

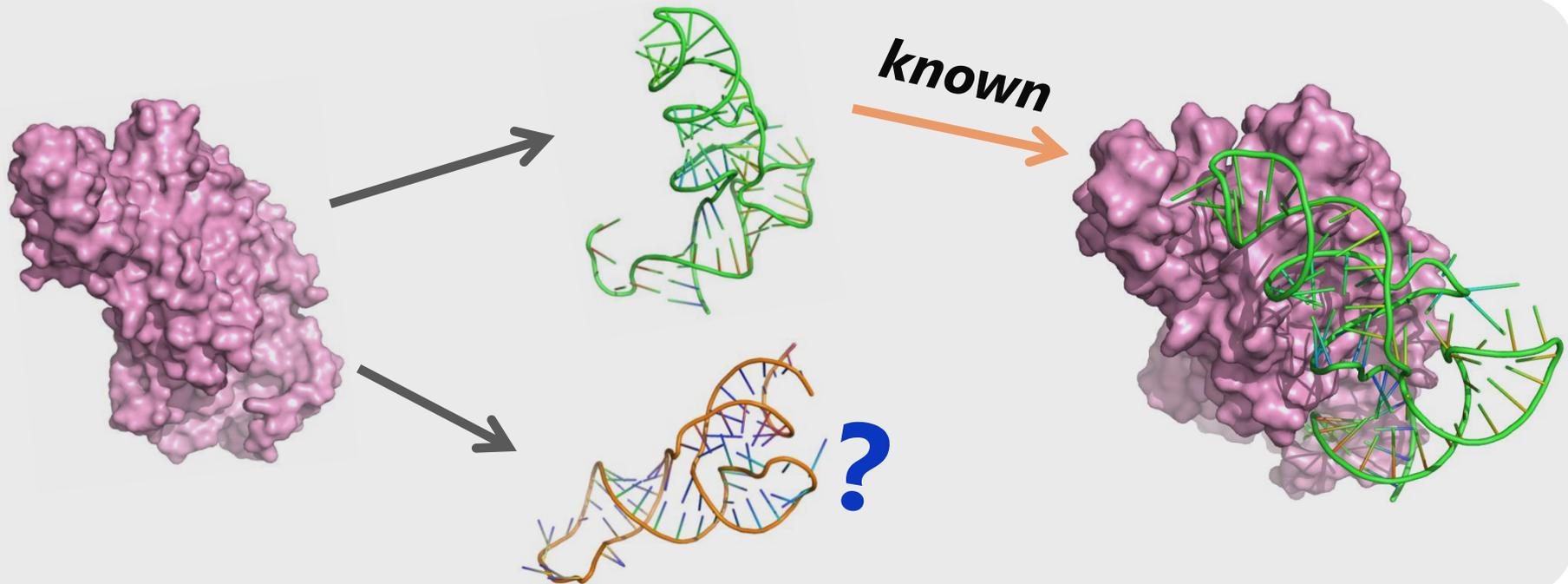
ドッキング計算に基づく網羅的 タンパク質-RNA間相互作用予測

大上 雅史 松崎 由理 内古閑 伸之
石田 貴士 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

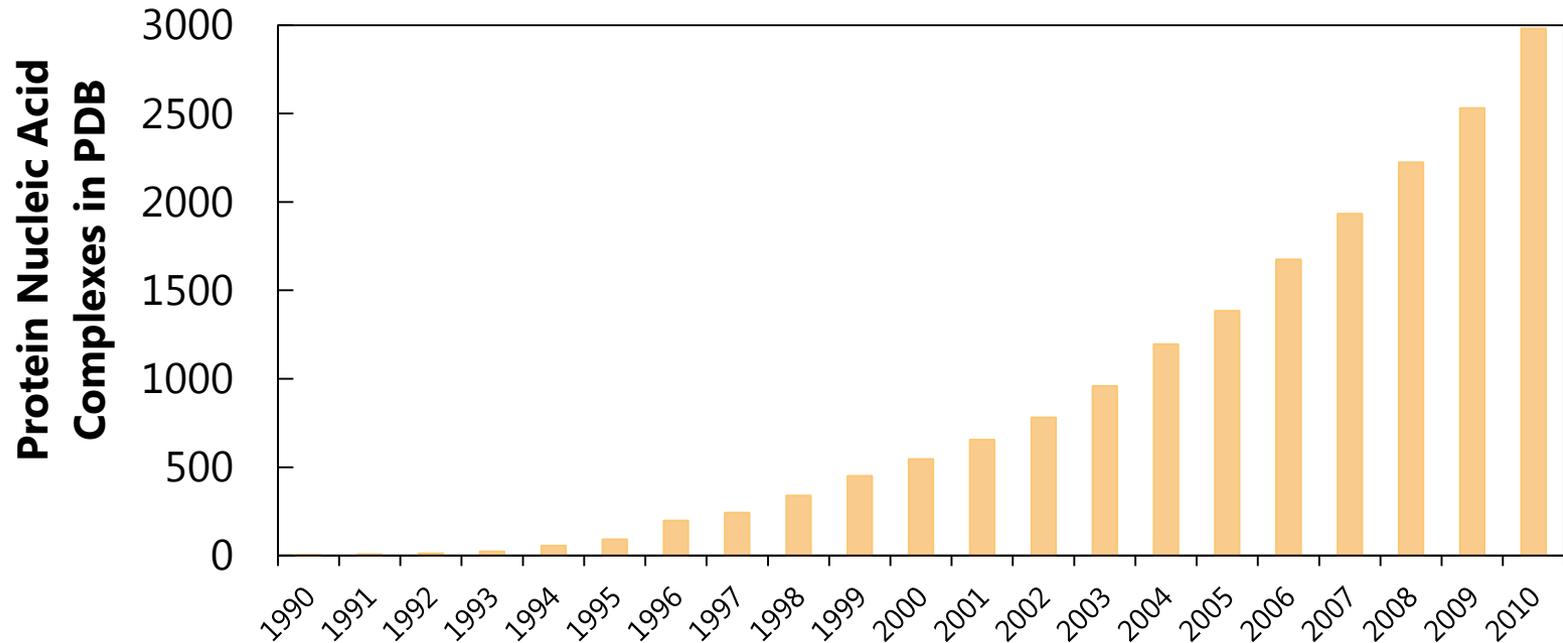
はじめに

- **タンパク質-RNA間相互作用** (Protein-RNA Interaction, PRI)
 - 細胞システム等の理解に重要
- **タンパク質-RNA間相互作用予測**
 - タンパク質がどのRNAと相互作用するかを予測



