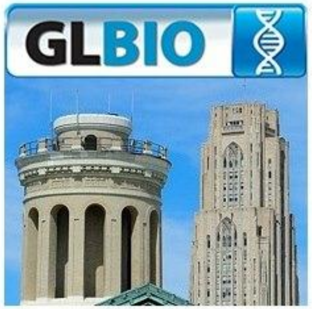
東京工業大学
Tokyo Institute of Technology



Highly Precise Protein-Protein Interaction Prediction Based on Consensus Between Template-Based and *de Novo* Docking Method

Masahito Ohue[†,‡] **Yuri Matsuzaki**[†] **Takehiro Shimoda**[†]
Takashi Ishida[†] **Yutaka Akiyama**[†]

[†] Grad. School of Information Sci. and Eng., Tokyo Institute of Technology, JAPAN
[‡] Research Fellow of the Japan Society for the Promotion of Science, JAPAN

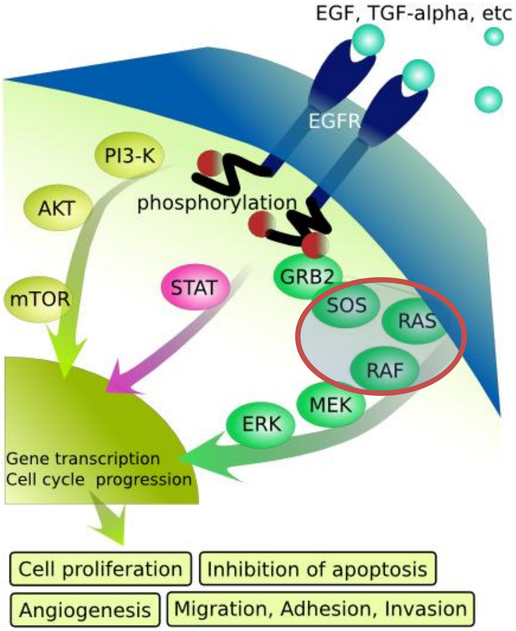
The Great Lakes Bioinformatics Conference 2013 May 15, 2013

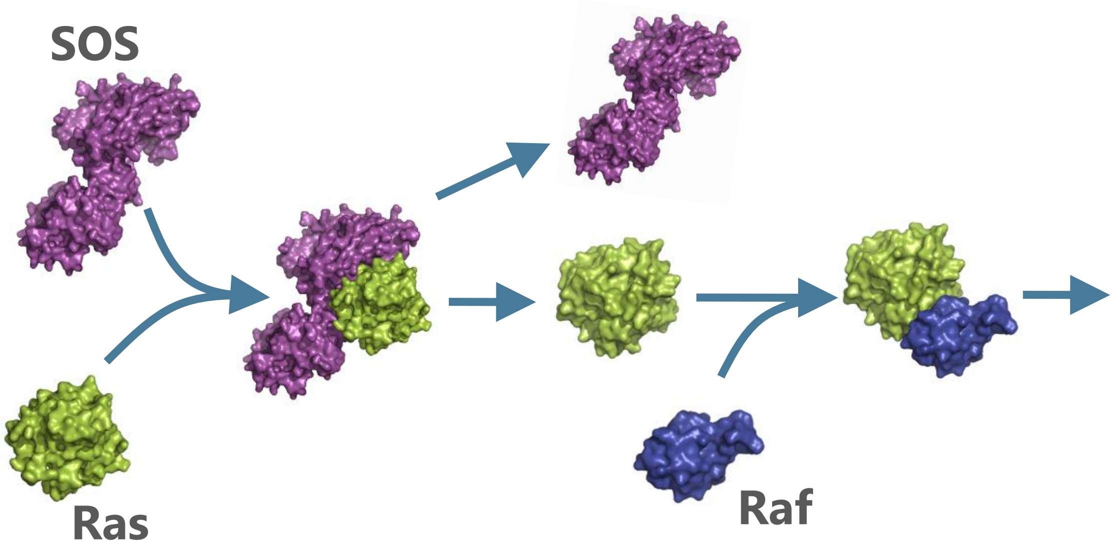
Introduction

Background

- Proteins play a key role in biological events

ex) Signal transduction



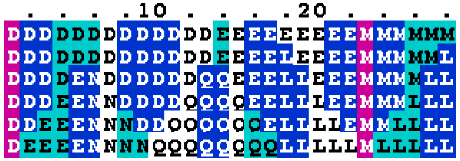


http://en.wikipedia.org/wiki/Epidermal_growth_factor_receptor

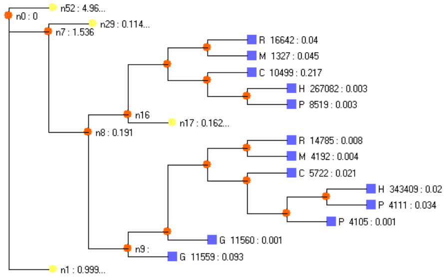
Predicting protein-protein interaction (PPI) networks is one of the main topics in systems biology

