AKIYAMA Laboratory
Bioinformatics

東京工業大学
Tokyo Institute of Technology

Improvement of the Protein-Protein Docking Prediction by Introducing a Simple Hydrophobic Interaction Model: an Application to Interaction Pathway Analysis

Masahito Ohue[†] [‡] Yuri Matsuzaki[†] Takashi Ishida[†] Yutaka Akiyama[†]

[†] Graduate School of Information Science and Engineering,
Tokyo Institute of Technology
[‡] JSPS Research Fellow

The 7th IAPR International Conference on
Pattern Recognition in Bioinformatics
November 9th, 2012, Tokyo, Japan

PRIB 2012

Summary

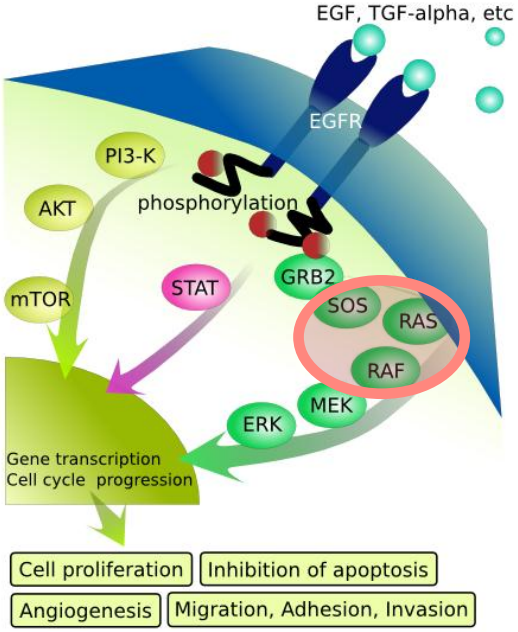
- **MEGADOCK is a PPI network prediction system using all-to-all protein docking calculation**
  - Required a rapidly calculation for all-to-all prediction
  - MEGADOCK is 8.8 times faster than ZDOCK
- **We proposed new docking score model added a hydrophobic interaction keeping rapidness**
- **We achieved the better level of accuracy without increasing calculation time**

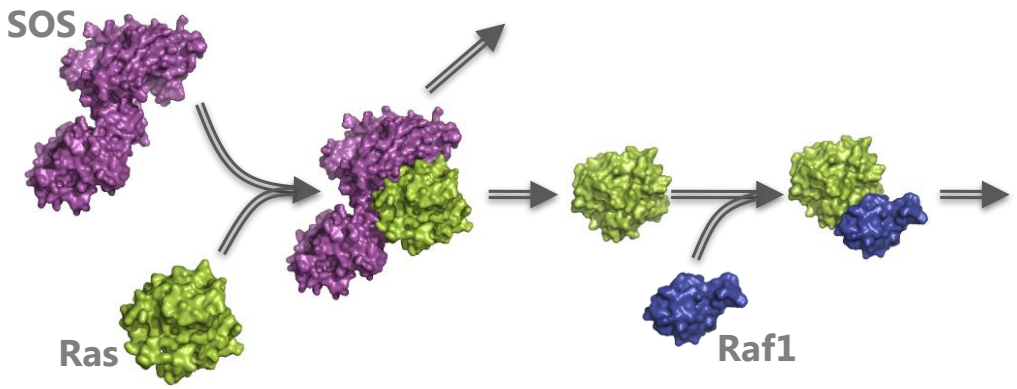
Introduction

Background

- **Proteins play a key role in biological events**

ex) Signal transduction





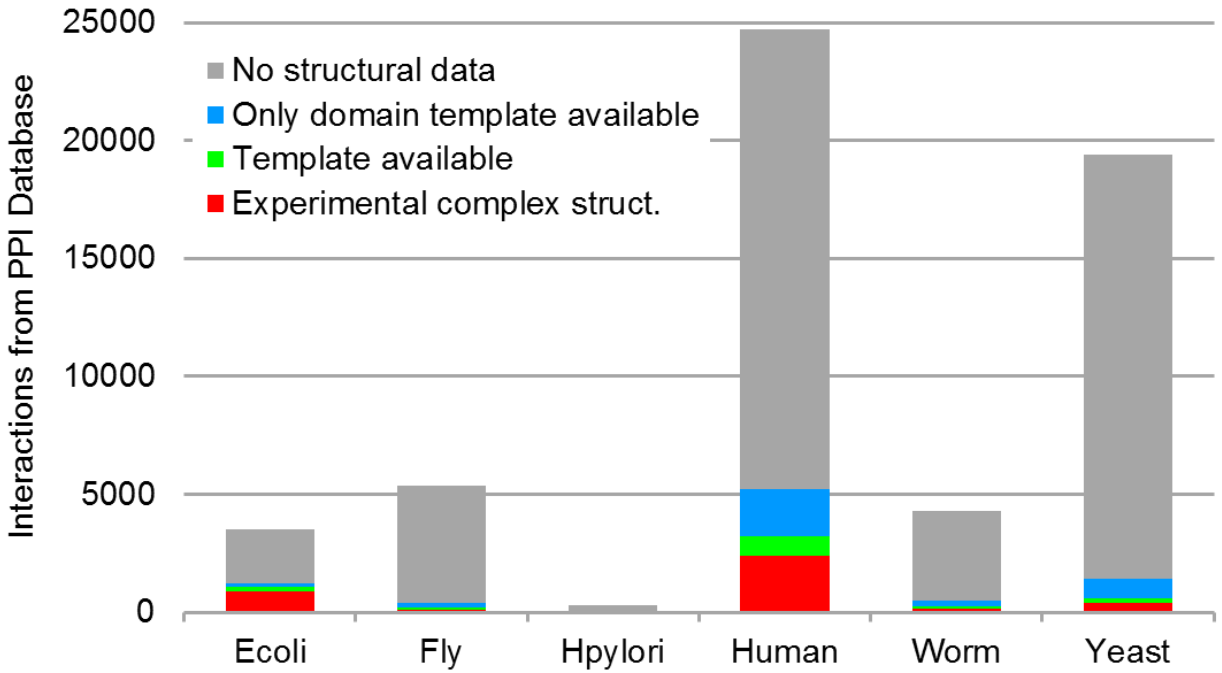
http://en.wikipedia.org/wiki/Epidermal_growth_factor_receptor

Many proteins display their biological functions by binding to a specific partner protein at a specific site

Background

- **PDB crystal structures are currently increasing but interacted structures are not enough**

Structural coverage of interactions Modified from Stein A, 2011.



While there is structural data for a large part of the interactors (60–80%), the portion of interactions having an experimental structure or a template for comparative modeling is limited (less than 30%)

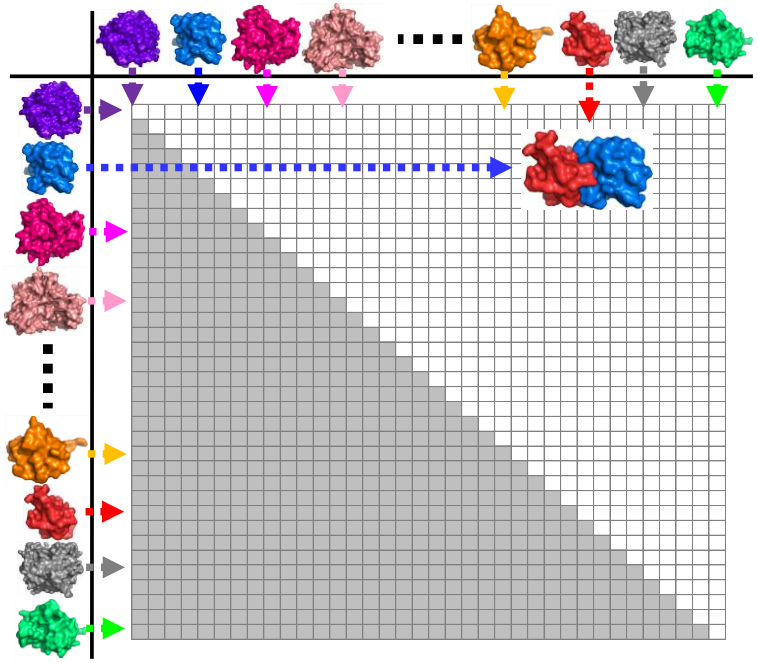
→ Computational protein docking is very important

