3D-1D 法の可能性

国立遺伝学研究所 太田元規

Essential ideas in the protein 3D-1D compatibility analysis are briefly introduced. The 3D-1D alignment was revolutionary in the establishment of the methodology. The method is applicable for structure prediction, stability evaluation of point mutants, deducing functional sites etc. In these applications it appears free of the protein folding problem, however, this is not the case if we try to design de-novo sequence.

structure prediction / structure-sequence alignment / three-dimensional (3D) profile / protein stability / protein folding

1. はじめに

昨今、タンパク質の立体構造予測法の新手の切り札として3D-1D法が注目されている¹⁾。筆者はここ数年、(少なくとも本人の自覚においては)主に3D-1D法の抱えている基礎的な問題に取り組んできた。その立場から見ても、3D-1D法はまだまだ解釈をはっきりさせなくてはならない不確かな要素を抱えているものの、タンパク質の解析手法として認知され、その適用範囲を拡大しているように思う。ここでは3D-1D法のシンパの一人としてこの方法を、特に3Dプロフィールをめぐる問題を中心に紹介したい。

2. 3D-1D法の起源と歴史

3D-1D法の原点は1つの論文に凝縮されているように思う。Bowie et al. (1990) *Proteins*, 7, 257-264。ここで彼らは1つの発想の試みを行う。構造と配列をアラインメントする、という発想である²⁾。アラインメントとは、タンパク質やDNAの類縁関係を推し測る場合に用いられる工程の一つで、対応するサイト同士を併記する行為を言う。この論文以前では配列と配列または構造と構造を並べることをアラインメントと言った。ここで初めて、構造と配列というタンパク質の表裏を成す異質なものの同士を並べる手法が提案されたのである。やり方はとても簡単というか稚拙でさえある。アミノ酸の疎水性を見て、配列を疎水性を反映した文字列に変換する。構造はタンパク質内部への埋もれ度を

見て、それを反映した文字列に変換する。文字列同士の“スコア表”があれば2つをアラインメントすることは動的計画法 (Dynamic Programming : DP) を使えば容易い。

その後MITにいたBowieはUCLAに移り、3D-1D法を世界に知らしめたScience論文を上梓する³⁾。Bowieのスコアは基本的にはサイト周りの環境のみに依存する一体型の関数であったが、これとは独立にSipplはベアワイズの経験的関数の作成形式を提案し、アラインメントをしないいわゆるSippl testで、天然配列が天然構造を認識できることを示した^{4)・5)}。そしてSipplの形式をDPの形式にはめ込むべく、Jonesらのダブル・ダイナミック・プログラミング (DDP)⁶⁾やGodzikらの凍結近似 (Frozen Approximation)⁷⁾などの解法が提案されていく。この地点で3D-1D法の黎明期は終わりを告げ、多くの関数 (スコア表) が提案される百花繚乱状態への突入が始まり、それは現在へと尾を引いている。

3. 3D-1D法とは

3D-1D法は名の通り、3D (構造) と1D (配列) の適合具合を問うことによりタンパク質の種々の性質を解析する方法である。ここでは、マッチの度合いを測る関数が重要となる。まず、構造と配列の関係がクリアで冗長性のないPDB (Protein Data Bank ; タンパク質の構造を集めたデータベース) の構造を用意する。この構造群から、経験的な構造と配列の適合性関数を作

Frontiers in the structure (3D)-sequence (1D) compatibility analysis
Motonori OTA
National Institute of Genetics

成する (必ずしも経験的である必要はない、それが有用だから主流になっている)。具体的には、例えばLeuとAlaの距離依存の適合性関数を作成する場合、LeuとAlaの距離関係をこれらの構造群で勘定する。LeuとAlaの側鎖の位置関係をみて、それらがXÅのものがY個、とカウントする。個数がわかればそれを適当な変換を用いてエネルギーライクな関数表にすることができる。これを20×20のアミノ酸種で行なえば、こういったアミノ酸がこういった位置関係にあるのが総合的に好まれるのが数値として与えられる。このように作成したものが適合性関数である。