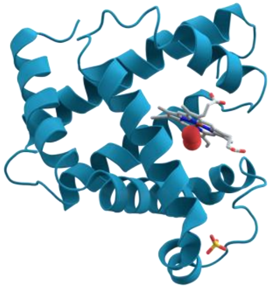
Structure-based PPI prediction

- **PPI prediction methods**
  - Sequence-based method
  - Evolutional information-based method
  - Tertiary structure-based method



```
. . . .10 . . . .20 . . . .
DDDDDDDDDDDDDEEEEEEEEEMMMMMM
DDDDDDDDDDDDDEEEELEEEEMMMMML
DDDDENDDDDDDQQEEELLEEEMMMMLL
DDDEENNDDDDQQQQEELLLEEMMMLLL
DDEEENNNDDOOOOOOELLLLEMMLLLL
DEEEENNNNQQQQQQQQLLLLLMLLLLL
```

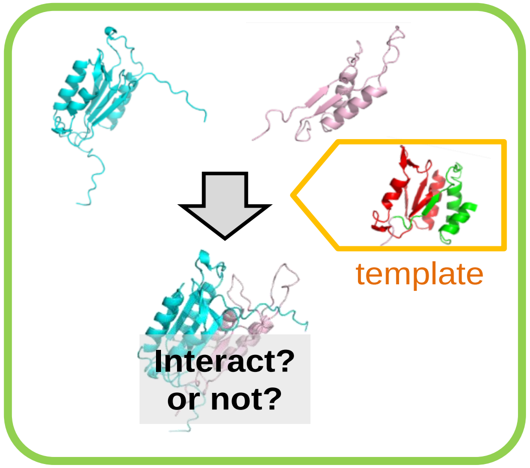




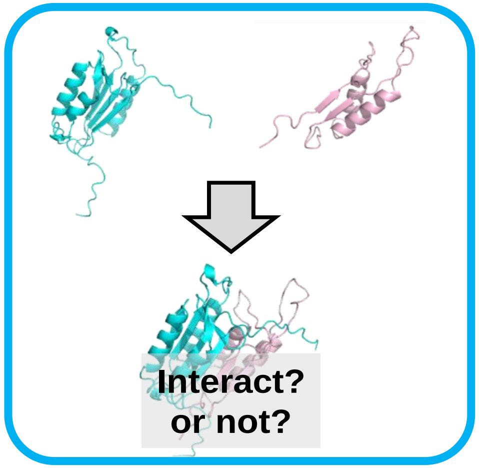
- **Why use tertiary structure?**
  - Functions are determined by structure
  - Predicted protein complex structures are useful for structure-based drug design

Structure-based PPI prediction

- **Structure-based PPI prediction method**
  - **Template-based method**
    - e.g. PRISM [Tuncbag+2011]
    - Using monomeric structures & **solved complex structures** (from PDB)



  - ***De novo* docking method**
    - a.k.a. Template-free docking
    - e.g. MEGADOCK [Ohue+2013]
    - Using only monomeric structures



Template-based method

- Concept

  known complex structures or interface architectures can be used to model the complex formed between two target proteins

- Pros and Cons

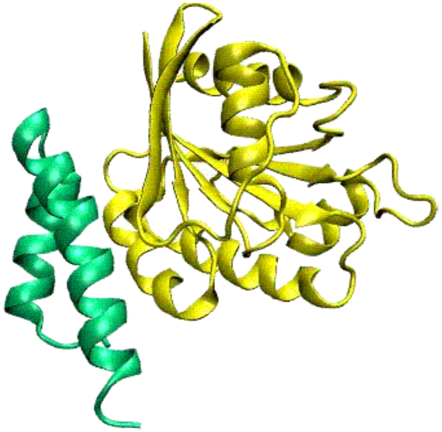
  ☺ Good prediction accuracy

  ☹ Narrow applicable range (require the template set)

