1ZF-8

立体構造情報を用いたドッキング計算による大規模タンパク質-RNA 間相互作用予測手法

大上 雅史 † 松崎 由理 † 秋山 泰 †
† 東京工業大学 大学院情報理工学研究科 計算工学専攻

1 序論

タンパク質 RNA 間相互作用（Protein-RNA Interaction; PRI）は細胞内システムの理解に重要であり，RNA 結合タンパク質の結合対象となる RNA を認識する機構を解明することは計算生物学における課題の 1 つとなっている [1]．我々はタンパク質同士の相互作用に対し，タンパク質剛体ドッキング手法を用いた予測手法を既に提案しているが，本研究では当該システムをタンパク質-RNA 間相互作用予測に拡張し，78 例のタンパク質-RNA 複合体に対する 78×78 の all-to-all 予測を行った．元の 78 の結合ペアを正例とし，その他の 6,006 ペアを負例として予測精度を検証した結果，F 値 0.465 での予測に成功し，また偽陽性と判定されたペアの中には十分相互作用可能性の検証の余地があるペアが発見された．

2 方法

2.1 MEGADOCK

MEGADOCK システム [2] は，我々が開発している網羅的タンパク質間相互作用予測システムであり，FFT によるタンパク質剛体ドッキング手法を利用している．real Pairwise Shape Complementarity(rPSC) と呼ばれる形状相補性スコア [2] と，CHARMm19 原子電荷パラメータと FTDock 静電ポテンシャル [3] による静電的相互作用スコアを組み合わせることで，高速なドッキング計算を可能としている．

MEGADOCK システムはタンパク質間相互作用予測のために開発したものであり，PDB 形式で構造情報が得られている分子は原理上全て扱うことができる．しかし，CHARMm19 ではアミノ酸関連原子の電荷値しか定義されていないため，RNA などのタンパク質以外の分子ではドッキング計算における静電的相互作用項が計算できない．そこで本研究では RNA を扱うために，デオキシリボ核酸およびリボ核酸関連原子の電荷値が定義されている CHARMm27 の原子電荷パラメータを用いて，MEGADOCK システムのドッキング計算を拡張した．

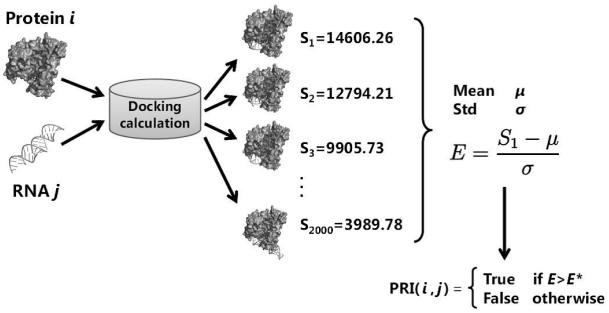

図 1: MEGADOCK による PRI 予測の概略

2.2 タンパク質-RNA 相互作用予測法

図 1 に本稿で提案するタンパク質-RNA 間相互作用 (PRI) 予測法の概略を示す．まず，予測対象のタンパク質と RNA のペアに対して剛体ドッキング計算を行う．剛体ドッキング計算では，RNA の結合位置の良さを表す評価値であるドッキングスコア S_k が得られ，そのドッキングスコアの良いものから順に 2,000 個複合体構造が出力される．それらのドッキングスコアを元に，以下の式によって PRI 予測のための評価値，E-value を決定する．

$$E = \frac{S_1 - \mu}{\sigma} \tag{1}$$

ここで S_1 は 1 位のスコア，μ と σ はそれぞれ S_1 から $S_{2,000}$ までの 2,000 個のドッキングスコアから計算される平均と標準偏差である．この E-value の値から，ある閾値 E^* を超えるかどうかで相互作用の判定を行う．

$$\mathrm{PRI}(i,j) = \begin{cases} \text{True} & \text{if } E > E^* \\ \text{False} & \text{otherwise} \end{cases} \tag{2}$$

2.3 データセット

Protein Data Bank（PDB）に含まれるタンパク質–RNA 複合体から，PISCES サーバ [4] によって，解像度 3.0Å 以下，相互の配列相同性が 30% 以下の X 線結晶構造のみを含むデータセットを構築した．得られたタンパク質–RNA 複合体の数は 78 個であった．データセットに含まれる PDB ID と RNA の種類を表 1 に示す．

A Large-scale Protein-RNA Interaction Prediction Method by Using Docking Calculation

† Masahito OHUE(ohue@bi.cs.titech.ac.jp)
† Yuri MATSUZAKI(y_matsuzaki@bi.cs.titech.ac.jp)
† Yutaka AKIYAMA(akiyama@cs.titech.ac.jp)
Graduate School of Information Science and Engineering, Tokyo Institute of Technology (†)
Ookayama 2-12-1-W8-76, Meguro-ku, Tokyo, 152-8550 Japan.

Copyright ©2011 Information Processing Society of Japan.
All Rights Reserved.

