

附章1

発展的トピック

“Advanced topics”

<要旨>

最初に、2023年に発表された Arai & Arai による論文を紹介している。彼らは、『“天然接触相互作用”が2つの局所構造間にも働く』という、より拡張した仮定をモデルに導入して、実験結果をより正確に再現するという新たな成果を達成した。

ここでは、彼らの論文に準じて、タンパク質のフォールディング過程の統計力学モデル（“NILS モデル”）の拡張版を定式化しようと試みている。次の3つのトピックを取り上げている：『Topic I. “NILS モデル”による分配関数の“転送行列表示”』，『Topic II. “NILS モデル”を拡張した新たなモデルの分配関数』，『Topic III. i 番目のアミノ酸残基が天然状態にある確率』

これらの定式の具体的な数値計算は気軽に実行できる状況ではなくスーパーコンピュータが必要であり、若い研究者の新たな挑戦を期待していると述べている。



2023 年、衝撃的な論文が発表された！

2023 年、衝撃的な論文が Ooka & Arai によって発表された^{*)}。彼らは、タンパク質分子のフォールディング過程において、“天然接触相互作用”^{**)}が“局所構造内のみで働く”と仮定する簡単なモデル(“WSME モデル”，及び同等の“NILS モデル”^{***)})を拡張して、『“天然接触相互作用”が2つの局所構造間にも働く』と仮定する新たなモデル(“WSME-L モデル”)を提案し、そのモデルによる分配関数から計算した Φ 値が実験結果を見事に再現することを示した：

^{*)} Koji Ooka & Munehito Arai, Accurate prediction of protein folding mechanisms by simple structure-based statistical mechanical models, Nature communications, 14:6338, 1-17, 2023.

^{**)} 『タンパク質分子の天然構造において、互いに接触しているアミノ酸残基間の相互作用』をいう(“Go potential”とも呼ばれている)。

^{***)} 『Non-Interacting Local Structure model』の略。

タンパク質分子のフォールディング・メカニズムの本質をついた、しかも、より簡単なモデルを構築して得られた理論的結果が実験結果をある程度再現できれば、我々は、そのモデルによって現象の本質を理解したということになるが、もし実験結果が再現できないときには、最初の簡単なモデルを拡張した新たなモデルが必要になる(『タンパク質分子のフォールディングの統計力学』；第10章, p.153, Table1 参照)。

全ての“天然接触相互作用”を考慮する場合の分配関数を求めることは、現在のコンピュータの性能ではまだ困難であると思われるが、Ooka & Arai は、『“天然接触相互作用”が局所構造内だけでなく、2つの局所構造間にも働く』という限定的な、しかし当初の仮定よりもより拡張した仮定をモデルに導入して実験結果をより正確に再現した。このことは、タンパク質分子のフォールディング過程における“天然接触相互作用”の重要性を示唆しており、フォールディング理論研究の新たな展開を予感させる。

ここでは、Ooka & Arai による拡張されたモデルに準じて、我々が提案している“NILS モデル”(タンパク質のフォールディング過程の統計力学モデル)の拡張版を定式化しようと試みた。次の3つのトピックを取り上げる：

『Topic I. “NILS モデル”による分配関数の“転送行列表示”』

『Topic II. “NILS モデル”を拡張した新たなモデルの分配関数』

『Topic III. i 番目のアミノ酸残基が天然状態にある確率』

彼らは新たなモデルによる定式の数値計算を実行しているが、次のように言及している：『60 残基程度のタンパク質は通常の PC では十数秒で済むが、200 残基を超えるとスーパーコンピュータが必要で、その計算時間が1時間を超える。』

我々の“NILS model”の拡張版の定式の具体的な数値計算も決して気軽に実行できる状況ではない。この分野に興味をもつ若い研究者が、このトピックの内容を参考にして、スーパーコンピュータを用いて、新たに挑戦してくれることを切に願っている。

Topic I . “NILS モデル”による分配関数の“転送行列表示”

＜要旨＞ “NILS モデル” による漸化式を転送行列で表し，最終的な分配関数を“転送行列表示”に変換する．

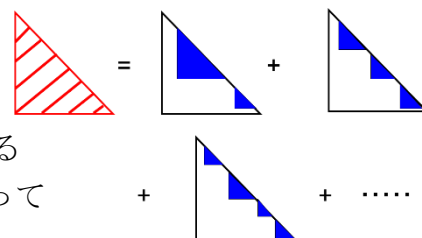


“NILS モデル” とは？

“NILS モデル”の仮定（『タンパク質分子のフォールディングの統計力学』；第4章参照）：

- 3次元格子に単一の球(ユニット)で表された N 残基からなるタンパク質を考慮する．
- タンパク質には唯一の特異的な立体構造(“天然構造”)が規定されているものとする．
- **局所構造の定義**：連続した残基が部分的に天然構造と同一の立体構造をとり，その領域の両端の残基が非天然状態にあるとき，その連続した領域を“局所構造”と呼ぶ．従って，局所構造内では“天然接触相互作用”のみが働いていることになる．ここで，“天然接触相互作用”とは，タンパク質分子の天然構造において，2つのアミノ酸残基が隣り合った格子点を占有するときに両者の間に発生する相互作用をいう．両者のアミノ酸残基を天然接触ペアと呼ぶ． i 番目から j 番目の残基からなる局所構造を $[i, j]$ と表記する．また，局所構造には“鎖エントロピーの損失”が発生する．
- 局所構造の最小単位：連続する4残基を最小の局所構造とする．
- **天然状態の定義**：($i-1$) 番目から ($i+2$) 番目までの残基が局所構造とっているとき， i 番目の残基が天然状態にあるとする．従って，アミノ酸残基の状態を考慮する範囲は， $i=2, 3, \dots, N-2$ であり， $(N-3)$ 個のアミノ酸残基について天然状態，及び非天然状態を考慮することになる．
- “NILS モデル”の仮定と分配関数：

