

リランキングを用いた タンパク質ドッキングの精度向上と 網羅的タンパク質間相互作用予測手法への応用

大上 雅史¹

松崎 裕介¹

松崎 由理¹

佐藤 智之²

秋山 泰¹

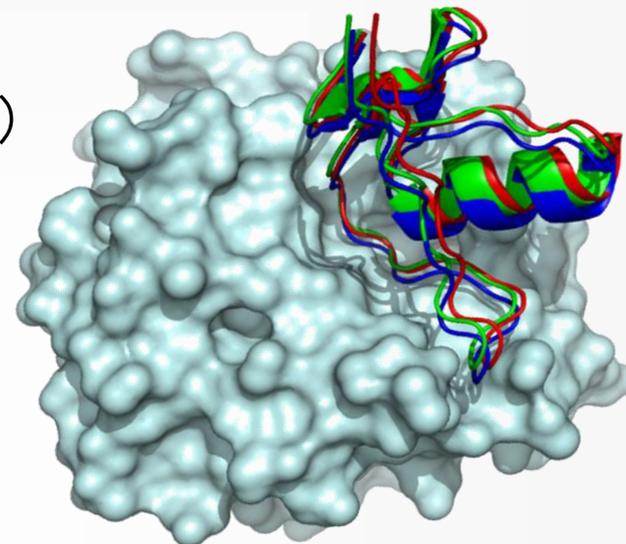
¹東京工業大学 大学院情報理工学研究科 計算工学専攻

²みずほ情報総研株式会社

本研究について

- **タンパク質間相互作用 (PPI) ネットワーク**

- 生命現象における中心的役割を担う
- 創薬ターゲットとして注目 (PPI阻害薬)



- **計算機によるPPI予測の発展**

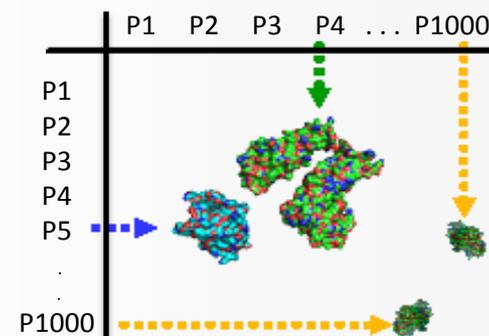
- 配列モチーフからの予測
- 文献情報からの予測
- ドッキングシミュレーション
 - 分子動力学法
 - **形状相補性に基づいたタンパク質ドッキング (PPD) ← 本研究の対象**

- **目的: タンパク質ドッキングによるPPIネットワーク予測**

- 増えつつある立体構造データの活用
- PPI予測システムの確立へ

本研究の概要

- タンパク質の立体構造情報を用いてPPIを予測する
 - 大量のタンパク質群同士の総当たり(網羅的)な予測
 - MEGADOCKを用いたドッキング予測
 - 1ペアに対し, 複数のドッキング予測構造を返す
 - ZRANKによるエネルギー計算
 - 予測構造の再評価, リランキング
 - 構造のクラスタリング, PPI判定



- 実験
 - ドッキング予測実験
 - 正解構造が得られているペア44組の複合体再現実験
 - PPI予測実験
 - 総当たり(44 × 44通り)での相互作用予測精度の検証実験

- all-to-all タンパク質ドッキング専用システム
 - 計算の高速性に特化した網羅的計算専用PPDシステム
 - 現行Ver.は2.1 (M.Ohue, *IPSJ SIG-BIO* 17(11), 2009)
 - 形状相補性 G と静電的相互作用 E を考慮

$$\mathbf{R}(l, m, n) = G_R(l, m, n) + iE_R(l, m, n)$$

$$\mathbf{L}(l, m, n) = G_L(l, m, n) + iwE_L(l, m, n)$$

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

- 形状相補性 G は独自のrPSC Score(Ohue, 2009)を利用
- 静電的相互作用 E はCHARMMとFTDockのポテンシャルを利用

予測出力数の改良

- MEGADOCKやZDOCKの予測構造数
 - $N^3 \times 3600$ (回転角15deg.刻みの場合)
 - ZDOCKでは各角度ごとに最良のスコアとなるもの1つのみ報告
 - 最大3600個, 標準で2000個
 - MEGADOCK2.1では各角度ごとの出力数 t を指定可能
 - 出力される予測構造数は最大 $3600 \times t$ 個
- 仮説:「相互作用を有するタンパク質ペアはドッキング予測における結合部位が空間的に密に分布する」
 - 出力数を増やせば相互作用予測の精度向上につながるはず
 - 既存のZDOCKでは対応できない
(ただしドッキング予測システムとしては優秀である)

