

カスケード選択型分子動力学シミュレーションに基づく低解像度タンパク質構造の精密化

原田 隆平

筑波大学計算科学研究センター

1. はじめに

平成30年度(後期)スーパーコンピューター若手・女性利用者推薦制度の採択課題「カスケード選択型分子動力学シミュレーションに基づく低解像度タンパク質構造の精密化」の研究成果を報告する。

2. イントロダクション

タンパク質は極めて複雑な構造と精緻な機能をもつ高分子化合物であり、分子認識、情報伝達、酵素反応など、生体内でおこる様々な生命現象の根幹をなす。立体構造と機能の間には大きな相関(構造-活性相関)があることが期待されており、これまでX線結晶構造解析・中性子回折実験や核磁気共鳴法(NMR)などの実験的手法により、数多くのタンパク質の立体構造が明らかにされてきた。最近では、X線自由電子レーザーやクライオ電顕等の実験により、タンパク質動態から機能と構造変化の関係が徐々に明らかになりつつある。そのような状況の中、実験事実をミクロスコピックなレベルで理解する理論計算に期待が集まっている。

近年のスーパーコンピュータの発展、および解析手法の進展が相まって、生体内反応解析の分野は格段の進歩を遂げている。特に、分子動力学シミュレーション(Molecular Dynamics Simulation: MD)は、生体分子の動態を調べる上で極めて有効な研究手段である。しかしながら、生体内反応はミリ秒〜秒、時にはさらに長い時間かかる極めて遅いプロセスであり、レアイベントを再現するためには同等のシミュレーション時間が必要であるが、現在のMDシミュレーションではサブミリ秒程度が限界であるため、その抽出が極めて難しい。我々はこれまで、超並列カスケード選択型分子動力学シミュレーション(Parallel Cascade Selection MD: PaCS-MD) [1]を開発し、タンパク質の折りたたみやドメイン運動などの遅い過程の構造変化の解析を行ってきた。PaCS-MDは、(1)複数の短時間MDシミュレーション、(2)初期構造の選択を繰り返すことによって、初期構造から終構造への構造遷移を誘起する。様々な計算性能の比較を通して、従来のMDシミュレーションやMulticanonical MD [2], Accelerated MD [3]などの構造サンプリング法と比較してより平易な実装で比較的簡単にレアイベントを抽出可能であることを実証してきた。

従来のPaCS-MDでは、初期構造の選択時に終構造などの情報を事前に与えていた。同じアルゴリズムを用いて、実験から得られる低解像度の構造データ(例えば SAXS や Cryo-MD)と MD シミュレーションで得られたタンパク質構造に基づく理論計算データの差を小さくする構造を探索することが可能である。

本研究では、実験データとの類似度を用いて PaCS-MD を実行することにより低解像度の構造データを再現する MD 構造を探索し、高分解能のタンパク質構造を再構築する計算手法を開発した [4]。また、実験データとの非類似度を利用することで、構造遷移を促進する計算手法も合わせて開発したので、その研究成果を報告する。

3. 手法設計

3.1 PaCS-MD

PaCS-MD は、「初期構造の選択」と「選択した初期構造からリスタートする短時間 MD シミュレーション」による「構造リサンプリング」を繰り返すことにより遷移経路を探索する。本手法の特徴的な点として、バイアスをかけることなく構造間の遷移経路を探索可能である。従来法では、構造遷移を促進するため系の温度を高温にして MD シミュレーションを実行するか、系の特定の部分に外力や外部ポテンシャルといった「バイアス」を付加することがほとんどであった [5,6]。バイアスは、付加する位置や付加する大きさ、向き、時間などパラメーターを最適化した上で適用しなければならないので、パラメーター最適化の準備計算が必要となる。また、系のサイズが大きくなるにつれて、パラメーターの最適化自体が困難となる場合もあるため、構造探索の計算時間よりもパラメーターの最適化にかかる計算時間の方が長くなる可能性もある。本手法は、反応座標を設定してしまえば、遷移確率が高い初期構造を選択し、短時間 MD シミュレーションにより構造リサンプリングを繰り返していだけなので、バイアスは使用しない。更に、パラメーターを最適化するための準備計算も必要としない。本手法により効率的に構造探索が可能であるのは、構造リサンプリングを通して「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。図 1 に本手法の概念図を示す。図 1 に示す様に、始構造からスタートして終構造へ遷移する可能性が高い候補構造を反応座標に基づき選択し、短時間 MD シミュレーションをリスタートさせるサイクルを繰り返すことで、徐々に構造分布が終構造近傍に近づいていく。手法適用にあたり様々な反応座標が考えられるが、最も単純な例として、終構造が分かっている場合には、終構造に対する平均自乗距離(Root Mean Square Deviation: RMSD)が考えられる。構造リサンプリングにより得られるトラジェクトリを RMSD の値でランキングし、値の小さい順にスナップショットを複数個選択し、短時間 MD シミュレーションの初期構造として構造探索していくことにより、徐々に終構造へ近づいていく。