『局所構造間，局所構造とランダム・コイル領域間，及び，ランダム・コイル領域間には全ての相互作用は働かない』と仮定する．従って，“NILS モデル”による分配関数は，イラストのように，ポリペプチド鎖に沿って形成される，あらゆる可能な局所構造の組み合わせを考慮すれば求められる．



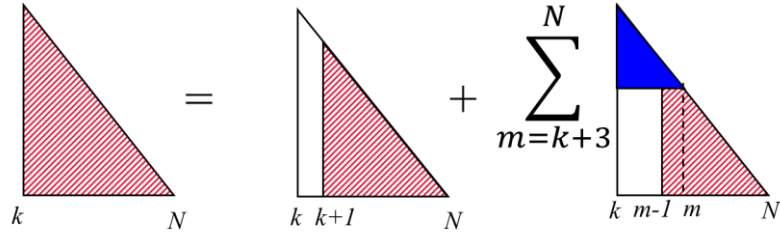
“NILS モデル”による分配関数を求めるための漸化式

“NILS モデル”による分配関数を求めるための漸化式は次式で書ける：

$$Z_{k,N} = Z_{k+1,N} + \sum_{m=k+3}^N f(k,m)^{-1} \exp\{-\beta E(k,m)\} Z_{m-1,N} = \sum_{m=k+2}^N w_{k,m} Z_{m-1,N}$$

$$(k = N-3, N-4, \dots, 3, 2, 1)$$

上式の漸化式は、下図のイラストで表示できる：



ここで、

$Z_{k,N}$: $(k \sim N)$ 残基からなる仮想的なタンパク質分子に対する補助的分配関数 (auxiliary partition function) を表す.

$Z_{N-1,N} = Z_{N-2,N} = 1$ とする.

$w_{k,m} \equiv f(k,m)^{-1} \exp\{-\beta E(k,m)\}$ (統計重率を表す.)

ただし、 $w_{k,k+2} = 1$ ($\because f(k,k+2) = 1, E(k,k+2) = 0$)



漸化式を転送行列に変換する

まず、具体的な k の値で転送行列 (the transfer matrix) 表示を試みる：

・ $k = N - 3$ のとき

$$Z_{N-3,N} = \sum_{m=N-1}^N w_{N-3,m} Z_{m-1,N} = w_{N-3,N-1} Z_{N-2,N} + w_{N-3,N} Z_{N-1,N}$$

上式を転送行列で表示する：

$$\begin{pmatrix} Z_{N-3,N} \\ Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix} = \begin{pmatrix} w_{N-3,N-1} & w_{N-3,N} \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix} = Q_3^2 \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix}$$

3 次元 3×2 次元 2 次元

ただし、

$$Q_3^2 \equiv \begin{pmatrix} w_{N-3,N-1} & w_{N-3,N} \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

・ $k = N - 4$ のとき

$$Z_{N-4,N} = \sum_{m=N-2}^N w_{N-4,m} Z_{m-1,N} = w_{N-4,N-2} Z_{N-3,N} + w_{N-4,N-1} Z_{N-2,N} + w_{N-4,N} Z_{N-1,N}$$

上式を転送行列で表示する：

$$\begin{pmatrix} Z_{N-4,N} \\ Z_{N-3,N} \\ Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix} = \begin{pmatrix} w_{N-4,N-2} & w_{N-4,N-1} & w_{N-4,N} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Z_{N-3,N} \\ Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix} = Q_4^3 \begin{pmatrix} Z_{N-3,N} \\ Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix} = Q_4^3 Q_3^2 \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix}$$

4 次元 4×3 次元 3 次元

ただし,

$$Q_4^3 \equiv \begin{pmatrix} w_{N-4,N-2} & w_{N-4,N-1} & w_{N-4,N} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\vdots$$

• $k = k$ のとき

$$Z_{k,N} = \sum_{m=k+2}^N w_{k,m} Z_{m-1,N} = w_{k,k+2} Z_{k+1,N} + w_{k,k+3} Z_{k+2,N} + \cdots + w_{k,N} Z_{N-1,N}$$

$$\begin{pmatrix} Z_{k,N} \\ Z_{k+1,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = \begin{pmatrix} w_{k,k+2} & w_{k,k+3} & \cdots & w_{k,N} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & & & 1 \end{pmatrix} \begin{pmatrix} Z_{k+1,N} \\ Z_{k+2,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = Q_{N-k}^{N-k-1} \begin{pmatrix} Z_{k+1,N} \\ Z_{k+2,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix}$$

(N-k) 次元 (N-k)×(N-k-1) 次元 (N-k-1) 次元

ただし,

$$Q_{N-k}^{N-k-1} = \begin{pmatrix} w_{k,k+2} & w_{k,k+3} & \cdots & w_{k,N} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & & \cdots & 1 \end{pmatrix}$$

Q_{N-k}^{N-k-1} は (N-k-1) 次元の列行列を (N-k) 次元の列行列に転送する行列である.

