



Title	MSDC-MDによる天然変性蛋白質の熱力学解析と構造予測
Author(s)	下山, 紘充
Citation	サイバーメディアHPCジャーナル. 2021, 10, p. 35-39
Version Type	VoR
URL	https://doi.org/10.18910/83286
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

MSDC-MD による天然変性蛋白質の熱力学解析と構造予測

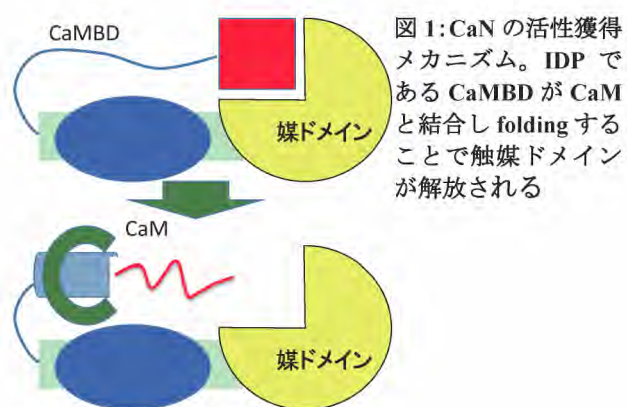
下山 紘充

北里大学 薬学部

1. 序論

多くの蛋白質は溶液中で秩序だった三次元構造へと折れ畳むことが知られている。これは天然構造と呼ばれ、蛋白質の機能と深く関わっている。その一方で、蛋白質の中には溶液中にあるにも関わらず秩序だった構造を持たない蛋白質が存在する（天然変性蛋白質、以下 IDP）。IDP の中には特定の対象蛋白質と結合することでフォールディングし、秩序だった天然構造を獲得するものが存在する。この現象は *coupled folding and binding* と呼ばれ、蛋白質が情報伝達を行う手段として関心が集まっている。

カルシニューリン(CaN)は細胞内でセリンやスレオニンを脱リン酸化する“分子スイッチ”であり、多くの細胞内反応に関わる重要な蛋白質として知られている。CaN は単独では触媒ドメインが塞がれているが（図1上）、カルモジュリン(CaM)とカルモジュリン結合部位(CaMBD)が結合することによって CaMBD が folding を起こし、結果的に活性を獲得する（図1下）。



蛋白質の性質を調べる手段として、分子動力学(MD)は非常に有用な手法である。しかし蛋白質は多谷構造を持つ自由エネルギー地形(FEL)によって特徴付けられる複雑な系であり、これらの谷は高い自由エネルギーバリアによって隔てられている。この

ため通常の全原子(AAM)-MD はいわゆる *kinetic trapping* と呼ばれる問題に悩まされることになる。

このような問題を解決するため、拡張アンサンブル法が開発されてきた。例えばレプリカ交換法[1]、マルチバーリック-マルチサーマル法[2]、マルチカノニカル法[3]、Wang-Landau 法[4]、MetaD[5]などである。他にも反応座標の特定の部分をサンプルするアンブレラ法[6]や、反応座標の特定の経路に沿ってサンプルするストリング法[7]などがある。

MD を高速化するもう一つの方法は疎視化モデル(CGM)の活用である。CGM は複数の原子を一つの質点として扱う簡易的なモデルであり、精度に劣るが速度に勝る。近年では AAM の精度と CGM の速度を両立させるマルチスケールシミュレーション(MSS)が注目を集めている。MSS は主に三つのタイプに分けることができる。一つ目は AAM と CGM を同時に用いて蛋白質をモデリングする混成モデルである。二つ目は、まず AAM-MD を行い、その結果を用いて CGM の力場を改良して CGM-MD を行うタイプである。force-matching, fluctuation-matching 法が有名である。三つ目は、まず CGM-MD を行い CGM 構造を取得し、その構造情報を元にして AAM-MD を実行するタイプである。

筆者は以前の研究で三つ目のタイプの MSS をマルチカノニカル法と組み合わせ、エネルギー空間上でのサンプリング効率を 10 倍にした[8]。本研究では Multi-scale Divide-and-conquer MD (MSDC-MD) という新しい手法を提唱した。具体的にはアンブレラサンプリングを用い、エネルギー空間の代わりに主成分空間上で MSS を実行した。結果として CaMBD の FEL の算出に成功し、本来単独では得られない CaMBD の天然構造のサンプリングに成功した。また CaMBD が IDP である理由を熱力学的に解明した。

