

経験的に求めた蛋白質の立体構造ポテンシャル

(Empirical Inter-residue potentials for Protein Structures)

宮澤三造 (群馬大学工学部共通講座 376 群馬県桐生市天神町 1-5-1)

e-mail: miyazawa@smlab.sci.gunma-u.ac.jp

近年、蛋白質のアミノ酸配列および立体構造データの蓄積に伴い、構造予測の一環として、また配列比較では同定できないような遠縁の蛋白質の同定に利用すべく、蛋白質におけるアミノ酸配列と立体構造の適合性の研究が盛んになってきた。適合性を計るメジャーとして、また立体構造エネルギーの近似としての両側面から、立体構造データから抽出されたアミノ酸残基レベルでの経験的残基間ポテンシャルについて考察する。

1 はじめに

生物に含まれる、20種のアミノ酸が数百ペプチド結合により線形に結合した鎖状分子である蛋白質は、DNAの塩基配列により定まるアミノ酸配列をもち、生理的条件下、その配列によって決まる固有の3次元構造をとる。蛋白質の立体構造は、原子の揺動を別にして、唯一の状態からなるユニークな構造である。アミノ酸配列から蛋白質の立体構造を予測することは長年の課題であるが、まだ確立した方法論はない。X線解析、NMR等の実験手段の進歩により、また、分子動力学計算により、多くの蛋白質で静的、動的な構造が明らかになりつつある。しかし、蛋白質の変性状態や、蛋白質の変性状態からネイティブ状態までの広範な構造空間に関しては、ほとんどわかっていない。というのも、位相空間を記述する座標変数が多数で

あること、またエネルギー面が複雑であること、蛋白質の折り畳みはミリ秒から秒のオーダーの現象であるが、現在の分子動力学計算で記述できる時間スケールはたかだかなノ秒のオーダーであること等が、その困難さの原因となっている。蛋白質の折り畳み過程の記述において、分子構造表現の単純化、それに伴う相互作用ポテンシャルの単純化が要求されよう。

一方、これまで数百種類の蛋白質の構造が判明しているが、その解析を通じ、自然界にある蛋白質の(骨格)構造には、(多分進化の産物として)、1000 [1] から 7900 [2] 程度の限られた数のプロトタイプがあるとの説が提出されている。もしこの説が正しいなら、アミノ酸配列から蛋白質構造の予測にあたって、探索する蛋白質構造の構造空間を、限られた数の骨格構造に限ることができよう。しかし、よく用いられる点電荷近似による経験的原子間ポテンシャル(たとえば CHARMM)であっても、溶媒効果を考慮しないと、全く異なる蛋白質構造においてアミノ酸残基を交換しエネルギーの最小化を行なうと、自己の構造より低いエネルギー状態に到達することがある [3]。このことは、蛋白質構造のエネルギー評価において、溶媒の考慮が必須であることを示す。しかし、与えられたアミノ酸配列について、プロトタイプである多数の骨格構造の各々に関し、その自由エネルギーを評価するための方法としては、経験的原子間ポテンシャルを用いた、溶媒を含む蛋白質構造のエネルギー計算は、計算時間がかかり過ぎ実際的ではない。幸いなことに、アミノ酸残基レベルでの非常に単純化した経験的ポテンシャルであっても、各アミノ酸の疎水性を考慮に入れたものであれば、蛋白質立体構造とアミノ酸配列間の適合性をある程度正しく計れることがわかっている。そこで、アミノ酸配列と蛋白質構造との適合性判定に適した判定関数が工夫されている。配列と構造との適合性の判定は、構造予測の一環として意味あるだけでなく、配列比較では同定できないような遠縁の蛋白質

を同定できるという利点をもつ。なぜなら、分子進化の過程で、蛋白質構造はアミノ酸配列よりも保存的であるからである。よって、蛋白質構造形成に関与している主要な相互作用を適切に近似するアミノ酸残基レベルでの経験的ポテンシャルの実用的価値は高い。

ここでは、立体構造エネルギーの近似として、また適合性を計るメジャーとしての両側面から、既知の蛋白質構造データから抽出されたアミノ酸残基間の経験的相互作用ポテンシャルについて考察する。

2 立体構造ポテンシャル

ここでは特に、立体構造形成に与る相互作用ポテンシャルについて考える。通常は、定められた相互作用に基づく系に関しその振舞を記述し実験との比較を行うわけであるが、ここでの問題は逆に、蛋白質立体構造における各アミノ酸残基の空間的配置から、アミノ酸残基間の経験的相互作用ポテンシャルを得ようとする試みである。以下では、特に記述されない限り、エネルギーの単位は kT もしくは RT とする。

2.1 化学反応近似によるアミノ酸間接触エネルギーの見積もり

最初の試みは、1976年 Tanaka & Scheraga [4] により、立体構造予測のための経験的相互作用ポテンシャルとして提案された。彼等は、アミノ酸間相互作用は短距離力からなるとし、アミノ酸間接触を化学反応系と仮定し、多数のアミノ酸間接触からなる系の標準自由エネルギーを、2体の標準自由エネルギーの和として近似した。また、既知の蛋白質構造におけるアミノ酸間接触は、熱平衡状態にあると仮定し、アミノ酸間接触の標準自由エネルギーを、既知の蛋白質構造におけるアミノ酸間接触数から評価した。すなわち、 i と j