上式は, 次のように表示できる:

$$\begin{pmatrix} Z_{k,N} \\ Z_{k+1,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = Q_{N-k}^{N-k-1} \begin{pmatrix} Z_{k+1,N} \\ Z_{k+2,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = Q_{N-k}^{N-k-1} Q_{N-k-1}^{N-k-2} \begin{pmatrix} Z_{k+2,N} \\ Z_{k+3,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = \cdots$$

$$= Q_{N-k}^{N-k-1} Q_{N-k-1}^{N-k-2} \cdots Q_4^3 Q_3^2 \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix}$$

$$\therefore \begin{pmatrix} Z_{k,N} \\ Z_{k+1,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = \prod_{j=k}^{N-3} Q_{N-j}^{N-j-1} \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix}$$



“NILS モデル” による分配関数の “転送行列表示”

上式において、 $k=1$ のとき分配関数を転送行列で表示する：

$$\begin{aligned}
 Z_{1,N} &= \sum_{m=3}^N w_{1,m} Z_{m-1,N} = w_{1,3} Z_{2,N} + w_{1,4} Z_{3,N} + \cdots + w_{1,N} Z_{N-1,N} \\
 \begin{pmatrix} Z_{1,N} \\ Z_{2,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} &= \begin{pmatrix} w_{1,3} & w_{1,4} & \cdots & w_{1,N} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & 1 \end{pmatrix} \begin{pmatrix} Z_{2,N} \\ Z_{3,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = Q_{N-1}^{N-2} \begin{pmatrix} Z_{2,N} \\ Z_{3,N} \\ \vdots \\ Z_{N-1,N} \end{pmatrix} = \cdots \\
 &= Q_{N-1}^{N-2} Q_{N-2}^{N-3} \cdots Q_4^3 Q_3^2 \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix}
 \end{aligned}$$

◎ 最終的な分配関数は、次のように転送行列で表示できる：

$$Z_{1,N} = (1, 0, \dots, 0) \prod_{k=1}^{N-3} Q_{N-k}^{N-k-1} \begin{pmatrix} Z_{N-2,N} \\ Z_{N-1,N} \end{pmatrix}$$

ただし、

$$Q_{N-k}^{N-k-1} = \begin{pmatrix} w_{k,k+2} & w_{k,k+3} & \cdots & w_{k,N} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & & \cdots & 1 \end{pmatrix}$$



“Red House in Winter”

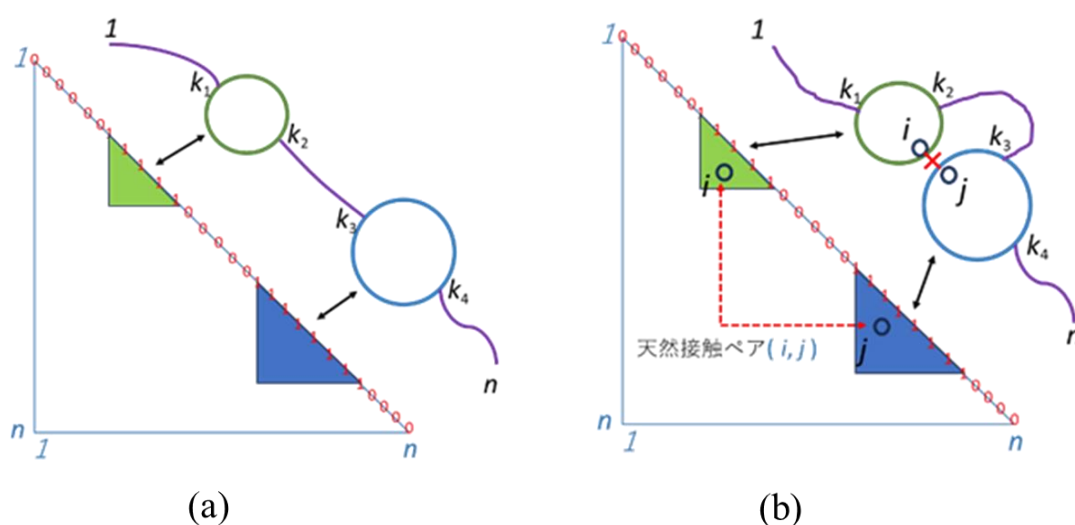
Topic II. “NILS モデル”を拡張した新たなモデルの分配関数

<要旨> 『“天然接触相互作用”が2つの局所構造間に働く』という新たなモデルにより分配関数を求める方法を定式化する.



新たなモデルとは？

新たなモデルは、タンパク質分子のフォールディング過程において、次の2つのケースの両方を考慮したモデルである：(a) 局所構造間、局所構造とランダム・コイル領域間、及び、ランダム・コイル領域間の全ての相互作用を無視する場合（図 a），(b) 任意の2個の局所構造間において“天然接触相互作用”のみが働いている場合（図 b）.



この両方を含む系の分配関数 Ξ は、次の2つの分配関数の和である：

$$\Xi = Q_{NILS} + Q_{I-2LS}$$

ただし、

Q_{NILS} : “NILS モデル” (Non-Interacting Local Structure model) による分配関数：

『局所構造間、局所構造とランダム・コイル領域間、及び、ランダム・コイル領域間の全ての相互作用を無視する』と仮定した場合の分配関数

Q_{I-2LS} : “I-2LS モデル” (Interacting-2 Local Structures model) による分配関数：

『任意の2個の局所構造間に“天然接触相互作用”のみが働く』と仮定した場合の分配関数

まず，“NILS モデル”による分配関数 Q_{NILS} を定式化する．



分配関数 Q_{NILS} の定式化

(1) “NILS モデル”による分配関数 Q_{NILS} を求めるための漸化式は次式で表される：

$$Z_{j,N} = Z_{j+1,N} + \sum_{m=j+3}^N f(j,m)^{-1} \exp\{-\beta E(j,m)\} Z_{m-1,N} = \sum_{m=j+2}^N w_{j,m} Z_{m-1,N}$$

$$(j = N-3, N-4, \dots, 3, 2, 1)$$

ここで，

$$Z_{N-1,N} = Z_{N-2,N} = 1$$

$$w_{j,m} \equiv f(j,m)^{-1} \exp\{-\beta E(j,m)\}$$

ただし， $w_{j,j+2} = 1$ （ $\because f(j,j+2) = 1, E(j,j+2) = 0$ ）

(2) 漸化式に表れる統計重率 $w_{j,m}$ を次のように置き換える：

$$w_{j,m} = f(j,m)^{-1} \exp\{-\beta E(j,m)\} = e^{-\{E(j,m) + T \ln f(j,m)\}/T}$$

