

総説

アミノ酸座位間における共進化に基づく残基間コンタクト予測: タンパク質立体構造予測にむけて

宮澤三造

Residue-residue interactions that fold a protein into a unique three-dimensional structure and make it play a specific function put structural and functional constraints in varying degrees on each residue site. Coevolution between closely-located sites caused by such selective constraints is recorded in amino acid orders of homologous sequences and also in the evolutionary trace of amino acid substitutions. A challenge for predicting residue contacts through coevolving site pairs is to extract direct dependences between sites by removing phylogenetic correlations and indirect dependences through other residues within a protein or even through other molecules. Recent attempts, particularly by detecting co-substitutions, are reviewed.

selective constraint on amino acids / coevolution between residue sites / inverse Potts problem / inverse covariance matrix / substitution mapping / partial correlation coefficients

1. はじめに

相同タンパク質の分子進化の履歴は、壮大な進化の時間スケールのもとでなされたアミノ酸置換の実験結果であり、進化生物学だけでなくタンパク質構造を理解する上でも貴重なデータである。アミノ酸置換突然変異に作用する構造/機能保存的淘汰圧は、アミノ酸鎖をユニークな立体構造に折り畳む残基間相互作用に起因する。構造/機能上有害な残基置換は引続き生じた残基置換により補償されたもの(補償置換)のみが遺伝子プールに固定され、(1)多様な相同配列(平衡分布)と(2)分子進化系統樹の各分枝に生じた残基置換の軌跡(動的過程)にアミノ酸座位間での共進化の痕跡として記録される。このような共進化は、主に立体構造において近接する残基間の相互作用に起因するから、残基間の共進化を検出することにより近接残基対(残基間コンタクト)を予測できると考えられる。

座位間での共進化は分子進化学の黎明期以来当然のこととされ、配列データの増加とともに共進化を裏付ける残基置換も多数報告されてきた。とは言え、残基置換の相関から立体構造における残基間コンタクトを十分な精度で予測するには、(a)統計的に十分なデータサイズが必要であった。しかし、ゲノム/メタゲノム計画の進展とともに、タンパク質ドメインデータベース Pfam には、数千の相同配列からなるファミ

リーも多数見出される現在、(b)見掛けの残基間相関から、(b-1)他残基を介しての間接的相関、(b-2)系統進化に起因する偏りを除去し、残基間相互作用に起因する直接的相関だけをいかに抽出するかが近年の課題であった(文献1,2を参照)。最近、予測した残基間コンタクトからの立体構造予測が可能なるほどに予測精度があがってきた¹⁾⁻⁷⁾。座位間共進化の推定は(1)置換過程の平衡分布である多重配列アライメントにおけるアミノ酸タイプの座位間相関¹⁾⁻⁹⁾、(2)置換の動的過程である分子進化系統樹の各分枝に生じた残基置換の座位間相関¹⁰⁾に基づく2方法に大別される(図1参照)。以下では、(2)の方法を中心に紹介する。

2. 多重配列アライメントにおける座位間でのアミノ酸タイプの相関に基づく共進化座位対の推定

配列長 L, N 配列からなる多重配列アライメントを $A \equiv (A_1, A_2, \dots, A_L)$ とする。ここで座位 i のアライメントベクターは $A_i \equiv (A_{i1}, A_{i2}, \dots, A_{iM})'$ と定義する。'は転置を意味する。欠失を含む $q (= 21)$ 種のアミノ酸を a_k で表す。($A_{ii} \in \{a_1, \dots, a_q\}$)。 A における座位 i における a_k の頻度は $f_i(a_k) = \sum_j \delta(A_{ij}, a_k) / N$ 、残基対の頻度は $f_{ij}(a_k, a_l) = \sum_j \delta(A_{ij}, a_k) \delta(A_{jl}, a_l) / N$ である。(δ は Kronecker 記号。) 観測値 $f_i(a_k), f_{ij}(a_k, a_l)$ を満たす配列集団における配列 $\sigma = (\sigma_1, \dots, \sigma_L)$ 、($\sigma_i \in \{a_1, \dots, a_q\}$) の確率 $P(\sigma)$ を考える。