関数ができれば、配列を構造に実際に乗せてみて適合具合を問う。遠い類縁関係にあるタンパク同士では、配列の置換以外に欠損や挿入は当然起こりうることであるから、それらを許しながらどうやって配列を乗せていくか、が問題となる。これが3D-1Dアラインメントである。側鎖間相互作用関数のような2体ポテンシャルを用いると作業は厄介なので、先に少し触れたDDPという方法を用いる場合もあるが、構造を3Dプロフィールという表に変換し、これを利用する方法もある。3Dプロフィールは配列のホモロジーサーチなどで利用する置換スコア・マトリックス (PAM250, BLOSUM, JTTなど) をサイト毎に展開したものと同一形式をもつので、ノーマルなDPをそのまま使ってアラインメントを行うことができる⁹⁾。アラインメントができてしまえば、構造と配列の適合スコアを導出することは容易だ。このようにして例えば、PDBに登録されている (代表的) 構造をスキャンすれば、予測をしたいタンパク配列がどの既知構造と良くマッチするか評価することができる。3D-1D法による立体構造予測はこのようにして行なわれる。

4. 3Dプロフィールとタンパク質の安定性

筆者らはアラインメントを得る形式の手だてとして3Dプロフィールを利用している。3D-1D法の可能性を考える場合、3Dプロフィールは1つの出発点になると思う。

3Dプロフィールを凍結近似のもとで作成する手順は以下の様なものである (図1)⁹⁾。まず対象とする構造にその天然配列を乗せておく。次に注目するサイトに乗っているアミノ酸を除去し、20種類のアミノ酸を順々に乗せてみる。乗せる度にサイト周りが作る環境との適合度を前述の適合性関数で評価する。評価値をそれぞれのアミノ酸毎N端サイトからC端サイトまで記載すれば、各アミノ酸の各サイトへの適合度を表わした数値表ができる (凍結近似というのは、サイト周

ACDEFGHIKLMNPQRSTVWY

(a) Structure

1	HVIDGLMKASPRCTEQYNFW
2	LIWMYTVPFGNASRHCKDEQ
3	DGSNAEPKTCQHVMIRLYFW
4	DENGPKSQTACHVRILMYFW
5	GSPKDANTCEVHQRILMFYW
6	WGAEYSDNQTHLCPFMIVKR
7	WFVYIEMLAQSTGDKHNCPR
8	DNKGTQSEHPARCYVLMIFW
9	WASTVMGQCPHFNDIYKER
10	WVILMYFQRTKCEHPNDASG

⋮

(b) Profile

図1 3Dプロフィールの作成法。(a) 対象構造に天然配列を乗せ、各サイト周りの構造環境を構成する。注目するサイトに穴をあけて20種類のアミノ酸を挿入し、構造環境との適合度を評価する。(b) 評価した適合度をサイト毎に記載する。ここではアミノ酸を適合度順に左から右へ並べた文字プロフィールを示した。反転文字は天然配列を表わす。天然配列の適合順位が悪くないことが一見してわかる。

りの環境を構成する場合に側鎖を天然配列に“凍結”させる、という意味である)。

3Dプロフィールは元来アラインメントを得るための数値表であったが、筆者らはこれを各サイト毎適合度順に並べたアミノ酸列として表記してみた (図1b)。このようにすれば、コンピュータへのフィードとしてではなく、人間が目で見てもさまざまな情報を直観的に把握することができる。例えば特定の属性を持つアミ

図2 (a) 従来型の側鎖間相互作用関数で作った3Dプロフィール。(b) 規格化法を改良した側鎖間相互作用関数で作った3Dプロフィール。W, G, 天然配列には模様をつけて表示している。

ノ酸に注目し、それらに模様をつけて表示すると一層効果的である。従来利用されていたベアワイズの適合度関数を用いてこのような表記をした場合に、奇妙な現象が観察された。見かけの適合度にはアミノ酸種によるバイアスがかかっている、というものである。つまり、Trpは常に適合度が高い傾向があり、Glyは常に低い。これは明らかに不思議な現象である(図2a)。そこでいったん3Dプロフィールを作成した後、アミノ酸種毎に適合度を規格化するような形式を用いることにした。さまざまな構造周りの環境という集合を作り、各アミノ酸についてこれに対する平均適合度を求め基準値とする規格化である。このようにすると、適合度のアミノ酸種による片寄りは正され、それぞれが適宜サイト環境に適合するようになった(図2b)(詳述すると、3Dプロフィールにアミノ酸種のバイアスがか

かるという現象は筆者らの関数形式に側鎖間のコンタクト情報が加味されていたことに起因するようだ。通常側鎖がコンタクト状態になくても何らかの理由でそこにあるとして統計に加算する形式を使うが、この場合あまり関係のないものがノイズとして入ってくるので感度は落ちる。しかし、コンタクト/ノンコンタクトというフィルタを用いたつてとして、規格化にもある種の工夫が要請される)。またその結果、天然配列の天然環境に対する適合度は概ね良い方に評価されるようになった。つまり“天然配列は天然構造により良く(必ずしも最適ではないが)適合している”という通念も満たすようになった⁹⁾。