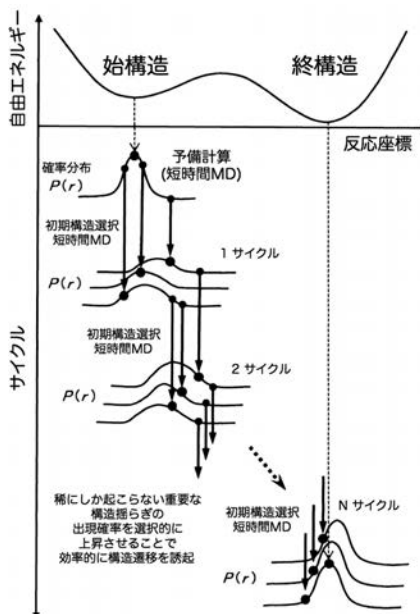


図 1: PaCS-MD の概念図

稀にしか起こらない重要な構造種らぎの出現確率を選択的に上昇させることで効率的に構造遷移を誘起

3.2 データ駆動型 PaCS-MD

SAXS データを参照した PaCS-MD では、スコアリング関数は理論的 SAXS データと実験的 SAXS データの間の平均自乗差 (χ)

$$\chi = \left\{ \frac{1}{K-1} \sum_{i=1}^K \left[\frac{\mu I_m(q_i) - I_{\text{target}}(q_i)}{\sigma(q_i)} \right]^2 \right\}^{1/2}$$

によって定義される。ここで K は $I_{\text{target}}(q)$ 内のデータ点の総数であり、 $\sigma(q)$ は標準偏差である。 $I_m(q)$ は PaCS-MD でサンプルされた構造の SAXS データの平均であり、 μ は最大値を揃えるための比例定数である。 $q = 4\pi \sin(\theta/\lambda)$, 2θ は散乱角であり、 λ は波長である。

一方、EM データを参照した PaCS-MD ではスコアリング関数は理論的 EM データと実験的 EM データの間の相互相関関数(CC)

$$CC = \frac{\sum(\rho_{m,i} - \bar{\rho}_{m,i})(\rho_{\text{target},i} - \bar{\rho}_{\text{target}})}{\sqrt{\sum(\rho_{m,i} - \bar{\rho}_{m,i})^2} \sqrt{\sum(\rho_{\text{target},i} - \bar{\rho}_{\text{target}})^2}}$$

によって定義される。ここで、 $\rho_{m,i}$ は、 i 番目のボクセルで PaCS-MD によってサンプルされた m 番目の構造の電子密度マップであり、 $\rho_{\text{target},i}$ は対応するターゲットの電子密度である。上付きバーは全てのボクセルに対する平均値を表す。

データ駆動型 PaCS-MD では、従来型の PaCS-MD で初期構造の選択に用いてきた指標である RMSD の代わりに、上述の χ や CC を指標に用いることで、実験データを出来るだけ再現するタンパク質構造を初期構造に選択していく。データ駆動型 PaCS-MD の概念図を図2に示す。図2に示す様に、サイクルを重ねていく毎に実験データとの誤差が小さく、あるいは実験データとの相関が高くなるように初期構造が選択されていく。最終的に、誤差と相関係数にしきい値を設けることで、実験データを再現する高解像度のタンパク質構造が得られる。