立体構造の現状

- **タンパク質-RNA複合体の結晶構造が増加**
 - 立体構造を用いた研究が増えつつある

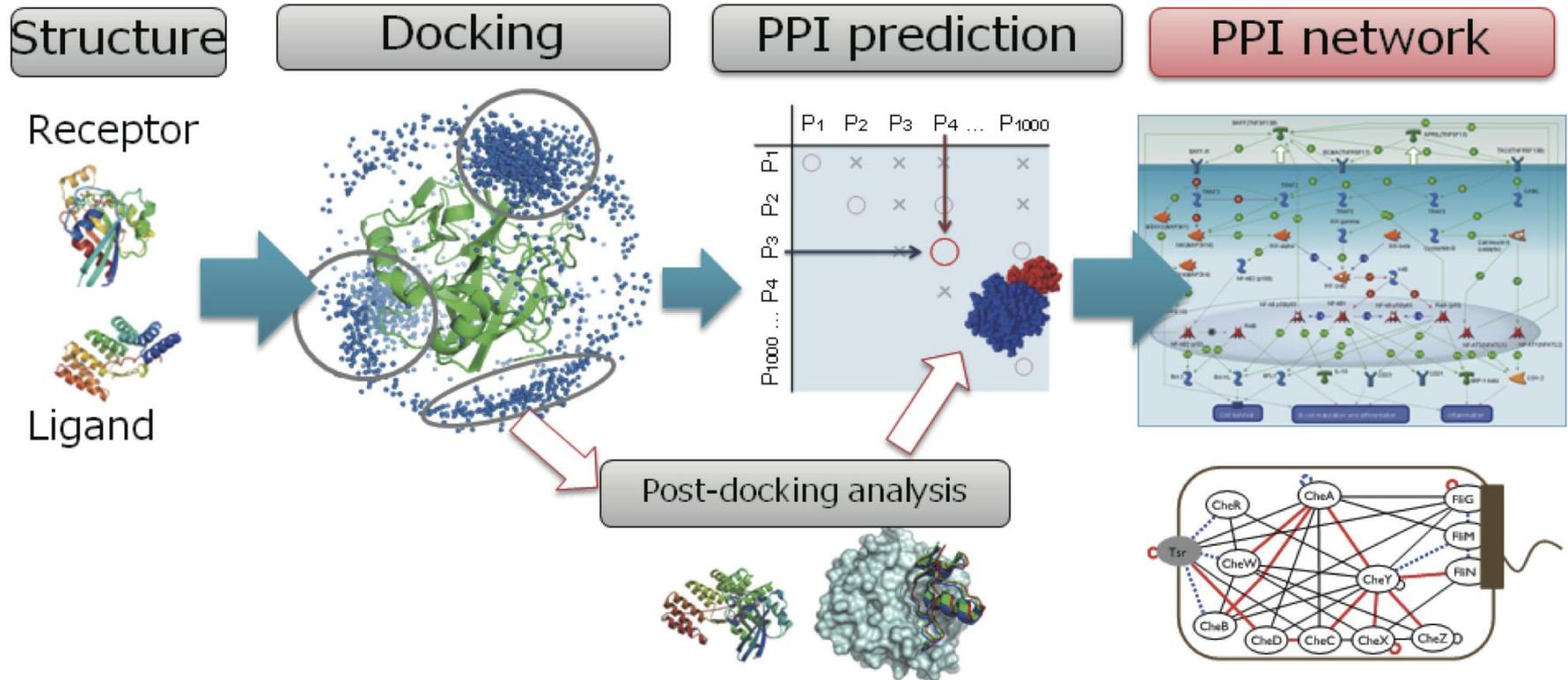


- **本研究の目的**
 - 立体構造を用いて、複数のタンパク質・RNAとの間の相互作用関係を網羅的に予測する

既存研究

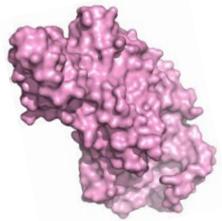
MEGADOCK

タンパク質間相互作用(PPI)の網羅的予測のためのシステム

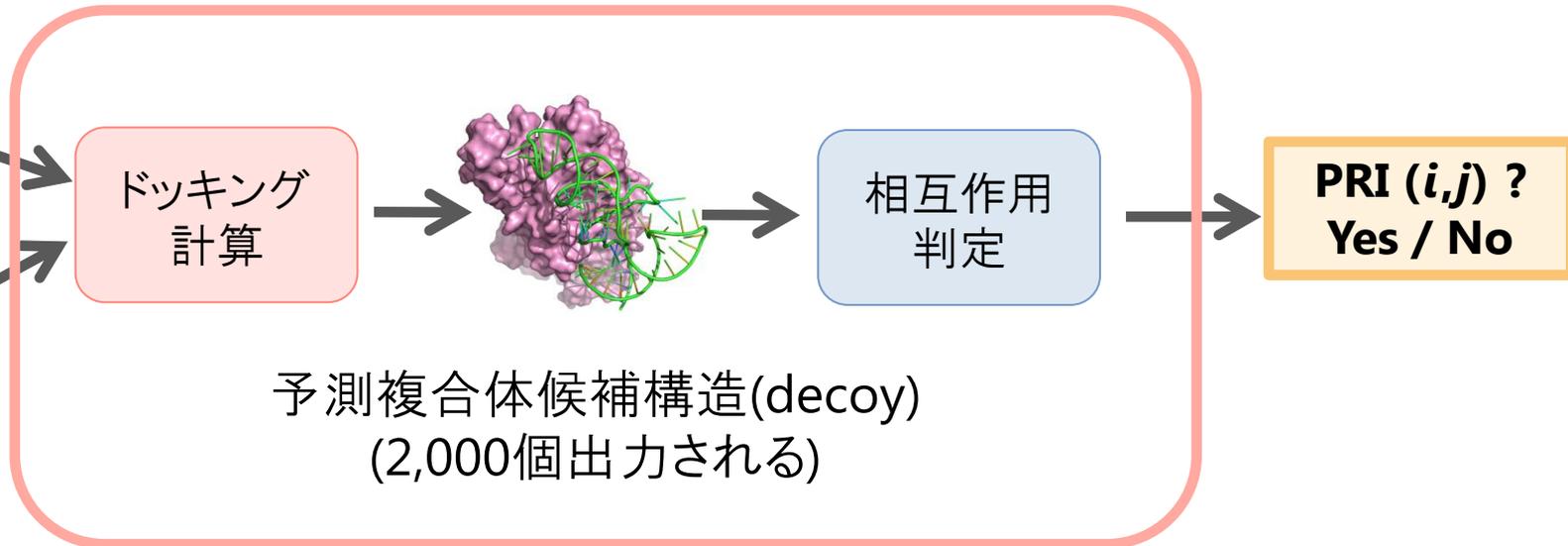


予測方法

タンパク質 i



RNA j



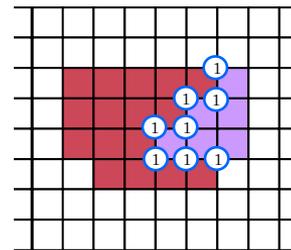
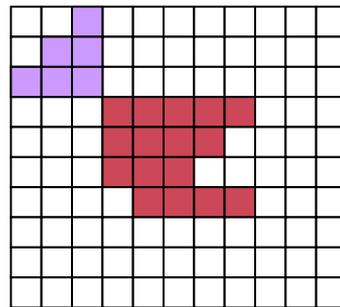
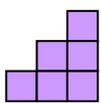
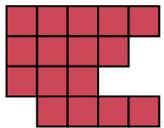
MEGADOCK System ver.2.5

ドッキング計算

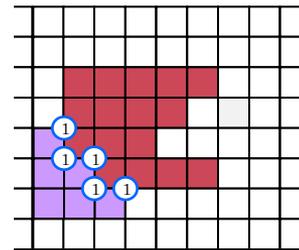
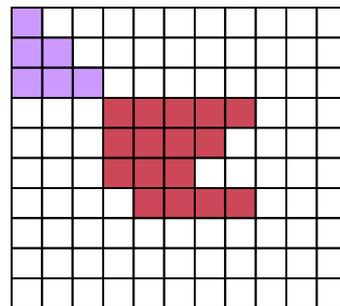
- **剛体ドッキング** (Grid base & Scoring function) (Katchalski-Katzir E., *et al.*, *PNAS*, 1992.)
 - 構造の形状相補性等に基づいて評価
 - タンパク質/RNAをボクセル空間で表現
 - 特定のスコアリングモデルに基づいてスコアを付与
 - それらの畳み込み和で評価関数を表す

タンパク質
(Receptor)

RNA
(Ligand)



best score = 8



best score = 5

Example of translational search in 2D

回転角: 15度刻み
3600通りの回転パターン

Docking rPSCスコア

- **rPSC (real Pairwise Shape Complementarity)** (Ohue M, 2010.)