種アミノ酸の接触反応の標準自由エネルギーは

$$\Delta G_{ij}^0 = -\ln\left[\frac{X_{ij}}{X_i X_j}\right] \quad (1)$$

ここで X_{ij} は i と j 種アミノ酸間接触のモル分率である。 X_i はいずれのアミノ酸とも接触していない i 種アミノ酸のモル分率である。

$$\sum_i \left(\sum_{j \leq i} X_{ij} + X_i \right) = 1 \quad (2)$$

ただし、アミノ酸に属する任意の原子が互いに、1つの水分子もその間に介在できない程隣接した時、2個のアミノ酸は接触していると定義した。また、鎖状分子としての影響を排除するため、アミノ酸鎖の上で4残基以上離れたアミノ酸間接触のみを対象とした。疎水性等のアミノ酸の性質を定性的に反映した接触エネルギーが得られた。

2.2 ベーテ近似によるアミノ酸間接触エネルギーの見積もり

蛋白質立体構造で観測されるアミノ酸間隣接頻度は、アミノ酸間相互作用とどの程度整合しているであろうか？鎖状分子としての形状は、アミノ酸鎖の上で互いに近くに位置しているアミノ酸程、隣接する確率を増加させる。また i 番目の残基と j 番目の残基間の接触 (i - j 接触) と i と j に充分近い残基間での接触 ($(i+k)$ - $(j+k+l)$ 接触) には、正の相関があるであろう。一方、アミノ酸間の排除体積効果は負の相関をもたらす。蛋白質立体構造におけるアミノ酸接触に関し、アミノ酸配列の及ぼす影響を否定できない。しかし、アミノ酸間相互作用と矛盾するようなアミノ酸間接触が多数存在するとしたら、不都合はないだろうか？そのような蛋白質は安定であろうか？蛋白質は長い分子進化の過程を経て来ている。その間多くのアミノ酸置換が生じたことが、蛋白質アミノ酸配列の比較から判明している。(ヘモグロビン分子で約 $2 \cdot 10^{-7}$ アミノ酸置換 / 年 / 残基)。

そのような分子進化の過程で、アミノ酸間相互作用と矛盾するようなアミノ酸間接触は淘汰されるのではなかろうか？ 一方、これまでの蛋白質構造の研究、特に蛋白質の 2 次構造予測の研究において、多くの異なる側面からの知見に基づく 2 次構造予測がいずれも 60% 近くの前測率を達成する事実等から、蛋白質において各種相互作用は概ね整合している (consistency principle) [5]、とする仮説が Go により提案されている。

以上のような考察に基づき、Miyazawa & Jernigan [6, 7] は、蛋白質構造を各アミノ酸が格子点に位置する格子構造としてとらえ、各種アミノ酸に特有な相互作用のもとアミノ酸が格子点を自由に動けるような仮想的系を考え、多数の蛋白質構造における統計的性質としてのアミノ酸間接触頻度は、このような仮想的系の熱平衡状態におけるアミノ酸間隣接頻度と一致する、と仮定した。アミノ酸鎖としての性質は平均場としてのみ考慮される。

蛋白質内でのアミノ酸間相互作用は、主に疎水性相互作用、ファン・デア・ワールス力、水素結合、そして静電相互作用からなり、静電相互作用を除き短距離力である。それ故、この仮想系におけるアミノ酸間相互作用は、最隣接相互作用のみと仮定する。この仮想系は、Ising スピン系と類似である。ただ、スピン系の場合、上向き、下向きの 2 種類のスピンからなるのに比べ、この系は 20 種類のアミノ酸の混合系である。大きな違いは、この系は通常 100 - 600 程度のアミノ酸から構成されている、という点である。つまり個々の蛋白質に対応するこの仮想系に熱力学的極限を当てはめることは不可能である。しかし一方、ここには多種類の蛋白質に対応して数百の類似な系が存在する。

言うまでもなく、3次元 Ising スピン系の厳密解はいまだ得られていない。多くの近似的解法が知られているが、残基間接触頻度から接触エネルギーを計算するには、相互作用ポテンシャルから最隣接

対の統計平均を顕に見積もる近似でなければならない。そこで、ランダム混合を仮定する平均場近似を第0近似とするなら、第一近似とも言えるベータ近似を採用した。

最隣接相互作用を仮定すると、蛋白質の全残基間接触エネルギーは以下のように表現される。

$$\begin{aligned} E^c &= \sum_i \sum_j E_{ij} n_{ij} \\ &= \sum_i (2E_{i0} - E_{00}) q_i n_i / 2 + e_{rr} n_{rr} + \sum_i \sum_j (e_{ij} - e_{rr}) n_{ij} \quad (3) \end{aligned}$$

ここで i 種、 j 種アミノ酸間の相対的な接触エネルギー e_{ij} は、絶対相互作用エネルギー E_{ij} から以下のように定義される。

$$e_{ij} \equiv E_{ij} + E_{00} - E_{i0} - E_{0j} \quad (4)$$

0 は空格子点を意味し、空格子点は溶媒 (水) で占有されているとする。 n_{ii} は i 種アミノ酸同士の隣接対の数、 $2n_{ij}$ は i 種、 j 種アミノ酸の隣接対の数である。また

$$n_{r0} \equiv \sum_i n_{i0} \quad n_{rr} \equiv \sum_i \sum_j n_{ij} \quad (5)$$

とする。 q_i は各アミノ酸と隣接し得るアミノ酸数 (最隣接格子点数 / 配位数 (coordination number)) である。 e_{rr} は平均的アミノ酸間の接触エネルギーに相当する。(式(6)で添字 i, j を r に置き換えた $\exp(-e_{rr}) \equiv \overline{n_{rr} n_{00}} / \overline{n_{r0} n_{0r}}$ で定義される。) コンフォメーションに依存するエネルギー (コンフォメーションエネルギー) としては式(3)の第2項、第3項を考慮すれば充分である。第2項は蛋白質を凝縮 (collapse) させるのに必要なエネルギー (collapse energy) であり、第3項は蛋白質構造内でのアミノ酸の配置エネルギーに相当する。

