

生命系の構築原理と進化

ー進化する分子ネットワークとしての生命

東京医科歯科大学大学院生命情報科学教育部

田中 博

第1章 生命系の構築原理の確立に向けて

本章では生命系の構築原理とその進化を複雑系生物学から解明し、普遍的な生命系理論の構築の可能性を考えてみる[1][2]。

1. 1 生命系の構築原理に向けた基礎的考察

「生命とはなにか」という問いで目指されるもの

生命科学の近年の著しい発展にもかかわらず、我々はいまだに「生命とはなにか」という問いに対して、明確な答えを持っていない。それだけでなく、この問いをどのような形で問えればよいかについてさえ手探りである。

生命とは何かという問いに対して、生命を様々な特徴の集まり、例えば、自己複製するとか、代謝があるとか、自己調節や適応などの基本的な特徴を列挙することは、一つの答えかたであるが、答えられた方はなにか釈然としない。我々が期待している生命に対する問いの解答は、自然の中の階層として「生命」を「普遍的な存在のクラス」として明確にする定義である。物理・化学的な階層と生物の階層では、その構成要素となる分子や物質に「違い」があるわけではないが、そこには大きな「非連続性」があると思われる。

また、この問いは同時に、生命系がどのようにして非生命的な物質階層から生成し得るかについての問いでもある。地球史的な偶然事象があるにしても、その生成の原理的可能性が物理学的可解性の上に提示されなければならない。

さて、この問いに対して、第一に言えることは、生命と非生命の違いは、構成分子の違いに帰着するものではなく、「生命を構成している物質間に成り立つ関係のあり方」にこそ存在するということである。もちろん生命起源的 (biogenic) な高分子も存在するが、それはあくまでも生命系のあり方のなかにおいて機能を実現するのであって、その逆ではない。それゆえ、上記の言表をわれわれの生命把握の基本命題として採用してよいと思われる。関係生物学を唱える Rosen[3]は、生命を考えると「物質を捨て去って、それらを成り立たせる組織体制をとりあつかえ」(Throw away the matter, keep the underlying organization)と言っている。

それでは、生命系を普遍的に特徴づけ、我々がそれによって生命系の基本を理解したと言明できる普遍形式となにか。生命を特徴付ける多くの性質の中で、生命系を非生命から峻別し、それから生命の諸性質が導出される生命系特有の「物質集合系の間に働く関係のあり方」とはなにか。それを1つの原理

的概念として表しえるものであろうか。

境界を自己決定し統合化された超分子的集合構造としての生命

まず生命に対する記述的考察からはじめよう。対象としては生命の最小実現であるバクテリアを思い浮かべてもよい(図1)。現在生存している最小の遺伝子数 486 をもつ生物 *Mycoplasma* もバクテリアである。

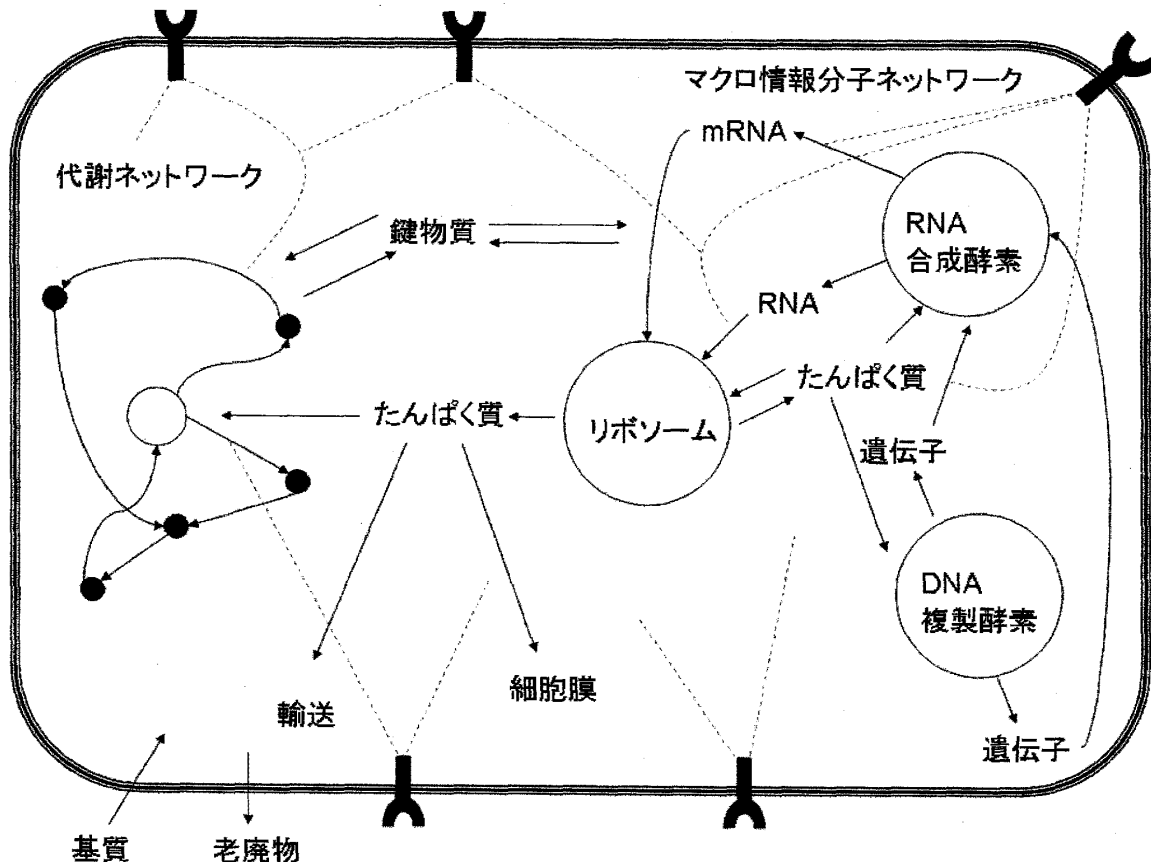


図1 最小生命系の構成 (Mittenthal[4])

誰しもが認めることであるが、生命は種とか社会とか集団的関係性において存在するが、具体的な生命は、明確な「境界をもった個体」として観察される。生命は自らの境界を自己決定し、世界を「外」と「内」に分ける境界によって閉じられた連続した領域に定在する。また、生命は、自分自身を存続するために外部から必要な物質やエネルギーを収集して、自己の構成要素に化学的に変化させ不断に活動する代謝サイクルを働かして自己を維持し再生産していることもよく知られた事実である。

このように、生命系は周囲の環境とエネルギーと物質の交換がある「開放系」でありながらも、明確な境界に囲まれた空間的な閉域(operational clousure)において、統合化された超分子的な集合構造(supermolecular organized structure)[5]として、自己を絶えず再生産している系である。このような系を例えば Maturana[6]や Varela[7]は、概念的に自己創出的統一体(autopoietic unity)と呼んだ。それでは、このように膨大な反応系を秩序付け、生命系を統一化している原理、そしてそれを担う機構とは

何であろうか

生命の秩序化原理としての基礎としての大域的な非平衡循環構造

生命系に固有な秩序原理を考究する前に、まず「生命」のように一見熱力学の第2法則に抗して、自己の秩序を維持している系の物理的可能性を論じておこう。もちろんこれは生命系構築の基礎であって、生命系の固有の構築原理を必要十分に記載するものではない。

物理的法則で構築される秩序には結晶のような平衡構造だけでなく、非平衡構造がある。これは開放系に特有な構造で、「非平衡性が維持されている」場であって、循環構造という秩序が自発的に形成される。エネルギーを外部から取り入れ、内部で発生したエントロピーを外部に排出する循環過程を働かせることによって系自体のエントロピー増大を抑える、いわゆるネグエントロピー系である。このような非平衡循環構造が可能であるためには、

- (1)非平衡性が維持される、より大域的な系の内部にある、
- (2)開放系である、
- (3)自己触媒反応を含め円環反応系が含まれる、

必要がある（「Morowitzの原理」[8]）。生命系において、部分的には平衡構造（生体膜など）も利用されている。しかし、生命系の大域的秩序形成において、その基礎的部分は「非平衡的な循環構造形成の原理」によって担われている。代謝反応系などのいわゆる本来の化学反応系は本質的に円環的な構造であり、非平衡的な物質・エネルギーの流れの中であって、自己再帰的な反応系(autocatalytic cycle)、すなわち非平衡循環系が形成されている。

我々生命系の場合、(1)の条件は、「太陽-冷たい宇宙」という非平衡場に生命系が存在するという事実にある。生命圏、そしてその中にある生命系を維持できるのは冷たい宇宙にエントロピーを捨てられるからであり、さらにそれを可能にしているのは宇宙が拡大していて非平衡性が增大している(hung-up現象)からである。すなわち、Brillouin-Layerの式[9]

$$I = S_{max} - S$$

で、宇宙の可能的エントロピー S_{max} が膨張のため増加しているので、現実のエントロピー S が追いつけず、その差である秩序 I が拡大しているからである。

1. 2 生命に特有な秩序原理——「生命系の情報原理」

「情報」に基づく秩序化

非平衡循環構造は、それだけでは生命圏だけでなく地球圏などにも見られるも秩序形成の原理であり、生命は、この原理のもとに、太陽光から始まる円環的反応系、すなわち光合成や酸化的磷酸化などの循環的反応過程を、ATPを媒介分子として巧みに連鎖して多重の円環構造（Wicken[10]の生命系の抗エントロピー構造）を構成している(図2)。

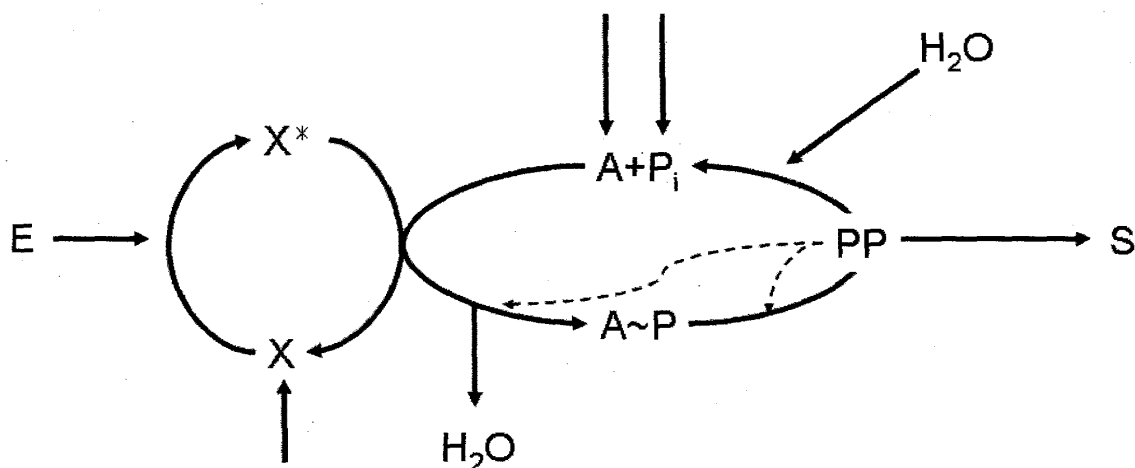


図2 Wicken の抗エントロピー構造

E は太陽光、X は光受容体、A はアミノ酸、 P_i は無機リン酸、PP はタンパク質

生命系においてこの循環的反応系は、通常の自己触媒反応、すなわちある分子が自分の反応の触媒になる反応だけでなく、集合として自己触媒的、すなわち、集合に含まれる分子が集合のどれかの反応に対して触媒的であり、その意味では自己強化的反応系とも言うべき発展形であるが、原理そのものは生命に固有のものではない。それでは、本来の生命系において始めて現れる構築原理とはなにか。

非生命的自然においては物理的な力に起因する物理的秩序化が物質の存在形式を決定しているが、生命系においては、現象の複雑さにより物理的法則性だけでは生命系という超分子集合構造を統合化することは可能ではない。あらたな秩序原理を構築する必要がある。

物理的な世界において現象が複雑になると物理的因果性では状態が唯一に決定せず、物理的多義性が出現する。このような多義的状态、物理的非決定性において状態を決定するのはその物理系自身ではなく、それを含む大域系においてであり、物理的因果性とは別の論理である。物理的非決定性が生じている系にあって各状態の確率的な実現のエントロピー S が可能な最大値 S_{max} より低い系は、確率構造がランダム性から歪んでおり「情報」が生成されていて、「情報による」状態決定の機構が示唆される^{*注1}。

生命系は、下位の部分系は物理化学的な過程や非平衡の構造化に委ねたとしてもそれらを調節し、系全体としての統一性(unity)を達成する大域的な「生命系特有の秩序化機構」を発展させた。それは一言で言えば、「生命系の情報原理」とも呼ぶべき原理で、非生命系ではこれまで存在しなかった「情報」を用いた秩序形成、すなわち information-based な秩序系の構築である。

注1：本稿では情報の生成に関しては、物理系の複雑性が増すと物理的に一意に決定できない物理的多義性を持った現象（例えば核酸の4塩基の自由選択性など）、「選択構造」による状態選択の歪みから情報の生成を考え、情報による秩序形成過程を論じたが、複雑性を増すとカオス化する手前に情報が発生する領域が存在するという「カオスの縁」論的な説明も可能である。