Template-based method

- Processes of **PRISM**[Tuncbag+2011]
  1. Surface extraction
  2. Structural matching to the template complex structure
  3. Refinement and Energy calculation

Green and Yellow : Template
Cyan and Pink : Target



http://prism.ccbb.ku.edu.tr/prism_simulation.html

De novo docking method

- Concept

  Using only monomeric (unbound) protein structures

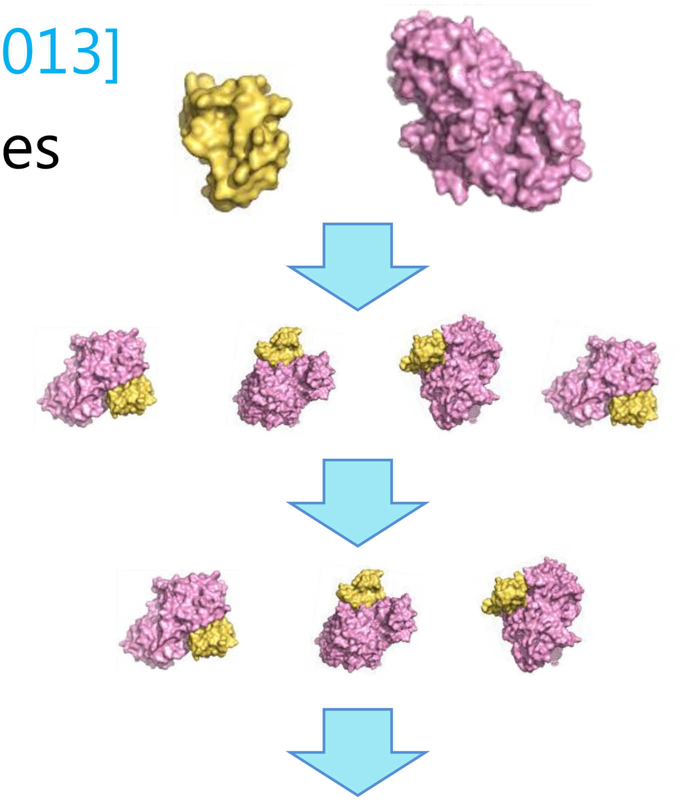
- Pros and Cons

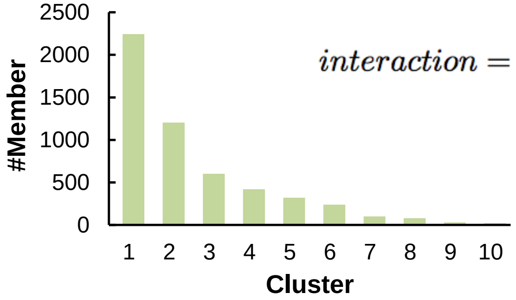
  ☺ Wide applicable range (not require template complexes)

  ☹ Results contain a lot of false-positives

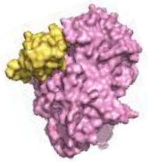
De novo docking method

- Processes of **MEGADOCK**[Ohue+2013]
  1. Generate candidates of complexes with coarse-grained docking
  2. Re-scoring using atom-level energy calculation
  3. Post-processes (clustering, PPI decision)





$$\mathit{interaction} = \begin{cases} \mathit{true} & \text{if } E > E^* \\ \mathit{false} & \text{otherwise} \end{cases}$$