ただし，無次元量： $\frac{kT}{\epsilon_0} = T^*$ より， T^* を改めて T と置き換えている．

<コメント>

統計重率に表れる $f(j,m)$ は， j 番目のアミノ酸残基から m 番目のアミノ酸残基までのポリペプチド鎖のセグメントが局所構造を形成したときのエントロピー損失を表している．この量を正確に見積ることは難しい．一つの例として，ある長さのポリペプチド鎖が，3 次元格子で取りうる全ての立体構造の数を（排除体積効果を考慮して）具体的に計算し，その値からの概算式（ $f(j,m) = 1.4048 \times (4.750)^{m-j-2}$ ）を採用する方法がある（『タンパク質分子のフォールディングの統計力学』；第 4 章，p.44 参照）．

(3) 次のような 2 つの変数 λ, ξ を導入する：

λ ：天然状態にある残基数を数えるための変数（最終的に $\lambda = 1$ ）

ξ ：局所構造 $[j,m]$ に固有の量（整数化）を加算していくための変数

これらの 2 つの変数 λ, ξ を用いると，統計重率 $w_{j,m}$ は次のように書ける：

$$w_{j,m} = \lambda^{m-j-2} \xi^{h(j,m)}$$

ここで，

$$\xi = e^{-1/T}, \quad h(j,m) = \{E(j,m) + T \ln f(j,m)\} \times 100$$

<コメント>

- ・ j 番目のアミノ酸残基から m 番目のアミノ酸残基までのポリペプチド鎖のセグメントで形成された局所構造を $[j, m]$ と表現する.
- ・ λ のべき乗：局所構造における“天然状態にある残基数(η)”に対応
- ・ ξ のべき乗 $h(j, m)$ ：局所構造 $[j, m]$ に固有の量(整数化するため 100 を乗じておく)
- ・ T : $T = T_m$ (転移温度)
- ・ $E(i, j)$: 局所構造 $[i, j]$ 内のエネルギー値
- ・ $f(j, m) = 1.4048 \times (4.750)^{m-j-2}$

(4) 上式の漸化式は次のように書ける：

$$Z_{j,N} = \sum_{m=j+2}^N \lambda^{m-j-2} \xi^{h(j,m)} Z_{m-1,N}$$

<コメント>

- コンピュータのメモリーに作業用領域を確保し、その領域を利用して、漸化式に従って順次、分配関数の一部の出力・処理・入力を繰り返す.
- 漸化式を変数 λ, ξ の、それぞれの“べき乗”で分配関数を整理して最終的な分配関数を求める手順は次の通りである(『附章 2“NILS モデルの漸化式より分配関数を求める！—Fortran プログラムと解説—』参照)：
- ① まず、 $j = N - 2$ のとき ($Z_{N-2,N}$) の処理を行う(“作業メモリー”に出力する)：

$$Z_{N-2,N} = \lambda^0 \xi^0$$

- ② 次に、 $j = N - 3$ のとき ($Z_{N-3,N}$) の処理を行う(“作業メモリー”に出力する)：

$$Z_{N-3,N} = \lambda^0 \xi^0 + \lambda \xi^{h(N-3,N)}$$

- ③ 漸化式を次のように書き換える：

$$Z_{j,N} = \lambda^{N-j-2} \xi^{h(j,N)} + \sum_{jj=j+2}^{N-2} \lambda^{jj-j-1} \xi^{h(j,jj+1)} Z_{jj,N} + \lambda^0 \xi^0 Z_{j+1,N}$$

この漸化式に従って、順次 ($j = N - 4, N - 5, \dots, 3, 2, 1$)、出力・処理・入力を繰り返す. 最終的に $Z_{1,N}$ が求める分配関数 Q_{NILS} である.

(5) 最終的に、 λ, ξ の“べき乗”で整理された分配関数 (Q_{NILS}) は次式で表される：

$$Q_{NILS} = \sum_{\eta} \sum_h \Omega(\eta, h) \lambda^{\eta} \xi^h$$

ただし,

η : 天然状態にあるアミノ酸残基数 $\eta (0 \leq \eta \leq N-3)$

λ のべき乗 (最終的には, $\lambda = 1$ とする) は, “天然状態にある残基数(η)” と次のように対応関係にある; $\lambda^0, \lambda^1, \lambda^2, \dots, \lambda^{N-3} \leftrightarrow \eta = 0, 1, 2, \dots, N-3$

$\Omega(\eta, h)$: 与えられた η, h を満たす全てのコンフォメーション数

(6) 分配関数は次式のようにも書ける:

$$Q_{NLS} = \sum_{\eta} W_{NLS}(\eta) \lambda^{\eta}$$

ここで, 天然状態にある残基数が η である状態の統計重率の和である $W_{NLS}(\eta)$ は次式のように書ける:

$$W_{NLS}(\eta) = \sum_h \Omega(\eta, h) \xi^h$$

これより, 天然状態にある残基数の期待値 $\langle \eta \rangle$ は, 次式のように計算できる:

$$\langle \eta \rangle = (Q_{NLS})^{-1} \sum_{\eta=0}^{n-3} \eta W_{NLS}(\eta)$$

(7) 天然状態にある残基数が η であるときの自由エネルギーは次式で与えられる:

$$F(\eta) = -k T \ln W_{NLS}(\eta)$$

次に, “I-2LS モデル”による分配関数 Q_{I-2LS} を定式化する.