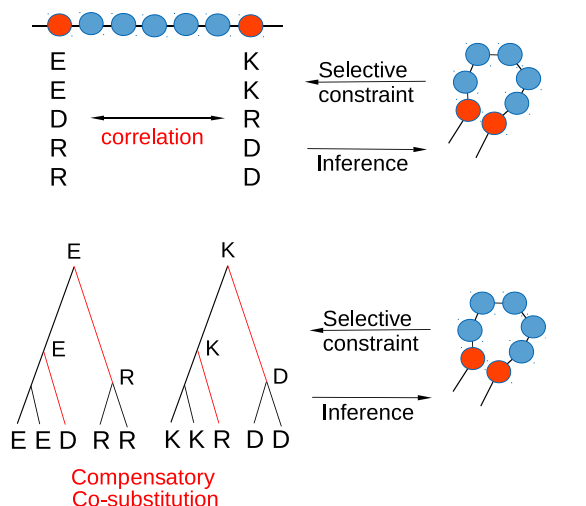


図 1 座位間共進化に基づく残基間コンタクト予測。上図は座位間でのアミノ酸タイプの相関、下図は共置換、補償置換に基づく残基間コンタクト予測を示す。

$$P(\sigma_i = a_k, \sigma_j = a_l) = \sum_{\sigma}^{(\sigma_i, \sigma_j) = (a_k, a_l)} P(\sigma) = f_{ij}(a_k, a_l) \quad (1)$$

$$P(\sigma_i = a_k) = \sum_{\sigma_j} P(\sigma_i = a_k, \sigma_j) = f_i(a_k) \quad (2)$$

エントロピー $(-\sum_{\sigma} P \log P)$ を最大にする $P(\sigma)$ は以下を満たす (最大エントロピー法)^{11), 12)}。

$$P(\sigma) = \frac{1}{Z} \exp \left(\sum_i \left(h_i(\sigma_i) + \sum_{j>i} J_{ij}(\sigma_i, \sigma_j) \right) \right) \quad (3)$$

ここで、 Z は分配関数、 $h_i(\sigma_i), J_{ij}(\sigma_i, \sigma_j)$ は、条件式 1 と 2 を満たすようなパラメータである。座位間での直接相関を表す $J_{ij}(\sigma_i, \sigma_j)$ は Direct coupling (DC) と名付けられた¹²⁾。式 3 において、式 1, 2 から $h_i(\sigma_i), J_{ij}(\sigma_i, \sigma_j)$ を推定する問題は、統計物理学で inverse Ising/Potts problem として、また神経回路網理論で Boltzmann machine として研究されてきた問題と同一でありその手法が応用できる^{9), 13)}。

$J_{ij}(\sigma_i, \sigma_j)$ の最も簡単な近似解法では、ベクトル $(\delta(A_{1i}, a_i), \dots, \delta(A_{Nj}, a_j))$ 間の共分散 $C_{ik, jl} \equiv P(\sigma_i = a_k, \sigma_j = a_l) - P(\sigma_i = a_k)P(\sigma_j = a_l)$ は線形応答より $C_{ik, jl} = \partial^2 \ln Z / \partial h_i(a_k) \partial h_j(a_l) = \partial P(\sigma_i = a_k) / \partial h_j(a_l)$ である。ゲージ不変性を考慮し、

$$J_{i,j}(a_k, a_l) = J_{i,j}(a_l, a_k) = h_i(a_k) = 0 \quad (4)$$

とし、式 3 を平均場近似により $P(\sigma_i = a_k) / P(\sigma_i = a_l) \sim \exp(h_i(a_k) + \sum_{l \neq q} \sum_j J_{ij}(a_k, a_l) P(\sigma_j = a_l))$ と J_{ij} の 1 次までで近似すれば、以下の近似解 (naive mean-field inversion (NMFI))¹³⁾ を得る。

$$J_{i,j}(a_k, a_l) = -(C^{-1})_{i,j} \quad (i \neq j, k, l \neq q) \quad (5)$$

($\zeta = (q-1)(i-1) + k, \eta = (q-1)(j-1) + l$). C^{-1} は共分散行列の $k, l = q$ の行と列を除去した $(q-1)L \times (q-1)L$ の部分行列の逆行列である。

