



## 遺伝子発現のZipf則

古澤 力\*

### Zipf's Law in Gene Expression

Key Words : Zipf's Law, Gene Expression, Universality, Complex Systems

#### 1. はじめに

分子生物学の発展により、細胞内の遺伝子、タンパク、代謝物質などについての詳細な情報が得られるようになった。しかし一方で、その網羅的な細胞内ダイナミクスの情報から、我々が何を理解出来るのかについては、不明瞭なままである。言うまでもなく、ある系について全てを網羅的に記述すること、それを理解することは別の営みである。なぜなら理解するということは、ある程度細部を無視して、薄目で全体像をみることから始まるからである。全ての細胞内化学反応の詳細が完全に明らかにされたとしても、それは我々が生命システムを理解したことと等価とは言えないであろう。

無論、それら網羅的解析が理解するために無意味であるというわけではない。理解のためには、網羅的な解析の先にあるものが重要となる。では、生物を理解するための研究はどうしたらよいのであろうか。まず、理解するには、個別の現象を記述するのではなく、ある現象の集合に共通する普遍的性質をひきだし、その普遍性のなりたつ範囲クラスとその普遍性の由来をひきださねばならない。そのためにはできるだけ簡単な設定で普遍的性質を探り、普遍的性質を成り立たせている論理を抽出することが必要である。

この考えの下で、生命システムにおける普遍的な性質を、理論と実験の両面から明らかにしようと研究を行っている。ここでは紙面も限られており、そのうちの一つである細胞内化学反応ネットワークが従う普遍的性質について簡単に紹介する。(モデル/解析の詳細は[1]を、思想面を含めた日本語による解説は[2]を参照)

#### 2. 遺伝子発現のZipf則

一般に細胞内の生化学反応は、代謝にせよ遺伝子を含む反応にせよ、非常に多くの成分を持ち複雑なネットワークを形成している。一方で、各成分の分子数は一般にそれほど多くはなく、そのような状況では個々の成分の分子数は大きくゆらぎ変動するであろう。にもかかわらず、細胞は状態を維持し、ほぼ同じものを複製していく。それでは、この細胞状態を再帰的に維持していくネットワークのダイナミクスには、なにか普遍的な性質があるだろうか。

そこで、以下で述べる確率的な化学反応のダイナミクスを持ち、分裂していく細胞モデルを用いて、「自分と同じものを複製し続けられる」細胞内ダイナミクスがどのような性質を持つかを考えることにする(モデルと結果の詳細は文献[1]を参照)。まず、細胞内には  $k$  種類の分子があり、それらの間で  $x_i + x_j \rightarrow x_m + x_j$  ( $x_i$  は  $i$  番目の種類の分子) といった形の触媒反応のネットワークが形成されているとする。このとき、触媒反応が起こる  $i, j, m$  の組合せは一定の確率  $p$  でランダムに決められるとする。また、一部の分子は細胞膜を透過できるとし、細胞膜の透過の速さつまり拡散係数  $D$  は分子の種類に依存せず一定であるとする。加えて、外部環境には触媒活性を持たない少数の成分(=栄養成分)が一定濃度で存在するとする。この系においては、細胞は環境から栄養の分子を取り込み、それが細胞内の触媒反応



\* Chikara FURUSAWA  
1972年10月生  
2000年東京大学大学院・総合文化研究科・博士後期過程修了  
現在、大阪大学・情報科学研究科・バイオ情報工学専攻、助教授、博士(学術)、理論生物学  
TEL 06-6879-7432  
FAX 06-6879-7432  
E-Mail furusawa@ist.osaka-u.ac.jp

ネットワークによって他の分子に変換されていき、膜を透過できない分子が作られていく。この反応がうまく進行すれば、結果として細胞内の総分子数は増加していく。ここで、細胞内の総分子数が一定値  $N_{max}$  を越えた場合には、細胞は分裂しランダムに選ばれた半分の分子から娘細胞が形成されるとする。

ここでは、系のコントロールパラメータとして細胞膜の拡散係数  $D$ 、つまり細胞が自分自身と外部環境をどれだけ「混ぜ合わせる」かの度合を採用し、細胞内ダイナミクスの変化を見ることにする。まず、 $D$  が小さい、つまり外部からの栄養の取り込み速度が遅い場合には、多くの成分の間での反応がほぼランダムに起こる。つまり(ランダムに作られた)反応ネットワークの使い方には何ら特別な構造はなく、結果として各成分の分子数はだいたい同じオーダーに落ち着く。