Background

- **Protein docking is useful of all-to-all protein-protein interaction prediction**

  - 1000 proteins' combinations
    →500,500 (1000x1001/2)
  - 500,500 PPI predictions require
    43,000 CPU days when using ZDOCK*
  - On 400 nodes x 12 cores CPU/node
    → still require 9 days

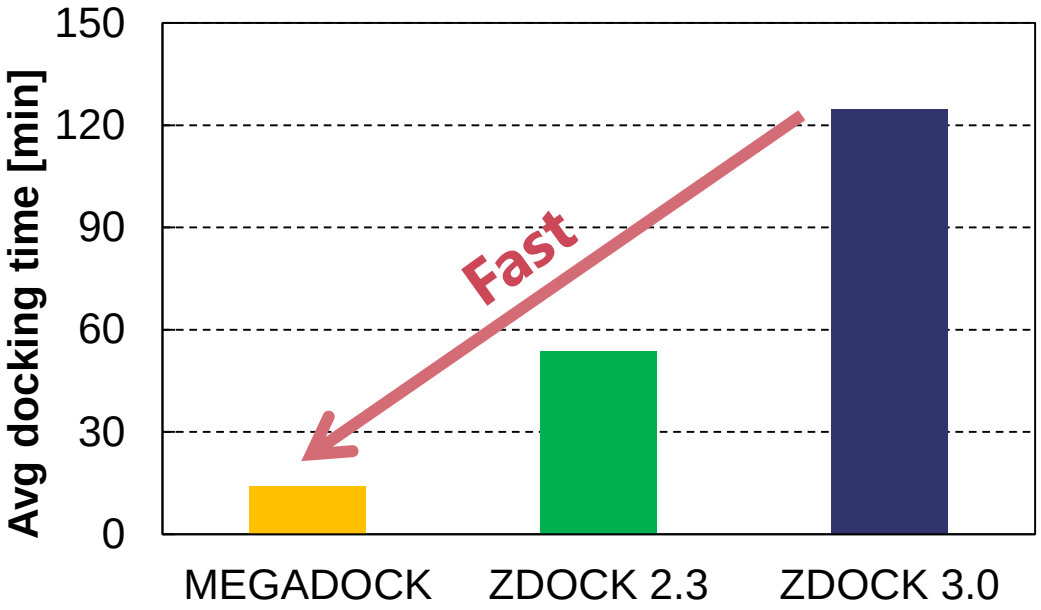
*Mintseris J, *et al. Proteins*, 2007.



Need a fast protein docking software for all-to-all protein-protein interaction prediction

Background

- MEGADOCK has a rapid protein docking engine



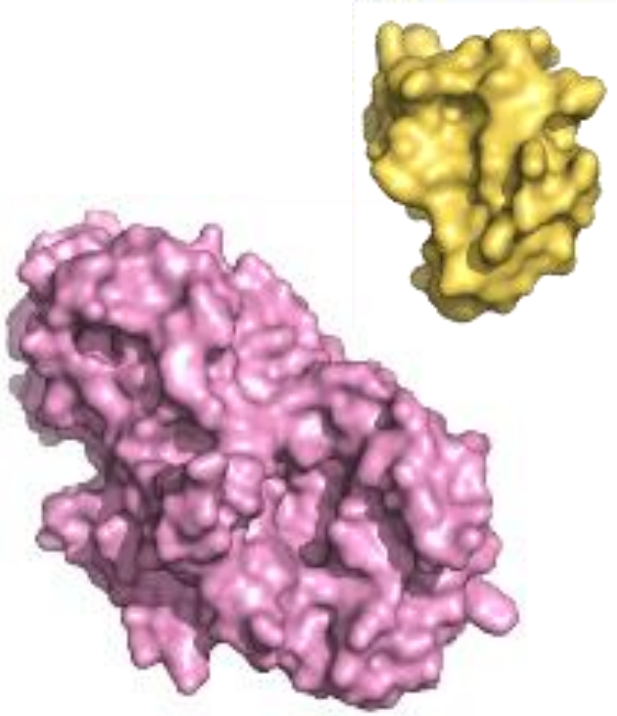
All-to-all PPI Prediction
in 1000 proteins

9 days → **1 day**
ZDOCK 3.0 MEGADOCK

- However, MEGADOCK's docking accuracy is worse compared to ZDOCKs
- Our purpose is developing more accurate protein docking method with rapidness

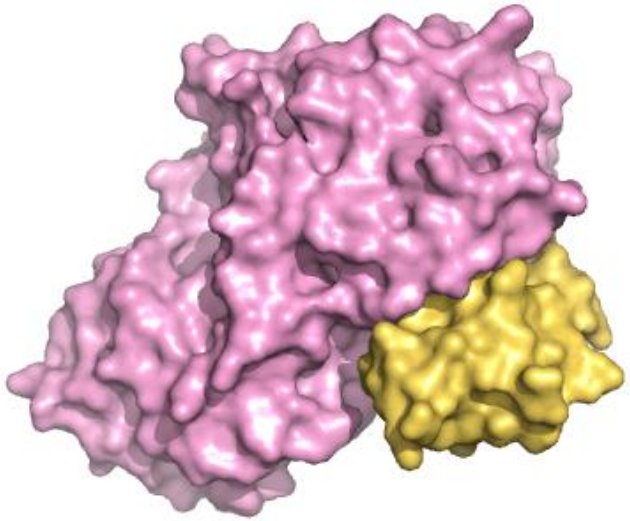
Protein Docking Problem

Protein monomeric structure pair

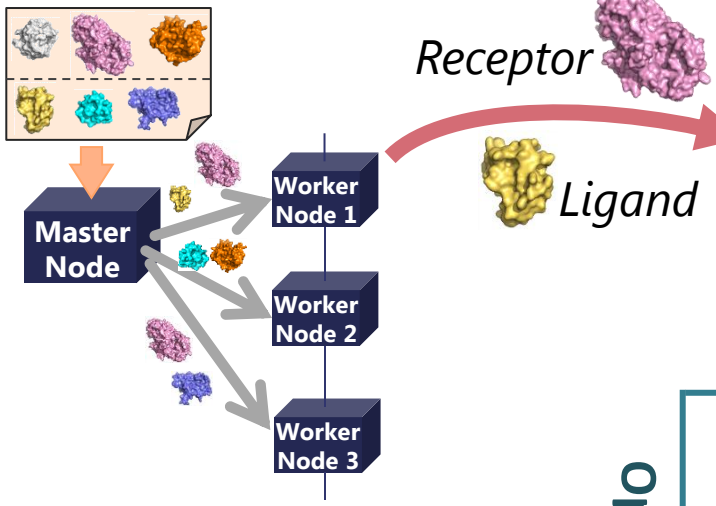


Docking calculation

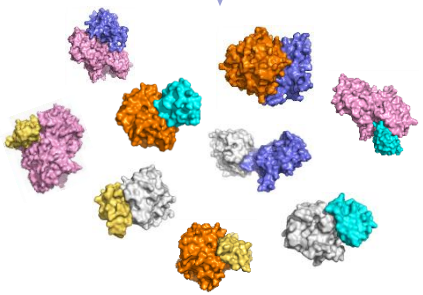
Protein complex (predicted structure)



