

コンジェリンのフォールド進化から ヴァーチャルフォールドシミュレーションへ

白井 剛

名古屋大学大学院工学研究科生物機能工学

E-mail: i45282a@nucc.cc.nagoya-u.ac.jp

はじめに

今回紹介させていただく話題は2つの部分から成る。前半部では、魚類レクチンであるコンジェリン I、II のX線結晶構造解析を紹介する。コンジェリン I と II は遺伝子重複で生じたアイソフォームであるが、構造解析の結果、両者の構造は一部異なっていることが分かった。この差をタンパク質のフォールド進化という観点から解説する。後半部は、コンジェリンの構造解析が発端となったバイオインフォマティクス研究の話題である。これは、計算機を使ってタンパク質のフォールド空間を全探索することを目的としたシミュレーションである。最終的な結果はまだ得られていない。今回は、シミュレーションの発想・実装の現状とこれまで得られた限定的な結果について解説し、ご批判を仰ぎたいと考えている。

コンジェリンの構造解析 -フォールド進化の発見-

ウナギなどのぬるぬるした体表粘液は外界からの物理的な衝撃に対する防御手段となるが、粘液中には様々なタンパク質因子が分泌されており、生物化学的防御を担う組織でもある。そのような体表粘液で先天性免疫を担う因子の一つとして、糖鎖結合タンパク質であるレクチンが挙げられる。マアナゴ体表粘液からはガレクチン（ガラクトシド結合特異性レクチン）ファミリーに属するコンジェリン I およびコンジェリン II の2種のレクチンがみついている。ガレクチンファミリーは動物界に広く分布し、その発現している組織によって発生・分化・形態形成・免疫など多様な生命現象に関与する多機能レクチンである。

コンジェリン I と II は、それぞれ特異的にホモ2量体を形成し、2価性糖鎖結合機能を発揮する。例えばコンジェリンは、海洋性バクテリアの凝集、寄生生物の胚発生阻害などの活性を示す。これがコンジェリンの機能が生体防御で

あると推定される根拠である。

コンジェリン I と II はアミノ酸配列で 50% の類似性を示すが、耐熱性・pH 依存性・ラクトース誘導体に対する糖鎖結合特異性などでは差が見られる。特に耐熱性は、コンジェリン I が 50 で 80 分間以上も安定であるのに対して、コンジェリン II は同じ条件で完全失活し、大きな違いを見せる。

このコンジェリンの遺伝子解析より分子進化的に面白いことが明らかになった。両コンジェリンの塩基配列を比較すると、翻訳領域の配列一致度(73%)の方が 5' および 3' 末端非翻訳領域の一致度(90%)より低く、また、翻訳領域でアミノ酸置換を伴う非同義座位での塩基置換が頻繁に認められた。非同義置換の塩基の相違度(KA)と同義置換の塩基の相違度(KS)の比 KA/KS 値を求めると 2.7 という異常に大きな値をとる(通常の遺伝子では 0.2 前後である)。これは、コンジェリンが加速進化していることを示しているが、通常このような進化は遺伝子が強い選択圧下に置かれた場合におけると考えられている。

このコンジェリンの立体構造を解明するために X 線結晶構造解析を行った。結果として、コンジェリン I の構造は 1.5 と比較的高分解能で決定できたが、解明された構造にはひとつ意外な点があった。それは、2 つのサブユニット間でストランドの交換(ストランド=スワップと呼ぶ)が起っていたことである。