生命系を統合する「分子認識/相互作用ネットワーク」

「情報」の秩序形成の概念を顕著に示す例は、Yagil[11]がタバコモザイクウイルスとバクテリオファージ T4 の自己構築に関して分析した区別である。前者は構造も簡単で一旦部分に分解しても物理的自己組織化作用でひとりでに集合して元に戻るが、それより複雑性の高い T4 では、配列情報（遺伝子）がなければ元に戻らない。すなわち、複雑性にはある閾値があって、それ以上では物理的法則性では系が構成できず「情報」が必要となる。

生命系が複雑で物理的法則性からは生成できず、遺伝子に基づいた「情報」に基づいて作られるということは現在では誰でも知っていることである。しかし、遺伝情報の集合であるゲノムは通常の建築物の「設計図」とは違い、完成したときの生体の形状や内部構造などが詳細に記載しているわけではない。記載されているのは機能タンパク質分子の作り方(アミノ酸配列)とタンパク質に翻訳されない RNA^{*注2}の配列情報だけである。

生命系では、発生の過程とともに、遺伝情報から翻訳されたマクロ分子が、生命系内に自然に「分子認識機能に基づく相互作用ネットワークを張り」、それが物理法則(Physical law)に替わって、生命系、すなわち個々の物理・化学過程を統合化する生命法則(Biogenic law)を形成する。したがって、ゲノムは、単なるマクロ分子の構築方法だけでなく、発生過程とともにこれらが現実化するための「期待構造」が非明示的に含まれている。発生が通常の条件下で進み、それにもなって適切な時期にマクロ分子が遺伝情報から翻訳され生産されると、これらは「自然と」生命系内に相互作用ネットワークを張り、生命系を統合化する主体的機序となる^{*注3}。そして「生命系の情報原理」とはゲノムの情報ではなくてこのように生命系に張られた分子相互作用ネットワークの情報に準拠した秩序形成作用である。

注2：non-coding RNA、ゲノムの40%を占め、遺伝子発現制御に関係するとされる

注3：ここには動物行動学でいう理性同型(ratiomorphic)な動物の認知(Rielf[12])と同じように「世界」に対する期待構造の上に成り立つのである「知」がある。自然には生起することが極めて稀な人為的な環境設定をすれば、たちどころこの動物の「知」に基づく失敗をする。生命系の分子認識/相互作用ネットワークも通常の条件下での発生過程の進行の期待の上に展開される。

分子認識/相互作用ネットワークは選択的構造を持つ

さて、このような分子認識/相互作用ネットワークの性質についてもう少し検討しよう。まずこのネットワークはマクロ分子による分子認識に基づいて構築されたネットワークである。分子認識においては、マクロ分子同士が照合する場合もあるし、情報伝達に低分子信号分子の授与を行う場合もあり、また核酸鎖に転写因子としてのタンパク質が結合して遺伝子を発現させる場合もある。いずれも、いわゆるパターン照合型実行が行われ、分子認識後のタンパク質の変容によって効果を生み出す。したがって、分子認識を担うマクロ分子は知識科学でいうところの知識の最小実現である「if-then」ルールの実行単位といえる。いずれにせよ、これは物理・化学層の物理・化学反応のネットワークとは異なり、認識を機能とするマクロ分子間の「情報」の伝達、情報による作用/秩序形成を目標とするネットワークである。

物理的な因果法則性は微小時間先の現象の未来を形成していく「前向きな決定」機構であるのに対して、「情報によって作られた秩序」の正当性はその系が存続できるか、選択によって「後ろ向きに」決定される。したがって生命を統合化する「分子認識/相互作用ネットワーク」は選択された構造をもつ。

そのあり方は、生命系の原理的あり方を形成するものである。

分子認識/相互作用ネットワークの志向的構造への展開

生命分子ネットワークは、生命系の全体を統合して境界も含めた統一体(unity)を形成するものである。統一体として自然に出現した瞬間から、統一体に加わる「選択」が始まる。すなわち、そのあり方に関して、生存を目的としてその目的に向けて系を適合的に組織化した系が選択される。したがって、選択された生命系を外から見ると、どこにも生命系のなかに目的を設定する実体的な機構はないが、系自体の構造が、あたかも virtual に目的が先に存在してそれに基づいて組織化された形式に見える。このようにして、生命系は志向的 (intentional) 構造を持つ。通常条件下での将来的世界への期待構造に合わせて系を組織化する、生命の志向的あり方が反転的に構築される。

ただ、生命の始原的形態であるバクテリアでは、我々人間のように本当の意味で意図をもつわけではない。バクテリアには意図を生成する実体機構、例えば中枢神経系のような器官は存在しない。周りの栄養素の濃度勾配を受容体タンパクで認識して情報伝達タンパク質にそれを伝え、それが転写因子となって鞭毛を動かすタンパク質を産出し化学走性を実現させる。バクテリアの生命分子ネットワークの中にこのようなパスが含まれて、「if-then」ルールを実行している。その意味では、志向的構造を持つというより志向性同型的 (intentionmorphic) な構造を持つという方が適切である。

さらに、生命分子ネットワークそのものも、系を統合化するためにネットワークの全体性がネットワーク自らにおいて明らかでなければならない。ネットワークのそれぞれのマクロ分子も、自分の作用の結果が翻って自らに戻ってくることによって自らの作用の効果の伝達を知り、それを基に自らの作用を調節することができる。その意味で分子ネットワークは作用的な閉域(operational closure)を作っていないなければならない。選択的な観点からも、反照的に自己へと再帰するネットワークは統一性をもたらし、環境への適合性が高い。自らの作用の情報が戻ってくる、情報再帰的構造が生命分子ネットワークの、発展し選択された構造として確立する。

このようなネットワークの情報再帰性および閉域性が、上に志向性構造の出現と同じように反転して、外から見るとネットワークはあたかも内部的に自己のイメージを有する形式を備えているように見える*注4。具体的に言うと分子認識すなわち分子的「if-then」ルールの集合が自己を構成する。もちろん生命の始原システムであるバクテリアでは、自己を担う実体的機構があるわけではない。あるのは分子認識ネットワークだけである。その意味ではこれを自己同型な(selfmorphic)システムと呼ぶのが適切である。ただこれは数学の自己同型(automorphism)と混乱しやすい。そこでここでは自己形式システムとよぼう。

注4：ネットワークが自らの全体性にかかわる再帰的なモデルや<自己>イメージについては、Kaufmann の自己触媒的集合[13]や Jerne の免疫系のイディオタイプネットワーク[14]でも考究された。

1. 3 「生命的自己」の展開へ

生命系の秩序原理の基礎命題について

さて、ここでこれまでに展開してきたことをまとめてみよう。

(1) 2つの生命系組織化の原理

非平衡循環構造形成の原理

生命系の土台は、非平衡性において形成される循環的反応系による秩序形成である。エントロピー代謝に基づき、非平衡的なエネルギー・物質流れにおいて形成された循環過程であるが、生命系の場合その発展形で、自己強化的循環システムである。物理化学法則に基づいて構築できる構造である。このような系はハイパーサイクル(Hypercycle (Eigen[15]))や自己触媒的ネットワーク (autocatalytic network (Kauffman[13]))などこれまでの理論生物学者が様々な概念で記述した生命系の特徴である。

情報に基づく秩序形成の原理

生命系の秩序原理として「情報に基づいた秩序形成」が行われる。具体的にはマクロ分子の分子認識に基づく相互作用情報伝達・実現ネットワークを形成する。これは生命起源的な法則ともいうべきもので、物理・化学層に浸透して、生命を構造づける主体的原理となる。

(2) 自己形式システムへの展開

分子認識/相互ネットワークの自己再帰的・志向的構造への選択的展開

生命系が情報に基づく秩序化によって1つの統一体(unity)として振舞うことが可能になると、系に対する選択的圧力が加わる。そのため、目的志向的な生命の応答・振る舞いを起こすように分子認識ネットワークが組織化される。

自己形式システムへの展開

また、自己再帰的な情報の流れを実現することによって、自らの行動に対して再帰的に関わるかのように見える。あたかも生命系が自己をもっているかの構造が形成される。

このようにネットワークの志向的構造の上に自己性を実現されることによって、生命系が自己形式システムとして展開される。もちろん、生命分子ネットワークは代謝や発生などの過程も過程自体を構成するだけでなく、バクテリア、例えば大腸菌では、受容体タンパク質と応答調節タンパク質からなる2成分制御(bacteria two-component system)は40ほどある。これは環世界(Umwelt: Uexkul)への「if-then」ルールが40ぐらいという大腸菌の「知性」のレベルを表すと考えられる。

「生命的自己」への発展—自己、志向性、世界への期待構造の基礎

以上のようなバクテリア的な生命系の構造が出来上がると、進化による複雑化により、生命の秩序系のあり方が明確化する方向へと進化する。進化的過程を通して、生命の自己構造が明確になり、将来において実現すべき可能的自己(目標の集合)と現実の自己との2重的関連性が形成され、このような志向的関係性の存在の基に「生命的自己」の自己性が、例えば免疫システムにおける自己と非自己の区別などによってさらに明確に決められていく。

生命系は、自己の全体像についての規範が絶えず現実の存在に対して一方で措定されており、それとの関連で現実の存在を変化させる意味で、自分の基準的なあり方を内部に持つ系である。バクテリアの段階でも自分の境界を自分で決めている、というところにある種の再帰性が存在する。「生命は自分自

身への関連性において存在する」という構造が実現する。

このような志向性構造に裏付けられた再帰的な自己構造のもとに、行為や情報の「意味」が発生する。分子的水準における記号性、情報とそれが担う分子の間に存在する無根拠性などいくつかの生命的自己に特徴的な現象が生成する。

生物哲学者が、主体性とか自律性とか概念で指し示しているものは、生命分子ネットワークが作る自己一同型的な情報の秩序から創発する概念である。

また、生命の情報統合系は、ゲノムにはマクロ分子の構築情報しか記載していないが、非明示的に対応すべき世界や生命系の状態に対する期待構造が含まれている。生命系の秩序構造は、自己の未来に対する志向性構造と世界に対する期待構造が非明示的に符号化(code)された自己形式系であるといえよう。

生命系の構築原理

生命系とは、分子認識相互作用ネットワークを介して、自己再帰的な情報系を物理的に実在化しものである

生命分子ネットワークのシステム進化について

バクテリアレベルの最小生命系を対象にして、生命系の秩序原理を考察した。生命は一般に2つの層の秩序化が絡み合ったものである。非平衡下で円環反応をカップリングして、自己触媒的な円環的反応の複合体のようなエネルギー・物質代謝反応ネットワークは、もちろん要所的に酵素やイオンポンプなどのマクロ分子が介在するが基本的には物理化学的反応のネットワークである。これに浸透して情報や統御を行うマクロ分子ネットワークが存在する。そして選択による志向的構造への展開によって、実体的な統括的自己はどこにも存在せず、ネットワークしか実在しないが、志向性あるいは自己一同型的な構造が完成する。

このような生命分子ネットワークは、生命の進化とともに複雑化する。かつて中立進化説や進化の総合学説が考えたように、遺伝子やタンパク質は個々に変異したり選択されたりするわけではない。それらは生命分子ネットワークの中に他のネットワークと関連付けられて存在するのであり、ネットワーク自体が複雑化する過程こそ生命の進化である。したがって、ネットワーク拘束性が強い遺伝子・タンパク質は変異しにくい、進化速度が小さいといえる。ネットワーク拘束性を、そのタンパク質が相互作用するタンパク質数（degree 指標）であたえるか、そのタンパク質をめぐる、他のタンパク質同士の連結性（cluster coefficient 指標）で表すと、進化速度はこれに負の相関をしめす。また、ネットワークの複雑化は遺伝子の重複による遺伝子重複の頻度は、ネットワーク拘束性の弱いところから、ネットワークの度数における末端部から始まっていくと考えられる。

システム進化生物学、あるいは分子ネットワーク進化論は、最近のシステムの生命科学の興隆とともに始まった。現在、分子ネットワーク進化論の原則を建設しつつある。グラフ論的指標で与えられるネットワーク拘束性と進化速度の負の関連性は、現在得られた原則の1つである。また、遺伝子重複がネットワーク拘束の弱い末端から起こり、系の構造を次第に変化させていくというのも原則になりうる。

第2章以後においては、これらについての我々の研究成果について報告する。

講演内容に関して

もちろん、生命分子ネットワークも漸進的に遺伝子重複を通して複雑化していく過程とともに、細胞内共生などの非連続的なモジュール進化をとげるときもある。発生調節ネットワークと進化、とくに形態的なマクロ進化との関係も、本来論じたいところであるが、紙幅の制限もあるので本テキストには掲載できない。これに関してはテキストとして拙著[1][2]を参照していただいて、講演において省かずに重要な点を紹介したい。また、ここで展開した生命系の志向性、すなわち一般に主体性とか自己根拠性とかよばれている事象は、素朴に生命を観察すれば、バクテリアさえ「心」を持っているように見える点からも、生命系の謎を提示する問題である。ここでは生命系が選択を受けて、分子ネットワークに基礎を置く情報秩序系が、志向性同型的な構造をとるとしたがこの点についてもより最新の成果を紹介したい。

