

# 分散型シミュレーションを用いて生体分子のレアイベントを捉える

原田 隆平 (筑波大学計算科学研究センター ryuhei@ccs.tsukuba.ac.jp)

生体分子は室温で熱的に揺らいでおり、様々な物理量を計算するためには信頼性の高い統計平均や、十分に長時間の時間平均を考慮しなければならない。生体分子の状態間にはエネルギー障壁が存在し、室温のエネルギースケールと比較して十分大きい場合が多い。ゆえに、生体分子の立場で考えると、エネルギー障壁を乗り越えて構造遷移することは稀にしか起こらない現象、つまり「レアイベント」である。一般的に、生体機能に関する生体分子の構造変化は長時間スケールで誘起されるレアイベントで特徴付けられる場合が多いため、生体機能を解析するうえでレアイベントを検出することは重要である。しかし、通常の分子動力学シミュレーション (MD) を用いて有限の計算時間で全ての取り得る状態をサンプリングすることは現実的でないため、レアイベントの検出は難しい。ゆえに、現在に至るまでレアイベントを捉えるために様々なサンプリング法が開発されてきた。また、サンプリング法の開発と異なる戦略として、計算機チップの開発や超並列計算でMDの高速化を図る研究も進んでいる。長時間MDが実行できれば、時系列データとしてレアイベントが手に入るため、生体機能を原子分解能の高解像度で解析することができる。しかし、高性能な計算機の利用は限られるため、長時間MDを実行することは依然として困難を伴う。

我々は長時間MDを実行する代わりに、短時間MDを複数の異なる初期条件からスタートし、サンプリングを稼ぐことでレアイベントを捉えるサンプリング法を開発した。我々のレアイベントサンプリング法の探索戦略は極めて単純である。具体的には、「重要な初期構造の選択」と「選択した初期構造から短時間MDのリスタート」から成

るサイクルを繰り返すだけである。特徴的な点として、バイアスをかけることなく生体分子のレアイベントを探索することができる。効率的にレアイベントサンプリングが可能である理由は、短時間MDに基づく分散型シミュレーションを通して、「稀にしか発生しない分子構造」の出現確率を選択的に上昇させているからである。

代表例である Outlier Flooding Method (OFLOOD) は、生体分子のレアイベントを記述する適当な反応座標を指定した後、短時間MDから得られる分子構造を構造空間に射影して構造分布を得る。次に、構造分布をクラスタリングすることで生体分子の状態分布を解析し、確率密度が疎 (スパース) な状態領域から初期構造を集中的に選択し、短時間MDをリスタートさせる。一般的に、クラスタに属さない密度が疎な状態領域は、「はずれ値 (Outlier)」と呼ばれる。OFLOODでは、出現確率が低い Outlierに該当する分子構造を初期構造として選択し、短時間MDをリスタートさせる構造探索を繰り返す。Outlierは、準安定状態間の遷移状態の近傍に存在している可能性が高いため、Maxwell-Boltzmann分布に従い初期速度を再生成して運動エネルギーを与え直すことで、隣接する準安定状態に容易に構造遷移すると期待され、効率的なレアイベントサンプリングが可能となる。適用研究の例としてOFLOODは、レアイベントとしてタンパク質のフォールディング過程をナノ秒オーダーの計算コストで捉えることに成功している。従来のMDを用いた場合、フォールディング過程を捉えるためにはマイクロ秒以上の計算コストが必要となるため、OFLOODを用いることで、かなりの計算コストを削減することができる。

—Keywords—

## 生体分子の分子動力学シミュレーション:

ニュートンの運動方程式 (微分方程式) を数値的に解くことで、生体分子のダイナミクスを原子座標の時系列データ (トラジェクトリ) として得ることができるシミュレーション手法。

## 分散型シミュレーション:

複数のプロセッサが同一のプログラムを実行し、各プロセッサは別々のデータを処理するモデル。生体分子の分子動力学シミュレーション (MD) の場合には、初期条件 (具体的には、初期構造や温度や圧力などのMDパラメータ) の異なる複数の短時間MDを各プロセッサで別々に実行することで様々なトラジェクトリを得ることが可能となり、生体分子の統計アンサンブルを稼ぐことができる。

## タンパク質フォールディング:

細胞内で翻訳されたタンパク質が、アミノ酸配列に対応する固有の立体構造 (天然構造) に自発的に折りたたまれる現象。 $n$ 残基のタンパク質は、およそ  $10^n$  通りの立体構造をとることが可能であり、網羅的探索によりタンパク質の天然構造を探し当てることは現実的に不可能である。このため、タンパク質フォールディングには特定の経路が存在すると考えられている。