

ドッキング計算に基づく 網羅的タンパク質-RNA 間相互作用予測

大上 雅史^{†1,†2} 松崎 由理^{†1} 内古閑 伸之^{†1}
石田 貴士^{†1} 秋山 泰^{†1}

タンパク質-RNA 間相互作用 (protein-RNA interaction, PRI) は、細胞システムの理解に重要であり、複数のタンパク質と RNA の相互作用関係を網羅的に予測する手法の確立が求められている。我々は立体構造情報を用いたタンパク質間相互作用予測システム MEGADOCK を開発してきたが、本研究では MEGADOCK に対して CHARMM27 核酸原子パラメータを導入することで、RNA 立体構造を扱えるように拡張した。Protein Data Bank より選出した 78 の RNA 結合タンパク質による 78 × 78 件の網羅的 PRI 予測を行い、元の構造に相当する 78 ペアを正例、その他の 6,006 ペアを負例として評価を行った結果、 F 値 0.465 での予測に成功した。

Docking-calculation-based All-to-all Protein-RNA Interaction Prediction

MASAHITO OHUE,^{†1,†2} YURI MATSUZAKI,^{†1}
NOBUYUKI UCHIKOGA,^{†1} TAKASHI ISHIDA^{†1}
and YUTAKA AKIYAMA^{†1}

Elucidating protein-RNA interactions (PRIs) is important for understanding cellular systems. We developed a PRI prediction method by using a rigid-body protein-RNA docking calculation with tertiary structure data. We evaluated this method by using 78 protein-RNA complex structures from the Protein Data Bank. We predicted the interactions for pairs in 78 × 78 combinations. Of these, 78 original complexes were defined as positive pairs, and the other 6,006 complexes were defined as negative pairs; then an F -measure value of 0.465 was obtained with our prediction system.

1. はじめに

タンパク質-RNA 間相互作用 (protein-RNA interaction, PRI) は、タンパク質合成や遺伝子発現制御などの細胞システムの理解に重要であり、細胞内のタンパク質と RNA の相互作用関係を網羅的に予測する手法の確立が求められている。

これまでに、PRI について既知の結晶構造を用いた統計的な解析を行った研究はいくつかあり、相互作用が起きるリボ核酸塩基とアミノ酸残基の組み合わせについての検証が行われている¹⁾⁻⁴⁾。また、Molfit⁵⁾ や FTDock⁶⁾、ZDOCK^{7),8)} などのタンパク質ドッキング手法を利用した、タンパク質-RNA 間剛体ドッキングに関する研究も少しずつ増えてきており⁹⁾、最近では、タンパク質複合体構造予測に関する国際コンテストである CAPRI¹⁰⁾ にもタンパク質-RNA 複合体が登場している^{11),12)}。このように、PRI 予測は様々な方面からのアプローチによって研究がされてきている。

一方で、たとえば Pérez-Cano らの研究⁹⁾ は、高々数個のタンパク質-RNA 複合体にしか手法を適用しておらず、複数のタンパク質-RNA 群に対しての網羅的な相互作用予測研究は行われていないのが現状である。Pérez-Cano らは、複合体状態の *bound* 構造と、*unbound* 構造の両方が得られている複合体構造を対象としたドッキング予測を行ったものであり、*bound*、*unbound* 共に構造が得られている例が少なく、多くの例での検証が困難であったと想定される。しかし、*bound* 構造の複合体群から得られたタンパク質群と RNA 群の総当たり解析であれば、現在得られている構造からでも多くの予測を行うことが可能となる。細胞内では多くのタンパク質と RNA が相互作用しており、ヒトゲノム中には RNA と相互作用するタンパク質が少なくとも 1,500 個はコードされているという推定もある¹³⁾ ため、多数のタンパク質-RNA 群を想定した網羅的な PRI 予測は、今後ますます重要になると考える。

本研究ではタンパク質ドッキング計算を応用した PRI 予測手法を提案する。我々は ZDOCK を用いた網羅的なタンパク質間相互作用 (protein-protein interaction, PPI) 予測の研究¹⁴⁾ や、計算の高速性を重視した独自の網羅的 PPI 予測システム MEGADOCK¹⁵⁾ の開発を進めており、本研究はこの MEGADOCK のタンパク質-RNA 間相互作用予測への拡張を行うものである¹⁶⁾。

†1 東京工業大学 大学院情報理工学専攻

Graduate School of Information Science and Engineering, Tokyo Institute of Technology