- [1] 田中博 (2002) 生命と複雑系 培風館
- [2] 田中博 (2007) 生命—進化する分子ネットワーク パーソナルメディア
- [3] Rosen, R. (1991), *Life Itself: A comprehensive inquiry into the nature, origin, and fabrication of life*, Columbia University Press.
- [4] Mittenthall, J.E. et al. (1993), "Designing bacteria," in *Thinking about biology*, Addison-Wesley.
- [5] Fleischaker, G.R. et al. (1994), *Self-production of supermolecular structures*. Kluwer Academic Publishers.
- [6] Maturana, H.R. and Varela, F.J. (1980), *Autopoiesis and cognition: The realization of the living*, Reidel.
- [7] Varela, F.J. (1994), *On defining life*, in *Self-production of supermolecular structures*, Kluwer Academic Publishers.
- [8] Morowitz, H.J. (1968), *Energy flow in biology*, Academic Press.
- [9] Brillouin, L. (1962), *Science and information theory (2nd ed.)*, Academic Press.
- [10] Wicken, J.S. (1987), *Evolution, thermodynamics, and information*, Oxford University Press.
- [11] Yagil, G. (1995), "Complexity analysis of self-organizing vs. a template-directed systems," in *Advances in artificial life*, Springer.
- [12] Riedl, R. (1981), *Biologie der Erkenntnis-Die stammesgeschichtlichen Grundlagen der Vernunft*, Verlag Paul Parey.
- [13] Kauffman, S.A. (1986), "Autocatalytic sets of proteins," *J. Theor. Biol.*, Vol. 119, pp.1-24.
- [14] Jerne, N.K. (1984), "Idiotypic networks and other preconceived ideas," *Immunol. Rev.*, Vol. 79, pp.5-24
- [15] Eigen, M., "Self-organization of matter and the evolution of biological macromolecules," *Naturwissenschaften*, Vol. 58, pp.465-523 (1971)

第2章 生命分子ネットワークの構造と進化

2. 1 生命分子ネットワークの構造

グラフ理論の幕開け—Koenigsbourg の7つの橋

グラフやネットワークの構造に「一般的な法則」があるとして、これを明らかにしようとする考え方が生まれたのは、「Koenigsbourg(ケーニヒスブルグ)の7つの橋」と呼ばれるような問題からである。この問題に18世紀の大数学者 Euler が関心をもち解決したところからグラフ理論が始まったといえる。

18世紀当時の Koenigsbourg の Pregel(プレーゲル)川には、図1のように7つの橋がかかっていた。「Koenigsbourg の7つの橋」の問題とは、どこの区域から出発してもよいとして、この7つの橋を1回ずつ渡ってすべての区域を訪れる周回路があるか、という問題である。数学者 Euler は長年解けなかったこの問題に関心をもって取り組み、この問題には本来解答がないこと、すなわちこのような条件を満たす経路は原理的にありえないことを示し、この問題に最終的に決着をつけた。

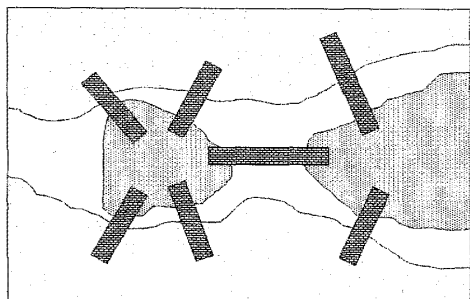
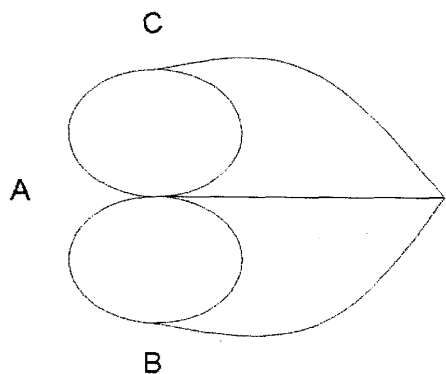


図1. Koenigsbourg の7つの橋

さて、問題を少し詳しく見よう。この問題の本質を、4つの島や地区を点で、7つの橋を線で表した図に描き直して、「この図の線をなどって一筆書きができるかどうか」という問題と同一であると見抜いた。要するに、経路の具体的な長さや角度は関係なく、点と線がどうつながっているかだけがここでは問題であったのである。この考えが数学におけるグラフ理論、ひいては

今日のトポロジーという数学の分野の始まりとなった。現在では「Euler 閉路」を見つける問題といわれているこの問題に対して、Euler は、簡単な性質を基礎にしてこの問題を考えた。

図2. 図1のトポロジー図



その解法は、一筆書きが途中で通過するノード（接点、頂点）については入る経路と出る経路の数が等しいということである。つまり、途中のノードに出入りするエッジ（リンク、枝、辺などとも）の総数は偶数である。ここから、「奇数個のエッジをもつノード(奇数点)は出発点であるか、終着点でなければならない」ということが導かれる。一筆書きができるためには、まず出発点が1つで終着点が1つの場合が考えられる。すなわち、全体で奇数点は2つであとはみな通過ノードである。つぎに、一筆書きで一巡してもとに戻る経路を考える。このときは、出発点と終着点が一致しなければならない。つまり出発点に出て行くエッジと入るエッジの両方がつながり、出発点は偶

数点になる。従って、ほかの通過点も合わせて奇数点はなくなる。

このことから Euler は「一筆書きが可能であるためには、奇数のエッジを持つノードが 0 か 2 でなければならない」という定理を得た。図 2 から明らかなように、4 つのノードすべてが奇数のエッジをもつ Koenigsbourg では一筆書きが本来不可能である。

グラフは、その構造のなかに隠れた特性をもち、これによって表される具体的な関係を制約する。Koenigsbourg の問題が 2 世紀以上も解決されなかったのは、橋が表す結合関係の根底に横たわるグラフの構造、すなわち奇数点が 4 つというグラフ構造のためである。Euler によって、ネットワーク構造などを表すグラフ理論は、この世界の関係の構造を理解するのに有効な新しい理論であるという考えが打ち出された。

グラフ理論は、Euler によって生み出された後、Cauchy(コーシー)、Hamilton、Cayley(ケイリー)、Kirchhoffsches(キルヒホッフ)、Polya(ポリア)といった、数学の他の分野にも大きな業績をもつ大数学者たちによって発展された。彼らによって「規則的な秩序をもった」グラフすなわち規則的グラフ(regular graph)について、現在知られていることはほとんどすべて見出されたといえる。近年になると、与えられグラフの性質を調べるという研究からグラフ自身の形成の過程を調べることに関心が移っていき、数学者 Erdos(エルドス、エルデシュとも)と Renyi(レニー)のランダムグラフ理論が出現した。

ランダムグラフ

ランダムグラフとは、これまでの「規則的なグラフ」とは違い、与えられたノードの集まりのなかから、ランダムに 2 つのノードを選び、それらをエッジでつなぐような操作で形成されるグラフである。

さて、Erdos と Renyi が明らかにしたランダムグラフの重要な性質とは、ノードの数に対して、エッジの数が半数になると、1 種の相転移が出現して、「巨大なコンポーネント」が突然形成されるという現象が起こるという性質である。図 3 はこれを図示したもので、わずか 20 個のノードであるが、この過程が一目瞭然である。物理的にはこの現象はパーコレーションともよばれ、例えば山火事などで、ある程度の数の木が着火しても通常は自然に鎮火して、森全体が焼け尽くすことはないが、ある数を越えて燃えたと全面的な山火事に導き、森林全部が焼け尽くすこととなる現象と同じである。ここで、エッジとノードの比率が 1.0 になると、ほとんどすべてのノードが結ばれることになる。

この 1957 年に発表された Erdos-Renyi の定理は、グラフ理論に画期的な分野を拓いたといわれる。その影響は、相転移やパーコレーションに理論におよぶ広範なものであった。

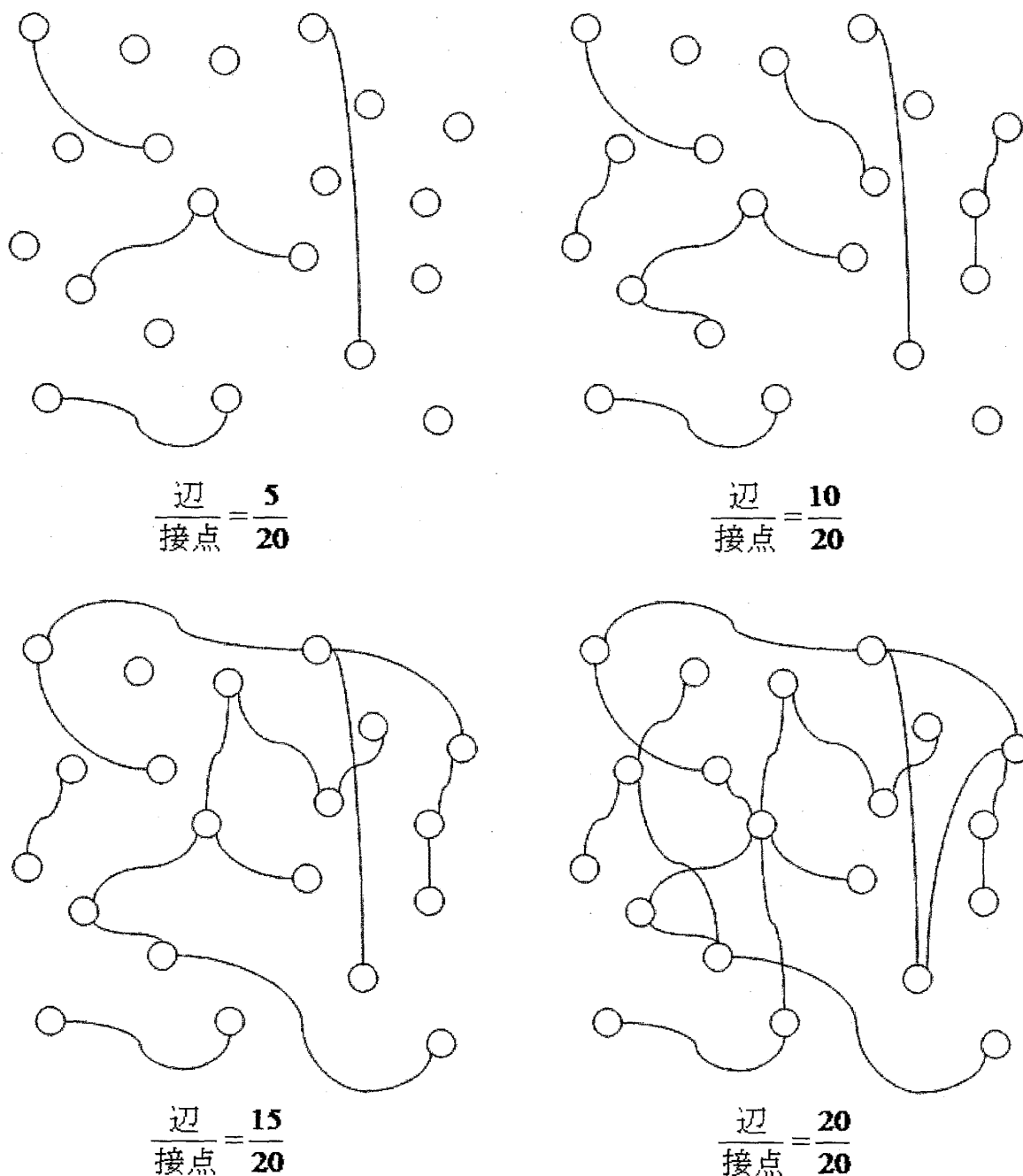


図3. Erdos - Renyi の定理の概念図

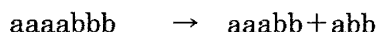
Kaufmann による Erdos の定理の生命起源への利用

Erdos の「巨大なコンポーネントの出現」についての理論は、多くの生命の理論研究者たちを刺激した。特に、生命という系がどのようにして非生命的な分子の集合から出現するかという、困難な問題に取り組んでいた幾人かの学者に大きなヒントを与えた。そのなかには、理論生物学分野の代表的研究者である Kaufmann がいる。Kaufmann は生命の最初の起源となる核酸やタンパク質などのマクロ分子

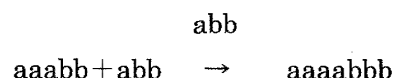
ポリマーが、同じ分子単位の連続によってなりたっている鎖状分子であることから、ポリマー間の結合や開裂の反応でお互いに関係しあっているポリマー集合を考えた。例えば、核酸ポリマーの場合で考えてみる。簡単のため 4 種類ある塩基が 2 種類から成り立つとして a と b と記載するとする。たとえば、5 塩基のポリマーであれば aaabb のように鎖状ポリマーを表せる。この長さの違うポリマーの集合を考える。ある長さまでのポリマーがすべて含まれる集合を考えても良い。たとえば、結合反応は、



という反応式であらわされ、逆方向の反応、



は、開裂する反応をあらわす。このポリマー集合で、それぞれのポリマーをノードと考えると、この反応関係をエッジで表すことができる。このポリマー集合のそれぞれのノードと枝は、結合や解裂の反応によってグラフのようにつながっていると考えることができる。ただ、この生命の始原反応系においては、すこしを複雑にさせる事情が存在する。それぞれの反応を触媒する分子もポリマーの集合内に存在することである。したがって、反応は、



