

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 11 日現在

機関番号：82626

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23500374

研究課題名(和文) 基質結合部位予測に向けたタンパク質局所構造の高速比較法の開発

研究課題名(英文) Development of ultra-fast comparison method of protein local structures towards the prediction of ligand-binding sites

研究代表者

富井 健太郎 (TOMII, KENTARO)

独立行政法人産業技術総合研究所・ゲノム情報研究センター・研究チーム長

研究者番号：40357570

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究では、タンパク質基質結合部位の粗視化と高速なソートアルゴリズムの適用により、大量の結合部位の比較を可能とする新たな手法を開発した。タンパク質立体構造データベース全体の規模の既知及び潜在的基質結合部位に対して開発手法を適用し、その比較結果を収載したデータベースPoSSuMを構築と公開を行った。現在PoSSuMは、550万の既知及び潜在的基質結合部位を比較して得られた4,900万の類似結合部位ペアを収載するまでに成長している。また、ドラッグリポジショニング、副作用予測などへの応用を目指し、ChEMBLのアッセイ情報へのリンクを付加した新たなデータベースPoSSuMdsの構築と公開を行った。

研究成果の概要(英文)：In this study, we have developed a novel method that can perform a large-scale comparison of protein-ligand binding sites by utilizing a coarse-grained representation of binding sites and a fast sorting algorithm. Using the method, we conducted a large-scale all-pairs similarity search for both known and potential binding sites in the PDB, and constructed the database, called PoSSuM, that listed the discovered similar site pairs. PoSSuM has grown to provide information related to 49 million pairs of similar binding sites discovered among 5.5 million known and putative binding sites. For pharmaceutical applications, such as predictions of side effects and drug repositioning, we have provided a new database, PoSSuMds (PoSSuM drug search) to catalog detected known and potential binding sites for approved drug compounds whose assay data were retrieved from ChEMBL.

研究分野：生命情報学

キーワード：タンパク質 基質結合部位 機能部位 低分子医薬品

1. 研究開始当初の背景

(1) 構造生物学等の著しい進展により、機能未知のタンパク質立体構造情報が、非常な勢いで蓄積されつつある。こうした状況により、タンパク質立体構造情報に基づく計算機を用いた網羅的な機能情報の付加システムの開発が、現代の生命情報科学において欠かせない課題の一つとなっている。実際タンパク質立体構造データベース Protein Data Bank (PDB)には、研究開始当初既に 6(現在 10)万件以上のタンパク質立体構造が登録されており、それらの中の膨大な数の既知及び潜在的基質結合部位の網羅的な比較・分類は、タンパク質と基質の結合に関する新たな特徴発見を促し、効率かつ精度の良い予測法の開発につながるものと考えられる。

(2) タンパク質の基質結合部位の予測には、構造類似性比較法が多く利用されているが、既存手法の多くは、部位間の類似度の計算に時間を要する点がネックとなり、既知及び潜在的基質結合部位全体を用いた総当たり比較を行うことは困難であり、新たな手法の開発が待たれていた。そのため、これまでタンパク質立体構造の網羅的な分類等の研究に従事してきた申請者は、連携研究者らの開発した入力数に計算時間が比例する非常に高速なソートアルゴリズムを応用した基質結合部位の総当たり計算を可能にする技術の開発を考案するに至った。

2. 研究の目的

PDB 中の大量の既知及び潜在的基質結合部位を網羅的に比較・分類することは、タンパク質と基質の結合に関する重要かつ新たな特徴発見を促し、効率かつ精度の良い予測法の開発につながるものと考えられる。しかし既存技術による網羅的比較は多大な計算時間を要するため、新たな手法の開発が待たれている。この問題を解決するために、結合部位構造の粗視化により抽出される特徴量と連携研究者らが開発した強力なソートアルゴリズムを組み合わせる事で、結合部位の総当たり検索を迅速に実行可能な程度の高速な比較法の開発を行う。また、機能未知タンパク質に対する基質及びその結合部位予測への開発手法の応用を目指す。