リランキング

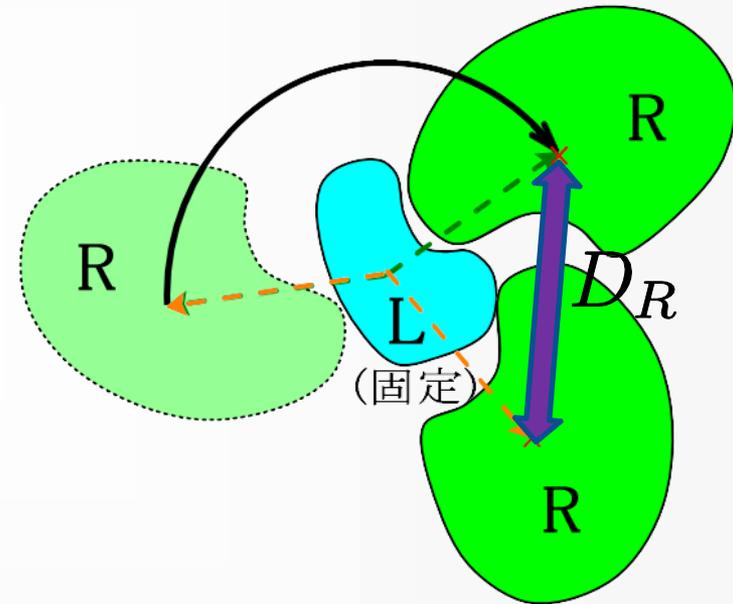
- ドッキングシステムの出力する予測複合体
 - 結合エネルギーが高く現実的でないものも含まれる
 - エネルギー計算を全ての候補構造について行うのが理想
 - しかし候補数が多すぎて不可能
 - 絞り込んだ候補構造群ならば少し精密なモデルを適用可能
- **ZRANK** (B. Pierce, *Proteins* 67(4), 2007.)
 - ZDOCK用の予測構造リランキングシステム
 - 限られた範囲内での擬似的なエネルギー計算
 - ファン・デル・ワールスエネルギー E_{vdW}
 - 静電的相互作用エネルギー E_{Elec}
 - 脱溶媒和エネルギー E_{DS}
 - 上位にドッキング正解構造を集中させることができる

クラスタリング

- 2つの予測構造間の非類似度 D

➡ $D = D_L + D_R$

- D_L : レセプターRを固定したときの
リガンド間のEuclid距離
- D_R : リガンドLを固定した時の
回転後のレセプター間の距離



- クラスタリング手法には群平均法が良い
 - 評価実験で示唆されている(Yusuke Matsuzaki, *IP SJ SIG-BIO* 13(5), 2008)