という形になる。したがってエッジも反応エッジと触媒エッジに分かれる。しかし原理的には、先程述べた、ランダムグラフの場合とおなじである。すなわち、ポリマー集合の大きさを与えるポリマーの最大長を大きくしていくと、必然的に反応関係（枝数）は組み合わせ的に増加し、Erdos・Renyi の定理と同じことが起こる。すなわち反応でつながった生命の情報マクロ分子のネットワークが出現する。すなわち、生体のポリマーの自己触媒集合から「お互いに反応系で連結した全体」としての生命系が出現するとした。

そのほかにも遺伝の相互作用による淘汰局面や遺伝子ネットワークについての Kaufmann のモデルもいずれもランダムグラフ理論に基づいてものである。とくに、生命の起源に関しては、Erdos の定理をそのまま使用したといえる。

Kaufmann 以外にも Erdos-Renyi の定理を使って生命系の創発について理論を構成した研究者も多い。生命が自己触媒反応するポリマーの集合のなかから 1 種の相転移として、互いに化学反応で連結し、閉じたポリマー集合として出現したという理論は魅力的である。

実在するネットワークはランダムネットワークではない

この様に、Erdos と Renyi のランダムグラフ理論は大きな成功を収めた。しかし、1990 年代に入ると、大部分の実在するネットワークはランダムグラフとは全く違う構造をとることがわかってきた。本テキストで取り上げるタンパク質間相互作用ネットワーク (PIN, Protein Interaction Network) を含む、実在するネットワークの共通する特徴としては、ランダムネットワークには見られない、スケールフリー性とスモールワールド性が挙げられる。この2つの性質については次章で説明する。更に、これら2つの性質以外の PIN に見られる構造についても説明する。

2. 2 タンパク質間相互作用ネットワーク (PIN) の構造について

Yeast Two-hybrid 法等のハイスループットな実験手法の開発により、タンパク質間相互作用を網羅的に同定することが可能となっている[1,2]。PIN では、タンパク質はノード、タンパク質間相互作用はエッジで表す。PIN の構造は主に以下の4つの性質で特徴づけられる[3]。

第1の特徴は、PIN は、上記で述べたランダムネットワークとは全く違う、スケールフリーネットワークである[4,5]ことである。ノードが有するエッジの数をそのノードの次数と呼ぶが、ランダムネットワークでは $P(k)$ は正規分布に従うのに対し、スケールフリーネットワークでは、あるノードが次数 k を持つ確率 $P(k)$ はべき乗則

$$P(k) \sim k^{-\gamma}$$

に従う[6]。つまりは、ごく少数のノードが極めて多くのエッジを有しているが、大多数のノードは1つないしは2つのエッジのみしか有していない。このようなスケールフリーネットワークの特徴としては故障に強く、攻撃に弱いという性質がある。すなわち、ランダムに選んだノードを取り除いても、ネットワークの性質自体は殆ど変わらないが、次数 k の高いノードを取り除いた場合はネットワークの構造は崩壊してしまう[7]。実際、酵母の PIN 中において、次数 k の高いタンパク質は、生存に必須のタンパク質が多いということが知られている[4]。

第2の特徴としては、PIN はスモールワールド(smallworld)性を有するということである[5,8]。スモールワールドの語源は、社会心理学者であるミルグラムが1960年代に行った実験に遡る[9]。ミルグラムはこの実験で、アメリカ在住の数百人に手紙を出し、それらの手紙が何人の手元を経由して目標の人物まで届くのかを調べた。ただし、その手紙には、手紙を受け取った人は、目標としている人物を直接知らないのであれば、その人物を知っていそうな友人に手紙を送るようにという指示が書かれていた。すると、平均約6人を経由することで、目標の人物に手紙が届くことが発見された。この事実が示すことは、所謂、「世間は狭い」ということであり、従って、スモールワールドという言葉がこの実験から生まれた。スモールワールド性を有するネットワークについてであるが、このようなネットワークではクラスター係数(後述)はランダムネットワークよりも非常に高く、平均最短経路長はランダムネットワークと同程度に小さい[10]。

ところで、ランダムグラフは非常に平均最短経路長が小さく、情報伝達が早いネットワークとして知られているが、PIN も生物が環境に適応するために早い情報伝達を必要としたために平均最短経路長が

小さくなっているのだと考えられる。また、クラスター係数についてであるが、クラスター係数はノード一つ一つについて決まる係数である。次数 k のノード i が存在し、そのノード i に直接結合している k 個のノードの間に m 本のエッジが存在していたとする。すると、ノード i のクラスター係数 C_i は

$$C_i = 2m / k(k-1)$$

で定義される[10]。平均クラスター係数が高いネットワーク中には、3ノードと3エッジで構成される三角形が多数存在する。以上の2つの性質、すなわち、スケールフリー性とスモールワールド性は、多様な実在のネットワークに共通するものである。実際、論文の共著関係のネットワーク[11]、WWW[6]、脳の機能相関ネットワーク[12]、代謝系ネットワーク[13]、そして遺伝子制御ネットワーク[14]等もこれら2つの性質を有していることが知られている。従って、PINを含むこれらの実在のネットワークは、ランダムネットワークとは全く異なった構造を採っている。

第3に、PINは非同類結合性(disassortativeness)を有するネットワークである[15-17]。非同類結合性とは、次数の値が近い2ノード間にはあまりエッジが張られていないという性質である。すなわち、高い次数の2ノードには殆どエッジは存在していない。一方、高い次数のノードと低い次数のノードの間には優先的にエッジが張られている。非同類結合性は、次数 k のノードに直接結合しているノードを全部ピックアップして、それらのノードの平均次数、 $\langle k_{nn}(k) \rangle$ (nn が意味するのは最近接ノード、すなわち nearest neighbor である) を調べることで定義できる。もし、ネットワークが非同類結合性を有するのであれば、 k の値が小さい時に $\langle k_{nn}(k) \rangle$ の値は大きく、 k の値が大きいために $\langle k_{nn}(k) \rangle$ の値は小さくなるはずである。実際、PINにおいて、 $\langle k_{nn}(k) \rangle$ の分布を調べてみると、 $\langle k_{nn}(k) \rangle$ の値は k が大きくなるに従いべき乗則($\langle k_{nn}(k) \rangle \sim k^{-\gamma}$)に従って減少している[15-17]。

ところで、PINでは次数の高いノードとその回りのノードが一体となって一つの機能単位を構成していると考えられている[17]。このことから、次数の高いノード同士が直接結合していないということは、機能ユニット同士はネットワーク空間上で混線しないようにうまく分けられていることを示唆する[17]。しかしながら、最近、我々は中程度の次数を持つノード同士は非同類結合性を示さず、それらの中程度の次数を持つノードの間に優先的にエッジが張られていることを発見した[18]。これら、中程度のノードは転写制御に関するものが多く、密接にお互いを制御する必要があるために同類結合性を示しているのだと考えられる。

最後に、PINは階層性を有している[15,19]。階層性とは、次数が低いノードはクラスター係数が高く、次数が高くなればなるほど、そのノードのクラスター係数は低くなるという性質のことを言う。階層性は、次数が k であるような全ノードのクラスター係数の平均値($\langle C(k) \rangle$)を調べることで確認できる。代表的な例としては、階層性を有するネットワークでは、 $\langle C(k) \rangle$ はべき乗則、

$$\langle C(k) \rangle \sim k^{-\beta}$$

に従う[15,19]。例えば、代謝系のネットワークでは、 $\langle C(k) \rangle \sim k^{-1}$ である[13]。ただし、PINでは、 $\langle C(k) \rangle$ の値は k が大きくなるにつれて低くなるものの、完全なべき乗則に従うわけでは無い[20]。階層性の存在するネットワークでは、次数の高いノードの回りのノードはお互いに結合しておらず、次数の高いノードの回りのノード間の距離は、ネットワーク空間上において遠く離れていると思われる。言い換えると、次数の高いノードはネットワークの一部分にだけエッジを集中して張っているのではなく、全体に

満遍なくエッジを張り、ネットワーク全体を制御していると思われる。一方、次数の低いノードはネットワークの局所にしかエッジを張っておらず、ネットワークの局所のみを制御していると考えられる。

2. 3 タンパク質間相互作用ネットワークの成長モデル

スケールフリーネットワークは、新たに生成されたノードが次数の高いノードへのエッジを優先的に獲得することにより構築される。このようなネットワーク成長モデルのことを、優先選択モデルと呼ぶ[21]。優先選択モデルは1999年に Barabasi と Albert によって提唱されたが[21]、優先選択とべき乗則との間の関係性はより以前に提唱されていた[22-24]。

優先選択モデルでは、

(1). 新たなノードを1つ追加する

(2). 新しいノードは、ネットワーク中に以前から存在していたノードの内 m 個のノードを選んで、それらに対しエッジを形成する。ただし、ノードの次数 k の値に比例する確率でこれら m 個のノードを選ぶ。

この(1)と(2)のプロセスを繰り返すことで、スケールフリーネットワークが構築される。この様な過程で構築されたネットワークの次数分布は

$$P(k) = 2m^2 / k^3$$

となる。優先選択モデルによって構築されるネットワークの $P(k)$ の分布の導出については参考文献[21]を参照されたい。

以下で説明するように、PIN の進化では、遺伝子重複がスケールフリー性の形成に主要な役割を果たしていると考えられる[3]。あるノードが遺伝子重複により複製された結果、新たなノードが生成されるが、これらの相同な2ノードは全く同じノードに対しエッジを張っている。また、ネットワーク中のノードがランダムに選ばれた場合、そのノードが高い次数のノードにエッジを張っている確率は高い。従って、ランダムに選ばれたノードが遺伝子重複された場合、新たに生成されたノードと次数の高いノードの間に優先的にエッジが張られる[3]。

遺伝子重複を基にしたネットワーク成長モデルは、最初に Sole らにより提唱された[8]。彼らのモデルでは、遺伝子重複、分化、そして少数のランダムなエッジ構築の全部で3つの過程を使用する。分化の過程では、遺伝子重複で複製されたノードの持っているエッジそれぞれが一定の確率で取り除かれる。これが意味するのは、遺伝子重複で作られたペアには、大きな変異が入りやすいということである。これが何故かについて以下のように説明する。

タンパク質 A が重複され、タンパク質 A' が作られたと仮定すると、タンパク質 A' はタンパク質 A と同じ機能を持っているはずである。従って、タンパク質 A が変異を起してその機能を変異させたとしても、タンパク質 A' がタンパク質 A の果たすべき機能を補うので、この変異は生物の生存にとって致命的なものとはならない。従って、遺伝子重複で作られたペアには、遺伝子重複されていないタンパク質と比べて、より大きな変異が許容できる訳である。ここで注意しなくてはいけないのは、遺伝子重複されたペアの内、大きな変異を受けるのは、どちらか一方だけであるということである[25]。その理由は、タンパク質 A、A' の両者に大きな変異が入ったのであれば、タンパク質 A が担っていた機能がネットワ

ーク中から消えてしまう。従って、このような変異は生物の生存にとって致命的である。実際の PIN でも、遺伝子重複されたペアのうち、どちらか一方のタンパク質に対し変異が頻繁に入り、このタンパク質がもう一方のタンパク質より多くのエッジを失うことが知られている[26]。Sole のモデルでは、スケールフリー性を再現できる。しかしながら、高い平均クラスター係数を再現することが出来ないことが指摘されている[8,27,28]。Sole のモデルで構築されたネットワークにおける、 $P(k)$ の分布の解析的な導出については参考文献[29]を参照されたい。

この問題を解決するために、最近、遺伝子重複に基づいたネットワーク成長モデルである、ヘテロ 2 量体化 (Heterodimerization; HD) モデルが提唱された[28,30]。このモデルでは、遺伝子重複によるノードとエッジの複製の他に、分化と HD の二つの過程を使用する。HD では、遺伝子重複で作られた相同な 2 ノード間に一定の確率でエッジが張られる。この時、これらの相同な 2 ノードが最近接ノードを共有するのならば、三角形が出来る (図 5 A)。従って、HD モデルは、PIN のスケールフリー性と高い平均クラスター係数を再現出来る。ところで、この HD エッジは、自己結合するタンパク質が遺伝子重複された時にのみ生成される。自己結合するタンパク質とは、自分自身と結合するための結合部位を持っているものである。ここで、あるタンパク質 A が自己結合するための結合部位を持っている場合を考えて欲しい。このタンパク質 A が遺伝子重複で複製されてタンパク質 A'が出来たとすると、タンパク質 A、A'ともに自分自身と結合するための結合部位を持っているはずであり、この部位は殆ど同じ構造を取っているはずである。従って、この場合、この結合部位を使って、タンパク質 A はタンパク質 A'と結合することが出来る。最近、Middendorff らが、機械学習の方法を応用して HD モデルを含む 7 つのネットワークモデルを比較検討し、HD モデルがショウジョウバエの PIN の進化過程を最も良く反映していることを示した[27]。しかしながら、ショウジョウバエの PIN は、酵母、線虫、人の PIN と比べて、データの信頼性が低いことが指摘されている[31]。

最近、我々は、HD モデルが酵母の PIN の進化を反映していないことを示した[20]。そこで、我々は、HD モデルを修正した優先的ヘテロ 2 量体化(Non-Uniform Heterodimerization; NHD)モデルを提唱した。酵母の PIN には幾つかの HD モデルでは再現できない特徴が存在するが、これらの特徴は NHD モデルで再現出来ることを示した[20]。

