

課題名 (タイトル) :

## 大規模タンパク質間ネットワーク推定に関する研究

利用者氏名 : ○宮野 悟, 秋山 泰, 石田 貴士, 内古閑 伸之, 大上 雅史,  
佐藤 智之, 藤原 康広, 松崎 由理所属 : 社会知創成事業 次世代計算科学研究開発プログラム  
次世代生命体統合シミュレーション研究推進グループ  
データ解析融合研究開発チーム  
東京工業大学 大学院情報理工学研究科

## 1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

本課題では、次世代生命体統合シミュレーションプロジェクトの委託研究として、大規模タンパク質間ネットワーク推定を行っている。

本研究では、生命体の理解に向けて重要な鍵の一つとなる「タンパク質間相互作用ネットワーク」の推定を可能とすることを目的として、大規模データ解析の手法に基づき、超高速に候補タンパク質間の相互作用の可能性を調べる計算アルゴリズムを開発するとともに、大規模並列計算機上で性能測定を行う。

これまでに、小規模な系を例題としてシステム生物学的な研究上の課題に対して本システムを適用して、新たなタンパク質相互作用の候補の提案やタンパク質-RNA 相互作用探索などを実施してきた (Matsuzaki, *et al.*, *JBCB* 2009, Ohue, *et al.*, *Genome Info* 2011 他)。その際には、約 100 のタンパク質立体構造を対象に、全タンパク質の組み合わせについて相互作用可能性の有無を総当たりで予測した。また、データ解析融合チームが掲げる「肺がんと薬」のテーマに関係の深いヒトの EGFR シグナル伝達系を例題として、 $500 \times 500 = 250,000$  規模の問題で予備実験を行った。現在では、データ解析融合チームの宮野研究室と連携して、遺伝子ネットワーク推定結果から得た探索対象タンパク質群を対象に、 $2,000 \times 2,000 = 4,000,000$  規模の問題に本課題で開発したタンパク質相互作用予測プログラムを適用し、その性能評価を進めている。

2011 年度より次世代機「京」での性能評価も進めているが、実用的な計算は継続して RICC で行っている。

## 2. 具体的な利用内容、計算方法

本研究では、以下の手順で網羅的なタンパク質間相互作用予測を行う。タンパク質ドッキング計算には独自で開発したソフト「MEGADOCK」(大上他、情処論 TOM, 2010) を利用する。

## i. 網羅的タンパク質-タンパク質ドッキング

対象とするタンパク質の全組み合わせについて、1 対 1 のドッキング計算を行い、3,600 個の複合体候補構造 (デコイ) を作成する。

ドッキング計算は、形状相補性に関する項  $G$  と静電的相互作用の項  $E$  の計算からなるドッキングスコアによってデコイを評価する。それぞれのタンパク質を 1 辺が  $1.2 \text{ \AA}$  の三次元ボクセル空間上に配置し、タンパク質の内部、表面、その他の種別によって、各ボクセル上に異なるスコアを代入する。形状相補性の項  $G$  には、我々が新規に提案した「real Pairwise Shape Complementarity (rPSC)」スコア (大上他、情処論 TOM, 2010) を用いる。既存手法と比べて計算時間の面で有利なモデルとなっている。また、静電的相互作用を、アミノ酸残基ごとに CHARMM19 に基づいて原子に電荷を与え、ボクセルに分割してボクセル電荷  $q(l, m, n)$  を決定し、静電的相互作用の項  $E_r(l, m, n)$ ,  $E_l(l, m, n)$  を決める (対象とするタンパク質ペアの一方をレセプターR、もう一方をリガンドL とする)。ドッキング予測の良さを表す評価値であるドッキングスコア  $S$  を

$$\begin{aligned}
 R(l, m, n) &= G_R(l, m, n) + iE_R(l, m, n) \\
 L(l, m, n) &= G_L(l, m, n) + iwE_L(l, m, n) \\
 S(\alpha, \beta, \gamma) &= \mathcal{R} \left[ \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N R(l, m, n) L(l \right. \\
 &\quad \left. + \alpha, m + \beta, n + \gamma) \right]
 \end{aligned}$$