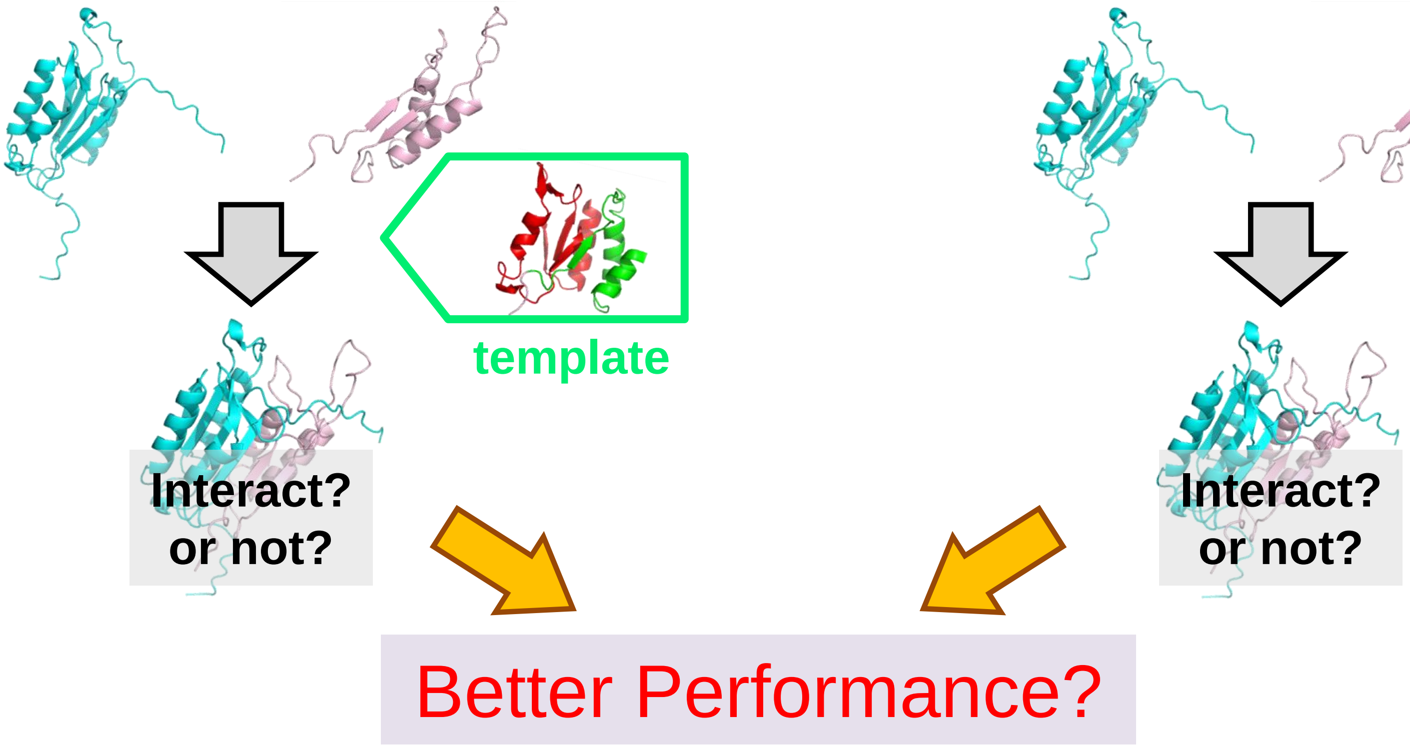


Aim

- **Combine two different PPI prediction methods**

PRISM
(template-based)
Tuncbag N, *et al. Nat Protocol,* 2011.

MEGADOCK
(*de novo* docking)
Ohue M, *et al. Prot Pept Lett,* 2013.



template

Interact? or not?

Interact? or not?

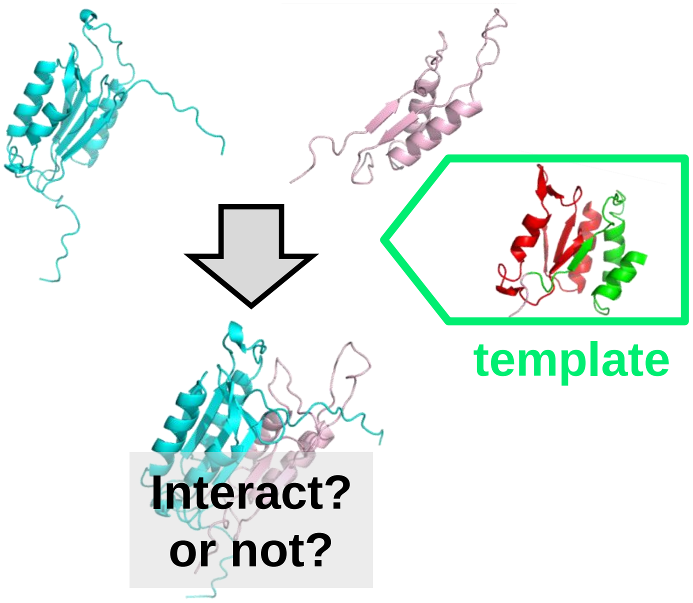
Better Performance?