2. 4 非一様ヘテロ量体化 [20] によるタンパク質間相互作用ネットワークの進化解析

モデルによる PIN のシミュレーションの詳細

酵母の PIN データ及び、タンパク質のアミノ酸配列データを MIPS データベース (<http://mips.gsf.de>)[32]からダウンロードした。また、相同な 2 つのタンパク質間の相互作用 (HD エッジ) の数を推定するために BLASTP プログラム[33]を用いた。

シミュレーションでは、ネットワーク中のノード数が酵母の PIN のノード数と等しくなるまで、NHD もしくは HD モデルに従ってネットワークを成長させた。これらのモデルでは全部で 2 つのパラメーター (複製されたエッジを消す確率 α および遺伝子重複で作られた相同な 2 ノードの間にエッジを張る確率 β) を使用している (図 5 A)。生成したネットワークの平均クラスター係数の値とエッジ数が、酵母の

PIN でのそれぞれの値とほぼ等しくなるように、これら 2 つのパラメーターの値を調整した。また、遺伝子重複により相同な 2 ノードが作られた直後に、それらのノードのアミノ酸配列に変異が起きると仮定した。従って、遺伝子重複により相同な 2 ノードが生成された場合、それらのノードと他の全てのノードとの間の進化距離を 1 増加させた。また、遺伝子重複で生成された相同な 2 ノード間の進化距離を 1 とした (図 5 B)。シミュレーションで生成したネットワーク中では、2 ノード間の進化距離が 3 (もしくは 4) より小さい場合、それらの 2 ノードを相同なペアと定義した。この時、HD 及び NHD ネットワーク中の相同なペア数は酵母の PIN の値とほぼ等しい(表 1)。

表1. NHD、HDネットワークとPINにおけるネットワーク統計量[20]

Model	d_T	α^a	β^a	n_H^b	n_{HI}^c	n_{HI}/n_H	$\langle k \rangle^d$	$\langle C \rangle^e$	$\langle L \rangle^f$
HD model	1	0.725	0.061	1,312 (11)	140 (12)	0.107 (0.009)	3.73 (0.09)	0.066 (0.006)	6.45 (0.14)
	2	-	-	3,031 (27)	269 (19)	0.089 (0.006)	-	-	-
	3	-	-	5,309 (43)	395 (25)	0.074 (0.005)	-	-	-
	4	-	-	8,337 (65)	514 (31)	0.062 (0.004)	-	-	-
	5	-	-	12,363 (92)	628 (42)	0.051 (0.003)	-	-	-
NHD model	1	0.745	0.028	1,308 (11)	52 (6)	0.040 (0.005)	3.74 (0.07)	0.066 (0.006)	6.23 (0.12)
	2	-	-	3,030 (22)	105 (11)	0.035 (0.004)	-	-	-
	3	-	-	5,315 (42)	157 (17)	0.029 (0.003)	-	-	-
	4	-	-	8,351 (61)	208 (21)	0.025 (0.003)	-	-	-
	5	-	-	12,373 (86)	259 (28)	0.021 (0.002)	-	-	-
Yeast PIN ^g				6,544	175	0.027	3.74	0.066	4.85
Random ^h							3.74	0.00096	6.27

括弧の中の数字は、100個のネットワークについて計算した標準誤差。-は上の値と同じ値であることをしめす。

a. シミュレーションで使用したパラメータ

b. 相同なペアの数。2 ノード間の進化距離が d_T 以下であれば、その 2 ノードを相同なペアと定義する。

c. 相同な 2 ノードをつなぐ HD エッジの数。

d. 平均次数。

e. 平均クラスター係数。

f. 平均最短経路長。

g. 自己結合エッジを除いた酵母の PIN。

h. 酵母の PIN と同じ平均次数 ($\langle k \rangle = 3.74$) およびノード数 ($N = 3,891$) を持つランダムネットワーク。ランダムネットワークの $\langle C \rangle$ および L は、数式 $\langle C \rangle = \langle k \rangle / N$ [42] および $L = \log N / \log \langle k \rangle$ [40] を用いて計算した。

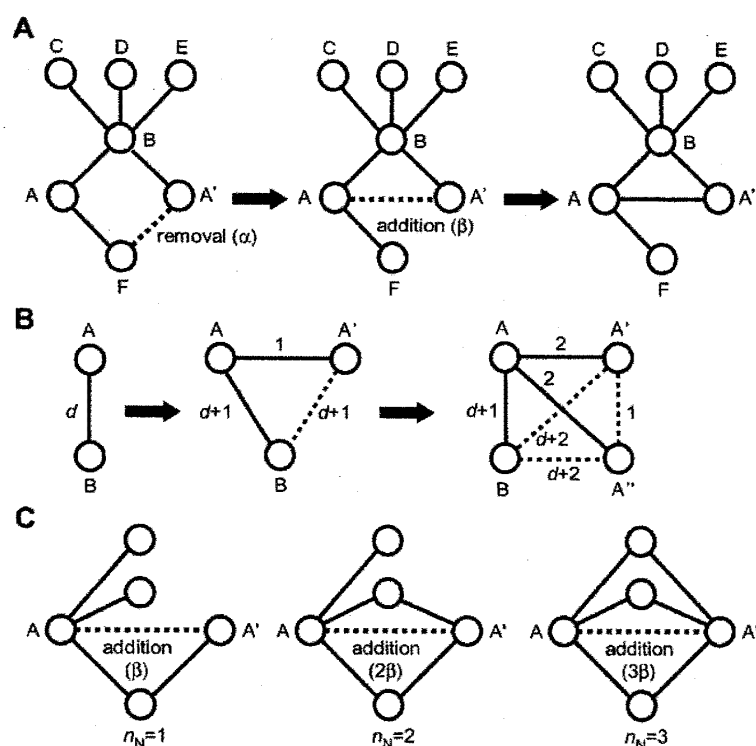


図5. シミュレーション[20]

(A) HD モデル。ノード A が遺伝子重複されてノード A' が生成される。ノード A' のそれぞれのエッジが一定の確率 α で取り除かれる (左図)。遺伝子重複で作られた相同な 2 ノードの内、どちらか片方のノード (この場合は A') だけしかエッジを失わないという点に注意して欲しい。ノード A とノード A' の間に、確率 β で HD エッジが張られる (中央図)。(B) 進化距離。ノードが遺伝子重複された時、相同なペアに変異が生じると仮定する。従って、この時、相同な 2 ノードとネットワーク中のその他のノードとの間の進化距離を 1 だけ増やす。ノード A とノード B の進化距離を d とする (左図)。A が遺伝子重複され A' が生成され時、ノード A とノード A' に変異が生じる。従って、この時、ノード A とノード B の進化距離は $d+1$ に、ノード A' とノード B の進化距離は $d+2$ となる (中央図)。右図は、ノード A とノード A' の間にエッジが張られる場合を示す。

ノード B の進化距離は $d+1$ となる。進化距離は 2 ノードの間にエッジがあるかどうかに影響されないと仮定している。(2 ノード間の点線は 2 ノード間にエッジが無いことを表し、直線はエッジが存在することを示す。) ノード A とノード A' の進化距離は 1 とする。そして、今度は、ノード A が遺伝子重複されてノード A' が生成された場合を考える。この時の 2 ノード間の進化距離は右図に示す通りになる。(C) NHD モデル。NHD モデルでは、相同な 2 ノード(A と A') の間に HD エッジが張られる確率は、これらのノードが共有する最近接ノードの数(n_N)に比例して増加する。

非一様ヘテロ量体化 (NHD) モデル [20]

シミュレーションの結果、HD モデルで酵母の PIN と同等の平均クラスター係数を再現するために必要な HD エッジの数は、酵母 PIN 中の HD エッジの数と比べて非常に多いことが判った(表 1)。従って、HD モデルは酵母の PIN の進化を反映していない。ところで、相同な 2 ノードが 3 つの最近接ノードを共有している場合、相同な 2 ノード間に HD エッジが張られれば、3 つの三角形が新たに生成される (図 5 C)。従って、相同な 2 ノードが多くの最近接ノードを共有している時に、相同な 2 ノード間に優先的にエッジを張るのならば、HD モデルよりも少ない HD エッジ数で、酵母の PIN と同等の平均クラスター係数を再現出来るはずである。そこで、我々は、HD モデルを修正した NHD モデルを提唱した。NHD モデルでは、遺伝子重複によるノードとエッジの複製の他に、分化と NHD を用いる。NHD では、遺伝子重複で作られた相同な 2 ノード間に HD エッジが張られる確率が、これらの 2 ノードが共有している最近接ノード数に比例する(図 1C)。例えば、相同な 2 ノードが 3 つの最近接ノードを共有していれば、確率 3β で相同な 2 ノード間に HD エッジを構築する(図 5 C)。ところで、もし相同な 2 ノードが n_N 個

の最近接ノードを共有しており、この m_N の値が極めて大きければ確率 β_{mN} は 1 を超えてしまう。そこで、もし $\beta_{mN} > 1$ となる場合は、相同な 2 ノード間に HD エッジを構築する確率を 1 とした。しかしながら、実際のシミュレーションで $\beta_{mN} > 1$ となるような場合は起こらなかった。NHD モデルで酵母の PIN と同等のクラスター係数を再現するのに必要な HD エッジ数は、実際の酵母の PIN での数とほぼ等しかった(表 1)。

NHD ネットワークでの平均最短経路長はほぼランダムネットワークに等しかった。従って、NHD モデルはスモールワールド性を再現出来る。しかしながら、大変興味深いことに、PIN の平均最短経路長は、ランダムネットワークでの平均最短経路長よりもかなり低い(表 1)。また、 $k < 50$ の範囲では、酵母 PIN での確率分布 $P(k)$ は NHD および HD ネットワークでの確率分布 $P(k)$ とほぼ等しかった(図 6 A)。ここで注目して欲しいのは、PIN、HD および NHD ネットワークにおいて $P(k)$ の分布は単純なべき乗則 $P(k) \sim k^{-\gamma}$ ではなく、exponential cut off 項が加わった $P(k) \sim (k_0 + k)^{-\gamma} e^{-k/k_c}$ に従うということである。大変興味深いことに、この PIN の $P(k)$ の分布は優先的選択モデルでは再現できないが、遺伝子重複を基にしたモデルで再現出来る。しかしながら、50 以上の次数を持つノードの数は、PIN(17)の方が NHD ネットワーク(5.8)よりもかなり多い(図 6 A)。

$\langle C(k) \rangle$ の分布についてであるが、NHD、HD ネットワークともにべき乗則 $\langle C(k) \rangle \sim k^{-\beta}$ に従う(図 6 B)。このことから、NHD、HD ネットワークともに階層性を有するネットワークであることが示された。また、 $\langle k_{nn}(k) \rangle$ についてであるが、NHD、HD、PIN ともにべき乗則 $\langle k_{nn}(k) \rangle \sim k^{-\gamma}$ に従う。しかしながら、NHD、HD ネットワークでの ν の値は PIN での ν の値に比べてかなり小さい(図 6 C)。この結果から、NHD ネットワークでは、低い次数のノードと高い次数のノードの間にエッジが多いが、PIN 中では、このようなエッジの数は更に多いことが判る。

ネットワーク中における、三角形の空間的な分布を調べるために、われわれは、あるエッジが m_T 個の三角形に含まれる確率 $P_T(m_T)$ を調べた(図 6 D)。例えば、図 5 C (中央図) の場合では、ノード A とノード A' をつなぐエッジにおける m_T の値は 2 となる。確率分布 $P_T(m_T)$ を導入することにより、ネットワーク中における三角形の空間的な偏りを統計的に測定することができる。酵母の PIN での $P_T(m_T)$ の分布は、HD ネットワークでの分布と大きく異なっていた。HD モデルでは、HD エッジが相同な 2 ノード間にランダムに生成されるので、HD ネットワーク中の三角形の分布はランダムである。従って、酵母の PIN での三角形の分布はランダムでは無く、三角形はネットワークの一部に集中して存在している。NHD ネットワーク中の $P_T(m_T)$ の分布は、酵母の PIN での分布と良く一致した。