MEGADOCK Docking Engine Overview



MPI Parallelization



Receptor

Ligand

Receptor voxelization

Receptor FFT

Loop for *Ligand* rotations

Ligand voxelization

Ligand FFT

Convolution

Score calculation (Inverse FFT)

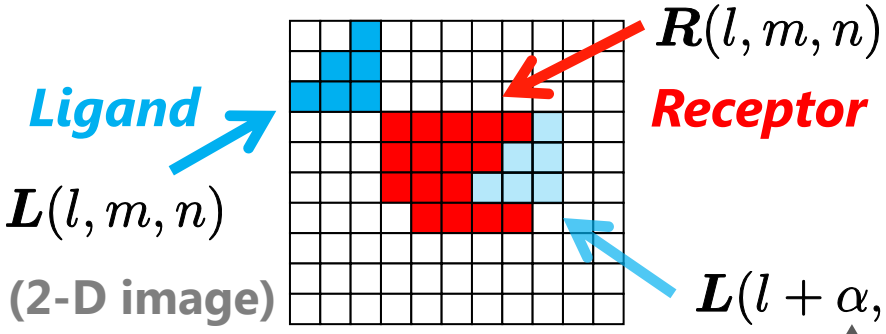
Save high scoring docking poses

OpenMP Parallelization

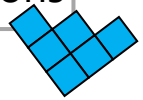
3-D Grid Convolution by FFT

- 3-D Grid Convolution

Katchalski-Katzir E, *et al. PNAS*, 1992.



$\Delta\theta$ is 15 degree → 3600 patterns of Ligand rotations



Docking Score

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

Convolution $\mathcal{O}(N^6)$

$$S(\alpha,\beta,\gamma)=\mathrm{IFFT}\left[\mathrm{FFT}[\boldsymbol{R}(l,m,n)]^{*}\mathrm{FFT}\left[\boldsymbol{L}(l,m,n)\right]\right]$$

Convolution using FFT

$$\mathcal{O}(N^3\log N)$$

Previous Score Model (MEGADOCK)

- **MEGADOCK's docking score** Ohue M, *et al. IPSJ TOM*, 2010.

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

Convolution

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

$$\boldsymbol{L}(l,m,n) = G_L(l,m,n) + iw_eE_L(l,m,n)$$

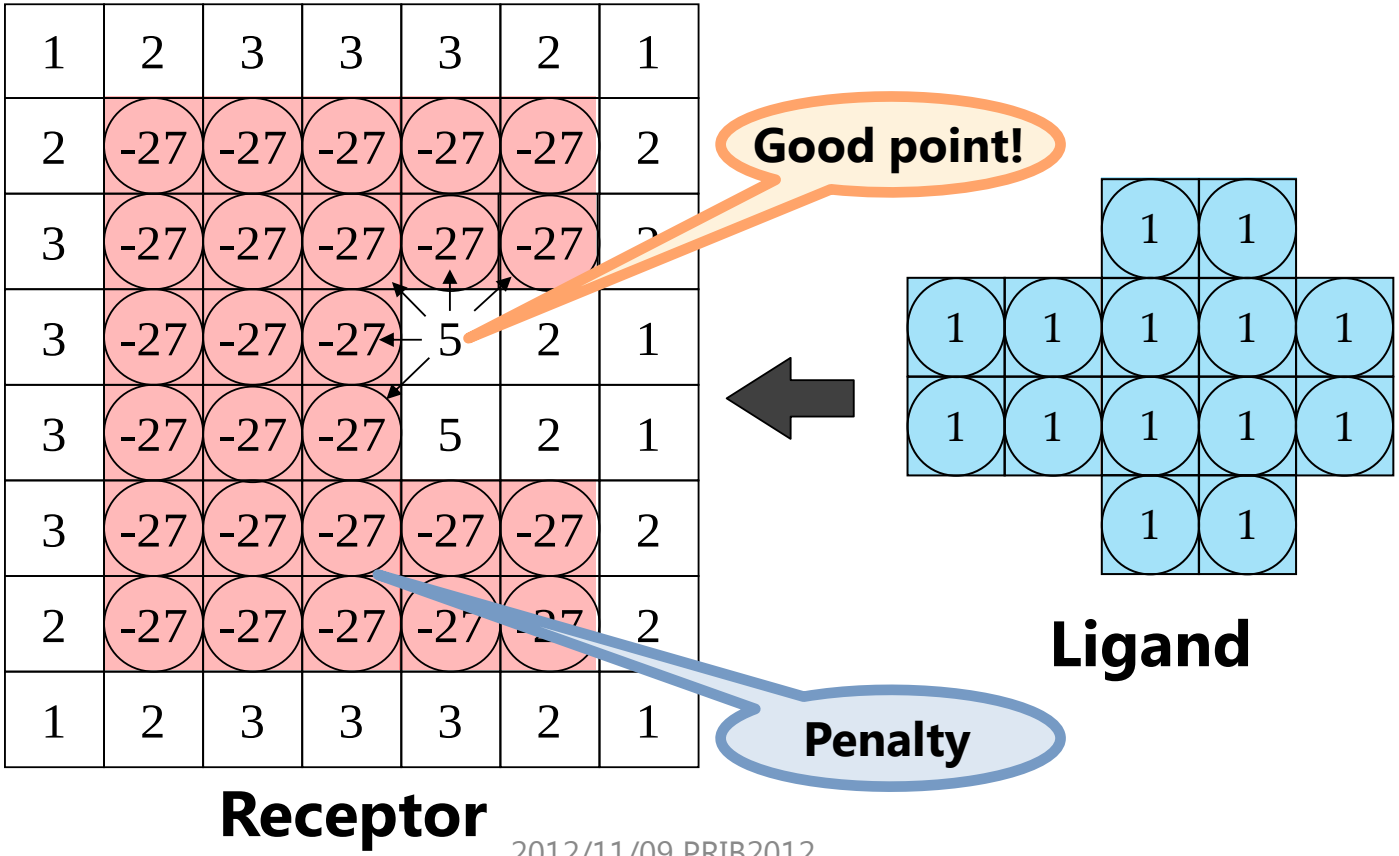
rPSC **ELEC**

– Calculate shape complementarity and electrostatics with only 1 convolution

rPSC $G_R(l,m,n),\ G_L(l,m,n)$

- **rPSC (real Pairwise Shape Complementarity)** Ohue M, *et al. IPSJ TOM*, 2010.
  - Shape complementarity score using only real value

(2-D image)



Comparison of Convolutions

- **ZDOCK 3.0 convolution** Mintseris J, *et al. Proteins*, 2007.

Shape Complementarity = Fit neatly + i Clash

Electrostatics = Coulomb force

Desolvation free energy 1 = atom *a*'s effect + i atom *b*'s effect

Desolvation free energy 2 = atom *c*'s effect + i atom *d*'s effect

⋮

Desolvation free energy 6 = atom *k*'s effect + i atom *l*'s effect

- **MEGADOCK convolution**

Docking Score = Shape Complementarity (**rPSC**) + i Coulomb force

Accurate and Fast Docking

- **For more accurate protein docking**
  - Considering **hydrophobic interaction**
  - Previous methods require high calculation cost
    - ZDOCK 2.3・・・2 convolutions
    - ZDOCK 3.0・・・8 convolutions
- **For keeping rapid calculation**
  - We should keep 1 time convolution
- **We proposed simple hydrophobic interaction model with keeping 1 convolution**

Materials and Methods

Proposed Model

- **Implementation of hydrophobic interaction**
  - Used **Atomic Contact Energy** score
- **Atomic Contact Energy (ACE)** Zhang C, *et al. J Mol Biol*, 1997.
  - The empirical atomic pairwise potential for estimating desolvation free energies
  - Used by ZDOCK, FireDock*, etc. *Andrusier N, *et al. Proteins*, 2007.