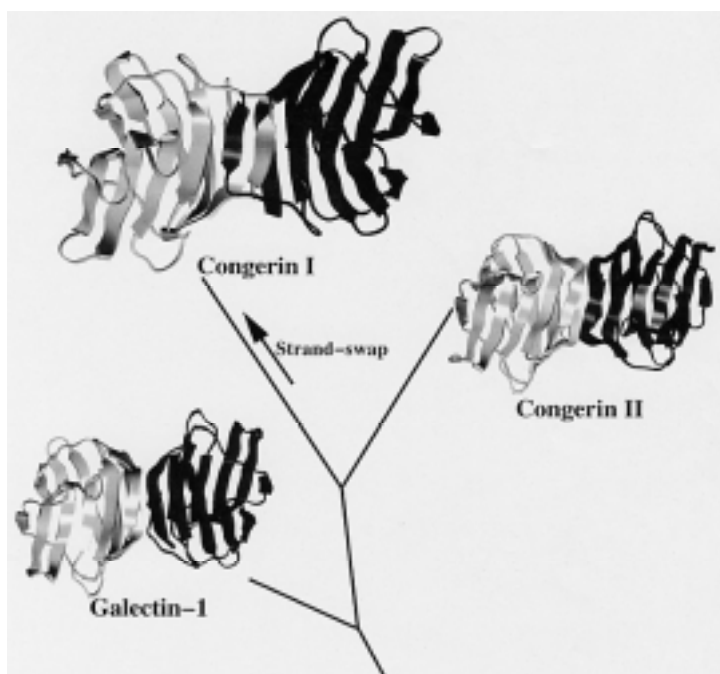


図 1 コンジェリンの構造と分子系統

図 1 はコンジェリン I と類縁タンパク質の立体構造を示したもので、2 量体のサブユニットを濃淡で分けて示している。これらのタンパク質はすべて、サブユニットにまたがるシート 2 枚を重ね合わせた形の分子骨格をもっている。コンジェリン I の前面中央部に位置する 2 本のストランドは、互いに他のサブユニットにはまり込むようにシートを作っているが、ウシ由来のガレクチン-1などはそ

のような構造はとっていない。ガレクチン-1では2枚のシートはすべて逆平行シートであるが、ストランド=スワップの結果、コンジェリンIでは部分的に平行シート構造に変化している。すなわち進化の過程でフォールドに変化が起こった事になる。

ストランド=スワップが進化の途上のどの時期に起きたかは、コンジェリンIにもっとも近縁であるコンジェリンIIの構造解析を続けて行うことで判明した(1.45 分解能)。図1の様にコンジェリンIIにはストランド=スワップは起こっていなかった。アミノ酸配列から推定した分子系統によると、この構造変化は両コンジェリン遺伝子の重複後に起こったことがわかる。

糖鎖結合部位はサブユニットにそれぞれ一カ所しか存在しないので、糖鎖のクロスリンクによる細胞凝集活性に2量体構造は欠かせない。これは抗体の抗原凝集活性に2量体構造が必要であるのと同じである。このことから、ストランド=スワップの意味は2量体構造の強化ではないかと考えられる。ストランド=スワップによりサブユニット間の接触面積や水素結合の数は倍以上に増加する。また、ペプチド鎖が入り組んだ構造になるので、2つのサブユニットを引き離しにくくなる。事実、上に述べたように細胞凝集活性の耐熱性はコンジェリンIの方がIIよりかなり高いことが分かっている。コンジェリンは体表粘液に分泌されるので、体外の環境にさらされて機能しなければならない。これがコンジェリンIの強化型の2量体構造が有利となり、自然選択された原因と考えられる。

タンパク質進化の点から特筆すべきは、選択圧のかかっている遺伝子の産物で立体構造の変化が認められたということである。微少な変化ではあるが、自然選択により"タンパク質の新規フォールド"が出現したことが示されたといえる。コンジェリンIはフォールド自体もまれに進化するという実例である。

ヴァーチャルフォールドシミュレーション -コンピュータでフォールド空間探索-

1980年代、構造解析されたタンパク質の数は飛躍的に増加した。それとともに、配列的にも機能的にも一見無関係であるのに立体構造の類似したタンパク質が多数発見された。フォールドレベルのタンパク質構造の種類は限られているのではないかと推測はこのあたりに端を発する。やがて、C. Chothiaが「タンパク質フォールドの総数は高々1,000個程度である」との説を唱えて衝

撃を与えることになる。これは、その時点で解析済みのアミノ酸配列と立体構造からの統計的な推定である。その後もいくつかの推定値が出された。それらの値は数千～約一万の間でばらつくものの、フォールド総数は遺伝子総数と比べてかなり少ないとの説は定着した。近年、タンパク質立体構造の網羅的解析を目的とした構造ゲノミクス研究が多数計画されているが、上記の説は重要な精神的背景となっているように思われる。

今回紹介するヴァーチャルフォールドシミュレーションの発想は、これまでと別のアプローチからタンパク質のフォールド空間を評価できないかというものである。構想は以下のような単純なものである。

- (1) コンピュータ上でタンパク質フォールドをランダム発生させる。
- (2) 「タンパク質らしさを備えている」と判断されるフォールドは保存し、残りは捨てる。
- (3) 新規フォールドの出現率が十分さがるまで 1-2 を繰り返し、収集されたフォールド(これをヴァーチャルフォールドと呼ぶ)を分類する。

この方法が従来の評価法と異なる点は、解析済みアミノ酸配列の分類による推定ではなく、物理的に実現可能と予想されるフォールドの総数はどのくらいか、という観点から評価する点である。発想は単純であるが、このシミュレーションの実装はそれほど容易ではない。以下図 2 のシミュレーション概略図に沿って解説する。