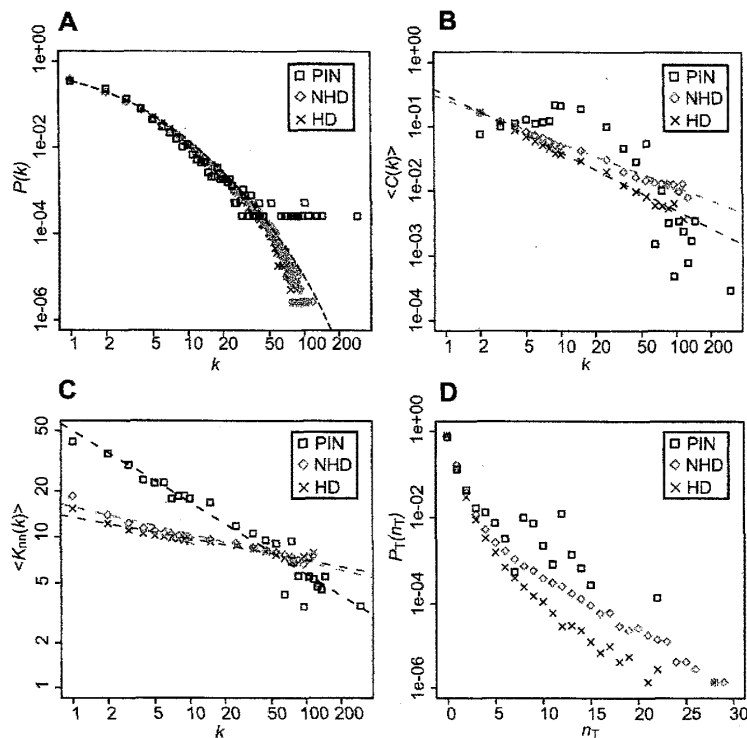


図6. NHDおよびHDモデルで構築されたネットワークの特徴

四角形はPIN、ひし形はNHDモデルで構築されたネットワーク、ばつ印はHDモデルで構築されたネットワークの結果を表す。NHDおよびHDモデルの結果については100回のシミュレーションの平均値を用いた。(A) 次数分布 $P(k)$ 。点線は $P(k) \sim (k_0 + k)^{-\gamma} e^{-k/k_c}$ ($\gamma=2.7$, $k_0=3.4$, $k_c=50$)を表す。(B) 平均クラスター係数 $\langle C(k) \rangle$ の分布。直線は $\sim k^{-0.68}$, および $\sim k^{-0.90}$ を表している。(C) $\langle k_{nn}(k) \rangle$ の分布。直線は $\sim k^{0.47}$, $\sim k^{0.18}$, および $\sim k^{0.14}$ を表している。(D) $P_T(m)$ の分布。

変異による遺伝子喪失の影響について[20]

NHDモデルには、変異による遺伝子喪失の過程は含まれていない。ここでは、この過程をNHDモデルにとりいれて(NHD+Eモデル)、その影響について調べてみることにする。NHD+Eモデルでは、ネットワーク中に新たなエッジが加わる度に、ランダムに選んだノードとそのノードの全てのエッジを確率 δ で取り除く。結果は図7と表2のようになり、NHDネットワーク中のノードをランダムに消しても殆ど結果に変化は生じなかった。従って、変異による遺伝子喪失の過程は（もし変異がランダムに起こるのであれば）、PINの進化にそれほど大きな影響を与えないと考えられる。

表2. NHDおよびNHD+Eモデルにより構築されたネットワークにおける統計量[20]

Model	δ^a	d_T	α^b	β^b	n_H^b	n_{H1}^b	n_{H1}/n_H	$\langle k \rangle^b$	$\langle C \rangle^b$	$\langle L \rangle^b$
NHD	0	3	0.745	0.028	5,315 (42)	157 (17)	0.029 (0.003)	3.74 (0.07)	0.066 (0.006)	6.23 (0.12)
	-	4	-	-	8,351 (61)	208 (21)	0.025 (0.003)	-	-	-
NHD+E	0.001	3	0.745	0.028	5,309 (40)	157 (16)	0.029 (0.003)	3.73 (0.09)	0.066 (0.006)	6.23 (0.13)
	-	4	-	-	8,336 (72)	208 (20)	0.025 (0.002)	-	-	-
	0.01	3	0.74	0.028	5,266 (39)	159 (15)	0.030 (0.003)	3.73 (0.08)	0.066 (0.006)	6.31 (0.14)
	-	4	-	-	8,253 (71)	211 (19)	0.026 (0.002)	-	-	-
	0.1	3	0.695	0.026	4,862 (48)	158 (16)	0.032 (0.003)	3.76 (0.10)	0.068 (0.007)	7.03 (0.39)
	-	4	-	-	7,446 (71)	210 (19)	0.028 (0.003)	-	-	-
Yeast PIN ^b					6,544	175	0.027	3.74	0.066	4.85

括弧の中の数字は、100個のネットワークについて計算した標準誤差。-は上の値と同じ値であることをしめす。

- a. 各ステップごとにネットワーク中からランダムにノードを選び、そのノードを確率 δ で取り除いた。
b. 表1を参照のこと。

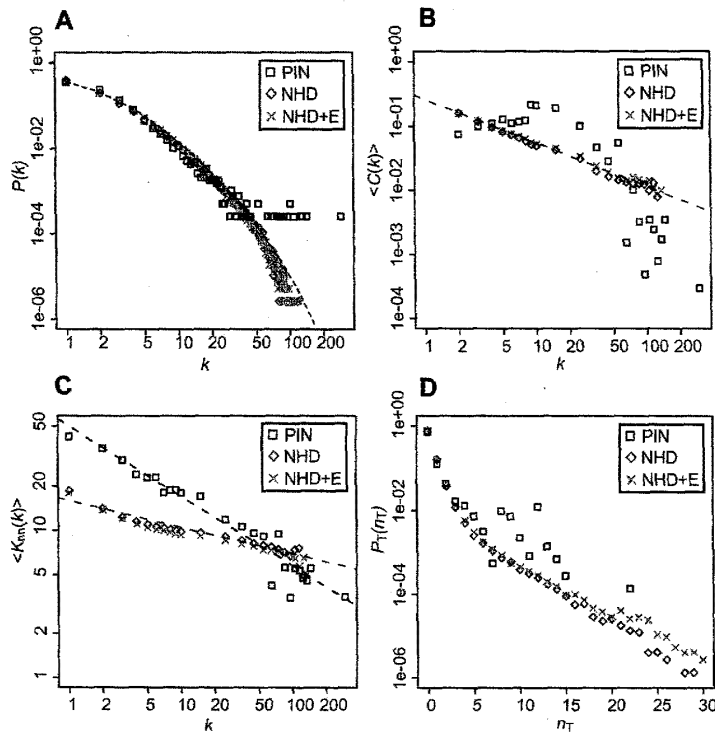


図7. NHD および NHD+E モデルで構築されたネットワークの特徴

四角形は PIN、ひし形は NHD モデルで構築されたネットワーク、ばつ印は NHD+E ($\delta=0.1$) モデルで構築されたネットワークの結果を表す。NHD および NHD+E モデルの結果については100回のシミュレーションの平均値を用いた。(A) 次数分布 $P(k)$ 。点線は $P(k) \sim (k_0 + k)^{-\gamma} e^{-k/k_0}$ ($\gamma=2.7$, $k_0=3.4$, $k_0=50$) を表す。(B) 平均クラスター係数 $\langle C(k) \rangle$ の分布。直線は $\sim k^{-0.68}$ を表している。(C) $\langle k_{nn}(k) \rangle$ の分布。直線は $\sim k^{-0.47}$ 、および $\sim k^{-0.18}$ を表している。(D) $P_T(m)$ の分布。

NHD モデルの生物学的正当性について [20]

我々は、NHD モデルにより、酵母の PIN の高い平均クラスター係数、少ない HD エッジ数、そして三角形の空間的な分布の偏りを再現できることを示した。このことは、NHD モデルが PIN の進化を反映していることを示唆する。

NHD モデルの生物学的正当性については、以下のように説明できる。HD エッジは自己結合するノードが遺伝子重複することで作られる。NHD モデルでは、一定の確率で自己結合するノードが選ばれて遺伝子重複すると仮定しているが、この確率は、酵母の PIN においてランダムに選ばれたノードが自己結合している確率とほぼ等しい。また、NHD モデルでは、自己結合するノードが遺伝子重複され、その時に作られた相同な 2 ノードが多数の最近接ノードを共有しているのならば、相同な 2 ノード間の HD エッジは優先的に保存されると仮定している。図 8A は 2 つの相同なノードが n_N 個の最近接ノードを共有する時にその 2 ノードの間に HD エッジが存在している確率 $P_{HD}(n_N)$ の分布を示している。この図から、NHD ネットワーク、PIN の両者ともに、 $P_{HD}(n_N)$ の値は n_N が増加するにつれて高くなっていることがわかる (HD ネットワークでは $P_{HD}(n_N)$ の値は n_N の値によらず一定)。このことは、HD エッジは PIN の進化の仮定で非一様に保存されていることを示唆する。

ところで、NHD モデルにおける HD エッジの優先的保存に関する仮定は、PIN では次数が高いノードの進化速度は遅いという事実[34-36]により、以下の様に正当化される。ノードの進化速度が遅い場合、そのノードが他のノードと結合する際に使用する結合部位は変異を起こし難いと考えられる。従って、高い次数のノードの結合部位は変異を起こし難いと推測される。酵母の PIN では、相同な 2 ノードが共有している最近接ノードが多い場合、相同な 2 ノードの次数が高いことが確認できた (図 8B)。従って、相同な 2 ノードが多く最近接ノードを共有する場合、相同な 2 ノード間の HD エッジを構成する結合部位は変異を起こし難く優先的に保存されると考えられる。

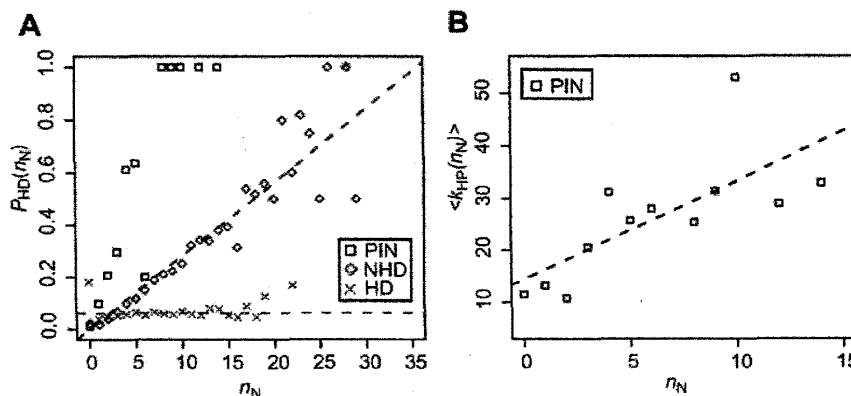


図 8. シミュレーションで作成したネットワークおよび PIN のにおける HD エッジについて

黒い四角形は PIN、赤いひし形は NHD モデルで構築されたネットワーク、緑のばつ印は HD モデルで構築されたネットワークの結果を表す。

NHD および HD モデルの結果については 100 回のシミュレーションの

平均値を用いた。(A) $P_{HD}(n_N)$ の分布。(相同な 2 ノードが n_N 個の最近接ノードを共有する時、その相同な 2 ノード間に HD エッジが張られている確率を $P_{HD}(n_N)$ と表す。)図は $d_t=3$ の場合の結果について表している。直線の傾きは 0.028 (赤) と 0 (緑) である。図に表示はしなかったが、 $d_t=4$ の場合の結果は $d_t=3$ の場合と殆ど変わらなかった。(B) $\langle k_{HP}(n_N) \rangle$ の分布。 $\langle k_{HP}(n_N) \rangle$ は、相同なタンパク質と n_N 個の最近接ノードを共有しており、かつ HD エッジに結合されているような全タンパク質の平均次数を表す。直線は回帰直線である ($r=0.73$)。

PIN の平均最短経路長と次数の高いノードの数との関連について [20]

PIN は一般的にスモールワールドネットワークと考えられてきた[5,8]。しかしながら、上記のように、PIN の平均最短経路長はランダムネットワークよりもかなり小さい。このような、ランダムネットワークよりも小さい平均最短経路長を持つようなネットワークはウルTRASモールと呼ばれている[37,38]。ところで、スケールフリー性を持ったランダムネットワークはウルTRASモールである事が知られている[37,38]。そこで、Maslov らの提唱した方法を使用して[17]、PIN のエッジを $P(k)$ の分布を保ったままランダムに繋ぎ変え、スケールフリーランダムネットワークを作成した。このランダム化した PIN の平均最短経路長は、ランダム化していない PIN の値と殆ど変わらなかった。しかし、ランダム化した PIN の平均クラスター係数は、ランダム化していない PIN よりも非常に小さい値となった。以上のことから、PIN はスケールフリーランダムネットワークとは全く違う構造を取っているものの、大変興味深いことに、ウルTRASモールである。

では、何故、PIN はウルTRASモールなのであろうか？PIN から次数の高いノードを取り除いた場合、PIN の平均最短経路長は非常に大きくなることが知られている[39]。従って、PIN の非常に小さい平均最短経路長は、PIN 中に多数の次数が大きいノードが存在することが原因ではないかと考えられる。（次数が 50 以上のノードは PIN では 17 個であるが、NHD ネットワークでは 5.8 個である。）実際に、次数が 50 以上のノード全てを、PIN 及び NHD ネットワークの両方から取り除いたところ、PIN の平均最短経路長(6.13)と平均クラスター係数の値(0.063)は、NHD ネットワークでの値(6.51 と 0.060)と殆ど等しくなった。このことから、次数の高いノードの数が、PIN の平均最短経路長に多大な影響を与えていることがわかる。

まとめ

NHD モデルは2つのパラメーターのみを使用する単純なモデルであるが、従来の HD モデルでは説明できなかった酵母の PIN の様々な特徴を捉えることができる。NHD モデルが PIN の進化を良く再現するという事実は、タンパク質の結合部位の変異はネットワーク構造からの拘束を受けるということである。次の章では、PIN のネットワーク構造が、タンパク質のアミノ酸配列の変異のしやすさに対し、どのような影響を与えているのかについて精査し議論する。