2. 蛋白質と手法

2.1 CaM-CaMBD の複合体

CaMBD は 24 残機の長さであり、CaN の天然構造内において無秩序化しているが、CaM と結合することで α -helix 構造を取ることがわかっている (PDB ID 4Q5U)。図 2 に 4Q5U の構造を示す。

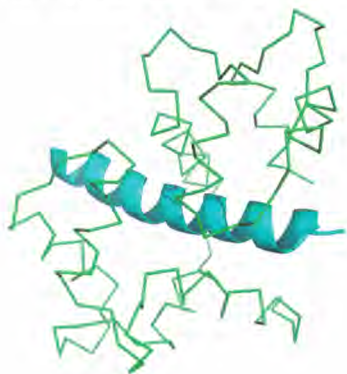


図 2 :
4Q5U の構造。
CaMBD (シアン)
CaM (緑)

2.2 MSDC-MD の手続き

MSDC-MD の手続きは主に 4 つに分けることができる (図 3)。

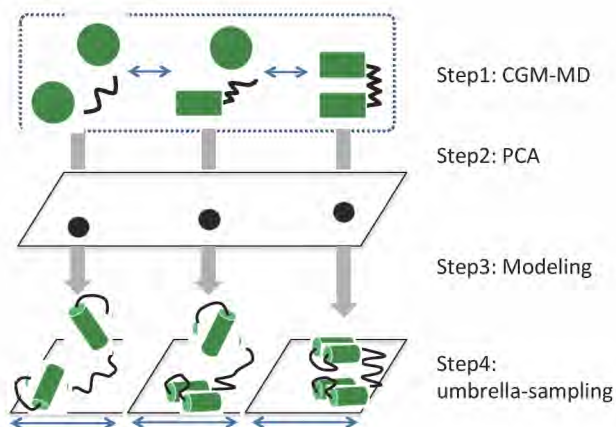


図 3 : MSDC-MD の手続き概要

(ステップ 1) 広い構造空間をサンプルするため、CGM-MD を実行する。本研究では UNRES server を用いた。UNRES は残機あたり二つの相互作用サイトしか持っていないため AAM に比べて 1000-4000 倍早い MD が可能であるとされている。FF2 力場を用いた Langevin Dynamics を 100ns 相当行なった。構造は 1ns 毎に取得した。初期構造は完全に伸びきった構造から始めた。なるべく広い範囲の構造をサンプルするため、T=290, 300, 320, 340, 400 の CGM-MD を行なった。

(ステップ 2) ステップ 1 の CGM トラジェクトリーを一纏めにし、CGM 構造の $C\alpha$ 位置に位置高速をかけながら、主成分分析 (PCA) にかけた。回転と並進の影響を取り除くため、RMSD フィッティングを T=290, 1ns の構造に対して行なった。N-,C-末端の揺らぎが大きいため、フィッティングは 5-19 残機に対して行なった。RMSD フィッティングは以下のように書ける。

$$\begin{aligned} \mathbf{r}'_i &= R(\mathbf{r}_i - \mathbf{Q}) \quad \text{for } i = 1 \dots 24 \\ \mathbf{Q} &= \frac{1}{15} \sum_{i=5}^{19} \mathbf{r}_i \\ R &= (A^t A)^{1/2} A^{-1} \end{aligned}$$

ここで、 $\mathbf{r}_i = (x_i, y_i, z_i)$ は元の座標、行列 R は 3×3 の回転行列であり、行列 A は以下のように書ける

$$A_{ij} = \sum_{k=5}^{19} (r_{ki} - Q_i)(s_{kj} - P_j) \quad \text{for } i, j = 1 \dots 3$$

$$P = \frac{1}{15} \sum_{i=5}^{19} \mathbf{s}_i$$

ここで s_{kj} は T=290, 1ns の座標である。フィッティング後の座標 $\mathbf{X}' = (\mathbf{r}'_1, \dots, \mathbf{r}'_N)$ を PCA にかけることで固有ベクトルと固有値を算出した。構造の第 n 固有値に対する射影は $v_n = \sum_{i=1 \dots 3N} c^{(n)}_i \cdot X'_i$ である。ここで $\mathbf{c}^{(n)}$ は X' の分散共分散行列の第 n 固有ベクトルであり、 $(A)_{ij} = \langle (X'_i - \langle X'_i \rangle) \cdot (X'_j - \langle X'_j \rangle) \rangle$ ($\langle \dots \rangle$ は熱力学平均) に対して以下の方程式を満たす。

$$\Delta \mathbf{c}^{(n)} = \lambda_n \cdot \mathbf{c}^{(n)}$$

(ステップ 3) v_1, v_2 平面上に射影した構造を「小範囲」に分割した (図 4 点線)。各小範囲に含まれる CGM 構造を MODELLER プログラムによって AAM 構造へ復元した。この際 $C\alpha$ 構造に位置拘束をかけ、CGM 構造の情報が失われないようにした。なお説明のために各小領域には図 4 のように番号を振ってある。

