

タンパク質エネルギー地形における構造多様性と 多次元エネルギー地形の新しい可視化手法

星野恭子^a, 松永康佑^a, Gareth J. Lyllance^b,
Roy L. Johnston^b, David J. Wales^c, 小松崎民樹^{ad}
a 神戸大学理学部, b Birmingham 大化学,
c ケンブリッジ大化学, d JST/CREST

1 Introduction

「エネルギー地形」という観点は、自由度の大きな系の動力学や熱力学的性質を調べる上で、重要な鍵となるものである [1–3]。それらの系の動力学、熱力学的振る舞いは、 N 個の原子からなる系の場合、 $3N$ 個のカルテシアン座標の関数として表されるポテンシャルエネルギー面 (Potential Energy Surface: PES) により決定されると考えられる。しかしながら、PES を基に自由度の大きな系の物理的性質を議論するには、二つの困難な問題がある。まず第一に、原子の数が増えるにつれて、一般に固定点の数が指数関数的に増加する [4]。また第二に、多次元の PES の特徴を、問題の本質的側面を正確に捉えたまま低次元で可視化することが困難である [5]。高次元の PES を低次元で可視化する方法として現在最も有効であると思われる方法は、Becker と Karplus [6] により提案された disconnectivity graph という手法である。disconnectivity graph は、サンプルされた minimum 構造と one rank saddle 構造のデータベースを基に構成される。ここで、one rank saddle とはヘシアンが負の固有値を一つだけ持つ固定点のことを言う。離散化された各エネルギー値において、データベース内の構造は、superbasin と呼ばれる各集合に分類される。superbasin とは、データベース内の one rank saddle 構造の情報を基に、各エネルギー値以下で移ることの可能な構造の集合として定義される。disconnectivity graph では、この superbasin をグラフにおける node として表現し、エネルギー値を上下させることで見えてくる superbasin 間の包含関係を link として表現する。このようにして出来上がった disconnectivity graph は、PES の大域に渡った構造間の「繋がり」に関する情報を与えてくれる。本報告の目的は、PES の有効な近似を行うために必要な自由度の数を調べるとともに、また得られた計量的な指標を disconnectivity graph に盛り込むことで、多次元 PES の新たな可視化手法を提案することにある。我々が用いたモデルは、Honeycutt と Thirumalai [7] によって提案された、3 種類のビーズ 46 個からなる BLN モデルと呼ばれる粗視化タンパク質モデルである。BLN モデルは global minimum 構造周りでフラストレーションを持つように設計されており、効率的には折れ畳まないことが知られている [9]。もう一つのモデルは、BLN モデルの global minimum 構造においてコンタクトをとっていないビー

ズ間、つまり、ある一定距離以上に離れているようなビーズ間のすべての引力を取り除くことによって作られた Gō-like BLN モデルと呼ばれるモデルである [8]。Gō-like BLN モデルは、global minimum 構造を保持するのに整合性のある相互作用のみを残しているので、BLN モデルに比べてよりフラストレーションの小さい PES を持っている。

2 Method

今回我々は、既に先行研究 [9] により得られている、BLN モデルに対して 500 点の minimum 構造と 636 点の saddle 構造、Gō-like BLN モデルに対しては 520 点の minimum 構造と 844 点の saddle 構造からなるデータベースを用いた。以後、2 種類の構造を総称して単に固定点と呼ぶことにする。これらのデータに対して、まず McLachlan の 'best fit' アルゴリズム [10] を使って全体の並進と回転の自由度を取り除く。

1. 重心が原点にセットされた座標系において、データベース内の N_{SP} 個の構造の集合 $\{\mathbf{q}(k)\}$ における集団平均 $\langle \mathbf{q} \rangle$ で定義される「参照構造」を計算する。ここで、 $q_i(k)$ は k 番目の固定点の i 番目のカルテシアン座標。($i = 1, 2, \dots, 3N, k = 1, 2, \dots, N_{\text{SP}}$)
2. 構造 $\mathbf{q}'(k)$ からなる新しい集合をつくる。 $\mathbf{q}'(k)$ は、 $\mathbf{q}(k)$ と $\langle \mathbf{q} \rangle$ との差異の和 s を最小化することで得られる。

$$s(k) = \frac{1}{2} \sum_{i=1}^{3N} (q_i(k) - \langle q_i \rangle)^2. \quad (1)$$

3. $\{\mathbf{q}'(k)\}$ に対して新たな集団平均構造を計算する。
4. 新たな集団平均構造を新たな参照構造として再定義し、集団平均構造の変化がある閾値内に収束するまでステップ 2 に戻る。

