

研究種目：基盤研究（B）
研究期間：2007～2011
課題番号：19300102
研究課題名（和文） 候補サイトの空間分布に注目した信頼度の高いタンパク質ドッキング判定手法の開発
研究課題名（英文） An exhaustive all-to-all protein-protein interaction prediction algorithm based on 3-D shape complementarity and its statistical distribution on protein surface.

研究代表者

秋山 泰 (AKIYAMA YUTAKA)
東京工業大学・大学院情報理工学研究所・教授
研究者番号：30243091

研究代表者の専門分野：バイオインフォマティクス
科研費の分科・細目：情報学・生体生命情報学
キーワード：蛋白質、ドッキング、高性能計算、バイオインフォマティクス

1. 研究計画の概要

各種の生物のゲノム情報が急速に読み取られ、生体細胞を構成する要素としての遺伝子やタンパク質の列挙的な理解が大きく進展した。2000年頃からは、これらの要素間のシステム上での関係を研究する「システム生物学」(Systems Biology)が勃興し、遺伝子制御ネットワーク、タンパク質相互作用ネットワーク、化合物の代謝ネットワークなど、各種の細胞内ネットワークを実験計測と計算の融合により求める研究が深化している。

本研究は、タンパク質-タンパク質間相互作用(Protein-Protein Interaction; PPI)ネットワークの解明を支援する新たな計算論的な手段の提供を目的として、タンパク質立体構造を活用したドッキング計算に基づき、任意の2つのタンパク質構造間での相互作用の可能性の有無を網羅的かつ高速に判定する計算手法の開発を行う。生体内ではタンパク質間の会合を通じて、複雑な機能を有する複合体の形成、シグナル伝達、物質の細胞内輸送など多くの現象が制御されており、PPIを知ることは「生命機能のシステム的理解」を支える最も重要な柱の一つである。PPIを調べるための様々な実験計測手法(LC-MS/MS, Y2H, FRET等)が発展し、網羅的な解析プロジェクトも進みつつあるが、タンパク質の組合せは膨大であることから、計算論的な手法による予測への期待は大きい。

本研究では、提案者が開発中の独自のタンパク質ドッキングソフトウェアの評価関数および計算アルゴリズムに改良を重ねて、我が国最高速の並列計算機を活用した網羅的な評価実験を行う。計算の中核は主に構造表面の幾何学的な相補性に注目するもので

あるが、静電相互作用をはじめとする物理化学的相互作用についても計算量を増大させないよう慎重に工夫しながら計算に加える。

PDBデータベースに登録済のタンパク質立体構造から、システム生物学上のテーマ等に基づき重要度の高い集合を選び出し、その変形構造なども考慮した上で、少なくとも数百×数百通りの網羅的なドッキング計算を実施し、最終的には1000×1000(百万通りのペア)級を越える計算の提供を目指す。

ドッキングの解を選ぶ際、自由エネルギーの見積値といった従来の物理化学的な概念とは異なる視点も重視し、特に候補となるドッキングサイトのタンパク質表面での分布の特徴やその統計的有意性に注目して、計算によって得られた解の「信頼度」を数値的に提示する方法を開発する。従来の技術では、比較的良くドッキング現象を捉えた正しい計算結果が、スコア数値だけは高いが生物学的には到底信頼できないドッキング予測と混在しており、両者の判別が困難であった。もしもこの両者をきちんと切り分けて提示することができれば、利用者は可能性の高いペアだけを選び、実験へのフィードバックを的確に行うことが可能となろう。「信頼度」の定量的な目安を与えることができれば、その閾値をユーザが選ぶことにより、感度(Sensitivity)と選択度(Specificity)のバランスを変化させて、信頼度の高い予測や、逆に新規の相互作用の発見につながる感度の高い予測を行うことが可能となる。

配列解析におけるBLASTのように、ユーザがクエリ構造を入力すると、既知構造との間で網羅的計算を行い、相互作用候補のタンパク質を順位表示する機能の提供も目指す。

2. 研究の進捗状況

(1)計算手法の改良

Katchalski-Katzir の表面形状相補性の計算に、物理化学的相互作用を計算効率を落とさず加える手法を検討し、新規に rPSC (real PSC)法を提案した。Zhang らの ZDOCK で採用された PSC 関数と類似し、深いクレフト部との一致に高得点を与える手法だが、複素数を使わず実数空間で表現して、空いた虚数部に CHARMM の力場から再定義した静電相互作用の簡易計算項を付加した。複素 FFT の回数が減ることにより ZDOCK3.0 に匹敵する予測精度を約 4 倍の計算速度で実現した。形状相補性項と静電相互作用項の比率は通常は固定しているが、タンパク質の特性に注目して動的に調整すると予測精度が向上することも示した。また ZDOCK 等の既存ソフトウェアとは異なり、回転角度毎の候補を複数保持するようにして、ZRANK によるランキングの後処理をこれら加えることによりさらに予測精度が向上した。