– 実数値のみで表される形状相補性スコア

$$G_R(l, m, n) = \begin{cases} \# \text{ of R atoms within } (3.6\text{\AA} + \text{R atom } r_{vdW}) \text{ open space} \\ -27 \text{ inside of the R} \end{cases}$$

$$G_L(l, m, n) = \begin{cases} 0 & \text{solvent accessible surface layer of the L} \\ 1 & \text{solvent excluding surface layer of the L} \\ 2 & \text{core of the L} \\ 0 & \text{open space} \end{cases}$$

$$\text{rPSC}(\alpha, \beta, \gamma) = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N G_R(l, m, n) G_L(l + \alpha, m + \beta, n + \gamma)$$

※ **R** : Receptor (protein)
L : Ligand (RNA)

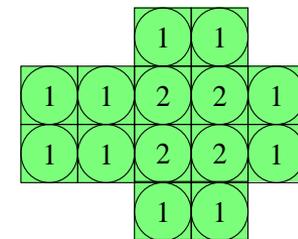
Previous Model: **PSC** (ZDOCK)

Chen R, Weng Z. *Proteins*, 2003.

Using **complex number**

1	2	3	3	3	2	1
2	-27	-27	-27	-27	-27	2
3	-27	-27	-27	-27	-27	2
3	-27	-27	-27	5	2	1
3	-27	-27	-27	5	2	1
3	-27	-27	-27	-27	-27	2
2	-27	-27	-27	-27	-27	2
1	2	3	3	3	2	1

Protein



RNA

Docking Electrostaticsスコア

- FTDock potential with CHARMM19 parameter

- voxel i (l, m, n)の電荷を決定する

- 規則に従ってタンパク質原子の電荷を付与

q_{atom}

- voxelに分割 $q_{\text{atom}} \rightarrow q(l, m, n)$

- voxel i の電界 φ_i を計算する

$$\varphi_i = \sum_{j \in \mathcal{V}} \frac{q_j}{\varepsilon(r_{ij})r_{ij}}, \quad \varepsilon(r) = \begin{cases} 4 & (r \leq 6\text{\AA}) \\ 38r - 224 & (6\text{\AA} < r < 8\text{\AA}) \\ 80 & (8\text{\AA} \leq r) \end{cases}$$

$$\text{ELEC}(\alpha, \beta, \gamma) = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N E_R(l, m, n) E_L(l + \alpha, m + \beta, n + \gamma)$$

$$E_R(l, m, n) = \begin{cases} \varphi(l, m, n) & (\text{entire voxel excluding core}) \\ 0 & (\text{core of the receptor molecule}) \end{cases}$$

$$E_L(l, m, n) = q(l, m, n)$$

MEGADOCK評価関数

- 評価関数(ドッキングスコア)

$$S(\alpha, \beta, \gamma) = \Re \left[\sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N R(l, m, n) L(l + \alpha, m + \beta, n + \gamma) \right]$$

$$R(l, m, n) = G_R(l, m, n) + iE_R(l, m, n)$$

$$L(l, m, n) = G_L(l, m, n) + iwE_L(l, m, n)$$

rPSC

Electrostatics

- RNA分子を扱うための拡張

- MEGADOCK(ver. 2.2)はタンパク質にしか対応していなかった

- 核酸分子への対応

- 新たにCHARMM27の電荷パラメータの追加

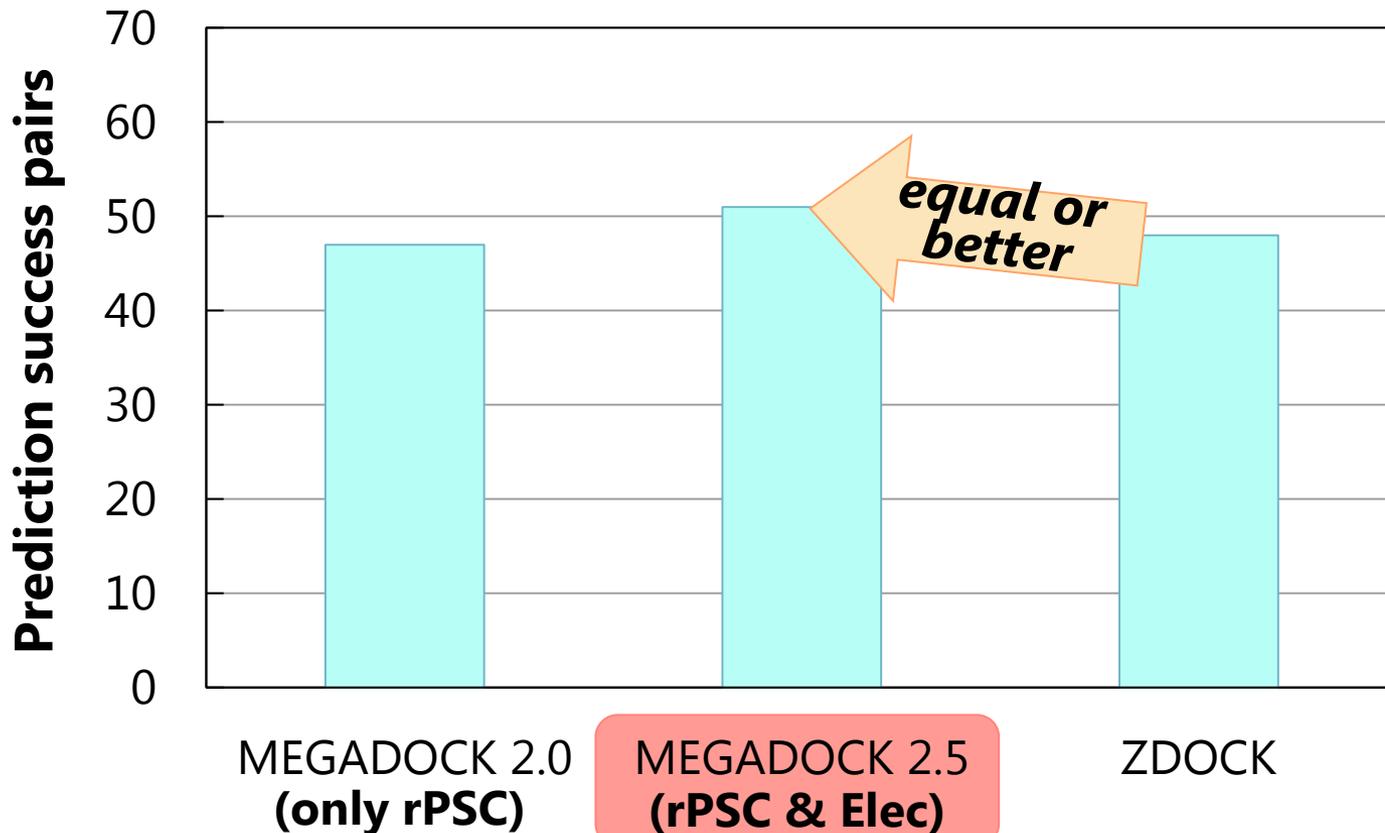
- 核酸分子に含まれる原子のvan-der-Waals半径パラメータの追加

→MEGADOCK ver. 2.5

再ドッキング精度(Protein-RNA)