ベータ近似により、相対的な接触エネルギーと最隣接対の統計平

均に関し、以下の関係式を得る。

$$\exp(-e_{ij}) = \frac{\overline{n_{ij}n_{00}}}{n_{i0}n_{0j}} \quad (6)$$

ここで $\overline{n_{ij}}$ は n_{ij} の統計平均を意味する。つまり、よく知られているように、ベータ近似では隣接対に関しあたかも化学平衡近似 ($i-0 + 0-j \rightleftharpoons i-j + 0-0$) を適用したのと同じ結果となる。この意味では、2.1の方法に類似している。

ここで、溶媒分子間の接触数 n_{00} をどうやって計算するかは後回しにして、 n_{00} を消去した以下の方程式を考えよう。

$$\exp(e_{ij} - e_{rr}) = \frac{\overline{n_{ij}n_{r0}n_{r0}}}{n_{i0}n_{j0}n_{rr}} \quad (7)$$

凝縮 (collapse) エネルギーを差し引いた接触エネルギー ($e_{ij} - e_{rr}$) は $\overline{n_{00}}$ によらないことに注意してもらいたい。

アミノ酸間接触の定義は、近似の荒さとの整合性を考え、アミノ酸側鎖の中心間の距離が、アミノ酸間の動径分布関数から求めた最隣接に対応する $R_c = 6.5\text{\AA}$ 内の場合、互いに接触しているとした。ただし、立体構造においても隣接する確率の高い、アミノ酸鎖で両隣の残基は接触残基に含めないものとする。アミノ酸と溶媒間の接触数 n_{i0} は、 $\sum_j n_{ij} = (q_i n_i)/2$ の関係式より求める。 n_i は i 種アミノ酸の数である。配位数 q_i は残基の平均体積から定めた。接触エネルギーを式 (7) から評価するには、一蛋白質内での接触数の統計平均 $\overline{n_{ij}}$ を多数の蛋白質に関する n_{ij} の和 (N_{ij}) に置き換えて、見積もる。(実際には蛋白質のアミノ酸数、成分の違いを考慮する補正項を含む。)

一接触あたりの凝縮 (collapse) エネルギー e_{rr} の見積もりは、 n_{00} もしくは n_0 、すなわち系のサイズを見積もる必要がある。蛋白質が鎖状分子であることの性質は、ただ系のサイズとしてのみ考慮する。具体的には、アミノ酸間相互作用がないような仮想的蛋白質を、要素間結合の向きが自由な線形鎖 (Freely jointed chain) として近似し、

その系での接触数を見積もり、系に含まれる溶媒の数 n_0 を評価した。

2.2.1 実験値との比較

接触エネルギーの推定値は、実験から得られた各種アミノ酸側鎖の疎水エネルギーと比較された。また、アミノ酸配列と蛋白質構造の適合性判定に使用した結果、最低エネルギー状態としてのネイティブ構造の同定、および、配列比較と同程度に、互いに適合する配列と構造の識別が可能であった。

図 1 は、各種アミノ酸が蛋白質構造に完全に埋め込まれた際の接触エネルギー [7] を、アミノ酸側鎖を水から有機溶媒へ移動させた際の転移自由エネルギーの評価値 [8] と比較したものである。ここでは、グリシンを基準にとってあるから、接触エネルギーは e_{rr} の寄与を含まない。疎水性残基以外のアミノ酸の比較は無意味だが、参考のため示してある。疎水性残基 (Ala, Val, Ile, Leu, Phe) に関する値はほぼ一致する。(しかし、この転移自由エネルギーの評価値は、従来の評価値のほぼ 2 倍に等しく、通常転移自由エネルギーの評価値 [9] と比較すると、接触エネルギーからの見積もりは、約 2 倍ほど大きい [6, 7].)

一方、アミノ酸置換による蛋白質安定性の変化の測定においては、ある座位で他 19 種のアミノ酸へ置換した際の変性自由エネルギー変化 (Unfolding free energy change、変性させるのに必要な自由エネルギーの変化) は置換による疎水エネルギー変化と相関するものの、回帰直線の傾きは座位により異なり、よく引用される疎水エネルギー変化の評価値 [9] と比較するとその 3 倍以上も大きな変性自由エネルギー変化も報告されている [10]。図 2 は、tryptophan synthase α における Glu-49 を他 18 種のアミノ酸へ置換した際の変性自由エネルギー変化 [10] を、期待される接触エネルギー変化 [6] に対してプロットした図である。疎水エネルギー変化 [9] に対する回帰直線の傾

き 3.7 に較べ、接触エネルギー変化の場合、傾きは 0.88-1.2 となり、もったもな結果を得る [11]。よって、接触エネルギーの見積もりはこれまで見積もられた疎水エネルギーのほぼ上限に対応するが、一概に、大きすぎるとは言えない。また、蛋白質内環境は必ずしも有機溶媒と同一ではないことも考慮しなくてはならない。

以上 2 つの比較において、温度 T としては、本来は蛋白質の転移温度を用いるべきだが、実験が行われた時の温度 (室温) を用いている。

凝縮 (collapse) エネルギー e_{rr} の推定値は、誤差が大きいと思われるが、 $-2.34kT$ から $-2.55kT$ であり、1 残基あたり $-4.7kT$ から $-5.4kT$ に相当する [6, 7]。ランダムコイル状態からネイティブ状態への転移に伴う側鎖のコンフォメーションエントロピーの変化は $-1.7kT$ /残基程度である [12]。よって、凝縮エネルギーのこの推定値は、主鎖のコンフォメーションエントロピー変化が $-3kT$ から $-3.7kT$ /残基程度であることを示唆している。

