



実験室内分子進化過程の熱力学および情報論的解釈

産業技術総合研究所 生命情報科学研究センター

(現在) 埼玉県中小企業振興公社, 埼玉バイオグループ (JST) 相田拓洋

埼玉大学工学部機能材料工学科 伏見 謙

In vitro protein evolution is comprehended as a hill-climbing process on a fitness landscape in sequence space, just like the protein folding is comprehended as a valley-descending process on an energy landscape in conformation space. In this review, based on a mathematical model for *in vitro* molecular evolution, we demonstrated an analogy between evolutionary dynamics and thermodynamics and then many concepts in evolutionary dynamics became comprehensible. By introducing thermodynamics-like concepts such as “free fitness”, it is interpreted that evolution is driven in the direction in which free fitness increases. In this process, the evolving population climbs the fitness landscape by absorbing “fitness information” as the negative entropy from the surroundings as an experimental set-up.

in vitro evolution / fitness landscape / information / thermodynamics / sequence space

1. はじめに

本稿で述べる実験室内分子進化¹⁾とは、配列情報のランダム変異導入 (mutagenesis) と淘汰 (selection) からなるダーウィン進化の原理を、タンパク質や核酸などの生体高分子に適用することで起こる人工的な分子進化のことである。この分野は、1967年の Spiegelman らの実験に端を発し、単純化された進化系の中に生体高分子の創出原理の探求を目指した理学的側面²⁾と、産業や医療などへの貢献を目的とした工学的側面³⁾を車の両輪として進展してきた。しかし、近年では、分子設計の道具としての意義が強調され、この分野が始動した当時に掲げられた「実験科学の視点から進化の本質に迫ろう」とする精神がなおざりにされているように思える。確かに実験室内分子進化が、機能性核酸の創出やタンパク質機能の改変などの道具として多大な成功を取めていることは事実であり、目的とする機能性生体高分子がいとも簡単に得られるように考えられている。現在、当分野における理論研究は、「目的の分子を効率良く得るためには、突然変異率や集団サイズをどのように設定すればよいのか？」などの進化戦略

に関する実験理論研究が主流である。しかし、本稿では、このような実験理論は別の機会にゆずり、実験室内分子進化の当初の精神にたちかえり、より基盤となる生体高分子の情報創出の問題に切り込んでみた。

まず導入として、実験室内分子進化を理論的に捉える上で重要な概念をいくつか紹介する。淘汰過程における評価の指標として表現型分子が与える数値を適応度 (fitness) とよぶ。これは具体的には生体高分子の安定性、触媒活性、リガンド分子との親和性や複製反応速度定数などの物性量であるが、より広義に解釈して、「表現型分子が有する (実験者が要求する) 価値の大きさを数値化したもの」と定義できる。実験室内分子進化の実験系は、進化する実体としての表現型分子あるいは表現型・遺伝子型分子を指す進化系 (evolution system) と、その周囲の外部環境 (surroundings) に分離できる。たとえば、ある受容体タンパク質に結合するペプチドを進化的に創出する場合、ペプチド分子が進化系であり、溶媒の条件も含めてこの受容体分子表面の形状や電荷の分布が外部環境である。ここでの適応度は受容体との結合自由エネルギーに負号を付けたものと定義すればよい。実験者は適応度を適切に評価

An Interpretation of Evolutionary Dynamics in *in vitro* Molecular Evolution from Thermodynamical and Informational View Point

Takuyo AITA^{1,*} and Yuzuru HUSIMI²

¹Computational Biology Research Center (CBRC) National Institute of Advanced Industrial Science and Technology (AIST)

*Present affiliation: Rational Evolutionary Design of Advanced Biomolecules (REDS) Group/JST, Saitama Small Enterprise Promotion Corporation SKIP

²Department of Functional Materials Science, Saitama University

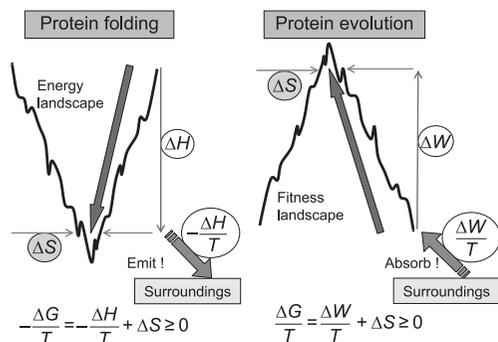


Fig. 1 An analogy between protein folding and protein evolution. In protein folding, a folding polypeptide descends the energy landscape by emitting the thermal entropy $-\Delta H/T (= \Delta S_{sur})$ to the surroundings. In protein evolution, we interpret that the evolving sequences climb the fitness landscape by absorbing "fitness information" $\Delta W/T (= \Delta I_{in})$ as the negative entropy from the surroundings, which is defined by an experimental set-up. (カラー図は電子ジャーナル <http://www.jstage.jst.go.jp/browse/biophys/>参照)