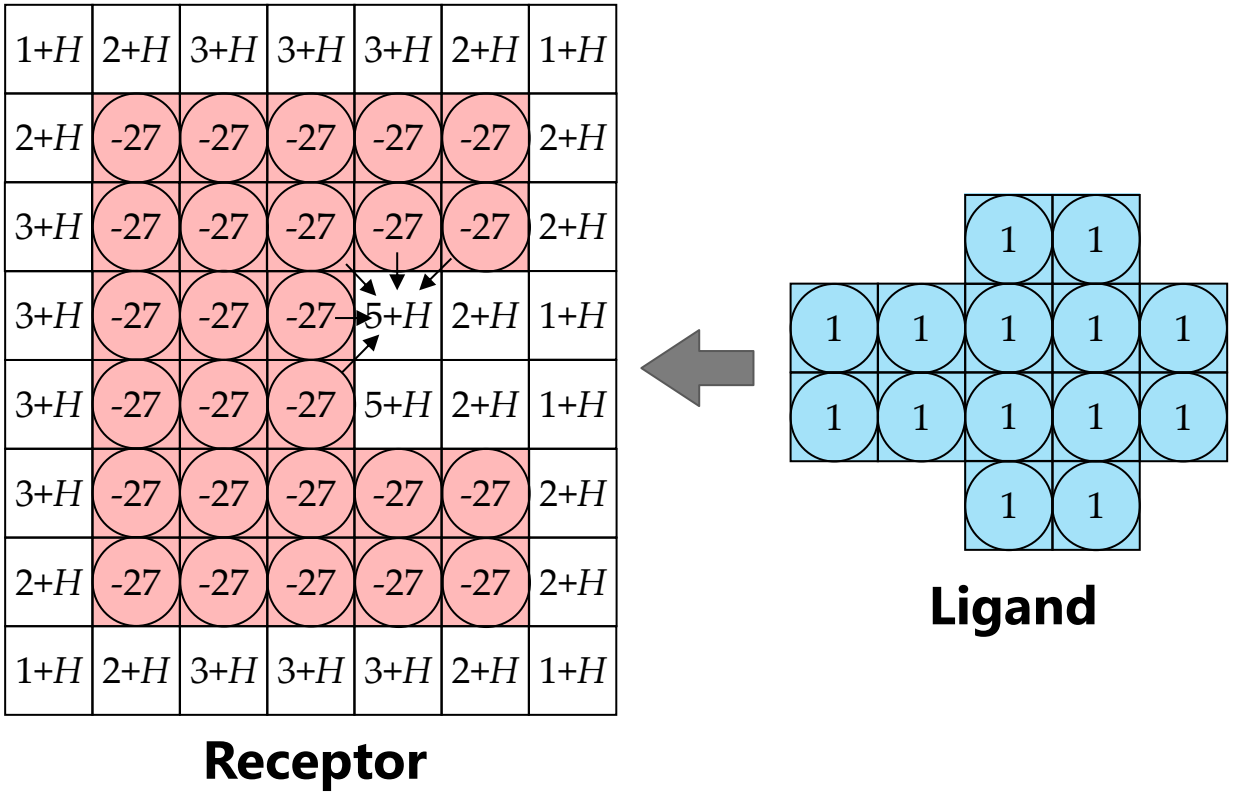
- **How to add the ACE to MEGADOCK?**
  - Add the averaged ACE value (**non-pairwise**) of surface atom of Receptor

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

Modified rPSC Model

Modification of rPSC Model

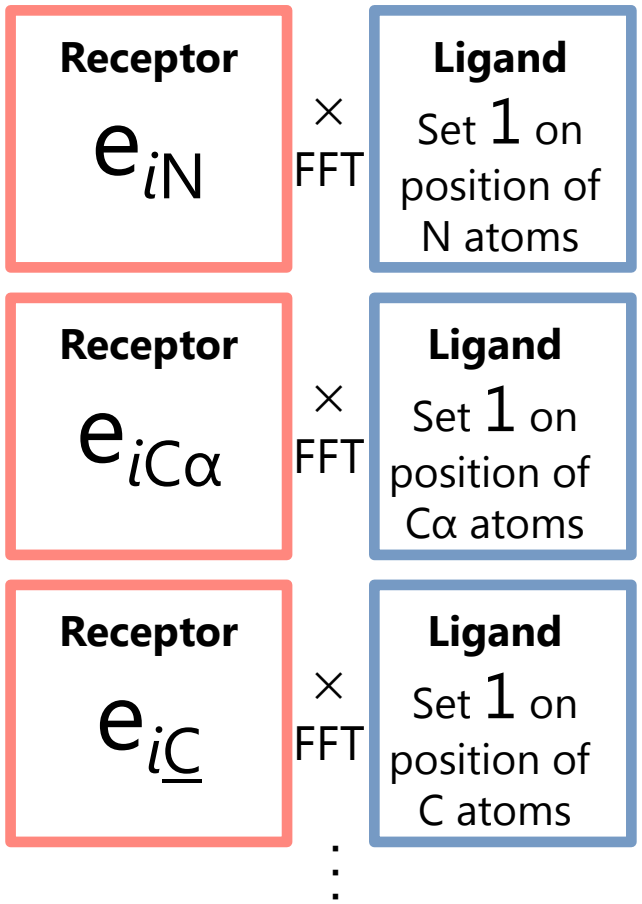
- Considered hydrophobic surface of the receptor using non-pairewise ACE



$$H(l,m,n) = \begin{cases} \text{Sum of near atoms' ACE} & \text{(space)} \\ 0 & \text{(inside of the receptor)} \end{cases}$$

Pairwise Potential

- Using table of contact energy

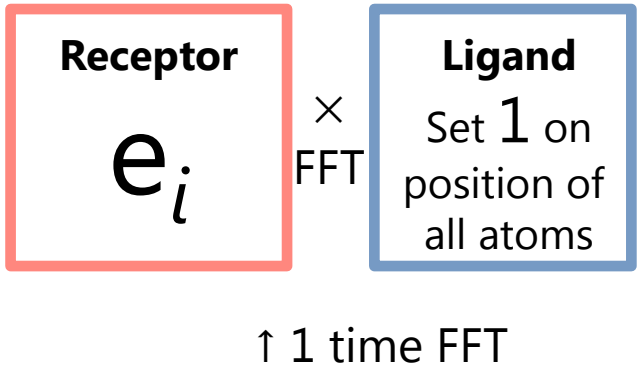


e_{ij}	N	Cα	C	…
N	-0.724	-0.903	-0.722	…
Cα	-0.119	-0.842	-0.850	…
C	-0.008	-0.077	-0.704	…
…	…	…	…	…

←Need (#type of atoms/2) times convolutions

Averaged (Non-Pairwise) Potential

- Using averaged contact energy



e_{ij}	N	Cα	C	…	e_i
N	-0.724	-0.903	-0.722	…	-0.495
Cα	-0.119	-0.842	-0.850	…	-0.553
C	-0.008	-0.077	-0.704	…	-0.464
…	…	…	…	…	…