Protein-RNA redocking test (78 pairs)

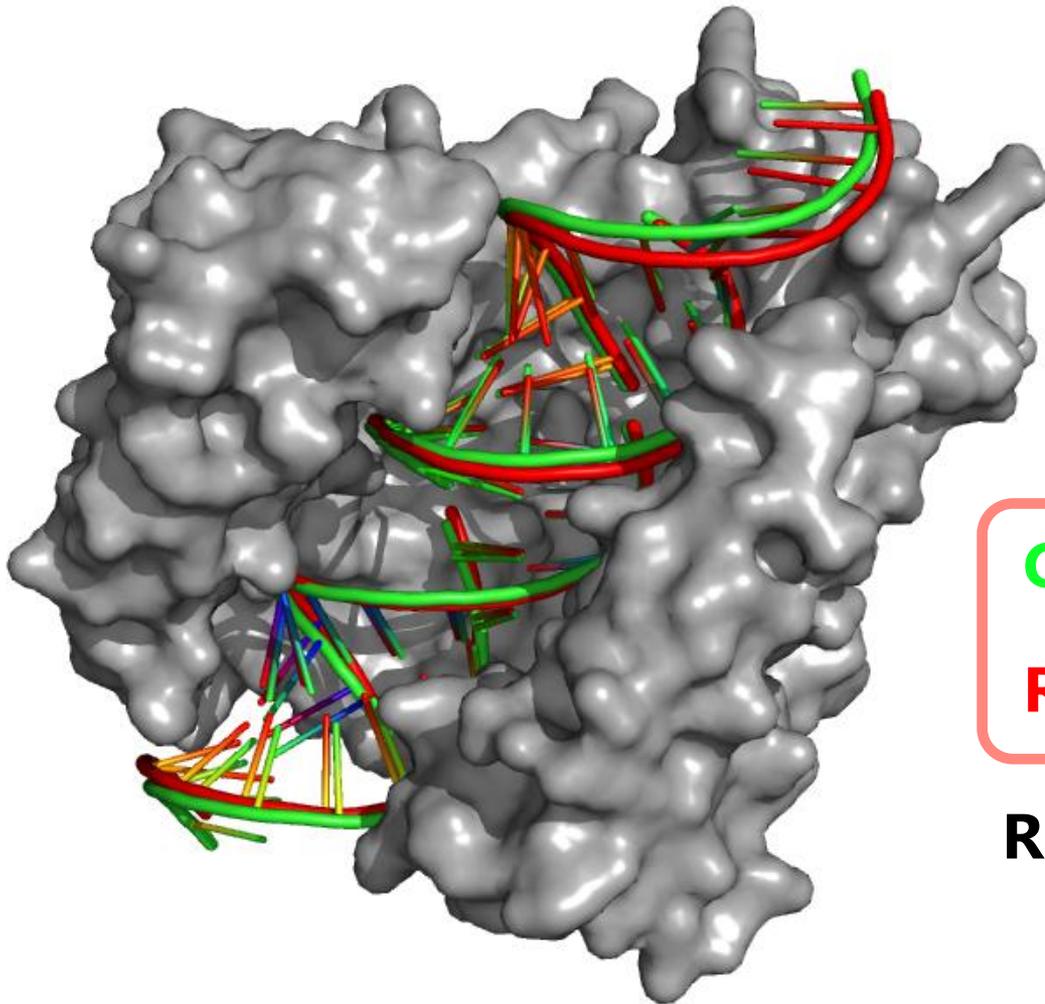
- PDBのタンパク質-RNA複合体を分離して再びドッキング
- (タンパク質側をフィットさせたときの)RNAのRMSDが5Å以下のとき「成功」



Predicted complex

PDB ID: 2NUG

RNase III from *Aquifex aeolicus*
complexed with ds-RNA



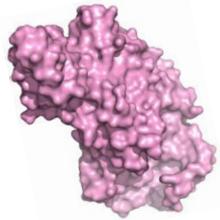
Green : MEGADOCK 2.5
(1st ranked decoy)

Red: X-ray structure

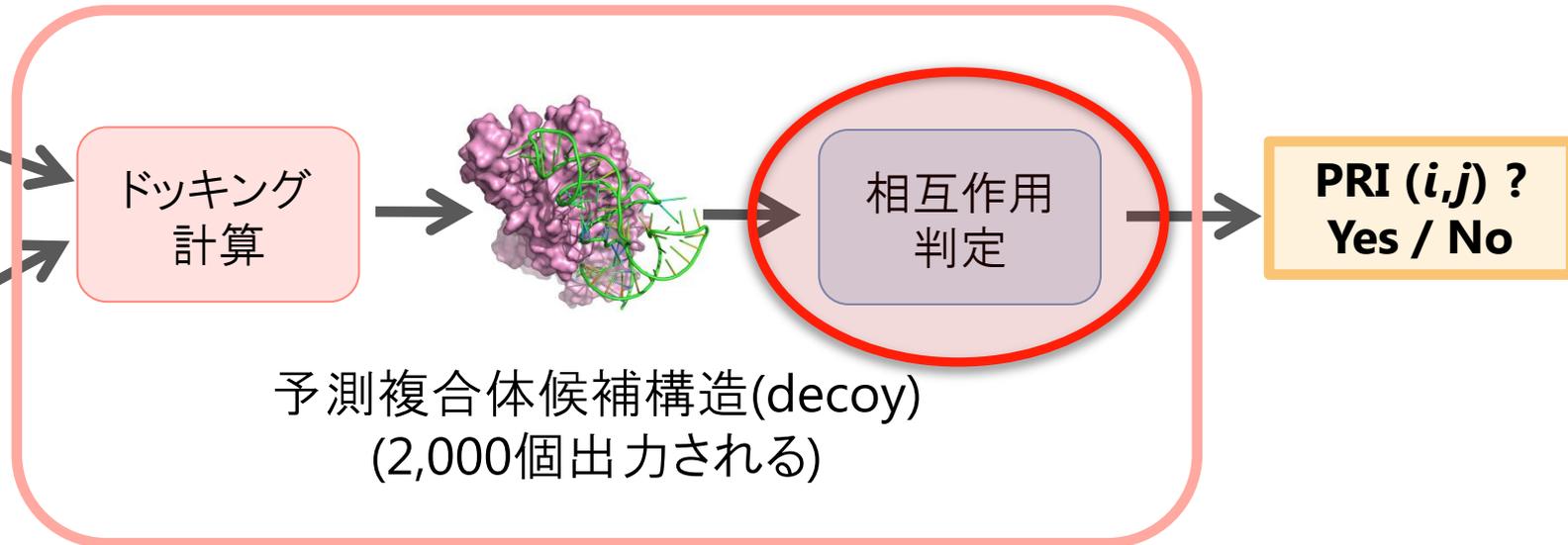
RNA RMSD = 1.45Å

予測方法

タンパク質 i



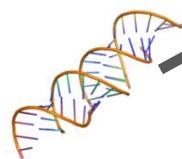
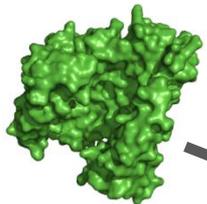
RNA j