2.3 平均力ポテンシャルを用いたアミノ酸間相互作用ポテンシャルの評価

もう一つの試みは、蛋白質構造における各種アミノ酸間の動径分布関数から求めた平均力ポテンシャルによって、アミノ酸間相互作用ポテンシャルを近似する方法である。Sippl [13] は、各種アミノ酸間の正味の相互作用ポテンシャルを、距離 (d) の関数である相対的な平均力ポテンシャルとして、以下のように評価した。ただし構造データからの評価においては、少数サンプルへの考慮が払われた。

$$\begin{aligned}\Delta E_k^{ij}(d) &\equiv E_k^{ij}(d) - E_k(d) \\ &= -\ln \frac{f_k^{ij}(d)}{f_k(d)}\end{aligned}\quad (8)$$

ここで $f_k^{ij}(d)$ は、アミノ酸配列で k 残基離れた i 種、 j 種アミノ酸が、立体構造において距離 d のところに位置する頻度である。 $f_k(d)$ は、アミノ酸の種類を無視した時のその頻度である。 $E_k^{ij}(d)$ はそのようなアミノ酸間に働く平均力ポテンシャルである。

言うまでもなく、平均力ポテンシャルは注目する 2 アミノ酸だけでなく他のアミノ酸からの多体効果を含む。例えば、剛体球からなる系の動径分布関数は、おおよそ直径の整数倍の位置にピークを持つ。剛体球密度が高いほどピークは顕著である。よって、動径分布関数からもとめた ($E_k^{ij}(d)$ に相当する) 2 体の平均力ポテンシャルも、実際の剛体球間相互作用には存在しない複数のピークを持つ。よって、式 (8) では、正味の相互作用ポテンシャルを、相対的な平均力ポテンシャルとして評価している。しかし、異なる半径の 2 種類の剛体球からなる系の場合、式 (8) を適用して求められる相対的な相互作用ポテンシャル ($\Delta E_k^{ij}(d)$) は、真のポテンシャルにはない複数のピークを持つ。このような簡単な例でも明らかのように、式 (8) から求められる平均力ポテンシャルは多体効果を含み、このような平均力ポテンシャルの和として、実際の相互作用エネルギーがどの程度近似可能なものか不明である。

このタイプの経験的ポテンシャルは、主にアミノ酸配列と蛋白質構造の適合性判定のための判定関数として提案され、既知の蛋白質構造のなかでネイティブ構造を最低エネルギー状態として同定 [14, 15]、また、適合する配列と構造の識別が、通常の配列比較以上に可能であったと報告されている [16, 17]。

2.4 上記 2 つの近似の比較

最隣接相互作用を仮定した 2.2 と異なり、2.3 の Sippl 型ポテンシャルでは相互作用における距離依存性が平均力場として評価され

ている。それ以外には、**2.2** と **2.3** のポテンシャルにはどのような相違があるのであろう。式 (8) の定義に従い、以下のような関係式を考えよう。ただし、以下では簡単のため添字 k は省略した。

$$\begin{aligned}
& \Delta E^{ij}(d) + \Delta E^{rr}(d) - \Delta E^{ir}(d) - \Delta E^{rj}(d) \\
&= -\ln\left[\left(\frac{2\overline{n_{ij}(d)}}{n_i n_j}\right) / \left(\frac{2\overline{n_{rr}(d)}}{n_r(n_r-1)}\right)\right] + \ln\left[\left(\frac{2\overline{n_{ir}(d)}}{n_i(n_r-1)}\right) / \left(\frac{2\overline{n_{rr}(d)}}{n_r(n_r-1)}\right)\right] \\
&\quad + \ln\left[\left(\frac{2\overline{n_{rj}(d)}}{(n_r-1)n_j}\right) / \left(\frac{2\overline{n_{rr}(d)}}{n_r(n_r-1)}\right)\right] \\
&\simeq -\ln\left[\frac{\overline{n_{ij}(d)n_{rr}(d)}}{n_{ir}(d)n_{rj}(d)}\right] \\
&= e_{ij}(d) + e_{rr}(d) - e_{ir}(d) - e_{rj}(d) \tag{9}
\end{aligned}$$

ここで $n_{ij}(d)$ は距離 d のところにある i - j 対の数である。また $e_{ij}(d)$ は以下で定義される。

$$\exp(-e_{ij}(d)) \equiv \frac{\overline{n_{ij}(d)n_{00}(d)}}{n_{i0}(d)n_{0j}(d)} \tag{10}$$

上式はベータ近似により導かれた式 (6) に対応するが、ここでは $e_{ij}(d)$ の定義式であり、意味が異なることに注意願いたい。また以下の関係式が得られる。

$$e_{ij}(d) - e_{rr}(d) \simeq \Delta E^{ij}(d) + \Delta E^i(d) + \Delta E^j(d) \tag{11}$$

$$e_{ir}(d) - e_{rr}(d) = \Delta E^{ir}(d) + \Delta E^i(d) \tag{12}$$

$$\Delta E^i(d) \equiv -\ln\left[\left(\frac{\overline{n_{r0}(d)}}{n_r}\right) / \left(\frac{\overline{n_{i0}(d)}}{n_i}\right)\right] \tag{13}$$

