

研究開発課題別事後評価結果

1. 研究開発課題名

高精度タンパク質相互作用予測システムの開発

2. 代表研究者名

清水 謙多郎（東京大学大学院農学生命科学研究科 教授）

3. 代表研究者による成果開発概要

3-1. 研究目的

ゲノム解析、プロテオーム解析を通して、生体の内外における様々な現象が多面的な相関をもって関連づけられるようになった。同時に複数の現象をつなぐ情報・エネルギー等の流れも明らかとなってきたが、それらの多くは、タンパク質と他の分子の間の相互作用（以後、タンパク質間相互作用と総称する）を介したものである。しかしながら、生化学的研究手法やX線結晶構造解析、NMR解析などによるタンパク質間相互作用解析には、一般に複雑な手順・高額な装置・研究者の熟練等が必要である。一方、バイオインフォマティクス的手法を用いた現在の相互作用予測・解析システムでは、詳細な機能の解析を行うのに十分な精度を得ることが難しく、生命システムの解明には、タンパク質間相互作用の解析が重要であるが、生化学実験や構造決定などの実験的手法は、多くの時間とコストを必要とする。そこで本課題では、従来よりも高精度のタンパク質間相互作用予測システムの開発を行う。また、ドッキング予測と分子動力学シミュレーションを芳香環水酸化ジオキシゲナーゼ(ROS)のコンポーネント間の相互作用解析に適用し、それらの特異性と電子伝達能との関係を解析する。

3-2. 研究開発概要

本研究課題では、高速かつ高精度のタンパク質間相互作用予測システムを開発した。タンパク質-タンパク質、タンパク質-低分子の相互作用予測・解析を行うため、(1) 相互作用部位予測、(2) ドッキングシミュレーション、(3) 高精度複合体モデリングとダイナミクス解析の3つのアプローチから研究を行った。(1)については、タンパク質-タンパク質間相互作用予測の手法の開発を行い、さらに、タンパク質とリガンド、糖の相互作用部位予測、相互作用解析を行って、原子レベルに至るタンパク質間相互作用の総合的な研究を行った。タンパク質間相互作用予測、タンパク質-糖鎖結合予測については、タンパク質の配列情報のみから機械学習SVMを用いて予測する手法を開発した。また、タンパク質の構造が得られているとき、リガンドが結合する空間的位置を予測する手法を開発した。本件については、先行課題で開発した手法の改良を行うとともに、リガンドが既知の場合は、相互作用の条件をより詳細に設定して、予測精度の向上を図る手法を開発した。さらに、粗視化分子動力学シミュレーションにより、タンパク質-リガンドの相互作用を解析する手法を開発した。ドッキングシミュレーションについては、先行課題で、球面調和関数と新規に設計した正規直交基底関数での級数展開による高速内積計算を使ったアルゴリズムを開発しており、本研究課題では、剛体ドッキングの結果得られた多数の複合体構造（候補構造）の中からネイティブに近い構造を選択する手法の開発を行った。データベースについては、タンパク質-リガンド複合体、タンパク質-糖鎖結合タンパク質のデータベースを

構築した。

実験解析グループでは、3種の細菌に由来する電子伝達系の異なるカルバゾール水酸化酵素 (CARDO; 芳香環水酸化ジオキシゲナーゼ [Rieske non-heme iron oxygenase] の一種) を材料に、コンポーネント間相互作用の機能・構造解析を行い、ドッキングシミュレーションの結果との比較・検討を行った。

