

# 構造妥当性評価を用いて生体分子の構造 を効率的に探索する分子シミュレーション技術の開発

## Developments of Efficient Sampling Methods for Exploring Protein Structures with Structural Validation

保田 拓範

筑波大学 理工情報生命学術院 生命地球科学研究群

生物学学位プログラム 博士後期課程

### 1. 研究目的

タンパク質のダイナミクスはその機能に直結している。そのため、タンパク質の機能を理解する上で、その動的な性質を解明することは有効である。全原子分子動力学シミュレーション (MD) は、タンパク質のダイナミクスを高い時間的分解能で観測することができるため、タンパク質の機能解明に広く用いられてきた。一方で、タンパク質が取りうる構造は多様かつ複雑であるため、通常分子動力学シミュレーションではタンパク質がとりうる立体構造を過不足なく探索することは依然として難しい。この問題を解決するため、ソフトウェア面からは様々な Enhanced-sampling 法 (ES 法) の開発が行われてきた。

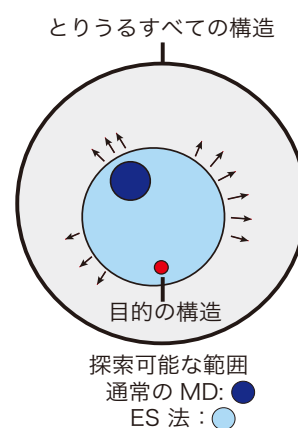


図 1: 通常の MD と ES 法の探索範囲の比較

ES 法は、エネルギーや温度、リサンプリングなど多様な外部バイアスを加えることにより、長時間スケールにおいてとりうるタンパク質の構造を効率的に探索することができる。(図-1) しかしながら、ES 法では外部バイアスの制御に失敗すると構造の崩壊を引き起こしたり、必要以上に広い構造空間を探索したりする問題が存在し、外部バイアスをどのように制御するかは課題であった。そこで、ES 法における外部バイアスを制御する新規アルゴリズムの開発を行うことを目的とした研究を行った。

### 2. 研究成果の内容

研究成果として、タンパク質の構造妥当性をチェックしながら、外部バイアスを徐々に付加するアルゴリズムを開発した。具体的には、結晶構造の精度評価に用いられてきた構造因子 (G-factor) に着目し、外部バイアスを制御するアルゴリズム“G-factor-based External Bias Limiter (GERBIL)”を提案した。さらに、GERBIL を様々な ES 法と組み合わせることで構造探索の効率化に成功した。これらの結果から、ES 法によって生成されたトラジェクトリに対して構造妥当性評価を行うことで、外部バイアスを適切に制御できることを示すことができた。

より詳細には、提案アルゴリズムの検証として、異なる計算スキームを有する 2 種類

の ES 法 (Biased Sampling Method: BS 法、Adaptive Sampling Method: AS 法) へ GERBIL を適用した。前者については、Accelerated MD (aMD) を選択した。aMD は Boost Potential を加えることにより構造探索を効率化するが、Boost Potential の強さを適切に制御することが難しい。そこで、aMD に GERBIL を適用した aMD-GERBIL により Boost Potential の強さを制御した。結果として、Boost Potential を適切に制御することで、対象としたタンパク質の多様な安定状態を同定することができた。また、後者については、Outlier Flooding Method (OFLOOD) を選択した。OFLOOD はトラジェクトリを構造空間に射影して得られる分布の外縁に該当するタンパク質構造から MD をリスタートするリサンプリングを繰り返すセレクションバイアスにより効率的な構造サンプリングが実現する。しかし、結晶構造から離れた構造空間を探索するため、生体内ではとれない構造をサンプリングしてしまう。この問題を回避するため、OFLOOD に GERBIL を適用した OFLOOD-GERBIL によりセレクションバイアスを制御した。結果として、構造妥当性を評価しなかった場合と比較し、OFLOOD-GERBIL の探索領域は長時間 (10  $\mu$ s) MD が探索した構造空間と良い一致を示した。以上から、AS 法においても構造妥当性評価を利用することで外部バイアスを適切に制御できることが示された。

### 3. 学際共同利用プログラムが果たした役割と意義

本研究で使用した分子動力学計算は生体分子を構成する全ての原子について、運動方程式の時間積分を行なうことで生体分子のダイナミクスを追跡する。そのため、演算を行なう上で必要な計算コストは非常に多く、本プロジェクトで使用するソフトウェアである GROMACS や AMBER は GPU の使用が、大幅な演算速度の上昇に寄与した。また、本プロジェクトで提案した新規アルゴリズムでは、独立かつ並列的に計算を実施することでパラメーター探索を行っている。そのため、多くの計算ノードを有する Cygnus は必要不可欠であった。以上が、高性能の GPU と多くの計算ノードを有する Cygnus の利用が本プロジェクト推進の上で果たした役割と意義である。

### 4. 今後の展望

今回の研究から、開発したアルゴリズムは多様な ES 法に対して適用可能であることがわかった。そこで、今回開発したアルゴリズムを、実際に生体内で機能を有するタンパク質に適用することでその機能解明を行っていく予定である。

### 5. 成果発表

#### (1) 学術論文

保田拓範 “学生優秀発表賞受賞寄稿「Enhanced sampling における構造妥当性の評価は外部バイアスを適切に制御する」”, 分子シミュレーション学会誌「アンサンブル」(2023).

保田拓範, 原田隆平, “Parallel Cascade Selection Molecular Dynamics (PaCS-MD) に基づくタンパク質の遷移経路探索”, 分子シミュレーション学会誌「アンサンブル」(2023).

(2) 学会発表

保田拓範、森田陸離、重田育照、原田隆平、第 46 回分子生物学会(1P-242)、神戸国際会館、2023 年 12 月

保田拓範、森田陸離、重田育照、原田隆平、第 61 回日本生物物理学会年会(3P-202)、名古屋国際会議場、2023 年 11 月

(3) その他

| 使用計算機                       | 使用計算機に<br>○ | 配分リソース※ |      |
|-----------------------------|-------------|---------|------|
|                             |             | 当初配分    | 追加配分 |
| Cygnus                      | ○           | 40000   | 0    |
| Wisteria/BDEC-01            |             |         |      |
| ※配分リソースについてはノード時間積をご記入ください。 |             |         |      |