

簡易疎水性相互作用モデルによる タンパク質間ドッキング予測の高精度化

大上 雅史^{† ‡} 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

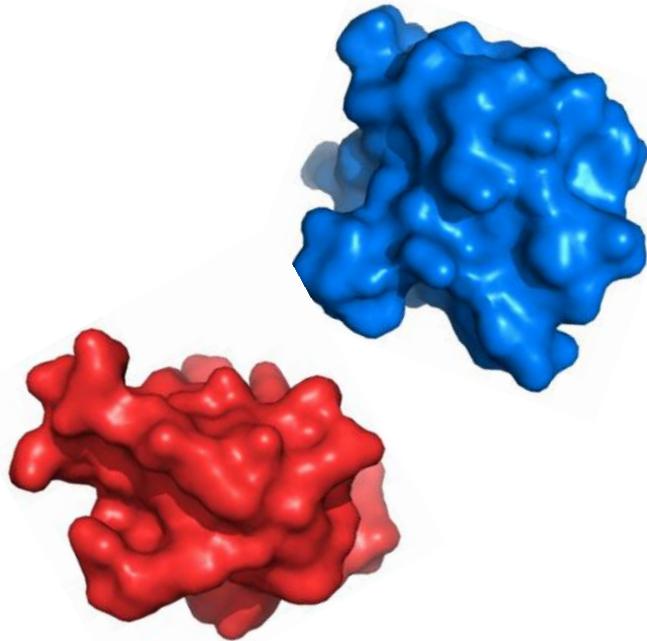
‡ 日本学術振興会 特別研究員

本研究の概要

- **ドッキングソフトウェアMEGADOCKの改良**
 - 従来は形状相補性と静電相互作用の2項で計算
- **新たに疎水性相互作用を考慮**
- **計算時間を増やさずに精度の向上に成功**

タンパク質間ドッキング予測

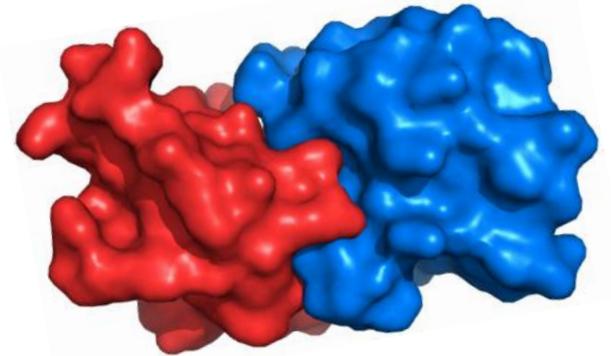
タンパク質ペア



ドッキング計算



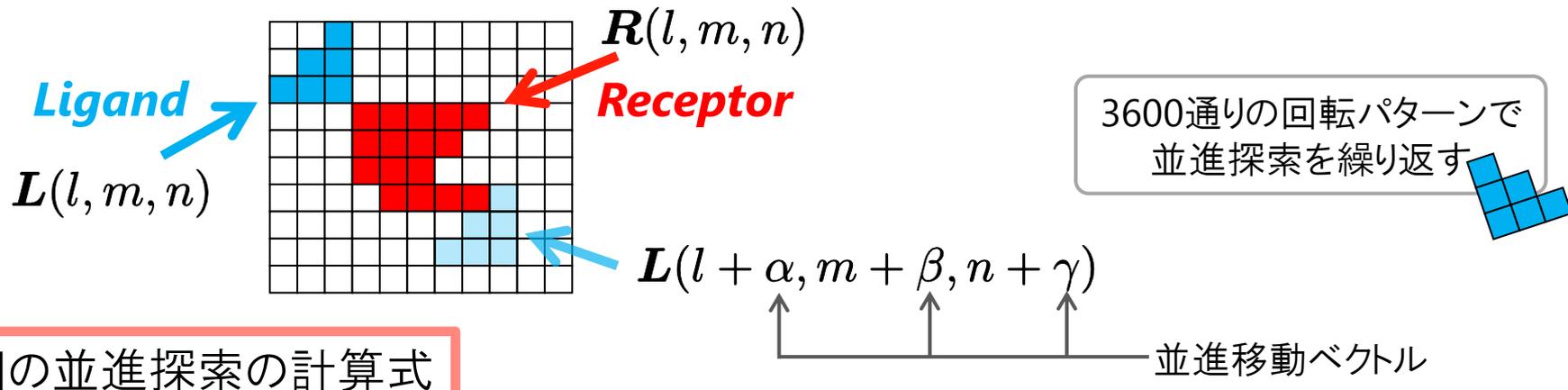
タンパク質複合体
(候補構造)



ドッキング計算

- **Katchalski-Katzirアルゴリズム** Katchalski-Katzir E, et al. PNAS, 1992.