これで座位間直接相関が求まったことになるが、実際の系では、逆行列が可能な (各座位で q 種すべてのアミノ酸を含む $(q-1)L$ 種類以上からなる) 配列アライメントを得ることはほぼ不可能である。そこで、(a) $f_i(a_k), f_{ij}(a_k, a_l)$ を pseudo counts により補正^{1), 2)}, (b) 共分散の shrinkage method による補正⁸⁾ 等が用いられた。なお、各座位のアミノ酸間相関 ($J_{ij}(a_k, a_l)$) から i, j のみに依存する座位間相関を評価するために、(a) $P_{ij}^D(a_k, a_l) \propto \exp[J_{ij}(a_k, a_l) + h_i'(a_k) + h_j'(a_l)]$ に基づく相互情報量 (Direct Information (DI))^{1), 2), 12)},

$$DI_{ij} \equiv \sum_k \sum_l P_{ij}^D(a_k, a_l) \log \left[P_{ij}^D(a_k, a_l) / (f_i(a_k) f_j(a_l)) \right] \quad (6)$$

(h' は、 (i, j) 毎に P_{ij}^D が式 2 を満たすよう定める。) (b) $J'_{ij}(a_k, a_l) \equiv J_{ij}(a_k, a_l) - J_{ij}(a_k, \bullet) - J_{ij}(\bullet, a_l) + J_{ij}(\bullet, \bullet)$ の l_1 ⁸⁾ ないしは l_2 ⁹⁾ ノルム (S_{ij}^N) に基づくスコア $S_{ij}^{CN} \equiv S_{ij}^N - S_{\bullet}^N S_{\bullet}^N / S_{\bullet}^N$ が用いられた。(•はその添字に関する平均を意味する。) なお、(b)の方が予測精度が向上するとの報告がある⁹⁾。

一方、Ekeberg et al.⁹⁾ は

$$-l(\mathbf{b}, \mathbf{J}) \equiv \frac{-1}{N} \sum_i \sum_j \log P(\sigma_i = A_{ij} | \forall j \neq i, \sigma_j = A_{ij}) \quad (7)$$

をペナルティ関数 (\mathbf{b} と \mathbf{J} の l_2 ノルムの 2 乗の 1 次結合) とともに最小化する方法 (pseudo-likelihood maximization) により \mathbf{b}, \mathbf{J} を評価し、NMFI より予測性能が高いと述べている。また、Taylor et al.³⁾ は相互情報量行列 ($MI_{ij} \equiv \sum_k \sum_l f_{ij}(a_k, a_l) \log [f_{ij}(a_k, a_l) / (f_i(a_k) f_j(a_l))]$) を共分散行列とみなしその逆行列から計算される偏相関係数をスコアとして用いた。なお、Bayesian Network に基づくモデルは予測性能が落ちる。

以上、直接相関を評価する方法について述べたが、系統進化に起因する偏りに関しては、いずれの方法においても、配列に関する平均を重み付け平均に置き換えることで対処された。

3. 座位間共置換に基づく共進化座位対の推定

筆者は、多重配列アライメントから最尤法により推定された分子系統樹を用い、各分枝に生じた残基置換の座位間相関を評価することにより、共進化座位対の

推定を試みた¹⁰⁾. この方法は、(a) 時間軸を考慮にされることで他座位を介しての間接的相関の寄与が小さく座位間での共進化が高精度で評価できる、(b) 配列間の系統進化を適切に考慮可能という利点を持つ.

