

## アミノ酸配列データの解析

—生物の中での物理学と情報学の結合—

美宅成樹

東京農工大学工学部

DNAの配列とタンパク質のアミノ酸配列には重大な違いがある。配列を変えたときに、つまり情報の内容を変えたときに、DNAでは高分子自体の性質が変わらないのに対して、アミノ酸配列が変わると、タンパク質の立体構造、機能、その他の性質が大きく変わるのである。したがって、タンパク質のアミノ酸配列の解析では、高分子の立体構造を作る物理化学的過程についての理解が重要なのである。それに関連して、極性相互作用と疎水性相互作用が対照的な性質を持っていて、立体構造の形成過程で、それらが分業しているということを考慮する必要がある。ここでは、アミノ酸配列の解析に対する次のような物理化学的アプローチを紹介する。(1) 分子間力の性質、(2) 生体高分子における分子間力の分業、(3) 膜タンパク質の立体構造の予測。

## Analysis of Amino Acid Sequence.

Physics and Informatics in Biological Systems

Shigeki Mitaku

Tokyo University of Agriculture and Technology, Faculty of Technology

Nakacho 2-24-16, Koganei, Tokyo 184, Japan

There is essential difference between DNA sequence and amino acid sequence in physical point of view. When the sequence is changed, the physical properties as well as the structure of DNA polymer do not change, whereas those of polypeptides (proteins) vary according to the change in the sequence. Therefore, the understanding of physico-chemical processes in polymer solutions is quite important for the theoretical analysis of amino acid sequences. Particularly, the different role of polar interactions and hydrophobic interaction has to be taken into account. Here, our approach to this problem is surveyed under the following sections: (1) the properties of intermolecular forces, (2) different role of intermolecular forces in biopolymer solutions and (3) molecular structure prediction of membrane proteins.

## 1. はじめに

生物を構成する分子の中で最も重要なものにDNAとタンパク質がある。DNAは4種類、タンパク質では20種類の文字で構成されていて、情報の中身としては変換可能であるため、その違いについて余り議論されることが少ない。しかし、DNAとタンパク質の配列の間には本質的な違いがある。すなわち、DNAでは塩基対が単位となって情報の担体となっているが、塩基対はA-TでもG-Cでもその長さが同じであって、文字を書き換えても情報担体としての高分子の性質に違いがほとんどないのである。情報自体とそれを書き込んでいる物質の性質を切り離して考えることができるわけである。これは、情報の担体としては優れた性質（というよりむしろ当然の性質）となっていて、そのために情報の保存・複製などを行うためにはただひとつの装置を用意すればすむのである。これに対して、タンパク質の場合はそこに情報が書き込まれているということは同じであるが、文字を書き換えるとタンパク質の構造、機能など高分子の様々な性質が大きく変化してしまう。それは、タンパク質が情報処理のための分子ではなく、情報を用いてできた分子機械であって構造が機能に合わせて変化しなければならないという事情による。つまりタンパク質では情報の変化によって、構造が変わることは合目的なのである。DNAの性質が中立な突然変異を保証しており、タンパク質の性質は配列のランダム化を阻止しているわけである。そのバランスでパターンマッチングなどの現象が起こってくる。このように、アミノ酸配列と立体構造との間の物理化学的な関係を明かにすることは、ゲノム知識情報処理分野における越えなければならない最大の課題のひとつである。

## 2. 分子間力の性質

タンパク質の立体構造は、その分子状のひもに働く分子間力のバランスによってできている。分子間力のバランスでできているという意味では、低分子の結晶も基本的には同じであるが、高分子の場合は色々な種類の分子間力が同時に各部の間に働く点で単純な結晶などとは異なっている。そのような極めて複雑なタンパク質の立体構造形成の問題を完全に理解するには、まず分子間力の性質をよく知ることが必要である。分子状のひも自体を作る化学結合を別とすると、タンパク質の立体構造形成に関与する分子間力には疎水性相互作用、性電相互作用、水素結合、ファンデルワールス力などがある。簡単にそれぞれの性質をまとめると、以下のとおりである。

(1)疎水性相互作用： 非極性基のまわりの水の構造は秩序を持っていることが知られており、それが原因となって凝集現象が起こる。これを疎水性効果あるいは疎水性相互作用と呼んでいる。ある意味で見かけの相互作用であって、水の中では強いが、水がなければ働かない。

(2)静電相互作用： クーロン力で、電荷同士の相互作用である。水の誘電率が非常に大きいので、水の中では弱い。この相互作用は分子間力の中では比較的強い部類にはいる。

(3)水素結合： 電気双極子同士の相互作用が生体高分子では多いが、その内

でも水素をはさんだ形での電気双極子間の相互作用が強い。これを水素結合と言う。

(4)ファンデルワールス力： 全ての原子に働く力で、短距離になると強くなる。蛋白質をコンパクトにするのにはファンデルワールス力がかなり寄与している。

### 3. タンパク質における分子間力の分業

タンパク質の立体構造には階層があるが、一般に構造に階層がある場合、構造形成の主役となる相互作用が階層によって異なっていることが多い。つまり、相互作用が分業していることが多いのである。タンパク質の階層構造の形成もそうで、特に膜タンパク質では顕著である。