†2 日本学術振興会特別研究員

Research Fellow of the Japan Society for the Promotion of Science

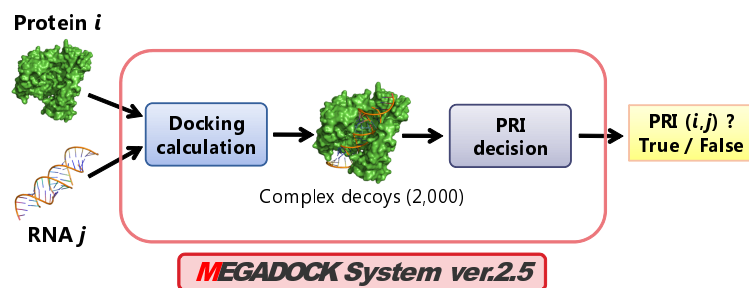


図 1 提案する MEGADOCK ver.2.5 の概略図．タンパク質 i と RNA j のペアに対して，FFT Grid ベース⁵⁾ のドッキング計算を行い，評価値の高い候補構造を上位 2,000 個抽出する (Docking calculation)．それらの候補構造の評価値の分布を用いてペア (i, j) の相互作用の有無を予測する (PRI decision)．

Fig. 1 Process flow for MEGADOCK ver.2.5, the PRI prediction system proposed in this paper. This system calculates FFT-based⁵⁾ rigid-body docking by using the given protein i and RNA j pair, generates 2,000 high-ranked decoys (Docking calculation), and detects the interacting (i, j) pair from docking score distributions (PRI decision).

2. 提案手法

本研究では，従来 PPI 予測を行っていたシステムを RNA 分子へ拡張した新たな PRI 予測システム (MEGADOCK ver.2.5) を提案する．MEGADOCK システムの構成・概略を図 1 に示す．また，以下では使用したデータセットと MEGADOCK システムについて説明する．

2.1 データセット

本研究で用いたデータセットは，Protein Data Bank (PDB) に含まれるタンパク質-RNA 複合体のうち，1) 解像度が 3.0Å 以下，2) タンパク質配列相同性が 30%以下，3) X 線結晶構造である，という条件に含まれるものを使用した．選出の際には PISCES サーバ¹⁷⁾ を利用した．選出された構造数は 78 であった．表 1 に RNA の種類ごとの PDB ID リストを示す．

2.2 MEGADOCK：ドッキング構造生成

ここでは，図 1 に示した Docking calculation における処理について説明する．

構 成

MEGADOCK 2.5 のドッキング計算は，MEGADOCK 2.2 の FFT Grid ベースの剛体

表 1 使用した 78 個のタンパク質-RNA 複合体の PDB ID のリスト．
Table 1 List of the PDB IDs of the 78 protein-RNA complexes used.

RNA type	PDB ID of the protein-RNA complex
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, 2ZIO
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)

ドッキング計算¹⁵⁾ に RNA 分子のパラメータを導入することで実現している．MEGADOCK 2.2 のドッキング計算の評価値は rPSC スコアと静電的相互作用スコアの重みつき和であるが，静電的相互作用を計算する際の原子電荷パラメータに CHARMM19^{18),19)} によるアミノ酸原子パラメータを用いていた．MEGADOCK 2.5 では，核酸や脂質の力場が定義された CHARMM27²⁰⁾ によって，DNA・RNA 原子電荷パラメータを新たに導入した．導入にあたり，ZDOCK 2.3⁸⁾ に DNA パラメータを導入してタンパク質-DNA ドッキングを行った研究²¹⁾ を参考にした．

性能評価 (複合体構造予測)

MEGADOCK 2.5 のドッキング予測性能を確認するため，表 1 の複合体に対して再ドッキングを行い，タンパク質側を固定したときの RNA の全原子を用いた RMSD (RNA-RMSD) によって複合体構造予測精度を確認した．ZDOCK のデフォルト設定⁸⁾ と同様に，RNA は 15 度刻みで 3,600 通りの回転パターンをとりながらタンパク質の周りを並進探索し，各回転パターンごとに最も良い評価値を得た構造を選択する．それら 3,600 個の構造でさらに評価値の高い上位構造の中に RNA-RMSD が 5Å 以下の構造 (near-native decoy) が含まれているかどうかで評価を行った．比較対象として，形状相補性のみでのドッキング

表 2 MEGADOCK と ZDOCK の再ドッキング結果．表 1 の 78 複合体のうち，near-native decoy が各順位範囲に含まれていた複合体数をカウントしている．

Table 2 Results for protein-RNA re-docking test of MEGADOCK and ZDOCK.

