

# 物理化学的相互作用の導入による

## 網羅的タンパク質間相互作用予測システムの高精度化

大上 雅史<sup>†</sup> 松崎 裕介<sup>†</sup> 松崎 由理<sup>†</sup>

佐藤 智之<sup>‡</sup> 秋山 泰<sup>†</sup>

<sup>†</sup>東京工業大学 大学院情報理工学研究科 計算工学専攻

<sup>‡</sup>みずほ情報総研株式会社



# 目次

- **本研究の説明**
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめと今後の課題

# 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されている
- **計算機によるPPI予測の発展**
  - Molecular Dynamics
  - 配列モチーフからの予測
  - 文献情報からの予測
  - **Protein Docking**
    - 形状相補性に基づいたタンパク質ドッキング
- **タンパク質ドッキングによるPPIネットワーク予測**
  - MEGADOCKの改良によるドッキング精度の向上

# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめと今後の課題

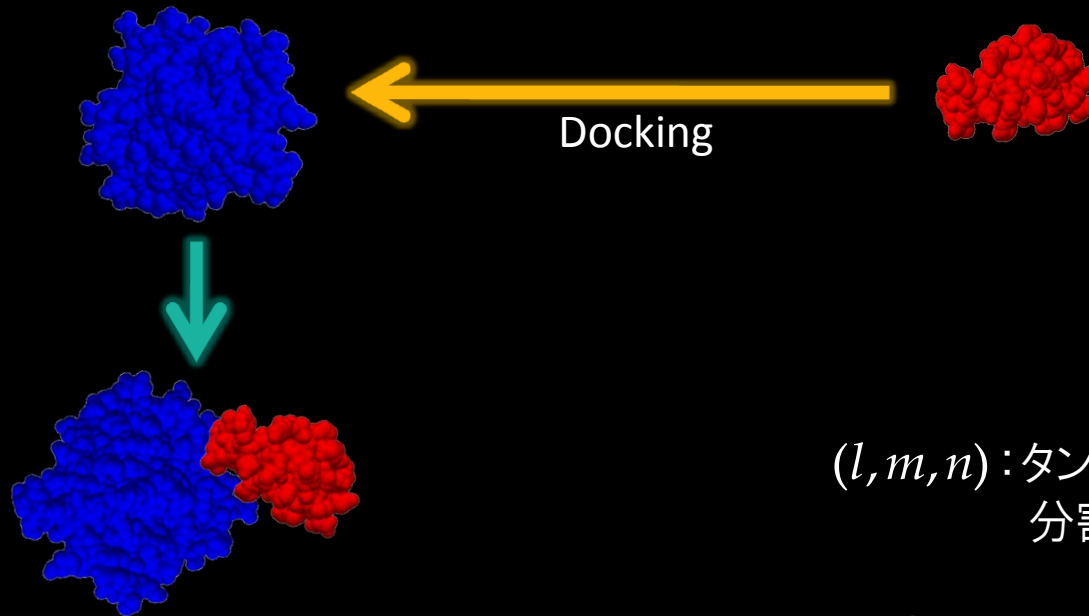
# タンパク質ドッキング

- 代表的なドッキングソフトウェア
  - MolFit (E. Katchalski-Katzir 1992～)
    - FFT Baseの形状相補性アルゴリズムを提案
  - FTDock (H.A. Gabb 1997)
    - MolFitの形状相補性スコアに静電的効果を追加
  - ZDOCK (Z. Weng 2002～)
    - 形状相補性, 静電, 疎水効果を考慮 (ver.2.3)
    - 自由エネルギー計算のFFTを追加 (ver.3.0)

# 形状相補性の計算

Receptor Protein  $\bar{a}$

Ligand Protein  $\bar{b}$



Complex  $\bar{c}$

$(l, m, n)$ : タンパク質を3次元voxellに  
分割したときの1つのvoxel

$(\alpha, \beta, \gamma)$ : Ligandの平行移動

$$\bar{c}_{\alpha, \beta, \gamma} = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \bar{a}_{l, m, n} \times \bar{b}_{l+\alpha, m+\beta, n+\gamma}$$

畳み込み和で表わされる

# MolFit

- 形状相補性計算にフーリエ変換(DFT)を利用

$$A_{o,p,q} = \text{DFT}[\bar{a}_{l,m,n}], \quad B_{o,p,q} = \text{DFT}[\bar{b}_{l,m,n}]$$

$$C_{o,p,q} = A_{o,p,q}^* B_{o,p,q} \quad \bar{a}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the Rec.} \\ \rho & \text{inside the Rec.} \\ 0 & \text{outside the Rec.} \end{cases}$$

$$\bar{c}_{\alpha,\beta,\gamma} = \text{IDFT}[C_{o,p,q}] \quad \bar{b}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the Lig.} \\ \delta & \text{inside the Lig.} \\ 0 & \text{outside the Lig.} \end{cases}$$

K-Kスコア

- 高速フーリエ変換(FFT)を用いて高速化が可能

$$O(N^6) \quad \longrightarrow \quad O(N^3 \log N)$$

# FTDock

- 形状相補性 (MolFitと同じ) と静電的相互作用を考慮
- Algorithm (Electrostatics)

- 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  $q_{\text{atom}}$
  - voxel に分割  $q_{\text{atom}} \rightarrow q_{l,m,n}$
- voxel  $i$  の電界を計算する

$$\varphi_i = \sum_j \frac{q_j}{\varepsilon(r_{ij}) r_{ij}}, \quad \varepsilon(r_{ij}) = \begin{cases} 4: & r_{ij} \leq 6\text{\AA} \\ 38r_{ij} - 224: & 6\text{\AA} < r_{ij} < 8\text{\AA} \\ 80: & r_{ij} \geq 8\text{\AA} \end{cases}$$

- Electrostatics スコア

minimum cutoff  $r_{ij} < 2.0\text{\AA} \rightarrow r_{ij} = 2.0\text{\AA}$

**Rec.**  $E_{\alpha}^{l,m,n} = \begin{cases} 0 & \text{core of molecule} \\ \phi^{l,m,n} & \text{surface and excluded core} \end{cases}$

core of molecule  
surface and excluded core

**Lig.**  $E_{l,m,n}^b = q_{l,m,n}^b$

**Complex**  $E_{\alpha,\beta,\gamma}^c = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N E_{l,m,n}^a E_{l+\alpha,m+\beta,n+\gamma}^b$



# ZDOCK

- 形状相補性, 静電的相互作用, 疎水性相互作用(Ver.2.3)
  - Ver.3.0ではさらに計算を複雑化
- 形状相補性には**PSCスコア** (pairwise shape complementarity)を使用