ここで $\Delta E^i(d)$ は一種の疎水エネルギーである。

さて、距離 d に関し荒い見方を採用し、 $d \leq R_c$ と $d > R_c$ の範囲に分けて考えた時、 $e_{ij}(d \leq R_c)$ はベータ近似により求められた接触エネルギー e_{ij} に相当する。

$$e_{ij}(d \leq R_c) = e_{ij} \equiv E_{ij} + E_{00} - E_{i0} - E_{0j} \tag{14}$$

よって式 (11) と (14) から、 $\Delta E^{ij}(d)$ は、 $\Delta E^i(d)$ と蛋白質が球状に凝縮するのに必要な凝縮エネルギー ($e_{rr}(d)$) を含んでいないことがわかる。結局 Sippl のポテンシャルを用いた場合は、式 (3) を参照して、コンフォメーションルエネルギーは以下のように表現するのが適当であろう。

$$E^{conf} = \sum_d e_{rr}(d)n_{rr}(d) + \sum_d \sum_i \sum_j (e_{ij}(d) - e_{rr}(d))n_{ij}(d) \quad (15)$$

$$= \sum_d e_{rr}(d)n_{rr}(d) + \sum_d \sum_i \sum_j (\Delta E^{ij}(d) + \Delta E^i(d) + \Delta E^j(d))n_{ij}(d) \quad (16)$$

けれども、前節で指摘したように、相互作用ポテンシャルの距離依存性を平均力場として評価することの是非は、議論の余地がある。

疎水相互作用に起因する $\Delta E^i(d)$ はアミノ酸残基に強く依存する。よって、アミノ酸残基によらない凝縮 (collapse) エネルギーは別にして、残基間相互作用エネルギーを正しく評価するには、 $\Delta E^i(d)$ の評価が必須である。同様な指摘は Sippl [20] によってもなされている。Sippl の相互作用ポテンシャルとともに用いられた $\Delta E^i(d)$ に相当する疎水エネルギーとしては、以下のようなポテンシャルがある。

$$\Delta E^i(h) = -\ln\left(\frac{f^i(h)}{f(h)}\right) \quad (17)$$

$f^i(h)$ は i 種アミノ酸の h に関する頻度分布、 $f(h)$ はアミノ酸の種類を無視した際のその頻度分布である。 h としては、アミノ酸の溶媒への露出面積 (accessible surface area) [15]、露出比 (accessibility) [16]、半径 r と $r + \Delta r$ 内にある他アミノ酸の数 [17]、等が用いられた。これらの関数は、いずれも疎水エネルギーを反映しているが、 $\Delta E^i(d)$ とは異なる。

3 経験的残基間ポテンシャルは真の相互作用をどの程度反映しているか？

以上紹介した経験的ポテンシャルは、残基間接触に関して、ボルツマン分布を仮定している。また、蛋白質の鎖状分子としての性質を無視している。これらの仮定の適否、またエネルギーポテンシャルの推定方法の適否は、検討されねばならない

3.1 蛋白質構造における残基間接触頻度分布は、ボルツマン分布か？

生理的条件下における個々の蛋白質の立体構造は構造の揺らぎを別にすればユニークな構造であり、多数の構造からなるアンサンブルではない。よって、蛋白質立体構造における隣接アミノ酸対の頻度のような統計量は、決して、蛋白質立体構造におけるある特定の残基対の構造に関する時間軸にそった頻度分布あるいは空間多型の頻度分布を意味しない。前節の接触エネルギーの評価においては、アミノ酸接触対頻度のような統計量を、蛋白質構造という環境の下における位相平均として考えている。また残基間相互作用とその他の相互作用は残基の位置および蛋白質には依らないと仮定し、アミノ酸接触対頻度から残基間相互作用の経験的エネルギーを評価した。

注目する p 番目の残基 / 残基対の構造を ρ_p で表し、その相互作用エネルギーを ϵ , 他の相互作用エネルギーを E とする。注目する残基、残基対の構造を ρ に固定しまた E を固定した時の状態密度を $\Omega(E, \rho_p)$ とする。 Z_p を p 番目の残基 / 残基対の分配関数とすると、多数の蛋白質構造における、そのような残基 / 残基対の構造 ρ の存在確率は以下のようになる。

$$\begin{aligned} & \text{残基対の構造 } (\rho) \text{ の存在確率} \\ &= \frac{1}{\sum_p 1} \sum_p \frac{1}{Z_p} \exp\left(-\frac{\epsilon(\rho)}{kT}\right) \int \Omega(E, \rho_p) \exp\left(-\frac{E}{kT}\right) dE \end{aligned}$$

$$= \exp\left(-\frac{\epsilon(\rho)}{kT}\right) \left[\frac{1}{\sum_p 1} \sum_p \frac{1}{Z_p} \exp\left(-\frac{f(\rho_p, T)}{kT}\right) \right] \quad (18)$$

ここで、 $\epsilon(\rho) + f(\rho_p, T)$ は平均力場のエネルギーである。蛋白質の転移温度付近で、括弧の中の項が、選ばれた蛋白質構造の組に依らず、 ρ だけの関数で近似できれば、蛋白質構造から計算されるこの種の統計量は位相平均とみなせ、転移温度 (T_m) を温度定数として持つボルツマン分布 ($\exp(-\tilde{\epsilon}(\rho)/(kT_m))$); $\tilde{\epsilon}$ は平均力場) をなす。残基間接触頻度とか残基の 2 次構造分布のような統計量においては、アミノ酸配列への依存性は無視できると筆者は考えている。

