

「配列空間の“島モデル”再考」

国立遺伝学研究所 西川 建

天然タンパク質のアミノ酸配列と立体構造の関係を概念図として表したのが「島モデル」である [1]。島モデルでは次のことが表現されている。

- A 1. タンパク質の配列空間は任意のアミノ酸配列を表現しうる超多次元空間であり、ある一定の配列をもつタンパク質は配列空間の1点にプロットされる。このとき類似した配列は近くに、そうでない配列は遠くに位置づけられる。
- A 2. 相同なアミノ酸配列をもつ天然タンパク質は族（ファミリー）を構成し、配列空間ではクラスター（つまり「島」）をなす。
- A 3. 同じファミリーに属するタンパク質は同一の立体構造（フォールド）をもち、ファミリーが違えば立体構造は異なる。したがって、1つの島には1つの立体構造が対応する。
- A 4. 配列空間において、島どうしは「海」によって隔てられている。海の領域には立体構造は対応しない。すなわち、海の領域にプロットされる配列はいわゆる「フォールドできない配列」に当たる。

このように島モデルは、タンパク質のアミノ酸配列（クラスターをなす天然配列）と立体構造（構造の異同、フォールドするか否か）の関係を全体的に捉えたタンパク質一般の描像であるといえるが、このような描像はすでによく知られている次の事実と符合する。

- B 1. 立体構造とアミノ酸配列は1対1の対応関係ではなく1対多の関係にあり、配列が少々変わっても立体構造（骨格構造）は変化しない。
- B 2. 変異タンパク質において観測されるように、変異導入の結果変化するのは立体構造ではなく、構造の安定性およびタンパク質の活性などである。島モデルの言葉で云えば配列の変異は島の上での移動に対応し、安定性や活性は場所によって異なる。また安定性が低下しすぎるともはやフォールドできなくなる（海に落ちる）ことを意味する。
- B 3. 配列をデザインして造られた人工タンパク質は容易にフォールドしてくれない（結晶化しない）。このことは島のサイズに比して海の広大さを示唆する。
- B 4. タンパク質の分類においては、ファミリーを超えるような分類枠はない。言い換えると、異なるファミリー間の系統関係は一般に不明である。島モデルでは、島どうしが不連続的に分布し、たがいに海で隔てられていることに対応する。
- B 5. 分子進化の過程で1つの天然配列が変異して、他の島に飛び移ることはほとんど起らない。つまり、それぞれの天然配列は事実上1つの島に閉じ込められる。原核生物と真核生物のように遠く離れた種の間でも多数の同族タンパク質が見つかる理由は、このような「閉じ込め効果」によってよく説明できる。

以上のように、島モデルはタンパク質の配列と構造の関係を直観的に捉えたナイーブな描像にすぎないが、タンパク質の基本的な諸特徴をよく「説明」しうるモデルになっていると考える。このようなモデルを最初に提起してからすでに10年が経過する [2] が、この間のタンパク質研究の進展は島モデルを支持し豊富化しさえすれば、島モデルを否定し、あるいは修正を迫るようなものではなかったといってよい。

たとえば、1992年にタンパク質の基本フォールドは1000個ほどにすぎないとした有名な Chothia の論文が発表された [3]。Chothia のいう基本フォールドは広い意味のタンパク質ファミリーに対応し、それはまさしく島モデルでいう島に相当する。ファミリー（島）が不連続であるからこそ、その個数が数えられるのであり、連続分布であれば「個数」は意味をなさない。当時、私は島の総数が推計できるとは思ってもよらず、データベースとゲノム解析のデータからファミリー数を算定してみせた Chothia の腕前にいたく感動した思

い出がある。また、1994 年頃になると格子模型を使ったフォールディング・シミュレーションが盛んになって、フォールドする配列としない配列の違いはどこにあるか、ということが初めて問題にされるようになった [4]。これにより foldable, unfoldable という言葉が生れ、フォールドしない配列という表現がやっと市民権を得た（それ以前は「海」の領域の意味が伝わらずにとまどう事が多かった。「フォールドしない配列」など考えたこともないという人が如何に多いかが痛感された）。一方、この 10 年間で多様化した人工タンパク質への挑戦も同じく unfoldable 配列と天然配列の違いを際立たせる効果をもたらしている。しかし、私の見るところ「海」の広大さを自覚している研究者はまだまだ少ないようだ。この点は後で再び触れることにしたい。

一方、島モデルには大きな欠点がある。それは、天然タンパク質を実際に配列空間にプロットすることが出来ないことだ。その意味で配列空間は「イマジナリー（仮想的）な空間」である。研究会の場では「なぜ仮想的なのか？」と質問されて適切に即答できなかった。ので、あらためて以下に説明しておきたい。

なぜ実際の天然配列がプロットできないかと言えば、配列どうしを比較し、類似したものは空間内の近い点に、違いが大きいものは遠くにプロットされる必要があるが、そのような適切な比較のためには配列どうしのアラインメントが必須である。ところが、異なるファミリーに属するような配列どうしは上述（項目 B 4）のように系統関係が不明であり、アラインメントそのものが不能になってしまう。したがって、それらに共通の「座標系」を定義することはできない。もう少し丁寧にいえば、天然タンパク質の島（ファミリー）ごとに座標系を定義し個々の配列をプロットすることは可能であろうが、多くの島々を同じ空間内で同時に俯瞰するようなプロットは原理的に不可能である。島モデルの真髄はもちろん後者にこそあるので、それが現実には不可能だとすると島モデルはあくまでも仮想的な概念図にとどまり、具体的な「データ解析」にはむすびつかないという限界性がある。