MEGADOCK System ver.2.5

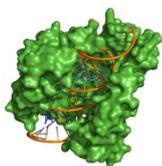
相互作用判定

タンパク質 i

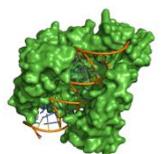


RNA j

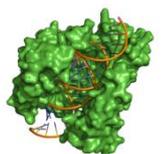
ドッキング
計算



$$S_1 = 14606.26$$

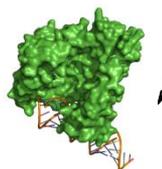


$$S_2 = 12794.21$$



$$S_3 = 9905.73$$

⋮



$$S_{2,000} = 3989.78$$

Mean μ
S.D. σ

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



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

ドッキング計算結果からの相互作用判定

• 相互作用判定の手順

1. MEGADOCKによって2,000個の候補構造を生成
 - 各角度ごとにスコア最大値をとる平行移動位置を出力
 - 得られた3,600個の候補構造の上位2000個を対象とする
2. PRIスコア E を計算する

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

S_1 : 1位の候補構造のドッキングスコア

μ : 1位から2000位までのドッキングスコアの平均

σ : 1位から2000位までのドッキングスコアの標準偏差

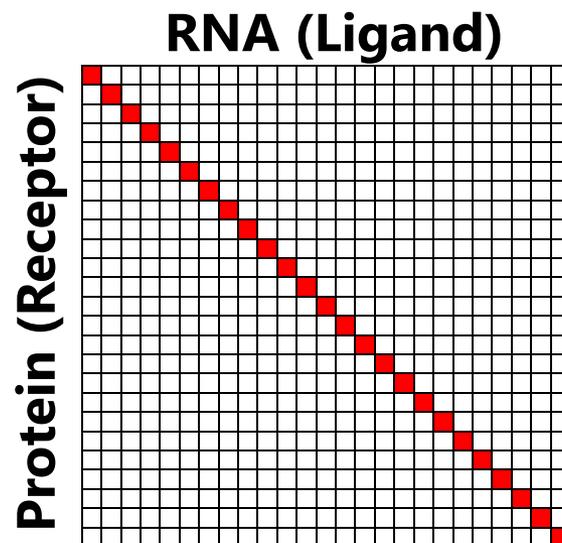
3. 閾値 E^* を超えるかどうかで相互作用しているか否かを予測

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

網羅的PRI予測 評価実験

• 評価方法

- PDB中のタンパク質-RNA複合体結晶構造78個を利用
 - PISCESサーバ(Wang, 2005)で配列冗長性を排除したもの
 - 解像度3.0Å以下
 - 相互の配列一致度30%以下
 - X線結晶構造
- $78 \times 78 = 6,084$ 通りの網羅的計算
- 対角成分の78ペアを正例,
それ以外の6,006ペアを負例とした



使用したPDB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, 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)

予測結果の評価について

予測結果

	R _A	R _B	R _C
P _A	○	○	○
P _B	×	○	×
P _C	○	×	×

正しいPRI

	R _A	R _B	R _C
P _A	○	×	×
P _B	×	○	×
P _C	×	×	○

評価

	R _A	R _B	R _C
P _A	TP	FP	FP
P _B	TN	TP	TN
P _C	FP	TN	FN

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

例 (TP=2, TN=3, FP=3, FN=1)