PPI予測の方法

MEGADOCK 網羅的ドッキング予測

- 出力複合体数は $2000 \times t$ 個

ZRANK リランキング

- $2000 \times t$ 個をリランキングして, 2000個取り出す

クラスタリング

- 2000個の予測をクラスタリングにかける

PPI評価値の計算

- メンバ数が多いクラスタの中で, 最大のドッキングスコア(z値)を持つものを評価値 E とする

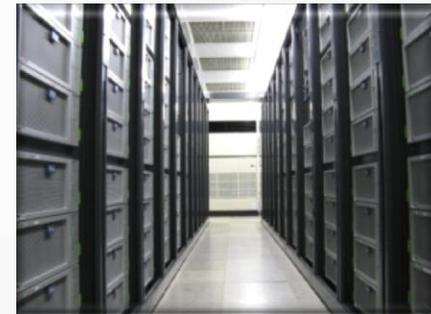
PPI判定

- 評価値 E がある閾値 E^* 以上となるタンパク質ペアを「相互作用する」と判定

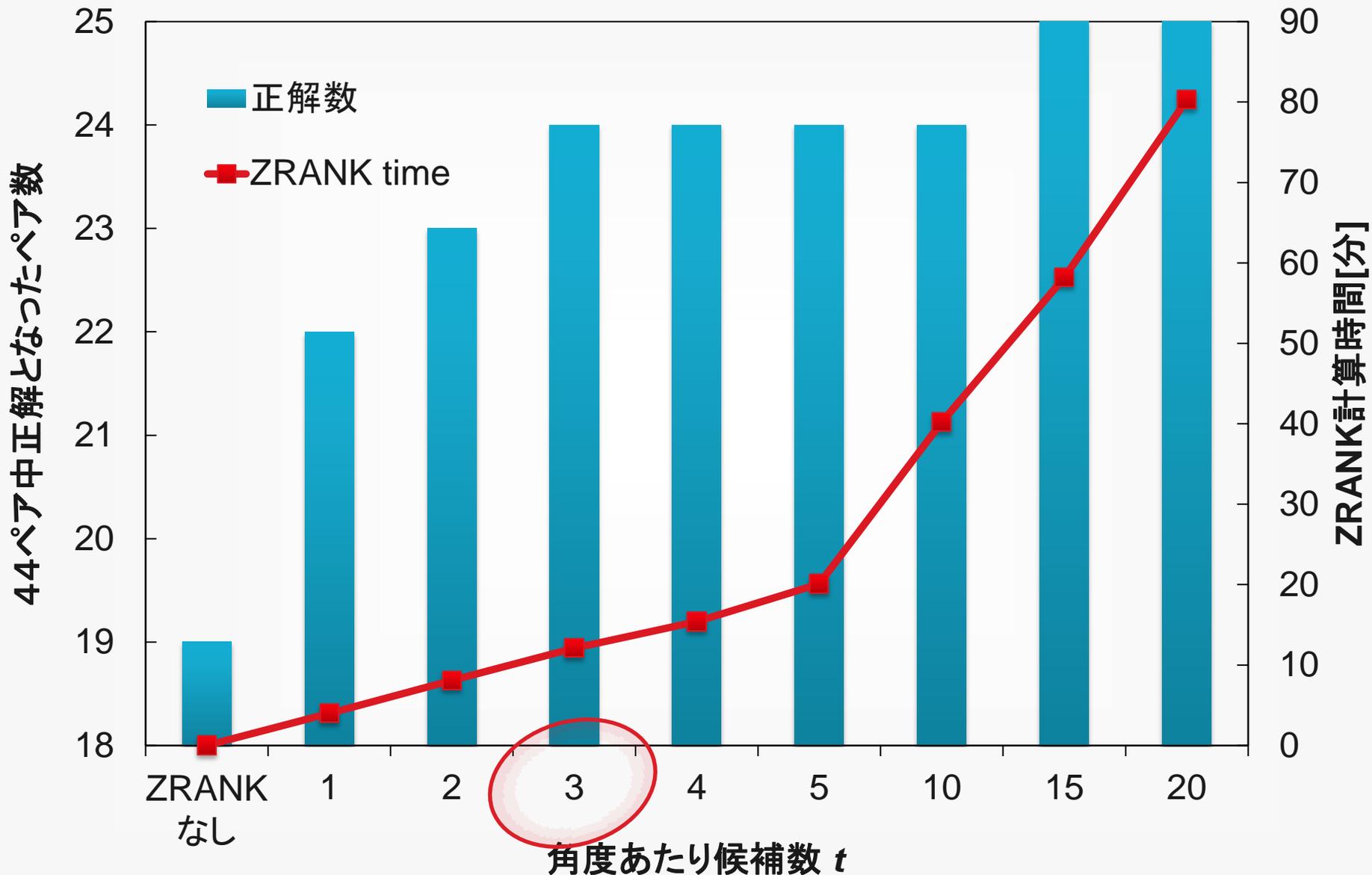
実験1 ZRANK&MEGADOCKの性能

- Protein-Protein Benchmark2.0によるドッキング予測
 - モノマー同士のペア44例を使用
 - X線結晶解析による正解構造とのリガンドRMSDによる評価
 - L-RMSDが5Å以下の予測構造を「**正解構造**」とする
 - 複数の予測のうち、**1位が正解構造だったペアを「正解」とする**
 - 角度ごとの出力数 t を変えて出力し、ZRANKをかける
 - $t = 1, 2, 3, 4, 5, 10, 15, 20$
 - TSUBAME supercluster systemを利用
 - ZRANKの平均計算時間

