

⑥ Protein Sequence Analysis by Parallel Inference Machine

M.Ishikawa*, M.Hoshida, M.Hirosawa, T.Toya, K.Onizuka, K.Nitta(ICOT,日本)

発表要旨

タンパク質を構成するアミノ酸配列の類似性抽出をする、マルチプルアライメントという課題について、並列推論マシンを用いた高速実行を試みた。従来、マルチプルアライメントは、実用的な時間内に実行できるアルゴリズムによる解の品質が十分でなく、もっぱら熟練した生物学者の手作業で行われてきた。そんな中で昨年、反復改善法という、小規模な課題であれば高品質な解を与えるアルゴリズムが発表された。我々はそのアルゴリズムに、並列探索評価法と限定探索手法を拡張することによって、実用的な時間内に高品質な解を安定して得られる並列限定反復改善法を開発して。これによって、実用的なマルチプルアライメントの課題が、並列計算機で自動的に解決できる可能性が明らかになった。

質疑応答

質問：マルチプルアライメントの全体の評価値はどのように算出しているのか？

回答：先ほど示したDayhoffのマトリクスによって、文字対の評価値とする。各カラムに存在するすべての文字対にわたって、それらの文字対の評価値を加えて、カラムの評価値とする。それを全カラムにわたって加えて全体の評価値とする。この方法は、この分野で一般的に使われている方法である。

質問：N本のマルチプルアライメントを行うのに、N次元のダイナミックプログラミングを行えばいいのではないかと？

回答：理論的には、それで最適なマルチプルアライメントが得られるのであるが、配列の長さのN乗の計算量が必要であり、実用的には4次元以上のダイナミックプログラミングは不可能である。それは並列計算機を用いても原理的には難しく、効果的な近似法を開発することが望まれている。