Materials and Methods

Consensus prediction method

- **Combine two different PPI prediction methods**

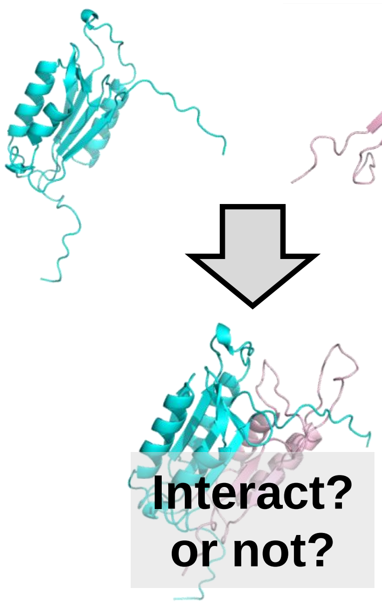
PRISM
(template-based)
Tuncbag N, *et al. Nat Protocol,* 2011.



template

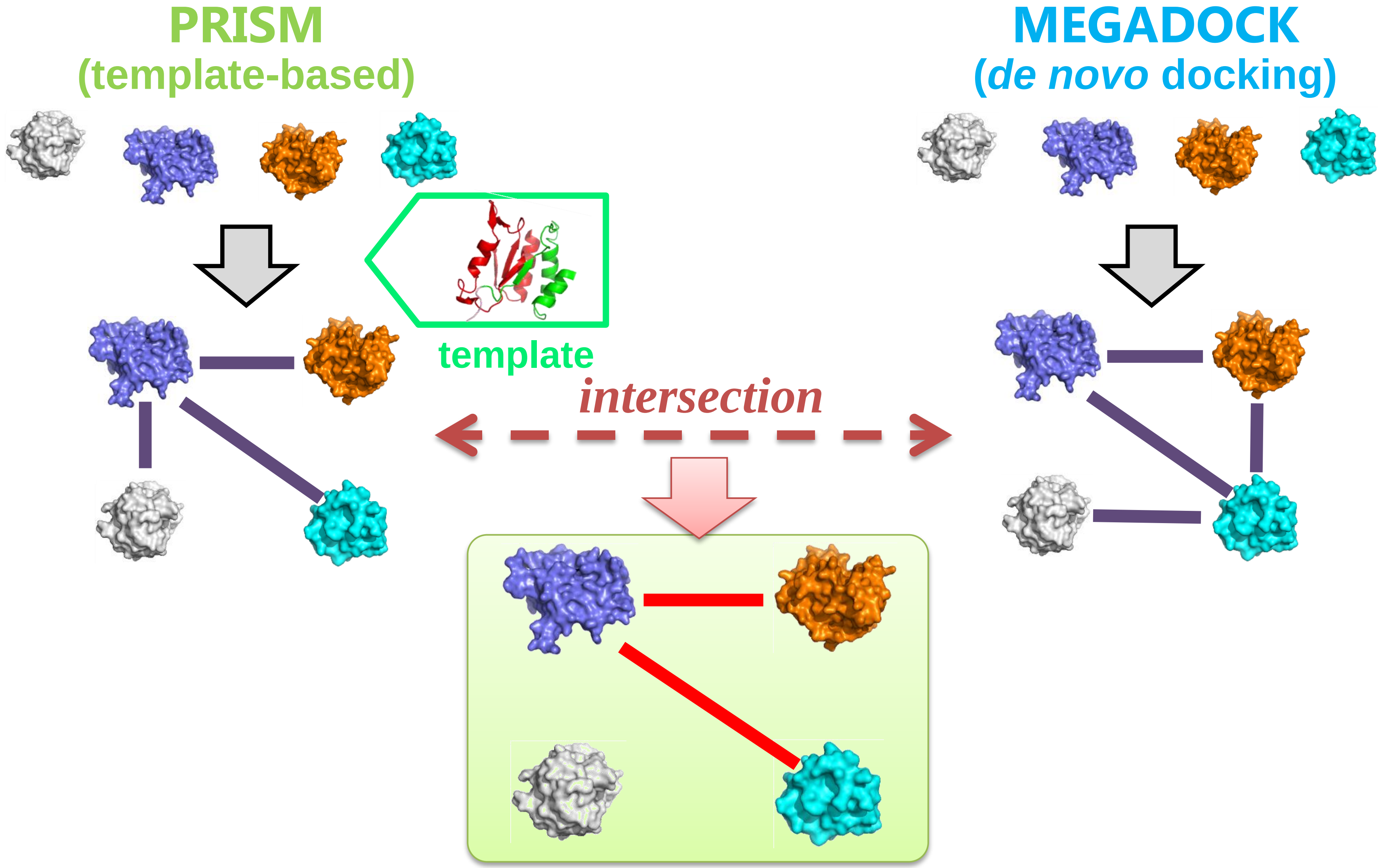
Interact?
or not?