各分枝における(座位間相関が共置換を示唆する)置換確率や(補償置換を示唆する側鎖の体積変化などの)置換に伴う特徴量を各置換の尤度をもって評価する. ここでは、アミノ酸置換モデルより赤池情報量規準(AIC)が小さいコドン置換置換モデル(Θ)¹⁰⁾を用い、(a) コドン置換は各座位で独立に生じ、(b) 各座位でのコドン置換は、置換率一定の time-reversible Markov 過程であり、(c) 配列は定常状態にあると仮定する. この時、系統樹 T 、多重配列アライメント A の尤度 $P(A|T, \Theta)$ は、各座位の尤度の積となる; $P(A|T, \Theta) = \prod_i P(A_i|T, \Theta)$. 分子系統樹 T における枝 b でコドン κ が λ へ置換する確率は、枝 b の左ノード $v_{bL} = \kappa$ (右ノード $v_{bR} = \lambda$) を含む左(右)部分系統樹の尤度 $P_{bL}(P_{bR})$ と、 κ の平衡頻度 f_κ 、枝長 t_b で κ から λ へ置換する確率 $P(\lambda|\kappa, t_b, \Theta)$ の積に等しい.

$$P(A_i, v_{bL} = \kappa, v_{bR} = \lambda | T, \Theta) = P_{bL}(A_i | v_{bL} = \kappa, T, \Theta) \times f_\kappa P(\lambda | \kappa, t_b, \Theta) P_{bR}(A_i | v_{bR} = \lambda, T, \Theta) \quad (8)$$

コドンの平衡頻度はアライメントにおけるアミノ酸頻度から縮退コドンは等頻度と仮定し評価する. よって、枝 b における κ から λ への置換に伴う変化量 $\Delta_{\kappa, \lambda}$ の期待値は、

$$\Delta_{ib}(A_i, \hat{T}, \hat{\Theta}) \equiv \sum_{\kappa, \lambda} \frac{\Delta_{\kappa, \lambda} P(A_i, v_{bL} = \kappa, v_{bR} = \lambda | \hat{T}, \hat{\Theta})}{P(A_i | \hat{T}, \hat{\Theta})} \quad (9)$$

$$P(A_i | T, \Theta) = \sum_{\kappa} \sum_{\lambda} P(A_i, v_{bL} = \kappa, v_{bR} = \lambda | T, \Theta) \quad (10)$$

となる. ここで、 $(\hat{T}, \hat{\Theta}) = \arg \max_{T, \Theta} P(A|T, \Theta)$ は、系統樹と置換モデル変数の最尤推定値である.

共進化、補償置換を検出するために $\Delta_{\kappa, \lambda}$ として以下の諸量を計算した. (a) アミノ酸置換確率 (s): $\Delta_{\kappa, \lambda}$ は同義置換には 0, 非同義置換には 1 と定義, (b) 置換に伴う側鎖の体積変化 (v), (c) 電荷変化量 (c), (d) 水素結合能 (hb): 各残基にはドナー (アクセプター) として可なら 1 (-1), 不可なら 0 を割り振り, ドナーおよびアクセプター能の変化量の和, (e) 置換に伴う側鎖の疎水度 (h) の変化量, (f) $\beta(b)$, (g) turn 形成能 (t) の変化量, (h) 残基の芳香族性 (ar), (i) 分枝性 (br), (j) 架橋能 (cl), (k) イオン性 (ion) の変化量 (詳細は文献 10 を参照).

共置換、補償置換を検出するために、置換ベクター

$\Delta_i = (\dots, \Delta_{ib}(A_i, \hat{T}, \hat{\Theta}) - \Delta_{i, \dots}, \dots)$ の座位間での相関係数 ($C_{ij} = \Delta_i \Delta_j / (\|\Delta_i\| \|\Delta_j\|)$) を考える. 残念ながらこの相関係数はほとんど全ての座位間で統計的に意味ある程大きな正値をとる. なぜなら、座位 i での置換率を μ_i とすると $\Delta_{ib} \sim 1 - \exp(-\mu_i \hat{t}_b) \sim \mu_i \Delta_{i, b}$ であり、枝長さ \hat{t}_b を介して強い正の相関を持つからである. 幸いなことに $\hat{t}_b \ll 1$ であれば、 Δ_{ib} は座位に関する平均 ($\Delta_{i, b}$) に比例する. よって、系統樹に起因するこの相関を除去すべく、 i, j 以外の座位における Δ_{kb} ($k \neq i, j$) へ線形多重回帰した際の残余ベクター $\prod_{\perp[\Delta_{k \neq i, j}]} \Delta_i$ と $\prod_{\perp[\Delta_{k \neq i, j}]} \Delta_j$ 間の相関係数 (C_{ij}) を考える. ($\prod_{\perp[\Delta_{k \neq i, j}]}$ は $\{\Delta_{k \neq i, j}\}$ に直交する空間への射影演算子である.) この相関係数は統計学では偏相関係数として知られている. 偏相関係数を用いることにより、 i, j 以外の座位を介した間接的な相関も(線形依存の範囲内で)除去することができる. なお、不変座位は対象から除外する.