表 1: 78 個のタンパク質-RNA 複合体 PDB ID リスト

RNA type	Complex PDB ID
tRNA	1ASY, 1B23, 1F7U, 1FFY, 1H3E, 1H4S, 1K8W, 1N78, 1Q2R, 1QF6, 1QTQ, 1R3E, 1SER, 1TFW, 1U0B, 2AZX, 2B3J, 2BTE, 2CT8, 2FK6, 2FMT, 2GJW, 2I82, 2R8S, 2ZZM, 3EPH, 3FOZ
mRNA	1KNZ, 1M8X, 1WPU, 1WSU, 1ZBH, 2ANR, 2F8K, 2HW8, 2IPY, 2J0S, 2PJP, 3K62
rRNA	1FEU, 1MZP, 2ASB, 2BH2, 3AEV
ssRNA	1FXL, 2BX2, 2JLV, 2R7R, 2VNU, 3FHT, 3I5X, 3IEV
dsRNA	1N35, 2AZ0, 2NUG, 2ZKO, 3EQT
siRNA	1SI3, 2BGG, 2F8S, 2ZI0
SRP RNA	1HQ1, 1JID, 1LNG
viral RNA	1F8V, 2E9T, 2QUX, 3BSO
RNA aptamer	1OOA, 3DD2, 3EGZ
others	3IAB(ribozyme), 2GXB(Z-RNA), 1A9N, 2OZB(snRNA), 1SDS, 3HAX(snoRNA)

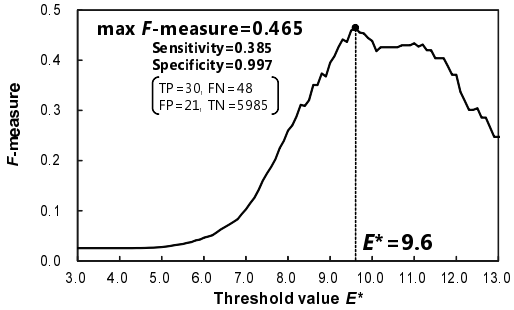

図 2: 78×78 予測結果（閾値 E^* に対する F-measure）

3 結果と考察

表 1 の 78 組のタンパク質-RNA から 78 × 78 通りの全ての組み合わせについて，MEGADOCK による PRI 予測を行った．タンパク質-RNA ペアの相互作用予測が正しいかどうかは，元々の結合ペア同士のみ正例とし，それ以外のペアは負例と仮定して判定した．予測結果が真陽性，真陰性，偽陽性，偽陰性のペアの数をそれぞれ TP，TN，FP，FN とし，予測性能は感度と選択度の調和平均である F-measure によって評価した．F-measure は次式で表される．

$$F\text{-measure} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

予測結果を決める E-value の閾値 E^* については，3.0 から 0.1 刻みで 13.0 まで変化させて精度を確認した．結果を図 2 に示す．$E^* = 9.6$ のときに F-measure が最大となり，$F = 0.465$ を得た．図 3 はこの相互作用予測結果に対する ROC 曲線である．図 3 における対角線はランダムに相互作用を予測したときの結果に相当し，MEGADOCK による予測がランダムな相互作用の予測に比べて有意に精度が優れていることが分かる．図 4 は 78 × 78 の全てのペアの予測結果をヒートマップ形式で表したものである．E-value が E^*，即ち 9.6 を超えるものは黒いセルで表されており，予測における正解である対角線上のセルをうまく予測できていることが視覚的に見て取れる．

4 結論

本稿では剛体ドッキング計算に基づくタンパク質-RNA 間相互作用予測手法を提案し，78 × 78 通りの網羅的予測を行い，精度を確認した．今後の課題としては，偽陽性と判定されたペアに対する相互作用可能性の検証，アンサンブルドッキングとの組み合わせなどが挙げられる．

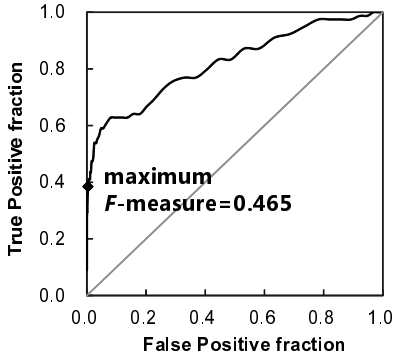

図 3: 78 × 78 予測結果における ROC 曲線

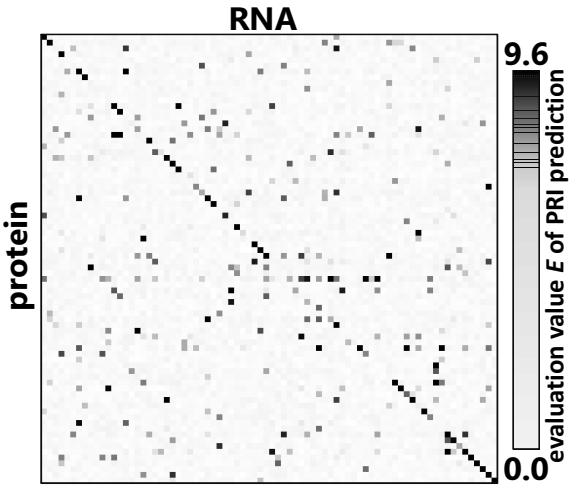

図 4: 78 × 78 PRI 予測結果マップ．セルの順番は縦横それぞれ PDB ID の辞書的昇順である．

参考文献

[1] Peréz-Cano, L., Solernou, A., Pons, C. and Fernández-Recio, J., Structural Prediction of Protein-RNA Interaction by Computational Docking with Propensity-based Statistical Potentials, *Pac Symp Biocomput 2010*, 293-301, 2010.

[2] 大上雅史，松崎由理，松崎裕介，佐藤智之，秋山 泰，MEGADOCK：立体構造情報からの網羅的タンパク質間相互作用予測とそのシステム生物学への応用，情処論数理モデル化と応用, 3(3): 91-106, 2010.

[3] Gabb, H.A., Jackson, R.M. and Sternberg, M.J.E., Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information, *J Mol Biol*, 272: 106-120, 1997.

[4] Wang, G., and Dunbrack, R.L., PISCES: Recent Improvements to a PDB Sequence Culling Server, *Nucleic Acids Research*, 33(Web Server issue): W94-8, 2005.

Copyright ©2011 Information Processing Society of Japan.
All Rights Reserved.