3 - 3 . 研究成果概要

タンパク質の配列情報のみから機械学習 SVM を用いて糖鎖結合タンパク質を予測する手法では、配列特徴として、アミノ酸の3つ組の出現頻度を用い、その結果、AUC の値で 0.80 を超える高い予測精度が得られることができた。さらに、糖鎖結合タンパク質の分類予測の手法を開発しており、C型レクチンの予測に関しては accuracy の値として 83.0% という結果を得ている。タンパク質-リガンド相互作用部位予測では、タンパク質周辺にプローブ点を配置し、タンパク質とそれらの間の相互作用エネルギー値が低いプローブ点をクラスターリングしてリガンドの存在する空間的位置を予測する手法を開発した。従来、bound 構造からの予測で precision ≥ 0.25 (25%以上、予測部位と実際のリガンドの存在部位が重なっているもの) の割合が 0.8 であったところ、0.857 に向上させることができた。リガンドが既知の場合、相互作用を調整することにより、予測精度を向上させることができることを示した。また、粗視化分子動力学シミュレーションにより、タンパク質-リガンドの相互作用を解析する手法を開発し、実際のリガンド結合部位を予測するとともに、リガンド周辺の溶媒分子の動きから結合機序を解析することができた。ドッキングシミュレーションについては、先行課題で開発したシステムが生成する候補構造に対し、精密な相互作用エネルギーの計算を行い、それらの係数の最適化を図ることで、予測順位の改善を試みた。その結果、上位 2,000 中に near-native 構造があったターゲットの多くについて順位を改善することができた。タンパク質-リガンド複合体、タンパク質-糖鎖結合タンパク質のデータベースを構築した。これらのデータベースは、構造が既知のタンパク質において、相互作用部位、相互作用の種類、単体・複合体の構造、他のデータベースへのリンク情報を登録している。タンパク質-リガンド相互作用については、PDB のすべてのリガンド複合体構造に対して、octree で構造を表現し、様々な条件で高速にリガンドの検索が行えるようにした。また、実験研究者と連携して、芳香環水酸化ジオキシゲナーゼ (Rieske non-heme iron oxygenase, ROS) のコンポーネント間相互作用・電子伝達機構の解明に関する研究を行った。

4 . 事後評価結果

4 - 1 . 当初計画の達成度

当初の研究開発計画から見た進捗状況や達成度等

タンパク質相互作用予測については高い予測精度を得ることができ、当初の計画通りに進んだといえる。

当初計画では想定されていなかった新たな展開

分子動力学シミュレーションに粗視化モデルを導入したこと、脂質結合タンパク質予測の手法の開発も行ったことなど、計画外の進展があったがこれらは好ましいものと評価できる。

4 - 2 . 知的財産権、外部発表（論文等）等研究開発成果の状況

外部発表（論文等）

論文、学会発表を含めて成果の情報発信が活発になされている。未発表になっている手法、ツールに関する論文も早急に発表することを期待する。

データベースの公開状況

成果プログラム5つ中2つが未公開であり、全体として公開が遅れている。

知的財産権

特許出願はなかった。本研究分野では、特許化より、論文やサーバなどによる公開の方が、成果の情報発信としての有効であると考えられる。

4 - 3 . 研究開発成果の公開による波及効果

データベース、プログラムや論文、学会発表など、これまでの成果公開によるバイオインフォマティクス領域およびライフサイエンス分野全体への影響や効果について

タンパク質相互作用関連の予測に関しては既存手法より高い予測精度を出しており、当該分野の発展に影響を与えつつあると思われる。今後のさらなる発展が期待される。

国内外のデータベース高度化・標準化の動向と比較して本課題の意義

相互関連の予測手法は従来の方法論の延長線上にあるが、細かな改良が施されており既存手法を上回る予測精度が達成されているので高く評価できる。糖鎖結合タンパク質に特化したデータベースは他にもほとんどないと思われ有用性が高いと思われる。

4 - 4 . 成果の実用化の可能性及び成果から予想される波及効果

成果を今後発展展開することによりバイオインフォマティクス領域およびライフサイエンス分野全体に影響や効果を及ぼすと期待できるか。

開発した予測手法、システムは従来の方法論に基づいているものの既存手法より良い予測精度をあげており、バイオインフォマティクス研究の発展に貢献することが期待できる。また、開発された予測プログラムについては、今後引き続き予測精度や使い勝手の改善を行いながら、利用しうる研究者へ周知する広報の努力も必要である。

今後、成果の生命科学、研究基盤、産業への貢献度

タンパク質-タンパク質相互作用データベースやタンパク質-リガンド相互作用データベースは、今後の創薬などの産業への応用が期待される。

4 - 5 . 総合評価

研究開発計画通りの成果が見られ、ライフサイエンス分野の情報基盤整備に貢献した。

粗視化モデルを導入した分子動力学シミュレーションや、糖鎖との相互作用など、タンパク質に関わる独自の相互作用予測方法を開発し、高い精度を達成している点が評価できる。

