

タンパク質のフラグメント分割に基づく折り畳み計算の効率化

鈴木正昭, 奥田洋司

東京大学人工物工学研究センター

{msuzuki,okuda}@race.u-tokyo.ac.jp

タンパク質の機能同定のために、迅速な立体構造解析がポストゲノムの重要な研究課題となっている。天然構造探索計算の高速化のため、局所アミノ酸配列-構造相関を考慮した多次元レプリカ交換法を提案し、 β ヘアピンペプチドを用いてその有効性を検討した。

1 はじめに

タンパク質分子は生理環境下でアミノ酸配列に応じて特有の立体構造に折り畳まれ、また、その構造は生化学的機能と密接に関係するため重要である。折り畳みシミュレーションによる構造予測においては、「タンパク質の天然構造は系の自由エネルギー最小状態に対応する」という Anfinsen の主張に基づき、MD や GA 等を用いた最適化計算により天然構造探索を行う。アミノ酸配列情報のみを用いた構造予測は未だ計算科学における最もチャレンジングな問題の一つである。

構造探索の高速化にあっては、一つ一つの構造に対するエネルギー関数評価の高速化と、無数の準安定状態の中から効率良く天然構造を見出す戦略とが必要となる。著者らはこれまで、前者に関しては、生体分子動力学計算コードのベクトル並列機向け最適化およびその地球シミュレータ上での性能評価を実施している[1]。また、後者に関しては、タンパク質における局所アミノ酸配列-構造相関の存在を考慮し局所構造予測能の向上に注目したフラグメントレプリカ交換法(FREM)を提案している[2]。本稿では、局所構造予測能と大域構造予測能を同時に担保する、FREM と従来のレプリカ交換法(REM)を組み合わせた改良 REM についてその有効性を述べる。

2 手法

REM[3]では、異なる温度を持ち互いに相互作用しないオリジナルの系のレプリカを複数用意する。各々のレプリカについて温度一定シミュレーションを独立かつ同時に進めながら途中で温度を交換することで、準安定状態への捕縛を回避し、広く構造空間を探索可能である。FREM ではタンパク質に対し各々数残基からなるフラグメントを

定義する。一定間隔ごとランダムに選択されるターゲットフラグメント F に対してフラグメントのエネルギーのみに基づき局所的なレプリカ交換計算を行うことを繰り返す。予測対象が α ヘリックスを多く含む場合特に少数のレプリカで効率的に予測が可能であることを確認している。

一方、配列上で離れたアミノ酸間の相互作用が支配的である β シート・リッチなタンパク質の構造予測には、大域構造探索能が要求される。そこで、FREM におけるフラグメントの長さを温度に続く二つ目のレプリカ・パラメータとしてFREMを2次元REMへ拡張し、従来REMと組み合わせることを考える。以下、拡張したFREMをMFREM(Multiple Fragment-size REM)と呼ぶ。MFREMの概略を図1に示す。MFREMは、フラグメント長の短いレプリカ群により局所構造の重点的な予測を、フラグメント長の長いレプリカ群により大域構造の安定化を図るものである。

MFREMでは、レプリカに一对一に対応するハミルトニアンが次式のように定義される。

$$E_{m,\zeta}(\mathbf{q}^i) \equiv S_m \sum_{k \in \text{fragment } F} E_{AA}^k(\mathbf{q}^i) + \sum_{k \in \text{fragment } F} E_{AA}^k(\mathbf{q}^i) \quad (1)$$

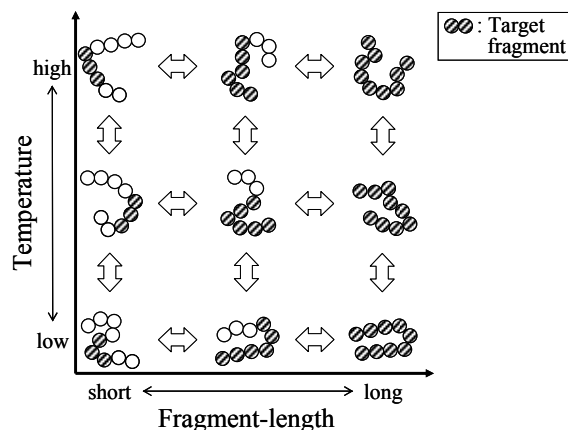


図1. MFREMの概略図。

ここで、 $E_{AA}^k(\mathbf{q}^i)$ はレプリカ i ($i=1, \dots, N_{\text{replica}}$)の k 番目 ($k=1, \dots, N_{AA}$)の残基のポテンシャルエネルギー、 N_{ζ}^l はフラグメント長 ($\zeta=1, \dots, N_{\text{rep}}^l$: N_{rep}^l はフラグメント長方向のレプリカ数)、 S_m はスケーリング因子 ($m=1, \dots, N_{\text{rep}}^T$: N_{rep}^T は温度方向のレプリカ数)である。このとき、温度方向に隣り合うレプリカ i と j の交換が起こる確率 $w(\mathbf{x}^i | \mathbf{x}^j)$ は式(2), (3)で表される。

$$w(\mathbf{x}^i | \mathbf{x}^j) = \begin{cases} 1 & \text{for } \Delta \leq 0 \\ \exp(-\Delta) & \text{for } \Delta > 0 \end{cases} \quad (2)$$

$$\Delta \equiv \beta \left\{ \begin{aligned} & [E_{m(i),\zeta}(\mathbf{q}^j) + E_{n(j),\zeta}(\mathbf{q}^i)] \\ & - [E_{m(i),\zeta}(\mathbf{q}^i) + E_{n(j),\zeta}(\mathbf{q}^j)] \end{aligned} \right\} \quad (3)$$