今後、 $\mathbf{q}(k)$ を並進回転を取り除いた構造空間とする。 $\mathbf{D}$ を $3N \times N_{\text{SP}}$ 行列とする。その要素 D_{ik} は、 k 番目のサンプル構造において（質量により重み付けした） i 番目のカルテシアン座標 $q_i(k)$ の平均構造 $\langle q_i \rangle$ からのずれとして定義する。すなわち、

$$D_{ik} = \Delta q_i(k) = q_i(k) - \langle q_i \rangle, \quad (2)$$

ここで

$$\langle q_i \rangle = \frac{1}{N_{\text{SP}}} \sum_{k=1}^{N_{\text{SP}}} q_i(k), \quad (3)$$

であり、 $3N \times 3N$ の分散共分散行列 $\mathbf{R}$ は以下のように定義される。

$$\mathbf{R} = \frac{1}{N_{\text{SP}}} \mathbf{D} \mathbf{D}^T. \quad (4)$$

ここで、 $\mathbf{D}^T$ は $\mathbf{D}$ の転置行列である。 $\mathbf{R}$ を対角化する固有ベクトルなる行列 $\mathbf{U}$ をつかって主成分 Q_i を定義する。

$$\mathbf{R} \mathbf{U} = \mathbf{U} \mathbf{r} \quad (\mathbf{U}^T \mathbf{U} = \mathbf{I}), \quad (5)$$

$$Q_i = \sum_{j=1}^{3N} u_{ij} \Delta q_j. \quad (6)$$

対角化された行列 $\mathbf{r}$ の i 番目の要素である固有値 r_i は、 i 番目の集団座標である主成分 Q_i の分散と等しい。普通、 r_i は大きい順に並べ替えられている ($r_1 \geq r_2, \dots, \geq r_{3N}$)。この線形変換は主成分解析 [11] と言われ、データの揺らぎ、あるいは分布を最もよく表わすための線形の集団座標を決定する。我々は、PES をより低次元で近似的に表現するために上記のデータベースに対して主成分解析を行う。主成分から元のカルテシアン座標へは、次のように再現される。

$$\Delta \mathbf{q} = \mathbf{U}^{-1} \mathbf{Q}. \quad (7)$$

ここで、 $\mathbf{U}\mathbf{U}^{-1} = \mathbf{U}^{-1}\mathbf{U} = \mathbf{I}$ なので、

$$\Delta q_i = \sum_{j=1}^{3N} u_{ji}^{-1} Q_j = \sum_{j=1}^{3N} u_{ij} Q_j \approx \sum_{j=1}^{N_r} u_{ij} Q_j. \quad (8)$$

このようにして、分散の大きい第 1 主成分から $N_r (\leq 3N - 6)$ 番目の主成分までを用いて $\mathbf{q}$ を近似することができる。得られた近似構造を元に、PES を再構成する。

3 Results and Discussion

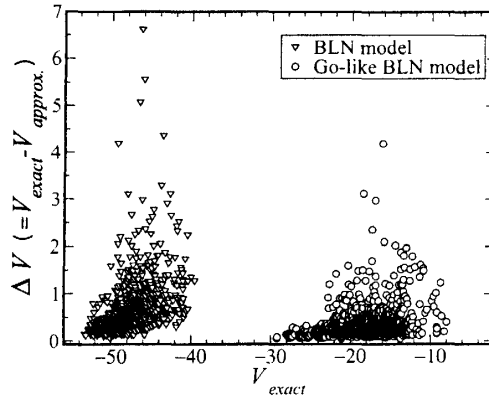


図 1: 92 主成分を用いた V_{approx} と正確な V_{exact} とのずれ

図 1 に、両モデルのデータベース内の全固定点について、 $N_r = 92$ (全自由度の 3 分の 2) として計算した近似ポテンシャルエネルギー V_{approx} と正確なポテンシャルエネルギー V_{exact} の差異を示した。両モデルとも、固定点のエネルギーが増加するにつれ、近似したエネルギーと正確なエネルギーとのずれも大きくなる。特にエネルギー的に global minimum に近い固定点に対して、オリジナルのポテンシャルよりも、Go-like BLN ポテンシャルのほうがエネルギーのずれが小さく、近似において取り除いた分散の小さい 46 個の主成分に対して比較的頑健である。また、エネルギーと同様に、様々な基準で測った構造の違いもまた global minimum 付近で、BLN モデルよ