(ステップ 4) 複数の AAM-MD を復元した AAM から始めた。使用プログラムは改造版 GROMACS で、温度 300K、圧力 1atm になるよう Berendsen 法を用

いた。各 MD は 90ns 程度行なった。この際以下のアンブレラポテンシャルを加えた。

$$\phi(v_m) = \begin{cases} \frac{1}{2}k(v_m^{\min} - v_m)^2 & \text{for } v_m < v_m^{\min} \\ 0 & \text{for } v_m^{\min} \leq v_m \leq v_m^{\max} \\ \frac{1}{2}k(v_m - v_m^{\max})^2 & \text{for } v_m^{\max} < v_m \end{cases}$$

ここで $m = 1$ or 2 , $k = 0.7\text{kJ/mol/nm}^2$, v_m^{\min} と v_m^{\max} は「サンプリング領域」の上限と下限である。サンプリング領域は隣の小範囲と重複するよう設定されており、例えばサンプリング領域 1、2 のパラメータは $v_1^{\max(1)} - v_1^{\min(2)} = 5\text{\AA}$ と設定されている。このため、サンプリング領域 1、2 の相対重率は $v_1^{\min(2)} \leq v_1 \leq v_1^{\max(1)}$ に含まれるサンプル数 N_1 と N_2 から $\omega_{12} = \log(N_2) - \log(N_1)$ と求めることができる。

ここで力は以下のように書ける。

$$-\nabla\phi(v_m) = \begin{cases} k(v_m^{\min} - v_m)\nabla v_m & \text{for } v_m < v_m^{\min} \\ 0 & \text{for } v_m^{\min} \leq v_m \leq v_m^{\max} \\ k(v_m - v_m^{\max})\nabla v_m & \text{for } v_m^{\max} < v_m \end{cases}$$

R と Q は $X_i = (r_1, \dots, r_N)$ に依存するため微分される。

$$\begin{aligned} \frac{\partial v_m}{\partial X_i} &= \sum_{j=1}^N \sum_{k=1}^3 c_{3j+k}^{(m)} \frac{\partial X'_{3j+k}}{\partial X_i} \\ &= \sum_{j=1}^N \sum_{k=1}^3 c_{3j+k}^{(m)} \left\{ \sum_{l=1}^3 \frac{\partial R_{kl}}{\partial X_i} (X_{3j+l} - Q_l) \right. \\ &\quad \left. + R_{kl} (\delta_{i,3j+l} - Q_l) + R_{kl} \left(X_{3j+l} - \frac{\partial Q_l}{\partial X_i} \right) \right\} \end{aligned}$$

3. 結果と議論

まず CaMBD の CGM-MD を行い PCA で解析した。構造の散布図が図 4 である。ここで重要なのは小領域 17 の天然構造 (大きな+マーク) が CGM 構造によって囲まれている点である。これらの CGM は天然構造に似通っているため、AAM に復元しても天然構造に似た構造になると期待できる。これらの AAM 構造は天然構造をサンプルする上で有益な初期構造になる。ただし CaMBD が IDP であるため、このまま AAM-MD を実行しても全ての構造は無秩序化してしまう。そこでアンブレラポテンシャルを