と定義する。 $(\alpha, \beta, \gamma)$  はリガンドの平行移動ベクトルである。

MEGADOCK ではドッキングスコアを、リガンドを平行移動させながら全空間における畳み込み和として計算する。ある回転角に対してボクセルサイズ  $N$  に対して  $N \times N \times N$  通りの平行移動を行う。その中から最も良いドッキングスコアを持つリガンドの平行移動ベクトルを、その角度におけるドッキング結果として返す。回転角のサンプリングにおいては、3,600 通りの回転パターンで計算を行う。よって、1 つの複合体について計算されるドッキング結合部位は、ボクセルサイズが  $N$  のとき、 $3,600 \times N^3$  通りとなる。なお、計算時間は単純に畳み込み和をとると  $O(N^6)$  だが、離散フーリエ変換 (DFT) と逆離散フーリエ変換 (IFT) を用いて、

$$\begin{aligned}
 S(\alpha, \beta, \gamma) \\
 = \text{IFT}[\text{DFT}[R(l, m, n)] * \text{DFT}[L(l, m, n)]] \quad \text{と}
 \end{aligned}$$

し、高速フーリエ変換 (FFT) を用いることで  $O(N^3 \log N)$  に削減することが可能となる。 $z^*$  は  $z$  の複素共役を表す。rPSC を用いることにより、既存の代表的ソフトウェア ZDOCK などが用いている複素数による表現モデルに比べて離散関数の虚数部に空きができる分、他の物理化学的相互作用の導入が可能となり、かつ FFT の計算回数の増加を回避し、計算の高速化を図っている。

#### ii. デコイのリランキング

各タンパク質ペアについて、上位数千個のデコイを ZRANK によりリランキングする。複数の方法を用いてエネルギースコア計算を行う。

#### iii. デコイのクラスタリング

デコイの構造類似性や相互作用プロファイルを利用したクラスタリングを行う。これまでに RMSD をもとにした構造類似性の評価方法を用いてデコイのクラスタリングを行ってきた。さらに、相互作

用に着目した類似性によってデコイをクラスタリングする。この手法では、残基単位での相互作用の有無のパターンについて、Tanimoto index を用いて線形時間で類似度を計算する。

#### iv. タンパク質間相互作用 (PPI) 評価と判定

クラスタリング結果を分析し、各タンパク質ペアについてタンパク質相互作用の有無を判定する。本研究で特に大規模な計算が必要となるのは 2-i のドッキングである。ドッキングについてはプログラムの並列化が可能である。並列計算による計算効率を測定し、MEGADOCK の性能を評価する。

### 3. 結果

#### i. タンパク質間相互作用予測システムの実データへの応用

宮野研究室提供による、肺がん薬関連の遺伝子データ (1,500 遺伝子) から得たタンパク質群に MEGADOCK を適用し、2,000 × 2,000 規模の網羅的タンパク質間ドッキングを行った。現在はこの結果を解析して新規のタンパク質間相互作用の候補を抽出する計算を行っている。

#### ii. アラニン変換による前処理評価

これまでタンパク質間相互作用予測手法として相互作用プロファイルを用いたクラスタ解析手法を開発してきた。デコイのグループ分割法について問題があったため、相互作用部位予測に適した方法とパラメータの探索を行った。

#### iii. リドッキング手法

相互作用プロファイルを用いて正解相互作用を含む代表相互作用プロファイルを定義した。この情報を用いてドッキング探索空間を縮小し、再度ドッキングすることによりある程度限定された空間内で新たなデコイ構造を多数取得することができた。リドッキング段階を経ることでより正解に近いデコイを得られることが示唆された。

#### iv. 並列化、計算効率向上

OpenMP によるノード内の回転角度並列に加え、MPI によるレセプターごとのドッキングの並列化を実装した。このハイブリッド並列化したコードを RICC 上でチューニングし、次世代機「京」で大規模並列計算を実行することが可能になった (2012 年 3 月現在 80,000 コア以上を達成)。