その状態から  $D$  を大きくしていくと、各成分の分子数の違いが増大する。栄養成分からまずある成分がつくられ、そこから別な成分がつくられる、あるいはその成分が別の反応を触媒するというカスケード構造が現われ、各成分の分子数はそのカスケードの順に応じて減少する。つまり栄養成分の流入が多いので、そこから他の成分への流れがそのカスケードに沿って形成されてくる。

しかし、ある程度以上  $D$  を大きくすると、栄養成分の流入速度が、それを反応によって他の成分に変えていく速度よりも大きい状況となる。この場合、分裂とともに栄養以外の成分の割合が減少していき、最終的には栄養成分のみが存在する状態になり、分裂が止まる。つまりこの状況では、細胞は過剰に外部環境を取り込みすぎたため、自分自身を維持することが出来なくなってしまふ。以降、この継続して成長ができなくなる臨界の  $D$  の値を  $D_c$  とする。

次に、このダイナミクスの変化を特徴づける量として、図1(a)に細胞の増殖速度と自己複製の正確さを拡散係数  $D$  の関数としてプロットした。図1(a)に示されるように、増殖速度と自己複製の正確さは共に臨界点  $D=D_c$  において最大値を取る。このことは、分子数がゆらいでいる状態においても臨界点においては比較的正確な自己複製が可能であることを示し、また同時に、色々なパラメータを持った細胞が淘汰している系においては、この臨界点近傍にある細胞が生き残ってくることを強く示唆する。なぜな

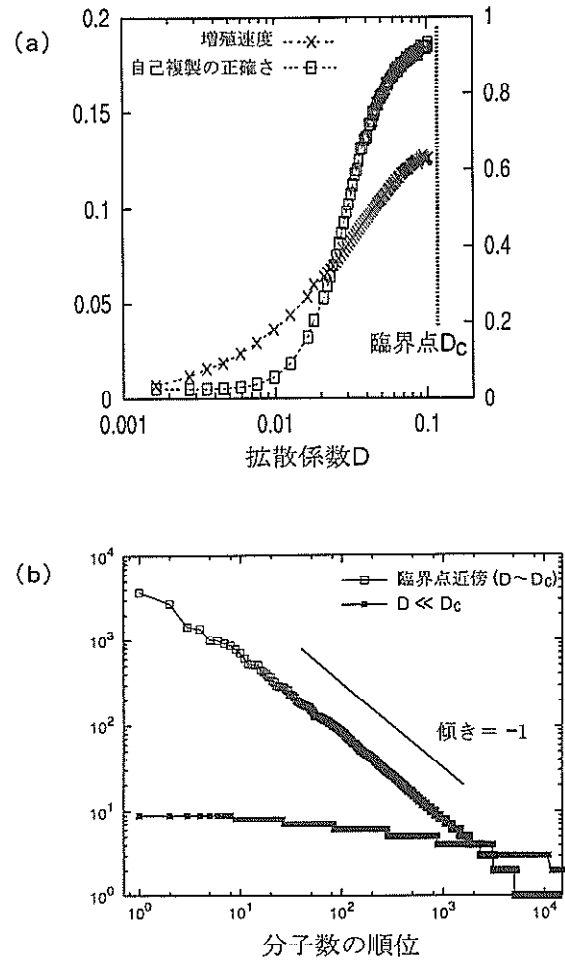


図1 (a)：細胞の増殖速度と自己複製の正確さを拡散係数  $D$  の関数としてプロットしたもの。ここで、分裂速度は分裂に要した時間の逆数としている。また、自己複製の正確さは、分裂ごとにどれだけ細胞の成分組織が維持されているかを表す量として、分裂時の各成分の分子数を要素とする  $k$  次元のベクトルを用い、その母細胞と娘細胞の間の内積で定義した。つまり、世代間で成分組成が維持されている場合には、このベクトルは  $k$  次元空間で同じ方向を向くことになるのでその内積は 1 となり、また組成が次の世代に伝わらない場合には 0 に近くなる。拡散係数が臨界点  $D_c$  を越えた細胞は増殖できないことに注意。  
(b)：細胞分裂時の各成分の分子数を多い順に並べ、その分子数と順位の関係を両対数表示したもの。臨界点  $D_c$  近傍のダイナミクスと  $D \ll D_c$  の状態のダイナミクスをそれぞれ示している。このシミュレーションでは、分子の種類  $k=5 \times 10^6$ 、反応パスが存在する確率  $p=0.022$ 、分裂時の総分子数  $N_{max}=5 \times 10^5$  としている。