Gutin 等 [18, 19] は、格子点の各々が残基に相当する 27 格子点からなる立方格子を用い、ランダムヘテロ鎖におけるエネルギー最小構造において、残基間接触頻度がその接触エネルギーに関してボルツマン分布をなすか否かを調べた。ランダムヘテロ鎖は、可能な全ての残基対に対し接触エネルギーを正規分布に従いランダムに付与することにより生成する。立方格子ペプチドの全コンパクト構造を生成することにより、個々のランダムヘテロ鎖のエネルギー最小構造を決定する。これらエネルギー最小構造における残基間接触頻度は、その接触エネルギーに関し、ボルツマン分布に従うという結果が得られた。また温度定数は、これらランダムヘテロ鎖において最少エネルギー構造が安定な構造となるための転移温度 (次節で定義するガラス転移温度 (T_g)) にほぼ等しかった。

けれども、このランダムヘテロ鎖のコンパクト構造のエネルギー分布は正規分布であり、最少エネルギー状態と励起状態間にエネルギーギャップをもつ蛋白質とは異なる。蛋白質構造における相互作用は、変性状態とネイティブ状態間での 2 状態転移をもたらすことから想像つくように、つよい協同性をもつ。鎖状分子としての性質も含めアミノ酸配列そしてその蛋白質構造に特有な相互作用と残基間相互作用が整合しているか否か、無相関と見なせるかは更に検

討されねばならない。

一方、Thomas & Dill [21] は、実際の蛋白質で観測される残基間接触頻度の解析から、残基間接触頻度はボルツマン分布に従うものの、接触エネルギーは等しいとするとその見かけの温度は、蛋白質間で大きく異なるという結果を得た。

3.2 相互作用エネルギーの推定方法は適切か？

Thomas & Dill [21] は 2 次元正方格子を用い、2 種類のアミノ酸からなる長さ $L = 11$ から 18 までの格子蛋白質を考え、1) あらかじめ定められた接触ポテンシャルを用い、 2^L の可能なすべての配列の各々について、コンパクトという条件の下、最低エネルギー状態を求め、2) ユニークな格子蛋白質構造を決定、3) この格子蛋白質を用い、2.2 と 2.3 に記述した方法でポテンシャルを推定した。接触ポテンシャルの推定値は、格子蛋白質の長さ、アミノ酸組成に依存した。また、実際の蛋白質でも接触エネルギーの推定値が蛋白質のアミノ酸組成および蛋白質構造の表面体積比に依存する、との結果を報告している。Sippl の平均力ポテンシャルに関しては、真のポテンシャルには存在しない波形状の距離依存性を持つことを示した。この波形状の距離依存性は多体効果に起因するものであろう。

一方、接触ポテンシャルの推定にベータ近似を用いているため、当然ながら、実際の蛋白質で観測される、表面積 ($\propto L^{2/3}$) の鎖長 (L) の $2/3$ 乗依存性は再現できない。表面積の期待値は、鎖長に比例してしまう [6]。よって、接触ポテンシャルの推定値は、蛋白質の長さに依存することとなる。蛋白質は 50 - 500 程度の長さの分布を持つが、推定値は平均値として考えざるを得ない。

3.3 折り畳み可能な格子蛋白質を用いたエネルギー推定方法の検証

Thomas & Dill [21] の検証では、ユニークな格子蛋白質を用いている。これらのユニークな格子蛋白質のほとんどは、蛋白質とは異なりランダムヘテロ鎖における最小エネルギー構造同様、励起状態間にエネルギーギャップをもたない [22]。しかし、蛋白質が適当な時間内に最低エネルギー状態に折り畳まれるためには、最低エネルギー状態と変性状態間のエネルギーギャップが十分大きくなければならない。

Mirny & Shakhnovich [23] は、蛋白質構造の条件の中から、最低エネルギー状態というだけでなく、折り畳み可能性という条件をも満たす 3 次元格子蛋白質を用いて、類似の検証を行った。1) 27 格子点からなる立方格子蛋白質においてコンパクトな構造をランダムに 200 選ぶ。2) その各々に対し、蛋白質構造データより推定された残基間接触ポテンシャル [6] を “真” のポテンシャルとして仮定し、20 種のアミノ酸からなる、折り畳み可能な最適なアミノ酸配列を決定。3) 接触ポテンシャルを推定。4) 格子蛋白質を折り畳むモンテカルロシミュレーションを実行。変性状態として、最低エネルギー状態である蛋白質構造と同等にコンパクトな、残基間接触数も等しい、しかし残基間接触は一様に分布しているような状態を考え、そのような蛋白質構造における平均エネルギーと最低エネルギーとの差を最大にするようなアミノ酸配列を、その蛋白質構造に折り畳み可能な最適なアミノ酸配列としてデザインした。

2.2 の方法で推定した接触ポテンシャルは仮定した接触ポテンシャルとよく相関 (相関係数 0.75) した。モンテカルロシミュレーションでは、“真” の接触ポテンシャルの場合は 100% 折り畳み可能、一方、推定値を接触ポテンシャルとして持つ格子蛋白質の場合はその 95% が最低エネルギー状態に到達できた。2.2 の方法による接触ポテン

シャルの推定値は”真”の値をある程度反映している、との結果は、3.2 [21] の結果と正反対である。この違いは、[21] ではすべて可能なアミノ酸配列を考慮したため、ユニークな最低エネルギー構造ではあるが、蛋白質としては不適な折り畳まれにくい構造が含まれたためである、と指摘されている。