– 3-DグリッドとFFTによるタンパク質間ドッキング予測



1回の並進探索の計算式

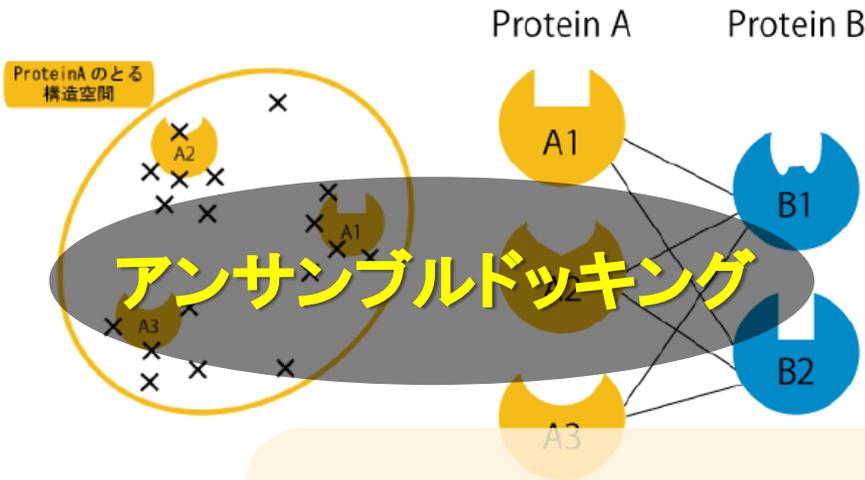
$$S(\alpha, \beta, \gamma) = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N R(l, m, n) L(l + \alpha, m + \beta, n + \gamma)$$

畳み込み和(相関関数)

$$S(\alpha, \beta, \gamma) = \text{IFFT} [\text{FFT}[R(l, m, n)] * \text{FFT}[L(l, m, n)]]$$

FFT-based Rigid Docking

ドッキング計算の応用例

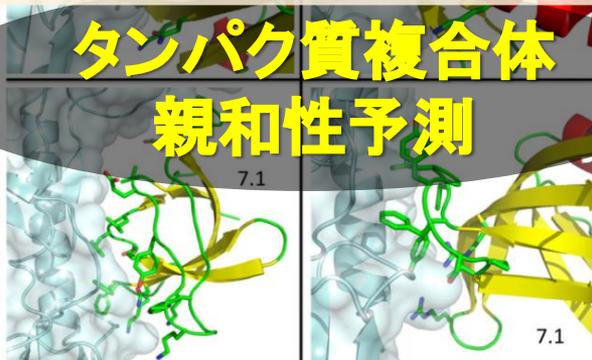


Venkatraman V, et al. *Proteins*, 2012.

タンパク質間相互作用
ドッキング計算
高速な計算が要求される

High average degree
Low average degree

タンパク質複合体
親和性予測



Fleishman S, et al. *JMB*, 2011.



Matsuzaki Y, et al. *JBCB*, 2009.

MEGADOCK

- **MEGADOCK**の評価関数(ドッキングスコア)

Ohue M, et al. *IP SJ TOM*, 2010.

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

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

1回の畳み込み演算

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

rPSC

ELEC

- 1回の畳み込み演算で2つの効果を計算
- 既存ソフトウェア(ZDOCK3.0*)より高速

*Mintseris J, et al. *Proteins*, 2007.