分配関数 Q_{I-2LS} の定式化

(1) “I-2LS モデル”の仮定:

“I-2LS モデル”は, 図のように,

任意の2個の局所構造 ($[u, v]$ と $[s, t]$) 間に

“天然接触相互作用”のみが働くと仮定する

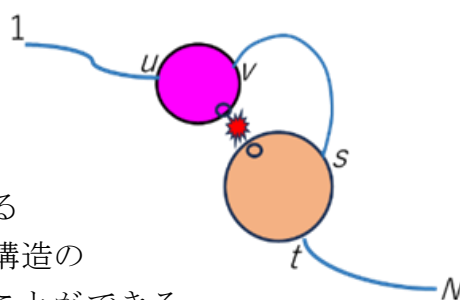
モデルであり, あらゆる可能な2個の局所構造の

組み合わせを考慮すれば分配関数を求めることができる.

任意の2個の局所構造間に “天然接触相互作用” が

存在しない場合は, すでに “NLS モデル” によって分配関数が求められているので

除外する必要がある. さらに, 3個以上の局所構造が同時に形成されているときの局所構造間の “天然接触相互作用” も考慮しない.



(2) 次のようなエネルギー値を仮定する：

$E(i, j)$: 局所構造 $[i, j]$ 内のエネルギー値

$E(u, v; s, t)$: 2 個の局所構造 ($[u, v]$ と $[s, t]$) 間の相互作用のエネルギー値

(3) < 統計重率の定義 >：

- $w_{i,j} \equiv f(i, j)^{-1} \exp\{-\beta E(i, j)\}$

$f(i, j)$: 局所構造 $[i, j]$ を形成することによる鎖エントロピー損失

(例として, $f(i, j) = 1.4048 \times (4.750)^{j-i-2}$ とおく)

- $g_{i,j} \equiv \sigma(i, j)^{-1} \exp\{-\beta E(i, j)\}$

$\sigma(i, j)$: 相互作用している 2 個の局所構造 $[u, v]$ と $[s, t]$ があり, それらを連結する領域で新たに形成される局所構造 $[i, j]$ が失う鎖エントロピー損失
(例として, $\sigma(i, j) = \alpha f(i, j)$ とおく. ただし, α : scaling factor)

- $w_{[u,v;s,t]} \equiv \exp\{-\beta E(u, v; s, t)\}$

(4) < 4 つの変数 (u, v, s, t) の条件 >：

局所構造の定義より

$$u \geq 1, u + 3 \leq v, v - 1 \leq s, s + 3 \leq t \leq N$$

上の条件は以下のように書くこともできる：

$$4 \leq u + 3 \leq v \leq s + 1 \leq t - 2 \leq N - 2$$

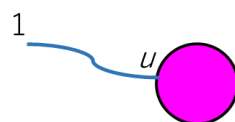
< コメント >

- 4 個のアミノ酸残基で形成される局所構造が, 局所構造の最小単位である.
- 条件 $v - 1 \leq s$: 「天然状態の定義」(3 次元格子模型において) より, 局所構造 $[u, v]$ と局所構造 $[s, t]$ の内部で天然状態にある残基は, それぞれ $(u + 1) \sim (v - 2)$ の範囲の残基, 及び, $(s + 1) \sim (t - 2)$ の範囲の残基であり, 残基 $u, (v - 1), v$ と残基 $s, (t - 1), t$ は非天然状態である. それ故, 2 つの局所構造が, 残基 $(v - 1), v$ と残基 s を共通の残基として重なることができることを表している.
- 4 つの変数 u, v, s, t の個々の範囲は次の通りである；

$$1 \leq u \leq N - 5, \quad u + 3 \leq v \leq N - 2, \quad v - 1 \leq s \leq N - 3, \quad s + 3 \leq t \leq N$$

(5) 次のような 3 種類の領域に対する分配関数を定式化する：

(a) N-末端側のポリペプチド鎖領域
の分配関数の定式化：



(a) N-末端側のポリペプチド鎖の領域

$$Z_{1,u+1} = \sum_{m=3}^{u+1} w_{1,m} Z_{m-1,u+1} = w_{1,3} Z_{2,u+1} + w_{1,4} Z_{3,u+1} + \cdots + w_{1,u+1} Z_{u,u+1}$$

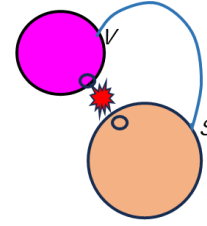
ただし,

$$Z_{u-1,u+1} = Z_{u,u+1} = 1, \quad w_{1,3} = 1$$

- (b) 相互作用している 2 個の局所構造を連結している
ポリペプチド鎖領域の分配関数の定式化:

$$Z_{v-1,s+1} = \sum_{m=v+1}^{s+1} g_{v-1,m} Z_{m-1,s+1}$$

$$= g_{v-1,v+1} Z_{v,s+1} + g_{v-1,v+2} Z_{v+1,s+1} + \cdots + g_{v-1,s+1} Z_{s,s+1}$$



(b) 相互作用している2個の
局所構造を連結している
ポリペプチド鎖の領域

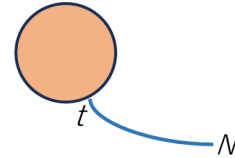
ただし,

$$Z_{s-1,s+1} = Z_{s,s+1} = 1, \quad g_{v-1,v+1} = 1$$

- (c) C-末端側のポリペプチド鎖領域の
分配関数の定式化:

$$Z_{t-1,N} = \sum_{m=t+1}^N w_{t-1,m} Z_{m-1,N}$$

$$= w_{t-1,t+1} Z_{t,N} + w_{t-1,t+2} Z_{t+1,N} + \cdots + w_{t-1,N} Z_{N-1,N}$$



(c) C-末端側のポリ
ペプチド鎖の領域

ただし,

$$Z_{N-2,N} = Z_{N-1,N} = 1, \quad w_{t-1,t+1} = 1$$

- (6) “I-2LS モデル” による分配関数 Q_{I-2LS} の定式化

“I-2LS モデル” による分配関数 Q_{I-2LS} は次のように表示できる:

$$Q_{I-2LS} = \sum_{u=1}^{N-5} \sum_{v=u+3}^{N-2} \sum_{s=v-1}^{N-3} \sum_{t=s+3}^N \delta(u, v; s, t) \{ (Z_{1,u+1}) \cdot w_{u,v} \cdot w_{[u,v;s,t]} \cdot (Z_{v-1,s+1}) \cdot w_{s,t} \cdot (Z_{t-1,N}) \}$$