Mirny & Shakhnovich [23] は、また、100 個の実際の蛋白質を使用し、類似の方法でアミノ酸配列の最適化、接触ポテンシャルの推定、比較を行っている。使用する蛋白質の数に関して接触ポテンシャルの推定値が収束することから、蛋白質データベースが 20 種類のアミノ酸間の接触ポテンシャルである 210 個のパラメーターの推定に十分であることが示された。一方、実際の蛋白質においては、接触ポテンシャルだけでは、容易に折り畳める程のエネルギーギャップを生成することは難しいそうであることが示唆されている。

4 蛋白質の折り畳み可能性

前節でも議論があったように、蛋白質構造における残基間接触頻度のような統計量がボルツマン分布をなすか否か、また推定方法の妥当性を検証するには蛋白質らしさを取り入れた系でなくてはならない。最近、蛋白質を特徴づける性質として折り畳み可能性が指摘された。一方、既知の蛋白質構造を最小エネルギー構造として持つだけでなく、これらの蛋白質構造が折り畳まれやすくなるような残基間接触ポテンシャルを求めようとする試みもある。以下で、蛋白質の折り畳み可能性に関する条件について簡単に考察する。[24, 25]。

変性状態から蛋白質構造への折り畳みは、1 msec から 100 sec、もしくはそれ以上の時間スケールで生じる。小さな蛋白質では、この折り畳み / 変性過程は、通常、2 状態転移である。従来、蛋白質を記述する簡単なモデルはすべて、(有限系である蛋白質に特徴的な) 2

状態転移を説明するもの、もしくはそれのみを考慮に入れて議論がなされた。以下で説明するモデルでは、2状態転移に加え、蛋白質がその折り畳みの過程で、エネルギー表面に存在する多数の極小値にトラップされずに最小エネルギー状態に到達するための条件を考察する。エネルギー表面に多数の極小値が存在するという点に関し、スピングラスとの共通点に着目する。

化学反応論における反応座標にならい、折り畳みの座標軸を考え、 x で表そう。多くの相互作用が関与しているから、 x 状態でのエネルギー分布を平均 $\bar{E}(x)$ 、分散 $\delta E^2(x)$ のガウス分布で近似 (random energy approximation) しよう。すなわち、状態密度 $\Omega(E, x)$ を、

$$\Omega(E, x) = \frac{1}{\sqrt{2\pi\delta E^2(x)}} \exp\left(\frac{S_0(x)}{k} - \frac{(E - \bar{E}(x))^2}{2\delta E^2(x)}\right) \quad (19)$$

と仮定する。ここで $S_0(x)$ は、配置エントロピーである。この時、最も確からしい状態のエネルギーは

$$E_{m.p.} = \bar{E}(x) - \frac{\delta E^2(x)}{kT} \quad (20)$$

であるから、 x に対応する状態の自由エネルギーは

$$\begin{aligned} F(x) &= E_{m.p.} - TS(E_{m.p.}, x) \\ &= \bar{E}(x) - \frac{\delta E^2(x)}{2kT} - TS_0(x) \end{aligned} \quad (21)$$

となる。よって転移が、ユニークなネイティブ状態 (x_n) と変性状態 (x_d) との2状態間で生ずるとすると、転移温度は以下のようになる； $\delta E^2(x_n) = 0$ 、 $S_0(x_n) = 0$ である。

$$kT_m = \frac{-\Delta E + [\Delta E^2 - 2\delta E^2(x_d)S_0(x_d)/k]^{1/2}}{2(S_0(x_d)/k)} \quad (22)$$

ここで ΔE は x_n と x_d 状態のエネルギー差 ($\Delta E \equiv E(x_n) - \bar{E}(x_d)$) とする。又、 $S(E_{m.p.}, x) = 0$ なる温度を考えよう。

$$kT_g(x) = \left[\frac{\delta E^2(x)}{2S_0(x)/k}\right]^{1/2} \quad (23)$$

この時、エントロピーが0であるから、系は x の状態の一つにトラップされてしまう。しかも、このような状態は唯一ではなく多数存在し得る。その内のどの微視的状态にトラップされるかは、系の履歴によるであろう。つまり $T_g(x)$ はガラス転移温度に相当する。

この簡単なモデルから、蛋白質が折り畳まれるためには、転移温度の比 $T_m/T_g(x)$ が、十分 1 より大きくなければならないことがわかる。そのためには、 $\Delta E(x_d)$ は大きく、 $\delta E^2(x)$ は小さいことが望ましい。つまり、変性に伴うエネルギー変化が大きく、系の相互作用エネルギーは minimum frustration の状態にあること、が蛋白質が折り畳まれるための必要条件となる。後者の条件は、各種相互作用における整合性原理 (Consistency principle) [5] と等価である。

最近、Li et al. [22] は、エネルギーギャップ大なる条件が配列、構造に課す制約に関し、興味深い研究を報告している。疎水、親水基の2種類の残基からなる 27mer の3次元格子蛋白質において、コンパクトコンフォメーションがコンパクトでないものより低いエネルギーをもつような接触エネルギーのもとでは、(a) 2^{27} 種類の配列の 4.75% がユニークな最低エネルギー状態をもつ。(b) 与えられたコンパクトコンフォメーションを最低エネルギー状態としてもつ配列の数 (N_s) は、基底状態と第一励起状態間の平均エネルギーギャップと正の相関を持つ。(c) N_s およびエネルギーギャップ大なる構造は、全コンパクト構造の中、高々 0.12% であった。

この事実は、コンパクト構造のうち一部の構造のみが折り畳み可能であり、また、折り畳み可能な構造はアミノ酸配列における突然変異にも比較的安定である、ことを示唆している。後者は、極めて類似な構造を持つ多数の異なるアミノ酸配列が見い出されている、という一般的事実と一致している。