するために、外部環境を適切に設定し制御しなければならない。また、すべての可能な配列 (= 遺伝子型) を包括的に扱うための概念上の空間を配列空間 (sequence space) とよび、各配列はこの空間中の各点で表現される。ここで、任意の2つの配列間の距離はハミング距離 (Hamming distance) で記述するのが一般的である。さらに、配列空間中の各点 x に、対応する適応度 $f(x)$ をプロットしてできるスカラー場を適応度地形 (fitness landscape) とよぶ。

本稿で扱う分子進化とは、高度な知性が存在しない状況下で、単にランダム変異と淘汰を繰り返すだけで、すべての可能な配列の集合から、実験者の設定した外部環境に適合した配列情報を引き出す過程、言い換えると、ダーウィン原理による外部環境からの情報獲得過程^{4),5)}を意味する。この情報獲得過程に関する原理を明確にするため、我々は分子進化の最も単純な系としての「制御された環境下での実験室内分子進化系」に焦点を絞り、実験と理論の両面から研究を行ってきた。

このような単純化された進化系では、生体高分子の進化過程は適応度地形上の山登り (適応歩行, adaptive walk) と表現できる (Fig. 1)。これは、タンパク質のフォールディング過程が、コンホメーション空間中のエネルギー地形上の谷下りと表現できるのと似ている。そこで、進化を既存の熱力学のアナロジーで捉えることで、今まで曖昧だった「外部環境からの情報獲得」に関する諸概念に意味づけができ、概念が体系的に整理できることが期待できる。進化と熱力学のアナロジー

に言及した先駆的研究として、巖佐の自由適応度の理論⁶⁾やEigenらの擬種の理論^{2),4),7)}がある。本総説では単純な数理モデルを解析した最近の我々の成果^{8),9)}を紹介したい。後半では、これまであまり議論されなかった「情報の価値的側面」に関して考察を行う。

2. 適応度地形と進化のモデル

まず、 v 個のアミノ酸残基から成るすべての可能なアミノ酸配列 (20^v種類) を考え、その各々の配列を20値 v 次元配列空間内の点として表現する。各点に当該配列の適応度をプロットしたのが適応度地形である。適応度のモデルとして、各残基が他の k 個の残基の影響を受けると仮定するNKモデル^{9),10)}を用いる。このモデルでは、アミノ酸配列 " $A_1 A_2 \dots A_v$ " の適応度 W が

$$W = \sum_{j=1}^v w_j(A_j | A_{j_1}, A_{j_2}, \dots, A_{j_k}) \quad (1)$$

で与えられる。ここで、 $w_j(A_j | A_{j_1}, A_{j_2}, \dots, A_{j_k})$ は、部位 j_i の残基 $A_{j_i} (i=1,2,\dots,k)$ が与えられたとき、残基 A_j が適応度 W に寄与する「部位適応度」(site-fitness) であり、ある上限と下限をもった一様分布から選んだ乱数とする。さらに、相互作用部位 j_i もランダムに選ばれるとする。式 (1) は、配列空間上に凸凹の重畳した山塊状の地形を形成する。 k の値が小さいほど凸凹の度合いは小さくなる。特に、 $k=0$ 場合は、各残基が独立に適応度に寄与するモデルであり、「富士山型地形モデル」とよばれる⁸⁾。

進化のモデルとして、次のモデルを用いる。 M 個の親配列が N 個の子配列を産出するとし、その N 個の子配列の中から適応度の高い順に M 個の配列を選び次世代の親配列とする。子配列の産出過程においては、ランダムな d 個のアミノ酸置換が生じるとする。すなわち、親配列とその子配列の間のハミング距離は d である。 M と N と d は一定とする。

$k \geq 0$ の場合の理論は $k=0$ の場合の理論に準ずるので、本総説では、 $k=0$ の場合 (富士山型地形モデル) の理論を中心に述べ、最後に $k \geq 0$ に関して言及する。本稿では、セクション3で、進化ダイナミクスを統計的に解析した結果を示し、セクション4で、その結果を熱力学的な変数を導入して解釈する。

