

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

大上 雅史¹

松崎 由理¹

松崎 裕介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

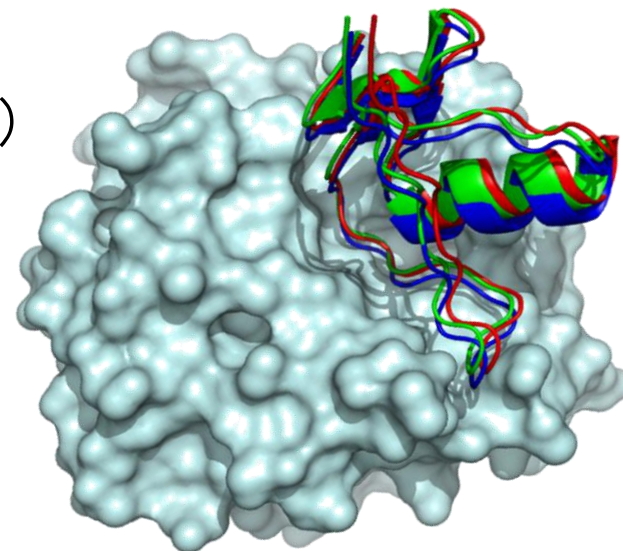
目次

- 研究背景
- MEGADOCKシステムの構成
 - ドッキング計算システム部の説明
 - PPI予測システム部の説明
- 計算機実験
 - ドッキング予測性能評価
 - 網羅的PPI予測性能評価
- システム生物学への応用

研究背景

- **タンパク質間相互作用 (PPI)**

- 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
- **構造情報を用いた予測**
 - 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