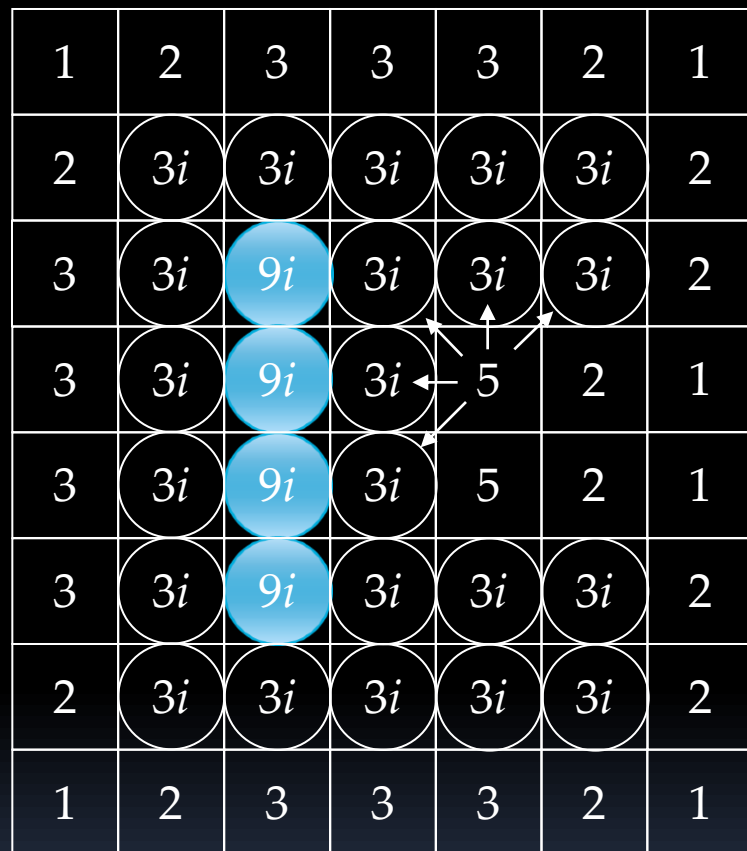
$$\mathcal{R}[\bar{a}_{l,m,n}] = \begin{cases} \# \text{ of R atoms within } (D + R \text{ atom } r) & \text{open space} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{H}[\bar{\rho}_{l,m,n}] = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if this grid is the nearest grid of a } \Gamma \text{ atom} \end{cases}$$

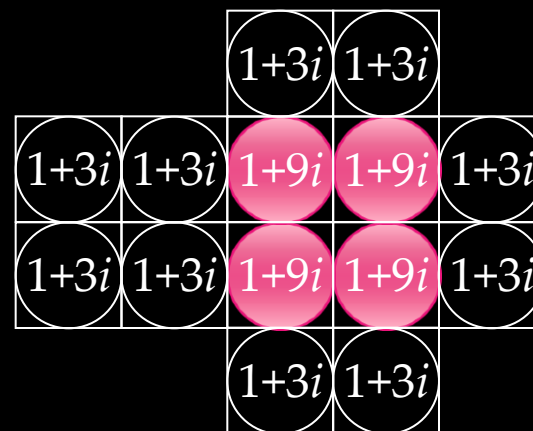
$$\mathcal{S}[\bar{a}_{l,m,n}] = \mathcal{S}[\bar{b}_{l,m,n}] = \begin{cases} 3 & \text{solvent } \textit{excluding} \text{ surface of the protein} \\ 9 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

$$\text{Score}_{\alpha,\beta,\gamma} = \mathcal{R} \left[ \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \bar{a}_{l,m,n} \bar{b}_{l+\alpha,m+\beta,n+\gamma} \right]$$

# ZDOCK PSC Score



Receptor PSC



Ligand PSC

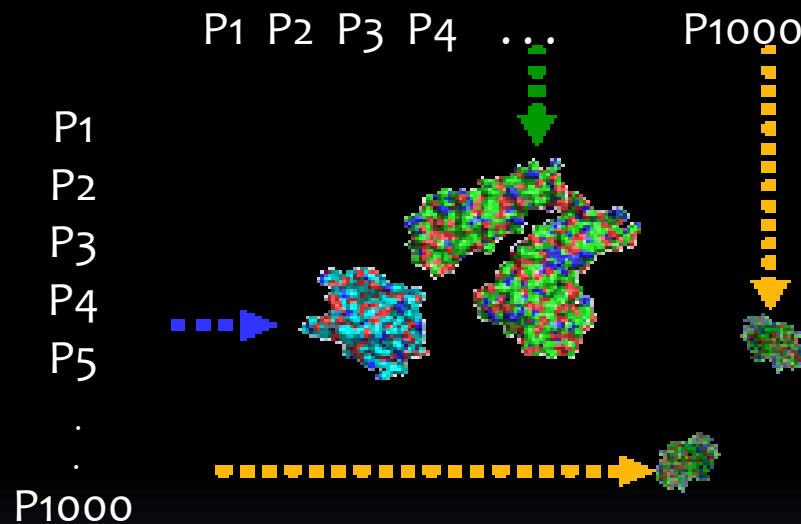
# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめと今後の課題

# all-to-all PPI Prediction

## ■ 網羅的PPI予測

- 多数のタンパク質からInteractionするペアを予測



- タンパク質ドッキングによって網羅的PPI予測を行う
  - 高速に計算できるドッキングシステムの欲求

# MEGADOCK System (ver. 1.0)

- all-to-all Dockingを高速に行うためのシステム
  - ドッキングに使うフーリエ変換の計算結果をライブラリ化
  - 形状相補性としてK-Kスコアを実装
    - MEGADOCK Ver.1.0 (Y. Akiyama, 2008)
- ZDOCKのPSCを使えることが理想
  - PSCは複素数表現である
    - 他の相互作用を考えた時に複素FFTの回数が増加する
    - 実数表現で性能の良い形状相補性スコアが欲しい
- rPSCスコアの提案

# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- **新形状相補性スコアrPSCの提案**
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめと今後の課題