5 . 主な論文発表等

1. W. Cao, K. Sumikoshi, T. Terada, S. Nakamura, K. Kitamoto, K. Shimizu: Computational Protocol for Screening GPI-anchored Proteins, *Proceedings of the First International Conference on Bioinformatics and Computational Biology (BICoB), Springer Lecture Notes in Bioinformatics Series*, **5462**, 164-175 (2009). (査読あり)
2. J. Inaba, S. Nakamura, K. Shimizu, T. Asami, Y. Suzuki: Anti-metatype peptides, a new molecular tool to monitor small ligands at high sensitivity and specificity. *Analytical Biochemistry*, **388**, 63-70 (2009). (査読あり)
- *3. W. Cao, J. Maruyama, K. Kitamoto, K. Sumikoshi, T. Terada, S. Nakamura, K. Shimizu.: Using a new GPI-anchored-protein identification system to mine the protein databases of *Aspergillus fumigatus*, *Aspergillus nidulans*, and *Aspergillus oryzae*, *J. Gen. Appl. Microbiol.* **55**, 5, 381-393 (2009). (査読あり)
4. S. Nakamura, K. Shimizu: Comprehensive analysis of sequence-structure relationships in the loop regions of proteins, *GIW 2009*, **23**, 106-116 (2009). (査読あり)
- * 5. S. Yamasaki, T. Terada, K. Shimizu, H. Kono, A. Sarai: A Generalized Conformational Energy Function of DNA Derived from Molecular Dynamics Simulations, *Nucleic Acids Res.*, **37**, e135 (2009). (査読あり)
- *6. R. Jono, Y. Watanabe, K. Shimizu, T. Terada: Multicanonical ab initio QM/MM Molecular Dynamics Simulation of a Peptide in an Aqueous Environment, *J. Comput. Chem.*, **31**, 6, 1168-1175 (2009). (査読あり)
7. K. Kadota, Y. Nakai, K. Shimizu: Ranking differentially expressed genes from Affymetrix gene expression data: methods with reproducibility, sensitivity, and specificity, *Algorithms for Molecular Biology*, **4**, 7 (2009). (査読あり)
8. T. Furuta, K. Shimizu, T. Terada: Accurate prediction of native tertiary structure of protein using molecular dynamics simulation with the aid of the knowledge of secondary structures, *Chemical Physics Letters*, **472**, 134-139 (2009). (査読あり)
9. K. Inoue, Y. Ashikawa, T. Umeda, M. Abo, J. Katsuki, Y. Usami, H. Noguchi, Z. Fujimoto, T. Terada, H. Yamane, H. Nojiri: Specific interactions between the ferredoxin and terminal oxygenase components of a class IIB Rieske nonheme iron oxygenase, carbazole 1,9a-dioxygenase. *Journal of Molecular Biology*, **392**, 436-451 (2009). (査読あり)
- *10. S. Someya, M. Kakuta, M. Morita, K. Sumikoshi, W. Cao, Z. Ge, O. Hirose, S. Nakamura, T. Terada, K. Shimizu: Prediction of carbohydrate-binding proteins from sequences using support vector machines, *Advances in Bioinformatics 2010*, 289301 (2010). (査読あり)
11. K. Urano, T. Daimon, Y. Banno, K. Mita, T. Terada, K. Shimizu, S. Katsuma, T. Shimada: Molecular defect of isovaleryl-CoA dehydrogenase in the skunk mutant of silkworm, *Bombyx mori*, *FEBS Journal*, **277**, 4452-4463 (2010). (査読あり)
- *12. O. Hirose, K. Shimizu: A self-organizing state space approach to infer time-varying causalities between regulatory proteins. Proceedings of International Conference on Information Technology in Bio- and Medical Informatics, *Lecture Notes in Computer Science*, **6266**, 158-171 (2010). (査読あり)