フォールドの省略表現

タンパク質は複雑な分子で、立体構造に含まれる情報量は非常に多い。このシミュレーションでタンパク質の原子座標そのままを扱うことは、計算速度・データ量の点から困難であるので、フォールドを省略化して扱うことにした。すなわち、ヘリックスまたはストランドをそれぞれ一本のベクトルとして表現し、2次構造以外の領域は無視する(ちなみに現時点では、それらベクトルの1次構造上の順番も無視している)。この省略化によって1個のフォールドの生成は、数個のベクトルの集合を生成する作業に単純化される。

タンパク質らしさの判断基準

「タンパク質らしさ」を評価することがこのシミュレーションの最難関の課題となる。とりあえず、既知の天然フォールドから引き出される経験則を判断

基準として用いることにした。基準は以下のものである

(1) ベクトルは衝突してはならない。

立体障害をさけるため、ベクトル間最短距離をモニターし、衝突距離にあるものは無条件で排除する。

(2) ベクトルの相対配置は物理化学的に好まれるものでなければならない。

個々の2次構造要素間の相対配置は、天然フォールドに認められるものを重視する。フォールド生成時に新たなベクトルを導入する場合は、天然フォールドから取り出したベクトル対を用い、一方をヴァーチャルフォールドに重ね合わせ、他方を取り込むことにする。

(3) ベクトルの密度はある程度高くなければならない。

衝突を避けるあまり隙間だらけのフォールドを作らないように、重心を中心とする球の半径と、球に含まれるベクトル数の分布が天然フォールドの統計に準拠するようモニターする。

(4) ベクトル和はある程度小さくなくてはならない。

フォールドは1本のペプチド鎖が折り畳まれた物であることを暗に取り込むため、天然フォールドのベクトル和分布に従うようモニターする。

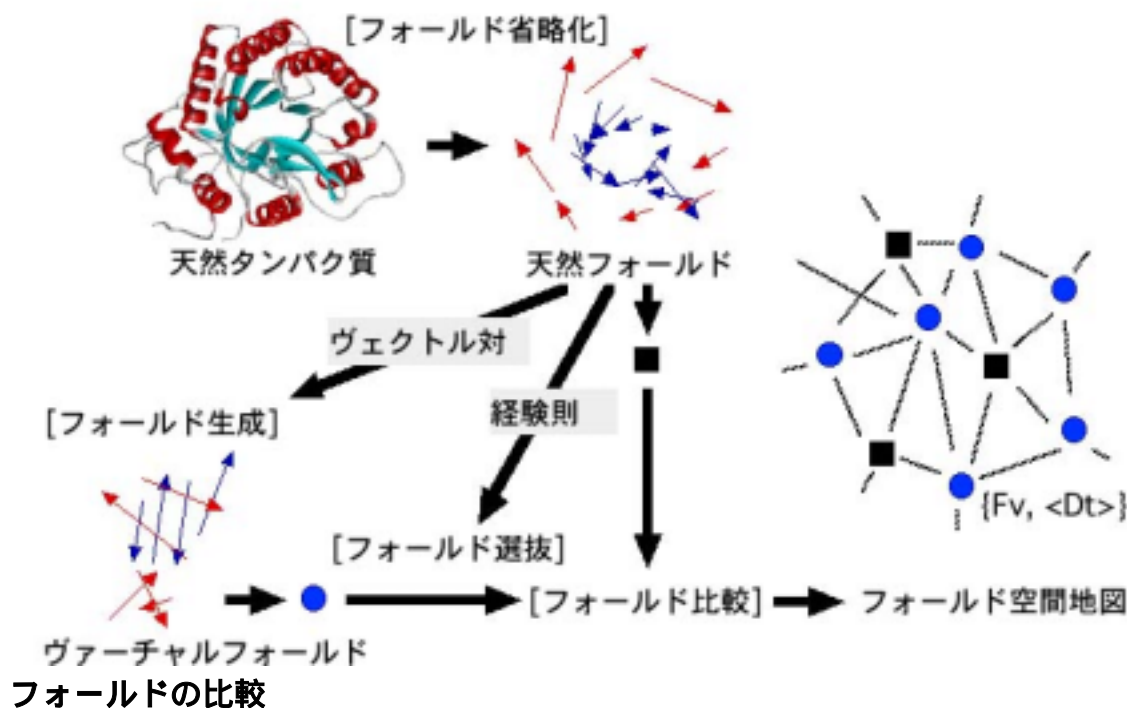


図 2 ヴァーチャルフォールドシミュレーション概要

ベクトル配置の比較は、対応するベクトルの中間距離(d ())とベクトルのなす角((rad))の自乗和 $Dt=d^2+58 \text{ }^2$ で行う(58は重み定数)。例えば N_a 本と N_b 本のベクトルを持つ2つのフォールドを比較したとき、両者の類似性は、重ね合わせ可能なベクトルの本数(N_c)の N_a または N_b に対する割合 $F_v = N_c/\max(N_a, N_b)$ と Dt の平均 $\langle Dt \rangle$ で表される。