その点で、配列空間はアミノ酸組成空間とは対照的である。組成空間は 20 次元空間として定義され、任意のタンパク質をそのアミノ酸組成に応じて実際にプロットして、互いの位置関係を目で見ることが出来る [5]。しかし、配列に比べてアミノ酸組成は大幅に縮退した情報であり、その上配列の一面だけを反映したものである。組成空間をもって配列空間を代用させるわけにはいかない。現に、組成空間において天然タンパク質は連続的に分布し、島状分布にはならない。

最後に、タンパク質の配列空間が島と海からなるという点をあらためて考えてみたい。一般に配列空間は核酸（RNA）やペプチドに対しても定義できる（歴史的には核酸の配列空間が最初に議論された）が、RNA やペプチドの配列空間では「島と海」という区別はなく、言うなればのっぺらぼうの空間である。当然ながらこの違いは、タンパク質だけが相転移（協同）的なフォールディング過程を経て特異的かつ強固な立体構造を形成するからにはかならない。生理的条件下でフォールドする配列とフォールドしない配列が峻別されていて、その中間の「部分的にフォールドする」ような配列はない。それが協同的ということだ。

立体構造に関しては通常いわれる「特異的な形」よりも、ここでは「かたさ」の方を強調しておきたい。「強固な構造」というと、ふつうは安定な構造を意味するが、ここではむしろ「ふにゃふにゃ」した柔軟な構造ではないことに力点を置きたい。ひも状高分子の中でタンパク質ほど強固な構造をつくれるものが他にあるだろうか。分子レベルのミクロの世界では熱ゆらぎによる激しい擾乱のため、巨大分子が一定の形を保つのは至難のはずである。もちろん、各種の低分子結晶などは硬くて安定だろうが、遺伝コードによって指定される化合物としては 1 次元のひも状分子でなければ具合が悪い。ところが、ひも状分子だと一般にふにゃふにゃして（人工高分子を見よ）、一定の形にはなりにくい。このような相矛盾する要請を同時に満たすものとして登場したのが、フォールディングによって強固な立体構造をつくりうるタンパク質だったのではないか。

分子レベルのミクロなスケールにおける強固な構造体は貴重である。おそらく、「形のある生命体」はタンパク質なくしてはあり得なかったのではないか。事実、リボソームやスプライスゾームなど、RNAとタンパク質からなる機能性複合体をみると、活性はむしろRNA側にあり（RNAワールドの名残り）、タンパク質は全体をささえる支持体の役割をしていると言われる。あるいは、ファージやウイルスの殻、尾部、脚などの構造体はすべてタンパク質製であり、さらに細胞一般では脂質二重膜からなる柔らかい細胞膜を裏打ちし、繊維状の骨格で支えて細胞に一定の形を与えているのはすべてタンパク質である。無常の原始スープの中で、タンパク質はさまざまな意味での足場となる貴重な存在だったのではないか。足場のまわりにより大きな構造体ができ、しだいに形のある生命体が出現するようになったと考えられる。

そうだとすると、タンパク質の進化においては、「フォールドするか否か」という点こそが第1に重要な自然選択の基準になったはずである。ポリペプチドのランダム変異によってフォールドする配列が出現するのは難しかったかもしれないが、なにせ時間だけはたっぷりあった。試行錯誤のすえに出現したフォールドできる配列はその他のものに比べて圧倒的に有用な価値をもち、それゆえにいったん出来た後は大事に保持されたにちがいない。その上、堅くて複雑な形をなす分子表面はリガンド結合や触媒活性のための格好の場を提供する。タンパク質は構造のみならず機能面においても有用性を発揮するようになって、多様な発展をとげるようになったに違いない。天然タンパク質が「島」にしか存在しない事実の背後には、以上のような進化的なストーリーが推論される。

参考文献

1. K. Nishikawa, "Island hypothesis: Proein distribution in the sequence space", *Viva Origino*, 21, 91-102 (1993)
2. 西川建「タンパク質のユニークさと生命の起源」生物物理、29, 320-323 (1989)、
「タンパク質の描像：島モデルとその生物学的意味」タンパク質の最前線（赤坂一之編、さんえい出版）pp. 5-23 (1991)
3. C. Chothia, "Proteins: one thousand families for the molecular biologist", *Nature*, 357, 543-544 (1992)
4. A. Sali, E. Shakhnovich, M. Karplus, "How does a protein fold?", *Nature*, 369, 248-251 (1994) など。
5. K. Nishikawa, Y. Kubota, T. Ooi, "Classification of proteins into groups based on amino acid composition", *J. Biochem.*, 94, 997-1007 (1983), H. Nakashima, K. Nishikawa, T. Ooi, "The folding type of a protein is relevant to the amino acid composition", *J. Biochem.*, 99, 153-162 (1986), H. Nakashima, K. Nishikawa, "Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies", *J. Mol. Biol.*, 238, 54-61 (1994) など。