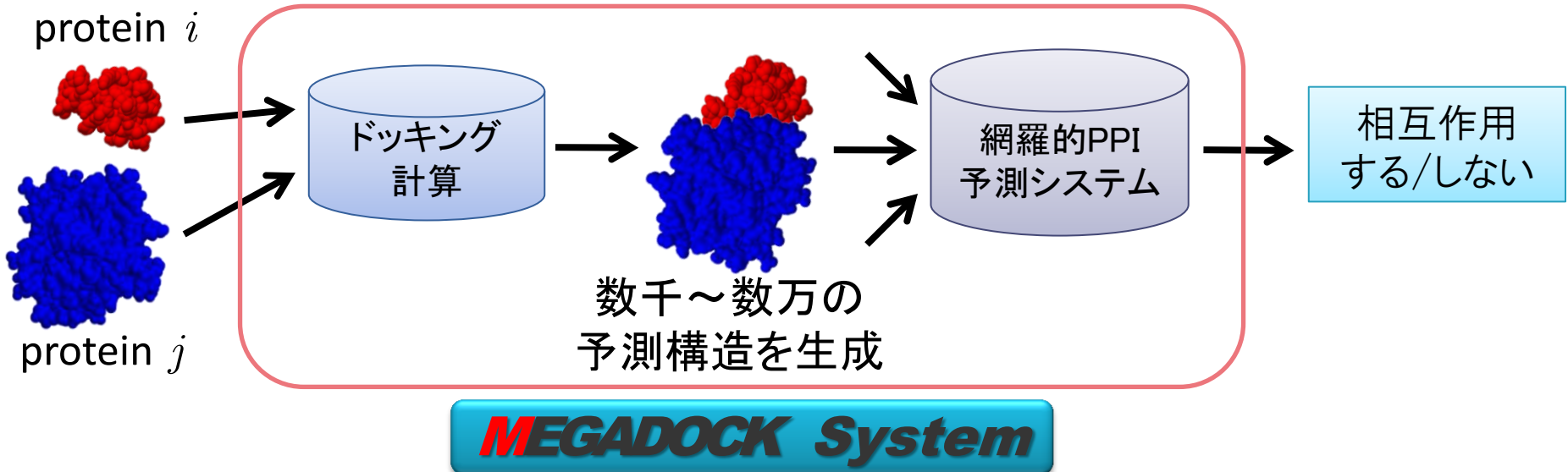
- **目的: タンパク質ドッキングに基づくPPIネットワーク予測**

- 増えつつある立体構造データの活用
- 1000 × 1000規模の網羅的計算への挑戦
- システム生物学への応用

目次

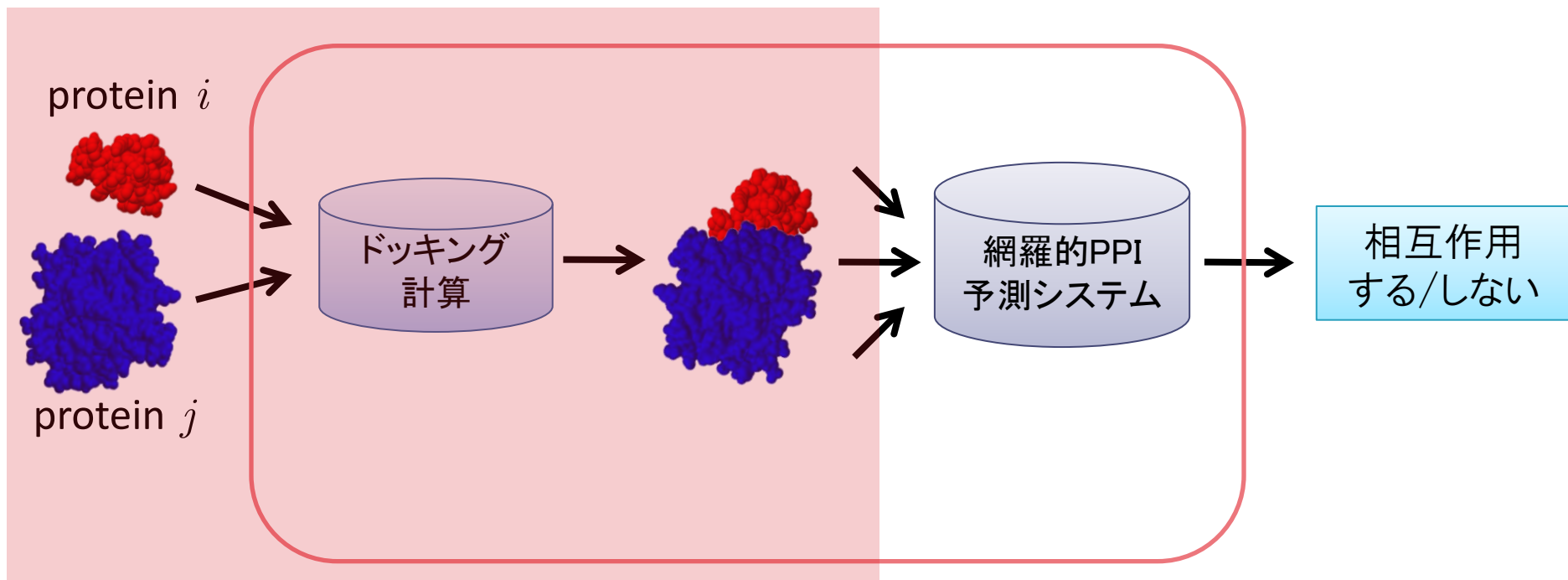
- 研究背景
- MEGADOCKシステムの構成
 - ドッキング計算システム部の説明
 - PPI予測システム部の説明
- 計算機実験
 - ドッキング予測性能評価
 - 網羅的PPI予測性能評価
- システム生物学への応用

システム構成



ドッキング計算システム部

ドッキング計算システム部

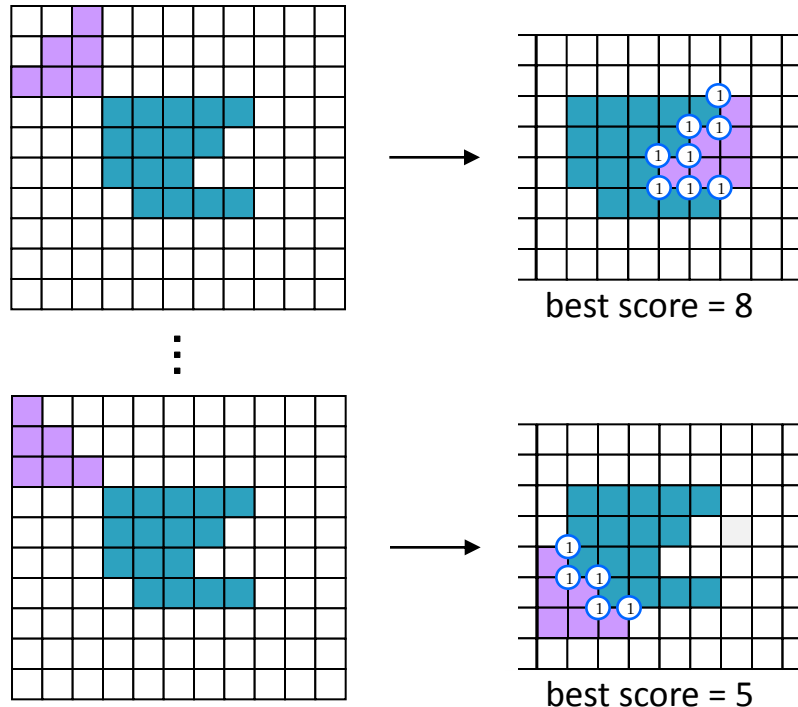


MEGADOCK System

ドッキング計算

- タンパク質ドッキング
 - タンパク質立体構造の形状相補性に基づいて評価
 - タンパク質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

15度刻み
3600通りの探索



[Example of translational search in 2D](#)

ドッキング計算の既存手法と問題点

- ZDOCK
 - 代表的なドッキングソフトウェア
 - ZDOCK 3.0 (J Mintseris, et al, *Proteins*, 69, pp.511-520, 2007)
 - ドッキング精度向上のために様々な効果を考慮
 - 形状相補性のスコア: Pairwise Shape Complementarity (PSC)
 - 複素数での表現
 - 物理化学的相互作用: 疎水効果, 静電的相互作用など
- 問題点
 - 計算時間が長い
 - 畳み込み計算に用いるFFTの回数が多く大規模な計算には向かない
 - 候補構造を絞り込みすぎる
 - 各角度ごとに最大のスコアを持つ構造を1つしか出力できない

提案手法:rPSC

- real Pairwise Shape Complementarity

- 形状相補性のスコアを**実数のみ**で表現

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

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

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

Receptor rPSC

		1	1		
1	1	2	2	1	
1	1	2	2	1	
		1	1		

Ligand rPSC

MEGADOCK:ドッキング計算システム部

- 実数表現のメリット

- 虚数項に空きができる
- 他の相互作用を追加してもFFTの回数が増加しない

- 評価関数 (ドッキングスコア)