4. まとめ

今年度はノード内並列とノード間並列の二段階の並列化（ハイブリッド並列化）の実装を完了し、大規模並列計算を実行可能とした。応用課題として、肺がんに関連するヒトの EGFR シグナル伝達系タンパク質間のドッキング計算を終了した。また、新たなポストドッキング解析手法の提案と評価を行った。

5. 今後の計画・展望

i. タンパク質間相互作用予測システムの改良

相互作用プロファイルを用いたリドッキングによりドッキング探索空間の縮小の手法を開発する。

ii. 実データへの応用

今年度実行した EGFR 系を対象としたドッキング結果を利用し、相互作用プロファイル解析などによる大規模な未知相互作用探索を進める。

6. 継続して利用する際に行う具体的な内容

これまでに研究に必要な計算の主要部分を終了した。次年度はタンパク質ドッキング結果の後処理のために簡易利用を申し込む予定である。

平成 23 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

Masahito Ohue, Yuri Matsuzaki, Yutaka Akiyama, Docking-calculation-based Method for Predicting Protein-RNA Interactions., *Genome Informatics*, 25(1):25-39, 2011.

Sarel J. Fleishman, Timothy A. Whitehead, Eva-Maria Strauch, Jacob E. Corn, Sanbo Qin, Huan-Xiang Zhou, Julie C. Mitchell, Omar N. A. Demerdash, Mayuko Takeda-Shitaka, Genki Terashi, Iain H. Moal, Xiaofan Li, Paul A. Bates, Martin Zacharias, Hahnbeom Park, Jun-su Ko, Hasup Lee, Chaok Seok, Thomas Bourquard, Julie Bernauer, Anne Poupon, Jerome Aze, Seren Soner, Sefik Kerem Oval., Pemra Ozbek, Nir Ben Tal, Turkan Haliloglu, Howook Hwang, Thom Vreven, Brian G. Pierce, Zhiping Weng, Laura Perez-Cano, Carles Pons, Juan Fernandez-Recio, Fan Jiang, Feng Yang, Xinqi Gong, Libin Cao, Xianjin Xu, Bin Liu, Panwen Wang, Chunhua Li, Cunxin Wang, Charles H. Robert, Mainak Guharoy, Shiyong Liu, Yangyu Huang, Lin Li, Dachuan Guo, Ying Chen, Yi Xiao, Nir London, Zohar Itzhaki, Ora Schueler-Furman, Yuval Inbar, Vladimir Patapov, Mati Cohen, Gideon Schreiber, Yuko Tsuchiya, Eiji Kanamori, Daron M. Standley, Haruki Nakamura, Kengo Kinoshita, Camden M. Driggers, Robert G. Hall, Jessica L. Morgan, Victor L. Hsu, Jian Zhan, Yuedong Yang, Yaoqi Zhou, Panagiotis L. Kastritis, Alexandre M. J. J. Bonvin, Weiyi Zhang, Carlos J. Camacho, Krishna P. Kilambi, Aroop Sircar, Jeffrey J. Gray, Masahito Ohue, Nobuyuki Uchikoga, Yuri Matsuzaki, Takashi Ishida, Yutaka Akiyama, Raed Khashan, Stephen Bush, Denis Fouches, Alexander Tropsha, Juan Esquivel-Rodriguez, Daisuke Kihara, P. Benjamin Stranges, Ron Jacak, Brian Kuhlman, Sheng-You Huang, Xiaoqin Zou, Shoshana J. Wodak, Joel Janin, David Baker, Community-Wide Assessment of Protein-Interface Modeling Suggests Improvements to Design Methodology, *Journal of Molecular Biology*, 414(2):289-302, 2011.

【国際会議、学会などでの口頭発表】

Masahito Ohue, MEGADOCK: An all-to-all protein-protein interaction prediction system based on tertiary structures information, Proc. of IIT Madras-Tokyo Tech Joint Workshop on Bioinformatics and Large Scale Data Analysis, Madras India, 15 July, 2011.

Nobuyuki Uchikoga, Analysis for generating representative interaction fingerprints of protein-protein docking decoys, Proc. of IIT Madras-Tokyo Tech Joint Workshop on Bioinformatics and Large Scale Data Analysis, Madras India, 15 July, 2011.