(2)解の空間分布への考慮

従来の類似研究では、定義したスコア関数の値のみに注目していたが、本研究では角度毎の候補数を増やし多数の準最適解を保持した上で、候補のタンパク質表面での空間分布の吟味を重視した。解の集中度を分析するため 6 種類の空間的クラスタリング法を試行し、後処理による精度向上を実現した。表面分布だけでなく、主鎖に沿った IFT プロファイル (内古閑、広川ら) の利用による候補解のクラスタリングも現在試行中である。

(3)システム生物学への大規模応用

バクテリアの細菌走化性系と、ヒト EGFR シグナル伝達系に関する大規模計算を実施した。細菌走化性では約 100×100 の網羅的計算を行った (JBCB 2009 論文)。EGFR 系では約 500×500 の網羅的計算を実施済である (発表準備中)。ドッキング計算に基づく相互作用予測では世界最大級の計算を実現したと考える。計算結果を Web 上でランキング形式で表示するツール群も開発した。

3. 現在までの達成度

①当初の計画以上に進展している。

(理由)

従来のようにスコア関数の上位解だけに注目するのではなく、候補解の空間分布に注目しようというアイデアは、多くのベンチマーク試験を通じてその有効性を示せた。

それらに加えて、rPSC 法の着想や、ランキング処理を多数の候補解集合に適用することなど、候補解を産出する計算方式でも大きな進展があり、先行する ZDOCK3.0 に匹敵する精度を約 4 倍の速度で実現できたことは計画以上の進展である。また $1000 \times 1000 =$ 百万通りの計算を行うと宣言した時

点では、まだ数百通りの計算すら膨大な時間がかかっていたが、様々な工夫と計算機の進歩により既に $500 \times 500 = 25$ 万通り級の計算までが実現している。この進展の速度は、当初の予想を大きく越えると評価して良い。

一方で、期間内にこのような超大規模処理が実現できてきた関係で、それらの膨大な計算結果からの知識抽出についての新たな興味深い課題も生じつつあると考えている。

なお本研究の成果に関連して平成 21 年度だけで 3 回の受賞を経験した (2 回は発表者に対する学生奨励賞、1 回は論文賞の内定)。これらは本研究の進捗状況が良好であることを、間接的に示す証左かと考えている。

4. 今後の研究の推進方策

EGFR 系では、東大宮野研究室で推定された関連遺伝子に基づき 1500×1500 規模の網羅的な構造ドッキングを実行する。複素 FFT 計算の部分を GPU 化し、東工大の Tsubame2.0 スパコン (H22 年 11 月稼働、最大 2.4PFLOPS) 上で定常的な計算サービスとして構築する。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

① Yuri Matsuzaki, Yusuke Matsuzaki, Toshiyuki Sato, Yutaka Akiyama, In silico screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis, *Journal of Bioinformatics and Computational Biology*, 7(6), 991-1012, 2009. 査読有
② Koki Tsukamoto, Tatsuya Yoshikawa, Yuichiro Hourai, Kazuhiko Fukui, Yutaka Akiyama, Development of an affinity evaluation and prediction system by using the shape complementarity characteristic between proteins, *Journal of Bioinformatics and Computational Biology*, 6(6), 1133-1156, 2008. 査読有

[学会発表] (計 37 件)

③ 大上雅史, 松崎裕介, 松崎由理, 佐藤智之, 秋山 泰, リンキングを用いたタンパク質ドッキングの精度向上と網羅的タンパク質間相互作用予測への応用, 情報処理学会バイオ情報学研究会 SIGBI0-20-3, 2010. 査読無
④ 松崎裕介, 大上雅史, 松崎由理, 佐藤智之, 関嶋政和, 秋山 泰, タンパク質の特性に基づく unbound ドッキングのための剛体予測手法の改良, 情報処理学会 SIGBI0-20-4, 2010. 査読無
⑤ 大上雅史, 松崎裕介, 松崎由理, 佐藤智之, 秋山 泰, 物理化学的相互作用の導入による網羅的タンパク質間相互作用予測システムの高精度化, 情報処理学会 SIGBI0-17-4, 2009. 査読無