	MEGADOCK 2.2	MEGADOCK 2.5	ZDOCK 2.3
1 位	47	51	48
2 位 ~ 10 位	10	13	15
11 位 ~ 100 位	8	4	3

計算を行う MEGADOCK 2.2^{*1} と，形状相補性と疎水効果によってドッキング計算を行う ZDOCK 2.3^{*2} によるドッキング予測も行った．表 2 に結果を示す．

表 2 は，各複合体ごとに，

- 1 位に near-native decoy が存在したもの
- (1 位には存在しなかったが) 2 位 ~ 10 位の中に near-native decoy が存在したもの
- (10 位までには存在しなかったが) 11 位 ~ 100 位の中に near-native decoy が存在したもの

をそれぞれカウントしたものである．

表 2 より，MEGADOCK の静電的相互作用を導入した効果 (MEGADOCK 2.2 から MEGADOCK 2.5 への性能向上) が確認された．また，ZDOCK 2.3 と MEGADOCK 2.5 を比べると，1 位が near-native decoy であった複合体数や，1 位 ~ 100 位までに near-native decoy が存在した複合体数が MEGADOCK 2.5 の方がわずかではあるが多く，タンパク質-RNA 間相互作用では疎水効果より静電的相互作用を考慮する重要性が示唆されていると考えられる．

図 2 に再ドッキングした構造をいくつか示した．図 2(a) は 2NUG，図 2(b) は 3EPH，図 2(c) は 3FOZ である．RNA は 2 つ表示しており，赤色の構造は実際の正解構造となる X 線結晶構造，緑色の構造は MEGADOCK 2.5 によって生成された 1 位の構造である．それぞれの RNA-RMSD は，(a)2NUG が 1.45Å，(b)3EPH が 2.25Å，(c)3FOZ が 3.03Å であり，再ドッキングがうまくいっていることがわかる．

2.3 MEGADOCK : PRI 判定

ここでは図 1 における PRI decision の部分である，ドッキング計算結果から対象ペアが

*1 RNA 電荷パラメータが含まれていないので，静電的相互作用の効果が 0 となる．

*2 本来は静電的相互作用も計算しているが，RNA 電荷を扱っていないため，MEGADOCK 2.2 と同様に静電的相互作用の効果が 0 となる．

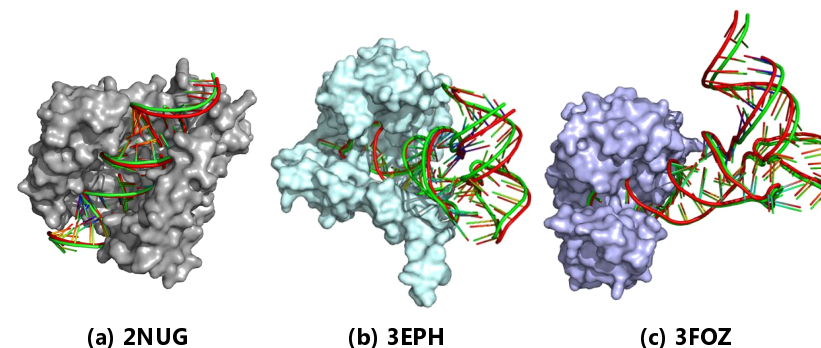


図 2 再ドッキングの結果の構造を示す．(a) は 2NUG，(b) は 3EPH，(c) は 3FOZ である．赤色の RNA はオリジナルの X 線結晶構造，緑色の RNA は MEGADOCK 2.5 による予測構造である．

Fig. 2 The structures after re-docking are shown for (a) 2NUG, (b) 3EPH, and (c) 3FOZ. In each figure, two RNA structures are shown: the green structure is the first ranked decoy generated by MEGADOCK, and the red structure is the original x-ray crystal structure.

相互作用するかないかを判定する手法について説明する．本研究で提案するドッキング計算結果から PRI 判定を行う手法の概要を図 3 に示す．

タンパク質 i と RNA j が相互作用する (True) かない (False) かを表す関数 $PRI(i, j)$ は以下の PRI スコア E を用いて決定される．

$$E = \frac{S_1 - \mu}{\sigma} \quad (1)$$

$$\mu = \frac{1}{D} \sum_{k=1}^D S_k, \quad \sigma^2 = \frac{1}{D} \sum_{k=1}^D (S_k - \mu)^2 \quad (2)$$

ここで S_1 は対象のタンパク質-RNA ペアの最も高いドッキング評価値， S_k は k 番目に高いドッキング評価値を表す． D はドッキング計算によって生成する候補構造数であり，本研究においては $D = 2,000$ とした．この PRI スコア E を用いて，閾値 E^* より大きいか小さいかによって，以下のように $PRI(i, j)$ を決定する．

$$PRI(i, j) = \begin{cases} \text{True} & \text{if } E > E^* \\ \text{False} & \text{otherwise} \end{cases} \quad (3)$$

