

# 23pPS-18 ゲノム配列比較とランダムウォーク

北里大 理

守 真太郎

Sequence Alignment and Random Walk Model

Dept. of Phys., KItasato Univ.

S. Mori

遺伝子の配列比較の問題を統計力学、特にランダム媒質中のランダムウォークの問題として定式化し理論的、数値的な解析を行う。配列比較は遺伝子のどこが似ていて、どこが似ていないかを見つける処理のことであり、計算生物学のもっとも基礎的な問題として研究されてきた。特に、ダイナミックプログラミングは、もっとも効率的な解法として知られ、そのアルゴリズムは、2 配列の比較の場合、2 次元空間上での再帰的な手続きに基づいている。ここでは、この問題を統計力学の手法、つまりあらゆる経路での重み付きの和を考え、ランダム系の統計力学の問題と捉える。例えば、一方の長さ  $L = 2$  の遺伝子  $\vec{T} = (T_1, T_2)$ 、 $\vec{S} = (S_1, S_2)$  を考える。ここで、 $T_1, T_2, S_1, S_2$  は  $A, T, G, C$  などの塩基を表している。この二つの遺伝子のアラインメントとは、塩基間にスペースをはさんだりすることによって、遺伝子の類似度を高めることであり、もっとも類似度の高いアラインメントを求めるのが目標となる。スペースについてはペナルティーとして  $h$ 、一致した塩基ペアには  $-J$ 、異なる塩基ペアには  $J$  というスコアを振り、類似度の高さをスコアの低さに置き換える。 $L = 2$  の場合、可能な組み合わせは 13 通りあり、それぞれのボルツマン因子の和として分配関数  $Z_L$  を導入すると

$$Z_2(\beta, h, J, \vec{T}, \vec{S}) = 6e^{-\beta 4h} + 2(e^{-\beta(2h+C_{22})} + e^{-\beta(2h+C_{11})}) + e^{-\beta(2h+C_{12})} + e^{-\beta(2h+C_{21})} + e^{-\beta(C_{11}+C_{22})} \quad (1)$$

となる。ここで、 $C_{ij}$  は塩基配列から決まる係数で

$$C_{ij} = \begin{cases} J & S_i = T_j \\ -J & S_i \neq T_j \end{cases} \quad (2)$$

である。こうして、類似度の高いアラインメントを求める問題は、上記のランダム系の基底状態の探索問題となる。本発表では、 $\vec{T}, \vec{S}$  がランダム変数である場合の系の統計的な振る舞いを DPRM として解析した結果について報告する。

- Introduction to Computational Molecular Biology, J.C.Setubal and J.Meidanis (1996).