

# 進化実験系の解析・評価技術・適応歩行技術の研究：中立ネットワークの探査と進化戦略の検討

## An *in silico* Exploration of the Neutral Network in Protein Sequence Space and Evolution Strategy

相田拓洋<sup>1,\*</sup>、伏見譲<sup>2</sup>

Takuyo Aita<sup>1,\*</sup>, Yuzuru Husimi<sup>2</sup>

<sup>1</sup> ノバルティスファーマ（株）筑波研究所、<sup>2</sup> 埼玉大学工学部機能材料工学科

<sup>1</sup>Novartis Pharma K.K., <sup>2</sup>Saitama University

### 概要

Considering that many amino acid sequences that fold into a common main chain structure form a percolation network in protein sequence space, we call the network "neutral network" for the structure. The neutral evolution is regarded as a walk on the neutral network. If a network for a structure A and a network for another structure B intersect each other in a special region in a sequence space, the walker can exchange one network to another network. That is, the evolving protein can change its shape from structure A to B and vice versa by accepting several mutations. Schultes *et al.* confirmed the existence of the neutral paths in the RNA sequence space in their experiment (Schultes & Bartel, 2000). For four distinct target structures ( $\alpha$ ,  $\beta$ ,  $\alpha/\beta$  and  $\alpha + \beta$ -types) with the same chain length of 108, we investigated whether there is the corresponding neutral network in the sequence space by using computer simulation. An exploring walk simulation suggested that the neutral network spans over the sequence space. Through another exploring walk simulation, we investigated contiguous regions between or among the neutral networks for the distinct protein structures and obtained the results that the closest approach distance between the two neutral networks ranged from 5 to 29 on the Hamming distance scale.

KEYWORDS: Experimental molecular evolution, Protein conformation, Sequence space

### 1 はじめに

進化分子工学は、機能性生体高分子の分子設計に関して、既存の合理的設計法である蛋白質工学とは異なり、ダーウィニズムに従う進化的手法を用いた設計法であり、現在様々な分野で広く活用されている。新規機能を有するタンパク質の設計に関しては、中立ネットワークを利用する方法が提唱されている。あるタンパク構造にフォールドするアミノ酸配列を、当該タンパク質構造の"中立配列"と呼ぶ。多数の中立配列が配列空間中でパーコレーションネットワークを形成している場合、このネットワークを"中立ネットワーク"と呼ぶ。分子進化は、中立ネットワーク上

の歩行と見なせるので、異なった中立ネットワーク間の位置関係が重要である。<sup>123</sup>

### 2 中立配列と中立セット

配列長が 108 であるが、構造が互いに異なる 4 種類のタンパク質、CH-domain (PDB コード:1bkr: $\alpha$  型)、IG variable domain (PDB コード:1lvfa: $\beta$  型)、thiore-

<sup>1</sup>Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST) 2-41-6 Aomi, Koto-ku, Tokyo 135-0064 Japan

<sup>2</sup>Email: aita.algo@aist.go.jp

<sup>3</sup>Phone: +81-3-3599-8057

doxin (PDBコード:2trxA: $\alpha/\beta$ 型)、barnase (PDBコード:1a2pA: $\alpha+\beta$ 型)の各構造に関して、中立配列を1300配列ずつ発生させた。ここで、中立配列の選択に関しては、経験的ポテンシャルを用いたエネルギー計算を行い、ある閾値以下のエネルギーをとる配列を中立配列とした。結果として得られた中立配列間のハミング距離は90-100程度であり、互いに配列空間の遠方に位置している。図1は、中立配列のセットから求めたプロファイルである。全体的な特徴として、グリシンとプロリンは2次構造 $\alpha$ ヘリックスや $\beta$ ストランドの境界付近に存在し、一方、アラニンやロイシンやフェニルアラニンは2次構造領域内に存在する傾向にある。さらに、同じ構造に関する中立配列セットに関して、任意の2つの中立配列間には、中立配列によるパス、すなわち、中立経路が存在することが分かった。すなわち、任意の中立配列から他の中立配列に、2配列を並置したときに対応するアミノ酸置換を順次導入することで、配列を遷移させることができる。ただし、このときパスの数(=アミノ酸置換の導入順序の場合の数)は $108!$ 通り存在するが、全てのパスが中立経路であるとは限らない。結果として、今回探索された配列空間において、中立ネットワークがパーコレートしていることが分かった。このネットワークの頂点の次数、すなわち中立配列の周囲に存在する中立配列の数は、周囲に存在する全ての配列のうちの5%–50%である。

### 3 異種ネットワーク間の接近

上述の4種類の蛋白構造の各々に関して、配列空間で固有の中立ネットワークを形成しているが、その異種ネットワーク間の接近について調べた。ここで、最接近歩行法を用いた。この歩行は、ネットワークA上の任意の点とネットワークB上の任意の点を、各自のネットワーク上を辿りながら、互いに接近させる歩行法である。この結果を図2に示した。結果として、ハミング距離5–29程度で接近していた。さらに、4種のネットワークA, B, C, D上の任意の点を、各自のネットワーク上を辿りながら、4点を互いに接近させる歩行を行った。その結果を図3に示した。結果として、4つのネットワークが最も接近するとき、そのインターチェンジはハミング距離20–30程度のボールである。これらのハミング距離は、蛋白全長の108に比べて小さな値である。一方、実験で扱える配列のオーダーは、 $10^{10} \sim 10^{12}$ 程度であり、これをハミング距離5のボール内に存在する全配列数( $\binom{108}{5}18^5 \approx 2 \times 10^{14}$ )に比べても少ない。よって、今