Sensitivity = 0.667

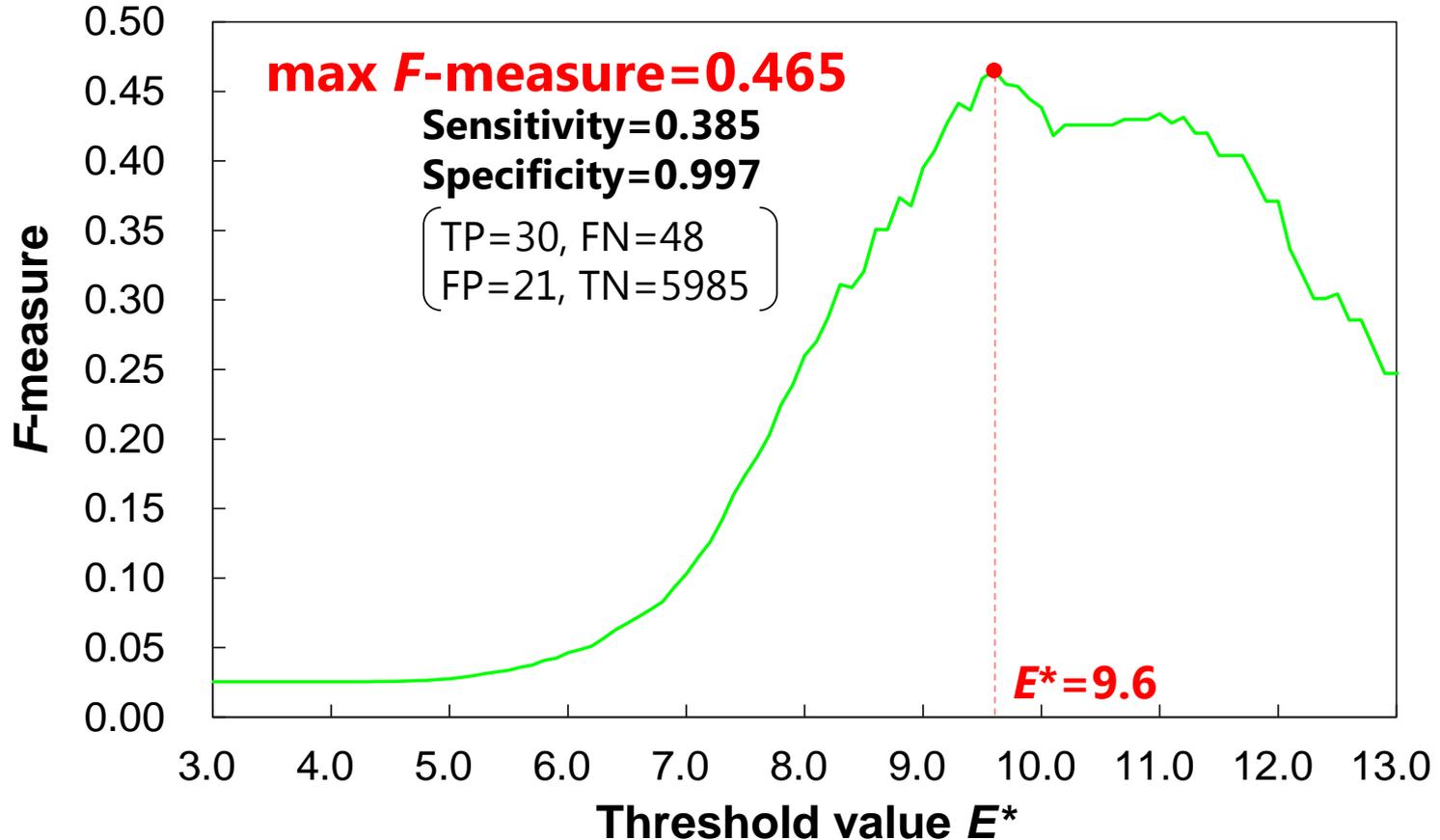
Specificity = 0.5

F-measure = 0.5

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

78 × 78 PRI予測結果

Threshold value E^* - F -measure plot



参考

MEGADOCKによる網羅的PPI予測(44 × 44 benchmark) → **F -measure=0.41**

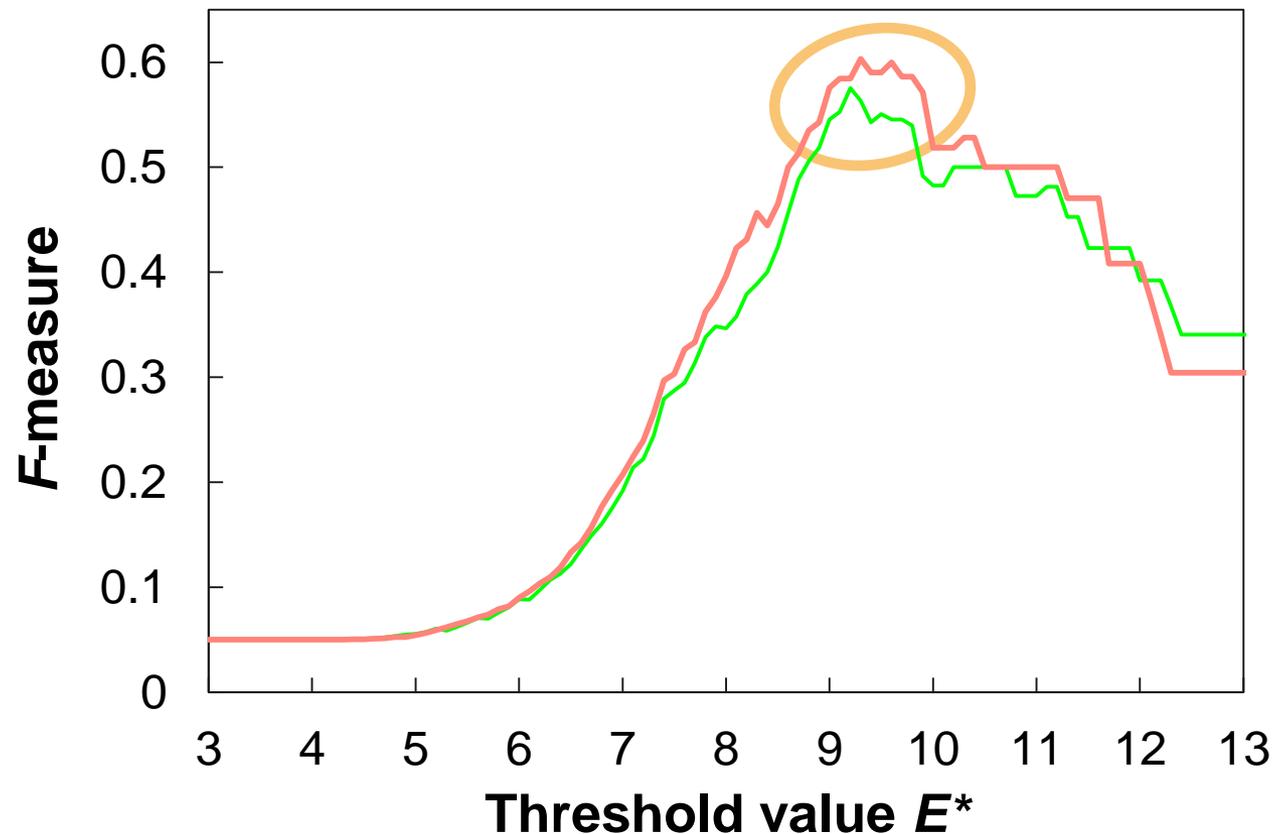
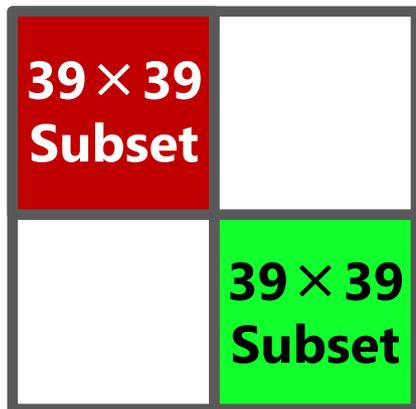
(Ohue M, 2010.)

78 × 78 PRI予測結果(2-fold)

Threshold value E^* - F -measure plot

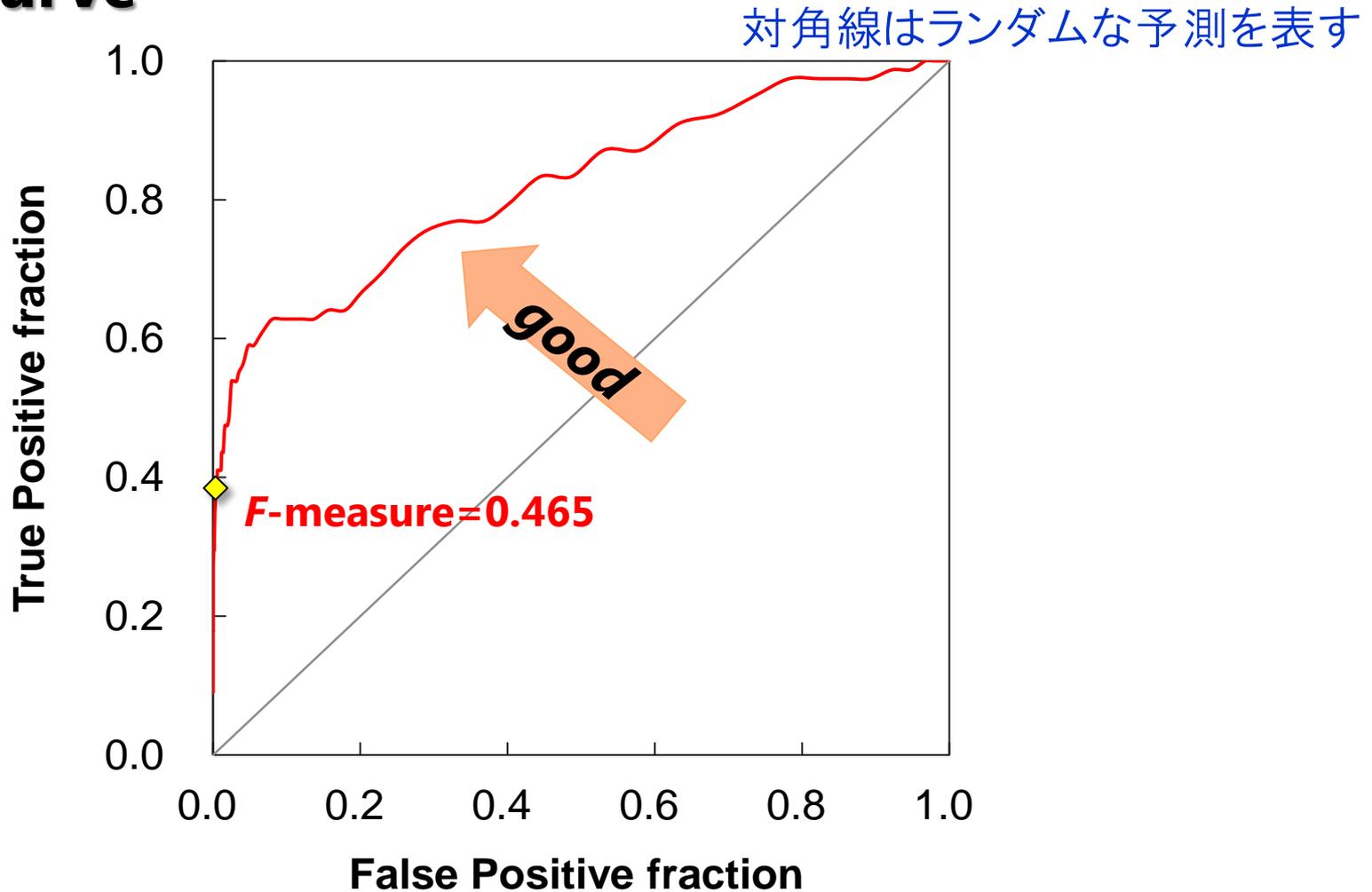
2-fold cross validation (2 dataset of 39 × 39)