$$C_{ij} = \frac{(C^{-1})_{ij}}{((C^{-1})_{ii}(C^{-1})_{jj})^{1/2}} \quad (11)$$

偏相関係数に基づき共進化座位対を以下の手順で推定した. (a) 座位間での(同じ枝で起きる)共置換は正の相関; $\rho_{ij}^s \equiv \max(C_{ij}^s, 0)$, (b) 他の諸量 (x) は共置換が前提であるから、便宜上、幾何平均をとる; $\rho_{ij}^x \equiv (\text{sgn } C_{ij}^x)(|\rho_{ij}^s C_{ij}^x|)^{1/2}$ (c) 体積、電荷、水素結合能に関しては、補償置換(負の相関)のみを考慮し、共進化スコア (coevolution score (CES)) を定義 (詳細は文献 10 を参照); $\rho_{ij} \equiv \max[\rho_{ij}^s, \max(-\rho_{ij}^v, 0), \max(-\rho_{ij}^c, 0), \max(-\rho_{ij}^{hb}, 0), |\rho_{ij}^h|, |\rho_{ij}^b|, |\rho_{ij}^t|, |\rho_{ij}^{ar}|, |\rho_{ij}^{br}|, \max(\rho_{ij}^d, 0), \max(\rho_{ij}^{ion}, 0)]$. ここで先に定義された 11 種の諸量 (a)-(j) をその順に ρ の上添字で示す. ρ_{ij} の値の大きい順に、定められた数の座位対を、共進化座位対として推定する.

4. 残基間コンタクトとしての予測精度

付録図 1 に Marks et al.²⁾ の Direct Coupling Analysis (DCA) 法 (ここでは手法に基づき NMFI-DI 法又は単に DI 法と記す) と共進化スコア (CES) 法¹⁰⁾ により予測された共進化座位対と残基間コンタクトを三角マップで示す. また、表 1 に正しく予測されたコンタクトの割合 (PPV) を、予測したタンパク質構造の PDB ID、用いた多重配列アライメントの Pfam ID とともに示す. 系統樹は、Pfam の FastTree による参照系統樹を真の系統樹とみなし、コドン置換モデルを用いて枝長を最尤推定. コンタクトの定義は、残基間での重原子の最小距離が 5 Å 内にある近接残基対でかつ 6 残基以上隔たった残基対 ($|i-j| > 5$) である. CES 法においては、

表 1
座位間共進化に基づくコンタクト予測.

Pfam ID/PDB ID	N	PPV ^{††}		PPV ^{††}	
		C_{ij}^{st}	C_{ij}^{st+}	ρ_{ij}^s	DI ^{§§}
Trans_reg_C/1ODD-A:156-232	7720	0.189	0.541	0.622	0.432
CH/1BKR-A:5-107	2960	0.053	0.439	0.491	0.439
7tm_1/1GZM-A:54-306	6302	0.008	0.290	0.306	0.169
SH3_1/2HDA-A:97-144	4160	0.241	0.621	0.655	0.552
Cadherin/2O72-A:113-212	7617	0.274	0.726	0.767	0.753
Trypsin/3TGI-E:16-238	6688	0.344	0.575	0.613	0.533
Kunitz_BPTI/5PTI-A:4-56	2130	0.216	0.514	0.486	0.541
KH_1/1WVN-A:7-69	5114	0.367	0.600	0.700	0.533
RRM_1/1G2E-A:41-111	7684	0.295	0.795	0.795	0.705
FKBP_C/1R9H-A:26-118	5695	0.197	0.667	0.727	0.697
Lectin_C/1SL5-A:273-379	4479	0.171	0.585	0.646	0.671
Thioredoxin/1RQM-A:1-103	7483	0.177	0.581	0.645	0.565
Response_reg/1E6K-A:8-121	7613	0.015	0.657	0.687	0.642
RNase_H/1F21-A:3-142	4782	0.132	0.407	0.407	0.549
Ras/5P21-A:5-165	6390	0.207	0.640	0.685	0.631