ら、その細胞は速く増えられ、さらにその組成を次世代によく伝えるからである。

そこで、この臨界点近傍において細胞内ダイナミクスがどのような性質を持つかさらに調べるため、図1(b)に細胞内の各成分の分子数のランク分布を示した。ここでランク分布とは、各成分の分子数を大きい順に並べて、横軸にその順位、縦軸に分子数を取ったものであり、それを両対数でプロットしている。図に示されるように、臨界点において分子数のランク分布は両対数プロットで傾きが-1の直線になる。つまりは分子数が順位の逆数に比例するというZipf則と呼ばれる関係が成り立っている<sup>[3]</sup>。これは臨界点に特徴的なダイナミクスの性質であり、臨界点から外れたより低い $D$ の状態では、前述のように各成分の量はだいたい同じオーダーとなり、ランク分布は傾きのゆるやかな曲線となる。

ここで注意して欲しいのは、このZipf則はモデルの細部に依存しない性質であるということである。例えば、 $k$ ,  $N_{max}$ ,  $p$ といった他のパラメータを色々変えてみても、常に臨界点 $D_c$ は依存し、そこでZipf則は同様に成り立つ。あるいは同様に、反応パスの密度である $\rho$ を変化させて臨界点 $k_c$ が存在し、そこではZipf則が成り立つ。さらには、モデルに変更を加えて、反応のレートや拡散係数が成分ごとに違い、とか反応の本数が広く分布している。ネットワークを完全にランダムでなくある程度構造を持たせるといった性質を持たせた場合においても、Zipf則は常に成り立つ性質である<sup>[4]</sup>。このことは、この臨界点での性質が、内部に化学反応のネットワークを持ち、外部から栄養を取り込みながら(再帰的に)分裂していくどのような系においても共通に現われる普遍的性質であることを強く示唆する。また、前述の増殖速度と自己複製の正確さの議論を考慮すると、進化の過程で生き残ってきた細胞が、この臨界点のダイナミクスの性質を共通に持つことが予想される。

では、実際の細胞においてこの性質を見出すことが出来るであろうか？ このことを調べるために、図2に細胞内の遺伝子発現量(mRNAの分子数)のランク分布をさまざまな組織や生物種について両対数でプロットした<sup>[5]</sup>。これらから読み取れるように、発現量はほぼ順位の逆数に比例している。ここでは、ヒトの様々な組織や線虫、酵母などの例を示したが、調べられている、ほぼすべての組織の細胞でこの法則が成り立っている。それは大腸菌、酵母、線虫か