Pros : high speed calculation
Cons: poor accuracy

Proposed Docking Score

- New docking score

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

$$\boldsymbol{R}(l,m,n) = G_R(l,m,n) + w_h H_R(l,m,n) + iE_R(l,m,n)$$

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

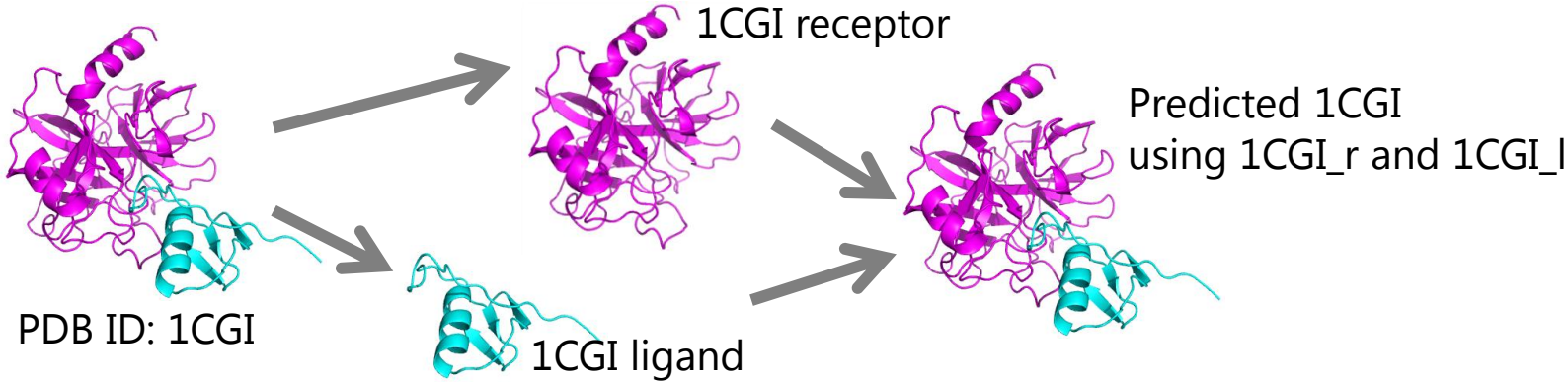
  - **Calculate 3 physico-chemical effects by only 1 convolution calculation**

Experimental Settings

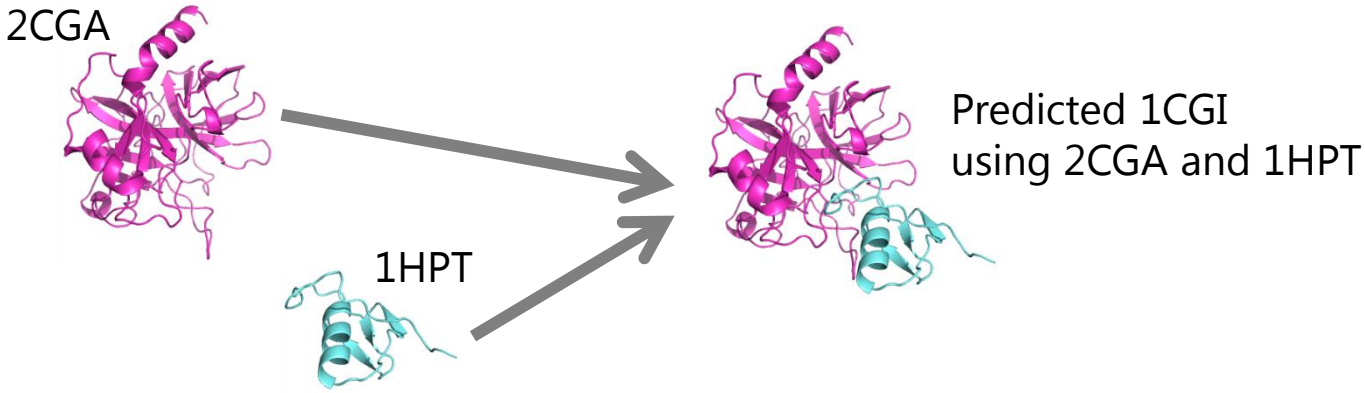
- **Dataset: Protein-protein docking benchmark 4.0**
  - 176 complex structures (include both bound/unbound form)

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

bound
(re-docking)

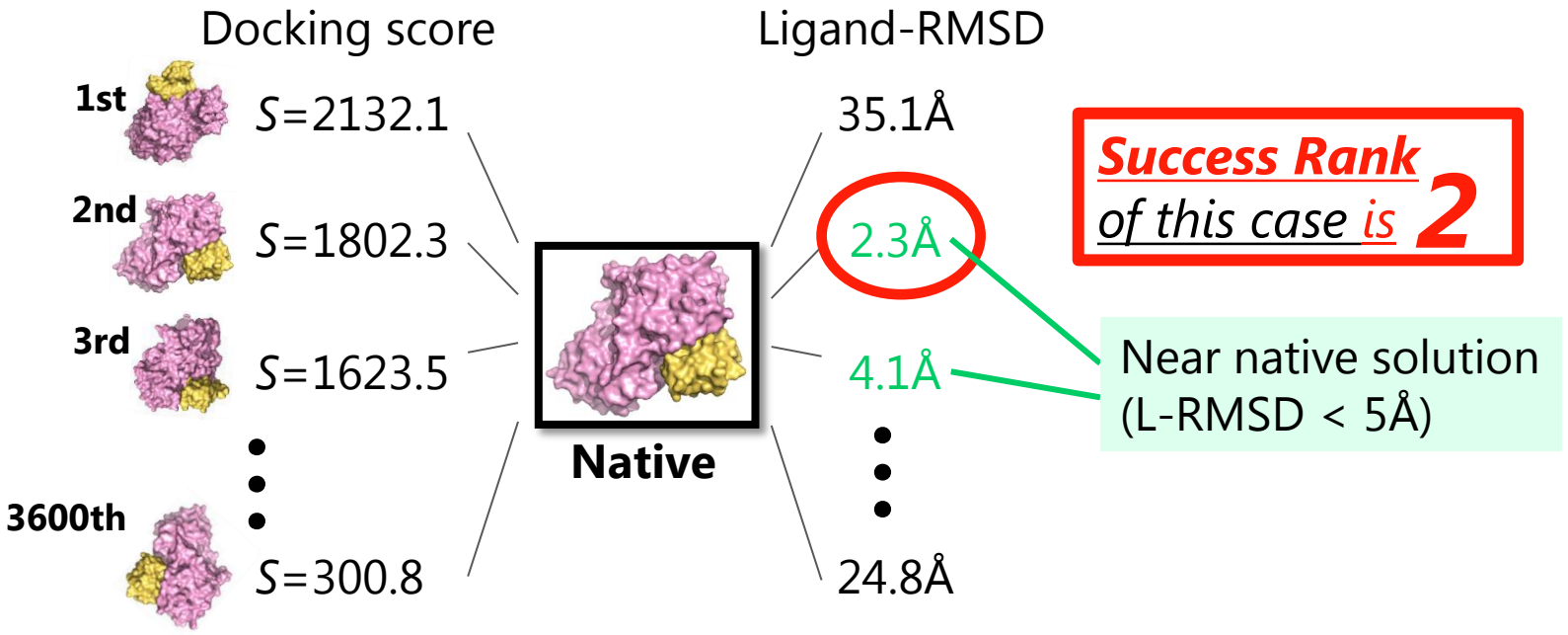


unbound
(real problem)