3. 進化ダイナミクスの統計的性質

親配列集団の平均適応度を \bar{W} として、この世代変化に注目する。今回のモデルに基づいて適応歩行 (進化) のシミュレーションを行った結果を Fig. 2 に示す。山麓に相当するランダム配列集団から出発した集団は、初期段階では急速に適応度の山に登るが、やがて定常状

態に達する。一方、山頂から出発した集団は山を下り同じく定常状態に達する。この定常状態における平均適応度 \mathcal{W} の世代平均を \mathcal{W}^* とする。この定常状態は、突然変異による配列空間内の拡散効果と、淘汰による配列多様性の収斂効果がバランスすることによって生じるので、この現象を「変異・淘汰・浮動平衡」とよぶ。ただし、配列空間中のある特定の点（＝配列）付近に定在するのではなく、配列空間中で適応度が \mathcal{W}^* の近傍をとる領域（言い換えると、適応度地形を高さ $W = \mathcal{W}^*$ で切断した断面）をランダムに渡り歩くのである。

定常値 \mathcal{W}^* を理論的に求めるに当たり、まず始めに、平均適応度 \mathcal{W} をもつ親配列集団から派生する子配列集団内の適応度の頻度分布を考える。統計数学を用いた解析結果から、親配列の平均適応度 \mathcal{W} からの適応度の変化量 ΔW の集団平均と標準偏差はそれぞれ

$$E[\Delta W|\mathcal{W}] = -\frac{d}{v}(\mathcal{W} - \mathcal{E}) \quad (2)$$

$$SD[\Delta W|\mathcal{W}] \approx \sqrt{\frac{2\gamma d}{v}} = \sigma\sqrt{2d} \quad (3)$$

で与えられ、その分布関数は正規分布に従う（導出は省略⁸⁾）。ここで、 v は配列の長さ、すなわち配列空間の次元数である。 \mathcal{E} と γ はそれぞれ配列空間全体にわたる適応度の平均と分散である。特に \mathcal{E} は、ランダム配列の適応度の期待値でもあるので、適応度地形の山麓における適応度に相当する。式 (2) から、親配列集団が山を登るにつれて、子配列集団内の適応度は親に比して有害になる傾向にあることがわかる。式 (3) の右辺の σ は適応度地形の特性に関するパラメータであり、 $\sigma \equiv \sqrt{\gamma v}$ で定義される。その物理的意味は、適応度地形上である方向にハミング距離 1 だけ移動したときの適応度変化の標準偏差である。（親配列集団の）平均適応度の 1 世代当たりの変化量 $\Delta \mathcal{W}$ を、本総説では“進化速度”とよぶことにする。子配列集団のサイズが $N < \binom{v}{d} 19^d$ の場合は、進化速度 $\Delta \mathcal{W}$ は確率的に記述される。親配列集団が平均適応度 \mathcal{W} をとるときの $\Delta \mathcal{W}$ の期待値と標準偏差は、正規分布に関する極値分布の理論から、それぞれ

$$E[\Delta \mathcal{W}|\mathcal{W}] = E[\Delta W|\mathcal{W}] + SD[\Delta W|\mathcal{W}] \times \zeta \quad (4)$$

$$SD[\Delta \mathcal{W}|\mathcal{W}] \approx \frac{2SD[\Delta W|\mathcal{W}]}{(\sqrt{M+1})\zeta} \quad (5)$$

で与えられる。ここで、 $\zeta (\geq 0)$ は淘汰圧の強さを表すパラメータであり、 N/M が大きな場合には

$$\zeta \approx \sqrt{2 \ln(N/M)} \quad (6)$$

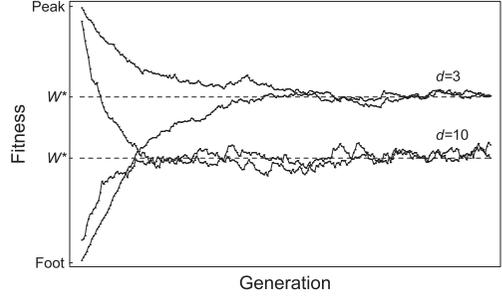


Fig. 2 Time course of the fitness \mathcal{W} in adaptive walks by a single walker. $k = 0$, $M = 1$, $N = 10$, $v = 140$. d -values are shown in the figure.