Nobuyuki Uchikoga, Yuri Matsuzaki, Masahito Ohue, Takatsugu Hirokawa, Yutaka Akiyama, Representative Interaction FingerPrint from decoys generated by rigid-body protein-protein docking, The 49<sup>th</sup> Annual Meeting of the Biophysical Society of Japan, Hyogo Japan, 16 Sep., 2011.

大上雅史, 松崎由理, 内古閑伸之, 石田貴士, 秋山 泰, ドッキング計算に基づく網羅的タンパク質-RNA 間相互作用予測, 情報処理学会 第 25 回バイオ情報学研究会, 2011-BIO-25:1-8, 沖縄, 24 Jul., 2011.

山本航平, 大上雅史, 内古閑伸之, 松崎由理, 石田貴士, 秋山 泰, タンパク質間相互作用予測における可視化手法の開発および予測精度の改善, 情報処理学会 第 25 回バイオ情報学研究会, 2011-BIO-25:1-8, 沖縄, 24 Jun., 2011.

ポスター発表

Masahito Ohue, Yuri Matsuzaki, Nobuyuki Uchikoga, Takashi Ishida, Yutaka Akiyama, Bound state structure estimation of protein-protein complex using rigid-body protein-protein docking method, The 10th Asia-Pacific Bioinformatics Conference (APBC2012), Melbourne, Australia, 17-19 Jan 2012.

Masahito Ohue, Yuri Matsuzaki, Nobuyuki Uchikoga, Takashi Ishida, Yutaka Akiyama, Protein Complex Structure Prediction with Normal Mode Ensemble and Physical Docking, Bioinformatics week in Odaiba 2011 (BiWO2011), Tokyo, Japan, 24-27 Jan, 2012.

Nobuyuki Uchikoga, Takatsugu Hirokawa, Yutaka Akiyama, Post-docking process by using representative interaction fingerprints, Bioinformatics week in Odaiba 2011 (BiWO2011), Tokyo, Japan, 24-27 Jan, 2012.

Masahito Ohue, Yuri Matsuzaki, Nobuyuki Uchikoga, Takashi Ishida, Yutaka Akiyama, Binding Conformation Prediction of a Protein Complex Using Physical Docking, GCOE Compview Final Symposium (Preliminary event of ISAAC 2011), Tokyo, Japan, 5 Nov. 2011.

Yuri Matsuzaki, Masahito Ohue, Nobuyuki Uchikoga, Takashi Ishida, Yutaka Akiyama, MEGADOCK: A fast protein-protein interaction prediction system by exhaustive docking, The 9th International Workshop on Advanced Genomics, Jul., Tokyo, Japan, 12-14 July, 2011.

Yuri Matsuzaki, Masahito Ohue, Nobuyuki Uchikoga, Takashi Ishida, Yutaka Akiyama, MEGADOCK: a Fast Protein-protein Interaction Prediction System by Exhaustive Docking, 12th International Conference on Systems Biology (ICSB 2011), Sep., Stockholm, Sweden, 15-18 June, 2011.

秋山 泰, 松崎 由理, 内古閑 伸之, 石田 貴士, 大上 雅史, 網羅的タンパク質ドッキング解析プログラム MEGADOCK の開発と応用, 次世代生命体統合シミュレーションソフトウェアの研究開発成果報告会 ISLiM 成果報告会 2011, 東京, 12/22, 2011.

松崎 由理, 内古閑 伸之, 大上 雅史, 石田 貴士, 秋山 泰, MEGADOCK: 網羅的な立体構造のドッキングによるタンパク質間相互作用予測アプリケーション, バイオスーパーコンピューティング サマースクール 2011 バイオスーパーコンピューティング, 兵庫, 9/26, 2011.

内古閑 伸之, 秋山 泰, Rigid-body protein-protein docking で正解相互作用をとらえるための手法の提案, バイオスーパーコンピューティング サマースクール 2011 バイオスーパーコンピューティング, 兵庫, 9/26, 2011.