Experimental Settings

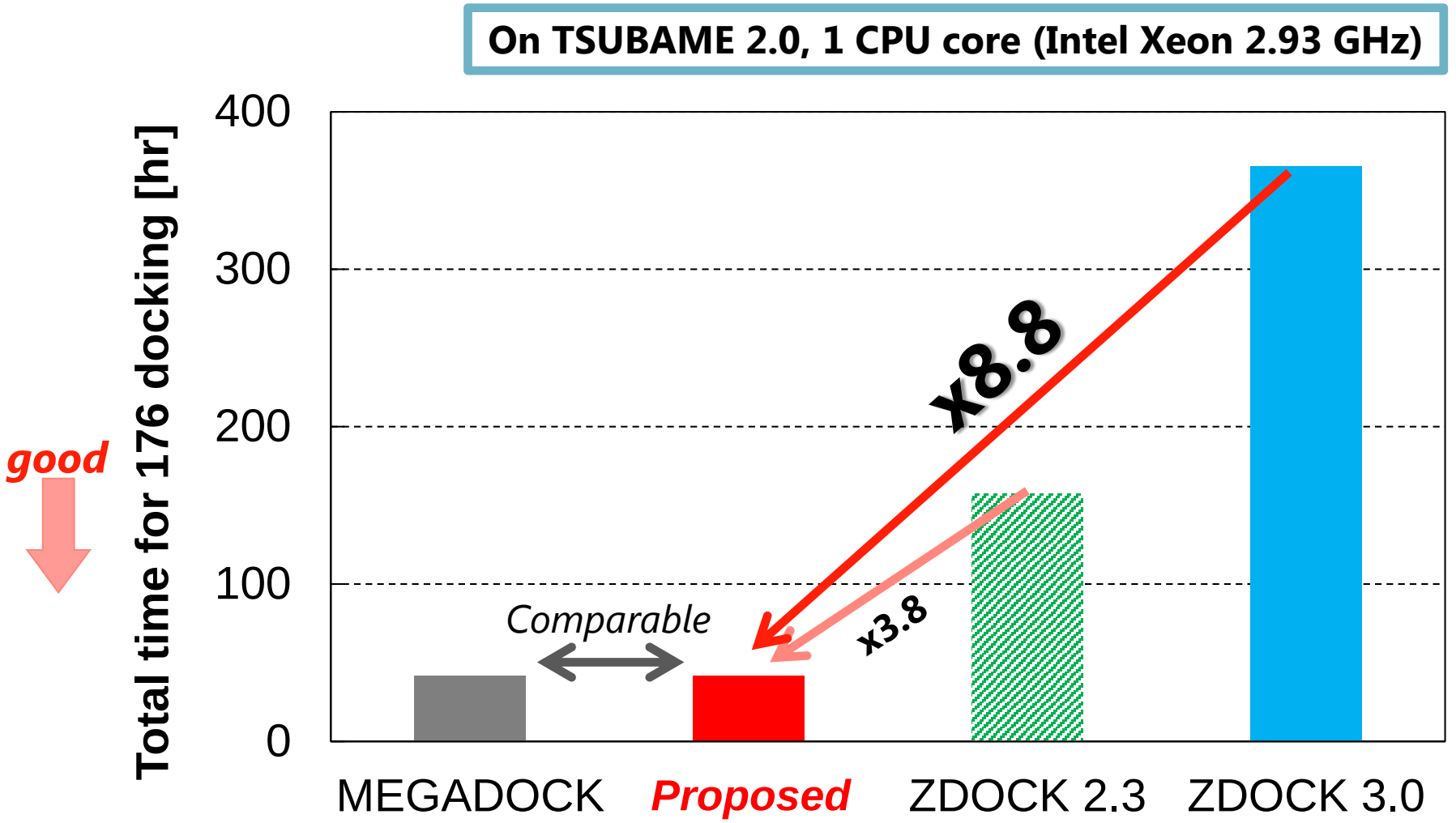
- How to evaluate accuracy? → Success Rank



Ligand-RMSD (root mean square deviation)
RMSD with respect to the X-ray structure, calculated for the all heavy atoms of the ligand residues when only the receptor proteins (predicted structure and X-ray structure) are superimposed.