である。式 (2) と (3) を式 (4) に代入すると、式 (4) は \mathcal{W} の単調減少関数であることがわかる。定常値 \mathcal{W}^* は $E[\Delta \mathcal{W}|\mathcal{W}^*] = 0$ を満たすので、この方程式を解くと \mathcal{W}^* は近似的に

$$\mathcal{W}^* = \mathcal{E} + \frac{\sqrt{\gamma v}}{\tau} \quad (7)$$

と求まる。分母の τ は配列空間の探索に関するパラメータであり、ハミング距離 d と、子集団サイズと親集団サイズの比 N/M の関数である。

$$\tau \equiv \frac{\sqrt{d/2}}{\zeta} \approx \sqrt{\frac{d}{4 \ln(N/M)}} \quad (8)$$

τ の分子は、突然変異による配列空間上の拡散効果を表し、一方、分母は淘汰による収斂効果を表している。さらに、式 (7) から、 τ が小さいと定常点 \mathcal{W}^* は山頂（適応度最大状態）に移行し、一方、 τ が大きいと定常点 \mathcal{W}^* は山麓（エントロピー最大状態）に移行する。よって、 τ は、温度のようなパラメータである。

進化速度の期待値を、“流束 (flux)” の意味を込めて改めて J と表記し、一方、標準偏差を Σ と表記する。定常点 \mathcal{W}^* 近傍では、式 (4) と (5) はそれぞれ近似的に

$$J \equiv E[\Delta \mathcal{W}|\mathcal{W}] = -\frac{d}{v}(\mathcal{W} - \mathcal{W}^*) \quad (9)$$

$$\Sigma \equiv SD[\Delta \mathcal{W}|\mathcal{W}] = \frac{4\sigma\tau}{\sqrt{M+1}} \quad (10)$$

となる（導出は省略⁸⁾）。

ここで補足しておくが、Fig. 2 に表れているように、適応度が減少するのは、次世代の親が「 N 個の子配列から選ばれる」モデルにしたがっているからであり、旧世代の M 個の親も選択の候補に含めるモデルの場合は、多少様相が変わってくる。まず、配列集団の適応度が減少することはなく、定常点 \mathcal{W}^* までは急速に山を登

る。この理由は、山麓の近くでは、歩行者 (=配列) の位置から周囲に延びる多数の道のうちで上り坂の占める割合が比較的多いため、これらの上り坂を見いだしやすいからである (式 (2) 参照)。しかし、適応度の山に登るにつれて上り坂の占める割合は徐々に少なくなり、定常点 \mathcal{W}^* を過ぎると歩行者は上り坂を見いだすことが確率的に困難になる。よって、進化速度 $\Delta\mathcal{W}$ はきわめて小さくなる。直感的に富士山型地形では簡単に山頂に辿り着けるイメージがあるが、パラメータ τ 次第では簡単に山頂にたどり着けるわけではない。

4. 進化ダイナミクスの熱力学および情報論的解釈

4.1 定義

ここから、以上述べた進化のダイナミクスを、熱力学的な変数を導入して解釈してみる (これから定義する変数と既知の熱力学変数との対応を **Table 1** にまとめた)。まず、適応度を W で表す。「独立偶然量の和は正規分布にしたがう」ので、配列空間全体にわたる適応度の頻度分布は以下の正規分布で与えられる。

$$\Omega(W) = \frac{20^v}{\sqrt{2\pi\mathcal{V}}} \exp\left(-\frac{(W-\mathcal{E})^2}{2\mathcal{V}}\right) \quad (11)$$

ある定数 k_L を用いてエントロピーを

$$S(W) = k_L \times \ln\Omega(W) \quad (12)$$

と定義し、さらに、平均適応度 \mathcal{W} をとる配列集団に関する“進化ポテンシャル” μ と“自由適応度” G をそれぞれ

$$\mu(\mathcal{W}) \equiv \mathcal{W} + T \times S(\mathcal{W}) \quad (13)$$

$$G(\mathcal{W}) \equiv \mu(\mathcal{W}) \times M \quad (14)$$

と定義する。ここで T はある定数である。式 (11) と (12) を用いると、 μ は上に凸の 2 次関数

$$\mu(\mathcal{W}) = -\frac{k_L T}{2\mathcal{V}} \left(\mathcal{W} - \left(\mathcal{E} + \frac{\mathcal{V}}{k_L T} \right) \right)^2 + \text{const} \quad (15)$$

