

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 3 日現在

機関番号：84420

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500277

研究課題名（和文）ペア予測から蛋白質—蛋白質相互作用の予測と分類

研究課題名（英文）Prediction of protein-protein interactions from residue-pair interaction prediction

研究代表者シャンダー アハマド（SHANDAR AHMAD）

独立行政法人医薬基盤研究所・創薬基盤研究部・研究員

研究者番号：80463298

研究成果の概要（和文）：蛋白質間相互作用部位及び相互作用相手の予測法と、蛋白質複合体を分類するための類似性指標を開発した。方法の中心はニューラルネットワークのアンサンブルを用いる点にある。予測モデルの精度は、クエリー毎にリアルタイムでモデルのトレーニングを行なうことで、著しく向上した。また、訓練データ内に蛋白質 Q が存在することで、蛋白質 P の予測精度がどれだけ向上するかをもって、蛋白質 PQ 間の類似性指標とした。ゲノム解析への応用のため、より高速な予測モデルを開発した。

研究成果の概要

（英文）：Methods for prediction of protein-protein interaction sites, interacting partners and for defining similarity between protein-protein complexes for classification have been developed. Core method is based on a neural network ensemble. Accuracy of the model is significantly improved by real-time training of models over query-specific data sets. Improvement in prediction accuracy of a protein P due to the presence of a protein Q in the training data is used as pair-wise similarity metric. A faster version of the model is developed for genomic scale application.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	1,700,000	510,000	2,210,000
2011 年度	800,000	240,000	1,040,000
2012 年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学。生体生命情報学

キーワード：バイオインフォマティクス

1. 研究開始当初の背景

配列に基づく蛋白質—蛋白質相互作用予測の問題には少なくとも二つの要素がある。す

なわち（1）結合部位予測のないセットから相互作用蛋白質パートナーを見つけること、及び（2）単一の配列から結合部位を見つけ

ること。これらの問題に対処する方法は主に単独で開発されてきた。このプロジェクトの主要な指標は蛋白質-蛋白質相互作用のあらゆる側面を予測する総合的アプローチを開発することである。

2. 研究の目的

主な目的は：

- (1) 単一の蛋白質ペアにおける残基ペアの相互作用の予測ができるシステムを開発すること。
- (2) 各蛋白質の相互作用部位を予測するためにペアワイズ予測を用いること。
- (3) 二つの蛋白質が実際に相互作用するかを予測するためにペアワイズ予測を用いること。
- (4) 分類するために使われた蛋白質間の類似性の指標としてそれぞれの結合