α , β , $\alpha + \beta$ と α/β フォールド毎に示す. 文献 10 の表を改変.

^{††} Positive predictive values (PPV = TP/(TP + FP)).

[‡] 置換確率ベクトル間の相関係数による予測. 両末端座位は除外.

^{†††} 置換確率ベクトル間の偏相関係数による予測. 両末端座位は除外.

[§] CES 法による予測;太字は DI 法より高い精度を示す.

^{§§} NMFI-DI (DCA) 法による予測;conservation filter²⁾を用いた.

置換ベクター間の相関係数 (C_{ij}^s) と偏相関係数 (C_{ij}^{st}) による PPV の比較から, 系統関係に起因する相関を除去することにより予測精度が有意に高まることを見て取れる. 共置換に基づく CES 法は, 他座位を介しての間接的相関の寄与は小さく, 系統関係に起因する相関の除去が肝要である. 方法に依らず, α 構造より β 構造のタンパク質の方が予測精度は高い傾向にある. また, 膜タンパク質 7tm_1 は予測精度が極端に低い. この差異は, 2次構造間のパッキングの差異によるものと思われる. CES 法は NMFI-DI 法より若干予測精度が高く, 特に膜タンパク質である 7tm_1 で顕著である. NMFI-DI 法は, 座位間でのアミノ酸タイプの相関に基づくため, α ヘリックス内の非近接残基対もコンタクトとして予測しがちである. 一方, CES 法は共置換による予測のため, アミノ酸タイプには強い相関が見られない膜タンパク質 (7tm_1) の場合も NMFI-DI 法よりは高い予測精度を示す (詳細は文献 10 を参照).

5. おわりに

タンパク質の構造決定は, 機能を理解する上で欠か

せない. しかし, 構造空間における未知のフォールドを探索すべく多大な努力にも拘らず, キュレイターによる Pfam-A データベース (release 26) では全 13672 ファミリーの 63%, また機能未知のドメイン (UDF) の 74% においてその全メンバーの構造が未決定である. UDF は Pfam-A においても増加しつつある. また, タンパク質ファミリー数の上限と見なし得るかもしれない (プログラムにより収集, 分類される) Pfam-B のファミリー数は 460125 に達する. 配列からの構造決定へのニーズは増すばかりである. しかし, タンパク質の立体構造予測は, フラグメントアセンブリー法や他の方法により構造空間を効率よくサンプリングすることが可能になり実用化の域に達したとは云え, 100 残基以上のタンパク質はいまだ容易ではない. ここで紹介した共進化に基づくコンタクト予測は配列長によらず予測精度が高い. Marks et al.²⁾ は, NMFI-DI 法による予測結果 (表 1 参照) に基づく距離制約を用い, 天然構造からの平均 2 乗距離 2.7–4.8 Å RMSD で予測可能であることを示した. その後, 膜タンパク質の構造予測, リガンド結合部位やタンパク質間相互作用部位などの機能部位予測に NMFI 法が相次いで適用されている^{4)–7)}. RNA 分子の立体構造予測 (塩基対, 塩基間コンタクト予測) への適用も可能だが, 塩基の種類はアミノ酸に比べ 4 種類と少数のため, 共置換に基づく CES 法が適している. 予測精度を上げるには, 多重配列アライメントの改善や分子間コンタクトも考慮する必要がある. 問題は多数の多様な相同配列を必要とするということである. 座位間相関は $L(L-1)/2$ 個の未知パラメータを含む. よって, アミノ酸タイプの座位間相関からの推定には, 方法に依らずなんらかの補正 (regularization) 無しでは少なくとも $N \geq L$ の配列が必要となる. 一方, CES 法では系統樹を考慮するため, 枝数 \geq 配列長 ($2N-3 \geq L$) が要求され, 前者の 1/2 程度になる. とは云え, 利用できる配列が少ない場合は shrinkage method による共分散行列への補正が同様に必要となる. なお, 不変座位は共進化座位の推定に関しなら情報を持ち得ず, 方法に依らず推定から除外されることに注意願いたい. しかし, DNA シーケンシングは次世代シーケンサーの登場により大量配列解析の時代となり, 各ドメインファミリーの配列はこれまで以上に増加傾向にある. 構造予測の一端を担える方法となり得るかも知れない.