となる。ここで $k_L T$ を、

$$k_L T = \sigma \quad (16)$$

を満たすと要請することで、進化ポテンシャル μ および自由適応度 G は $\mathcal{W} = \mathcal{W}^*$ で最大値をとり、かつ、この進化系の Lyapunov 関数となる (式 (28) 参照)。そこで、 k_L と T をそれぞれ次のように定義する。

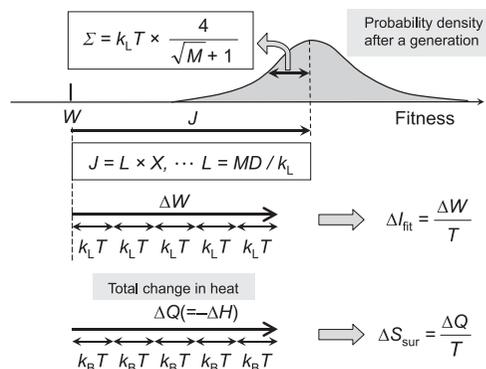


Fig. 3 (Upper) A probability density of the mean-fitness change after a single step walk (=generation) from the mean-fitness \mathcal{W} . J and Σ represent the expectation and standard deviation of the change. (Middle) The fitness-information change, ΔI_{fit} is the digitized mean-fitness change by the unit of uncertainty, $k_L T = \Sigma \times ((\sqrt{M} + 1)/4)$. (Bottom) The concept of ΔI_{fit} is analogous to the thermal entropy, $\Delta S_{\text{sur}} = \Delta Q/T = -\Delta H/T$, which is the digitized heat change by the unit of the fluctuation of energy per particle, $k_B T$. (カラー図は電子ジャーナル <http://www.jstage.jst.go.jp/browse/biophys/> 参照)

$$k_L \equiv \sigma, \quad (17)$$

$$T \equiv \tau \approx \sqrt{\frac{d}{4 \ln(N/M)}} \quad (18)$$

先に述べたように、 τ は温度のような性格のパラメータと解釈できるので、 T を“進化温度”と名付ける。一方、 σ は地形表面の凸凹の程度を表すので、 k_L を“地形定数”と名付ける。

さらに、“進化駆動力” X を次に定義する。

$$X(\mathcal{W}) = \frac{d}{d\mathcal{W}} \frac{\mu(\mathcal{W})}{T} \quad (19)$$

$$= -\frac{k_L}{\mathcal{V}} (\mathcal{W} - \mathcal{W}^*) \quad (20)$$

4.2 進化速度の期待値と標準偏差

再び、進化速度 (=平均適応度 \mathcal{W} の 1 世代当たりの変化量) $\Delta\mathcal{W}$ に着目する (Fig. 3 の上段)。式 (20) を式 (9) に代入すると、進化速度 $\Delta\mathcal{W}$ の期待値 J は定常点近傍において進化駆動力 X と次の線形関係を満たす。

$$J = LX, \quad L = \frac{MD}{k_L} \quad (21)$$

ここで、 L は線形輸送係数または易動度である。 D は配列集団が配列空間をランダム歩行した場合 (これは、 $N = M$ の場合に実現する) の、平均適応度 \mathcal{W} に関する

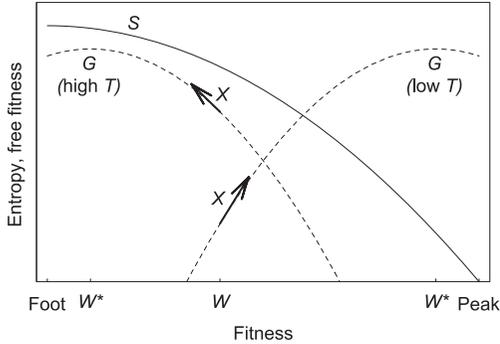


Fig. 4 Thermodynamical interpretation of evolutionary dynamics. The solid line and dashed line represent the entropy S and free fitness G , respectively, as a function of fitness. The right (left) dashed line is for the case of a low (high) evolutionary temperature T . W^* represents the stationary point, in which G takes the maximum. The vector (=arrow) represents the evolutionary force X that acts at the point W . Details are described in the text.