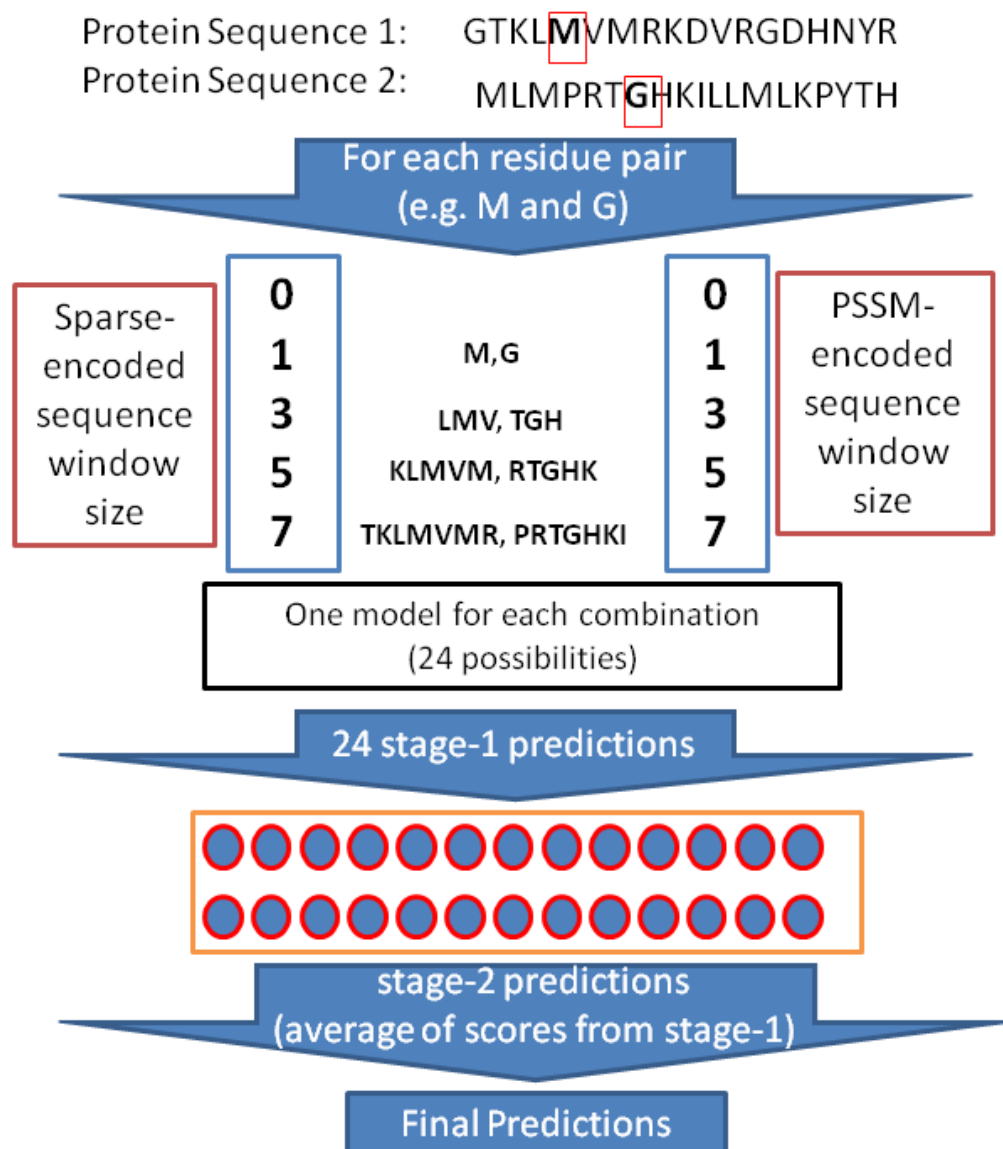
部位の予測可能性を用いること。

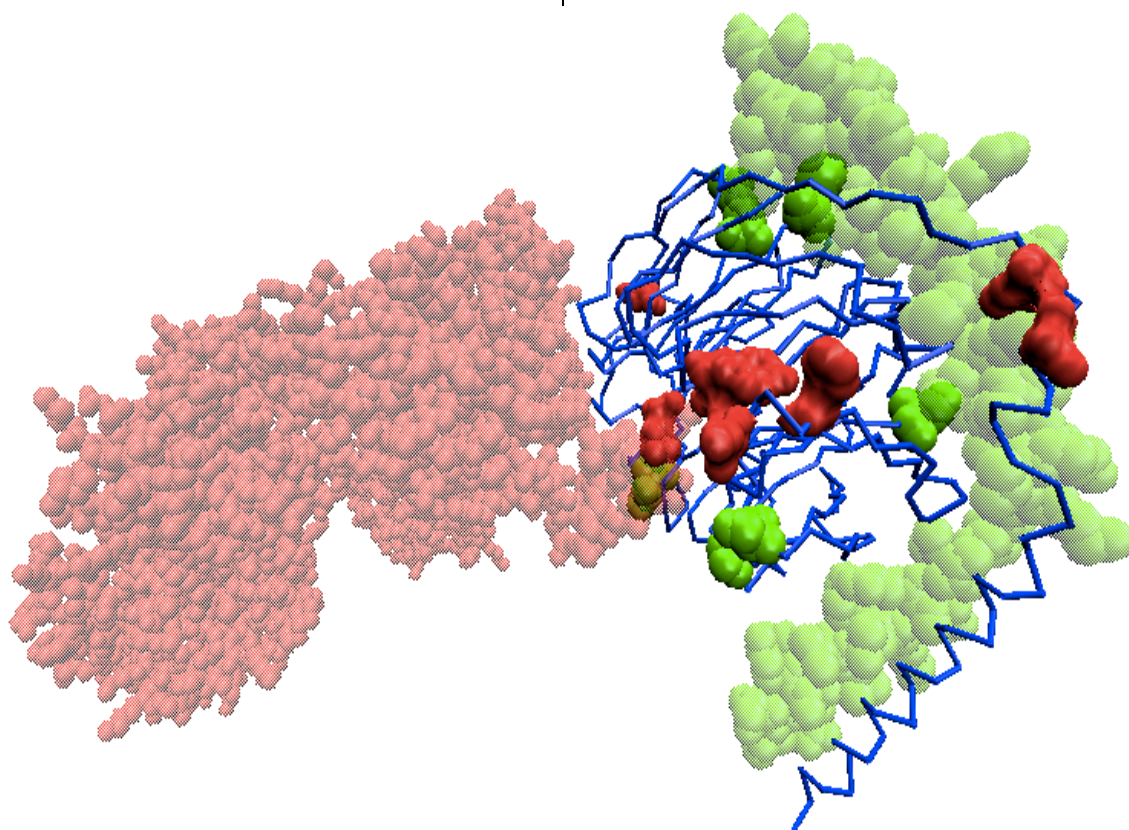
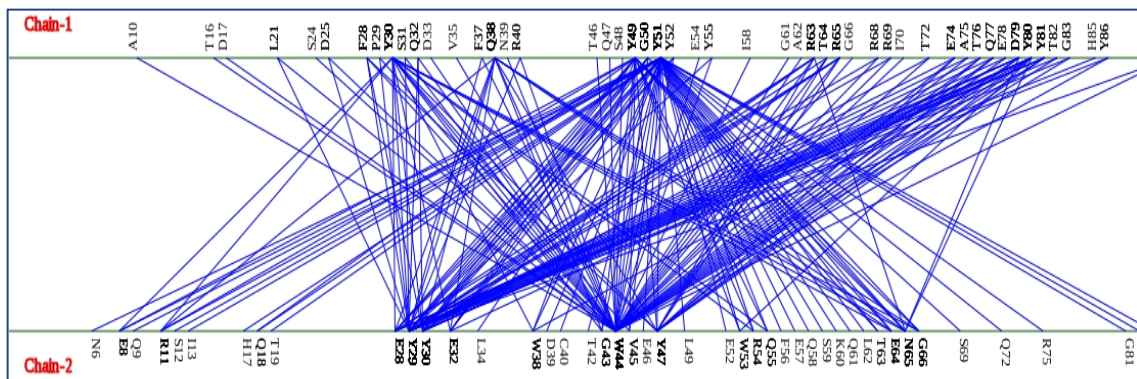
3. 研究の方法

相互作用残基ペア予測のためのニューラルネットワークに基づいたモデルを開発した。我々のアプローチは信頼度を付与するために複数の方法を開発することとそれらの出力値を平均化することであった。相互作用パートナーを予測するため、クエリ蛋白質の予測結合部位の進化的特徴と既知の相互作用のデータベースの相関が用いられた。

4. 研究成果

配列のみから相互作用残基ペア予測ができる、ニューラルネットワークアンサンブルが開発された。これは結合部位予測におけるデ





一タ不均衡に対処する新しいアプローチである。

唯一の入力データが蛋白質のアミノ酸配列ペアの情報であるウェブサーバーが開発された。我々の新しいアプローチを用いて、起こり得る全ての相互作用残基ペアが高精度で予測された。