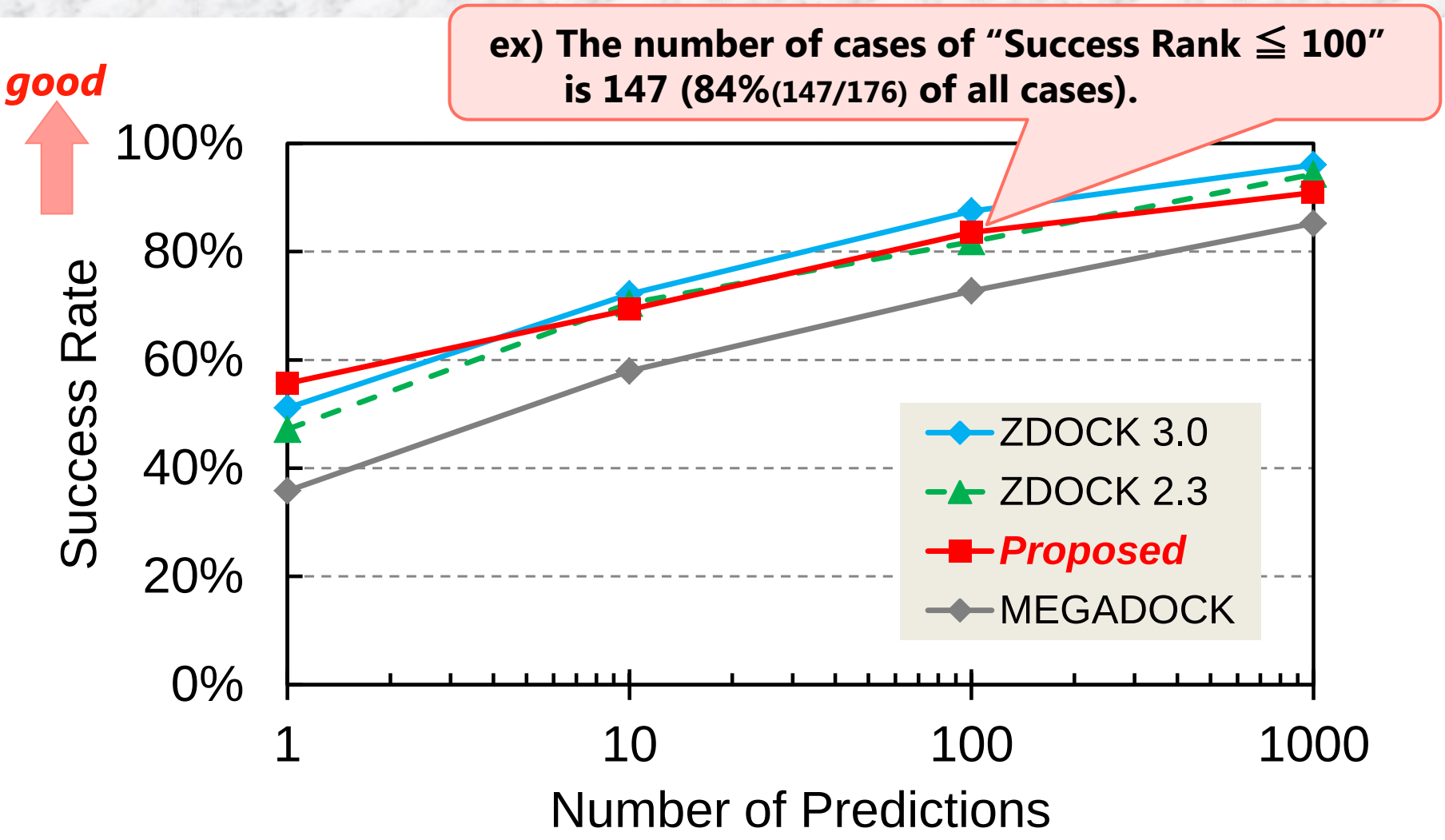
Results and Discussions

Docking Calculation Time



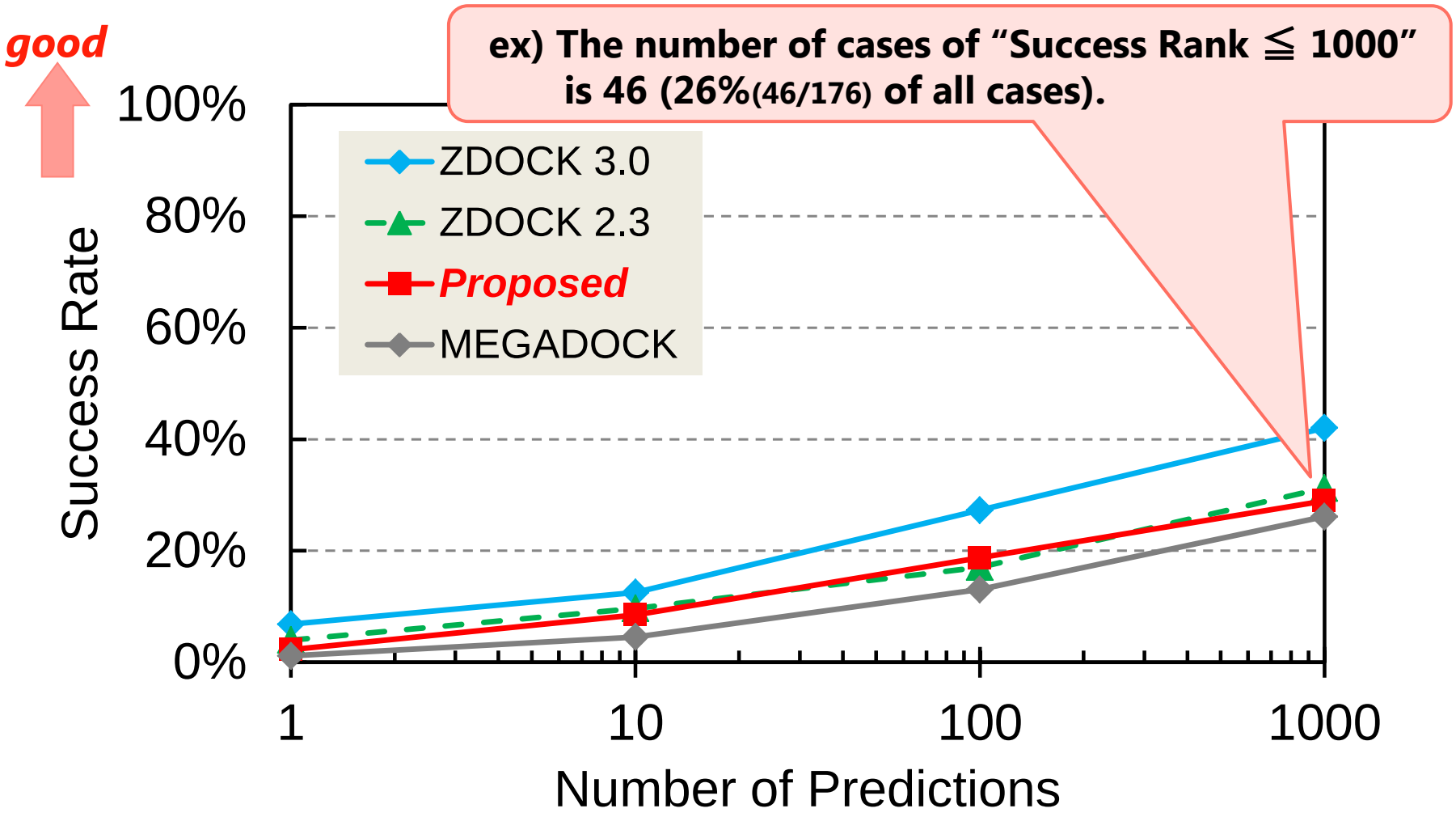
x8.8 faster than ZDOCK 3.0 / x3.8 faster than ZDOCK 2.3

Accuracy in Bound Set



Achieved the same level of accuracy as ZDOCK 3.0

Accuracy in Unbound Set



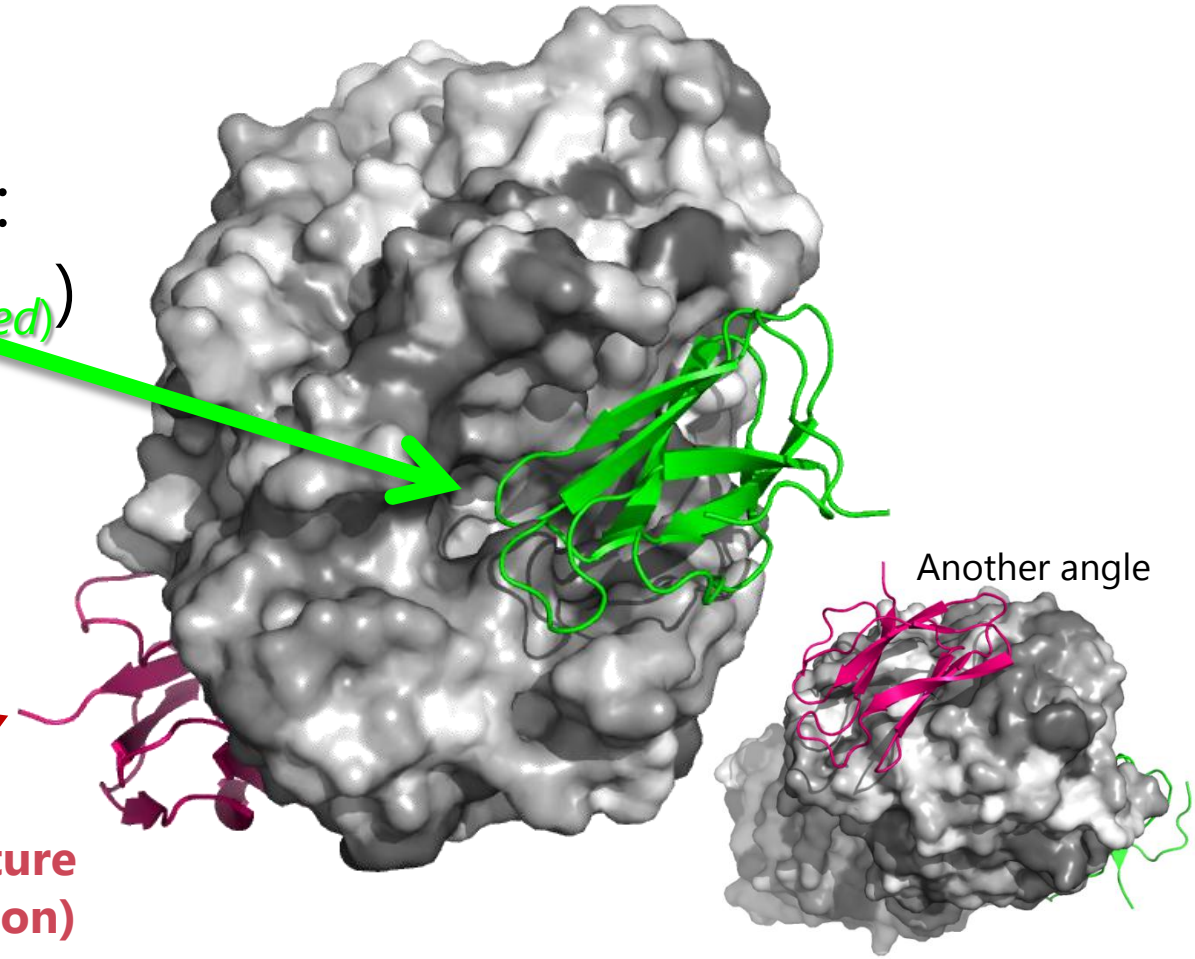
Achieved the same level of accuracy as ZDOCK 2.3

Prediction Example

PDB ID: 1BVN

(Unbound Success Rank:
11$_{(\text{MEGADOCK})} \rightarrow$ **1**$_{(\textit{Proposed})}$)