角度あたり候補数 t	1	2	3	4	5	10	15	20
計算時間[分]	4.0	8.1	12.1	15.4	20.1	40.2	58.2	80.3



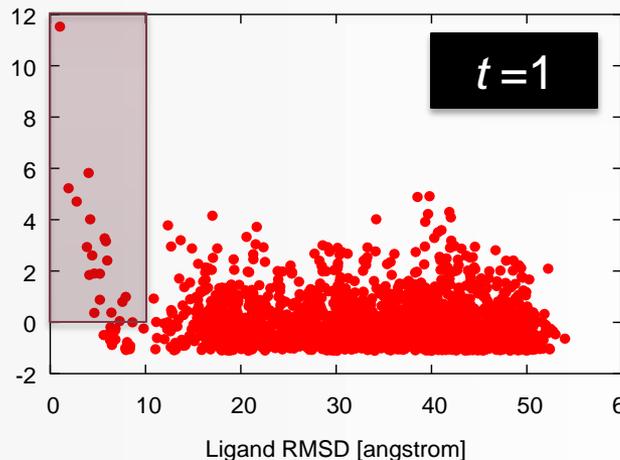
ドッキング予測結果



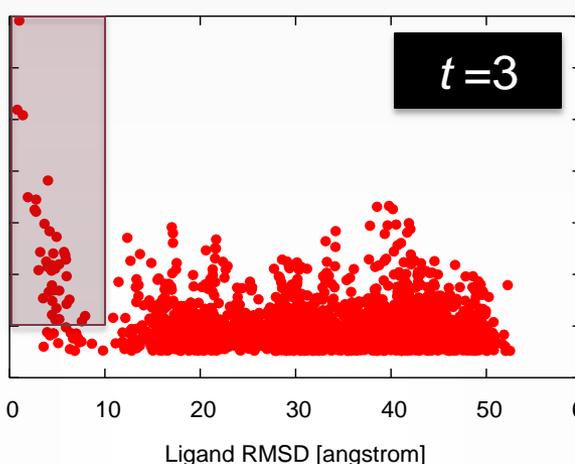
L-RMSDとドッキングスコアの分布

PDB-ID 1CGI

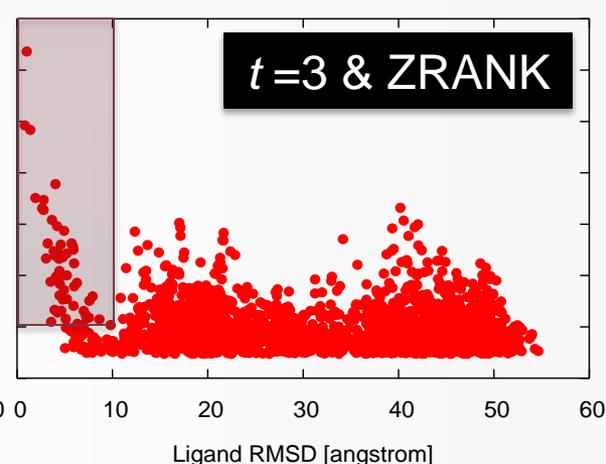
1CGI, (a)t=1



1CGI, (b)t=3

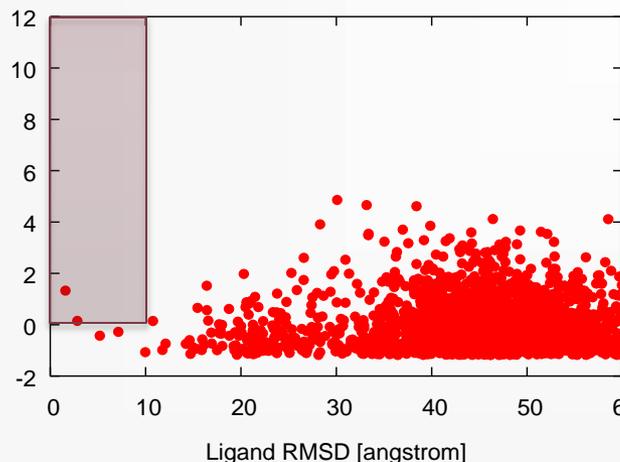


1CGI, (c)t=3 with ZRANK

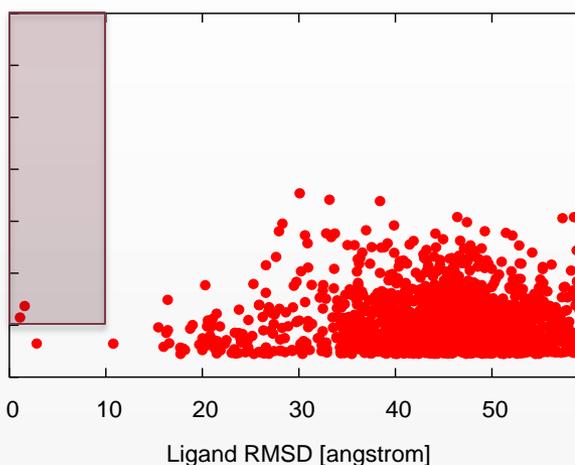


PDB-ID 1KTZ

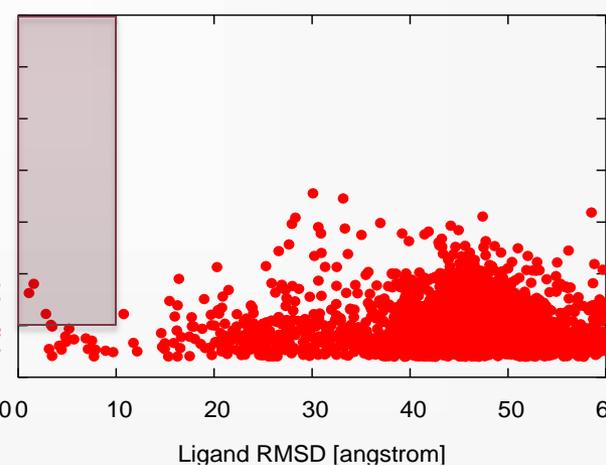
1KTZ, (a)t=1



1KTZ, (b)t=3

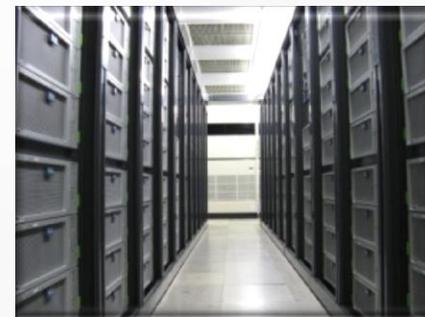
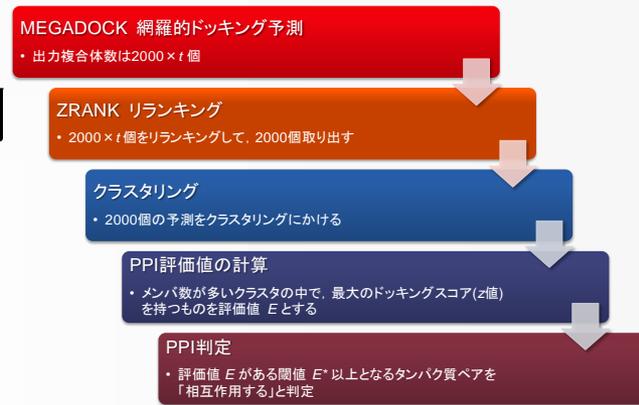


