

SCMS2.0によるタンパク質ポテンシャルエネルギー最小化の諸条件における評価

篠崎 隆宏^{1,a)} 関嶋 政和^{2,b)}

概要：タンパク質立体構造の ab initio な予測法としてこれまでに Slice Chain Max Sum 2.0 (SCMS2.0) アルゴリズムを提案した。SCMS2.0 についてのこれまでの評価はポリアラニンに限定されていたが、新たに実際のタンパク質のアミノ酸配列を用いた諸条件におけるエネルギー最小化結果について報告する。

1. はじめに

タンパク質の立体構造はその機能と深く関わる。このためアミノ酸配列から立体構造を予測するための様々な方法が研究されている。中でも既存の類似タンパク質構造情報を用いないアブイニシオ (ab initio) 構造予測は最も挑戦的な試みである。アブイニシオ構造予測は大きく分けてニュートン力学により原子運動を追跡する分子動力学 (MD) に基づいた方法と、分子のポテンシャルエネルギーにより定まるギブス分布をマルコフ連鎖モンテカルロ (MCMC) 法などによりサンプリングし探索する確率モデル法がある。タンパク質のような巨大分子では非熱平衡初期状態から熱平衡達成までに長い時間がかかり MD 法では原子運動追跡が困難となるのに対し、確率モデル法はギブス分布を直接探索することから効果的な適用が期待される。しかし実際には探索効率に問題があり、アブイニシオなタンパク質立体構造予測は未だ実現していない。

このような背景のもと、我々はこれまでに確率モデル法において効率的な探索を実現することを目的として Slice Chain Max-Sum (SCMS) 法およびその改良法である SCMS2.0 [1] の提案を行った。これまでの評価実験ではポリアラニンを構造予測の対象とし MCMC 法に対する優位性を示したが、本研究では実際のタンパク質のアミノ酸配列に SCMS2.0 を適用しその有効性について検討する。

2. 確率モデル法による立体構造予測

構造予測対象の分子を含むシステムのエネルギー E がシ

ステム状態を表す N 個の変数の組 $X = \{x_1, x_2, \dots, x_N\}$ により規定されるとする。温度 T が一定のカノニカルアンサンブルでは、システムが状態 X をとる確率密度は式 (1) に示すギブス分布により与えられる。

$$P(X) = \frac{1}{Z} \exp(-\beta E(X)). \quad (1)$$

ここで β は逆温度、 Z は確率密度を X の全ての値について積分したときに値が 1.0 となるための分配関数と呼ばれる正規化項 $\int_X \exp(-\beta E(X))$ である。熱平衡状態におけるシステムの自由エネルギーは $-\log Z$ に等しい。

予測対象の分子構造が N 個の変数のうちのはじめの M 個の変数の組 $Y = \{x_1, x_2, \dots, x_M\}$ により規定されるとすると、その確率密度は式 (2) により与えられる。

$$P(Y) = \frac{1}{Z} \int_{x_{M+1}, x_{M+2}, \dots, x_N} \exp(-\beta E(X)). \quad (2)$$

予測対象の分子が構造 Y をとる確率密度 $P(Y)$ は、 Y が与えられた条件でのシステムの自由エネルギーを $Z(Y) = \int_{x_{M+1}, x_{M+2}, \dots, x_N} \exp(-\beta E(X))$ と表すことにすると、 $P(Y) = \frac{Z(Y)}{Z}$ と書き直せる。すなわち確率密度 $P(Y)$ を最大とする分子構造は、自由エネルギー $-\log(Z(Y))$ を最小化する分子構造である。

原子座標を q 、運動量を p として、システムのエネルギーをハミルトニアンの形式で表すと運動エネルギー $K(p)$ とポテンシャルエネルギー $V(q)$ の和となる。分子構造に関わるのは q のみである。もしシステムが構造予測対象の分子のみから構成され Y が q と等しい場合、 $P(Y)$ の最大化はポテンシャルエネルギーの最小化と等しい。

3. SCMS2.0 アルゴリズム

SCMS では式 (3) に示す分子ポテンシャルエネルギーを仮定し、その最小値を与える原子配置を探索する。

¹ 東京工業大学
Tokyo Institute of Technology, Kanagawa 226-8502, Japan

² 東京工業大学
Tokyo Institute of Technology, Tokyo 152-8552, Japan

a) <http://www.ts.ip.titech.ac.jp>

b) <http://www.bio.gsic.titech.ac.jp>

$$V(\mathbf{r}) = \sum_{b \in \mathcal{B}} k_b (d_b^{eq} - d_b)^2 + \sum_{a \in \mathcal{A}} k_a (\theta_a^{eq} - \theta_a)^2 + \sum_{d \in \mathcal{D}} k_d (1 + \cos[n\phi_d - \delta_d]) \quad (3)$$