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

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

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

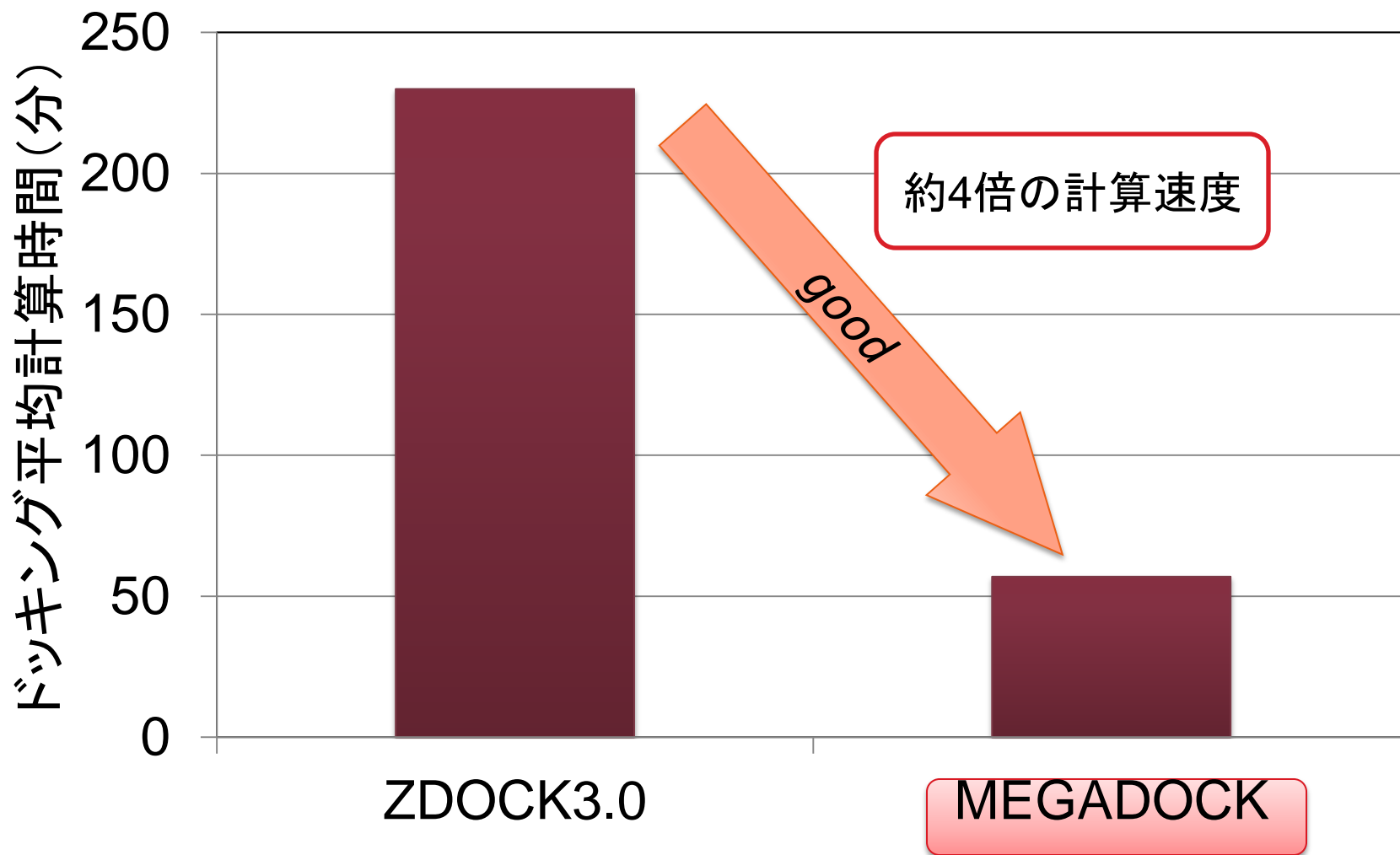
- 形状相補性: rPSC G
- 物理化学的相互作用: 静電的相互作用 E

- 次式によって計算可能

$$S(\alpha, \beta, \gamma) = \text{IFT} [\text{DFT}[\mathbf{R}(l, m, n)]^* \text{DFT}[\mathbf{L}(l, m, n)]]$$