文 献

- 1) Morcos, F. et al. (2011) Proc. Natl. Acad. Sci. USA **108**, E1293-E1301. doi: 10.1073/pnas.1111471108.

- 2) Marks, D. S. *et al.* (2011) PLoS One **6**, e28766. doi: 10.1371/journal.pone.0028766.
- 3) Taylor, W. R., Sadowski, M. I. (2011) PLoS ONE **6**, e28265. doi: 10.1371/journal.pone.0028265.
- 4) Hopf, T. A. *et al.* (2012) Cell **149**, 1607-1621. doi: 10.1016/j.cell.2012.04.012.
- 5) Marks, D. S. *et al.* (2012) Nature Biotech. **30**, 1072-1080. doi: 10.1038/nbt.2419.
- 6) Sułkowska, J. I. *et al.* (2012) Proc. Natl. Acad. Sci. USA **109**, 10340-10345. doi: 10.1073/pnas.1207864109.
- 7) Nugent, T., Jones, D. T. (2012) Proc. Natl. Acad. Sci. USA **109**, E1540–E1547. doi: 10.1073/pnas.1120036109.
- 8) Jones, D. T. *et al.* (2012) Bioinformatics **28**, 184-190. doi: 10.1093/bioinformatics/btr638.
- 9) Ekeberg, M. *et al.* (2013) Phys. Rev. E **87**, 012707. doi: 10.1103/PhysRevE.87.012707.
- 10) Miyazawa, S. (2013) PLoS One **8**, e54252. doi: 10.1371/journal.pone.0054252; BMC Evol. Biol. **13**, 257. doi: 10.1186/1471-2148-13-257.
- 11) Schneidman, E. *et al.* (2006) Nature **440**, 1007-1012. doi: 10.1038/nature04701.
- 12) Weigt, M. *et al.* (2009) Proc. Natl. Acad. Sci. USA **106**, 67-72. doi: 10.1073/pnas.0805923106.
- 13) Kappen, H. J., Rodríguez, F. B. (1998) Neural Comput. **10**, 1137-1156. doi: 10.1162/089976698300017386.



宮澤三造 (みやざわ さんぞう)

1975年名大理学研究科物理学専攻満期退学, 78年理学博士, 九大生物教室研究生, NIH visiting fellow/associate, 遺伝研助教授, 群馬大工学研究科准教授を経て, 2013年定年退職.

研究内容: バイオインフォマティクス.

E-mail: sanzo.miyazawa@gmail.com

URL: <https://www.sanzo.org/>

宮澤三造

用語解説

Pfam (<http://pfam.sanger.ac.uk>)

タンパク質ドメインの類似配列データベース。多重配列アライメント (seed) と Seed に対応するプロファイル隠れマルコフモデル (HMM), Seed HMM で収集されたアライメント (Full), そして FastTree により作成された系統樹データが利用できる。キュレーターによる Pfam-A とプログラムにより収集, 分類される Pfam-B がある。

(091 ページ)

(宮澤)

NMFI (Naive mean field inversion)

Ising/Potts モデルにおいて, サイト間相互作用 J に関し 2 次以上の項を無視して平均場近似を適用した場合の近似を naive mean field approximation といい, その逆問題の解を NMFI という。

(092 ページ)

(宮澤)

※本文中ゴシックで表記した用語を解説しています。