Receptor: α-amylase
Ligand: Tendamistat
(α-amylase inhibitor)



The 1st rank predicted structure by MEGADOCK (wrong position)

Hydrophobicity scale*

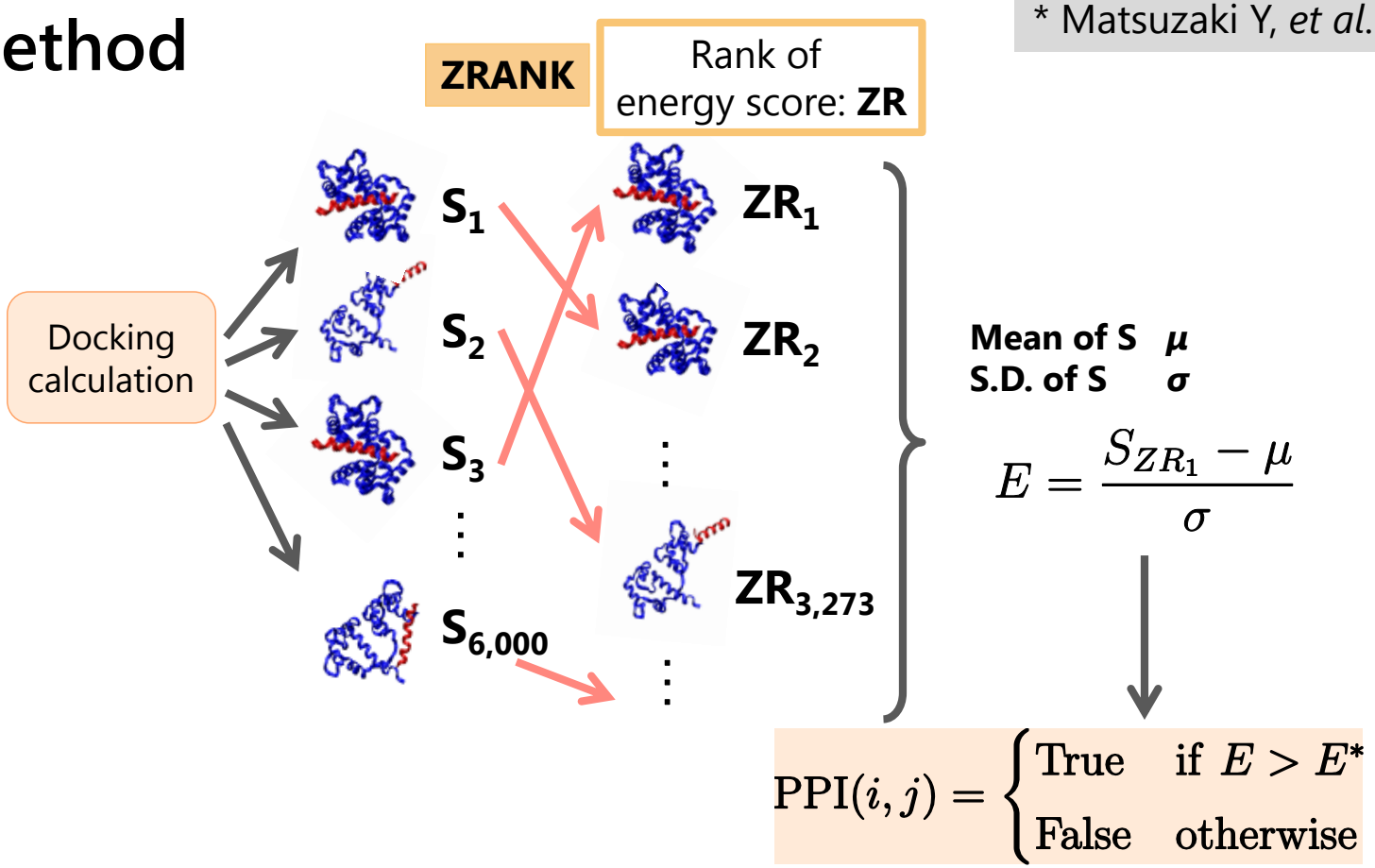
hydrophilic — hydrophobic

*Eisenberg D, *et al. J Mol Biol*, 1984.

Application to Pathway Analysis

- **Bacterial chemotaxis signaling pathway prediction**
  - Predicting "interact or not" from docking results*
- **Method**

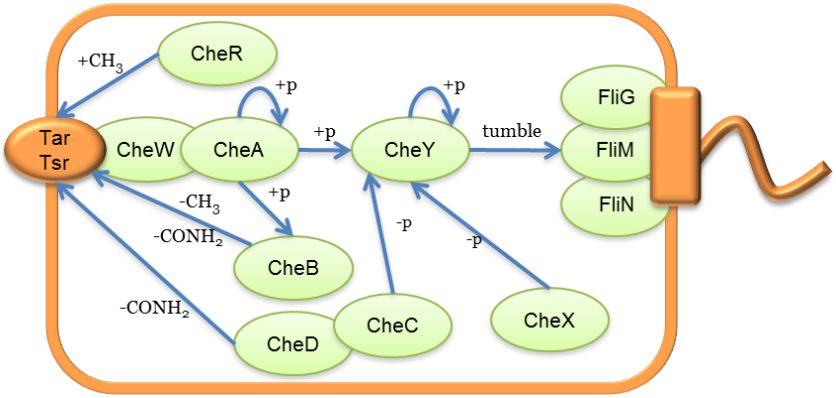
* Matsuzaki Y, *et al. JBCB*, 2009.

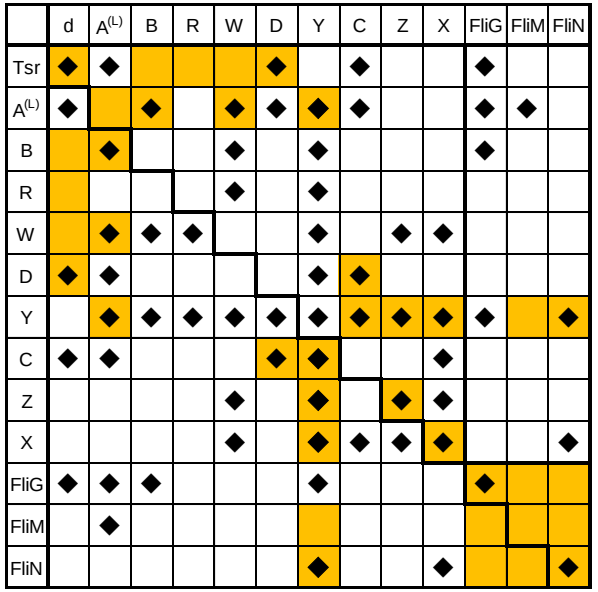


Application to Pathway Analysis

- **Bacterial chemotaxis signaling pathway prediction**
  - Predicting "interact or not" from docking results*

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





  - **Results** (F-measure)

MEGADOCK	*Proposed*	ZDOCK 3.0
0.43	0.45	0.49

Conclusion

Conclusion

- **We proposed novel docking score including simple hydrophobic interaction**
- **We achieved the better level of accuracy without increasing calculation time**
  - Same level of accuracy as ZDOCK 2.3
  - 8.8 times faster than ZDOCK 3.0
    3.8 times faster than ZDOCK 2.3
- **Future works**
  - Developing a hydrophobic interaction model considering both receptor and ligand protein
  - Applying to large PPI network and cross-docking of structure ensemble

Acknowledgements

- **Akiyama Lab. Members**

TECH chan

- **This work was supported in part by**
  - a Grant-in-Aid for JSPS Fellows, 23･8750
  - a Grant-in-Aid for Research and Development of The Next-Generation Integrated Life Simulation Software
  - Educational Academy of Computational Life Sciences

  from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).