計算時間

- 44のタンパク質ペアに対して計算したときの平均計算時間

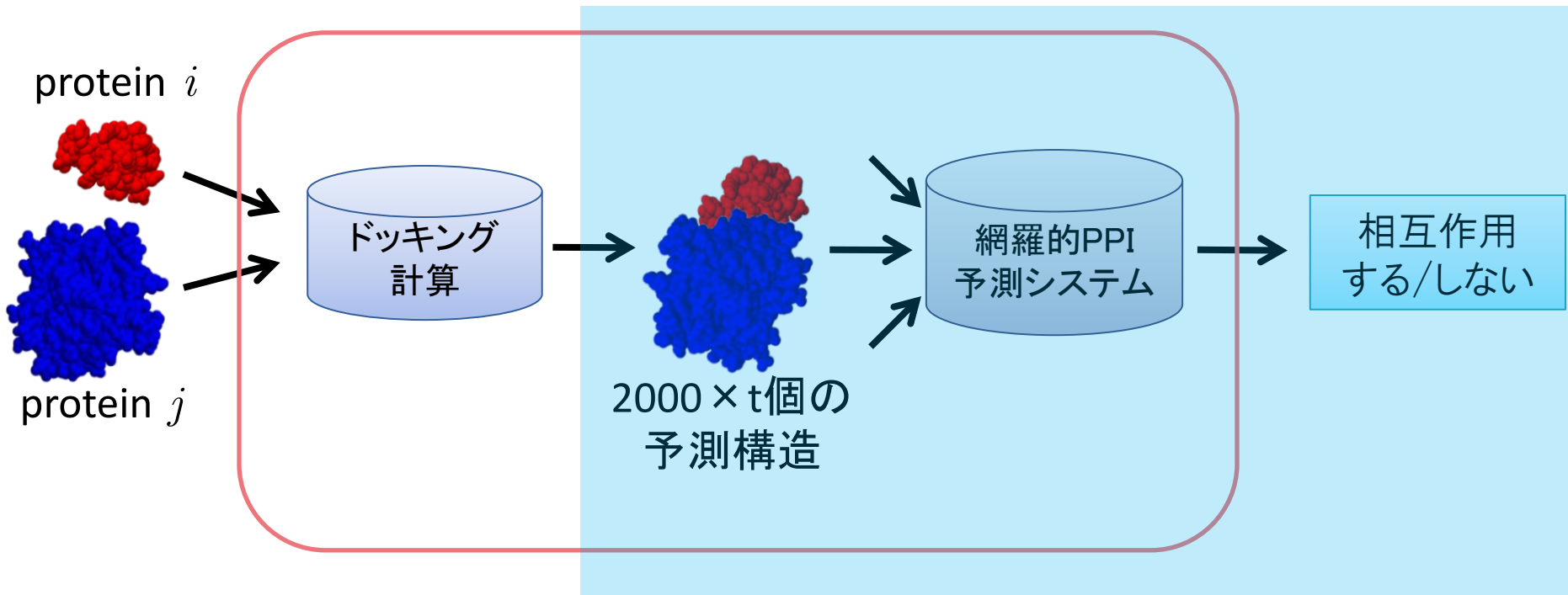


MEGADOCK:ドッキング計算システム部

	ZDOCK	MEGADOCK
形状相補性スコア	PSC(複素数)	rPSC(実数のみ)
角度ごとの出力数	1個	t 個(指定可能)
FFTの回数	複数回	1回のみ

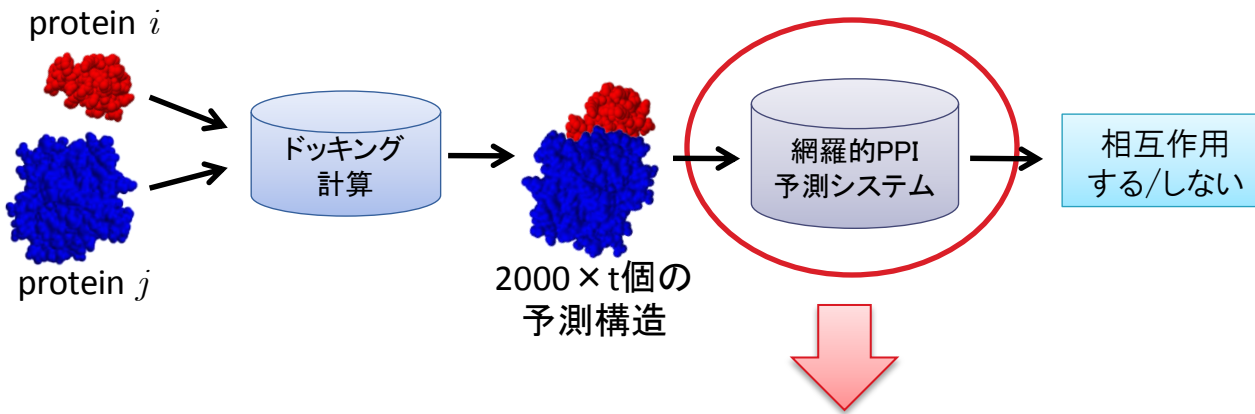
PPI予測システム部

PPI予測システム部



MEGADOCK System

ドッキング計算結果からのPPI予測の方法



1. エネルギー計算に基づくリランキング

2000 × t 個をリランキングして上位1000個を取り出す

2. 構造類似度に基づくクラスタリング

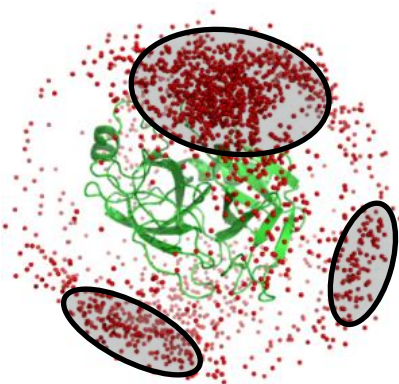
1000個の予測構造をクラスタ化する

3. PPI評価値の計算

大きいクラスタで最大のスコアを持つ構造を代表値として計算

4. PPI判定

3.で計算した評価値がある閾値以上となれば相互作用すると判定



目次

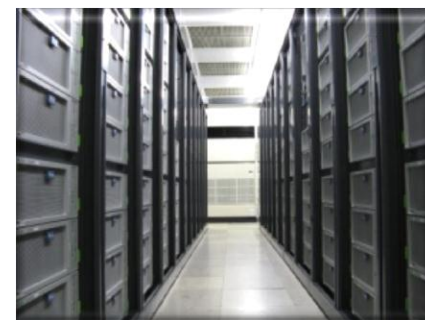
- 研究背景
- MEGADOCKシステムの構成
 - ドッキング計算システム部の説明
 - PPI予測システム部の説明
- 計算機実験
 - ドッキング予測性能評価
 - 網羅的PPI予測性能評価
- システム生物学への応用