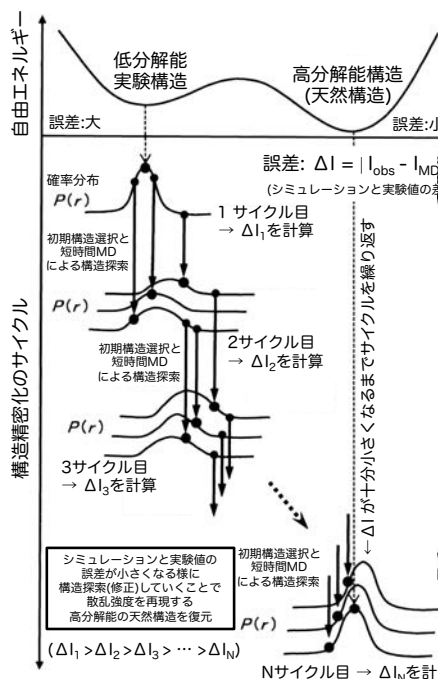


図2: データ駆動型 PaCS-MD の概念図

3.3 デモンストレーション

SAXS-駆動型PaCS-MDの応用例として、リジン-アルギニン-オルニチン結合(LAO)タンパク質を選択した。LAOタンパク質はOpen状態構造(PDBid : 2LAO) [7]とClosed状態構造(PDBid : 1LAF) [8]のX線結晶構造があり、 κ やCCを指標としてOpen-Closed構造遷移をSAXS-駆動型PaCS-MDにより再現した。SAXS-駆動型PaCS-MDの1サイクルは、NPT($T = 300$ K, $P = 1$ bar)アンサンブルのもとで100-psのMDシミュレーションを10個の独立な初期構造から初期速度を振り直して実行した。

SAXS-駆動型PaCS-MDの性能を示すために、SAXSのプロファイルと比較する。図3(a)にターゲットデータ(緑)、初期構造(青)、SAXS-駆動型PaCS-MDデータ(赤)をそれぞれ示す。初期構造では、 $q = 0.15$ から $q = 0.4$ においてターゲットデータから著しく外れていたが、SAXS-駆動型PaCS-MDから得られた理論SAXSプロファイルは、目標となる実験データのプロファイルを良好に再現し、ほぼ完全に重なっていた。SAXS-駆動型PaCS-MDによって得られた生成物(Closed状態構造)から測定した最小 C_{α} RMSDは 0.8\AA であり、Closed状態構造に関して高い分解能の構造が得られたことを示している。興味深いことに、SAXSデータが3次元の1次元平均回折パターンを表していたとしても、SAXS-駆動型PaCS-MDはレアイベントである構造遷移を誘起し、最終構造の細密化に成功している。この結果は、LAOタンパク質のOpen状態構造とClosed状態構造の二次構造のポーズのおよびトポロジーの類似性に起因している。したがって、本手法が低解像度の実験データから妥当な高分解能の原子レベルの構造を生じる可能性があることを示している。

EM-駆動型PaCS-MDに関しても、Open-Closed構造遷移が再現された。図3(b)は、Closed状態構造との C_{α} RMSDが最も小さいスナップショット(青)をX線構造(赤)・EMデータ(網掛け)に重ねたものである。得られた構造は、いくつかの可動性が高いループ領域を除いてX線構造とほぼ一致しており、EM-駆動型PaCS-MDの有効性が検証できた。

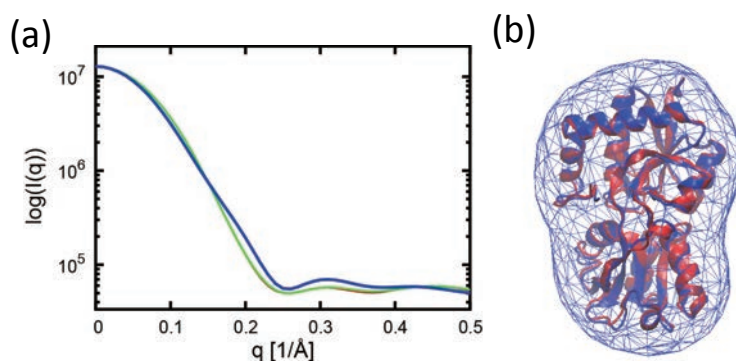


図 3. (a) SAXS のデータ(緑) : 実験, 青 : PaCS-MD の初期構造, 赤 : PaCS-MD の終構造, (b) Cryo-EM データ(網掛け), 青 : PaCS-MD の終構造, 赤 : X 線結晶構造

参考文献

- [1] R. Harada, A. Kitao, *J. Chem. Phys.* 139 (2013) 035103-035101-035110.
- [2] N. Nakajima, H. Nakamura, A. Kidera, *J. Phys. Chem. B* 101 (1997) 817-824.
- [3] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* 26 (2005) 1781-1802.
- [4] R. Harada, Y. Shigeta, *Phys. Chem. Chem. Phys.* 20 (2018) 17790-17798.
- [5] G.M. Torrie, J.P. Valleau, *J. Comput. Phys.* 23 (1977) 187-199.
- [6] G.M. Torrie, J.P. Valleau, *J. Chem. Phys.* 66 (1977) 1402-1408.
- [7] B.H. Oh, J. Pandit, C.H. Kang, K. Nikaido, S. Gokcen, G.F.L. Ames, S.H. Kim, *J. Biol. Chem.* 268 (1993) 11348-11355.
- [8] B.H. Oh, G.F.L. Ames, S.H. Kim, *J. Biol. Chem.* 269 (1994) 26323-26330.