疎水性相互作用の検討

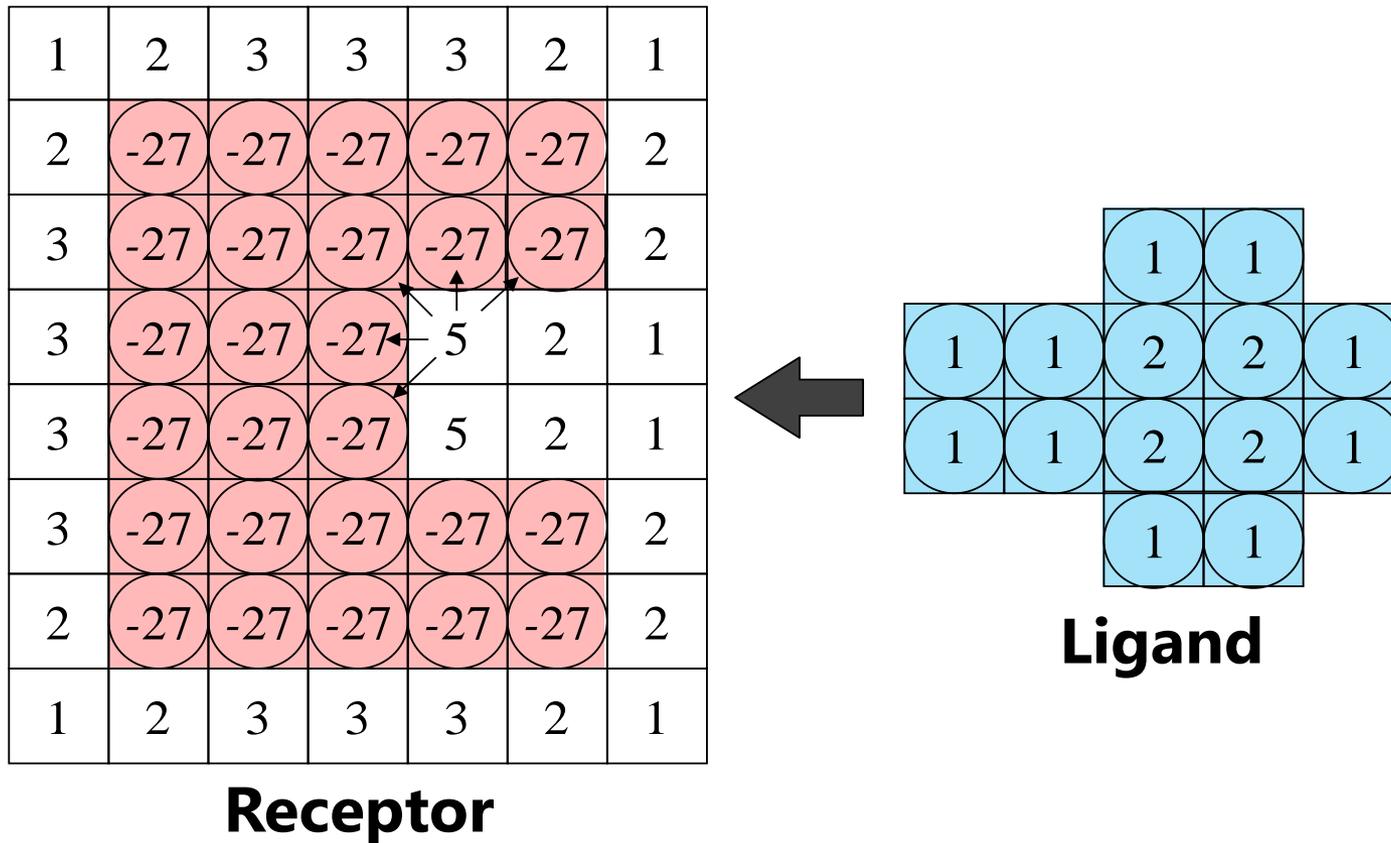
- MEGADOCKのスコア関数
 - 形状相補性(ファンデルワールス力)
 - 静電的相互作用(クーロン力)
- **第3の効果として疎水性相互作用を導入する**
 - タンパク質間相互作用の駆動力の1つ
 - 形状相補性と静電と疎水性を考慮したソフトウェア
 - ZDOCK 2.3, ZDOCK 3.0など
 - 複数回の畳み込み和を使っている
 - 例: ZDOCK 2.3・・・2回 Chen R, et al. Proteins, 2003.
 - ZDOCK 3.0・・・8回 Mintseris J, et al. Proteins, 2007.
- **目的: 計算時間を増加させずに疎水項を入れる**

疎水性相互作用の検討

- 疎水性相互作用を含むポテンシャル
 - 本研究では**Atomic Contact Energy**スコアを用いる
- **Atomic Contact Energy (ACE)** Zhang C, et al. *J Mol Biol*, 1997.
 - 脱水と自由エネルギーを推定する目的で作られた経験的原子ペアワイズポテンシャル
 - ZDOCK 2.3, FireDock*などで採用 * Andrusier N, et al. *Proteins*, 2007.
- どのように組み込むか
 - rPSCのレセプター側に平均化した値(non-pairwise型ACE)を組み込む

rPSCスコア

- **rPSC (real Pairwise Shape Complementarity)** Ohue M, et al. *IP SJ TOM*, 2010.
 - 実数値のみで表される形状相補性スコア



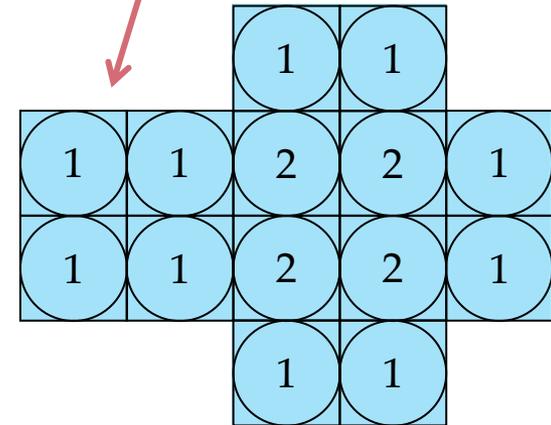
(提案手法)レセプター側のACE導入

- レセプターの疎水/親水面を考慮

1+H	2+H	3+H	3+H	3+H	2+H	1+H
2+H	-27	-27	-27	-27	-27	2+H
3+H	-27	-27	-27	-27	-27	2+H
3+H	-27	-27	-27	5+H	2+H	1+H
3+H	-27	-27	-27	5+H	2+H	1+H
3+H	-27	-27	-27	-27	-27	2+H
2+H	-27	-27	-27	-27	-27	2+H
1+H	2+H	3+H	3+H	3+H	2+H	1+H