ところで、現在、PIN の全構造のうち限られた構造のみしか判っていないと考えられている。PIN の未知の構造を明らかにして、それをより詳しく検討することで、未知の進化メカニズムを発見できるかもしれない。また、今後、精度の高い様々な生物種のタンパク質間相互作用ネットワークが利用可能となることが期待されており、それらを比較検討することで、よりリアリスティックな数理的進化モデルの構築が可能になると考えられる。

参考文献

- [1] Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S. et al. (2000), "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*," *Nature*, Vol. 403, pp.623-627
- [2] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. et al. (2001), "A comprehensive

- two-hybrid analysis to explore the yeast protein interactome," *Proc. Natl. Acad. Sci. USA.*, Vol. 98, pp.4569-4574
- [3] Barabasi, A.L. and Oltvai, Z.N. (2004), "Network biology: understanding the cell's functional organization," *Nat. Rev. Genet.*, Vol.5, pp.101-113
- [4] Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001), "Lethality and centrality in protein networks," *Nature*, Vol.411, pp.41-42
- [5] Wagner, A. (2001), "The yeast protein interaction network evolves rapidly and contains few redundant genes," *Mol. Biol. Evol.*, Vol.18, pp.1283-1292
- [6] Albert, R., Jeong, H. and Barabasi, A.L. (1999), "Diameter of the World-Wide Web," *Nature*, Vol.401, pp.130-131
- [7] Albert, R., Jeong, H. and Barabasi, A.L. (2000), "Error and attack tolerance of complex networks," *Nature*, Vol.406, pp.378-382
- [8] Sole, R.V., Pastor-Satorras, R., Smith, E.D. and Kepler, T.B. (2002), "A model of large-scale proteome evolution," *Adv. Comput. Syst.*, Vol.5, pp.43-54
- [9] Milgram, S. (1967), "The small world problem," *Psychol. Today*, Vol.2, pp.60-67
- [10] Watts, D.J. and Strogatz, S.H. (1998), "Collective dynamics of 'small-world' networks," *Nature*, Vol.393, pp.440-442
- [11] Newman, M.E.J. (2001), "The structure of scientific collaboration networks," *Proc. Natl. Acad. Sci. USA.*, Vol.98, pp.404-409
- [12] Eguiluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M. and Apkarian, A.V. (2005), "Scale-free-brain functional networks," *Phys. Rev. Lett.*, Vol.94, pp.018102
- [13] Williams, R.J., Martinez, N.D., Berlow, E.L., Dunne, J.A. and Barabasi, A.L. (2002), "Hierarchical organization of modularity in metabolic networks," *Science*, Vol.297, pp.1551-1555
- [14] van Noort V., Snel, B. and Huynen, M.A. (2004), "The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model," *EMBO Rep.*, Vol.5, pp.280-284
- [15] Vazquez, A. (2003), "Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations," *Phys. Rev. E.*, Vol.67, pp.056104
- [16] Costa, L.F., Rodrigues, F.A., Travieso, G. and Boas, P.R.V. (2007), "Characterization of complex networks: A survey of measurement," *Adv. Phys.*, Vol.56, pp.167-242
- [17] Maslov, S. and Sneppen, K. (2002), "Specificity and stability in topology of protein networks," *Science*, Vol.296, pp.910-913
- [18] Suzuki, Y., Hase, T., Nakagawa, S., Niimura, Y., Ogishima, S. and Tanaka, H. (Submitted)
- [19] Yook, S.H., Oltvai, Z.N. and Batabasi, A.L. (2004), "Functional and topological characterization of protein interaction networks," *Proteomics*, Vol.4, pp.928-942
- [20] Hase, T., Niimura, Y., Kaminuma, T. and Tanaka, H. (2008), "Non-uniform survival rate of

- heterodimerization links in the evolution of the yeast protein-protein interaction network,” *PLoS One*, Vol.3, pp.e1667
- [21] Barabasi, A.L. and Albert, R. (1999), “Emergence of scaling in random networks,” *Science*, Vol.286, pp.509-512
- [22] Simon, H.A. (1955), “On a class of skew distribution functions,” *Biometrika*, Vol.42, pp.425-440
- [23] D.J. de Solla Price (1965), “Networks of scientific papers,” *Science*, Vol.149, pp.510-515
- [24] D.J. de Solla Price (1976), “A general theory of bibliometric and other cumulative advantage process,” *Journal of American Society of Information Science*, Vol.27, pp.292-306
- [25] Kellis, M., Birren, B.W. and Lander, E.S. (2004), “Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*,” *Nature*, Vol.428, pp.617-624
- [26] Wagner, A. (2002), “Asymmetric functional divergence of duplicate genes in yeast,” *Mol. Biol. Evol.*, Vol.19, pp.1760-1768
- [27] Middendorff, M., Ziv, E. and Wiggins, C.H. (2005), “Inferring network mechanisms: The *Drosophila melanogaster* protein interaction network,” *Proc. Natl. Acad. Sci. USA*, Vol.102, pp.3192-3197
- [28] Ispolatov, I., Krapivsky, P.L., Mazo, I. and Yuryev, A. (2005), “Cliques and duplication-divergence network growth,” *New J. Phys.*, Vol.7, pp.145
- [29] Pastor-Satorras, R., Smith, R. and Sole, R.V. (2003), “Evolving protein interaction networks through gene duplication,” *J. Theor. Biol.*, Vol.222, pp.199-210
- [30] Vazquez, A., Flammini, A., Maritan, A. and Vespignani, A. (2003), “Modeling of protein interaction networks,” *ComplexUs*, Vol.1, pp.38-44
- [31] Patil, A. and Nakamura, H. (2005), “Filtering high-throughput protein-protein interaction data using a combination of genomic features,” *BMC Bioinformatics*, Vol.6, pp.100
- [32] Guldener, U., Munsterkotter, M., Oesterheld, M., Pagel, P., Ruepp, A. et al. (2006), “Mpaact: The MIPS protein interaction resource on yeast,” *Nucleic. Acids. Res.*, Vol.34, pp.D436-441
- [33] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990), “Basic local alignment search tool,” *J. Mol. Biol.*, Vol.215, pp.403-410
- [34] Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C. and Feldman, M.W. (2002), “Evolutionary rate in the protein interaction network,” *Science*, Vol.296, pp.750-752
- [35] Fraser, H.B. et al. (2003), “A simple dependence between protein evolution rate and the number of protein-protein interactions,” *BMC Evol. Biol.*, Vol.3, pp.11-
- [36] Fraser, H.B. (2005), “Modularity and evolutionary constraint on proteins,” *Nature Genetics*, Vol.37, pp.351-352
- [37] Chung, F. and Lu, L. (2002), “The average distances in random graphs with given expected degrees,” *Proc. Natl. Acad. Sci. USA*, Vol.99, pp.15879-15882
- [38] Cohen, R. and Havlin, S. (2003), “Scale-free networks are ultrasmall,” *Phys. Rev. Lett.*, Vol.90,

pp.058701

- [39] Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F. et al. (2004), "Evidence for dynamically organized modularity in the yeast protein-protein interaction network," *Nature*, Vol.430, pp.88-93
- [40] Albert, R., Barabasi, A.L. (2002), "Statistical mechanics of complex network," *Reviews of Modern Physics*, Vol.74, pp.47-97

第3章 生命ネットワークの進化におけるシステムの拘束とシステム進化生物学

システム進化生物学—システムの拘束性

従来行われてきた分子進化学というものは個々の遺伝子配列の変異に基づき、中立または選択進化を考察するという方法論に基づいていた。一方、システム進化生物学では遺伝子ないしタンパク質のネットワークを単位として進化の過程を考えていく方法論に基づいている。進化学というとただ単に過去に生物が辿ってきた遍歴を遡っていくだけの学問と思われがちであるが、最近では合成生物学などの生命システム工学の分野から生命システムのグランドデザインを解明できるのではないかと期待されている。

システム生物学の観点からすると生命は始原的な分子ネットワークとして発生し、これが徐々に複雑化することで進化していったと考えられる。その場合、多くの相互作用で結合した分子は変異や遺伝子重複が起これにくいと考えられる。このことは、ネットワークレベルでのシステムの拘束の存在を示唆するのではないだろうか？

これを検証するために二つの仮説を立てた。

まず1つ目は、システムの拘束が分子進化速度(遺伝子が機能分化する速さ)に影響を与えるのではないかとということである。分子進化速度とは一定期間に起こる核酸あるいはアミノ酸配列の変異であり、高い次数を持つタンパク質はシステムの中で重要であるため進化速度が遅いのではないかと考えられる。

二つ目はネットワークレベルで進化速度のようなものが見られるのではないかとという仮説である。つまり、密に連絡を取り合っているモジュール（個々クラスターを形成し、別々の機能を果たしている部分ネットワーク）は進化的拘束があるために進化速度が遅くなるのではないかと考えられる。

遺伝子レベルの進化的拘束性

それでは一つ目の仮説について検証する。

Fraser や Hahn らは、様々なネットワーク指標と分子進化速度の相関について調査した[1,2]。Fraser らは次数と進化速度の相関について、Hahn らは次数に加え、媒介中心性(Betweenness)と近接中心性(Closeness)の相関について調査した。

媒介中心性と近接中心性はいずれもネットワーク指標であり、媒介中心性は(1)式、近接中心性は(2)式で定義され、各々、特定のノードの媒介の度合いと他ノードとの近接の度合いを示している。すなわち、ノード n の媒介中心性 $C_b(n)$ は

$$C_b(n) = \sum_{s \neq n \neq t} (\sigma_{st}(n) / \sigma_{st}) \quad (1)$$

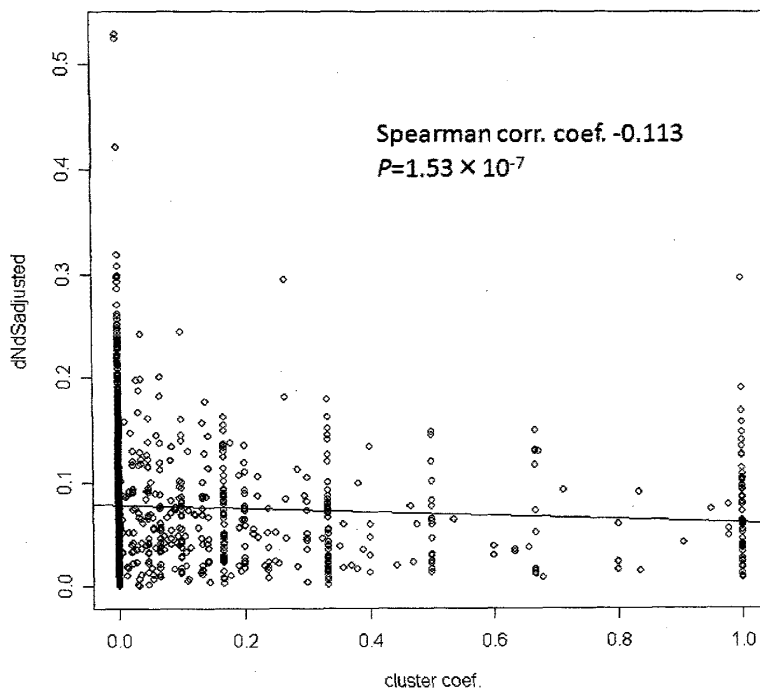
ここで、 $\sigma_{st}(n)$ はノード s, t をつなぐ最短経路においてノード n を通る回数、さらに σ_{st} はノード s, t 間の最短経路数である。また、ノード n の近接中心性 $C_c(n)$ は、

$$C_c(n) = 1 / \text{avg}(L(n, m)) \quad (2)$$

ここで、 $\text{avg}(L(n, m))$ は、2 ノード間の最短経路数の平均値である。

図1および表1に示すように、いずれの指標も進化速度に対し有意に弱い負の相関を示していることがわかる。しかし、一方で相関が弱いためこれを否定する意見も存在する[3]。

今回、我々は新たにクラスター係数（前章参照）と進化速度の相関を調査した(図1)。その結果、相関係数-0.113と低い値ではあるが、負の相関を示すことを発見した。



Makino らの報告によると、重複遺伝子は隣接ノードを共有している場合は進化速度が有意に遅い[4]。このことは、クラスター係数と進化速度の間に負の相関が存在することと矛盾しない。

以上の結果から、タンパク質がネットワークレベルで拘束を受ける場合、そのタンパク質の進化速度は遅くなる傾向があることがわかる。

また、進化速度の遅いタンパク質がどのような機能を有しているかを調べたところ、エネルギーやアミノ酸代謝、その他の代謝に関するタンパク質であることが判明した(図2)。