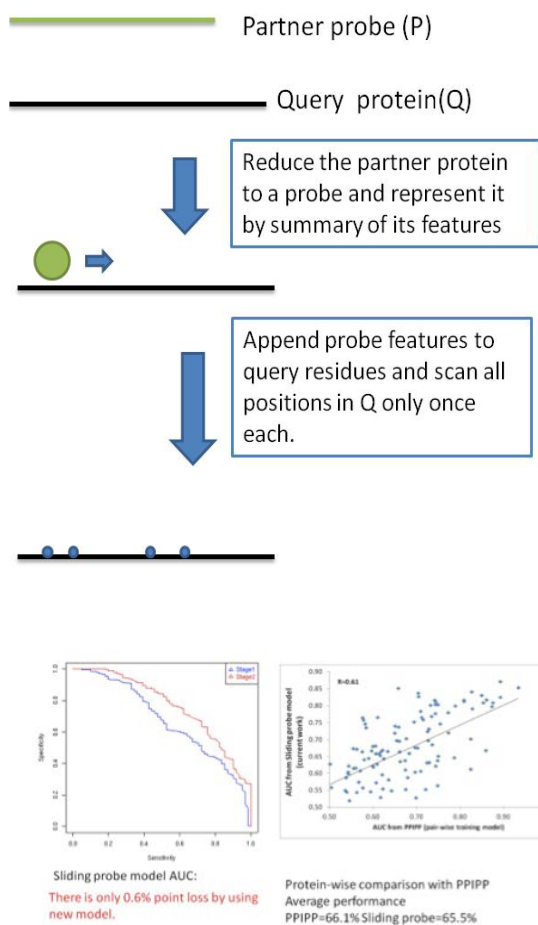
このモデルは 73% の高精度を実現した。
(AUC-ROC 曲線下面積により測定)

これらのペアワイズ予測を単一の蛋白結合部位に変換する手法も開発され、これは蛋白質-蛋白質相互作用部位を予測する全ての

既存手法を上回っていることが示された。また、ペアワイズ予測は異なるパートナーの異なる結合部位を予測することができることもわかった。これは本研究以前のいかなる予測の手法でも可能ではなかった。

速度を増すために、残基が互いにペアにならない、特定パートナー結合部位予測のさらなる方法が開発された。その代り、予測モデルを開発するために一つの蛋白質残基特徴が全パートナー蛋白質の特徴と結合された。このモデルは非常に速く、ペアワイズモデルと同等レベルの性能を有する。

このモデルは、オール対オールの残基ペアから始まったオリジナルモデルに十分匹敵する。



スライド式プローブモデルを用いて、全ゲノム（酵母）の蛋白質-蛋白質相互作用パートナーが予測された。蛋白質-蛋白質相互作用パートナーを予測するこの手法は非常に有効であることがわかり、80%に近い精度を確保している。（詳細は投稿準備中）ペアワイズ相互作用モデルは同じタスクにおいて性能ははるかに優れているが速度が遅く、相互作用パートナーの高精度推定のために少数の相互作用に応用することができる。

相互作用残基ペアの予測からの蛋白質-蛋白質複合体間の類似性の定義の新しい測定基準も開発された。この方法を用いて複合体間の多くの新しい関連性が検出された。成果は発表に向けて準備中である。

また、もしトレーニングの開始データセット

が類似のトレーニング例から構成されているならば結合部位予測モデルの性能は有意に改善されることがこの方法により示された。予測精度を改善するためのこの方法は、**Transfer of Training** フレームワークと呼ばれ、投稿準備中である。

