

遺伝的アルゴリズムの分子実装によるタンパク質進化

モデリング研究系 知的情報モデリンググループ
助教 染谷 博司

1 はじめに

報告者らは、タンパク質の進化過程を「タンパク質構造の“最適化”」とみなし、Sakamoto et al. (2005), 染谷他 (2007) にて、遺伝的アルゴリズム (GA) の分子実装によるタンパク質工学への接近法を提案している。ここでタンパク質工学とは、望ましい活性 (性質) を有する非天然タンパク質を開発することを指す。タンパク質工学は、医療・製薬・食品工学・環境科学など広い範囲で現在もその重要性を増しており、また、進化過程への理解を深める側面においても重要な研究課題である。Someya et al. (2009) では、(i) *in vitro* な実験環境における GA の分子実装とその数世代のタンパク質進化、および、(ii) ランダムウォークの確率モデルに基づく分子進化のプロトタイプモデリングとその計算機シミュレーション、を報告した。本稿ではこれらの報告内容を概説する。本稿の各図および記述の一部は Someya et al. (2009) から引用している。

なお本研究は、電子計算機以外にて GA を実装した一例としても、また、DNA Computing の一例としても、意義のある有用性を主張しうる点に関連研究との特に大きな違いがあり、将来発展への展望がある。

2 タンパク質工学と確率的最適化

タンパク質は固有のアミノ酸配列によって表現される。したがって、アミノ酸配列の組合せ空間を探索空間とみなすと、望みの活性を有するタンパク質の開発は、組合せ最適化問題のひとつとみなすことができる。しかし探索空間の広さは広大であり、また、そのほとんどは解として無意味 (infeasible) な空間であると考えられているため、極めて困難な最適化問題であると考えられる。そこで、本研究では、獲得を目指す新規タンパク質の改変元となる野生型タンパク質の周辺に探索領域をしばり、なおかつ、多峰性の適応度景観においても解の探索が可能な最適化手法が有効であると考え、GA の分子実装による接近法を提案している。概念図を Fig. 1 に示す。

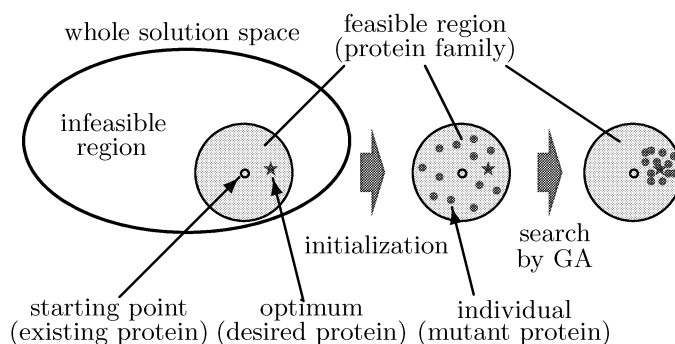


Figure 1: A diagram for the proposed search.

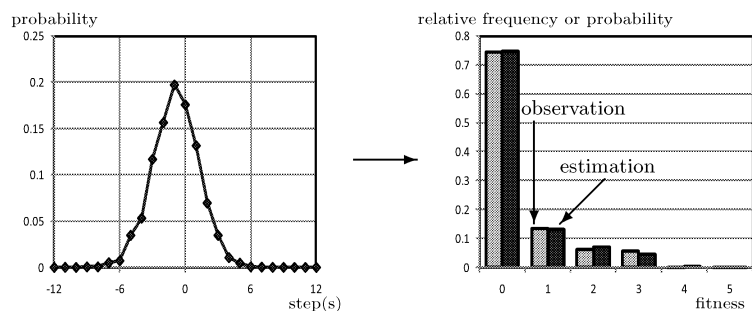


Figure 5: The left graph is an example of the probability distribution of the random walk steps in the probabilistic model, and the right one is the estimated fitness distribution derived from the one on the left. The model parameters, $f_{wt} = 15.9$, $\delta = 1.3$, $p_a = 0.4$ and $p_d = 0.6$, were determined by a roughly-discretized-complete search performed using Kullback-Leibler information as the fitting measure.

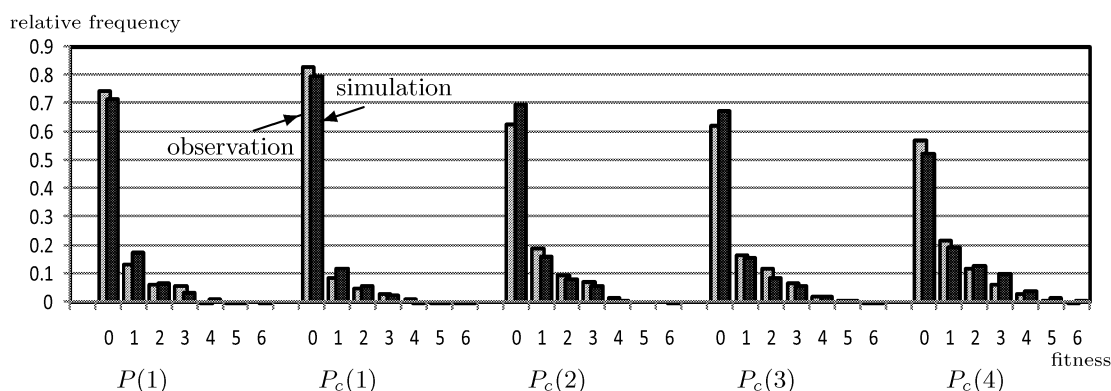


Figure 7: Comparison of the fitness evolution observed in the biological experiments and that which resulted from a computer simulation.

3 GAの分子実装によるタンパク質進化と確率モデル構築の試み

分子生物学的実験においては適応度分布の観察により着実な進化が確認された。また計算機シミュレーションでは、単純なランダムウォークに基づく確率モデルを採用しながらも、分子生物学的実験と比較的よく一致し、今後の進展への期待を抱くことができるものであった。Fig. 5に確率モデルの例を示す。Fig. 7に分子生物学的実験とシミュレーション実験の比較例を示す。前者の実験は一試行分のみであり、また、比較対象となりうる広く知られた既存の確率モデルは見当たらないため、主観的評価にならざるを得ないが、比較的よく一致していると感じられる。今後は、分子生物学的実験をさらに進め、より高い活性を有する多様な新規タンパク質の獲得を目指す。また、アミノ酸間の依存関係などの表現を確率モデルに取り入れ、より妥当性の高いモデルの構築を目指す。

関連する公表論文

Sakamoto, K., M. Yamamura, and H. Someya (2005). "Toward "Wet" Implementation of Genetic Algorithm for Protein Engineering". In C. Ferretti, G. Mauri, and C. Zandron, (eds), *DNA Computing (Proceedings of 10th International Workshop on DNA Computing)*, volume 3384 of *Lecture Notes in Computer Science*, pp. 308–318. Springer-Verlag Berlin Heidelberg.

Someya, H., K. Sakamoto, and M. Yamamura (2009). "Biologically-implemented Genetic Algorithm for Protein Engineering". In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2009)*, pp. 233–240, Montréal, Canada. ACM SIGEVO.

梁谷博司・山村雅幸・坂本健作 (2007). 「分子計算のための一点から開始される探索法」. 人工知能学会論文誌, **22**(4) pp.405–415.