## ■ real Pairwise Shape Complementarity

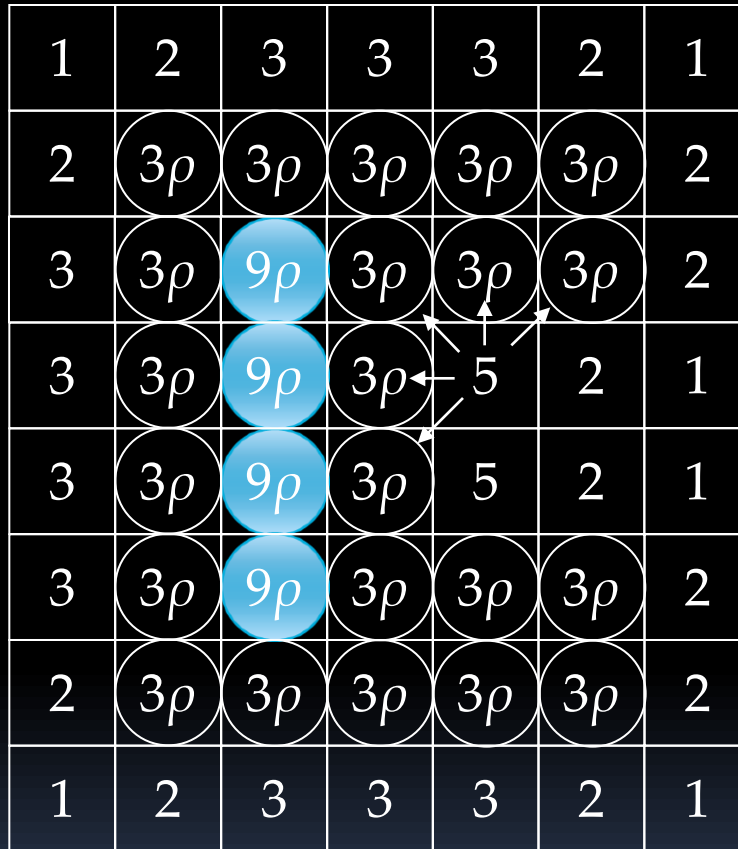
- ZDOCKのPSCを参考に、実数のみで表現

Rec.  $\bar{a}_{l,m,n} = \begin{cases} \# \text{ of } R \text{ atoms within } (D + R \text{ atom } r) & \text{open space} \\ 3\rho & \text{solvent excluding surface} \\ 9\rho & \text{core of } R \end{cases}$

Lig.  $\bar{b}_{l,m,n} = \begin{cases} 0 & \text{solvent accessible surface layer} \\ 1 & \text{solvent excluding surface layer} \\ \delta & \text{core of } L \\ 0 & \text{open space} \end{cases}$

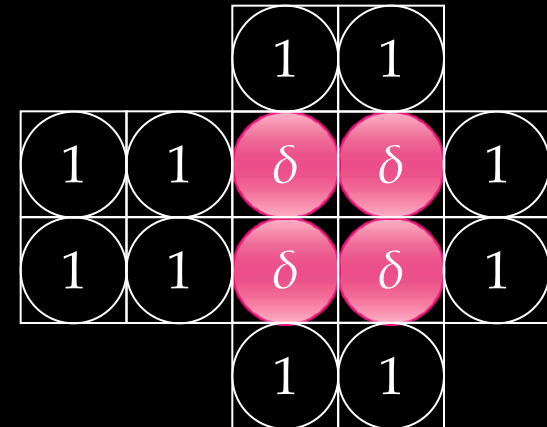
Complex  $\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \bar{a}_{l,m,n} \times \bar{b}_{l+\alpha,m+\beta,n+\gamma}$

# rPSC (real Pairwise Shape Complementarity)



Receptor rPSC

$\rho$ は負  $\delta$ は正



Ligand rPSC





# PSC と rPSC の比較

Receptor

PSC	Ligand		
	Core $1+9i$	Surface $1+3i$	Otherwise $0$
Core $9i$	-81	-27	0
Surface $3i$	-27	-9	0
Space near atoms $n$	$n$	$n$	0
Otherwise $0$	0	0	0

$$\rho = -3, \delta = 2$$

Ligand

Receptor

rPSC	Ligand		
	Core $2$	Surface $1$	Otherwise $0$
Core $-27$	-54	-27	0
Surface $-9$	-18	-9	0
Space near atoms $n$	$2n$	$n$	0
Otherwise $0$	0	0	0

# rPSCの利点

- 実数による表現
  - 虚数項に他の効果を入れられる
  - K-Kスコアよりも精密に形状相補性計算ができる
- 虚数項に何を入れるか
  - 静電的相互作用
  - 疎水性相互作用
  - その他

Dockingを考える上で重要な相互作用

# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめ

- FTDockの電界を利用
  - 静電的相互作用が実数で表現されている
  - rPSCと組み合わせて1回のFFTで計算できる

$$\bar{a}_{l,m,n} = G_{l,m,n}^a + iE_{l,m,n}^a$$

$$\bar{b}_{l,m,n} = G_{l,m,n}^b + i\omega E_{l,m,n}^b$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \left( \bar{a}_{l,m,n} \times \bar{b}_{l+\alpha,m+\beta,n+\gamma} \right)$$

$$\text{Score}_{\alpha,\beta,\gamma} = \Re \left[ \bar{c}_{\alpha,\beta,\gamma} \right]$$

$$= \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \left( G_{l,m,n}^a G_{l+\alpha,m+\beta,n+\gamma}^b - \omega E_{l,m,n}^a E_{l+\alpha,m+\beta,n+\gamma}^b \right)$$

rPSC

Electrostatics(重み付き)

- 電荷はCHARMM19を用いる

# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめと今後の課題

# ZDOCK Benchmarkによる評価実験

## ■ ZDOCK Benchmark 2.0を利用

- 目的: 複合体構造の正確な予測(システムの性能評価)
- 対象: 23個の複合体(先行研究で評価対象としたサブセット)