りも Gō-like BLN モデルのほうが小さいことが他の解析により判明した。

構造多様性

ここで、各エネルギー V における構造多様性を計る二つの量を導入する。一つ目の指標 $\sigma_s(V)$ 、は固定点の分布の分散で定義される。

$$\sigma_s(V) \equiv \frac{1}{N_{\text{SP}}(V)} \sum_{k \leq V} \sum_i^{3N} (q_i(k) - \langle q_i \rangle)^2. \quad (9)$$

ここで、 $N_{\text{SP}}(V)$ は任意のエネルギー V 以下の固定点の総数である。二つ目の指標 $D_s(V)$ 、はエネルギー V 以下の固定点の総分散を 99% 再現するのに必要とされる主成分の総数として定義する。図 2 はそれぞれのモデルの global minimum を基準とした V に対して $\sigma_s(V)$ (左)、 $D_s(V)$

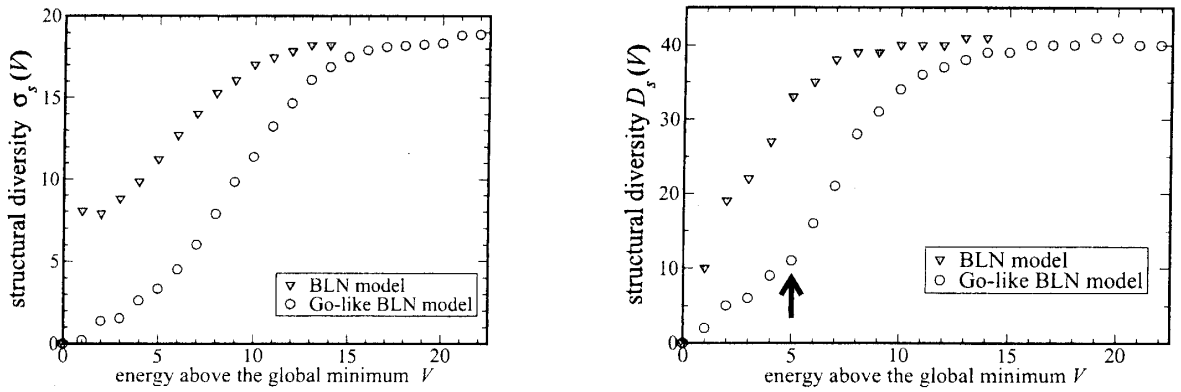


図 2: 構造多様性を表わす指標、 $\sigma_s(V)$ (左) と $D_s(V)$ (右)

(右) をプロットしたものである。構造多様性を表はかる二つの量は BLN モデルのほうが Gō-like BLN モデルよりも大きい。BLN モデルでは、 $V = 1\epsilon$ (ここで ϵ はモデルポテンシャルにおける Lennard-Jones 相互作用の谷の深さ) 付近で、一つ目の指標 $\sigma_s(V)$ は不連続に増加するが、二つ目の指標 $D_s(V)$ はむしろスムーズに増加する。一方で、 $D_s(V)$ は BLN モデルよりも Gō-like BLN モデルにおいて、構造分布の次元性に関する重要な変化を示している。BLN モデルにおいて $D_s(V)$ は global minimum から、ある収束する値まで単調に増加しているのに対し、Gō-like BLN モデルにおいては、 $D_s(V) \simeq 10$ まで増加が緩やかである。この $D_s(V) \simeq 10$ 以下となるエネルギー領域は、固定点のエネルギーと構造の両方が主成分空間の自由度の 3 分の 1 を取り除いたことによる影響をあまり受けない領域に相当する。

新しい disconnectivity graph

ここで、これまで用いた解析手法や概念を用いて Introduction で説明した disconnectivity graph を拡張する方法を提案する。まず、disconnectivity graph の各エネルギー V_i における node の横軸の座標には、その node が表わす superbasin 内に含まれる全固定点の第 1 主成分の平均の値を用いる。その上で、前項で述べた構造多様性の指標の一つ、すなわちある superbasin m に属する固定