MEGADOCK
(*de novo* docking)
Ohue M, *et al. Prot Pept Lett,* 2013.

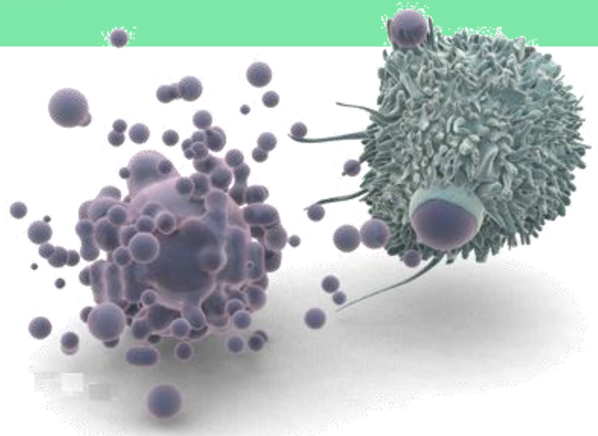


Interact?
or not?

Consensus prediction method



Dataset



- Human apoptosis PPI network
  - What is apoptosis?
    - Controlled cell death
    - Cause of diseases
      - Inhibited apoptosis → cancer, autoimmune diseases, viral infections
      - Enhanced apoptosis → AIDS, ischemic injury, neurodegeneration
  - 57 proteins from KEGG *hsa04210*
    - ${}_{57}C_2$ = 1,596 protein pairs
    - 137 known PPIs (positive samples)
    - 1,459 pairs may not interact (negative samples)

Evaluation of performance

- Definition

		fact	
		known	unknown
prediction	interaction	**TP**	**FP**
	no interaction	**FN**	**TN**

- Evaluation value

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

← **Reliability**

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

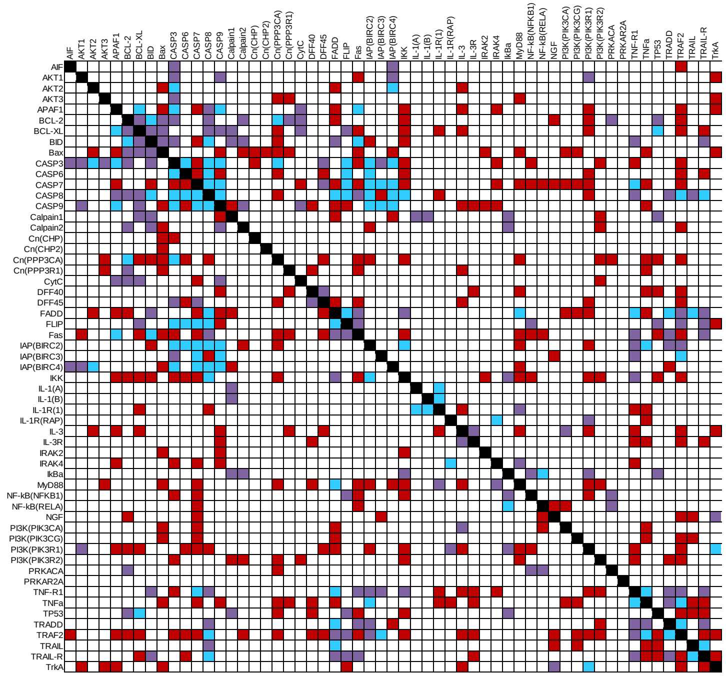
$$\text{F-measure} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}$$