### □ Docking Software

- MEGADOCK 1.0 (K-K)
- MEGADOCK 2.0 (rPSC)
- MEGADOCK 2.1 (rPSC+Elec)
- ZDOCK 2.3
- ZDOCK 3.0

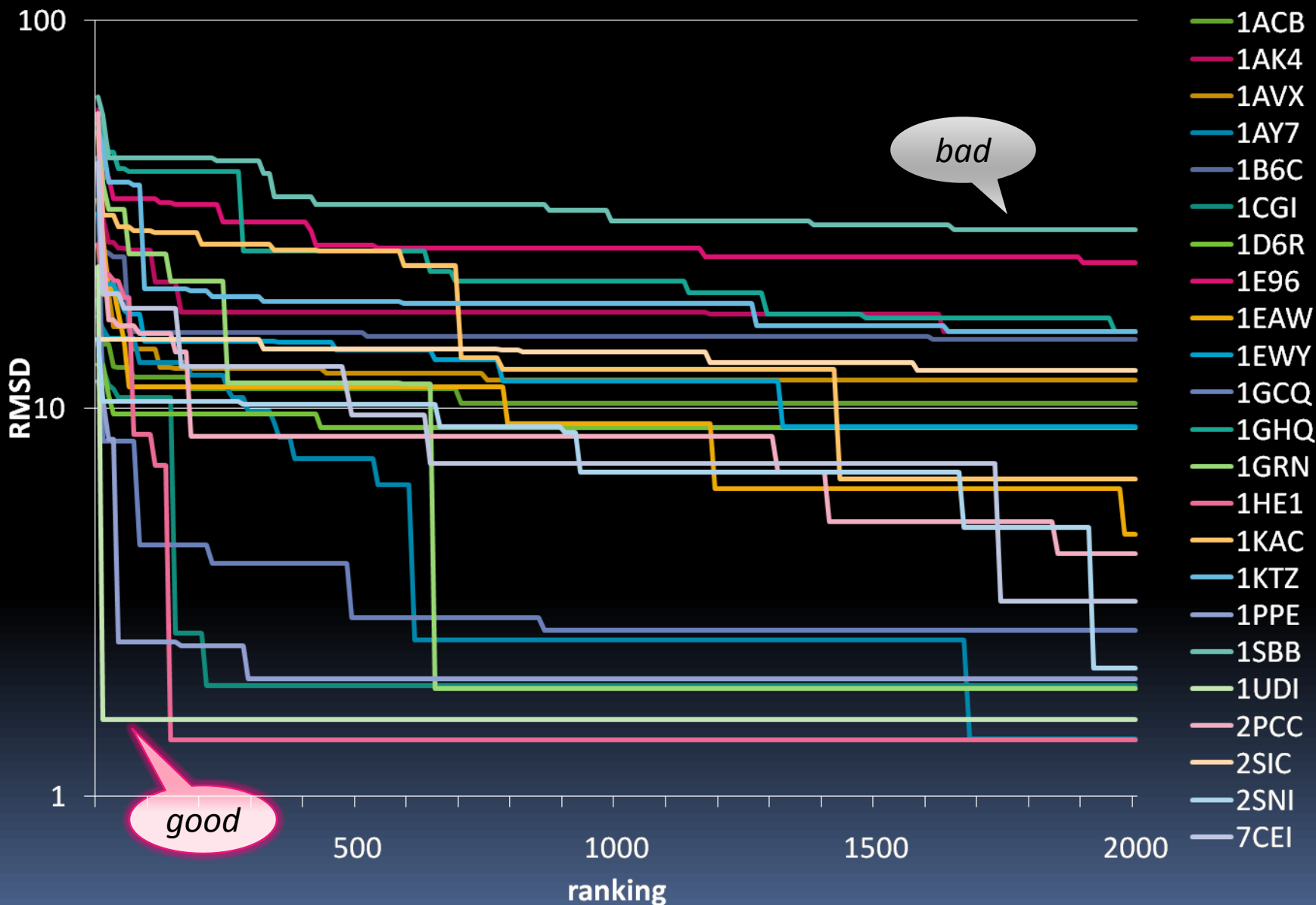
$$\rho = -3, \delta = 2$$

### □ 評価基準

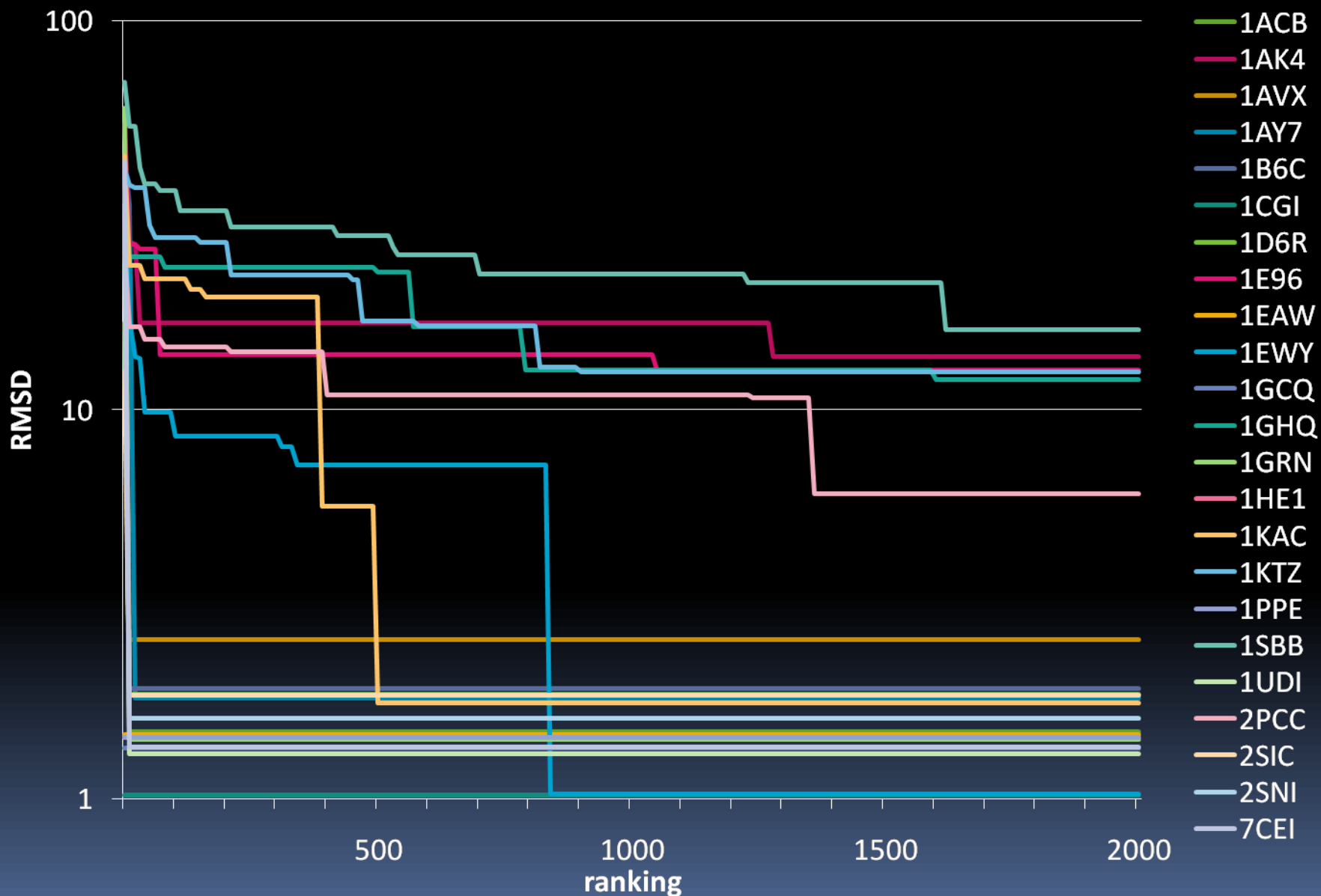
- 正解構造とのRMSD

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{r}_i^A - \mathbf{r}_i^B\|^2}$$