ただし,

$$\begin{aligned} \delta(u, v; s, t) &= 1: 2 \text{ 個の局所構造 } [u, v] \text{ と } [s, t] \text{ 間に “天然接触相互作用” がある場合} \\ &= 0: 2 \text{ 個の局所構造 } [u, v] \text{ と } [s, t] \text{ 間に “天然接触相互作用” が無い場合} \end{aligned}$$

分配関数 Q_{I-2LS} の $\{ \}$ の各項は 2 つの変数 λ, ξ を用いて, それぞれ次ように書ける:

$$Z_{1,u+1} = \sum_{m=3}^{u+1} w_{1,m} Z_{m-1,u+1} = \sum_{m=3}^{u+1} \lambda^{m-3} \xi^{h(1,m)} Z_{m-1,u+1}$$

$$w_{u,v} = \lambda^{v-u-2} \xi^{h(u,v)}$$

$$w_{[u,v;s,t]} = \xi^{h_{2E}[u,v;s,t]}$$

ここで,

$$\xi = e^{-1/T}, \quad h_{2E}[u, v; s, t] = E(u, v; s, t) \times 100$$

$$Z_{v-1, s+1} = \sum_{m=v+1}^{s+1} g_{v-1, m} Z_{m-1, s+1} = \sum_{m=v+1}^{s+1} \lambda^{m-v-1} \xi^{h_g(v-1, m)} Z_{m-1, s+1}$$

ここで,

$$\xi = e^{-1/T}, \quad h_g(j, m) = \{E(j, m) + T \ln g(j, m)\} \times 100$$

$h_g(j, m)$: 2 個の局所構造が相互作用しているとき、それらに挟まれたポリペプチド鎖において形成される局所構造 $[j, m]$ に固有の量 (整数化するために 100 を乗じておく)

$$w_{s, t} = \lambda^{t-s-2} \xi^{h(s, t)}$$

$$Z_{t-1, N} = \sum_{m=t+1}^N w_{t-1, m} Z_{m-1, N} = \sum_{m=t+1}^N \lambda^{m-t-1} \xi^{h(t-1, m)} Z_{m-1, N}$$

(7) “I-2LS モデル” による分配関数を求める手順

- ① まず, $Z_{t-1, N}$ を 2 つの変数 λ, ξ を用いて表す.
初期値: $Z_{N-2, N} = Z_{N-1, N} = 1$ に対して漸化式を実行する.
- ② $w_{s, t}$ を, 2 つの変数 λ, ξ を用いて表し, $w_{s, t} \cdot (Z_{t-1, N})$ を実行する. その結果を X_0 とする.
- ③ $Z_{v-1, s+1}$ を 2 つの変数 λ, ξ を用いて表し, 初期値: $Z_{s-1, s+1} = Z_{s, s+1} = X_0$ ($\equiv w_{s, t} \cdot (Z_{t-1, N})$) に対して $Z_{v-1, s+1}$ の漸化式を実行する.
- ④ $w_{[u, v; s, t]}$ を, 変数 ξ を用いて表し, $w_{[u, v; s, t]} \cdot (Z_{v-1, s+1}) \cdot w_{s, t} \cdot (Z_{t-1, N})$ を実行する.
- ⑤ $w_{u, v}$ を, 2 つの変数 λ, ξ を用いて表し, $w_{u, v} \cdot w_{[u, v; s, t]} \cdot (Z_{v-1, s+1}) \cdot w_{s, t} \cdot (Z_{t-1, N})$ を実行する. その結果を Y_0 とする.
- ⑥ $Z_{1, u+1}$ を, 2 つの変数 λ, ξ を用いて表し, 初期値: $Z_{s-1, s+1} = Z_{s, s+1} = Y_0$ ($\equiv w_{u, v} \cdot w_{[u, v; s, t]} \cdot (Z_{v-1, s+1}) \cdot w_{s, t} \cdot (Z_{t-1, N})$) に対して $Z_{1, u+1}$ の漸化式を実行する.
- ⑦ 整数変数 (u, v, s, t) の有効な全ての値 (の組) に対して、①～⑥を実行する.

(8) 最終的に分配関数 (Q_{I-2LS}) は次式で表される:

$$Q_{I-2LS} = \sum_{\eta} \sum_h \Gamma(\eta, h) \lambda^{\eta} \xi^h$$

ただし,

η : 天然状態にあるアミノ酸残基数 $\eta (0 \leq \eta \leq N-3)$

$\Gamma(\eta, h)$: 与えられた η, h を満たす全てのコンフォメーション数

分配関数は次式のようにも書ける:

$$Q_{I-2LS} = \sum_{\eta} W_{I-2LS}(\eta) \lambda^{\eta}$$

ここで,

$W_{I-2LS}(\eta)$ (天然状態にある残基数が η である状態の統計重率の和) は次式で表される:

$$W_{I-2LS}(\eta) = \sum_h \Gamma(\eta, h) \xi^h$$

これより『天然状態にある残基数の期待値 ($\langle \eta \rangle$)』は、次式のように計算できる:

$$\langle \eta \rangle = (Q_{I-2LS})^{-1} \sum_{\eta=0}^{n-2} \eta W_{I-2LS}(\eta)$$

(9) 更に, 天然状態にある残基数が η であるときの自由エネルギーは次式で与えられる:

$$F(\eta) = -k T \ln W_{I-2LS}(\eta)$$

(10) 最終的に, 新たなモデルにおける『天然状態にある残基数の期待値($\langle \eta \rangle$)』を求める式は次式で与えられる:

$$\langle \eta \rangle = (\Xi)^{-1} \left\{ \sum_{\eta=0}^{n-3} \eta W_{NILS}(\eta) + \sum_{\eta=0}^{n-3} \eta W_{I-2LS}(\eta) \right\}$$



“Bird over River”

Topic Ⅲ. i 番目のアミノ酸残基が天然状態にある確率

<要旨> 『 i 番目のアミノ酸残基が天然状態にある確率』の計算方法を定式化する.