では、3Dプロフィールは物理的には一体何を記述したものであるのだろうか? 筆者らは、3Dプロフィールの意味を問うべく1つの試みを実行した。構造と配列の適合性関数がエネルギー様の形式を持つ以上、それをもとに作成された3Dプロフィールはタンパク質の構造安定性と何らかの関係を持つものであろう。従来タンパク質の安定性は理論的には自由エネルギー振動法を利用した分子動力学の精密な計算に基づき議論されてきた。この時、サーモダイナミック・ダイアグラムという状態図を基に置換の及ぼす影響の評価が行なわれる。2つの配列についてdenatured state (非天然状態)とnative state (天然状態)を想定し、それらを準静的に結ぶ。各状態間のエネルギー差を組み合わせ、相対的な安定性を評価するのである。この形式に3Dプロフィールをあてはめる場合、タンパク質の非天然状態を仮定する必要性が生じた。そこで、前述した各アミノ酸の構造サイトへの平均適合度を各アミノ酸の非天然状態のエネルギーとおいた(タンパク質の非天然状態はさまざまな構造環境の集合体として表現されることになる)。この仮定により、3Dプロフィールに記載された数値を、非天然状態から天然状態への遷移エネルギーと見たてることができる。その結果得られた計算値と、リボヌクレアーゼHというタンパク質の多くの点突然変異体の安定性の実験値は有意な相関を示した(図3)。つまり、3Dプロフィールはタンパク質の安定性を記述したスコア表である(可能性が高いというか、それを理想とすべき)という1つのラインを提示することができた¹⁰⁾。

5. 3Dプロフィールと機能

配列をマルチプルにアラインメントすると保存部位の情報からさまざまな知見が得られるが、タンパク質は機能的高分子であるゆえ、そこには当然機能からの要請を含んだ成り行きが反映されている。モチーフと

図3 リボヌクレアーゼHIの点突然変異体の安定性の実験値と計算値との散布図。(a) すべての実験値と対応する計算値。相関係数は-0.51。(b) 埋もれているサイトでの実験値と対応する計算値。相関係数は-0.67。

呼ばれる機能部位などがこれにあたる。一方で3Dプロフィールは純粋に構造安定性を記述すべきものとして定義された。タンパク質は固有の構造を取って初めて機能が発現するものであるが、配列すべてが構造安定性(形成)に強く寄与しているわけではないであろう。機能部位には最終的に機能を果たすという役割が強く課せられているはずだから、その代償として構造安定性への寄与は免じられている可能性もある。つまり、保存サイトの3Dプロフィールを見ることによって、構造安定性への寄与を抑えてまで保存しないといけない純粋な機能部位をあぶり出すことが可能かもしれない。筆者らはこの可能性を、グロビンという配

図4 グロビン族完全保存残基(CD1; F, F8; H, 黒いバー)のサイト環境への適合度。値が小さいほど適合度が良い。Fは適合しているが、Hは適合しているとは言い難い。Hは安定性を犠牲にしてまでも保存してきた、純粋な機能部位である。

列の発散が激しいファミリーを対象に探ってみた。グロビンはヘムという補酵素を含むので、少しややこしい議論が必要なのであるが、配列の発散に耐えたいいくつかの比較的保存の良いサイトを構造からの要請も満たしているサイト(3Dプロフィールで適合度が良い保存サイト)と純粋に機能的要請からなるサイト(3Dプロフィールで適合度が悪いにもかかわらず保存しているサイト)とに分離することができた(図4)¹¹⁾。3Dプロフィールには、機能部位を裏から論じる種が隠されているのである。

6. 3Dプロフィールによる配列設計

図1bや図2bのような3Dプロフィールに列挙されたアミノ酸列はサイト周りの構造環境に対する適合度でソートされているので、(適合性関数などが理想的であるならば)一番左端にあるアミノ酸は構造安定性に最も貢献するだろう。だが、最も適合しているアミノ酸を構造に乗せよう一度3Dプロフィールを作成してみると、先ほどは最適アミノ酸であったものが今度はそうではない。前に構造環境を構成する際に利用した配列とは異なる配列で環境を構成したからである。そこで、こういう作業を何回か繰り返してみる。最適配列はたいがい2つの配列間で振動を始める。振動配列を合成しながらまた何回か3Dプロフィールを作成し直してみると、ついに振動がとまり、自らが作成した環