# RMSD – ranking グラフ (MEGADOCK 1.0, bound)

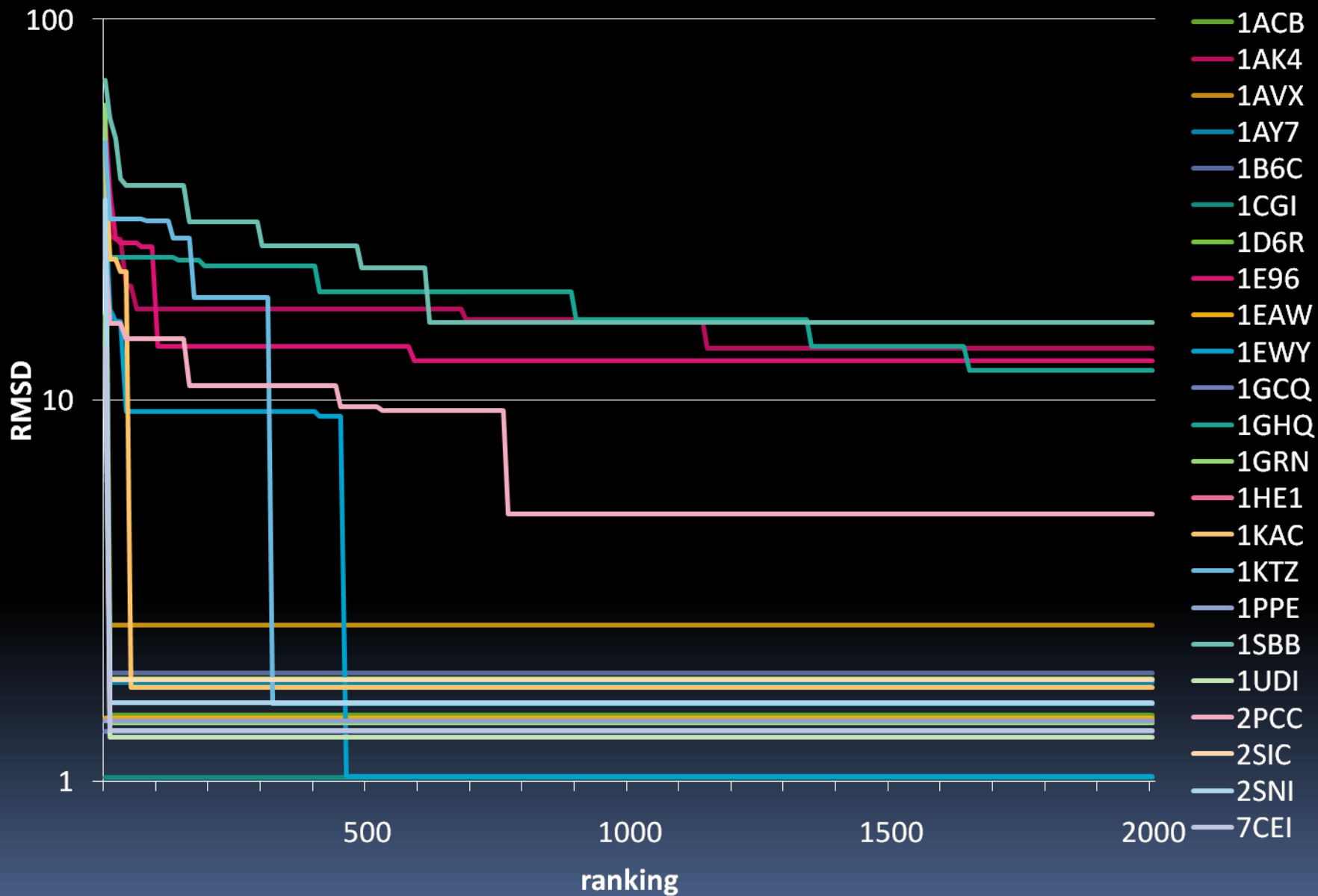


# RMSD – rankingグラフ (MEGADOCK 2.0, bound)