“ i 番目のアミノ酸残基が天然状態にある確率” : $p(i)$

i 番目のアミノ酸残基が“天然状態”にある確率 $p(i)$ は、次式で与えられる :

$$p(i) = \Xi^{-1} \{ Q_{NLS}^{(i-Native)}(i) + Q_{I-2LS}^{(i-Native)}(i) \}$$

$$(i = 2, 3, \dots, N-2)$$

ただし,

Ξ : 新たなモデルの分配関数

$Q_{NLS}^{(i-Native)}(i)$: “NILS モデル”において付加される, i 番目の残基が“天然状態”にある場合の分配関数

$Q_{I-2LS}^{(i-Native)}(i)$: “I-2LS モデル”において付加される, i 番目の残基が“天然状態”にある場合の分配関数

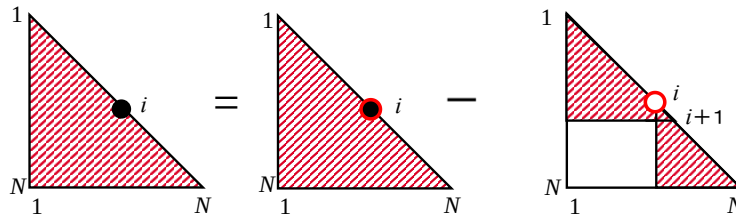
ここで, まず $Q_{NLS}^{(i-Native)}(i)$ の計算式を求めよう.

○ “NILS モデル”において付加される, i 番目の残基が“天然状態”にある場合の分配関数 $Q_{NLS}^{(i-Native)}(i)$ を求める :

下図は、“NILS モデル”において, i 番目の残基が“天然状態”にある場合の分配関数を求める方法をイラストで示している :

(i 番目の残基が“天然状態”にある場合の分配関数)

= (全体の分配関数) - (i 番目の残基が“非天然状態”にある場合の分配関数)



i 番目の残基が“非天然状態”にある場合の分配関数 $Q_{NLS}^{(i-Nonnative)}(i)$ は次式で表される :

$$Q_{NLS}^{(i-Nonnative)}(i) = Z_{1,i+1} \cdot Z_{i,N} = \left\{ \sum_{m=3}^{i+1} w_{1,m} Z_{m-1,i+1} \right\} \cdot \left\{ \sum_{m=i+2}^N w_{i,m} Z_{m-1,N} \right\}$$

故に、 i 番目の残基が“天然状態”にある場合の分配関数 $Q_{NILS}^{(i-Native)}(i)$ は次式で表される：

$$\begin{aligned}
 Q_{NILS}^{(i-Native)}(i) &= Q_{NILS} - Q_{NILS}^{(i-Nonnative)}(i) \\
 &= Q_{NILS} - \left\{ \sum_{m=3}^{i+1} w_{1,m} Z_{m-1,i+1} \right\} \cdot \left\{ \sum_{m=i+2}^N w_{i,m} Z_{m-1,N} \right\} \\
 Z_{1,i+1} &= \sum_{m=3}^{i+1} w_{1,m} Z_{m-1,i+1} = w_{1,3} Z_{2,i+1} + w_{1,4} Z_{3,i+1} + \cdots + w_{1,i+1} Z_{i,i+1} \\
 Z_{i,N} &= \sum_{m=i+2}^N w_{i,m} Z_{m-1,N} = w_{i,i+2} Z_{i+1,N} + w_{i,i+3} Z_{i+2,N} + \cdots + w_{i,N} Z_{N-1,N}
 \end{aligned}$$

次に、 $Q_{I-2LS}^{(i-Native)}(i)$ の計算式を求める。

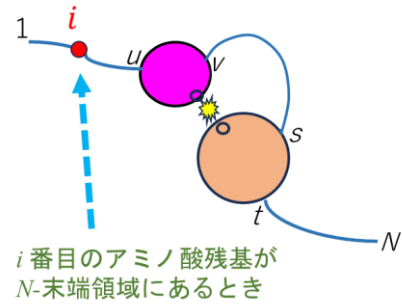
○ “I-2LS モデル” おいて付加される、 i 番目の残基が“天然状態”にある場合の分配関数 $Q_{I-2LS}^{(i-Native)}(i)$ を求める：

i 番目の残基が“天然状態”にある場合の分配関数 $Q_{I-2LS}^{(i-Native)}(i)$ は次式で表される：

$$Q_{I-2LS}^{(i-Native)}(i) = Q_{I-2LS} - Q_{I-2LS}^{(i-Nonnative)}(i)$$

ここで、 $Q_{I-2LS}^{(i-Nonnative)}(i)$ の計算式を求めるために、 i 番目のアミノ酸残基の部位を次の3通りの領域で考える（図参照）：