78 × 78 dataset



78×78 PRI予測結果

ROC Curve



78 × 78 PRI予測結果

**PRI map
(78x78)**

F-measure=0.465

Sensitivity=0.385

Specificity=0.997

TP=30, FN=48

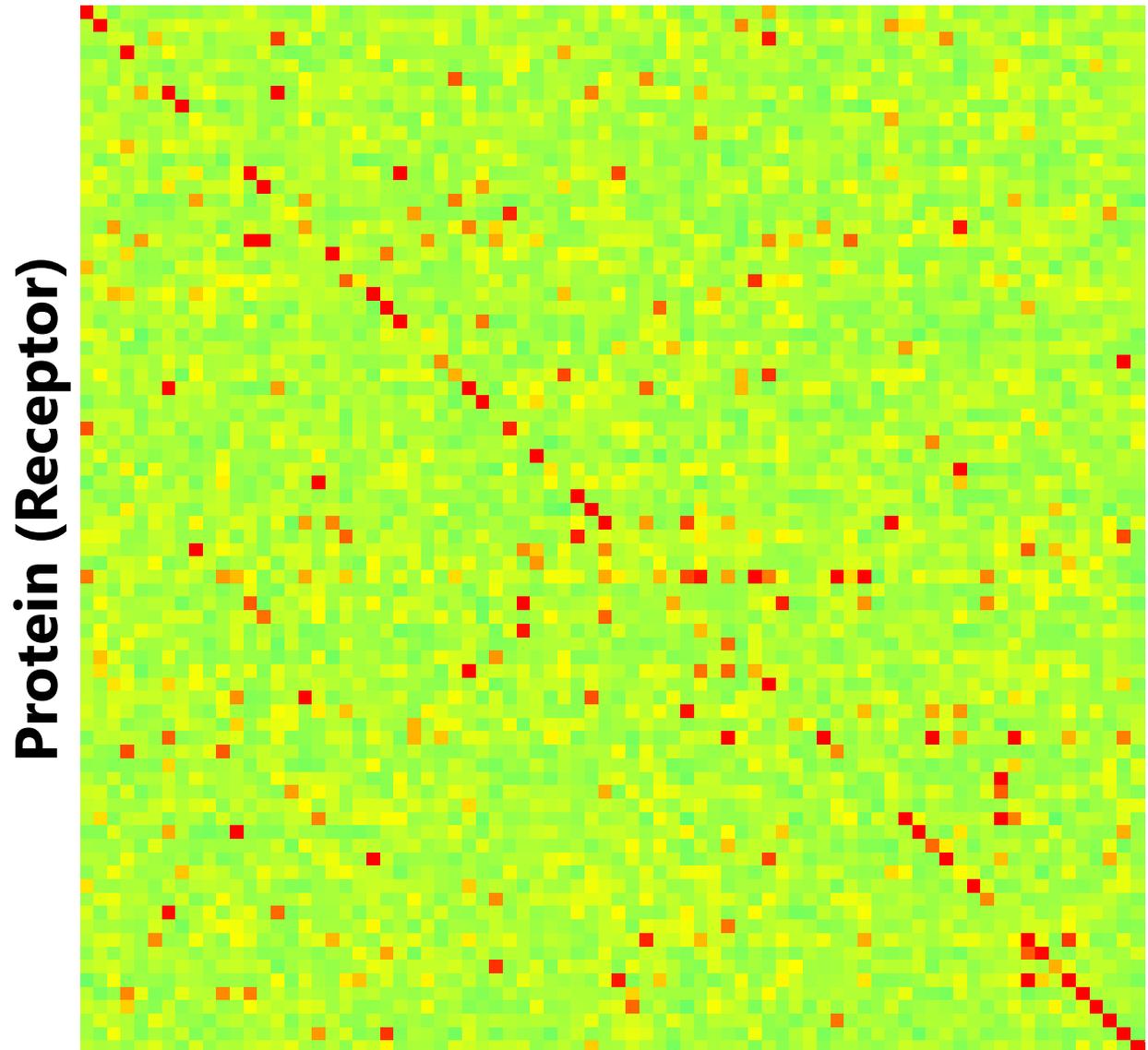
FP=21, TN=5985

PRI E-value



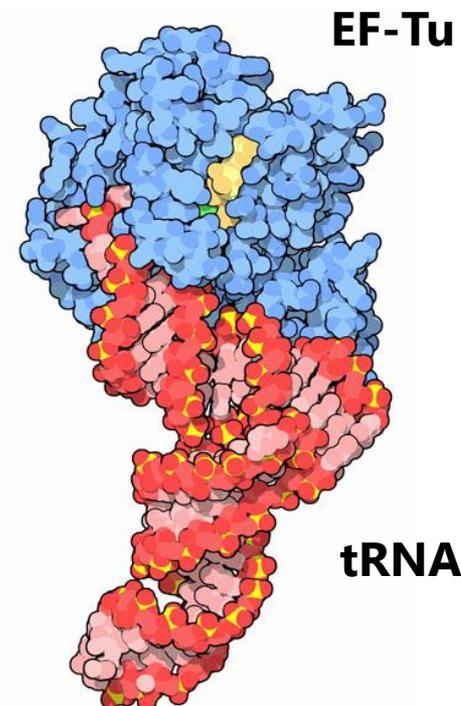
$E > 9.6 \rightarrow$ Positive

RNA (Ligand)



EF-Tu と tRNA の相互作用予測

- 伸長因子(Elongation Factor, EF)
 - 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うタンパク質
 - **EF-Tuタンパク質**
 - EF-TuがtRNAをリボソームに運ぶ
 - tRNAのコドンが正しいとEF-Tuの形が変化してGTPが切断, tRNA分離
 - EF-TuとEF-Tsが結合, 再利用できる形にする