境下では自分が最適アミノ酸となる配列が生成される。こういう配列は理論上は最も構造が安定な配列の解である可能性があるが、果たしてそうなっているのだろうか？

このようにしてグロビン構造を鋳型として生成された配列と天然配列の一致度は、25%程度であった。この値は遠い類縁関係にあるタンパク質間の配列一致度と同じ域であるが、だからと言ってグロビン型をとるタンパクであると結論づけるのは早急である。この配列は人工配列であり、進化の過程を経ていないので類縁性の経験則は適応できない。また、今まで安定性の予測はまだしも、タンパク質の配列を新規にデザインする試みはあまり成功を収めていない。3D-1D法は天然配列と天然構造の対応という枠組みで効果を発揮する方法であるから、その枠をはみだしたこのような計算結果は方法の可能性を吟味する上でも検証する必要があるし、その意義も大きいと思える。

筆者らは3Dプロフィールを繰り返し作成する方法で得られた解配列について（あまりにも大きな側鎖のぶつかりに対してはモデリングを行い除去した上で）、実際にそれを作成する試みを行なった¹²⁾。作成されたタンパクのCDスペクトルはヘム非存在下ではアボグロビンとほとんど同じであり、またヘム存在下ではそれを吸着した。

7. 自然への挑戦（結びにかえて）

3D-1D法はタンパク質の構造予測の1手法としてその名が知られるようになった。配列が既知構造ライブラリをサーチするという形でそれは実現した。ここではタンパク質の構造予測という主題がずっと抱えていた難問であるフォールディングの問題はスキップされている。また、3Dプロフィールはタンパクの安定性を示すものであると定義される（べき）ことを示した。そしてその自然な延長としてタンパク質の配列設計を行う指標ともなりうることを述べた。一方で3D-1D法の外枠は天然構造と天然配列との対応に規定されている。この枠に従う限り新規配列の設計へ一步を踏み出すことはできない。枠からの離脱というジャンクションは安定性解析からデザインに滑らかに接続するルートとは相入れないものがある。ここでフォールディング問題がまた登場する。天然配列はフォールドするものとして扱われてきた¹³⁾。その立体構造を議論する限り、巻かない（形を取らない）という状況は想定しな

くても良かった。しかし、設計を考える場合に巻く／巻かないは大きな問題となる。巻く存在としての天然配列を議論することなしに巻く存在としての人工配列は語れない。滑らかさを伴う離脱はフォールディングを内包したものである。その意味で逆のアプローチから3D-1D法はフォールディング問題に直面している。いまのところ設計が順調なのは嬉しいことであるが、このことはいずれ決着をつけなくてはいけないだろう。

「蛋白質は自然が長い時間をかけて慈しみ作りあげた傑作である（斎藤信彦）」¹⁴⁾。美しくて重い言葉だ。新たな視点から自然に挑戦していかなければと思う。

謝 辞

ここで述べた内容は金谷茂則（阪大）、磯貝泰弘と飯塚ラボの方々（理研）および西川建（遺伝研）各氏との協同研究です。また、この小論の内容についての確なコメントをいただいた査読者の方々に感謝いたします。

文 献

- 1) Lattman, E. E. ed. (1997) *Proteins*, suppl. 1, Wiley-Liss, Inc.
- 2) Bowie, J. U. et al. (1990) *Proteins* 7, 257-264.
- 3) Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991) *Science* 253, 164-170.
- 4) Sippl, M. J. (1990) *J. Mol. Biol.* 213, 859-883.
- 5) Hendlich, M. et al. (1990) *J. Mol. Biol.* 216, 167-180.
- 6) Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) *Nature* 358, 86-89.
- 7) Godzik, A., Skolnick, J., and Kolinski, A. (1992) *J. Mol. Biol.* 227, 227-238.
- 8) Needleman, S. B. and Wunsch, C. D. (1970) *J. Mol. Biol.* 48, 443-453.
- 9) Ota, M. and Nishikawa, K. (1997) *Protein Eng.* 10, 339-351.
- 10) Ota, M., Kanaya, S., and Nishikawa, K. (1995) *J. Mol. Biol.* 248, 733-738.
- 11) Ota, M., Isogai, Y., and Nishikawa, K. (1997) *FEBS Letters* 415, 129-133.
- 12) Isogai, Y. et al. : 第4回タンパク質立体構造の構築原理ワークショップ (1997, 三島)
- 13) 西川建 (1993) *Viva Origino* 21, 91-102.
- 14) 斎藤信彦 (1988) 数理科学 301, 75-82.