A から状態 B までに存在する遷移経路を探索することで、生体分子系への応用を考えた。これまで様々なレイイベントサンプリング法が開発されてきたが、多くは系の状態を無理やり不安定にするように外部バイアスに基づく戦略である。適切なバイアスを指定しない場合、普通では起こりえない構造遷移が誘起されてしまい、長時間ダイナミクスの予想には向いていない。一方 PaCS-MD は、既知の 2 状態間の遷移過程をバイアスなしで誘起可能な強力な手法である。本研究では、タンパク質の短時間 MD で長時間ダイナミクスを予測することを目的とし、PaCS-MD-GAN の開発を行った。

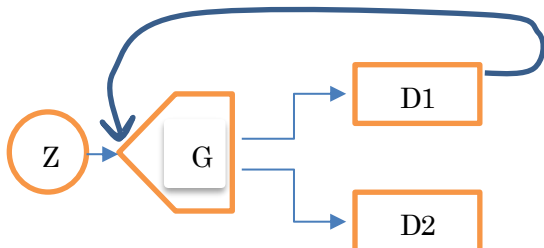
### 【方法】

PaCS-MD-GAN は 2 つのパートから構成される。

**PaCS-MD:** タンパク質の状態 A から状態 B までに存在する状態をサンプリングする。

**MD-GAN:** 深層生成モデルである GAN に基づき、深層生成モデルである GAN に基づき、複数の短時間 MD が生成するトラジェクトリを学習させ、長時間ダイナミクスを予測する。

### ・MD-GAN のネットワーク

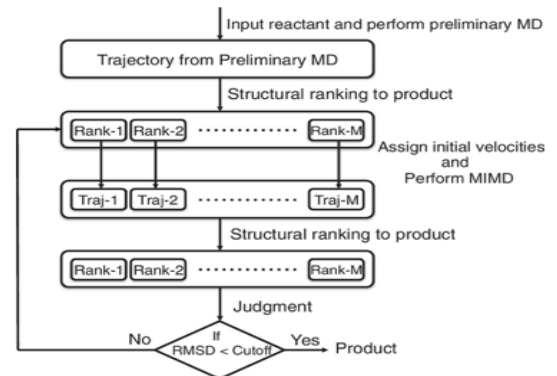


GAN (Generative Adversarial Nets)<sup>3</sup> は識別器と生成器の二つからなるネットワークで、生成器は識別器を騙すことができるように、識別器は本物か偽物かを判別できるように学習することで、本物の分布に近いものを生成できる。

MD-GAN は識別器を 2 つ使うことで、生成器がどんな出力を出すかを制御する。D1 で時系列データを学習させる。D2 では時系列データのつながりを学習させる。従来の時系列予想のモデルで

ある LSTM (Long Short Time Memory) と比べると、予測誤差がたまり発散、収束が起こること (exposure bias) が少ない。生成器と識別器は Multilayer Convolutional Network である。更に GAN の学習において問題となっているモード崩壊、勾配消失問題を防ぐために WGAN-gp を使用する。

### ・PaCS-MD のアルゴリズム



PaCS-MD によりタンパク質の状態 A-B 間の遷移経路を探索する。具体的には短時間 MD を利用し、目的構造 (B 状態) に近い分子構造を新たな初期構造に選択して、選択構造から短時間 MD をリスタートする。常に B 状態に近い分子構造から短時間 MD を繰り返すことで、目的構造へ至る遷移経路を生成する。

### ・分子動力学計算の実施

MD プログラム (Gromacs2019) を使い、 $T = 300 \text{ K}$ 、 $P = 1 \text{ atom}$  の条件で、T4 リゾチームの  $3.4 \mu\text{s}$  の長時間 MD を実行。

### 【結果】

結果は当日に公開します。

### 【考察】

従来の MD-GAN ではタンパク質への応用は困難であったが、PaCS-MD で状態間の遷移経路を探索することでそれを可能にした。しかし、PaCS-MD では遷移経路の近傍に存在する局所的な状態探索である場合多く、取り得る全ての状態を探索できている保証はない。しかし、従来のレイイベントサンプリング法の多くは外部バイアスを用いているので、今回の手法には向かない。何故ならば、外部バイアスによりアーティファクトが引き起こされ、異常な長時間ダイナミクスを予測するためである。そこで、より広い状態を合理的に探索可能なノンバイアスレイイベントサンプリング法として OFLOOD<sup>4</sup>が選択肢として考えられ、今後 PaCS-MD との併用することで、更なる予測精度の向上を目指したい。

### 【参考文献】

1. E. Katsuhiko, K. Tomobe, K. Yasuoka. 2018. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*.
2. H. Ryuhei, and A. Kitao. 2013. *Journal of Chemical Physics*.
3. Goodfellow, et. al. 2014. In *Advances in Neural Information Processing Systems*.
4. H. Ryuhei, T. Nakamura, Y. Takano, Y. Shigeta. 2015. *Journal of Computational Chemistry*.