

招待講演

## タンパク質配列解析の並列最適化問題解決

石川 幹人

Masato Ishikawa

松下電器産業（株）マルチメディアシステム研究所

タンパク質は、20種類のアミノ酸の文字配列として、特徴づけられる。その文字配列の解析により、生物学的な知見が得られる。最も典型的な配列解析手法のひとつは、マルチプルアライメントと呼ばれ、類似した性質を持つアミノ酸の文字を、複数の文字列間で縦に並べ揃える処理である。アミノ酸同士の類似性尺度が、経験的に定められていることから、マルチプルアライメントは、最適化問題として捉えることができる。この最適化問題の解決には、古くからダイナミックプログラミング（DP）の適用が試みられてきた。DPを用いれば、理論的には最適解が得られるのであるが、N本の文字配列のアライメントに、配列長のN乗のメモリと、N乗の計算時間が必要となる。すると、百文字以上あるアミノ酸の文字列について、現実的にDPが適用可能なのは、わずかに配列3本のアライメント（3次元DP）までである。そのため、マルチプルアライメントの問題解決には、準最適解を求める様々な近似解法や、その高速化手法が研究されてきた[1]。最も一般的な方法は、2次元DPの結果を組み合わせる方法であるが、実用規模の問題において、高品質な解を実用的時間内に求めるのは、依然として困難な課題であった。

我々は、（財）新世代コンピュータ技術開発機構において、同機構で開発された並列推論マシンを利用し、マルチプルアライメントの並列最適化問題解決の研究にあたった。最初の段階では、DPの内部処理の並列化を行った。3次元DPの並列パイプライン処理を実装し、台数に対して5割以上の速度向上効果が得られた[2]。そして、それらの配列3本のアライメント結果を組み合わせることで、マルチプルアライメントを形成した[3]。しかし、実用規模の問題では、実行時間に見合う十分な品質の解は得られなかった。次の段階では、シミュレーテッドアニーリングの手法をマルチプルアライメントの問題解決に導入し、並列に多点探索を行った[4]。しかし、定義した微小変形に対して、マルチプルアライメントの問題は規模が大きすぎ、現実的な時間内では小さな問題しか解決できなかった。ただし、他の方法によって得られたアライメント結果の修正には、効果があった[3]。

第3段階では、DPによる部分的な最適化を並列に繰り返して、全体のマルチプルアライメントを改善していく、並列反復改善法を開発し実装した[5]。これにより、実用規模の問題においても、高品質な解を実用的時間内に求めることが可能になった。さらに、並列反復改善法について、最良優先探索と多点分散探索とを比較検討した[6]。最終段階では、並列反復改善法を遺伝的アルゴリズムの枠組みにあてはめ、局所最適解に陥りやすい問題点を、軽減した。さらに、マルチ個体群の世代管理手法を導入することで、最良優先探索と多点分散探索とを融合した探索を実現した[7]。これにより、シミュレーテッドアニーリングにおける温度スケジューリングに相当する効果が形成できる。

我々は、並列反復改善法を装備したマルチプルアライメント編集ツール[8]も開発した。こうした諸々のソフトウェアは、<http://www.icot.or.jp/> より公開されている。

\* \* \*

- [1] 石川, 金久: 「ヒトゲノム計画と知識情報処理」培風館, pp65-97, 1995.
- [2] 戸谷, 星田, 石川, 新田: 情報処理学会プログラミング研究会 5-14, pp127-134, 1991.
- [3] 石川, 星田, 広沢, 戸谷, 鬼塚, 新田, 金久: 情報処理学会情報学基礎研究会 23-2, pp1-14, 1991.
- [4] 戸谷, 石川, 荒木, 星田: 情報処理学会記号処理研究会 69-1, pp1-8, 1993.
- [5] 星田, 石川, 広沢, 戸谷, 十時: 情報処理学会情報学基礎研究会 27-2, pp13-24, 1992.
- [6] 石川, 十時, 戸谷, 星田, 広沢: 情報処理学会論文誌, Vol.35, No.12, pp2816-2830, 1994.
- [7] 戸谷, 石川: 情報処理学会論文誌, Vol.36, No.11, pp2549-2558, 1995.
- [8] Ishikawa, Totoki, Tanaka and Hirotsawa: Bioinformatics and Genome Research, World Scientific, pp385-396, '95.

