

MEGADOCK : 大規模タンパク質間相互作用予測システムとその応用

大上 雅史 松崎 由理 石田 貴士 秋山 泰

東京工業大学 大学院情報理工学研究科 計算工学専攻

1. 概要

タンパク質間相互作用 (Protein-Protein Interaction, PPI) に関するネットワークの理解は、細胞内のシグナル伝達経路の特定や、それをターゲットとした創薬における重要課題となっている。本研究では、PPI を計算機で大規模にかつ高速に予測するため、並列計算機上での効率的な並列計算を可能にする PPI 予測システム “MEGADOCK” [1] を開発した。これにより、従来では実行が困難であった大規模な PPI ネットワークの予測が、大規模計算機の利用で容易に行えるようになった。

2. 手法

2.1 ハイブリッド並列化

MEGADOCK は大量のタンパク質ペアに対する計算を MPI/OpenMP によるハイブリッド並列で行うように実装されている。MEGADOCK のシステムの概要図を図 1 に示す。各タンパク質ペアの割り振りは MPI によるノード間並列によって、各タンパク質ペアごとの相互作用予測は OpenMP によるスレッド並列化によって並列計算を行っている。

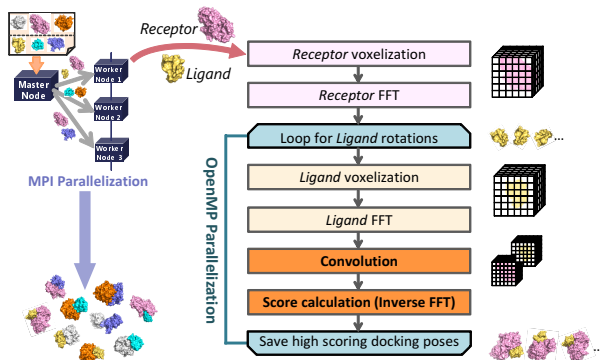


図 1 MEGADOCK の処理の概要。

2.2 タンパク質ドッキング計算

MEGADOCK はボクセル化による粗視化モデルに基いたタンパク質ドッキング計算を利用している (図 1 右側)。評価関数に形状相補性・静電相互作用・疎水性相互作用を考慮しており、高速フーリエ変換 (FFT) によってこれらを同時に計算することを可能とした [2]。FFT のライブラリは FFTW・FFTE・CSSL2 を検討し、FFT 点数に関わらず総合的に高速な計算速度を示した FFTW による実装を行った。

2.3 実装

MEGADOCK システムは C++ によって記述され、先に示したように MPI/OpenMP によるハイブリッド並列が実装されている。オープンソースで公開されており、<http://www.bi.cs.titech.ac.jp/megadock> から取得可能である。また、以下では東京工業大学の TSUBAME 2.0 での実行結果を示すが、本システムは理研 AICS の京でも利用可能 (http://www.islim.org/islim-dl_j.html) である。

3. 実験

3.1 ハイブリッド並列化効率の測定

TSUBAME 2.0 によって並列化効率の測定を行った。TSUBAME 2.0 (Thin ノード) は Intel Xeon 5670 (2.93 GHz, 6 コア) を各ノード 2 つずつ持っており、計 12 コアの利用が可能である。図 2(a) に MPI 並列の測定結果を示す。

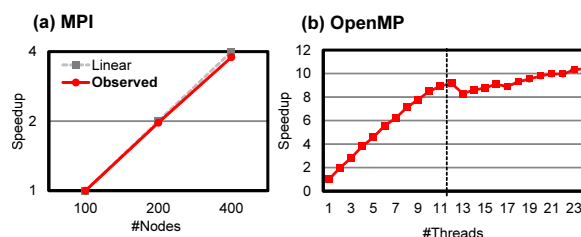


図 2 (a)MPI 並列の並列性能測定結果 (ストロングスケーリング)。(b)OpenMP スレッド並列の並列性能測定結果、破線の右側 (#Threads=13~24) はハイパースレッディング領域である。測定は (a)(b) ともに TSUBAME 2.0 Thin ノードによる。

ノード間並列ではストロングスケーリングが 100-200 ノードのときに 0.989, 200-400 ノードのときに 0.956 となった。また、図 2(b) に OpenMP によるスレッド並列の測定結果を示す。スレッド並列では 24 スレッド利用のときに 1 スレッド時の約 10.4 倍の計算速度となった。

3.2 パスウェイ解析への応用

ヒトの細胞死に関わるアポトーシス現象に関するパスウェイを構成するタンパク質群に対して MEGADOCK を適用した。パスウェイに該当する 57 個のタンパク質種に対応する 158 個の構造群に関して、 158×158 通りの総当たり計算を TSUBAME 2.0 を用いて行った。計算は約 500 ノード時間で完了した。結果を図 3 に示す。既知 PPI の再現率は 48% であり、また多くの新規 PPI 候補が発見された。

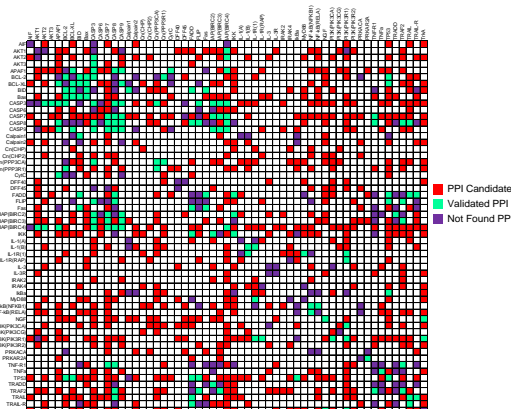


図 3 アポトーシスパスウェイへの適用結果。

文献

- [1] 大上雅史, 松崎由理, 松崎裕介, 佐藤智之, 秋山泰. 情処論 数理モデル化と応用, 3(3): 91-106, 2010.
- [2] Ohue M, Matsuzaki Y, Ishida T, Akiyama Y. In *Proc of PRIB2012, LNCS 7632*, 178-187, Springer, 2012.