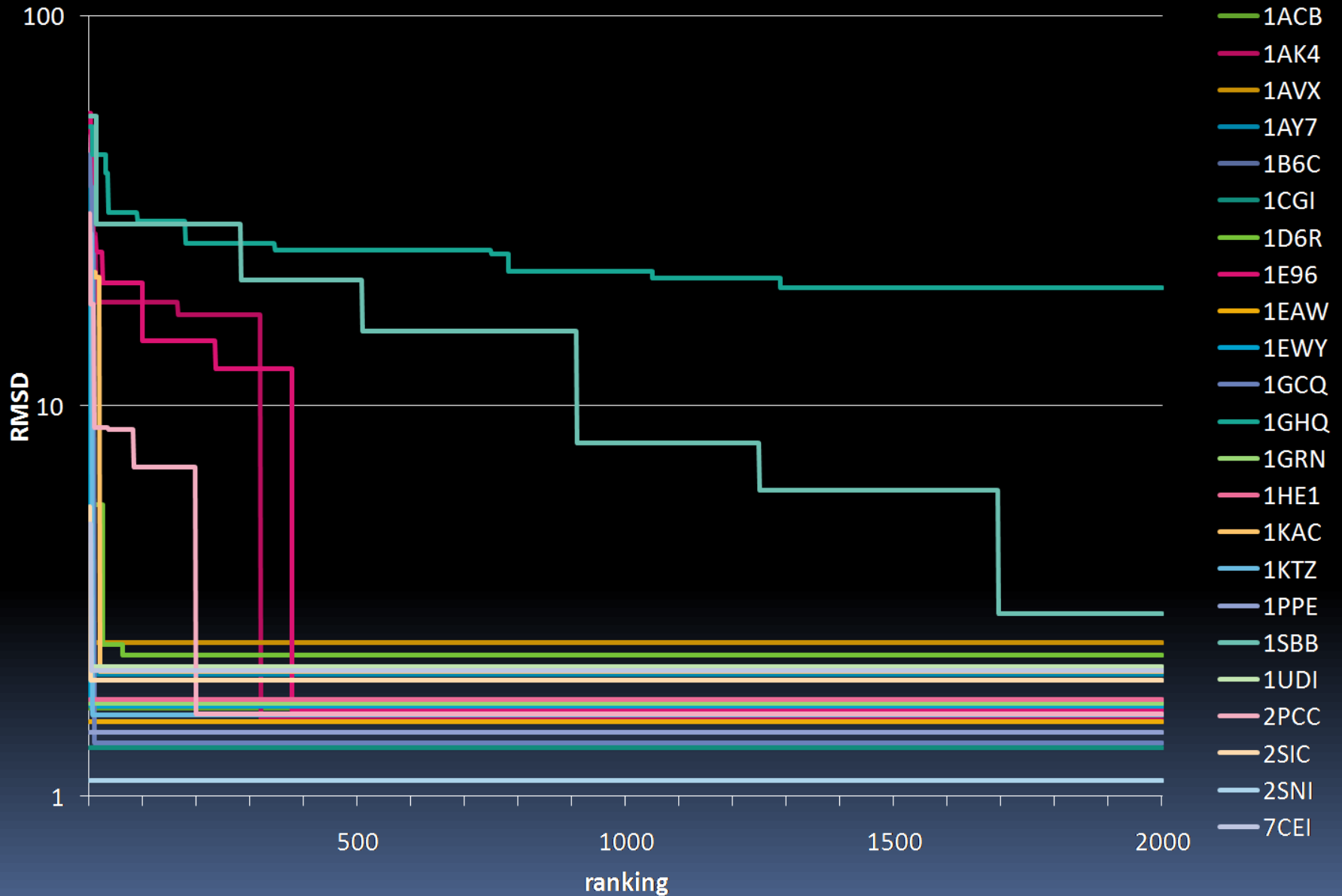




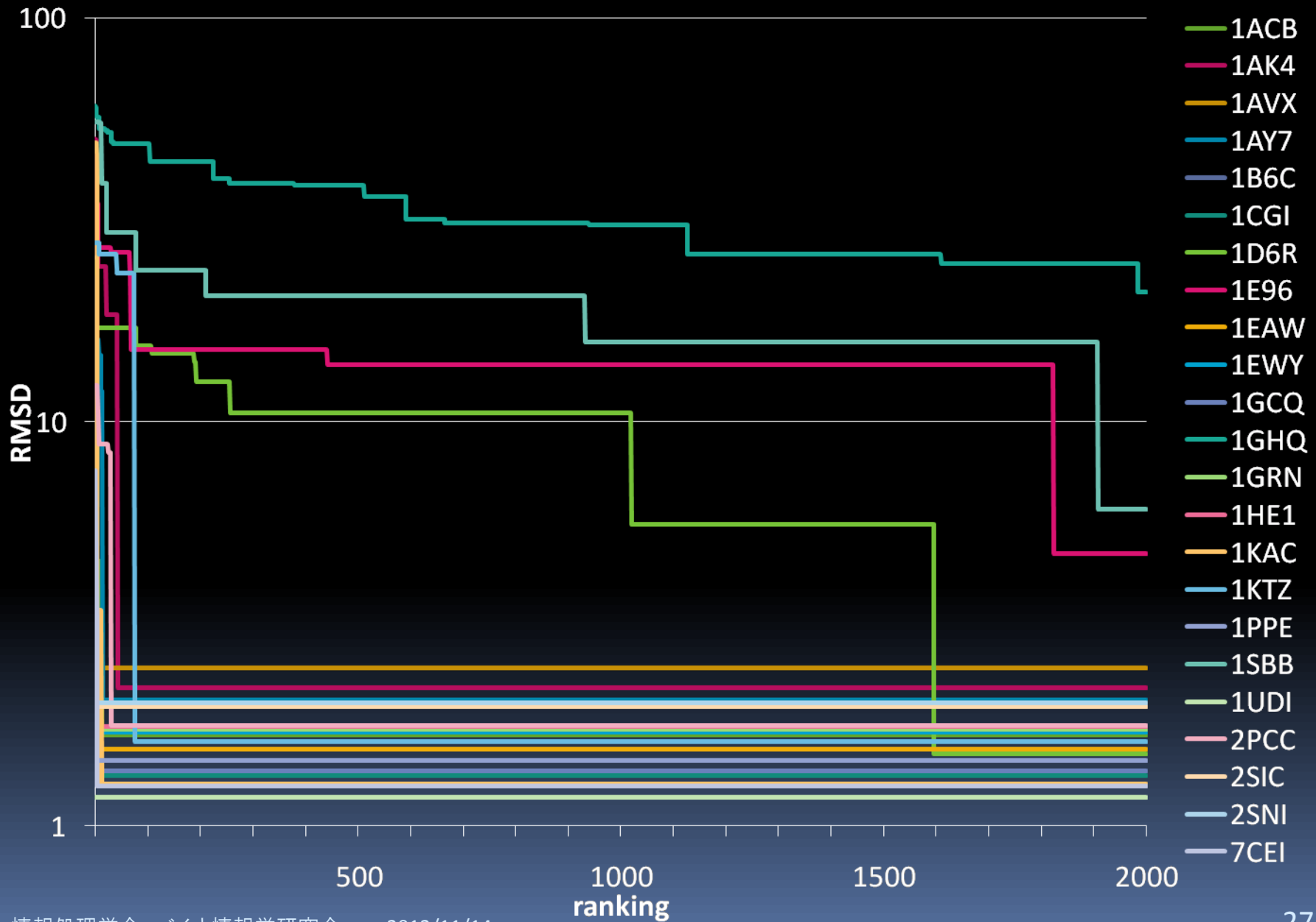
# RMSD – ranking グラフ (MEGADOCK 2.1, bound)