“拡散係数”であり、 $D = \sigma^2 dM$ と与えられる。式 (21) は、ブラウン運動理論におけるEinsteinの関係式とよく似た式になっている。一方、式 (10) で与えられる進化速度 ΔW の標準偏差は、式(16)の関係を用いることで、

$$\Sigma = k_L T \times \frac{4}{\sqrt{M+1}} \quad (22)$$

と書き直せる。 Σ は、1世代当たりの適応度変化の不定定性の度合いを表しており、一方、 $k_L T$ は統計熱力学における「1粒子当たりの熱揺らぎのエネルギーの大きさ」とアナロジーがとれる。したがって、式 (22) は両者の関係を表す式として合理的な結果であるように思える。

以上述べたことに基づいて進化を熱力学的に解釈すると次のようになる (Fig. 4)。集団は進化駆動力 X に作用されて、自由適応度もしくは進化ポテンシャルが最大になる状態に向かって進行する。進化駆動力 X は進化温度 T に依存する。今、集団が山の中腹にいとしよう。 $T = \infty$ の場合 (これは $N/M \approx 1$ の場合に実現する)、定常点 W^* は山麓であるので、負の力 ($X < 0$) が働き、集団は山麓の方向に導かれる。この場合はエントロピー最大状態が実現する。 T が低くなるにつれ定常点 W^* は高くなり、 $T \approx 0$ の場合 (これは $N/M \approx \infty$ の場合に実現する)、定常点は山頂であるので、正の力 ($X > 0$) が働き、集団は山頂方向に導かれる。この場合は適応度最大状態が実現する。ただし、 $T = 0$ の場合 (これは $d = 0$ の場合に実現する) は配列空間内を移動できないので、進化は起こらず単に初期集団内の淘汰が起こる。

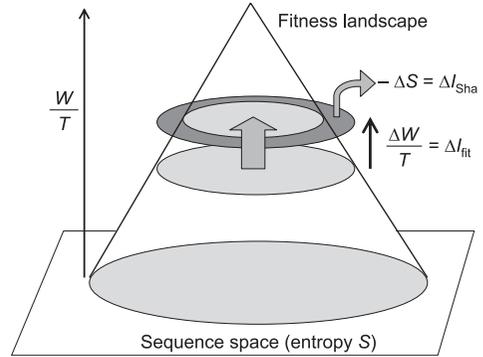


Fig. 5 Change of fitness information I_{fit} and Shannon information I_{Sha} after a single step walk (=generation). $\Delta I_{fit} = \Delta W/T$ is interpreted as a change of the height of the fitness landscape, while $\Delta I_{Sha} = -\Delta S$ is interpreted as a change of the area of the cross section. (カラー図は電子ジャーナル <http://www.jstage.jst.go.jp/browse/biophys/>参照)

(「集団の平均適応度は常に増大する」で有名な Fisher の原理は、淘汰のダイナミクスに限った内容である。)

一方、ウェットの実験においても上記の現象が確認されつつある。大阪大学の四方らのグループは、fd フェージコートタンパク質 (の一部) の感染能をランダム配列から進化させている¹¹⁾。集団サイズを $N = 10$ に固定して進化させたところ、数世代後に感染能がプラトーに達した。そこで、集団サイズを $N = 1000$ 程度に大きくしたところ、再び山を登り始め新たなプラトーに達した。これは、進化温度を低くしたことで、定常点がより高い位置に移り、進化駆動力が再び働きたと解釈できる。

4.3 進化過程で獲得した情報量

今、山麓から出発したポリペプチド集団が定常状態に向かって進化する過程を考える。そこで、次の3つの量を導入する。

$$I_{fit}(W) \equiv \frac{W}{T} - \frac{\mathcal{E}}{T} \quad (23)$$

$$I_{Sha}(W) \equiv S(\mathcal{E}) - S(W) \quad (24)$$

$$I_{bio}(W) \equiv \frac{\mu(W)}{T} - \frac{\mu(\mathcal{E})}{T} \quad (25)$$

$$= I_{fit}(W) - I_{Sha}(W) \quad (26)$$

ここで、 \mathcal{E} は山麓の適応度である。式 (24) の I_{Sha} は (集団が位置する) 高さ W をとる配列空間の測度に関する“Shannonの情報量”とよべる^{5), 12)} (Fig. 5)。一方、式 (22) の関係を用いると、 $\Delta I_{fit} = \Delta W/T$ には次のような物理的な意味がある。