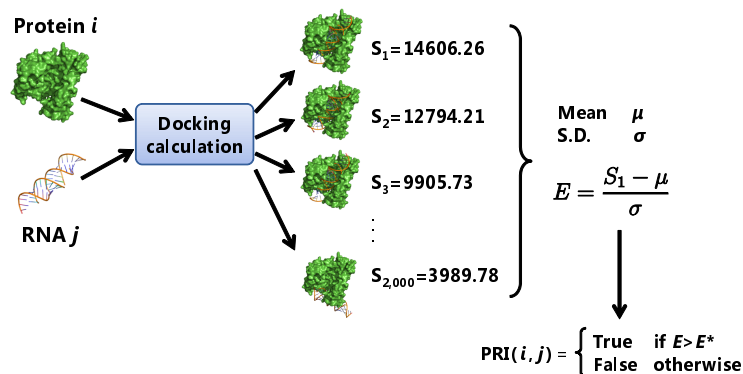


図3 ドッキング計算結果からの PRI 判定の概要。
Fig. 3 The workflow of PRI decision from docking calculation results.

3. 結果と考察

3.1 78 × 78 網羅的 PRI 予測

表1の78のタンパク質-RNAペアに対し、78 × 78の総当りドッキング計算を行い、網羅的PRI予測を行った。予測精度の評価として、元々のペアに相当する78のペアを正例、それ以外の6,006ペアを負例とし、true-positive(TP), true-negative(TN), false-positive(FP), false-negative(FN)の数を計上した。これらの値から、以下のsensitivityとspecificityを計算した。

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad \text{specificity} = \frac{TN}{TN + FP} \quad (4)$$

また、総合的な評価値としてF値を用いた。F値は以下に示す式で計算される。

$$F \text{ 値} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (5)$$

PRI予測のパラメータである閾値E*を3.0から13.0まで0.1刻みで変化させたときの予測結果におけるF値の推移を図4に示す。E* = 9.6のときにF値が最大0.465となる予測を得た。また、このときのsensitivityは0.385, specificityは0.997であった。対象の大きさや問題が異なるために直接比較することはできないが、44 × 44の網羅的PPI予測問題に対して、ZDOCKを用いて行った例¹⁴⁾ではF値0.43, MEGADOCKを用いて行っ

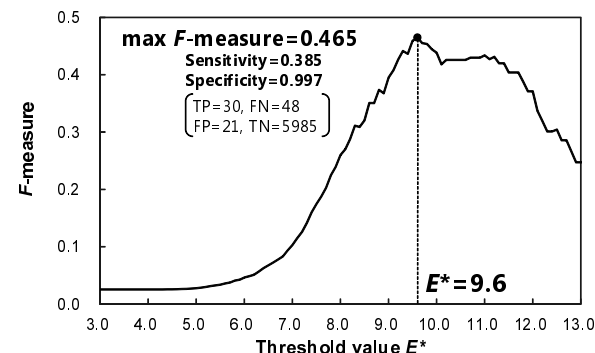


図4 78 × 78 データセットの予測結果。閾値E*に対するF値の変化を表している。最大はE* = 9.6のときで、F値0.465 (sensitivity = 0.385, specificity = 0.997)である。

Fig. 4 Results of the 78 × 78 prediction. This graph shows the change in the F-measure with respect to the threshold E*. The maximum F-measure is 0.465 when E* is 9.6, with a sensitivity of 0.385 and a specificity of 0.997.

た例¹⁵⁾ではF値0.415と報告されており、本研究の結果と近い値となっている。

パラメータE*による過学習の可能性を確認するために、表1の78のペアを分割した2つの39 × 39サブセットを構築し、2-fold cross validationによる予測を行った。結果を図5に示す。最大のF値をとるパラメータE*の値が両サブセットでほぼ共通していることから、過学習が起こっていないことが示唆される。

78 × 78予測におけるreceiver-operator characteristics (ROC) 曲線²³⁾を図6に示す。ROC曲線は、true-positive fraction (=sensitivity)とfalse-positive fraction (=1 - specificity)をプロットしたものであり、sensitivityとspecificityのトレードオフを表すものである。ランダムな予測では(0, 0)から(1, 1)の対角線になり、この対角線よりも左上の領域に曲線が位置すると、ランダムな予測よりも良いということが示される。本研究の結果は、ROC曲線が対角線よりも左上の領域に位置しているため、ランダムよりも良い予測であることが図6から見て取れる。さらに、定量的に評価する指標としてROC曲線の曲線下面積 (area under the curve, AUC) を計算したところ、0.821という値を得た。

全予測結果をヒートマップ形式で可視化したPRIマップを図7に示す。赤色のセルはEがE*(= 9.6)よりも高かったペアを示しており、正例ペアを示す左上から右下への対角線上のセルが、相互作用するものとして予測されていることが見て取れる。

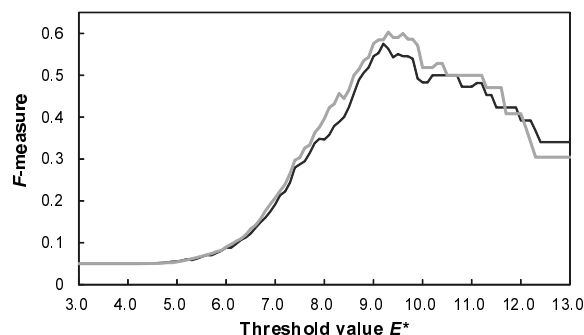


図 5 2-fold cross validation に見立てた 2 つの 39×39 サブセットにおける予測結果 . 閾値 E^* に対する F 値の変化を表している . 2 つのサブセット間で高い F 値を取る E^* の値がほぼ一致していることから , 過学習が起きていないことが示唆される .