Receptor

表面の1を利用



Ligand

$$H(l, m, n) = \begin{cases} \text{周囲のACE値の和 (外部)} \\ 0 \text{ (内部)} \end{cases}$$

*図では省略してあるが、重み w_h を乗じる

提案手法

- 新しいドッキングスコア

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

$$\mathbf{R}(l, m, n) = G_R(l, m, n) + w_h H_R(l, m, n) + i E_R(l, m, n)$$

$$\mathbf{L}(l, m, n) = G_L(l, m, n) + i w_e E_L(l, m, n)$$

- 1回の畳み込み演算で3つの効果を計算

評価実験

- **Protein-Protein Docking Benchmark 4.0を利用**

- 176個の複合体構造

Hwang H, *et al. Proteins*, 2010.

- bound set と unbound set で評価

- **bound** set: 複合体状態のPDBを取ってきて切り離したもの

- **unbound** set: それぞれ単体でのPDB

- **当たり(near-native構造)の基準**

- (正解構造との)レセプターフィット時のLigand-RMSDが5Å以下

- **比較する手法**

- 現MEGADOCK (ver. 2.5) ($w_h = 0, w_e = 1$ に相当)

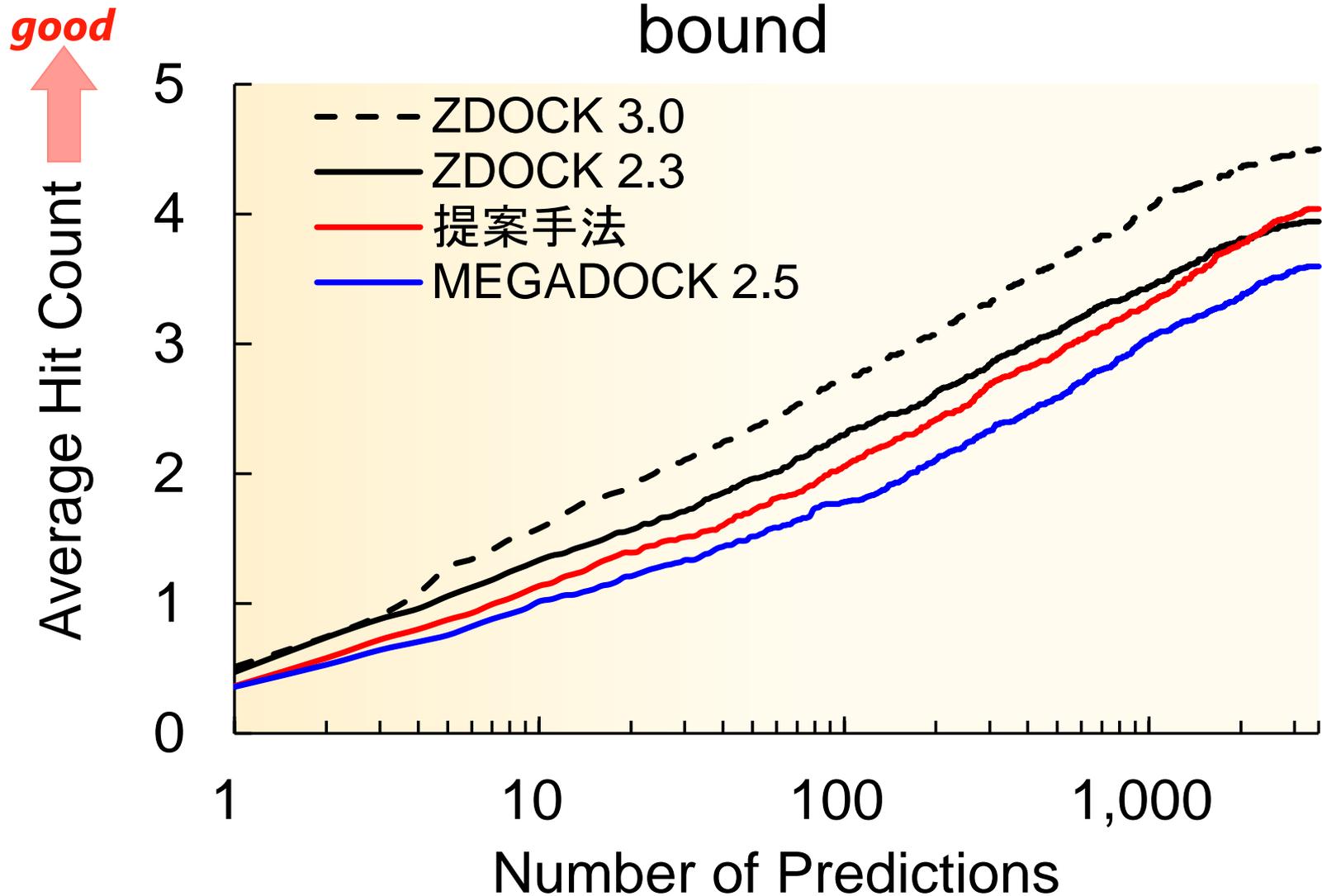
- 提案手法 ($w_h = 0.6, w_e = 1.15$)