用いて CaMBD をサンプリング領域内に閉じ込め、無秩序化を防ぎ天然構造のサンプリング確率を大幅に底上げする。

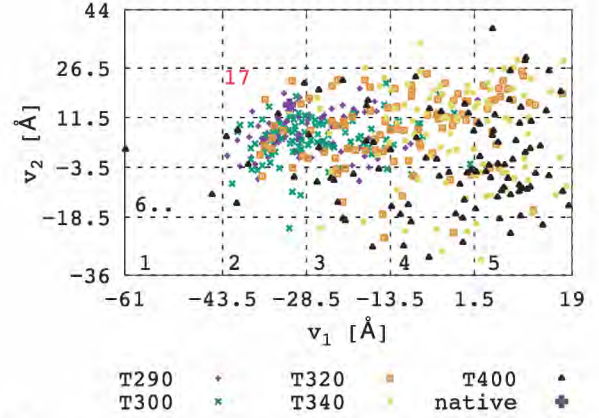


図 4 : CGM-MD の構造の散布図

主成分分析の結果得られた固有ベクトルを図 5 に示す。第一固有ベクトルは折れ曲がり、第二固有ベクトルはねじれ運動に対応していた。

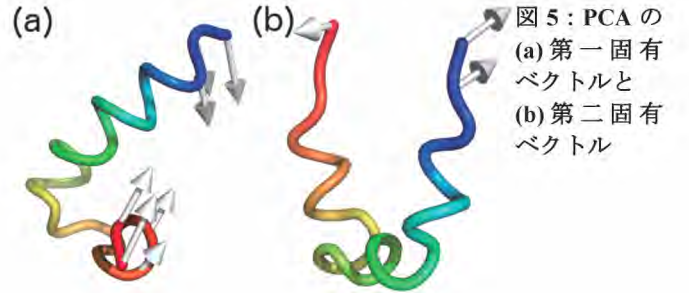


図 5 : PCA の (a) 第一固有ベクトルと (b) 第二固有ベクトル

各小領域にある CGM 構造を AAM 構造に復元し、AAM-MD を同時並列的に行なった。全体の FEL は各サンプリング領域の重率から再重率化を行うことで算出した (図 6)。比較のために通常の AAM-MD で得られた FEL を図 7 に示す。図 7 は $v_1 \leq -13, v_2 \leq -4$ の範囲においてサンプルが明らかに足りていないが、図 6 ははっきりと FEL の形状が見える。

CaMBD の天然構造は $-43.5 \leq v_1 \leq -28.5, 11.5 \leq v_2 \leq 26.5$ のくぼみに対応するが、図 6 からはそれよりも低い多谷構造が観測された。これは CaMBD が IDP である熱力学的理由であると言える。

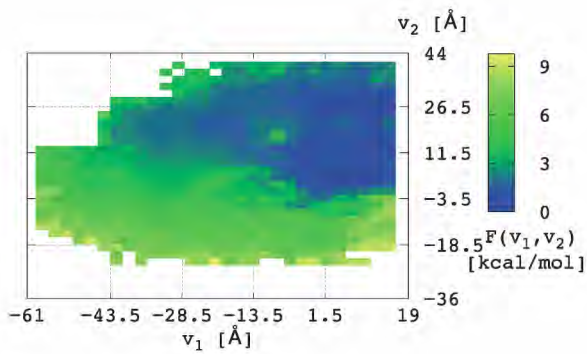


図 6 : MSDC-MD で得られた FEL

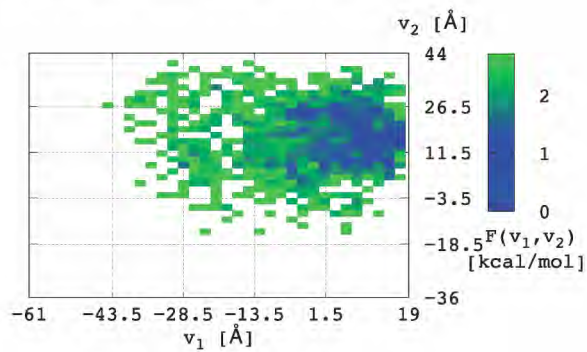


図 7 : 通常の AAM-MD で得られた FEL

また MSDC-MD の精度を比較するため、1 次元の FEL を算出した。MSDC-MD の結果と通常の AAM-MD 結果は統計誤差の範囲で一致していた。

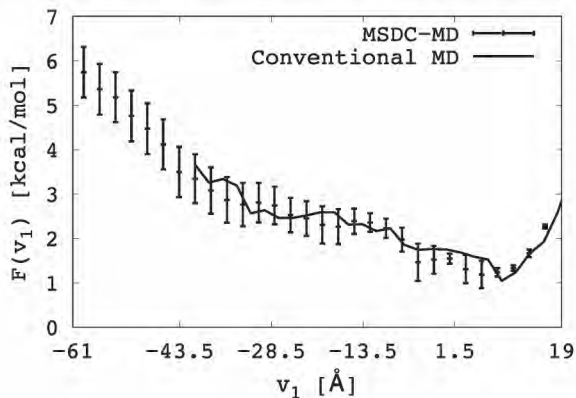


図 8 : 1 次元 FEL の比較。

サンプリングの効率を比較するため、TM スコアの時系列を比較した。TM スコアは二つの蛋白質の構造類似性を表す指標であり 0~1 の値をとる。1 に近いほど互いの構造は似ている。本研究では図 2 の CaMBD 構造に対する MSDC-MD (サンプリング領域 17) の構造の類似性と、同じく AAM-MD の構造の類似性の時系列を解析した (図 9)。