Fig. 5 Results of 2-fold cross validation prediction performed using the divided 39×39 subset. This graph shows the change of F -measure with respect to the threshold E^* . Because the value of E^* that yielded the maximum F -measure value was almost equal, it can be said that overfitting did not occur.

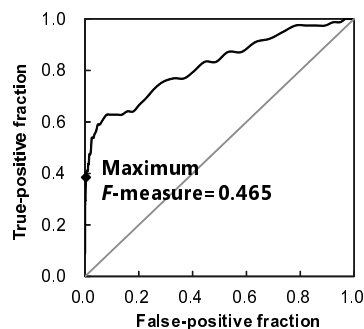


図 6 78×78 データセットの予測結果における ROC 曲線 . 曲線下面積 (AUC) の値は 0.821 であった .
Fig. 6 ROC curve of 78×78 dataset prediction results. The area under the curve (AUC) is 0.821.

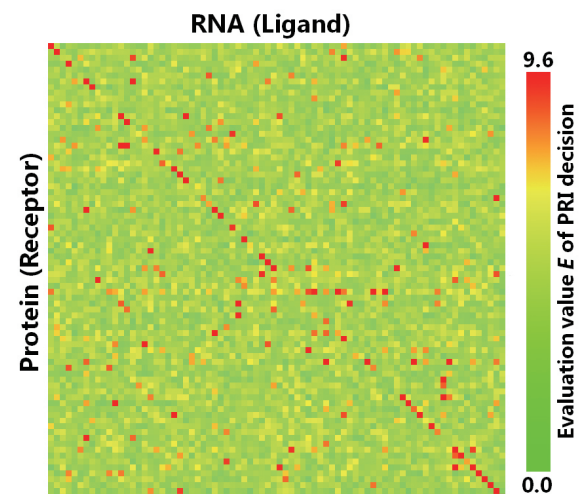


図 7 78×78 PRI 予測結果マップ . PRI スコア E が 9.6 を超えるものは全て赤色のセルで表されている . セルの並びは , 縦横それぞれ PDB ID の辞書的昇順となっている .

Fig. 7 78×78 map of PRI prediction results. The red cells are the cells for which the E is more than E^* ($= 9.6$). The cells have been arranged according to the PDB IDs, which have been arranged in alphabetical order for all axes.

3.2 False-positive の検討

21 個存在した false-positive ペアのうち , RNaseIII と dsRNA の複合体である 2NUG , および tRNA transferase と tRNA の複合体である 3EPH と 3FOZ が関係するペアについて構造を確認した . 対象とした結晶構造の PDB ID と説明を表 3 に示す . また , 表 4 にペアごとのスコア E の値と予測結果 , 図 8 に 2NUG のタンパク質と 2GJW , 2ZKO , 3EGZ のそれぞれの RNA をドッキングさせた構造を , 図 9 に 3EPH-3FOZ ペアと 3FOZ-3EPH ペアのドッキング結果をそれぞれ示す .

図 2(a) と図 8 を比較すると , 構造としては良く似かよっているように見える . タンパク質側が RNase ということもあり , dsRNA であれば結合すると予測されること自体は自然であり , この 3 例が高いスコア E を得ていることは , 構造からも納得できる結果であると考えられる . また , 図 2(b)(c) と図 9 においては , 同じ種類のタンパク質と RNA の組み合わせであるため構造は良く似ているが , 表 4 に示すように 3FOZ-3EPH は高いスコア E を得ているのに対し , 3EPH-3FOZ は $E = 9.2$ (相互作用しないと予測される) となってい

表 3 図 2, 図 8, 図 9 に示した PDB ID 名とその内容.

Table 3 PDB ID and the description of protein-RNA structures in Fig. 2, Fig. 8 and Fig. 9.