5. 主な発表論文等

（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7件）

- (1) Nagarajan R, Ahmad S, GromihaM, Novel approach for selecting the best predictor for identifying the binding sites in DNA binding proteins, *Nucleic Acids Research* (In Press) 2013 (査読有)
- (2) Tiwari P, Tripathi LP, Nishikawa-Matsumura T, Ahmad S, Song SN, Isobe T, Mizuguchi K, Yoshizaki K. Prediction and experimental validation of a putative non-consensus binding site for transcription factor STAT3 in serum amyloid A gene promoter. *Biochim Biophys Acta.* 2013 Jun;1830(6):3650-5. (査読有)
<http://dx.doi.org/10.1016/j.bbagen.2013.01.024>
- (3) Hutchins AP, Diez D, Takahashi Y, Ahmad S, Jauch R, Tremblay ML, Miranda-Saavedra D. Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions. *Nucleic Acids Res.* 2013 Feb 1;41(4):2155-70. (査読有)
<http://dx.doi.org/10.1093/nar/gks1300>
- (4) Ahmad S, Mizuguchi K. Partner-aware prediction of interacting residues in protein-protein complexes from sequence data. *PLoS One.* 2011;6(12):e29104. (査読有)
<http://doi.org/10.1371/journal.pone.0029104>
- (5) Ahmad S, Sarai A. Analysis of electric moments of RNA-binding proteins: implications for mechanism and prediction. *BMC Struct Biol.* 2011 Feb 1;11:8. (査読有)
<http://doi.org/10.1186/1472-6807-11-8>

- (6) Fernandez M, Kumagai Y, Standley DM, Sarai A, Mizuguchi K, Ahmad S. Prediction of dinucleotide-specific RNA-binding sites in proteins. BMC Bioinformatics. 2011 Nov 30;12 Suppl 13:S5. (査読有)
<http://doi.org/10.1186/1471-2105-12-S13-S5>
- (7) Firoz A, Malik A, Joplin KH, Ahmad Z, Jha V, Ahmad S. Residue propensities, discrimination and binding site prediction of adenine and guanine phosphates. BMC Biochem. 2011 May 13;12:20. (査読有)
<http://doi.org/10.1186/1471-2091-12-20>

[学会発表] (計 7 件)

- (1) A meta-analysis of public microarray datasets to detect differential expression of heat shock proteins in children, Shandar Ahmad, Biomarker discovery conference, Shoal Bay, Australia, December 13, 2012
- (2) A sliding-probe model for predicting partner aware protein-protein interaction sites, Shandar Ahmad, Kenji Mizuguchi, Genome Informatics Conference (GIW2012), Taiwan, December 6, 2012
- (3) Statistical analysis and prediction of kinked residue positions in helical membrane proteins, Shandar Ahmad, Kenji Mizuguchi, Japanese Society of Bioinformatics (JSBI) annual meeting 2010, Kyushu University, Fukuoka, December 13, 2010
- (4) Predicting Bio-molecular interactions in proteins and nucleic acids, International Interdisciplinary science conference, November 16, 2011, Jamia Millia Islamia University, New Delhi, India (invited talk by Shandar Ahmad)
- (5) Sequence-to-structure prediction errors can reveal conformational flexibility in helical membrane proteins, Shandar Ahmad, Hemjit Singh, Yogesh Paudel, Takaharu Mori, Yuji Sugita and Kenji Mizuguchi, Intelligent Systems in Molecular Biology (ISMB 2010), July 11, 2010, Boston, USA
- (6) Prediction of dinucleotide-specific RNA-binding sites in proteins using sequence and evolutionary information.

Michael Llamosa Fernandez, Yutaro Kumagai, Daron Standley, Akinori Sarai, Kenji Mizuguchi, Shandar Ahmad, Intelligent Systems in Molecular Biology (ISMB 2010), July 11, 2010, Boston, USA

- (7) Kinase Inhibition Modeling by Proteochemometrics and Docking Analysis. Michael Llamosa Fernandez, Akinori Sarai and Shandar Ahmad, Intelligent Systems in Molecular Biology (ISMB 2010), July 12, 2010, Boston, USA

[その他]

ホームページ等

相互作用残基ペア及び単一の蛋白質相互作用部位を予測するウェブサーバーが開発され公開されている。

<http://tardis.nibio.go.jp/netasa/ppipp/>

さらなるソフトウェアとデータを公開予定。現在、投稿準備中。

6. 研究組織

(1) 研究代表者 シャンダー アハマド

(SHANDAR AHMAD)

独立行政法人医薬基盤研究所・創薬基盤研究部・研究員

研究者番号：80463298