

タンパク質のモジュール構造と分子進化

郷 通子

最近、X線構造解析や二次元 NMR 法などを用いて多くのタンパク質の立体構造が次々と明らかにされている。その結果、一見多様に見えるタンパク質の多くはモジュールやドメインと呼ばれる基本的な構造ユニットの組合せから構成されていることが判明した。モジュールとは α ヘリックス、 β シートなどの2次構造とは独立に、立体構造が知られたタンパク質の原子座標を基に導かれた10～40前後のアミノ酸残基からなるコンパクトな構造ユニットである。また、細胞接着分子、血液凝固因子群、各種のレセプターなどでは2種以上の配列のリピートが多数回見られ、これらのリピートの単位は単独でも折り畳まれ機能を持つことが多い。これらのリピート単位をドメインと呼ぶ。

タンパク質は DNA 上に記された塩基配列の情報を基にして合成される。ところが、タンパク質をコードしている遺伝子のすべての塩基配列が情報を担っている訳ではない。真核生物の構造遺伝子には全くタンパク質の合成には関与しない部位“イントロン”が存在することが1977年に発見された。現在、イントロンは多くの真核生物で確認されているが原核生物についてはバクテリオファージを除いて発見されていない。なぜ、真核生物はこの一見無駄な塩基配列を持っているのだろうか？今のところ「エクソン(DNA上でタンパク質の情報を担った部位)の混成(シャッフリング)の介添え役として」という説が有力である。エクソンがタンパク質の機能ユニットをコードしているとすれば、エクソンの混成は新しい機能や複雑な機能を生むだろう。シャッフリングは少ない基本ユニットの組合せによって多様性を生み出すための経済性に富んだ方法である。しかしながら、もしDNA上のすべての塩基配列がタンパク質をコードしているとすれば、シャッフリングは容易には起こり得ない。なぜなら情報を担った部位で不等交叉が起きたり、他のエクソンが入り込んだりすることは、その部位がコードしているタンパク質の情報を破壊してしまい、ほとんどの場合進化的に有利に働かないからである。シャッフリングはDNA上にイントロンが存在するからこそ可能なのである。

タンパク質の機能はその立体構造に依存することが多い。もし、そうならばDNA上のエクソンがタンパク質の立体構造の基本単位であるモジュールに、イントロンがモジュール間の境界にそれぞれ対応していると予想される。そして、実際にイントロンがモジュール間の境界に対応している例が多くの場合に発見されている。ヘモグロ빈はサブユニット α 鎖と β 鎖が2本ずつ4つのヘムと共に会合した4量体である。ヘモグロ빈の立体構造はX線結晶解析によって決定されている。 α 炭素間の距離を等高線状に表した距離地図を用いて解

析した結果、 α 鎖と β 鎖は4つのモジュール(F1-F4)に分かれた。カエル、チキン、マウスなどは α 鎖、 β 鎖の遺伝子に各々2個のイントロンを持っているが、それらはモジュールF1とF2の境界、F3とF4の境界に対応していることが判明した。モジュールF2とF3との境界にもかつてイントロンが存在したとすればエクソンとモジュールとの1対1の対応がつく。実際、ダイズが持つグロビン遺伝子には3つのイントロンが存在し、中央のイントロンはまさにF2とF3の境界に対応している。このことは、グロビン遺伝子が動物と植物の分岐以前の10億年以上前から存在し、その当時にはエクソンとモジュールは完璧に1対1に対応していたことを物語っている。

生物はイントロンを利用して巧みにエクソンを混成しタンパク質を進化させてきた。しかしながら、イントロンは進化的な時間スケールで見ると失われていく傾向にある。イントロンがほとんど発見されない原核生物はその究極の形といえるだろう。一旦、複雑な機能を持つタンパク質をコードする遺伝子が出来てしまえば、無用に長い塩基配列は効率の良い増殖を行なえないという点で進化的に不利になる。モジュール混成による進化はタンパク質の初期進化の「創造の時代」に活発にみられたが、安定なドメインが完成した後にはドメインの重複と混成による「模倣の時代」になったと考えられる。