(a) $2 \leq i \leq u$ のとき
(i 番目のアミノ酸残基が N -末端領域であるとき)



$$Q_{I-2LS}^{(i-Nonnative)}(i)$$

$$= \sum_{u=1}^{N-5} \sum_{v=u+3}^{N-2} \sum_{s=v-1}^{N-3} \sum_{t=s+3}^N \delta(u, v; s, t) \{ (Z_{1,i+1}) \cdot (Z_{i,u+1}) \cdot w_{u,v} \cdot w_{[u,v;s,t]} \cdot (Z_{v-1,s+1}) \cdot w_{s,t} \cdot (Z_{t-1,N}) \}$$

ただし、

$\delta(u, v; s, t) = 1$: 2 個の局所構造 $[u, v]$ と $[s, t]$ 間に“天然接触相互作用”がある場合
 $= 0$: 2 個の局所構造 $[u, v]$ と $[s, t]$ 間に“天然接触相互作用”が無い場合

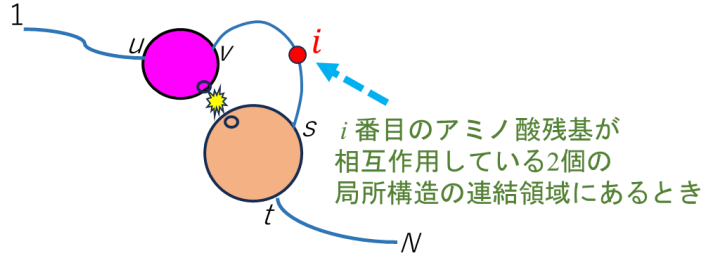
ここで、

$$Z_{1,i+1} = \sum_{m=3}^{i+1} w_{1,m} Z_{m-1,i+1} = w_{1,3} Z_{2,i+1} + w_{1,4} Z_{3,i+1} + \cdots + w_{1,i+1} Z_{i,i+1}$$

$$Z_{i,u+1} = \sum_{m=i+2}^{u+1} w_{i,m} Z_{m-1,u+1} = w_{i,i+2} Z_{i+1,u+1} + w_{i,i+3} Z_{i+2,u+1} + \cdots + w_{i,u+1} Z_{u,u+1}$$

(b) $v-1 \leq i \leq s$ のとき

(i 番目のアミノ酸残基が相互作用している 2 個の局所構造の連結領域にあるとき)



$$Q_{I-2LS}^{(i-Nonnative)}(i)$$

$$= \sum_{u=1}^{N-5} \sum_{v=u+3}^{N-2} \sum_{s=v-1}^{N-3} \sum_{t=s+3}^N \delta(u, v; s, t) \{ (Z_{1,u+1}) \cdot w_{u,v} \cdot w_{[u,v;s,t]} \cdot (Z_{v-1,i+1}) \cdot (Z_{i,s+1}) \cdot w_{s,t} \cdot (Z_{t-1,N}) \}$$

ただし,

$\delta(u, v; s, t) = 1$: 2 個の局所構造 $[u, v]$ と $[s, t]$ 間に“天然接触相互作用”がある場合

$= 0$: 2 個の局所構造 $[u, v]$ と $[s, t]$ 間に“天然接触相互作用”が無い場合

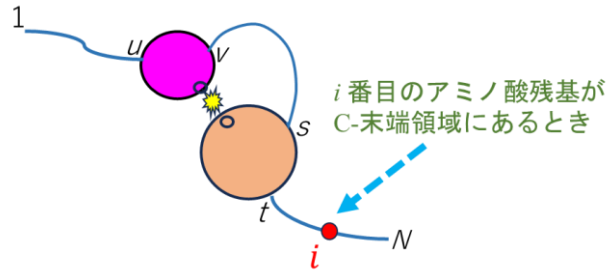
ここで,

$$Z_{v-1,i+1} = \sum_{m=v+1}^{i+1} g_{v-1,m} Z_{m-1,i+1} = g_{v-1,v+1} Z_{v,i+1} + g_{v-1,v+2} Z_{v-1,i+1} + \cdots + g_{v-1,i+1} Z_{i,i+1}$$

$$Z_{i,s+1} = \sum_{m=i+2}^{s+1} g_{i,m} Z_{m-1,s+1} = g_{i,i+2} Z_{i+1,s+1} + g_{i,i+3} Z_{i+2,s+1} + \cdots + g_{i,s+1} Z_{s,s+1}$$

(c) $t-1 \leq i \leq N-2$ のとき

(i 番目のアミノ酸残基が C-末端領域にあるとき)



$$Q_{I-2LS}^{(i-Nonnative)}(i)$$

$$= \sum_{u=1}^{N-5} \sum_{v=u+3}^{N-2} \sum_{s=v-1}^{N-3} \sum_{t=s+3}^N \delta(u, v; s, t) \{ (Z_{1,u+1}) \cdot w_{u,v} \cdot w_{[u,v;s,t]} \cdot (Z_{v-1,s+1}) \cdot w_{s,t} \cdot (Z_{t-1,i+1}) \cdot (Z_{i,N}) \}$$

ただし,

$\delta(u, v; s, t) = 1$: 2 個の局所構造 $[u, v]$ と $[s, t]$ 間に“天然接触相互作用”がある場合

$= 0$: 2 個の局所構造 $[u, v]$ と $[s, t]$ 間に“天然接触相互作用”が無い場合

ここで,

$$Z_{t-1,i+1} = \sum_{m=t+1}^{i+1} w_{t-1,m} Z_{m-1,i+1} = w_{t-1,t+1} Z_{t,i+1} + w_{t-1,t+2} Z_{t+1,i+1} + \cdots + w_{t-1,i+1} Z_{i,i+1}$$

$$Z_{i,N} = \sum_{m=i+2}^{u+1} w_{i,m} Z_{m-1,u+1} = w_{i,i+2} Z_{i+1,N} + w_{i,i+3} Z_{i+2,N} + \cdots + w_{i,N} Z_{N-1,N}$$

Memorandum