← **Overall accuracy**

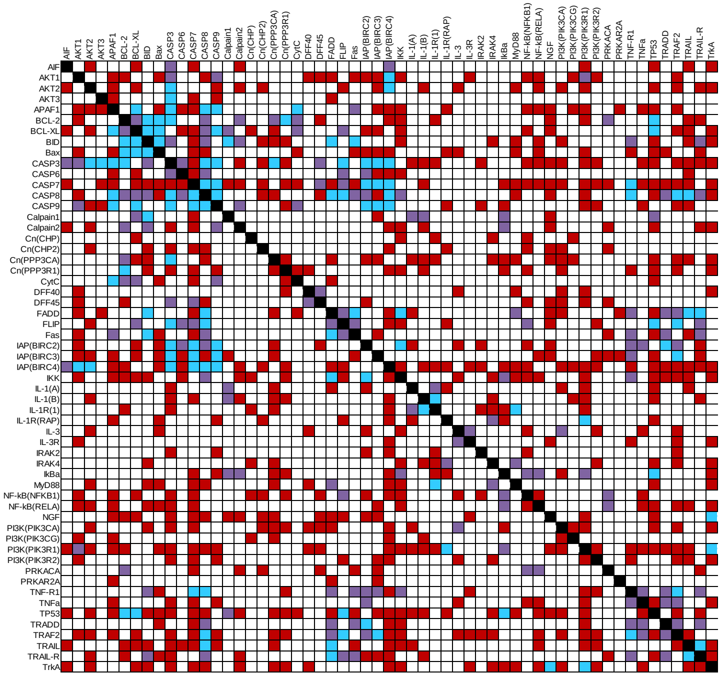
Results and Discussion

Comparison of two methods

PRISM(template)



MEGADOCK(*de novo*)

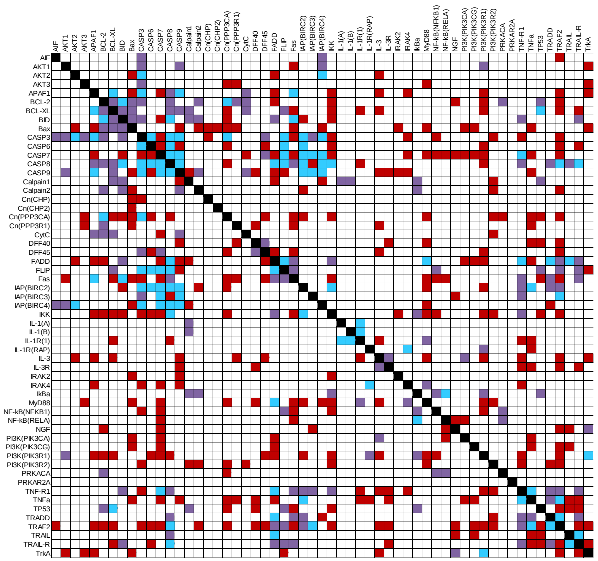


Method	Precision	Recall	F-measure
PRISM(template)	0.231	0.409	0.296
MEGADOCK(*de novo*)	0.145	0.453	0.220

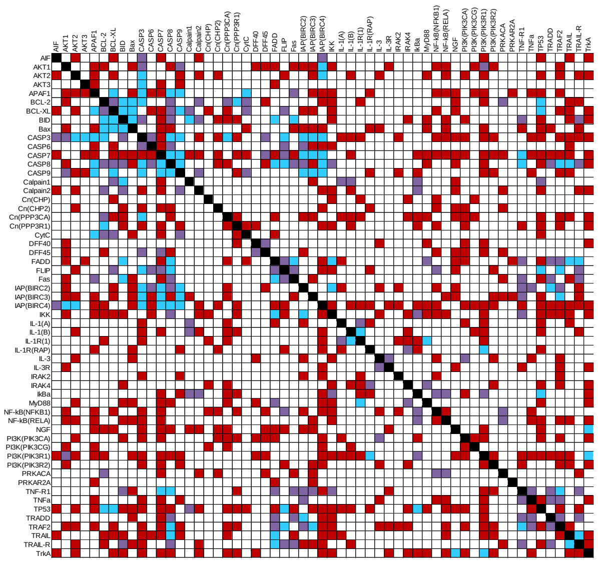
🟦 True Positive
🟥 False Positive
🟪 False Negative

Comparison of three methods

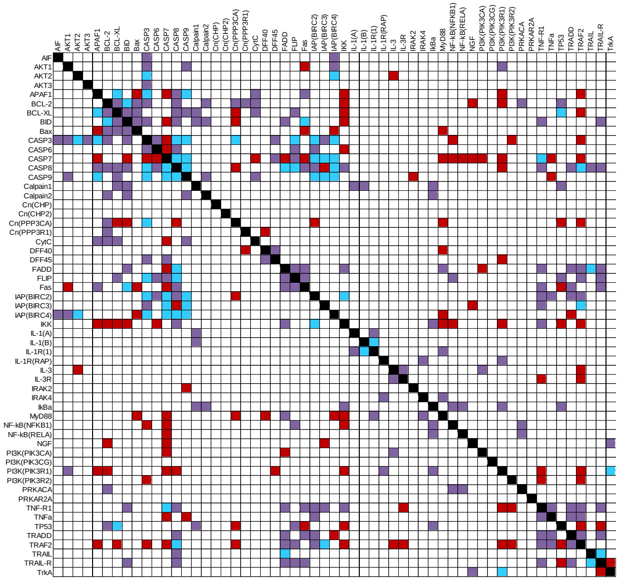
PRISM(template)