PDB ID	description
2NUG	RNase III from <i>Aquifex aeolicus</i> and dsRNA
2GJW	<i>Archaeoglobus fulgidus</i> tRNA-splicing endonuclease
2ZKO	NS1 protein of human influenza virus A and dsRNA
3EGZ	<i>Homo sapiens</i> u1 small nuclear ribonucleo-protein tetracycline aptamer and artificial riboswitch
3EPH	<i>Saccharomyces cerevisiae</i> dimethylallyl tRNA transferase and tRNA
3FOZ	<i>Escherichia coli</i> isopentenyl tRNA transferase and tRNA

表 4 図 2, 図 8, 図 9 に示したタンパク質-RNA ペアにおける相互作用予測結果.

Table 4 Interaction prediction results of protein-RNA pairs in Fig. 2, Fig. 8 and Fig. 9.

Pair (Protein-RNA)	PRI score E	Expected	Prediction	Prediction result
2NUG-2NUG	12.5		True	TP
2NUG-2GJW	9.8	×	True	FP
2NUG-2ZKO	9.9	×	True	FP
2NUG-3EGZ	11.1	×	True	FP
3EPH-3EPH	15.2		True	TP
3EPH-3FOZ	9.2	×	False	TN
3FOZ-3FOZ	17.9		True	TP
3FOZ-3EPH	12.5	×	True	FP

る. これらの例のように, 構造に大差がない場合で結果が異なることは, 生物種による微妙な違いの表れであると考えられ, 興味深い結果に思われるとともにより詳細な解析が求められる.

3.3 課題

RNA は柔軟な分子であり, PRI の予測を剛体ドッキングによるアプローチで行うことは適切でないように思われる. しかしながら, 剛体ドッキング計算を用いた網羅的予測は, 細胞内の PRI 現象における構造上の類似性などを理解するのに有用であると我々は考える. これに関連した今後の課題として, あらかじめ分子動力学法などによって構造のアンサンブルをとるなどして, RNA の構造の柔軟性を取り入れていくことが挙げられる. タンパク質間のドッキングにおけるアンサンブルドッキングの研究事例として, 複数の NMR コンフォメーションを利用して ZDOCK によってアンサンブルドッキングを実施した研究²²⁾がある. RNA のフレキシブル性を考慮すれば, RNA 側の構造アンサンブルを取ることは相互作用予測精度の向上に寄与すると考えられる.

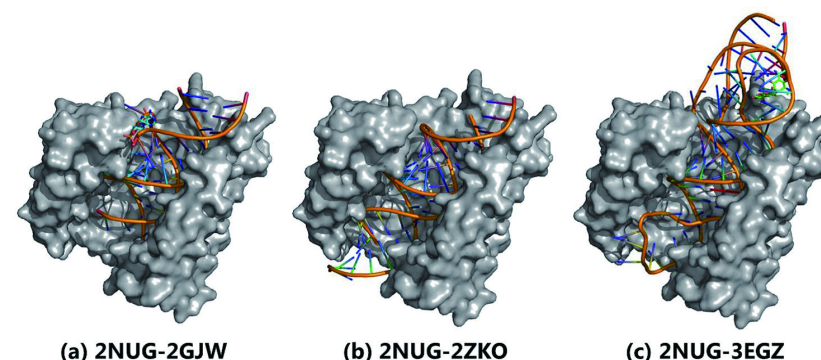


図 8 2NUG のタンパク質 (RNaseIII) と (a) 2GJW, (b) 2ZKO, (c) 3EGZ のそれぞれの RNA をドッキングさせた構造. 全て MEGADOCK 2.5 による 1 位のドッキング予測構造である.

Fig. 8 Protein (RNaseIII) of 2NUG and the RNA of the (a) 2GJW, (b) 2ZKO, and (c) 3EGZ docked structures. The structures are first ranked decoys generated by MEGADOCK 2.5.

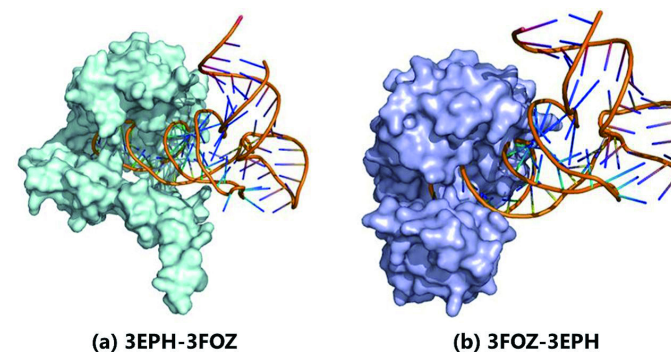


図 9 (a) 3EPH のタンパク質と 3FOZ の RNA のドッキング構造, (b) 3FOZ のタンパク質と 3EPH の RNA のドッキング構造. 全て MEGADOCK 2.5 による 1 位のドッキング予測構造である.

Fig. 9 (a) Protein of 3EPH and RNA for the 3FOZ docking structure and (b) protein of 3FOZ and RNA for the 3EPH docking structure. The structures are first ranked decoys generated by MEGADOCK 2.5.