- 重みパラメータは**bound** setを用いて最適化した

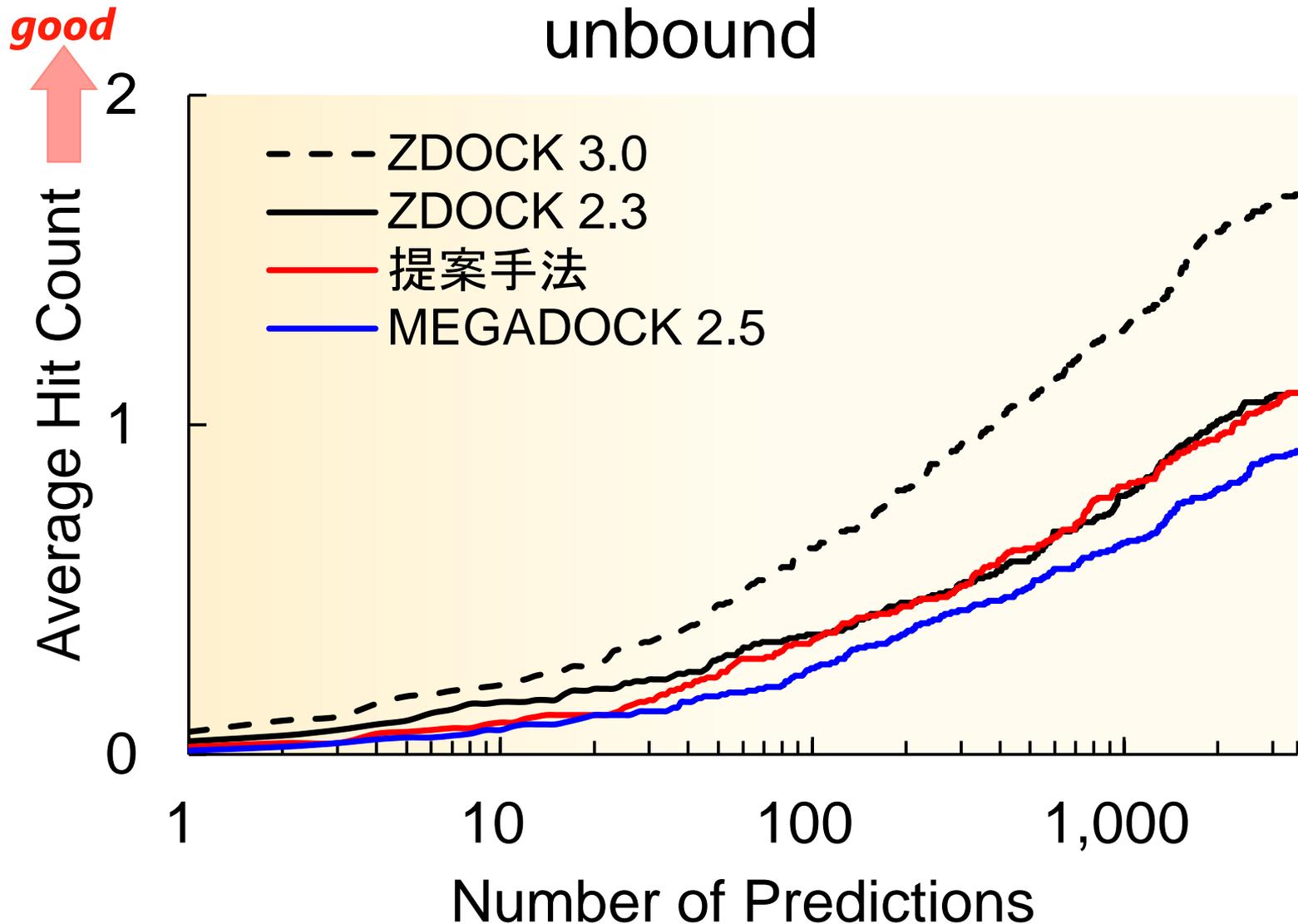
- ZDOCK 2.3

- ZDOCK 3.0

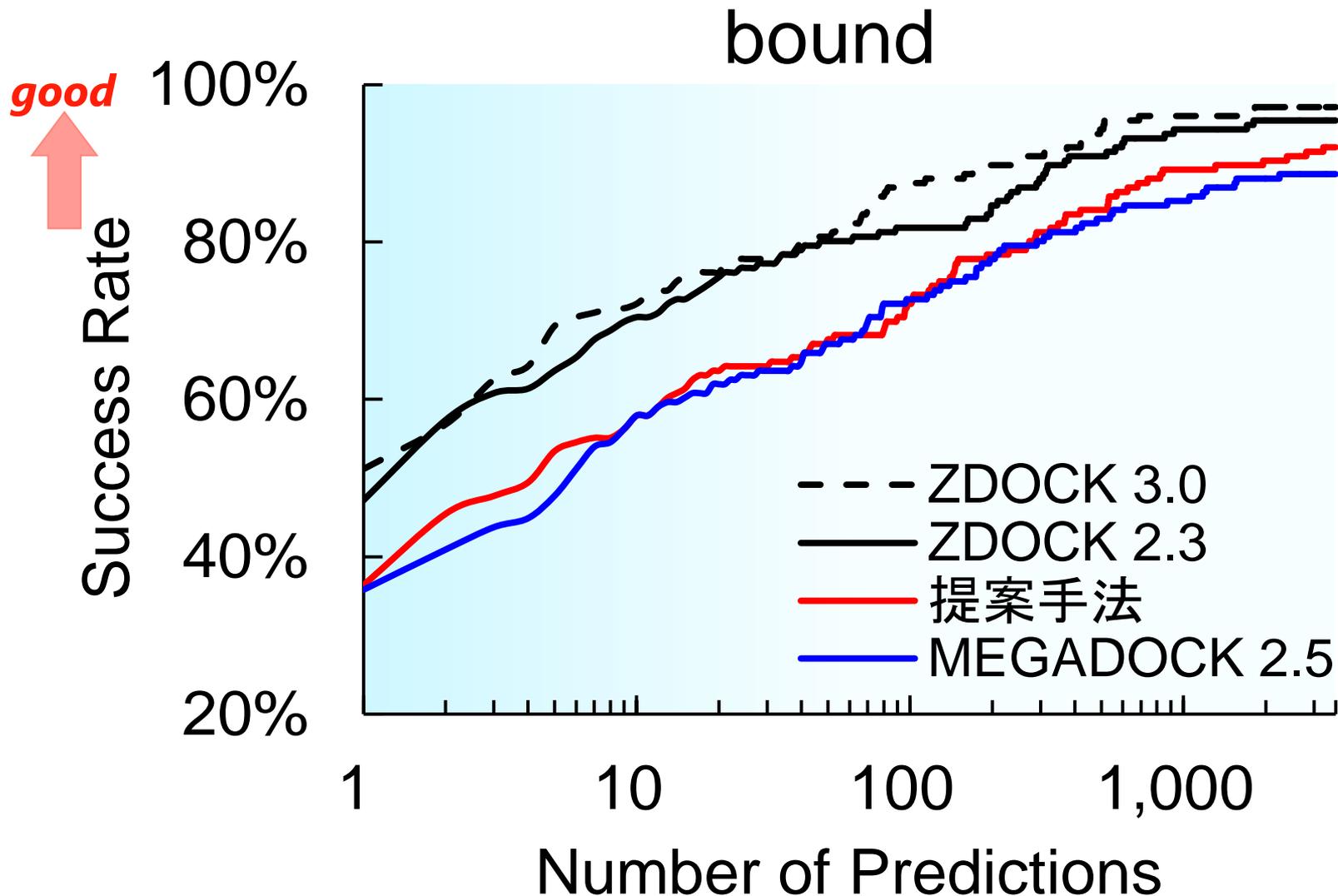
上位から取ってきた中に平均何個当たりがいるのか



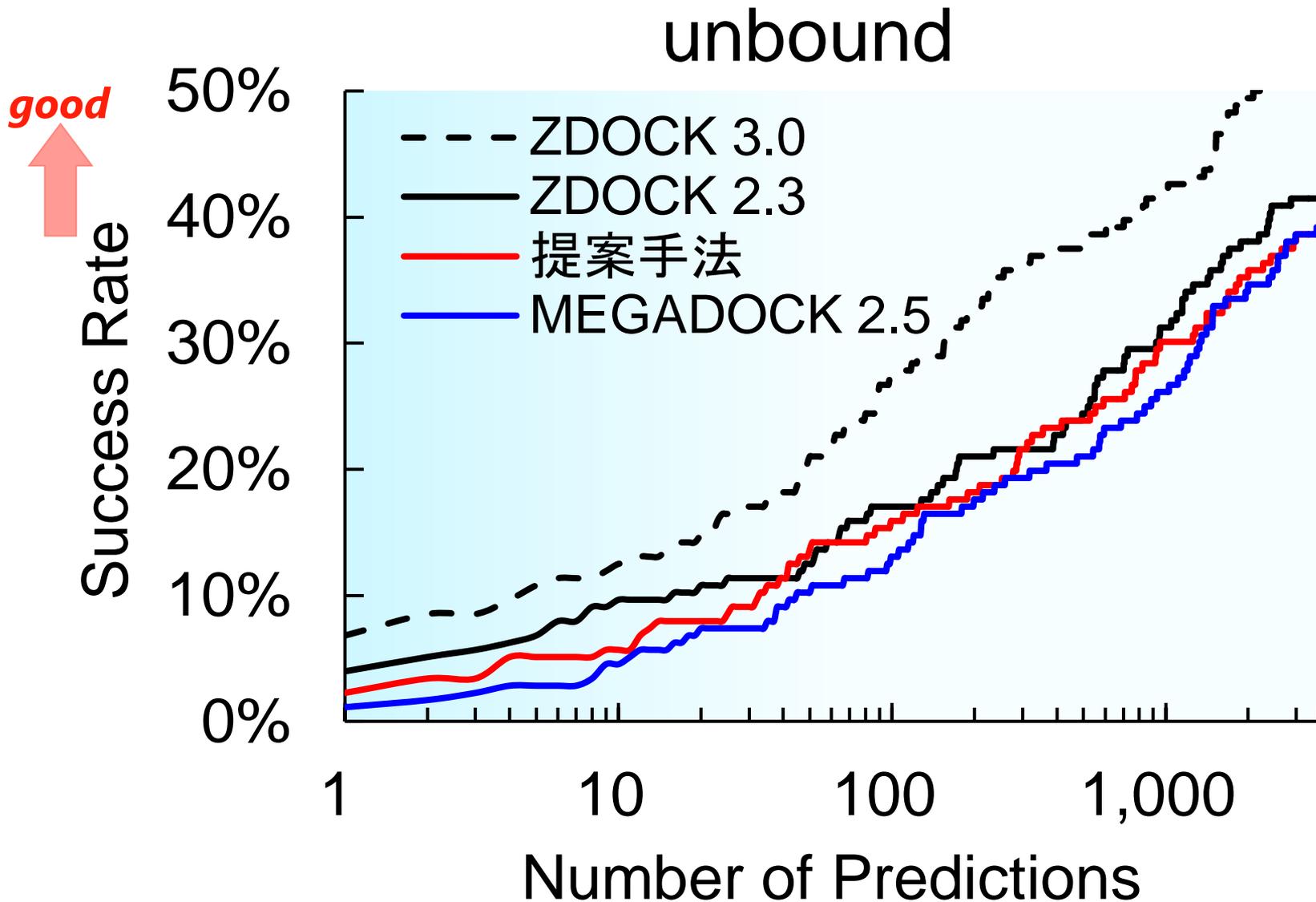
上位から取ってきた中に平均何個当たりがいろいろなのか