5 考察

アミノ酸配列と蛋白質構造との適合性を計るメジャーとして、また立体構造エネルギーの近似として、蛋白質構造から抽出した経験的残基間ポテンシャルについて簡単に考察した。残基間の排除体積効果のための斥力ポテンシャル、2次構造形成に与る2次構造ポテンシャルも、既知の蛋白質構造における頻度分布から経験的に求めることができる。ここでは紹介できなかったが、その種の統計分布から評価する経験的ポテンシャルとは異なる、既知の蛋白質配列に関して、その立体構造を最低エネルギー状態とするだけでなく、励起状態間のエネルギーギャップも最大にするような最適化ポテンシャルが提案されている [23, 26]。他の蛋白質構造との比較において、自己の蛋白質構造を最低エネルギー状態とするような最適化接触ポテンシャル [27, 28] は、他に比較して意味ある成功を治めなかった。エネルギーギャップを最大にするようなこの新しい最適化ポテンシャルが、どの程度立体構造エネルギーを反映するものか興味深い。

アミノ酸配列および蛋白質構造データの蓄積に伴い、配列と構造との適合性判定のための判定関数が、実用的側面から必要とされている。しかし、判定関数として最適なものは立体構造エネルギーを正しく反映したものであろう。その故、配列と構造との適合性の研究は、実用的側面だけでなく、蛋白質構造における相互作用を考える上でも意味がある。今後の発展に期待したい。

謝辞

本文を書くにあたって助言いただいた早稲田大学輪湖博士および査読者の方に感謝いたします。

参考文献

- [1] C. Chothia: *Nature* **357** (1992) 543.
- [2] C. A. Orengo, D. T. Jones and J. M. Thornton: *Nature* **372** (1994) 631.
- [3] J. Novotny, A.A. Rashin and R. E. Bruccoleri: *Proteins: Struct. Funct. Genet.* **4** (1988) 19.
- [4] S. Tanaka and H. A. Scheraga: *Macromolecules* **9** (1976) 945.
- [5] N. Go: *Annu. Rev. Biophys. Bioeng.* **12** (1983) 183.
- [6] S. Miyazawa and R. L. Jernigan: *Macromolecules* **18** (1985) 534.
- [7] S. Miyazawa and R. L. Jernigan: *J. Mol. Biol.* **256** (1996) 623.
- [8] K. A. Sharp, A. Nicholls, R. Friedman and B. Honig: *Biochemistry* **30** (1991) 9686.
- [9] Y. Nozaki and C. Tanford: *J. Biol. Chem.* **246** (1971) 2211.
- [10] K. Yutani, K. Ogasahara, T. Tsujita and Y. Sugino: *Proc. Natl. Acad. Sci. USA* **84** (1987) 4441.
- [11] S. Miyazawa and R. L. Jernigan: *Prot. Eng.* **7** (1994) 1209.
- [12] A. J. Doig and M. J. E. Sternberg: *Prot. Sci.* **4** (1995) 2247.
- [13] M. J. Sippl: *J. Mol. Biol.* **213** (1990) 859.
- [14] M. Hendlich, P. Lackner, S. Weitckus, H. Floechner, R. Froschauer, K. Gottsbachner, G. Casari and M. J. Sippl: *J. Mol. Biol.* **216** (1990) 167.

- [15] J.-P. A. Kocher, M. J. Rooman and S. J. Wodak: *J. Mol. Biol.* **235** (1994) 1598.
- [16] D. T. Jones, W. R. Taylor and J. M. Thornton: *Nature* **358** (1992) 86.
- [17] K. Nishikawa and Y. Matsuo: *Prot. Eng.* **6** (1993) 811.
- [18] A. M. Gutin, A. Ya. Badretdinov and A. V. Finkelstein: *Mol. Biol. (Russia), Engl. Transl.* **26** (1992) 94.
- [19] A. V. Finkelstein, A. Ya. Badretdinov and A. M. Gutin: *Proteins* **23** (1995) 142.
- [20] M. J. Sippl: *J. Comp.-Aided Mol. Design* **7** (1993) 473.
- [21] P. D. Thomas and K. Dill: *J. Mol. Biol.* **257** (1996) 457.
- [22] H. Li, R. Helling, C. Tang and N. Wingreen: *Science* **273** (1996) 666.
- [23] L. A. Mirny and E. I. Shakhnovich: preprint (1996).
- [24] J. D. Bryngelson and P. G. Wolynes: *Proc. Natl. Acad. Sci. USA* **84** (1987) 7524.
- [25] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes: *Proteins* **21** (1995) 167.
- [26] K. K. Koretke, Z. Luthey-Schulten and P. G. Wolynes: *Prot. Sci.* **5** (1996) 1043.
- [27] G. M. Crippen: *Biochemistry.* **30** (1991) 4232.
- [28] V. N. Maiorov and G. M. Crippen: *J. Mol. Biol.* **227** (1992) 876.

非会員著者の紹介: 宮澤三造氏は1948年東京生まれ。1970年東京都立大物理卒。名古屋大大学院博士課程修了後、九州大学生物研究生、米国国立保健衛生研究所客員研究員、国立遺伝学研究所助教授を経て、現在、群馬大学工学部助教授。専門は構造生物学。

図説

図 1. 各種アミノ酸側鎖の転移自由エネルギー [8] とそれに相当する接触エネルギー [6]。

図 2. Tryptophan synthase α subunit Glu-49 の置換に伴う変性自由エネルギー変化 [10] と接触エネルギー変化 [11]。



