

分担研究課題： 配列データ収集ソフトウェアの開発 および 配列解析

分担研究者： 宮澤 三造

所属： 群馬大学工学部

配列データ収集ソフトウェアの開発、およびホモロジー検索のための scoring matrix の研究を行った。

1 データ収集ソフトウェアの開発

現在、世界で解析された DNA 配列データは、GenBank/EMBL/DDBJ により収集入力され 3 ヶ月毎にリリースされている。また GenBank は Usenet という計算機ネットワーク上のネットニュースとして、EMBL は電子メールにより、新データを毎日リリースしている。当研究グループ以外で解析されたデータを収集するために、このようにして毎日計算機ネットワーク上にリリースされる配列データを自動的に取り込みデータベース化するソフトウェアの開発をおこなった。このシステムは電子メールによる検索解析システムとともに海外の研究者に利用されている。また、GenBank により開発された IBM/PC 用の配列データ入力ソフトウェア AuthorIn の NEC PC98 用に移植する作業を米澤氏 (帝人) のグループとともに行った。

2 ホモロジー検索のための new scoring matrix

一般に、ゲノム解析では機能の未知の配列が多量に解析されるようになる。DNA 配列や蛋白質配列の類似性検索は、そのような配列の機能の予測、また分子進化を考える上で欠くことができない。二つの配列の類似度は、対応する塩基もしくはアミノ酸間の類似度の和が最大になるような整合配列の類似度をもって定義される。よって塩基もしくはアミノ酸間の類似度 (scoring matrix) をいかに定義するかは、類似の機能をもつ配列や進化の上で近縁の配列を類似な配列として検出する上で重要である。塩基配列の場合は同一の塩基のみを類似なものとして定義するしか方法がなく、類似性検索能力が落ちる。アミノ酸配列の場合も同一のアミノ酸のみ類似なものとして定義する (identity matrix) と、60% 以上もアミノ酸が異なる二つの配列では類似か否かを判定することは困難である。検出度をあげるためもっとも良く使用されるのは、Dayhoff (1978) によって計算された PAM78 とよばれる類似度である。この類似度は、非常に近縁の相同蛋白質配列の比較から作成したア

ミノ酸置換行列に基づいた類似度であるが、使用した配列が主に hemoglobin, cytochrome C 等のため、あらゆる蛋白質配列の類似性検索にすぐれた性質を示すかは疑問である。そこでアミノ酸置換確率行列をアミノ酸置換による蛋白質立体構造不安定化の考察から評価し、ホモロジー検索のための scoring matrix として利用することを試みた。これは以下の2ステップからなる。

2.1 アミノ酸置換による蛋白質立体構造不安定化 (unfolding Gibbs free energy の変化) の予測

蛋白質構造の安定性に及ぼす単一アミノ酸置換の効果がいくつかの蛋白質で詳しく調べられている。Yutani et al. (1987) は tryptophan synthase α subunit で、また Matsumura et al. (1988) は bacteriophage T4 で、単一アミノ酸置換による unfolding Gibbs free energy ($G_{denatured} - G_{native}$) の変化を測定し、その変化量はアミノ酸残基の hydrophobicity (Nozaki & Tanford により測定された水から有機溶媒へアミノ酸を移す際の転移エネルギー) に比例することを示し、疎水相互作用の蛋白質構造の安定化に及ぼす重要性を指摘した。しかし、tryptophan synthase α subunit の場合アミノ酸置換による unfolding Gibbs free energy の変化は hydrophobicity の変化に較べ 3-4 倍もあり、その矛盾が指摘されている。私は以前蛋白質構造におけるアミノ酸残基間相互作用を、42 の蛋白質立体構造における各種アミノ酸残基間の接触頻度から接触エネルギー (residue-residue contact energy) の形で経験的相互作用ポテンシャルとして評価した。このポテンシャルを用い、単一アミノ酸置換による unfolding Gibbs free energy change を計算し、実験値と比較した。結果は 1) 疎水残基以外の残基も含め相関係数 0.8、また 2) 回帰直線の傾きは両方の蛋白質において 0.8-0.9 であった。Yutani et al. の解析では回帰直線の傾きは 3.7 程度であったのに比較して改善されている。理由は、1) 置換による denatured state の free energy の変化を考慮したこと。2) アミノ酸間接触エネルギーによる hydrophobicity の見積もりは Nozaki & Tanford の見積りに比較して約 2 倍大きいことによるものと思われる。この結果は、よく使用される Nozaki & Tanford による hydrophobicity の評価は過小評価である可能性を強く示唆している。

2.2 アミノ酸置換による蛋白質立体構造不安定化の評価に基づくホモロジー検索のためのアミノ酸置換確率行列

アミノ酸置換による蛋白質立体構造に及ぼす影響がある程度評価できたので、類似の方法で20種のアミノ酸間の相互置換による蛋白質立体構造の不安定化を平均の意味で評価し、それをアミノ酸間の類似度として蛋白質配列の類似性検索に利用することを試みた。類似性検索は1) 祖先を同じくする配列の検索, 2) もしくは蛋白質構造が類似の可能性のある配列を検索することのいずれかの目的をもってなされることが多い。後者の場合は、蛋白質構造の保持に関するアミノ酸間の類似度を scoring matrix として使用することが望ましい。アミノ酸を置換した際期待される unfolding Gibbs free energy change を scoring matrix として使用できよう。前者の場合は、蛋白質構造の保持は、蛋白質が機能を果たす上で欠かせず、淘汰圧として働くであろう。アミノ酸を置換した際期待されるエネルギー変化の Boltzmann's factor を fitness として使用し、一様な DNA 塩基置換率を仮定しコードン表に従ってアミノ酸が変異し蛋白質構造を保持する淘汰を受けるというモデルによりアミノ酸置換確率行列を計算した。250PAM (accepted amino acid mutations per 100 amino acids) に対応するアミノ酸置換確率行列を計算し、ランダムの際に期待される置換確率との比の log をとって、類似性検索の際の scoring matrix とした。これをホモロジー検索に使用した結果、Dayhoff等により評価された実際のアミノ酸置換に基づく scoring matrix の検出力と同等ないしは上回り、その有効性が確認された。

原著論文

1. Miyazawa, S. and Jernigan, R. L. (1992) Estimation of Protein Stability Changes for Single Site Substitution Mutants. To be submitted.
2. Miyazawa, S. and Jernigan, R. L. (1992) A New Substitution Matrix for Protein Sequence Searches Based on Contact Frequencies in Protein Structures. Submitted to Protein Engineering.

解説

1. 宮澤 三造 (1990) ゲノム解析とデータベース、実験医学、18, 42-48.