実験1 ドッキング計算性能評価

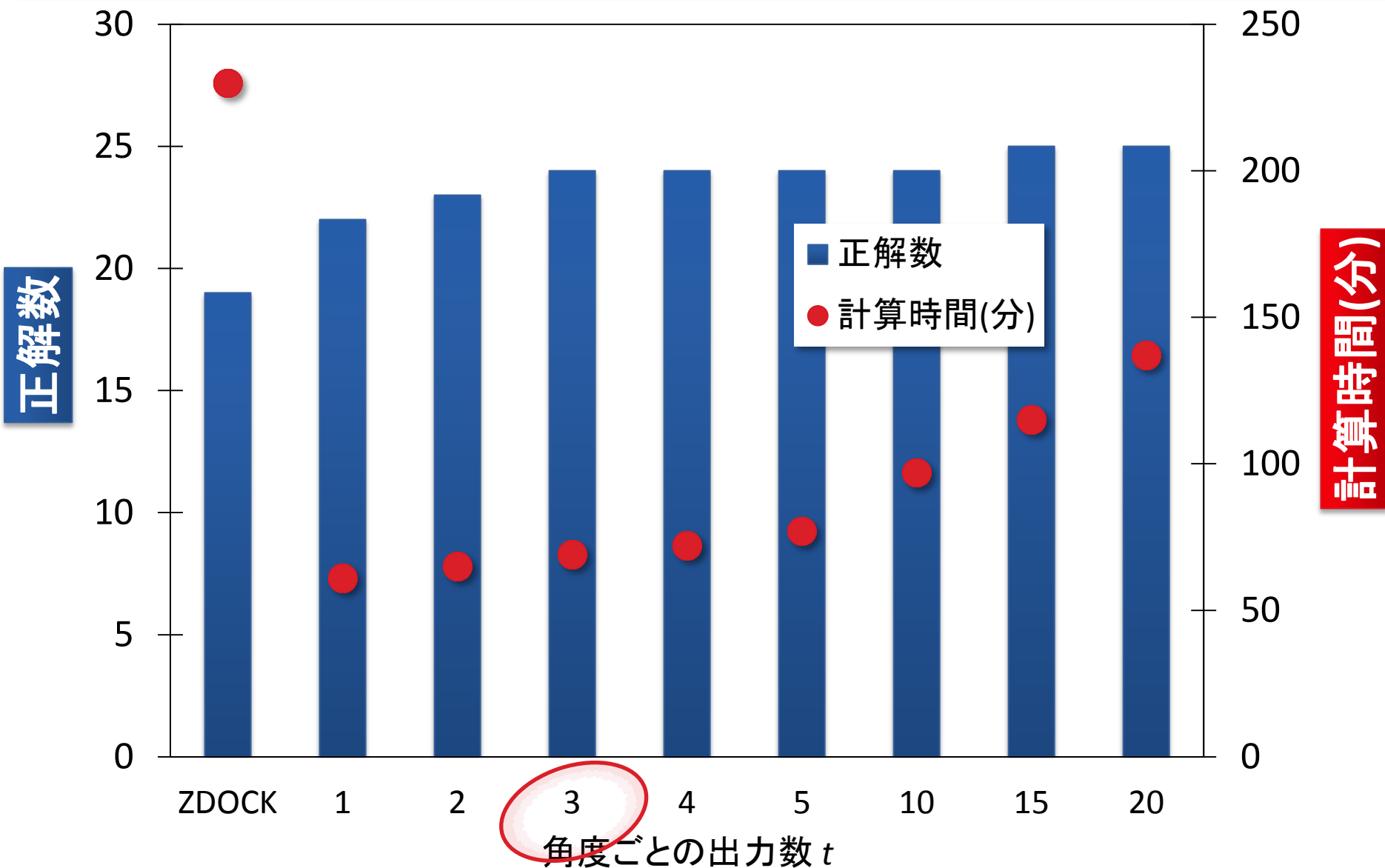
- ベンチマークデータによるドッキング計算性能評価
 - PPD benchmark2.0のモノマー同士のペア44例を使用
 - X線結晶解析による正解構造とのL-RMSDによる評価
 - L-RMSD:リガンドタンパク質の全原子を用いて計算したRMSD

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

- L-RMSDが5Å以下の予測構造を「正解構造」とする
- システム出力の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 を変えて実験
 - $t = 1, 2, 3, 4, 5, 10, 15, 20$
- TSUBAME supercluster systemを利用