ここで \mathcal{B} , \mathcal{A} および \mathcal{D} はそれぞれ原子結合集合, 結合角集合, 2面角集合である. 式中で各集合 \mathcal{B} , \mathcal{A} , \mathcal{D} に関する項はそれぞれ結合長 d_b , 結合角 θ_a , および 2面角 ϕ_d の関数となっており, k_b , d_b^{eq} , k_a , θ_a^{eq} , k_d , δ_d はそれらのパラメータである. 式 (3) は多数の要素ポテンシャルエネルギーの和となっており, ループを含む因子グラフとして表現できる. SCMS の最初のステップでは, その因子グラフを 3次元空間内での初期原子座標を手がかりにループを含まない線形構造のグラフに変換する. グラフの変数ノードは複数の原子の配置を表す連続変数であるが, これを次のステップにおいてサンプリングにより K 個の離散値で近似する. グラフ全体では変数ノード数 L に対して指数関数オーダー個 K^L の立体構造が表現されるが, max-sum アルゴリズムを用いることでそれら全ての中で最小のエネルギーを与えるものを線形オーダーコスト $O(K^2L)$ で探索することが主なアイデアである. 得られた原子配置を次のエポックの初期状態としてこのプロセスを繰り返し, 収束が得られたらその原子配置を最終結果として出力する.

SCMS2.0 は SCMS のポテンシャルエネルギーに Lennard-Jones ポテンシャルを追加することで原子間力を扱えるようにするとともに, 変数ノードでのサンプリングや max-sum による探索過程に準ニュートン法による最適化を導入する改良を行ったものである. これにより提案分布の分散を大きくとることが可能となり, 探索効率が大きく向上した.

4. 実験条件

Protein Data Bank において公開されているヒトオキシヘモグロビン (1HHO, 141 残基) [2] を対象とし, SCMS2.0 によるエネルギー最小化実験を行った. 比較手法として, 準ニュートン法による最適化付きの MCMC を用いた. 探索の初期構造として用いたのは, X 線結晶解析により決定された分解能 2.1 Å の原子座標である. ポテンシャルエネルギーとして考慮したのは結合長, 結合角, 2面角, および Lennard-Jones ポテンシャルである. 力場パラメータには Amber99SB を用いた. MCMC および SCMS2.0 におけるサンプリング時の提案分布は $[-2.0, 2.0]$ Å の一様分布とし, ファンデルワールスカの cutoff は 9.0Å とした. また, SCMS2.0 におけるスライス間隔は 50Å, ノード毎のサンプル数は 5 とした. 使用した計算機の CPU は, Intel Core i7 (3.20GHz) である.

5. 実験結果

図 1 に温度 500K で探索を行った場合の結果を示す. 横

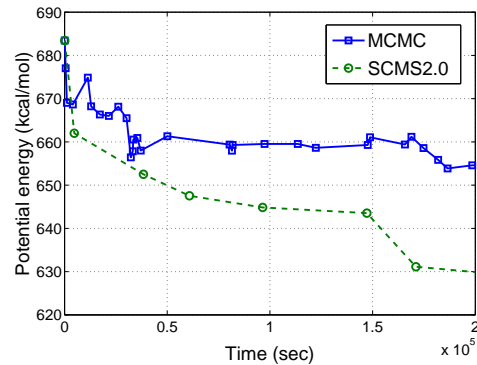


図 1 Potential energy minimization with 500K.

表 1 Potential energy minimization with various temperatures.

Temperature (K)	Method	Energy (kcal/mol)	CPU time (sec)
500	MCMC	653.85	2.0E5
500	SCMS	631.12	1.7E5
1000	MCMC	648.51	2.0E5
1000	SCMS	625.84	2.0E5
2000	MCMC	683.43	2.0E5
2000	SCMS	642.70	2.0E5

軸が計算時間, 縦軸が分子のポテンシャルエネルギーである. SCMS2.0 の方が最適化付き MCMC と比較して同じ計算時間でより低いエネルギーの原子配置を得ていることが分かる.

表 1 に温度を変えたときのエネルギー最小化結果を示す. エネルギーは, 探索時間が 2.0E5(sec) に一番近いエポック数まで探索を行った範囲での最小値である. いずれの温度でも SCMS2.0 の方が MCMC よりもよい結果となっている. 最小のエネルギー値は 500K の条件で得られた.

6. まとめと課題

SCMS2.0 を実際のタンパク質のアミノ酸配列に対して適用し, 最適化付き MCMC と比べて同じ計算時間でより低いポテンシャルエネルギーを与える原子配列を探索できることを示した. 今後の課題としては探索アルゴリズムの改良や溶媒効果のモデル化などがあげられる.

7. 謝辞

本研究は科研費 23650068 の助成を受けたものである.

参考文献

- [1] Naoto, I., Takahiro, S., Shiqiao, D., Sadaoki, F. and Masakazu, S.: A Study on the potential energy minimization of proteins by Slice Chain Max-Sum algorithm, *IPSJ SIG technical reports*, Vol. 2012, No. 20, pp. 1–8 (2012).
- [2] Shaanan, B.: Structure of human oxyhaemoglobin at 2.1 Å resolution, *Journal of Molecular Biology*, Vol. 171, No. 1, pp. 31–59 (1983).