4. EF-Tu タンパク質と tRNA の相互作用予測

本研究で提案した PRI 予測手法を実際の生物学的テーマに応用できるかを検討するための萌芽的な実験として、翻訳伸長因子 (elongation factor, EF) Tu タンパク質と tRNA の相互作用予測を試みた。EF タンパク質²⁴⁾ は、タンパク質翻訳においてポリペプチド鎖の伸長を促進させる機能を担っている。EF-Tu, EF-Ts, EF-G の 3 種類が存在するが、このうちの EF-Tu は、アミノアシル tRNA に結合してリボソームの活性部位に tRNA を正しく供給するという働きを持つ。生物に広く存在し保存されている EF-Tu であるが、生物種による相互作用現象の違いや、異種間の EF-Tu と tRNA による相互作用が実現するかどうかといったことはあまり研究されていない。異種生物種間において、MEGADOCK による相互作用予測結果によって差異が見い出せないかと考え、本研究では EF-Tu タンパク質構造と tRNA 構造を用いた PRI 予測を行った。

4.1 データセット

EF-Tu タンパク質の結晶構造として、1XB2 (*Bos taurus*), 1EFU (*Escherichia coli*), 1AIP (*Thermus thermophilus*) を PDB より選択した。これらは全て EF-Tu と EF-Ts の複合体となっているが、EF-Tu のみ切り離したものを利用した。また、EF-Tu は tRNA との結合状態と、EF-Ts との結合状態とで構造が異なるため、bound 構造に相当する既に tRNA と結合している結晶構造として PDB に登録されていた、2WRN (*T. thermophilus*), 1TTT (*Thermus aquaticus*) についても予測を行った。tRNA 構造については、2WRN と 1TTT の 2 種を用いた。

4.2 結果と考察

予測結果を表 5 に示す。2WRN の bound ペアによる予測が最もスコアが高く $E = 8.11$ であったが、これは 3.1 節で得られたパラメータ $E^* = 9.6$ を超えていないため、網羅的な相互作用予測の基準を用いれば相互作用しないものとして予測されるものである。したがって表 5 に示す全ての結果において相互作用しないと予測されることになるが、例えば bound ペア 1TTT-1TTT の PRI スコア $E (= 6.77)$ を基準とし、それ以上のスコアを得たペアについて相互作用する可能性があるかと判断するならば、2WRN-2WRN, 1AIP-2WRN, 2WRN-1TTT, 1XB2-2WRN, 1TTT-1TTT の 5 ペアが該当する。

また、全体的に *T. thermophilus* に関連する組が上位に位置していることから、*T. thermophilus* の EF-Tu と tRNA の親和性は他の種に比べて高いと予想される一方で、2.1 節に示したような冗長性を排除したデータセットではなく、配列や構造が類似しているもの同

表 5 EF-Tu と tRNA の相互作用予測結果。並びはスコア E の降順である。PDB ID の色は生物種との対応を示す。

Table 5 Results of EF-Tu-tRNA interaction prediction. The columns have been arranged according to the PRI score E in descending order. The color of PDB ID shows correspondence with the organisms (right table).

protein	RNA	score E	color table
2WRN	2WRN	8.11	1AIP <i>Thermus thermophilus</i>
1AIP	2WRN	8.04	1EFU <i>Escherichia coli</i>
2WRN	1TTT	7.54	1XB2 <i>Bos taurus</i>
1XB2	2WRN	7.15	2WRN <i>Thermus thermophilus</i>
1TTT	1TTT	6.77	1TTT <i>Thermus aquaticus</i>
1EFU	2WRN	6.25	
1TTT	2WRN	6.11	
1XB2	1TTT	5.89	
1EFU	1TTT	5.68	
1AIP	1TTT	5.63	

士の複合体によって構成されるデータセットを用意して、より多くのケースで PRI 予測を実施・比較する必要があると考えられる。

5. おわりに

本研究では、タンパク質間相互作用予測システム MEGADOCK を RNA 用に拡張したタンパク質-RNA 間相互作用予測システム MEGADOCK 2.5 を開発した。MEGADOCK 2.5 は剛体ドッキング手法を利用した高速な相互作用予測システムであり、その剛体ドッキングは ZDOCK 2.3 とほぼ同等の精度で行えることを確認し、静電的相互作用の寄与が大きいことが示唆された。また、本システムを PDB から抽出した 78 個のタンパク質-RNA 複合体による 78×78 通りの組み合わせに適用して相互作用予測精度を検証したところ、 F 値で 0.465 という値を得た。