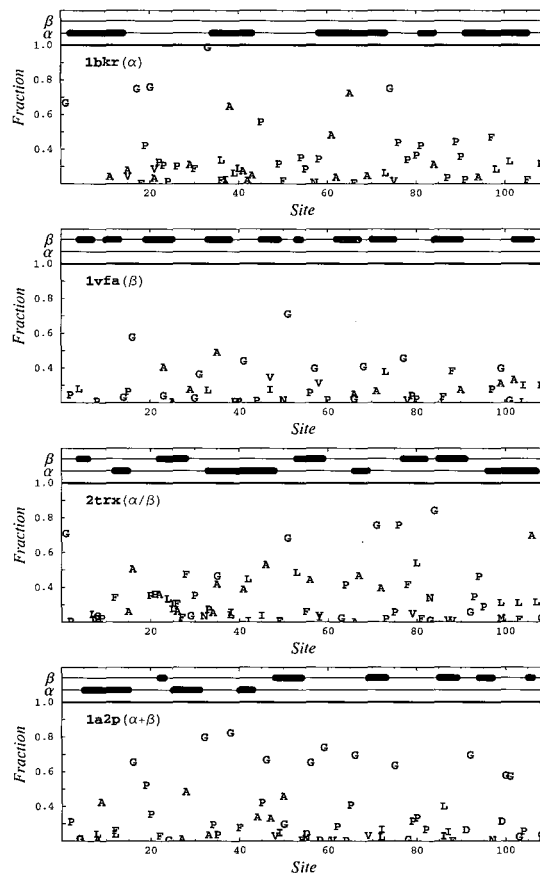


Figure 1: 中立配列のセットから求めたプロファイル。アミノ酸を1文字表記した。図の上を示した太線は、2次構造領域を示す。

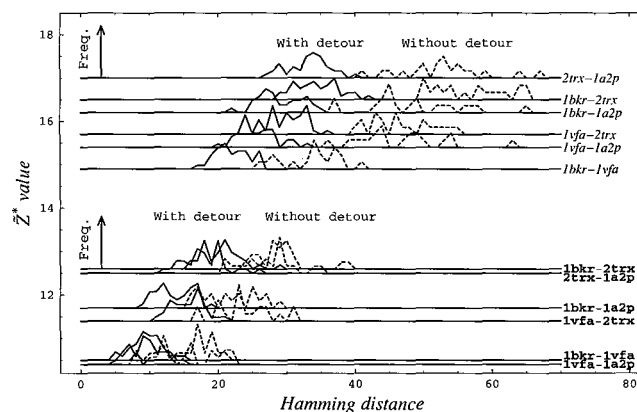


Figure 2: 2つの異種ネットワーク間の接近距離の分布。 $Z^*$ は中立配列を選択するときのエネルギー $Z$ スコアの基準値。黒の実線が迂回を許した場合。点線が迂回を許さない場合である。

回得られたハミング距離は、実験分子進化のインターチェインジになるほど小さな距離ではないと思える。しかしながら、実際問題としては、類似構造のネットワーク間の接近が重要であろう。この場合は、接近距離がより近くなると思えるので、インターチェインジを介してネットワーク間を乗り換えることで、劇的な進化が期待できる。

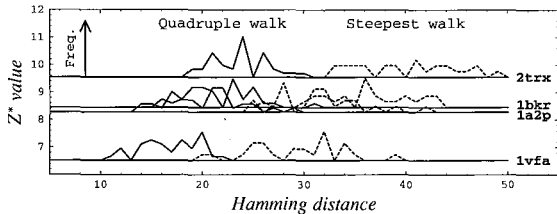


Figure 3: 4つの異種ネットワーク間の接近距離の分布。Zは中立配列を選択するときのエネルギーZスコアの基準値。黒の実線が最接近歩行の結果。点線がランダム配列からの最急勾配歩行の結果である。

## 4 謝辞

本研究は、通産省産業科学技術研究開発制度「加速型生物機能構築技術」の一環として、新エネルギー産業技術総合開発機構(NEDO)の委託(ノバルティスファーマ)およびバイオテクノロジー開発技術研究組合の再委託(埼玉大)を受けて、実施したものである。

## References

- [1] Babajide, A., Hofacker, I.L., Sippl, M.J. & Stadler, P.F. (1997). Neutral networks in protein space: a computational study based on knowledge-based potentials of mean force. *Folding & Design* **2**, 261-269.
- [2] Babajide, A., Farber, R., Hofacker, I.L., Inman, J., Lapedes, A.S. & Stadler, P.F. (2001) Exploring protein sequence space using knowledge-based potentials. *J.theor.Biol.* **212**, 35-46.
- [3] Blanco, F.J., Angrand, I. & Serrano, L. (1999) Exploring the conformational properties of the sequence space between two proteins with different folds: an experimental study. *J. Mol. Biol.* **285**, 741-753.
- [4] Fontana, W. & Schuster, P. (1998). Continuity in evolution: on the nature of transitions. *Science* **280**, 1451-1415.
- [5] Huynen, M.A., Stadler, P.F. & Fontana, W. (1996) Smoothness within ruggedness: the role of neutrality in adaptation. *Proc.Natl.Acad.Sci. USA* **93** 397-401.
- [6] Schultes, E.A. & Bartel, D.P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448-452.
- [7] Schuster, P. (1995). How to search for RNA structures: Theoretical concepts in evolutionary biotechnology. *J.Biotech.* **41**, 239-257.
- [8] Schuster, P. & Fontana, W. (1999). Chance and necessity in evolution: lessons from RNA. *Physica D* **133**, 427-452.