1KTZ, (c)t=3 with ZRANK

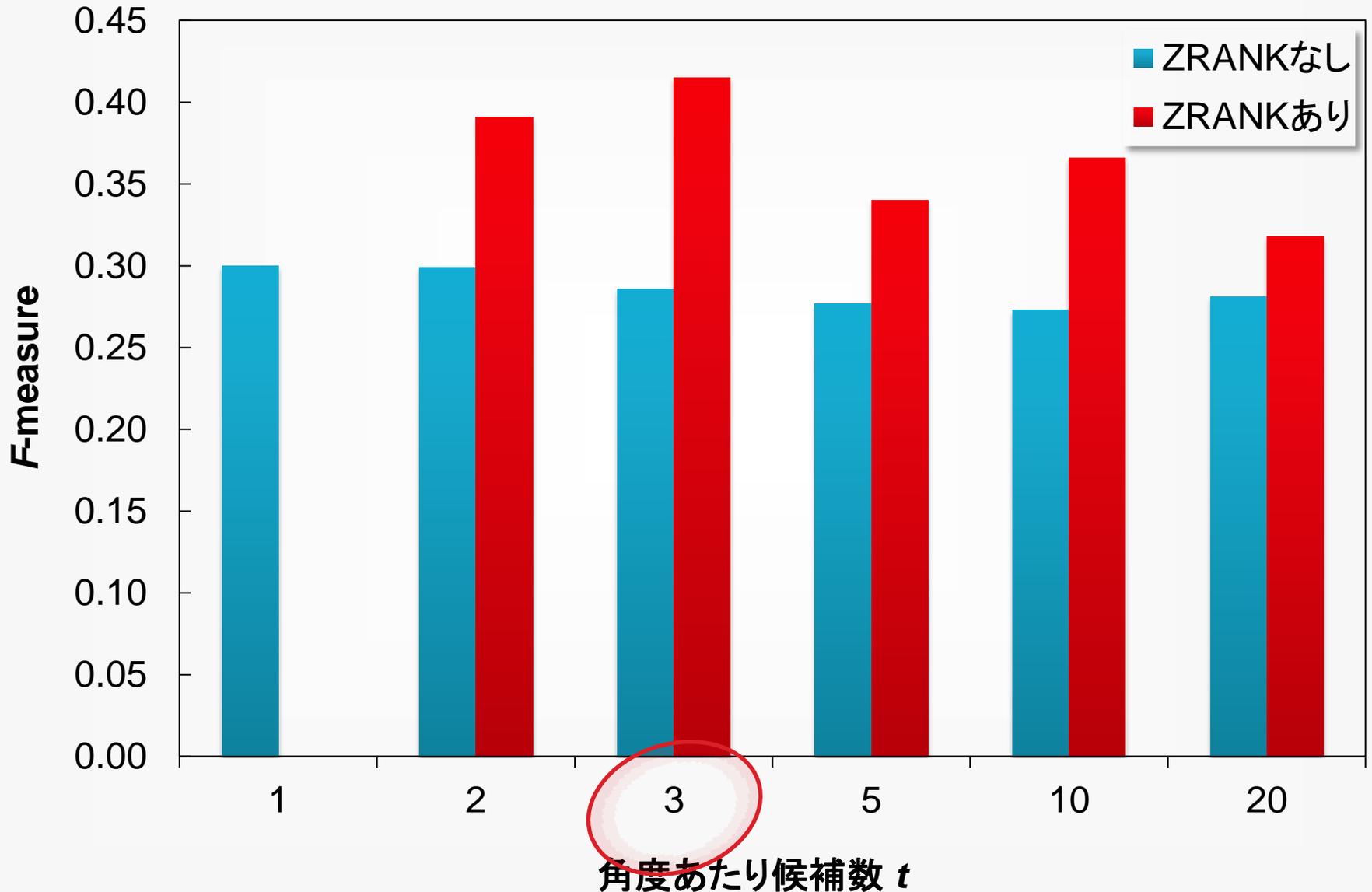


実験2 all-to-all PPI Prediction

- Protein-Protein Benchmark 2.0を使ったPPI予測実験
 - モノマー同士のペア44例を使用
 - 44 × 44の総当たりドッキングによる予測
 - PPI予測フローにしたがって予測を行う
 - F 値(F -measure)によって評価
 - 本来のペアの相手以外との複合体は存在しないもの(負例)とした
 - 角度ごとの出力数 t は1,2,3,5,10,20
 - TSUBAME supercluster systemを利用
 - 2000個のクラスタリング・・・32分