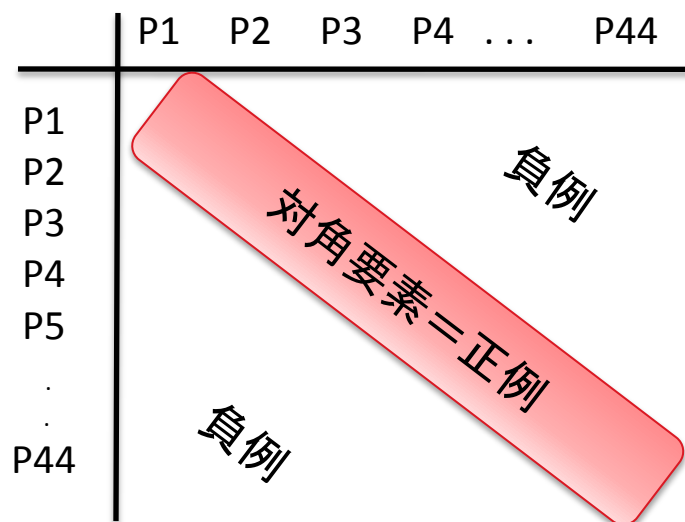


ドッキング計算性能評価



実験2 網羅的PPI予測実験

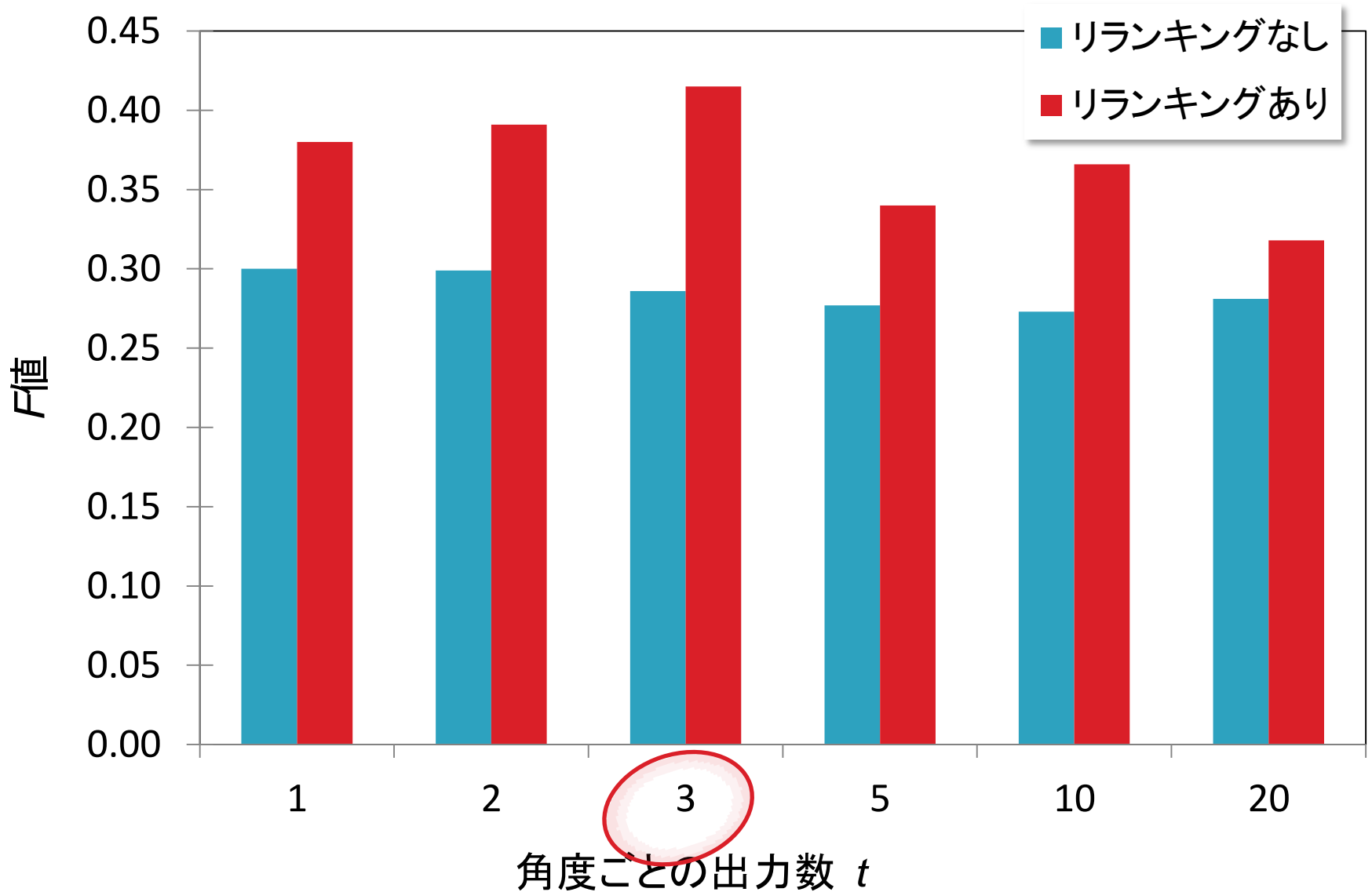
- ベンチマークデータを使ったPPI予測実験
 - モノマー同士のペア44例を使用
 - 44 × 44の総当たりドッキングによる予測



- F 値(F -measure)によって評価
 - 適合率と再現率の調和平均
- 角度ごとの出力数 t を変えて実験
 - $t = 1, 2, 3, 5, 10, 20$

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

相互作用予測結果



従来手法との比較

手法 (データセット: 44 × 44)	F値
スコア閾値を利用した方法(Ohue, 2009)	0.15
Boostingを用いた方法(Ohue, 2009)	0.25
ZDOCKを利用した方法(Matsuzaki, 2009)	0.43
本研究における手法	0.42