Table 1 An analogy between the thermodynamical system and our evolution system

Thermodynamics	Evolution
Number of moles: M	Number of walkers: M
Molar energy or molar enthalpy: H_m	Mean fitness of walkers: \mathcal{W}
Boltzmann constant: k_B	Landscape constant: $k_L = \sigma$
Molar entropy: $S_m \equiv k_B \ln \Omega$	Walker's entropy: $S \equiv k_L \ln \Omega$
Temperature: T	Evolutionary temperature: $T \approx \sqrt{d/4 \ln(N/M)}$
Chemical potential: $\mu \equiv H_m - TS_m$	Evolutionary potential: $\mu \equiv \mathcal{W} + TS$
Free energy: $G \equiv \mu \times M$	Free fitness: $G \equiv \mu \times M$
Generalized flux: J	Expectation of climbing rate: $J \equiv E[\Delta \mathcal{W} / \mathcal{W}]$
Generalized force: $X \equiv d(\mu/T) / dx$	Evolutionary force: $X \equiv d(\mu/T) / d\mathcal{W}$
Linear transport equation: $J = LX$	Linear relationship between J and X : $J = LX$
Diffusion coefficient: D	Diffusion coefficient: D
Einstein relation: $L = MD/k_B$	$L = MD/k_L$
Mean thermal energy of a particle: $k_B T \times const$	Standard deviation of climbing rate: $k_L T \times 4 / (\sqrt{M} + 1) = SD[\Delta \mathcal{W} / \mathcal{W}]$
$\Delta S_{system} / M = \Delta S_m$	Shannon information gain: $\Delta I_{Sha} = -\Delta S$
$\Delta S_{surroundings} / M = -\Delta H_m / T$	Fitness information gain: $\Delta I_{fit} = \Delta \mathcal{W} / T$
$\Delta S_{universe} / M = -\Delta \mu / T$	Biological information gain: $\Delta I_{bio} = \Delta \mu / T$

$$\frac{\Delta I_{fit}}{k_L} = \frac{\Delta \mathcal{W}}{k_L T} = \frac{\Delta \mathcal{W}}{\Sigma} \times \frac{4}{\sqrt{M} + 1} \quad (27)$$

$$I_{fit}(\mathcal{W}^*) \approx 2t^* \times k_L \ln(N/M) \quad (29)$$

$$I_{Sha}(\mathcal{W}^*) \approx t^* \times k_L \ln(N/M) \quad (30)$$

$$I_{bio}(\mathcal{W}^*) \approx t^* \times k_L \ln(N/M) \quad (31)$$

すなわち、 ΔI_{fit} は、進化速度 $\Delta \mathcal{W}$ を、その不確定性 Σ を単位として測った量と解釈できる (Fig. 3の中段)。これは、熱力学におけるエントロピー変化 $\Delta Q/T$ が、熱量 ΔQ を「1粒子当たりの熱揺らぎのエネルギーの大きさ」 $k_B T$ を単位として測った量であるのとよく似ている (Fig. 3の下段)。この熱力学とのアナロジーにしたがうと、 $\Delta I_{fit} = \Delta \mathcal{W} / T$ は進化システムが外部環境から吸収する負のエントロピーと解釈できるであろう (Fig. 1)。ここで、外部環境とは進化する対象 (すなわち集団の実体、この場合はポリペプチド集団) をとりまく実験環境 (例: アフィニティクロマトカラム) である。さらに負のエントロピー変化は獲得した情報量と等価であることから、我々は I_{fit} を外部環境から獲得する“適応度情報量”とよぶことにした (この適応度情報量の概念に関しては、「おわりに」でさらに言及する)。ところで、 ΔI_{bio} の期待値に関しては

$$E[\Delta I_{bio}] = JX = LX^2 \geq 0 \quad (28)$$

が成り立つので、 I_{bio} はこの進化系のLyapunov関数であり、進化は I_{bio} の最大状態に向かって進行すると結論できる。そこで、我々は I_{bio} を“生物学的情報量”とよぶことにした。

定常点 \mathcal{W}^* において、3つの情報量には次の関係が成り立つ。

ここで、 t^* は定常点にたどり着くまでの平均世代数であり、 $t^* \approx 2v/d$ である。右辺の $k_L \ln(N/M)$ は、「 N 個の中から M 個を選ぶ」という選択過程で獲得する (選択に関するShannonの)情報量であるので、 $t^* \times k_L \ln(N/M)$ は、定常状態までに獲得したこの情報量の総和である。よって、式(31)は、選択で獲得した情報量 $t^* \times k_L \ln(N/M)$ が、進化した生体高分子集団のもつ生物学的情報量 I_{bio} に変換されたと解釈できるので、順当な結果である。