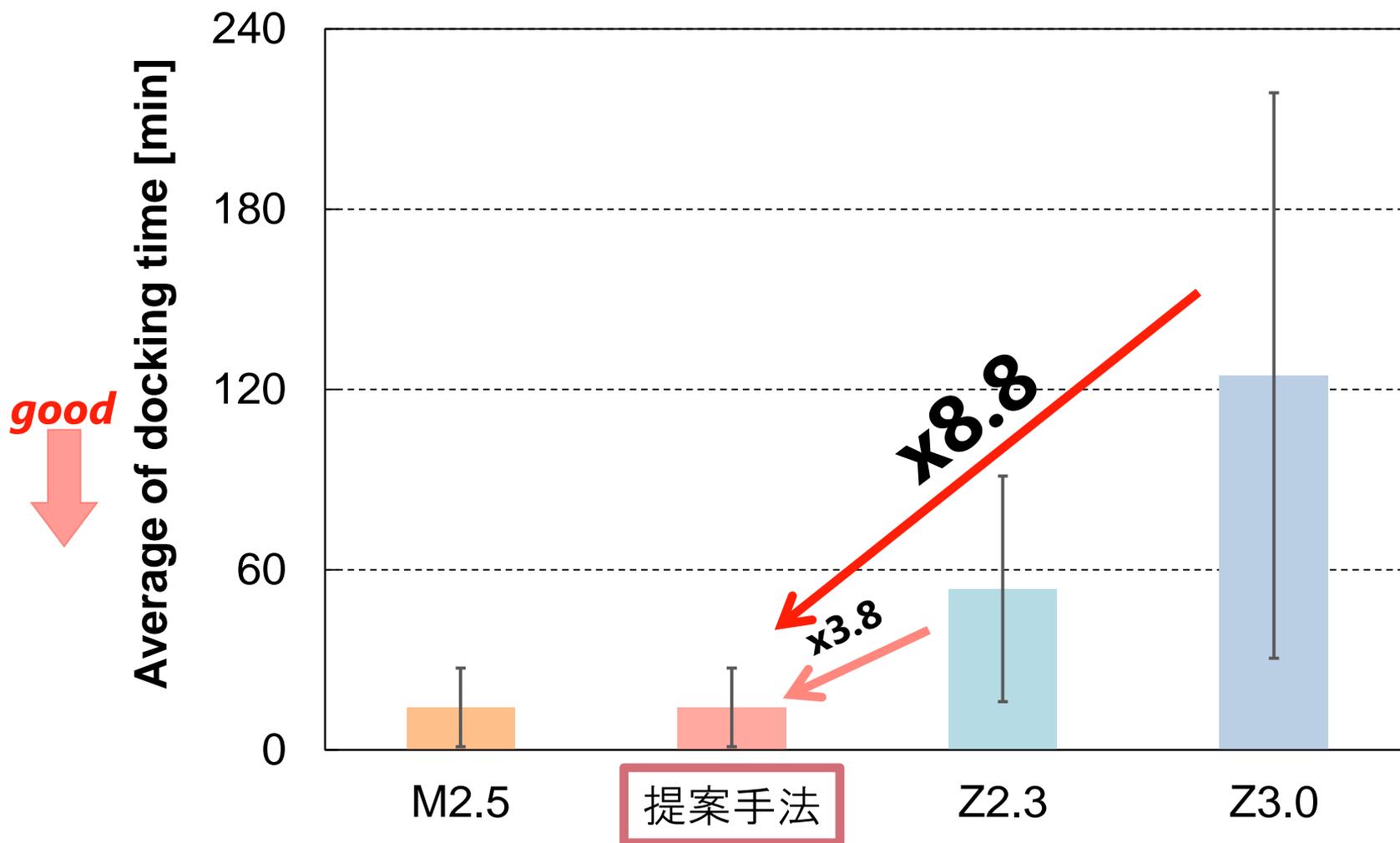
上から何個取れば当たりがいくのか



上から何個取れば当たりがいくのか



1ペアの平均計算時間(TSUBAME2.0 1core)



結果の例

PDB ID: 1BVN

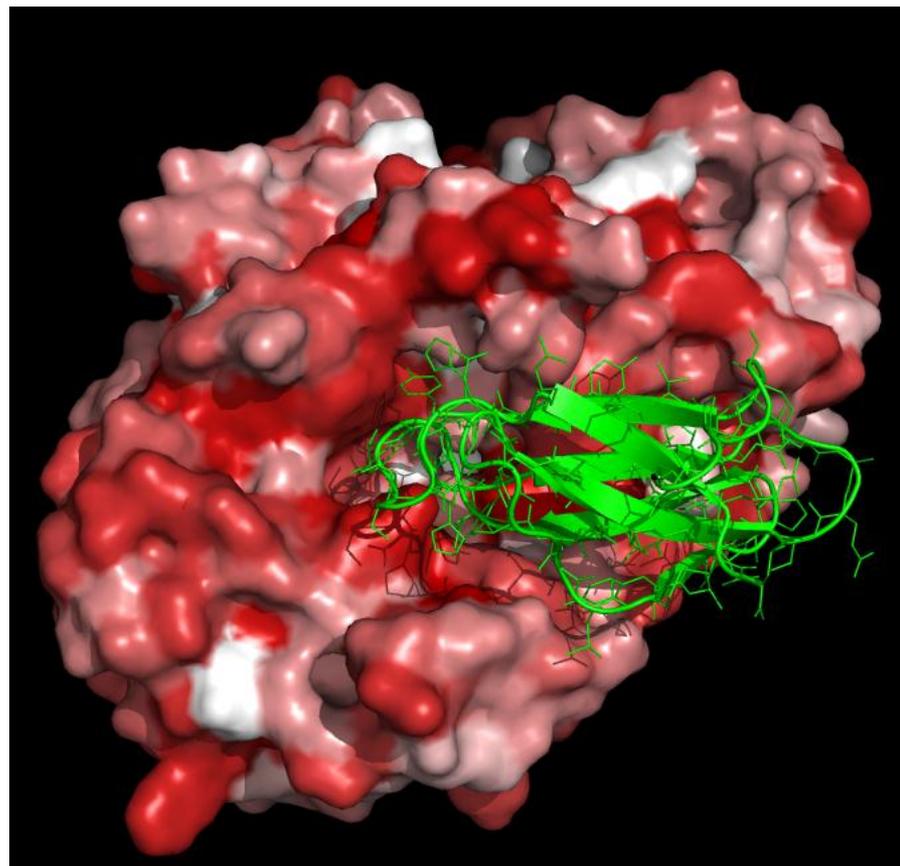
(unbound Rank:

11_(M2.5) → **1**_(+ACE))

Receptor: α -amylase

Ligand: Tendamistat

(α -amylase inhibitor)



Hydrophobicity scale*

親水性  疎水性

おわりに

• 本研究の成果

- レセプターのみを考慮した簡易型疎水項を提案
 - 畳み込み和の数を増やさずに既存モデルの拡張として導入
- 計算時間の増加を伴わずに精度の向上を確認
 - ZDOCK 2.3の3.8倍, ZDOCK 3.0の8.8倍の計算速度向上を達成
 - ZDOCK 2.3に近い精度の達成

• 今後の課題

- レセプター・リガンド両方ともに疎水項が入ったモデルの検討
 - 計算時間の増加を避けつつ更なる精度の向上を目指す
- パラメータ動的調整の検討
 - 2012-BIO-29(19) との連携
- タンパク質間相互作用ネットワーク予測への応用

謝辞

- **本研究は以下の支援を受けて行われました。**
 - 科学研究費補助金(特別研究員奨励費23・8750)
 - 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」
 - 日本学術振興会 博士課程教育リーディングプログラム
「情報生命博士教育院」