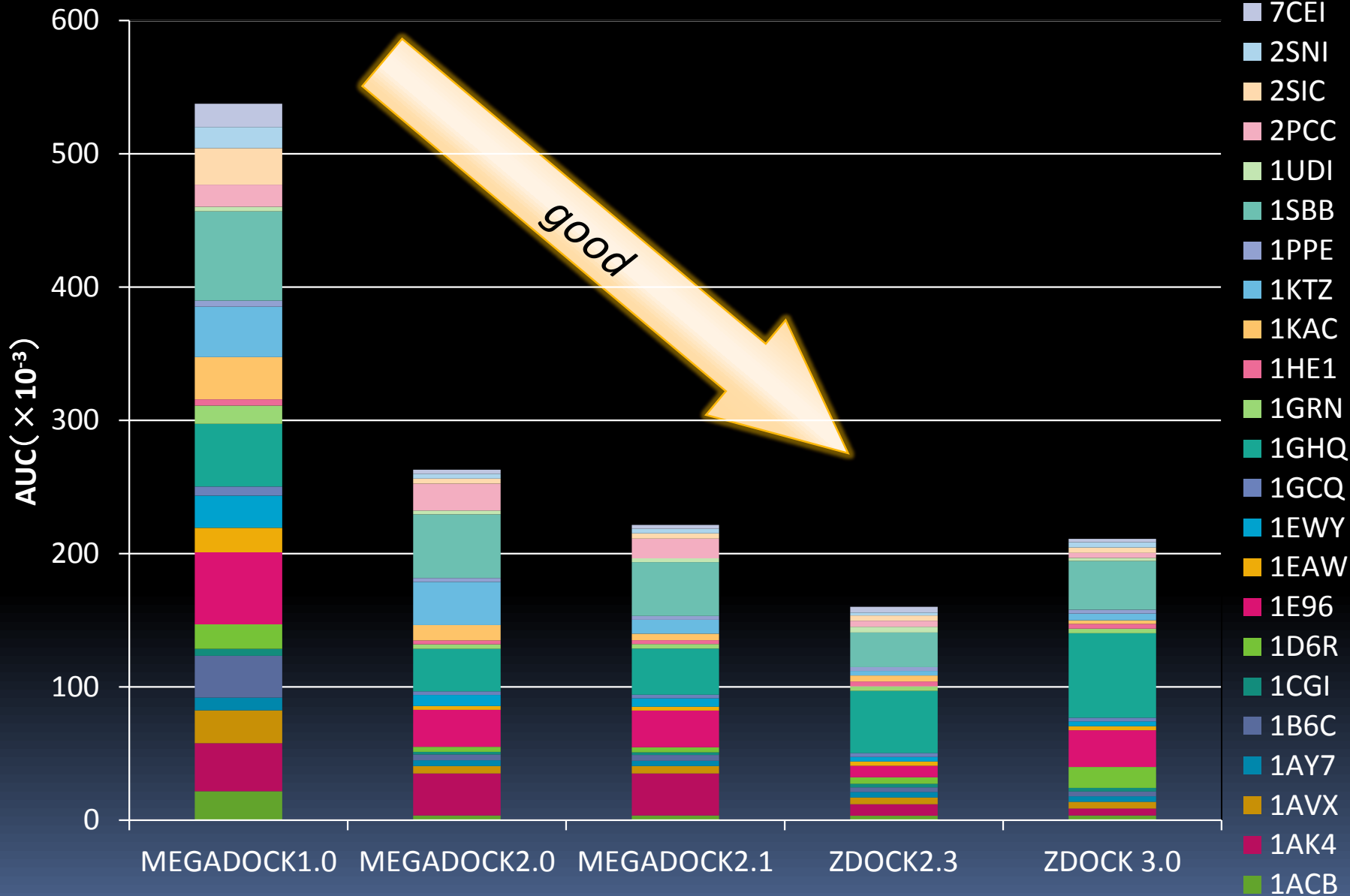
# RMSD – rankingグラフ (ZDOCK 2.3, bound)



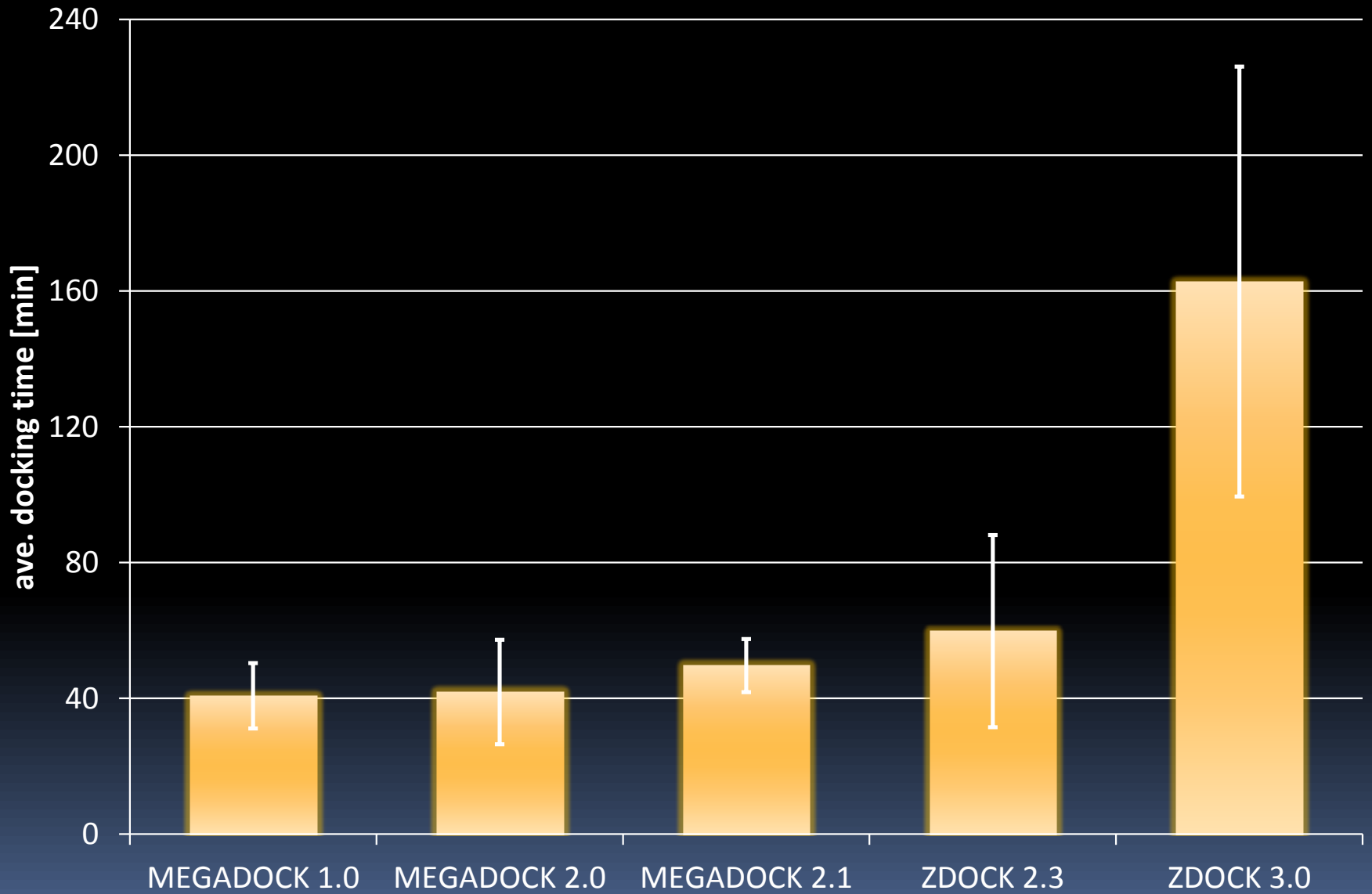
# RMSD – rankingグラフ (ZDOCK 3.0, bound)



# Area Under the Curve



# Docking time



# 1対1ドッキング評価の結論

- MEGADOCKの精度向上
  - Ver.1.0から2.0, 2.0から2.1での精度の改善が見られた
  - ZDOCKに迫る精度が得られた
- 計算時間増加の抑制に成功
  - Ver.2.0から2.1では10分程度の増加に抑えられた
  - ZDOCK3.0と比べると3倍高速

# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- まとめと今後の課題

# all-to-all Docking Prediction


- ZDOCK Benchmark 2.0
  - 目的: 相互作用相手の予測
  - 対象: 23個の複合体を総当たり
  - Docking Software
    - MEGADOCK 1.0 (K-K)
    - MEGADOCK 2.0 (rPSC)
    - MEGADOCK 2.1 (rPSC+Elec)
    - ZDOCK 2.3
    - ZDOCK 3.0
  - 評価基準
    - 相互作用“する”か“しない”かの判定



# 相互作用判定基準

- MEGADOCKの出力する1位の Score  $S_{ij}$  を  
表面積  $A_i, A_j$  で補正

$$\eta_{ij}^{(\log)} = \frac{\log S_{ij}}{\min\{A_i, A_j\}} \quad \eta_{ij}^{(\log \log)} = \frac{\log(\log S_{ij})}{\log(\min\{A_i, A_j\})} \quad \eta_{ij}^{(2/3)} = \frac{\sqrt[3]{S_{ij}^2}}{\min\{A_i, A_j\}}$$

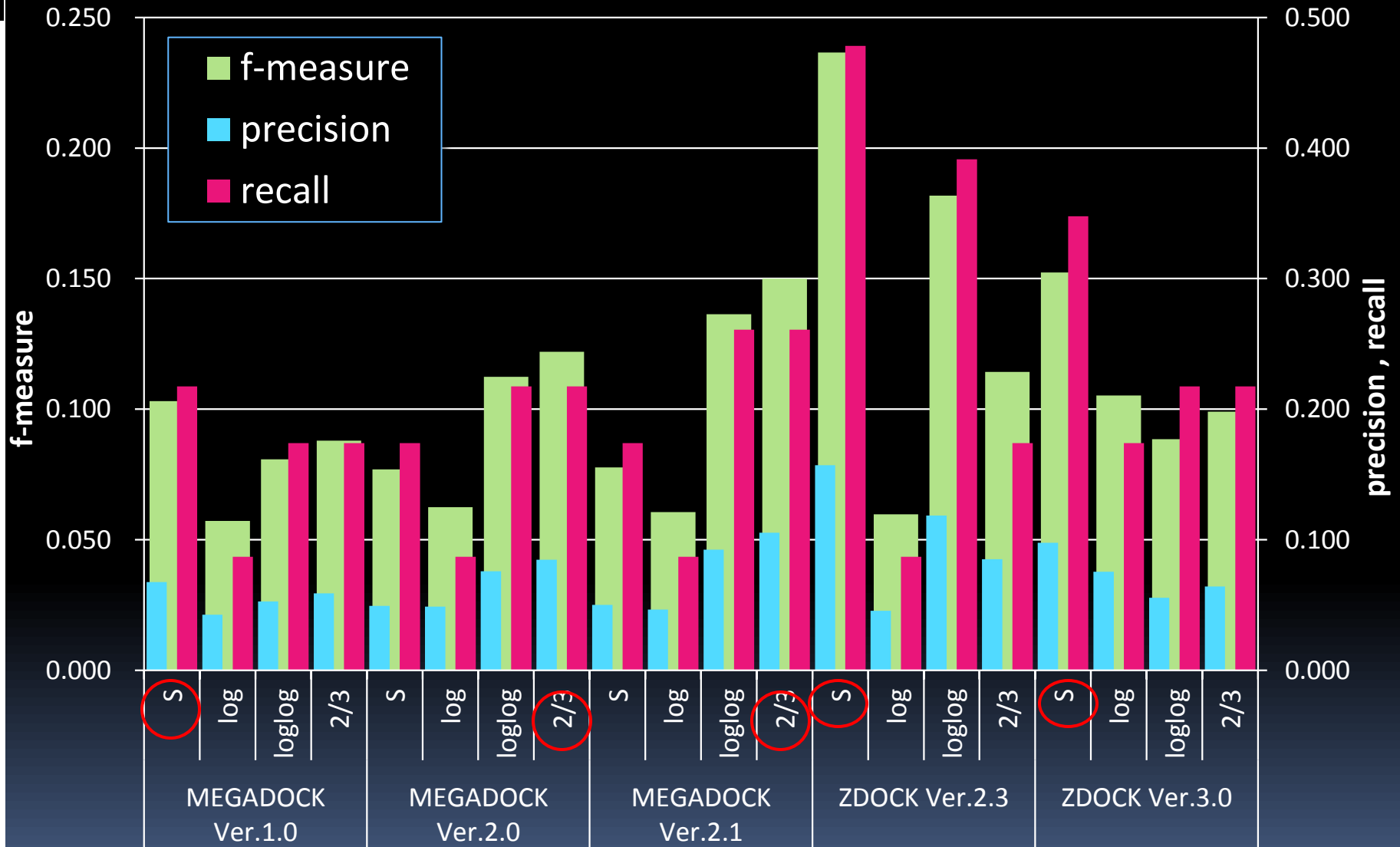

$$z_{ij}^{(n)} = \frac{\eta_{ij}^{(n)} - \mu^{(n)}}{\sigma^{(n)}}$$

z-scoreが1.0より大きければ「ペア( $i, j$ )は相互作用する」

正解ペアは $\{(i, j) \mid i=j\}$ とする。

相互作用判定でTP, FP, FN, TNを求め、f-measureによって性能評価

# all-to-all Docking Prediction



# all-to-allドッキング評価の結論

- f値による評価
  - MEGADOCKのバージョンアップによる精度向上が確認された
  - スコアを補正することで精度向上
    - MEGADOCKのスコアはタンパク質の大きさに依存している
  - 結果はZDOCKの方がよかった→改良の余地あり

# 目次

- 本研究の説明
- タンパク質ドッキングの従来研究
- 網羅的PPI予測とMEGADOCK
- 新形状相補性スコアrPSCの提案
- 静電的相互作用の導入
- 1対1ドッキング予測の性能評価
- 網羅的PPI予測の性能評価
- **まとめと今後の課題**

# まとめ

- MEGADOCKにrPSCと静電的相互作用を導入
  - 計算時間の増加を抑えつつ性能向上
    - ZDOCK 3.0の精度に近づいた
    - ZDOCK 3.0の3.3倍高速化
  - 網羅的PPI予測への適用
    - バージョンアップによる精度向上の確認
    - スコア補正による精度の向上

# 今後の課題

- rPSCの最適パラメータの推定
  - $\delta, \rho$  の精密な推定
- 相互作用判定の改良
  - クラスタリングの利用
    - 我々の先行研究で網羅的PPI予測性能の向上例あり
  - 機械学習の利用
    - ドッキングスコアから機械学習によるルール獲得
    - タンパク質の表面積や体積の情報も加味
- 生物系への適用
  - 大規模な系の相互作用ネットワーク推定へ