点の分布の全分散を 99% 再現するのに必要とされる主成分の数 $d_s(m)$ を新たに導入する。図 3 に、

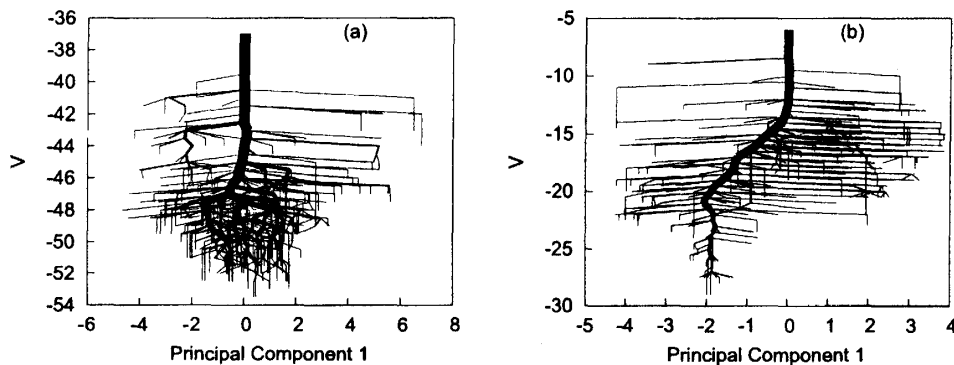


図 3: 直線の幅を $d_s(m)$ で定義した新しい disconnectivity graph。(a)BLN モデル、(b)Gō-like BLN モデル

これらの計量的な指標を盛り込んだ新たな disconnectivity graph を示す。図 3 では、link の線の太さがそれぞれの superbasin 内の構造多様性 $d_s(m)$ を表わしている。線の太さは、 $d_s(m)$ に比例して変化する。図 3 は、Gō-like BLN モデルの単一ファネル型エネルギー地形と、深い superbasin を多く含む BLN モデルのフラストレーションの大きいエネルギー地形との違いを、明らかにしている。図 3(b) の Gō-like BLN モデルでは、単一のファネル地形であることを意味する一本の線が顕著である。一方で、図 3(a) の BLN モデルでは、複数の太い直線が描かれており、複数の拮抗する superbasin が存在することがみてとれる。特に、それらの superbasin は、global minimum 付近においても、まだ構造空間に広く分布していることが新しく導入した指標によりわかる。反対に、Gō-like BLN では、global minimum 付近では、線が細くなり、その構造多様性も急激に減少している。この分布の小さい領域は、固定点のエネルギー（図 1）と構造の両方が分散の小さい方から 3 分の 1 の影響をあまり受けず、 V が小さくなるにつれ、superbasin の構造多様性が減少するような領域（図 2）と、完全に一致する。

4 Conclusions

我々は固定点間の計量関係をできるだけ保存するように多次元エネルギー地形を少数自由度空間に投影する新しい disconnectivity graph を提案し、タンパク質が持つエネルギー地形に関して構造多様性という概念を新規に導出した [12]。構造多様性は、端的に表現すると、PES を構成する固定点が $3N$ 次元空間においてどれくらい広範囲の多次元配位空間に分布しているかを評価するものである。我々は進化の所産として獲得された（とされる）ファネル型エネルギー地形をもつ Gō-like BLN モデルでは、階層的なエネルギー地形を有する BLN モデルに比べて、少数自由度に縮約可能であるエネルギー領域が global minimum 構造近傍に存在することを見出したが、そのような領域における構造多様性が顕著に低いことを解明した。この結果は、ファネル型エネルギー地形が、ポテンシャルにおける摂動に対して、global minimum 構造をより頑健に保持する可能性を示唆している。

参考文献

- [1] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, *Proc. Natl. Acad. Sci. USA* **1995**, 92, 3626.
- [2] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins: Struct., Func. Gen.* **1995**, 21, 167.
- [3] D. J. Wales, *Energy Landscape: With Applications to Clusters, Biomolecules and Glasses*, Cambridge University Press, Cambridge, **2003**
- [4] F. H. Stillinger and T. A. Weber, *Science* **1984**, 225, 983.
- [5] J. P. K. Doye, *Phys. Rev. Lett.* **2002**, 88, 238701.
- [6] O. M. Becker and M. Karplus, *J. Chem. Phys.* **1997**, 106, 1495.
- [7] J. D. Honeycutt and D. Thirumalai, *Biopolymers* **1992**, 32, 695.
- [8] H. Nymeyer, A. E. García and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 5921.
- [9] M. A. Miller and D. J. Wales, *J. Chem. Phys.* **1999**, 111, 6610.
- [10] A. D. McLachlan, *J. Mol. Biol.* **1979**, 128, 49.
- [11] O. M. Becker, J. A. D. MacKerell, B. Roux and M. Watanabe, editors, *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc., New York, **2001**.
- [12] T. Komatsuzaki, K. Hoshino, Y. Matsunaga, G. J. Rylance, R. Johnston, D. J. Wales, *J. Chem. Phys.* **2005**, 122, 84714.