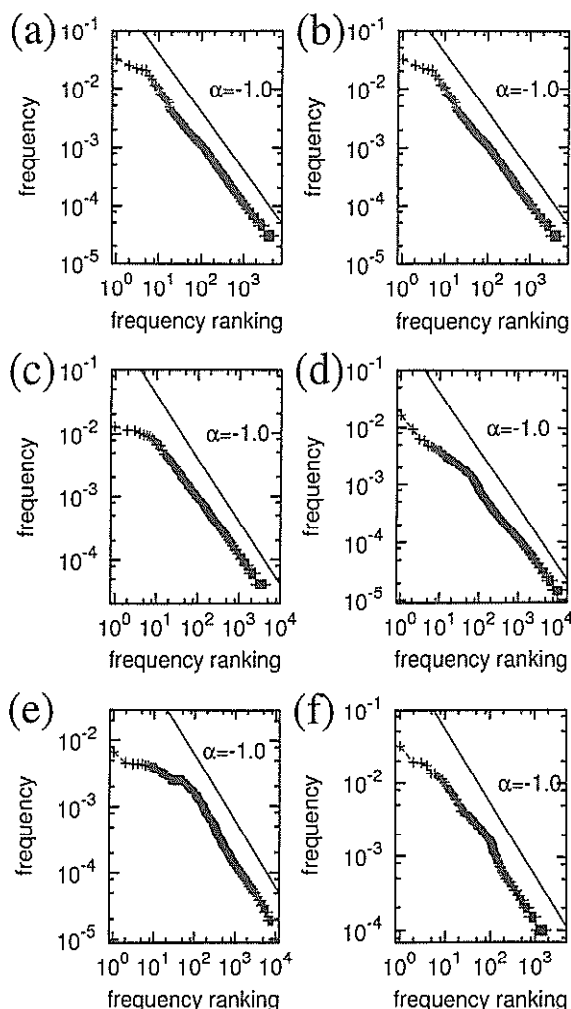


図2 遺伝子の発現量を多い順に並べ、量と順位の関係を両対数表示したもの。(a):ヒト肝臓、(b):ヒト腎臓、(c):ヒト大腸癌、(d):マウスES細胞、(e):線虫、(f):酵母をそれぞれ表示している。発現量は総発現量が1になるように正規化している。

ら始まり、マウス、ヒトの様々な組織の細胞、さらには、はい性幹細胞(ES細胞)、ガン細胞までも含んでいる。これらで両対数グラフでの傾きは-0.9から-1の間であり、Zipf則をほぼ満たしているといえるであろう。また、代謝フラックス解析の結果から、細胞内の主要代謝経路のフラックスの分布においても、同様なべき則が存在することが示唆されている<sup>[7]</sup>。これらの結果は、モデルによって見出された自己複製する系を持つ普遍的性質を、現実も細胞においても確認できたことを物語っている。

### 3. おわりに

上記の研究の他にも、細胞内の化学反応ネットワークのダイナミクスやゆらぎに着目してそこから普遍則を切り出し、細胞増殖、分化、発生、遺伝といった現象の理解を目指している。そのためには、上記の研究のように本質を抽出した計算機上のモデルとともに、そこから得られた普遍則を検証し、また新たな普遍則を見出すための実験も必要であり、詳細は省くがモデル・理解・実験の三位一体の研究を進めている([2]を参照のこと)。これら普遍則を切口とした生命システムの新たな理解は、今後の生物学の新たな基盤となると期待している。

### 参考文献

- [1] Furusawa, C. & Kaneko, K. (2003) *Phys. Rev. Lett.* 90, 088102
- [2] 生命とは何か 一複雑系生命論序説一, 金子邦彦(2003)東京大学出版会
- [3] こうした順位と量の逆比例関係はZipf則と呼ばれている。元々はZipfが自然言語での単語を出現頻度の多い順に並べてみて、この関係を見出した。その他にも、町の人口の分布やインターネットのトラフィックなどの多くの系においてこの関係は見出されている。
- [4] この細胞モデルに対し、遺伝子の転写・翻訳や代謝経路といった「現実」の細胞にある要素が入っていないから意味がないと考える方もいるかもしれない。しかし、そういった細部の要素を導入し、どこまでもモデルを精密にしても、この性質は同じように現われるであろう。ゆえに、それら細かい要素を加えることは現象の本質を曖昧にするだけであると考え、そのような手間は取らないというのが我々の立場である。
- [5] このデータはSAGE法で得られたものであるが、同様の結果がマイクロアレイを用いた実験でも確認されている。
- [6] この結果は遺伝子の発現量のみ結果であるが、代謝物質の量やタンパク質の量の分布、あるいはそれらを全て一つにして平等に1種類の分子だとみなして分子数の分布を見た場合においても、同様にZipf則が満たされるであろうと予想している。
- [7] Almaas, E. et. al. (2004) *Nature* 427, 839

この記事をお読みになり、著者の研究室の訪問見学をご希望の方は、当協会事務局へご連絡ください。事務局で著者と日程を調整して、おしらせいたします。

申し込み期限：本誌発行から2か月後の月末日

申し込み先：生産技術振興協会 tel 06-6395-4895 E-mail [seisan@maple.ocn.ne.jp](mailto:seisan@maple.ocn.ne.jp)

必要事項：お名前、ご所属、希望日時(選択の幅をもたせてください)、複数人の場合はそれぞれのお名前、ご所属、代表者の連絡先

著者の都合でご希望に沿えない場合もありますので、予めご了承ください。