

2P052

ペプチドの 3 次元構造の統計科学的予測

(岡大院・医歯薬*, 岡大・名誉**) ○真銅 隆至*, 玉懸 敬悦**, 永松 朝文*

1. 緒言

生体高分子の構造は, その高分子の機能に大きな影響を与えており, 構造を知る事は生理活性の機序を理解する上で, 不可欠と言ってもよい程に重要な意味を持つ. しかし第一原理的な電子状態理論による高分子の最安定構造の探索は, 仮令真空中の計算でも計算コストの面から現実的ではない.

そこで分子力学法や統計科学的な手法を含む経験的な手法を提案し, 性能のよいパラメータを決定する事は意味のある事であると思われる. 今回 我々は残基の置かれた環境同士の非類似度を 1 種の説明変数とした統計科学的な手法を提案する.

2. 手法

ペプチドの 3 次元構造の内重要な情報である骨格の情報(この場合はアミノ酸の N-CA-C 原子がなす平面と隣のアミノ酸の N-CA-C 平面とがなす角)を予測に於ける目的変数とした. 説明変数としては, 残基が置かれている環境(ここでは注目している残基自身のアミノ酸の種類と及び連なっている残基の種類, 1 次構造)同士の非類似度を採用した. 具体的には次の式で定義した.

$$Dissimilarity_{kl} = \sum weight_n \times DA_{0n}$$

ここで $weight_n$ は注目している残基 k と l から, それぞれ n コ分だけシフトしたアミノ酸の種類が, k 及び l での構造に与える影響の大きさを数値化したもので, n が 0 の場合(そのアミノ酸自身の種類)も考慮している. また DA_{0n} は k と l から n 残基だけ離れた処にあるアミノ酸同士の非類似度であり, 2 つのアミノ酸を指定したら 1 意に決まるものとした. $weight_n$ と DA_{0n} は予測能を極大にする様に決定するパラメータであり, ランダムに与えた値から逐次改善して求めていく. $Angle_k$ を予測する際のモデル式は次の様にした.

$$Angle_k = \frac{\sum_{k \neq l} Angle_l \frac{1}{Dissimilarity_{kl}}}{\sum_{k \neq l} \frac{1}{Dissimilarity_{kl}}}$$

誤差 S は, 平面同士の角度の全体から注目する角度 1 つだけを取り除いたときの, その角度の予測値と実験値との誤差の 2 乗を計算し, それを全ての角度について和を取り定義した. 誤差が最小のとき予測能が最大であると考え, 逐次改善は S を極小にする様行つた. 具体的に S は以下の式で表される.

$$S = \sum_k \left(\frac{\sum_{k \neq l} Angle_l \frac{1}{dissimilarity_{kl}}}{\sum_{k \neq l} \frac{1}{dissimilarity_{kl}}} - Angle_k \right)^2$$

極値探索は Taylor 展開の 2 次までの式で近似し, 改善の次のステップで極値を取る様計算した. またその際, 2 階微分値が負の場合はパラメータを更新する為に加える値の符号を反転させ, 極大値にトラップされない様にした.

また, 予測式を変えたときに, 柔軟にプログラムを変更出来る様に, パラメータによる微分を, アミノ酸の置かれている環境同士の非類似度による微分と, 環境の非類似度のパラメータによる微分の 2 段階に分けて計算を行った.

3. データ

データは PDB サイトから, NMR により決定された構造の中から, 鎖の数が 1 つで, 残基が 30 以内のもの, そして核酸, 糖, D 型アミノ酸やヘテロアトムを含まないものを探し, その中から選び出した.

計算結果の詳細は当日報告する.

参考文献

PDB サイト

http://www.pdbj.org/index_j.html