4倍の計算速度で同等の精度が得られた

参考

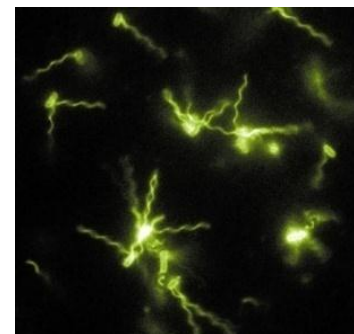
AEP(Yoshikawa, 2009) データセットは84 × 84	0.06
ZDOCKと機能情報を利用(Yoshikawa, 2010) データセットは12 × 12	0.47

目次

- 本研究について
- MEGADOCKシステムの構成
 - ドッキング計算システム部の説明
 - PPI予測システム部の説明
- 計算機実験
 - ドッキング予測性能評価
 - 網羅的PPI予測性能評価
- システム生物学への応用

システム生物学への応用

- 細菌走化性シグナル伝達系
 - システム生物学の典型的な問題
 - 分子の相互作用関係がよく理解されている
 - 一部未解明な問題も存在する



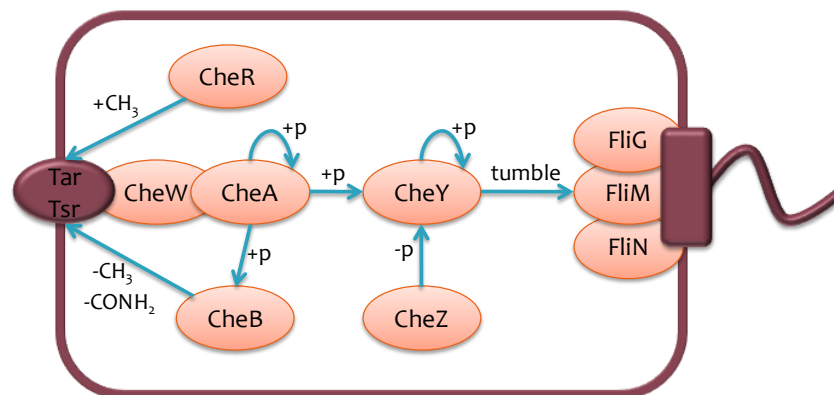
<http://www.mcb.harvard.edu/>

- 方法

- パスウェイ中の13種のタンパク質に対応する89構造を利用
- MEGADOCKによる網羅的PPI予測

細菌走化性系の13タンパク質

CheA CheB CheC CheD CheR
CheW CheX CheY CheZ Tsr
FliG FliM FliN

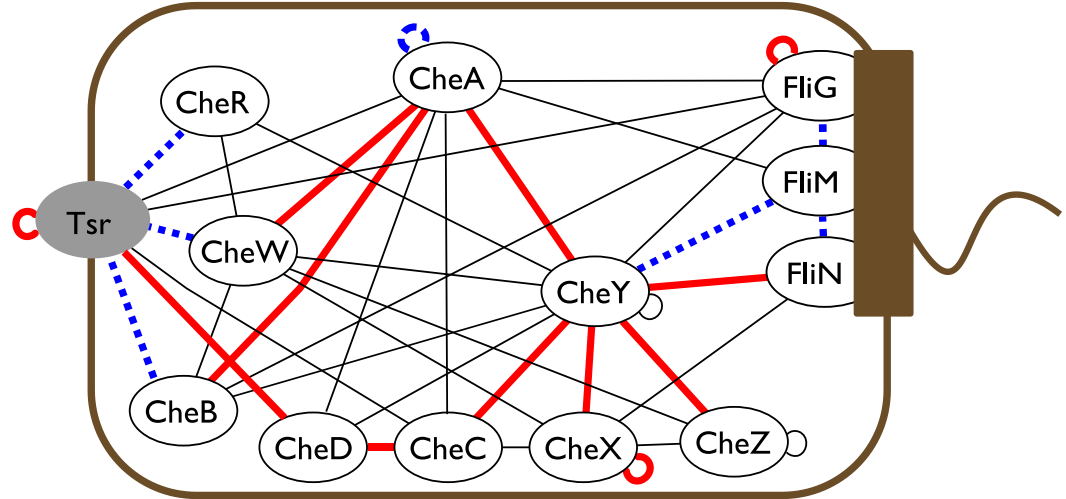


細菌走化性系PPI予測

- 結果