13. T. Hagiwara, S. Saito, Y. Ujiie, K. Imai, M. Kakuta, K. Kadota, T. Terada, K. Sumikoshi, K. Shimizu, T. Nishi: Development of a retention time prediction method for HPLC and application to metabolome analysis, *Bioinformatics*, **5**, 6, 255-288 (2010). (査読あり)
14. T. Umeda, J. Katsuki, Y. Ashikawa, Y. Usami, K. Inoue, H. Noguchi, Z. Fujimoto, H. Yamane, H. Nojiri: Crystallization and preliminary X-ray diffraction studies of a ferredoxin reductase component of carbazole 1,9a-dioxygenase from *Novosphingobium* sp. KA1. *Acta Cryst.*, **F66**, 712-714 (2010). (査読あり)
15. T. Umeda, J. Katsuki, Y. Ashikawa, Y. Usami, K. Inoue, H. Noguchi, Z. Fujimoto, H. Yamane, H. Nojiri: Crystallization and preliminary X-ray diffraction studies of a terminal oxygenase of carbazole 1,9a-dioxygenase from *Novosphingobium* sp. KA1. *Acta Cryst.*, **F66**, 1480-1483 (2010). (査読あり)
16. 野尻秀昭: 環境細菌の難分解性環境汚染物質分解能発現機構の解明, 極限環境生物学会誌, **9**, 4-10 (2010). (査読あり)
17. M. Morita, T. Terada, S. Nakamura, K. Shimizu: BUDDY-system: A web service for constructing a dataset of protein pairs between ligand-bound and unbound states, *BMC Research Notes*, **4**, 143 (2011). (査読あり)
18. K. Sumikoshi, T. Terada, S. Nakamura, K. Shimizu: Protein-Protein Docking Using Multi-layered Spherical Basis Functions, *Proceedings of the 2011 International Conference on Bioscience, Biochemistry and Bioinformatics 2011*, 342-347 (2011). (査読あり)
19. H. Nojiri: Structural and molecular genetic analyses of the bacterial carbazole degradation system, *Biosci. Bioelectron. Biochem.*, in press. (査読あり)

* は当課題の成果として主要な論文

参考

論文発表	国内 1 件、海外 18 件
招待、基調講演	国内 1 件、海外 1 件
口頭発表	国内 15 件、海外 1 件
ポスター発表	国内 28 件、海外 3 件
特許	国内 0 件、国際 0 件

成果プログラムおよびデータベース等

タンパク質-リガンド結合データベース

PDB に登録されたタンパク質-リガンド複合体構造に関して、単体のタンパク質から複合体、リガンドから複合体、複合体から単体のタンパク質とリガンドを、さまざまな付加条件で検索できるようにした汎用的なシステム。登録数は、1,683,940 タンパク質-リガンドペア。今後は、収集したデータをもとにした分類、解析を行う予定である。

平成 21 年 10 月 公開、<http://www.bi.a.u-tokyo.ac.jp/services/buddy/current/>

アクセス数 3,773 件 (ジョブ実行数 497 件)

糖鎖結合タンパク質データベース

PDB に登録されたタンパク質-糖鎖複合体構造に関して、配列・構造上の結合部位、アノテーション情報、機能分類、他のデータベースへのリンク、糖鎖結合部位に関する統計データなどを登録し、3D Viewer を用いて各複合体の立体構造を表示できるようにしている。登録しているタンパク質-糖鎖複合体構造は 1,148 で、現在の PDB のアノテーションに UniprotKB のアノテーションも加えてこの数になった。

タンパク質-リガンド結合データベース

PDB に登録されたタンパク質-リガンド複合体構造に関して、構造を octree で表現し、種々の条件で結合部位を効率的に検索できるようにしている。現在、このデータセットを用いて、糖鎖結合部位予測、脂質結合部位予測、リガンド結合部位における disorder 解析などの研究を行っている。登録しているタンパク質-リガンド複合体構造は 52,792。
未公開（平成 24 年 3 月までに公開予定）

タンパク質-リガンド結合部位予測プログラム

与えられたタンパク質の構造に対して、リガンドが結合する空間的位置を予測するプログラム。リガンド一般について予測する仕様であるが、現時、リガンドが既知の場合、さらに予測精度を向上させる手法を開発している。

糖鎖結合タンパク質予測プログラム

与えられたタンパク質が糖鎖と結合するかどうかを、アミノ酸配列情報のみから予測するプログラム。

平成 23 年 10 月公開、<http://bolero.bi.a.u-tokyo.ac.jp:8201/Lectin-Predictor/>
備考

旧バージョンは、平成 22 年 5 月 27 日に公開していたが、さらに手法の改良を行い、上記「糖鎖結合タンパク質データベース」と合わせて公開した。旧バージョンのアクセス数は、ログが消滅したため不明であるが、50 件以上であったことは確かである。