

For NIG, 12/17-12/18/99, (1999 年 12 月 6 日)

by Sanzo Miyazawa

蛋白質の配列・構造間の適合性判定; 配列・構造アライメントと評価関数

宮澤三造 (群馬大学工学部共通講座)
miyazawa@smlab.sci.gunma-u.ac.jp

前回は、私共が提案したアミノ酸残基レベルでの経験的エネルギーポテンシャル (アミノ酸接触エネルギー) について、(1) 実際の残基間相互作用を反映しているのか否か? (2) Sippl の考案した平均力ポテンシャルとの違い、に関する議論を中心にご報告した。今回は、エネルギー関数を適合性判定、特に配列・構造アライメントのための評価関数として使用するための問題点、および配列・構造アライメントによる配列・構造間の適合性判定について報告する。

経験的ポテンシャルはそのままではアミノ酸配列と立体構造の適合性評価関数としては使用できない (論文 2)。アミノ酸配列がある立体構造に適合するか否かは、その立体構造の安定性によって測られよう。立体構造の安定性は、その構造エネルギーだけではなく構造分配関数にも依存する。構造に適合する配列の検索ばかりではなく、配列に適合する構造の検索の際にも、gap を許す配列・構造アライメントを用いるならば、分配関数を考慮しなければならない。分配関数の評価に関して、蛋白質における最適なアミノ酸配列の設計に関する研究でも、議論されている。しかし、配列・構造アライメントに使用する評価関数としては、(1) アルゴリズムからくる制約 (2) データベース検索に適用できるための高速性を考慮しなくてはならない。データベース検索に使用できるような適合性評価関数について提案する。

配列アライメントでは、通常、評価関数は残基に関し 1 体のポテンシャルのみを含む。また 2 次構造予測の場合でも、多体のポテンシャルを含むがその相互作用レンジは限られている。よって評価関数の最適解は動的計画法で求めることができる。一方、配列・構造アライメントにおいては、2 体以上の相互作用を含むエネルギー関数を評価しなくてはならず、一般の構造予測同様 NP 完全の問題である。よって 2 体のポテンシャルをいかに (近似的に) 評価するが問題となる。

また、適合性が低いような場合には、評価関数の最小 / 最大に対応する解のみを考慮しても意味がない。配列・構造アライメントにおける評価関数はエネルギーである。配列アライメントであっても、評価関数の値は、likelihood である。論文 1 で議論したように、評価関数の最小 / 最大値に対応する解は most probable alignment であると言う意味以上のものではない。むしろ、個々の残基対がアラインする確率が重要である。アライメントの信頼確率は、分配関数から計算できる。分配関数は、(1 次元 ising model における分配関数を計算する手法の一つである) transfer matrix method により計算できる。(動的計画法は energy minimum を求める方法であり、transfer matrix method は分配関数を計算する方法である。) 我々は、最適アライメントよりもむしろアライメント確率の最大なる残基対からなるアライメント構築のアルゴリズムを以前提唱した (論文 1)。この方法は、アライメント中の個々の残基対のアライメント確率 (信頼度) を与え、類似性の低い場合に特に有用な方法である。アライメント確率の高い残基対のみを考慮すると、RMSD 値の低いアライメントが得られる。

ここでは、アライメント確率に基づく 2 体相互作用の近似的評価法および配列・構造アライメント構築法について報告する。

参考文献

1. Miyazawa, S.: A Reliable Sequence Alignment Method based on Probabilities of Residue Correspondences. *Protein Engineering*, Vol. 8, pp. 999-1009, (1995).
2. Miyazawa, S., and Jernigan, R. L.: An Empirical Energy Potential with a Reference State for Protein Fold and Sequence Recognition. *PROTEINS: Structures, Function, and Genetics*, Vol. 36, pp. 357-369, (1999).