3. 研究の方法

タンパク質基質結合部位の大量データに対する網羅的構造比較を達成するために、ソートアルゴリズムを応用する。このために、立体構造情報の高次元特徴ベクトル空間へのマッピング 符号つきランダム射影による次元圧縮と二値の数値列化 ソートアルゴリズムによる高速な探索という一連の試行と

改善を行い、新規高速比較手法を確立する。そして、開発手法を実際の PDB データに適用し、探索された結合部位ペアの妥当性や手法の高速化の度合い等について実効性の検証を行う。その後、開発手法による計算結果から得られた知見を整理したデータベースを構築し、公開する。また、機能未知のタンパク質に開発手法を適用し、基質及びその結合部位の予測を行う。さらに、結合部位の比較に基づく薬剤のオフ・ターゲット予測などに向けた研究を実施する。

4. 研究成果

(1) タンパク質の基質結合部位を構成アミノ酸残基の物理化学的特徴や二次構造などに応じて、高次元特徴ベクトル空間へのマッピングを行った。この特徴空間上で超高速近傍探索法 SketchSort の考え方を適用し、類似した性質をもつ結合部位ペアが類似した二値の数値列化されるように設定を行った。この数値列に高速なソートアルゴリズムを適用し、大量の比較から類似結合部位ペアを迅速に列挙可能な手法を確立した。

(2) 開発手法を PDB に登録されている既知および潜在的な基質結合部位からなるおよそ 120 万結合部位の比較に適用した。その結果、一般的な計算機でも実現可能(数十時間のオーダー)であることが明らかになった。

(3) 300 万を超える既知および潜在的な基質結合部位の比較から、タンパク質ファミリーやフォールド分類を超えて共通性をもつ新たな結合部位の例を発見した。また発見された既知および潜在的な基質結合部位の類似性から、幾つかの機能未知タンパク質の基質およびその結合部位予測を行った。この一部について、別の実験結果から正しい予測であることが明らかになった。

(4) 上記比較結果やタンパク質の構造および機能情報を付加したデータベース PoSSuM を構築し、そのウェブサイトを公開した。これまでに幾度か更新を行い、現時点では、およそ 550 万の既知および潜在的な基質結合部位の比較から、4,900 万の類似結合部位ペアの情報を収載したデータベースとなっている。

(5) 創薬研究などへの応用へ向け、PDB に登録されている低分子医薬品結合部位に着目した新たなデータベース PoSSuMds の構築と公開を行った。PoSSuMds では、既知低分子医薬品に関して、ChEMBL の化合物情報およびアッセイデータへのリンクが提供されており、結合部位の類似性とアッセイデータの両面から低分子医薬品の潜在的な基質結合部位への結合能の推定が可能となった。

5. 主な発表論文等

〔雑誌論文〕(計 4 件)

1. 池田 和由、富井 健太郎、PoSSuM: ポケット類似性情報に基づく合理的薬剤設計支援に向けて、日本薬学会 構造活性相関部会 SAR News、査読有、No. 28、2015、pp. 14-20、http://bukai.pharm.or.jp/bukai_kozo/news.html
2. Ito, J., Ikeda, K., Yamada, K., Mizuguchi, K., Tomii, K. PoSSuM v.2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs. Nucleic Acids Res. 査読有, Vol. 43 (Database issue), 2015, pp. D392-D398. DOI:10.1093/nar/gku1144
3. Ito, J., Tabei, Y., Shimizu, K., Tsuda, K., Tomii, K. PoSSuM: a database of similar protein-ligand binding and putative pockets. Nucleic Acids Res. 査読有, Vol. 40 (Database issue), 2012, pp. D541-D548. DOI:10.1093/nar/gkr1130
4. Ito, J., Tabei, Y., Shimizu, K., Tomii, K., Tsuda, K. PDB-scale analysis of known and putative ligand-binding sites with structural sketches. Proteins 査読有, Vol. 80, No. 3, 2012, pp. 747-63. DOI:10.1002/prot.23232

〔学会発表〕(計 13 件)