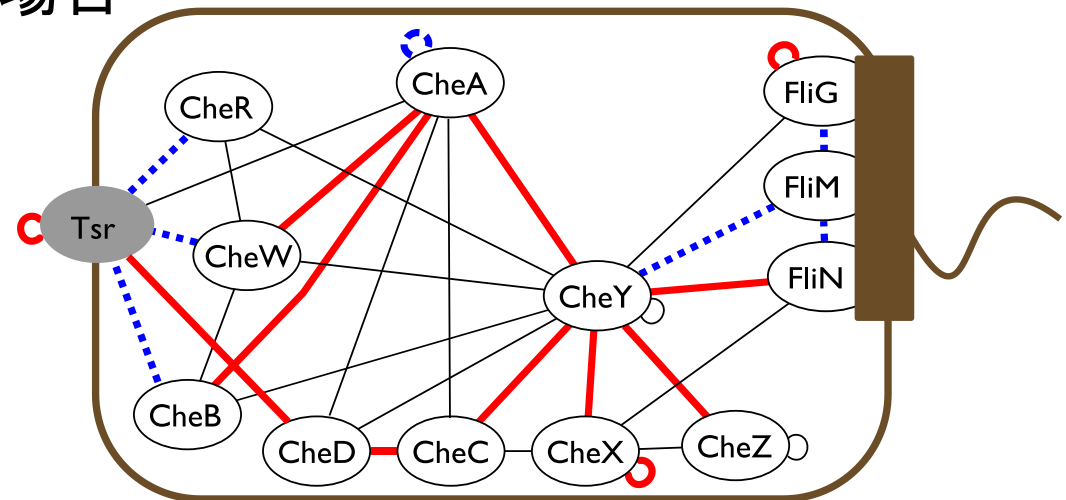
- 予測精度: F 値 0.44

赤太線: True Positive
青点線: False Negative
黒実線: False Positive



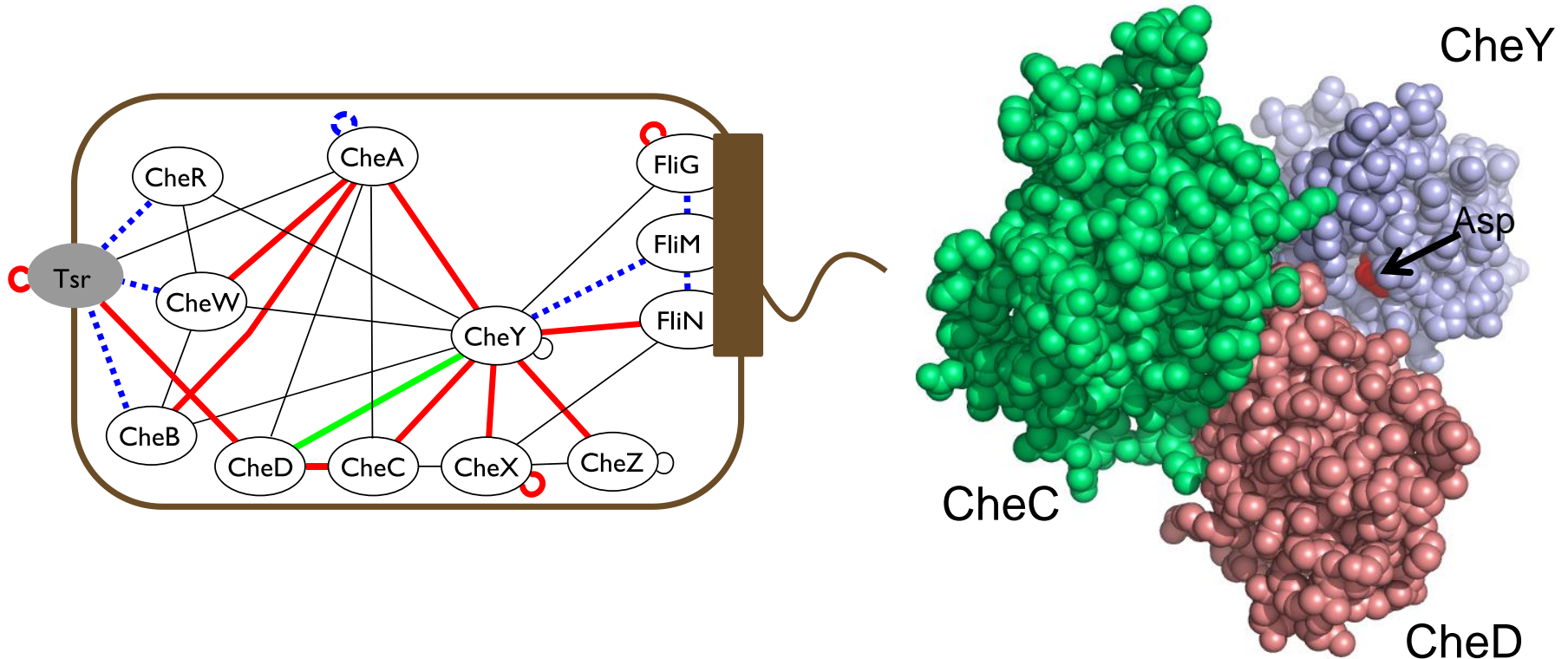
- 細胞内局在を考慮した場合

- F 値 0.52



細菌走化性系PPI予測

- CheD-CheY間相互作用を示唆
 - False Positiveの中で特にPPI評価値が高かったもの



- CheD-CheC, CheC-CheYの相互作用は知られている
- CheDは, CheCのCheYに対する脱リン酸化活性を高める

おわりに

- MEGADOCKを開発し、網羅的PPI予測を行った
 - rPSCの考案による高速な予測計算
 - ZDOCKと同等の精度を維持しつつ、4倍の高速化
 - 角度ごとの候補数を増やすことによる精度向上
 - 細菌走化性系への応用
 - ベンチマークと同程度の精度を得た
 - 未知のPPIの可能性を示唆
- 今後の課題
 - 構造予測手法との組み合わせ
 - 実問題への適用
 - ヒトEGFRシグナル伝達系(25万件の計算を予定)
 - β_2 ARシグナル伝達系

謝辞

- 本研究は以下の支援を受けて行われた
 - 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)