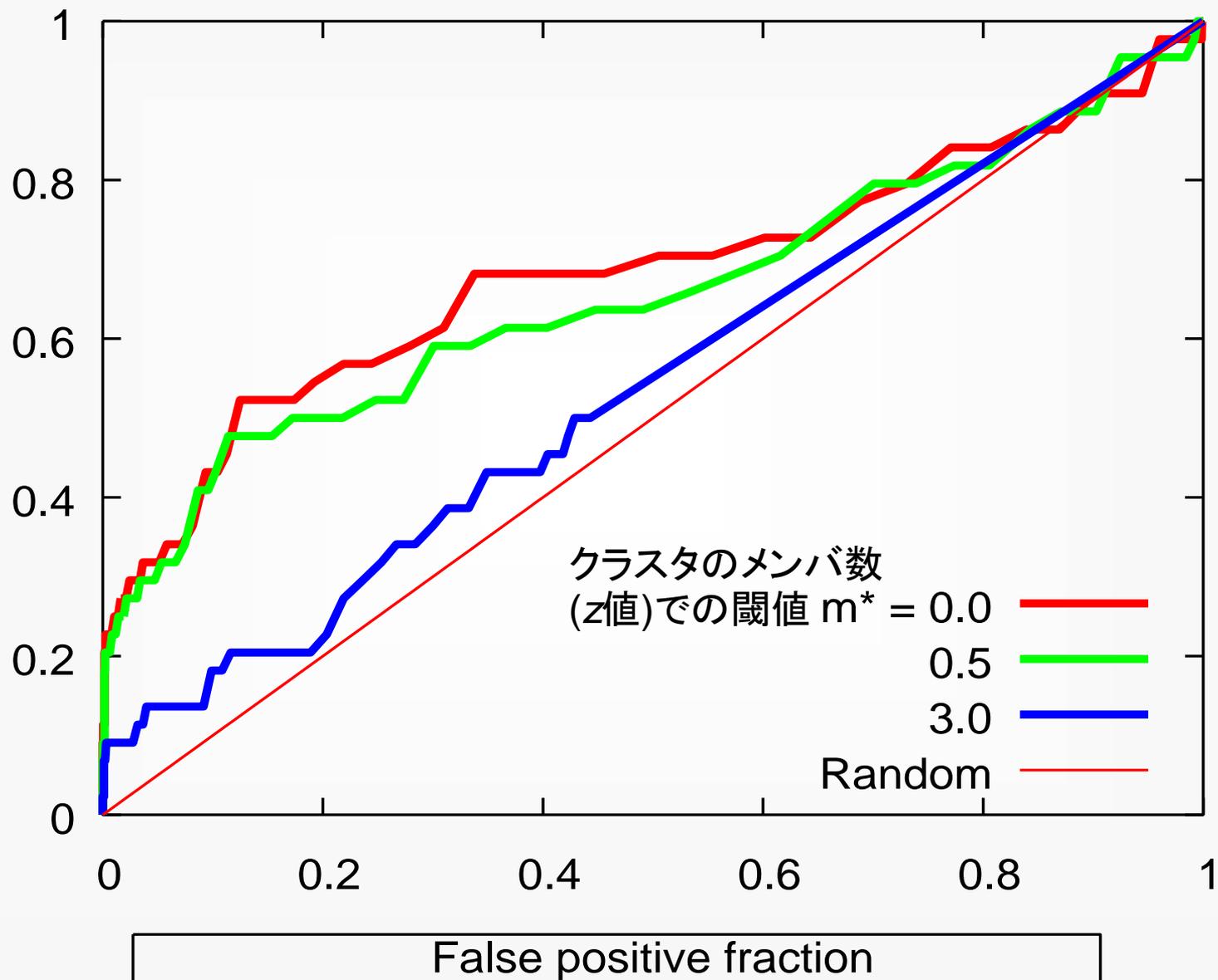
相互作用予測結果



従来手法との比較

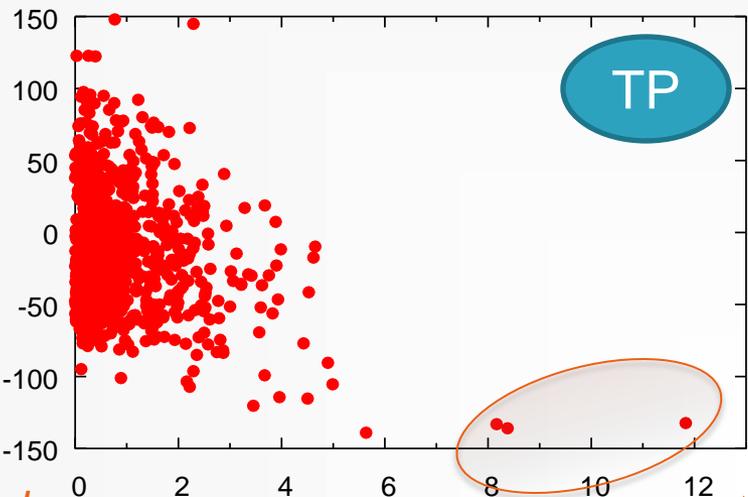
- F 値による比較
 - トップスコアを利用した方法(η -method) F 値=0.150 (Ohue, 2009)
 - Boostingを用いた方法 F 値=0.251 (Ohue, 2009)
 - 本研究における方法($t=3$) F 値=0.415
- 予測精度の向上が見られた
- ZRANKによるリランキングが網羅的PPI予測にも効果があることが実証された
- $t=3$ のときに F 値が最大となった
 - 計算時間は t に対して線型に増加するため、計算時間との兼ね合いから、本研究では $t=3$ が最も良いと判断した。