ここでも、天然フォールドを用いて比較のための判断基準を求めた。SCOPデータベースで、TIMバレル、グロビン、イムノグロブリン、サイトカイン、インターロイキン、SH3に分類されているタンパク質をベクトル表現し、すべての組み合わせで比較した場合に、およそ9割の組み合わせを拾い上げられる基準として、 $F_v > 2/3$ と $\langle Dt \rangle < 20$ を暫定的に採用した。

シミュレーション初速度からの推定

このシミュレーションを実行し、フォールド空間の全探査を行うには相当の時間がかかる。しかし、シミュレーションの初期段階でも、フォールド総数についての推定値を得ることはできる。ある時点までに得られた重複のないヴァーチャルフォールドの個数を v 、そのうち天然フォールドに類似したものの数を n 、既知の天然フォールドの数を N とすると、フォールド総数の推定値は $N(v/n)$ である。

上記のシミュレーションを実装し、500個のヴァーチャルフォールドを生成してみた。ヴァーチャルフォールド内で比較して重複を除いた個数は $v = 317$ であった。比較対象となる天然フォールドは、SCOPデータベースでフォールドレベルの分類群から1個ずつ代表を選抜した($N = 513$)。天然フォールドとの比較を行った結果、天然に類似したヴァーチャルフォールドは、上記 $\{F_v > 2/3, \langle Dt \rangle < 20\}$ の基準をそのまま適用すると51個、グラフィックスによる観察でさらに厳しく選抜すると14個であった。よってフォールド総数の推定値は、楽観的には約3,000個、厳しく評価すると約12,000個である。このままのパラメータでシミュレーションを続行すると、全フォールド空間の99%をカバーするのに必要な時間は2~8ヶ月と予想される。

展望

実際にはテストの結果から、いくつも改良を要する点が浮かび上がったため、今後もテストを繰り返しシミュレーション精度を向上させる必要がある。し

かし、テストの結果得られたフォールド総数の推定値が、意外なほど従来予想されていた値に近かったことは、一つの発見であった。

このシミュレーションから将来的に成果として期待でき、かつ他の方法では得られない情報が2つある。

(1)全フォールド空間のマップとモデル。

これまでフォールド総数についての予想はあるが、構造が未解明のフォールドについてモデルを作成することはされていない。よって全フォールド空間について構造類似性を基準にしたマップも作られていない。このシミュレーションは、省略化されたものにせよモデルを生産するので、それらの情報を提供できる。

(2)生物が採用してないフォールド。

生物は進化の過程で様々な遺伝子を発明してきたが、すべての可能性を試せたとはどうも考えられない。今後構造ゲノミクス研究は、生物が採用し保持しているフォールドの全貌を解明していくと期待されるが、天然資源として現存しないものは未知のまま残る。テストの結果は、その未知領域がそれほど大きくないことを示しているが、このシミュレーションにより、どのような未知フォールドが可能か予測することができる。

このシミュレーションはタンパク質構造科学の様々な局面に関連するため、説得力のある結果を示すのは容易でないが、なるべく多くの方の意見を参考に向上させてゆきたいと考えている。

コンジェリンの構造解析は東北大学農学研究科・村本光二、小川智久両博士との共同研究であり、両博士に感謝します。以下は、コンジェリン構造解析に関する参考文献です。

T. Ogawa, C. Ishii, D. Kagawa, K. Muramoto & H. Kamiya, "Accelerated evolution in the protein-coding region of galectin cDNAs, congerin I and congerin II, from skin mucus of conger eel (*Conger myriaster*)." *Biosci. Biotechnol. Biochem.* **63**, 1203-1208 (1999).

T. Shirai, C. Mitsuyama, Y. Niwa, Y. Matsui, H. Hotta, T. Yamane, H. Kamiya, C. Ishii, T. Ogawa & K. Muramoto, "High-resolution structure of the conger eel galectin, congerin I, in lactose-liganded and ligand-free forms: emergence of a new structure class by accelerated evolution." *Structure* **7**, 1223-1233 (1999).