1. 富井 健太郎、創薬等支援に向けたタンパク質の配列・構造・機能・進化解析、産総研・東京大学合同 PDIS ワークショップ、2014 年 12 月 18 日、産総研臨海副都心センター (東京都・江東区)
2. 富井 健太郎、ChEMBLdb を利用したタンパク質の低分子医薬品結合部位の類似性比較、バイオインフォマティクスとゲノム医療、2013 年 11 月 20 日、産総研臨海副都心センター (東京都・江東区)
3. 池田 和由、PoSSuM Updates and Integration With ChEMBL For Application of Drug Reuse, CBI 学会 2013 年大会 生命医薬情報学連合大会、2013 年 10 月 28-31 日、タワーホール船堀 (東京都・江戸川区)
4. 富井 健太郎、FORTE, PoSSuM and beyond, 「タンパク質立体構造予測の最前線」ワークショップ、2013 年 2 月 21 日、お茶の水女子大学 (東京都・文京区)
5. 富井 健太郎、機能解析のためのタンパク質基質結合可能部位の網羅的比較解析、CBRC2012、2012 年 11 月 2 日、産総研臨海副都心センター (東京都・江東区)
6. 富井 健太郎、PoSSuM: a database for predicting protein-ligand interactions, JSBi2012 年年会 生命医薬情報学連合大会、2012 年 10 月 14-17 日、タワーホール船堀 (東京都・江戸川

区)

7. 伊東 純一、PoSSuM: a database for searching similar pairs of known and potential ligand-binding sites in proteins, 第 50 回日本生物物理学会年会、2012 年 9 月 23 日、名古屋大学 (愛知県・名古屋市)
8. 富井 健太郎、PoSSuM: a database of known and potential ligand-binding sites in proteins, Bioinformatics 2012 - SocBiN, 2012 年 6 月 11-14 日、Stockholm (Sweden)
9. 伊東 純一、タンパク質の既知及び未知基質結合部位データベース PoSSuM、平成 23 年度第 11 回産総研・産技連 LS-BT 合同研究発表会、2012 年 1 月 31 日、産総研つくば中央 (茨城県・つくば市)
10. 富井 健太郎、PoSSuM: database to find similar ligand binding pockets of proteins, CBRC2011、2012 年 1 月 5 日、産総研臨海副都心センター (東京都・江東区)
11. 富井 健太郎、A novel ultra-fast comparison method for known and potential ligand-binding sites of proteins, ISMB/ECCB 2011, 2011 年 7 月 18 日、Vienna (Austria)
12. 伊東 純一、Pdb-wide discovery of similar ligand binding sites using structural sketches, 3DSIG 2011, 2011 年 7 月 15 日、Vienna (Austria)
13. 伊東 純一、Structural Sketch を用いたタンパク質リガンド類似結合部位の高速列挙とそのデータベース構築、第 11 回日本蛋白質科学会年会、2011 年 6 月 8 日、ホテル阪急エキスポパーク (大阪府・吹田市)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況 (計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ

<http://possum.cbrc.jp/PoSsum/>

アウトリーチ活動

富井 健太郎、タンパク質の既知及び推定基質結合部位の網羅的類似探索、平成 24 年度 HPCI セミナー/情報生命科学特別講義 II、2013 年 1 月 18 日、産総研臨海副都心センター(東京都・江東区)

6. 研究組織

(1)研究代表者

富井 健太郎 (TOMII, Kentaro)

産業技術総合研究所・ゲノム情報研究センター・研究チーム長

研究者番号：4 0 3 5 7 5 7 0

(2)研究分担者

()

研究者番号：

(3)連携研究者

津田 宏治 (TSUDA, Koji)

東京大学大学院・新領域創成科学研究科・情報生命科学専攻・教授

研究者番号：9 0 3 5 7 5 1 7

清水 佳奈 (SHIMIZU, Kana)

産業技術総合研究所・ゲノム情報研究センター・主任研究員

研究者番号：6 0 3 6 7 0 5 0

伊東 純一 (ITO, Junichi)

医薬基盤研究所・特任研究員

研究者番号：