図 9 では上下軸の時間スケールが異なることに注意していただきたい。通常の AAM-MD では 400 ナノ秒の計算を行っても TM スコアは 0.1~0.3 程度にとどまっている。一方で MSDC-MD では 0.3~0.6 の間を 90 ナノ秒の間に頻りに往復している。これはアンブレラポテンシャルが無秩序化を防ぎ、秩序構造のサンプリング確率を上げている結果であると考えられる。

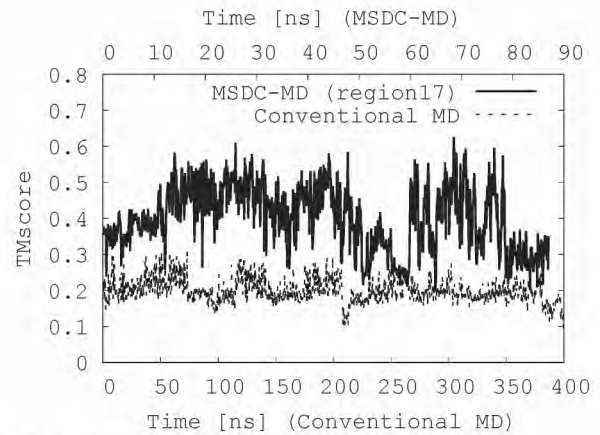


図 9 : TM スコアの時系列

本来 IDP であるはずの CaMBD で TM スコア~0.6 という値をとったのは驚くべき結果である。その構造を図 10 (a)に示す。短い MSDC-MD を CaM にも行い、CaN と CaM のドッキングを行った結果が図 b である。4Q5U に対する RMSD は 7.8 であったが、大まかな位置関係は図 2 と類似していた。

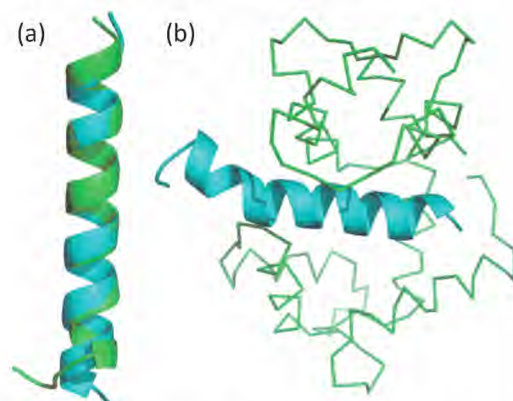


図 10 : (a)MSDC-MD の構造 (緑) と実験構造 (シアン)。 (b)ドッキングの結果。色は図 2 に同じ。

4. 終わりに

MSDC-MD は AAM の精度と CGM の速度とアンブレラサンプリングを組み合わせた手法である。本研究で MSDC-MD は IDP の秩序構造のようなレアイベントのサンプリングを効率的に行うことができることを示した。また自由エネルギーを算出することで、CaMBD がなぜ IDP なのかということも明らかにした。

本手法の利点は、本来二つの蛋白質を同時に扱わなければならないところを、別々の MSDC-MD に分けて計算し、ドッキングによって一つに統合できる点である。多体問題に対する計算コストを大幅に削減できると考えている。今後は本手法をさらに複雑なタンパク質に応用することで、生体反応の問題解明に役立てていきたいと考えている。

参考文献

- (1) Sugita Y. and Okamoto Y. *Chem. Phys. Lett.* **314**, 141–151 (1999).
- (2) Okumura H and Okamoto Y. *Chem. Phys. Lett.* **383**, 391–396 (2004).
- (3) Hansmann, U. H. E. and Okamoto Y. *J. Comput. Chem.* **14**, 1333–1338 (1993).
- (4) Wang, F. & Landau, D. P. *Phys. Rev. Lett.* **86**, 2050–2053 (2001).
- (5) Barducci, A. et al. *Phys. Rev. Lett.* **100**, 20603 (2008).
- (6) Kokubo, H. et al. *J. Comput. Chem.* **32**, 2810–2821 (2011).
- (7) Maragliano, L. et al. *J. Chem. Phys.* **125**, 24106 (2006).
- (8) Shimoyama, H. et al. *J. Chem. Phys.* **133**, (2010).