ROC曲線($t=3$ の場合)



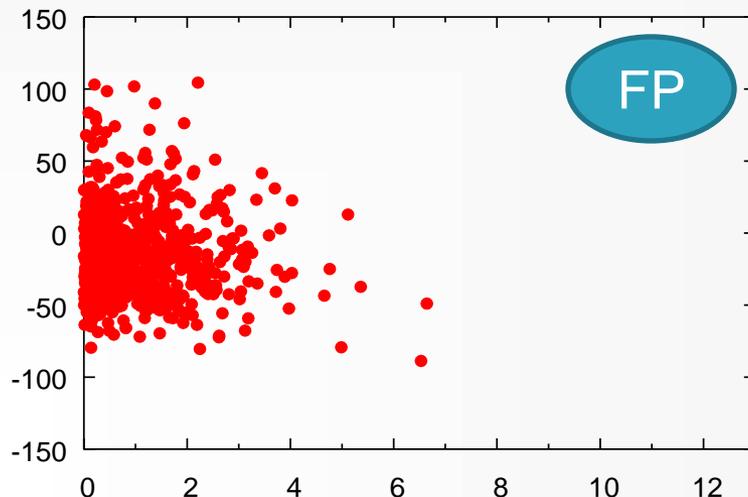
Docking Score vs. ZRANK Score プロット

1CGI_r-1CGI_I(TP), t=3



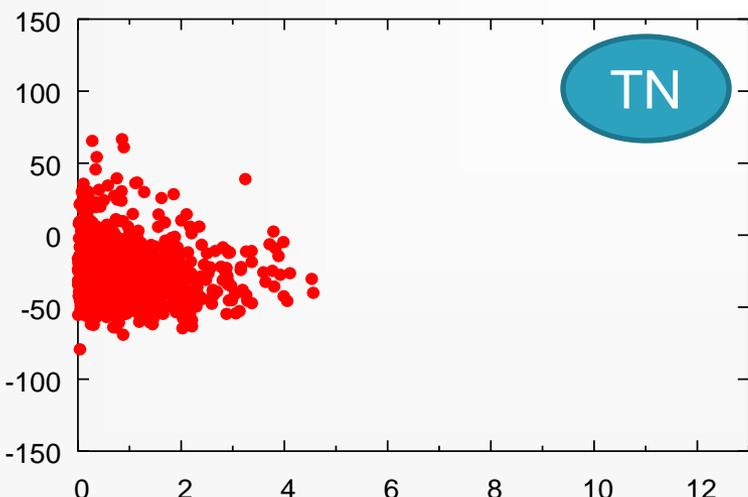
Z Score of MEGADOCK Docking Score

1UDI_r-1EAW_I(FP), t=3



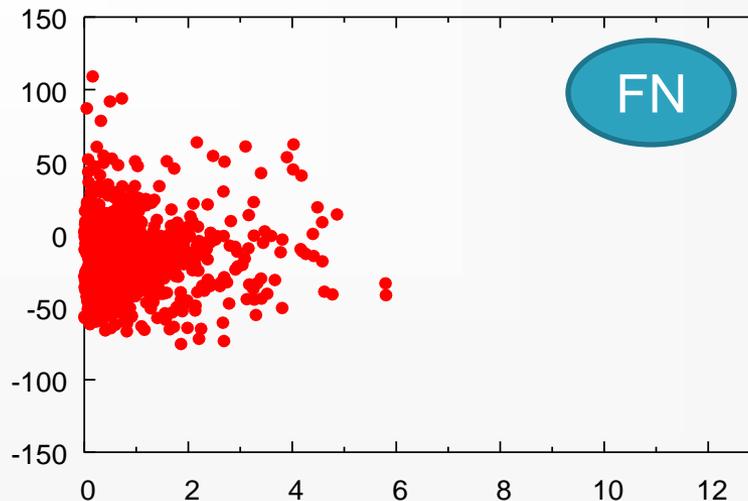
Z Score of MEGADOCK Docking Score

1E96_r-1PPE_I(TN), t=3



Z Score of MEGADOCK Docking Score

1EWY_r-1EWY_I(FN), t=3

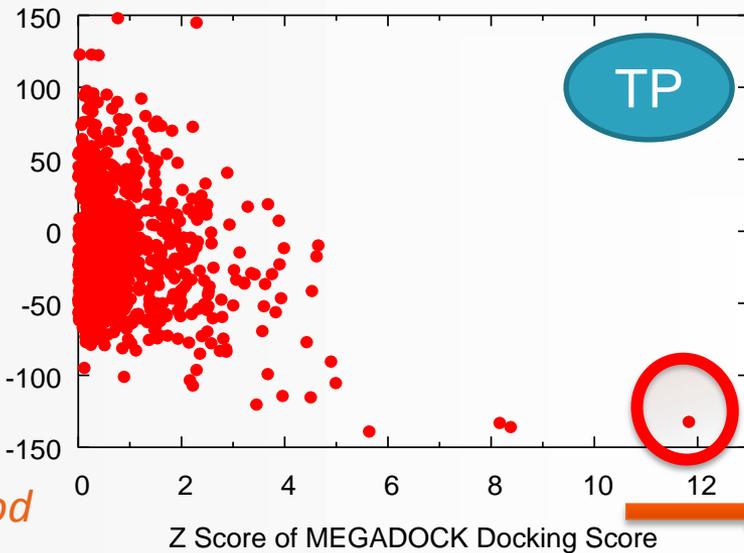


Z Score of MEGADOCK Docking Score

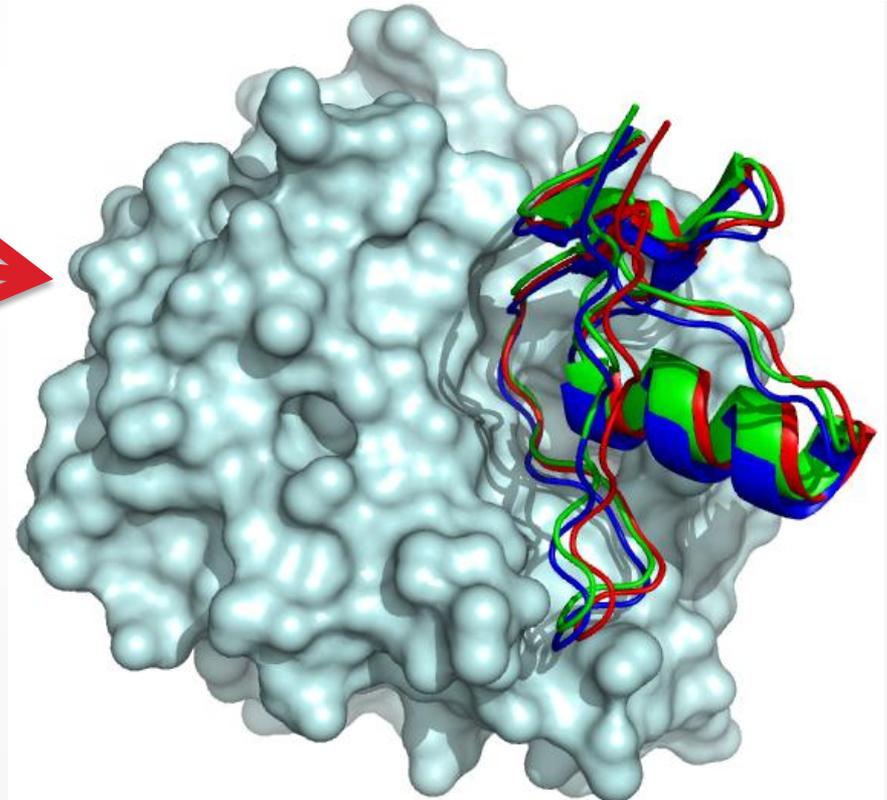
複合体の立体構造の例

Docking Prediction Example (PDB id = 1CGI)

1CGI_r-1CGI_l(TP), t=3