<http://www.pdb.org/pdb/101/motm.do?momID=81>

実際の応用のための萌芽的実験として
EF-TuとtRNAの相互作用予測を試みた

データセット

- 使用したPDBデータ

- EF-TuとtRNAの複合体 (EF-TuとtRNAの両方を使用)

- 2WRN *Thermus thermophilus* (tRNA^{Thr})
- 1TTT *Thermus aquaticus* (tRNA^{Phe})

- EF-TuとEF-Tsの複合体 (EF-Tu側のみ使用)

- 1AIP *Thermus thermophilus*
- 1EFU *Escherichia coli*
- 1XB2 *Bos taurus*

EF-Tu 5種, tRNA 2種

使用したデータの配列

タンパク質(EF-Tu)のSequence Identity

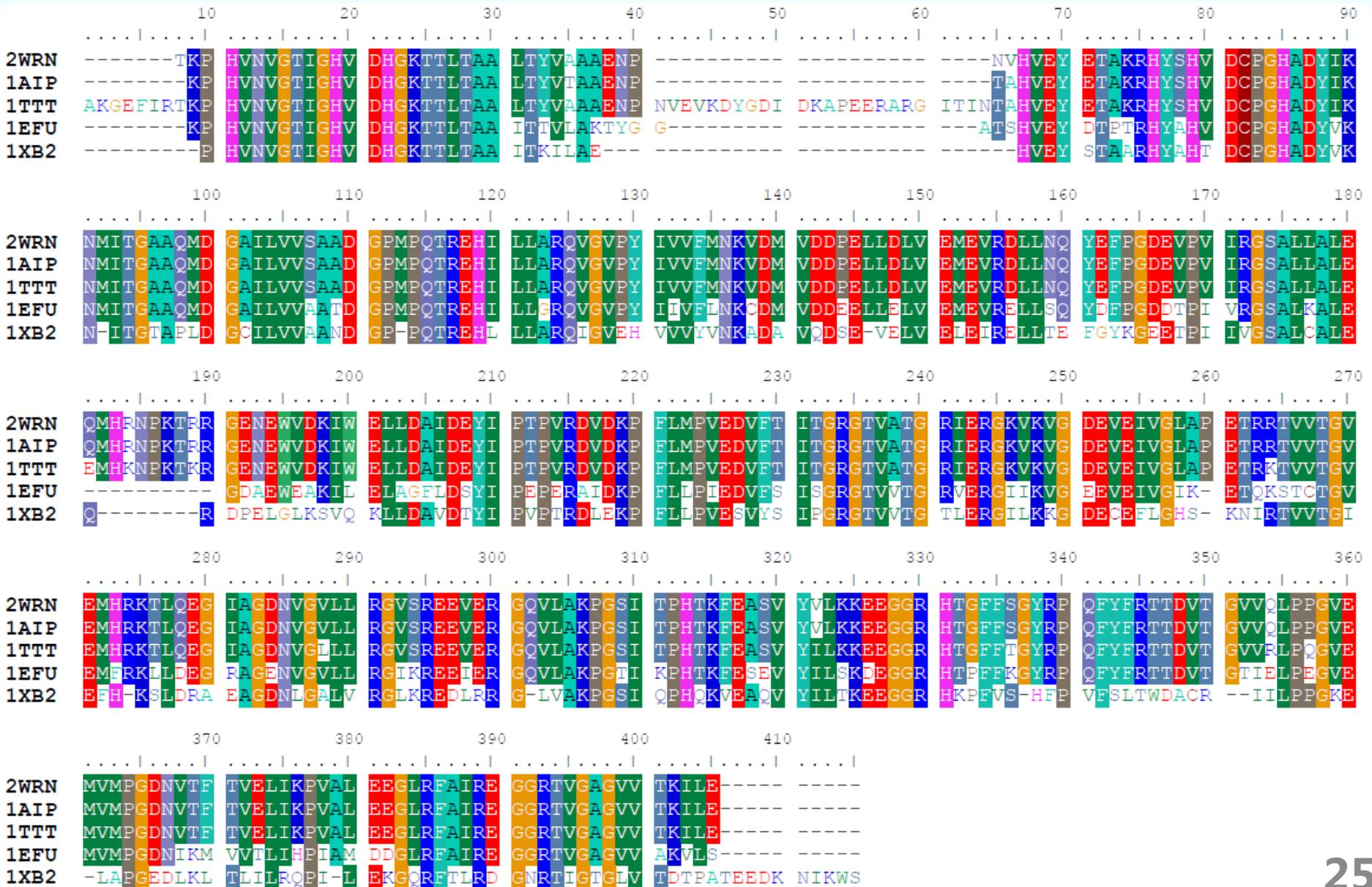
	<i>T. thermophilus</i>	<i>T. aquaticus</i>	<i>E. coli</i>	<i>B. taurus</i>	
	2WRN	1AIP	1TTT	1EFU	1XB2
2WRN		99%	92%	72%	57%
1AIP	99%		91%	72%	57%
1TTT	92%	91%		69%	52%
1EFU	72%	72%	69%		56%
1XB2	57%	57%	52%	56%	

tRNAのSequence Alignment

T. aquaticus **1TTT** GCGGAUUUAG CUCAGUUGG AGAGCGCCAG ACUGAAGAUC UGGAGGUC CU
T. thermophilus **2WRN** GCUGAUUAUAG CUCAGUUGGU AGAGCGCACC CUUGGU GAGG GU GAGGUC GG

 GUGUUCGAUC CACAGAAUUC GCACCA-
 CAGUUCGAAU CUGCCUAUCA GCACCAU **Seq.Id. 64%**

EF-Tuタンパク質の配列アラインメント



EF-Tu と tRNA の相互作用予測

EF-Tu	tRNA	Score E
2WRN	2WRN	8.11
1AIP	2WRN	8.04
2WRN	1TTT	7.54
1XB2	2WRN	7.15
1TTT	1TTT	6.77
1EFU	2WRN	6.25
1TTT	2WRN	6.11
1XB2	1TTT	5.89
1EFU	1TTT	5.68
1AIP	1TTT	5.63



Color table	
1AIP	<i>Thermus thermophilus</i>
1EFU	<i>Escherichia coli</i>
1XB2	<i>Bos taurus</i>
2WRN	<i>Thermus thermophilus</i>
1TTT	<i>Thermus aquaticus</i>

おわりに

- **ドッキング計算を用いたPRI予測手法を提案した**
 - 従来のタンパク質間相互作用予測システムの拡張を行った
 - 78×78網羅的PRI予測に適用し, F 値=0.465の予測を得た
- **今後の課題**
 - RNAの柔軟性を考慮した手法の考案
 - 構造サンプリングによるアンサンブルドッキングの利用
 - 可視化手法との連携(山本, SIGBIO25(31), 2011)

謝辞

- インド工科大学のM Michael Gromiha博士からEF-TuとtRNAに関する助言を頂きました。感謝致します。
- 本研究は以下の支援を受けて行われました。
 - 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」
 - 科学研究費補助金(基盤研究(B)19300102, 特別研究員奨励費23・8750)