図1. クラスター係数と進化速度の相関

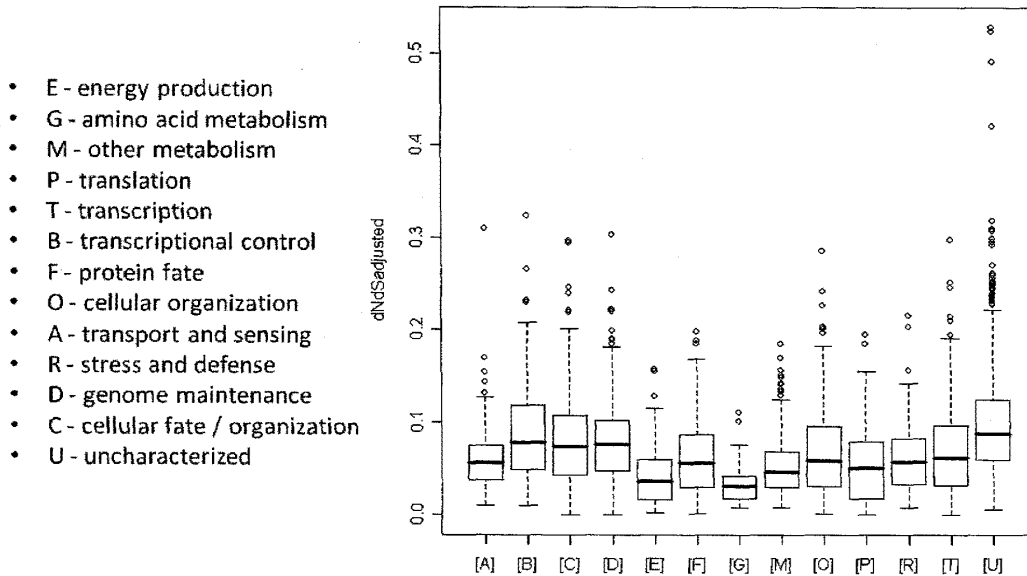


図 2. 機能分類ごとの分子進化速度比較

ところで、常に相互作用しているハブはパーティハブ、ある時間ごとに相互作用する相手を変え、全ての時間を通したときにハブに見えるものはデートハブと呼ばれている。そこで、我々は、デートハブおよびパーティハブの進化速度とクラスター係数を調査した (図 3)。すると、パーティハブの方が有意にデートハブよりも進化速度が遅く、クラスター係数が低いことがわかった。

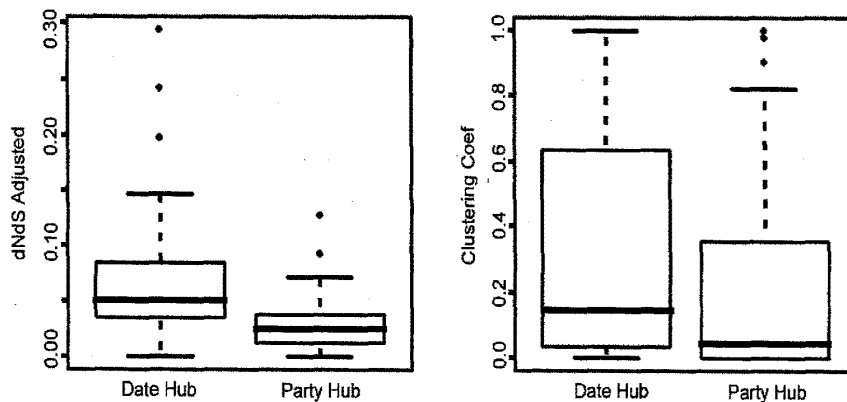


図 3. パーティハブとデートハブの比較

ネットワーク進化速度

次に、2つ目の仮説について検証する。

ここでもう一度分子進化速度について振り返る。分子進化速度という概念は分子時計の一定性を基にしている。(二種類の脊椎動物のアミノ酸配列を比較したときのアミノ酸置換数(進化距離)と、二生物種

の分岐時期をグラフにした際、両者の関係が直線関係になることから、進化は一定のペースで起こること、すなわち分子時計の一定性が確認された。) 分子時計の一定性が成立することから、進化距離を分岐時間で割ることで分子進化速度が求められるが、ネットワークの進化の場合、このような礎がない。よって、一部のネットワークの普遍性から進化可能性を考えられないかと考えた。そこで、現在求まるネットワーク指標に着目し、進化的な変化を観察していく。

Li らの報告[5]によると、近縁種で同定された重複遺伝子は非重複遺伝子と比較するとクラスター係数が有意に低いということが示されている。このことは、重複遺伝子は機能が重複しており、片方は機能的制約を受けないので他の遺伝子と多く相互作用しないことを示唆する。一方、高いモジュラリティを持つ遺伝子は重複しにくく、モジュラー性が遺伝子重複能に進化的拘束を与えていると推測される。

そこで私たちは近縁種以外の生物種で、構成タンパク質からオーソログを同定し以下の四つの進化的出現時期を推定した。

- Prokaryotes(原核生物): バクテリア
- Eukaryotes(真核生物): ヒト, ショウジョウバエ, 線虫, シロイヌナズナなど
- Fungi(菌類): 分裂酵母, カンジダ・アルビカンズ
- Yeast(出芽酵母): 出芽酵母

各々の時期に出現したタンパク質をネットワーク指標の面から比較した(図 4A,B)。

図 4A に結合次数を示す。進化的に古い始原的なネットワークは最低限のモジュールの組み合わせであると考えられるため、結合次数の高いタンパク質が多いと考えられるが、実際にはそういった傾向は存在せず、また Eukaryotes と Fungi の年代で比較的次數の大きいタンパク質が出現していることがわかる。

図 4B にクラスター係数を示す。Eukaryotes から Yeast の年代にかけてクラスター係数の高いものが徐々に少なくなっており、平均クラスター係数も同様に低下している。このことから、原始的な生物は最近になって出現した生物よりも密なネットワークを構成していたと推測される。また、媒介中心性と近接中心性についても同様に比較したが、特に関連性は見られなかった。

以上の結果からクラスター係数が高いほどシステムの拘束が強く、進化速度が遅くなることが示唆される。このことは、Li らの報告[5]とも一致する。

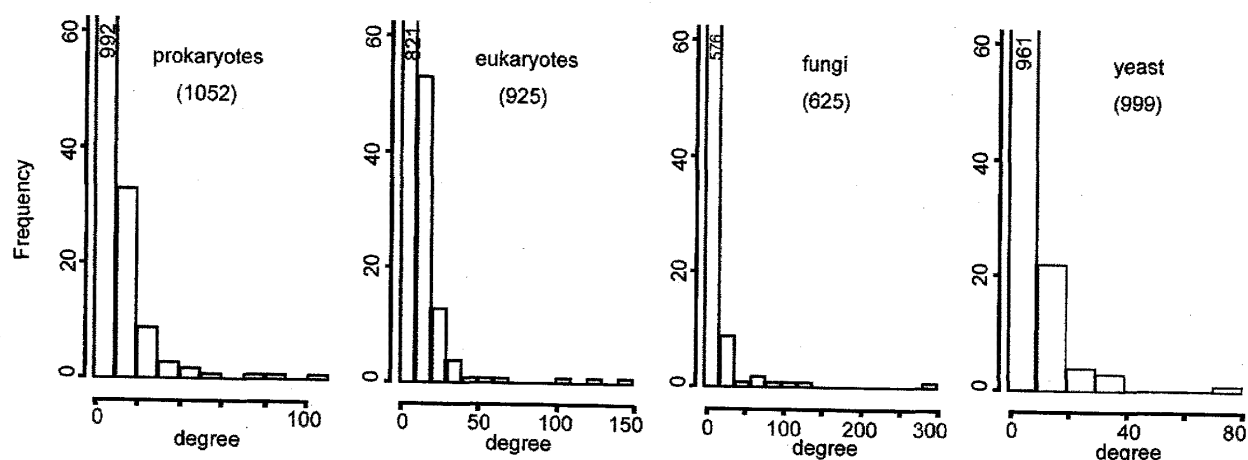


図 4A. 各進化的出現時期における次数分布。() 内にある数字は各々の出現時期に表れたタンパク質数を表す

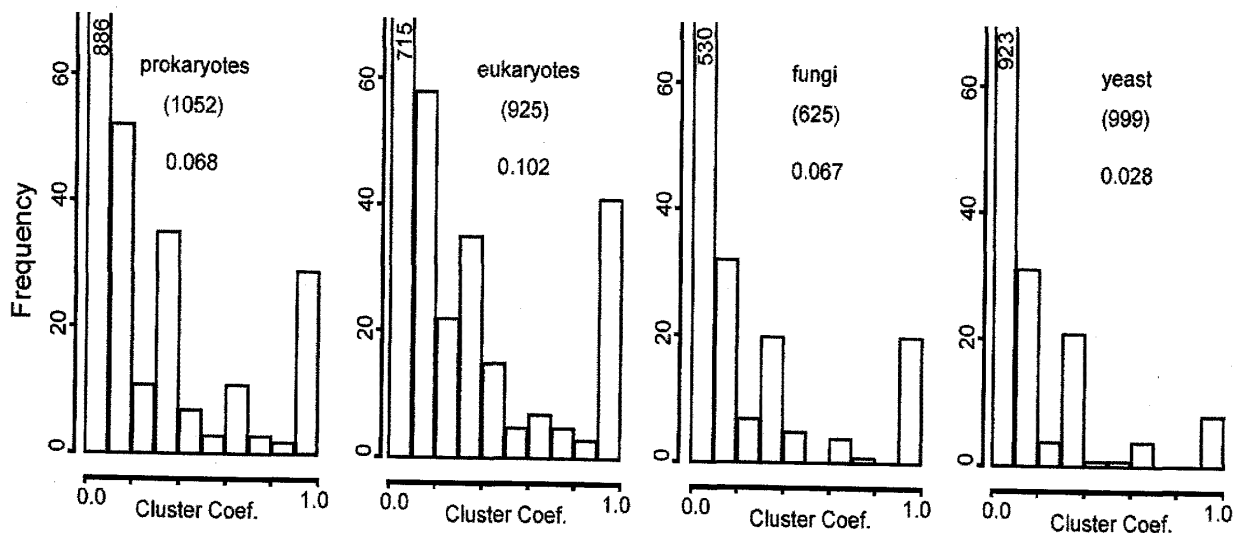


図 4B. 各進化的出現時期におけるクラスター係数分布。() 内にある数字は各々の出現時期に表れたタンパク質数を表す。また、中央にある数字は平均クラスター係数を表す

まとめ

二つの仮説を検証することにより、ネットワークとして重要な役割を持っているタンパク質の進化速度は遅く、さらにモジュラー性がシステムの拘束に関与していることが明らかになった。今後、様々な生物種の PIN が利用可能となることが期待されており、それらを比較することによりシステムの拘束性についてより詳細な解析が可能になると考えられる。

最後に システム進化生物学の確立に向けて

システム進化生物学の理論的な構築はまだ開始されたばかりである。今後解決しなければならない課題は多い。

分子レベルでの進化を、遺伝子単位に究明する従来の分子進化学を超えて、分子ネットワークの複雑化の過程としてみる「システム進化生物学」に関して、原理的な側面を明確化してゲノム科学・システム生物学時代に相応しい進化理論の構築を目指すために、必要な今後の研究の方向を述べてみよう。

① タンパク質間相互作用ネットワークによるシステム進化の法則の確立

——生物間の数理グラフ論的比較とシミュレーション

システム進化理論の法則を確立するために、タンパク質間相互作用ネットワークの結線構造に着目して、ネットワークの構造の諸特徴が、それに含まれる遺伝子の進化速度や遺伝子の重複頻度などをどう規定するか、諸生物の相互作用ネットワーク間の進化的比較やネットワーク進化シミュレーションを用いることによって、分子ネットワーク進化に関する諸法則を確立することが必要である。とくに生命ネットワークの特徴であるスケールフリー性を持ちつつ局所的に密な結合を示す性質の生成過程を明らかにする意義は大きい。

② 形態形成/発生調節系のシステム進化的研究

——マクロ形態学的進化論と分子進化学の橋渡しと新しい進化理論の確立

古生物学における形態的マクロ進化学と分子的進化の間隙を埋めるものとして、形態形成/発生を制御する遺伝子調節ネットワークの複雑化がボディプランの複雑化などを通してどのように生命体制の段階的進化をもたらしたのか、発生調節遺伝子族のネットワークの進化過程を通して追跡する必要がある。

③ 生命の大域的体制のシステム進化的研究

——生命の大域的体制の「入れ子」的進化と分子ネットワーク

さらに、より大域的な区別である原核生物、真核生物、多細胞生物の体制的進化についても、その「入れ子」的に進化原理を明確化し、対応する分子ネットワーク的な進化的展開を解明して、生命分子ネットワークにおける非連続的な大域的な進化に関するネットワーク進化理論の確立が待たれる。

これらの課題を通して、「システム進化生物学」すなわち分子ネットワーク進化理論の完成が待たれる。それは生命系のグランドセオリーとしての役割を果たすものと期待される。

参考文献

- [1] Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C. and Feldman, M.W. (2002), "Evolutionary Rate in the Protein Interaction Network," *Science*, Vol.296, pp.750-752
- [2] Hahn, M.W. and Kern, A.D. (2004), "Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks," *Mol. Biol. Evol.*, Vol.22, pp.803-806
- [3] Jordan, I.K., Wolf, Y.I. and Koonin, E.V. (2003), "No simple dependence between protein

evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly," *BMC Evol. Biol.*, Vol.3, pp.5

[4] Makino, T., Suzuki, Y. and Gojobori, T. (2006), "Differential evolutionary rates of duplicated genes in protein interaction network," *Gene*, Vol.385, pp.57-63

[5] Huang, L.L.Y., Xia, X., and Sun, Z. (2006), "Preferential duplication in the sparse part of yeast protein interaction network," *Mol. Biol. Evol.*, Vol.23, pp.2467-2473