5. おわりに

以上の理論は、各残基が独立に適応度に寄与する場合 ($k=0$) に関してのものであるが、相互作用がある場合 ($k \geq 1$) は、平均相互作用数 k が極端に大きくない場合に限り、残基置換数 d を実効的置換数 $d_{eff} = (1+k)d$ に置き換えるか、地形定数を $k_L = \sigma \times \sqrt{1+k}$ と再定義することで上記の理論的枠組みがほぼ適用できる⁹⁾。

最後に、適応度情報量 I_{fit} の概念について考察してみる。進化する対象 (ここでは生体高分子) は外部環境から「その環境に適応して生存し続けるための情報」を獲得する。冒頭で例にあげた「ある受容体タンパク質に結合するペプチドの進化的創出」に関する場合は、進化実験で創出された結合性ペプチドは、外部環境から「受容体に結合するための情報」を獲得したことになる。

ここまでよく聞く話ではあるが、この情報の価値を、「適応度変化である結合自由エネルギー変化 (ΔG) を進化温度 (T) で割ること、言い換えると、その変化の不確定性 (Σ) を単位として測ること」で定量化する。まとめると、ここで定義した“Shannon の情報量” I_{Sha} は情報の多様性に関する側面を表現するのに対し、“適応度情報量” I_{fit} は情報の価値的な側面を表現しているといえる。“生物学的情報量” I_{bio} はその両者を含んでいるのが特徴である。

今回の結論は、変異と淘汰の繰り返しによる配列空間中の運動と熱力学過程との間にアナロジーがとれ、その結果、分子進化過程が外部環境からの情報獲得過程であるという概念を明確にできたというものであるが、これに対する大きな問題は地形モデルへの依存性であろう。ここで述べた法則が現実の適応度地形について成立するか否かはまだわからないが、現在、先に述べた四方らの実験室内分子進化の実験データ¹¹⁾を当理論に基づいて解析しており、理論と実験がおおむね両立することが示唆された。今後は、これまで以上に理論と実験を連携して進化現象の本質に迫る必要があると思われる。

文献

- 1) Frances H. Arnold (ed) (2000) *Evolutionary Approaches to Protein Design*. *Advances in Protein Chemistry* 55, Academic Press.
- 2) 伏見 謙 (編) (2002) シリーズニューバイオフィジックス第8巻「生命の起源と進化の物理学」, 共立出版.
- 3) 伏見 謙 (1999) *化学と生物* 37, 678.
- 4) Eigen, M. (1992) *Steps towards life*, Oxford University press.
- 5) 田中 博 (2002) *生命と複雑系*, 培風館.
- 6) 巖佐 庸 (1991) *生物物理* 31, 27-32.
- 7) Eigen, M., McCaskill, J. S. and Schuster, P. (1989) *Adv. Chem. Phys.* 75, 149-263.
- 8) Aita, T., Morinaga, S. and Husimi, Y. (2004) *Bull. Math. Biol.* 66, 1371-1403.
- 9) Aita, T., Morinaga, S. and Husimi, Y. (2005) *Bull. Math. Biol.* 67, 613-635.
- 10) Kauffman, S. A. (1993) *The Origin of Order*. Oxford: Oxford University Press.
- 11) 林 勇樹, 卜部 格, 四方哲也: 第27回日本分子生物学会年会 (神戸, 2004年12月8-11日)
- 12) L. プリルアン (佐藤 洋訳) (2002) *科学と情報理論*, みすず書房.

実験室内分子進化過程の熱力学的および情報論的解釈

相田拓洋（あいた たくよう）

（財）埼玉県中小企業振興公社地域結集型共同研究グループREDSプロジェクト雇用研究員

1996年埼玉大学大学院理工学研究科博士課程修了、日本学術振興会特別研究員、ノバルティスファーマ（株）

契約社員、（独）産業技術総合研究所生命情報科学研究センター非常勤研究員を経て05年より現職。

研究テーマ：実験室内分子進化の理論研究、進化分子工学の方法論の開発

趣味：音楽鑑賞（クラシック音楽マニア）

連絡先：〒333-0844 埼玉県川口市上青木3-12-18, SAITEC, 7F

Email: sts028@saitama-j.or.jp

伏見 譲（ふしみ ゆずる）

埼玉大学大学院理工学研究科長

連絡先：〒338-8570 さいたま市下大久保255

E-mail: husimi@fms.saitama-u.ac.jp