タンパク質のアミノ酸配列がランダム構造をとって水の中にあつたとすると、水の中で強い相互作用（つまり疎水性相互作用）が働き、球状に折れ畳まったり、膜の中（極めて非極性の媒質）へ入っていったりする。つまり、あるアミノ酸配列が水溶性タンパク質になるか、膜タンパク質になるかは疎水性相互作用の働きによって決まっていると考えられる。一方、膜に入った膜タンパク質には強い極性の相互作用（静電相互作用と水素結合）が働き、その先の構造形成を駆動すると考えられる。すなわち膜タンパク質の高次構造は主に水素結合によってできているのである。このような分子間力の分業という現象を用いると、アミノ酸配列の立体構造形成を比較的容易に理解することができる。

### 4. 膜タンパク質の立体構造予測

アミノ酸配列の知識情報処理が様々な手法で行われつつある。しかし、先に述べたようにアミノ酸配列の場合は単なる文字列（単なる情報）と考えるわけにはいかない。文字列は物理化学的な現象（立体構造形成）を引き起こすのであるが、その現象は分子間力の働きによって起こる。そうだとすれば、アミノ酸配列を単純な20種類の文字による文章と考えるよりは、アミノ酸配列をそれから得られる分子間力の配列に直し、それを知識情報処理にかけべきである。しかしながら、長距離相互作用については単純にインデックス化することができないので、常にアミノ酸配列を単純な数値列に直すことができるとは限らない。このことを注意して、私達は膜タンパク質の立体構造予測を行っている。

以下、研究の内容を紹介するが、階層構造と分子間力の分業ということに関連して、もうひとつ指摘しておかねばならないことがある。それは構造を見るときに分解能の問題である。例えば、次に述べるように、あるアミノ酸配列が膜に入るかあるいは水溶性タンパク質となるかということは、数千個の原子の一個一個が決定的な意味を持っているわけではない。それよりもむしろ全体の平均的な性質やかなり大きな部分の平均的な性質によって決まっている。このことは階層構造の各階層で言えることで、予測をするときにどの階層を注目するかということによって適切な分解能が違ってくるのである。

(1)膜タンパク質の判別： 任意のアミノ酸配列が与えられたときにまず大きな意味での形態（水溶性タンパク質か膜タンパク質か）を知る必要がある。これは最も荒い階層であって、これは疎水性相互作用が重要である。疎水性インデックスを用いて疎水性の数値列に変換し、それから情報処理をすることになる。こういう一種の前処理を行うと、情報処理の手法に関わらず、かなりよい成績が得られる。私達は、先に述べたように、この階層は長距離の平均的な性質が重要で、分解能を落として解析することが良いと考えている。そこで、フーリエ変換による長周期のスペクトル密度を用いて判別したところ、およそ97パーセントの判別率が得られた。

(2)膜タンパク質の二次構造の予測： これはやはり疎水性相互作用によって、80パーセント程度の判別ができる。

(3)膜タンパク質の三次構造予測： 膜の中で膜貫通ヘリックス同士が相互作用して配置や傾きを決める段階である。この段階では膜の中での極性相互作用が重要であると同時に、前の階層よりは高い分解能の解析が必要である。これには二つの難しさがある。ひとつは極性相互作用が長距離相互作用であるということと、もうひとつは分解能を上げなければならないことである。まだ、この階層の予測は誰も成功していないが、私達は膜タンパク質が膜貫通ヘリックスの束になっているという事実を用いて、ヘリックス間のジグソーパズル問題としてとこうと試みている。このために、まず各ヘリックスのまわりの極性の相互作用場をエネルギー計算によって評価した。すなわち、注目するヘリックスと場を計算するためのプローブヘリックスのエネルギー計算を行ったのである。それによって、ヘリックスの軸方向および円周方向の場の形を求め、それをできるだけ合わせるようにジグソーパズルを解こうとしているのである。場の形まではいくつかのタンパク質で明かになり、ジグソーパズルの問題になり得るという印象を得ている。まだ最終的な結果は得ていないが、この解析でも情報処理までの前処理が極めて重要であるということを強調しておきたい。

#### 4. おわりに

生物（ゲノム）情報では情報学と物理学が2つの意味で関係している。第一に、対象であるDNA配列やアミノ酸配列が情報を持つと同時に、物理的な存在としての高分子に書き込まれている。したがって、考え方の上で、情報学的な考えと物理的考え方の両方が必要であると思われる。アミノ酸配列について言えば、より物理的な知識が必要となってくる。第二に、アミノ酸配列の解析のために物理的な考え方を推し進めていくと、最終的な評価の手法として情報学の助けがどうしても必要となってくる。今後、ゲノムの知識情報処理において、情報学と物理学の結合がますます緊密になっていくと期待される。