ここで β は逆温度である。また、フラグメント長方向に隣り合うレプリカ I と J の交換が起こる確率 $w(\mathbf{x}^I | \mathbf{x}^J)$ は、同様に、式(2), (4)で表される。

$$\Delta \equiv \beta \left\{ \begin{aligned} & [E_{m,\zeta(I)}(\mathbf{q}^J) + E_{m,\eta(J)}(\mathbf{q}^I)] \\ & - [E_{m,\zeta(I)}(\mathbf{q}^I) + E_{m,\eta(J)}(\mathbf{q}^J)] \end{aligned} \right\} \quad (4)$$

3 数値実験と考察

β ヘアピン構造を最安定構造に持つ10残基のペプチドを用い、構造探索性能比較を行った。レプリカ数は8とし、各レプリカは200-700Kの温度をとる。特にMFREMに関して、 $(N_{\text{rep}}^l, N_{\text{rep}}^T) = (2, 4)$ 、フラグメント長を4および10残基とした。AMBER96力場を用い、1nsの計算を実施した。各手法より得られた200Kにおけるポテンシャルエネルギーヒストグラムを図2に示す。FREMとMFREMとを比較すると、両手法とも従来REMより低いポテンシャルエネルギー構造をサンプリ

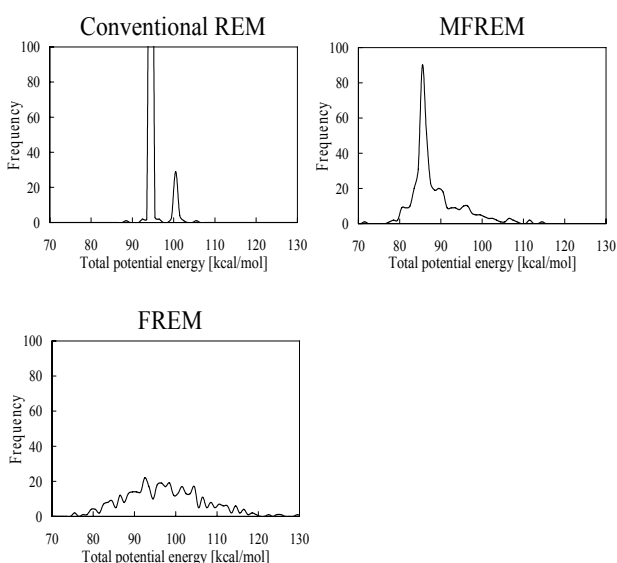


図2. 各REMから得られた200Kにおけるポテンシャルエネルギーヒストグラム。

ングしているが、その頻度はMFREMの方が高い。図3に各REMから得られた200Kにおける構造を示す。FREMに対し、MFREMでは同時に分子の大域構造を大きく変化させるようなダイナミクスも生じ得ることで、結果として、より効率的に β ヘアピン構造を得ることに成功している。

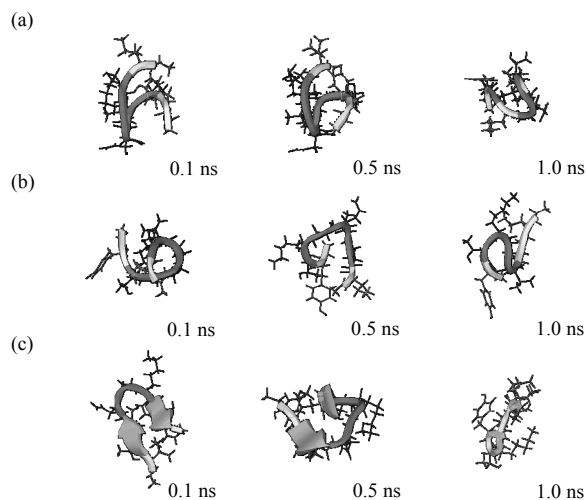


図3. 各REMから得られた200Kにおける β ヘアピンペプチドの構造 ((a) 従来REM, (b) FREM, (c) MFREM)。

4 まとめ

局所構造予測能と大域構造予測能を同時に担保する、局所アミノ酸配列-構造相関を考慮した多次元レプリカ交換法を提案し、その有効性を示した。今後、粗視化モデルに基づくレプリカ群の導入による精度と計算時間の両立などが考えられる。レプリカ交換法をベースとするこれらの手法は計算粒度が大きく柔軟な負荷分散も可能であり、グリッドを含む大規模並列計算機環境で高い実行性能の達成が可能なアプリケーションである。

参考文献

- [1] M. Suzuki and H. Okuda, "Acceleration of biomolecular dynamics simulations on the Earth Simulator", International Journal of Computational Methods (accepted)
- [2] 鈴木正昭, 奥田洋司, 「タンパク質立体構造予測のための改良レプリカ交換法」, 日本応用数学会論文誌 16 卷, 3 号, 255-264, 2006
- [3] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding", Chemical Physics Letters 314, 141-151, 1999