Blue : ZDOCK 3.0 (1st, RMSD=1.33 Å)
Green : MEGADOCK 2.1 (1st, RMSD= 1.02 Å)
Red: X-ray struct. (Correct Docking Pose)

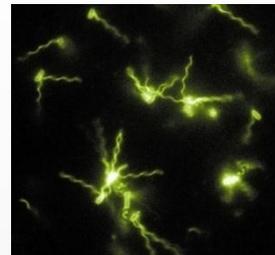


おわりに

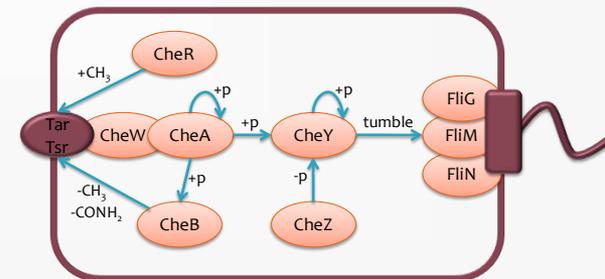
- MEGADOCKとZRANKを組合せて精度向上を図った
 - ドッキング予測性能が向上した(正解数19/44→24/44)
 - 計算時間などの兼ね合いで $t=3$ を最良とした
 - PPI予測についても, 従来手法より良いF値=0.415を得た
- ➡ ZRANKはドッキング・PPIともに効果があることを実証した

- 今後の課題

- ドッキング出力数やクラスタリングのパラメータの最適化
- 実問題への適用
 - 細菌走化性系
 - ヒトEGFRシグナル伝達系
 - 肺ガンに関わる重要な系



<http://www.mcb.harvard.edu/>



謝辞

- 本研究は以下の支援を受けて行われたものである
 - 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)