今後の課題として、構造の柔軟性に対応する手法の開発や、EF-Tu と tRNA をはじめとする実際の生物学的なテーマに応用していくことが挙げられる。

謝 辞

本研究の計算機実験は東工大 学術国際情報センター Tsubame2 スーパーコンピュータによって行われた。また、インド工科大学 Gromiha 博士から EF-Tu に関する助言を頂いた。感謝の意を表す。本研究は、文部科学省 最先端・高性能汎用スーパーコンピュータ

の開発利用「次世代生命体統合シミュレーションソフトウェアの研究開発」, 科研費(基盤研究(B)19300102, 特別研究員奨励費 23-8750)の支援を受けて行われたものである。

参 考 文 献

- 1) Ellis, J.J., Broom, M. and Jones, S., Protein-RNA Interactions : Structural Analysis and Functional Classes, *Proteins*, 66(4), 903–911, 2007.
- 2) Jeong, E., Kim, H., Lee, S. and Han, K., Discovering the interaction propensities of amino acids and nucleotides from protein-RNA complexes. *Molecules and cells*, 16(2), 161–167, 2003.
- 3) Lejeune, D., Delsaux, N., Charlotiaux, B., Thomas, A. and Brasseur, R., Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure, *Proteins*, 61(2), 258–271, 2005.
- 4) Pérez-Cano, L. and Fernández-Recio, J., Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins, *Proteins*, 78(1), 25–35, 2010.
- 5) Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. and Vakser, I.A., Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques, *Proc. Natl. Acad. Sci. USA*, 89(6), 2195–2199, 1992.
- 6) Gabb, H.A., Jackson, R.M. and Sternberg, M.J., Modelling protein docking using shape complementarity, electrostatics and biochemical information, *Journal of Molecular Biology*, 272, 106–120, 1997.
- 7) Chen, R. and Weng, Z., A novel shape complementarity scoring function for protein-protein docking, *Proteins*, 51(3), 397–408, 2003.
- 8) Chen, R., Li, L. and Weng, Z., ZDOCK: an initial-stage protein-docking algorithm, *Proteins*, 52(1), 80–87, 2003.
- 9) Pérez-Cano, L., Solernou, A., Pons, C. and Fernández-Recio, J., Structural prediction of protein-RNA interaction by computational docking with propensity-based statistical potentials, *Pacific Symposium on Biocomputing 2010*, 293–301, 2010.
- 10) Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J.E., Vajda, S., Vakser, I. and Wodak, S.J., CAPRI: a Critical Assessment of PRedicted Interactions, *Proteins*, 52(1), 2–9, 2003.
- 11) Lensink, M.F. and Wodak, S.J., Docking and scoring protein interactions: CAPRI 2009, *Proteins*, 78(15), 3073–3084, 2010.
- 12) Pons, C., Solernou, A., Pérez-Cano, L., Grosdidier, S., and Fernández-Recio, J., Optimization of pyDock for the new CAPRI challenges: Docking of homology-based models, domain-domain assembly and protein-RNA binding, *Proteins*, 78(15), 3182–3138, 2010.
- 13) Keene, J.D., Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome, *Proc. Natl. Acad. Sci. USA*, 98(13), 7018–7024, 2001.
- 14) Matsuzaki, Y., Matsuzaki, Y., Sato, T. and Akiyama, Y., *In silico* screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis, *J Bioinform and Comput Biol*, 7(6), 991–1012, 2009.
- 15) 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, 情報処理学会論文誌 数理モデル化と応用, 3(3), 91–106, 2010.
- 16) Ohue, M., Matsuzaki, Yuri and Akiyama, Y., Protein-RNA Interaction Prediction Method based on a Docking Calculation, *Genome Informatics*, 25, 2011 (in press).
- 17) 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.
- 18) Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *Journal of Computational Chemistry*, 4, 187–217, 1983.
- 19) Reiher, III., W.H., Theoretical studies of hydrogen bonding, *Ph.D. Thesis at Harvard University*, 1985.
- 20) MacKerell, Jr. A.D, Banavali, N. and Foloppe, N., Development and current status of the CHARMM force field for nucleic acids, *Biopolymers*, 56(4), 257–265, 2001.
- 21) Fanelli, F. and Ferrari, S., Prediction of MEF2A-DNA interface by rigid body docking: a tool for fast estimation of protein mutational effects on DNA binding, *Journal of Structural Biology*, 153(3), 278–283, 2006.
- 22) Yi, H., Qiu, S., Cao, Z., Wu, Y. and Li, W., Molecular basis of inhibitory peptide maurotoxin recognizing Kv1.2 channel explored by ZDOCK and molecular dynamic simulations, *Proteins*, 70(3), 844–854, 2010.
- 23) Zweig, M.H. and Campbell, G., Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, 39, 561–577, 1993.
- 24) Andersen, G.R., Nissen, P. and Nyborg, J., Elongation factors in protein biosynthesis, *Trends in Biochemical Sciences*, 28, 434–441, 2003.