MEGADOCK(*de novo*)



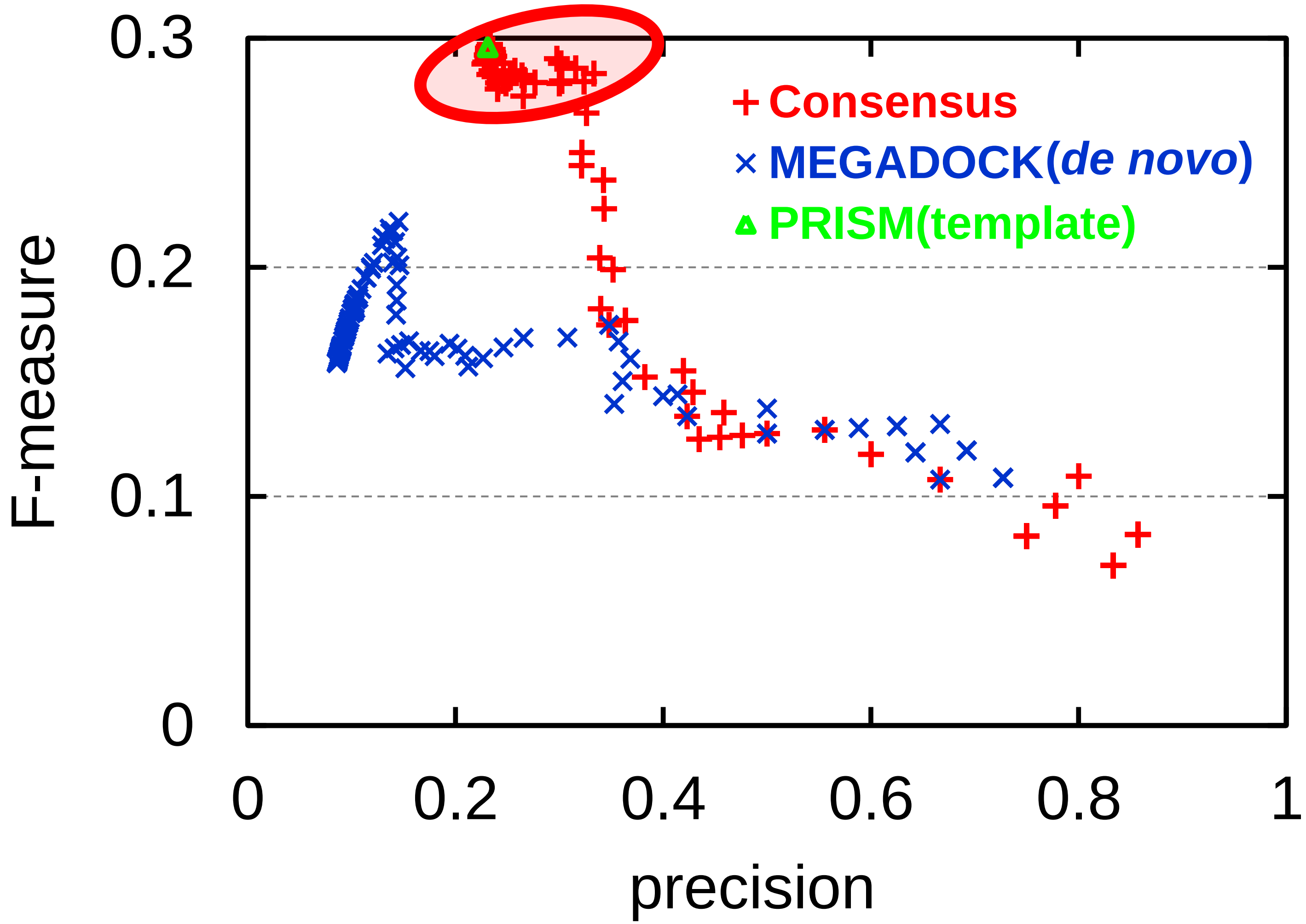
Consensus



Method	Precision	Recall	F-measure
PRISM(template)	0.231	0.409	0.296
MEGADOCK(*de novo*)	0.145	0.453	0.220
Consensus	**0.333**	0.248	0.285

🟦 True Positive
🟥 False Positive
🟪 False Negative

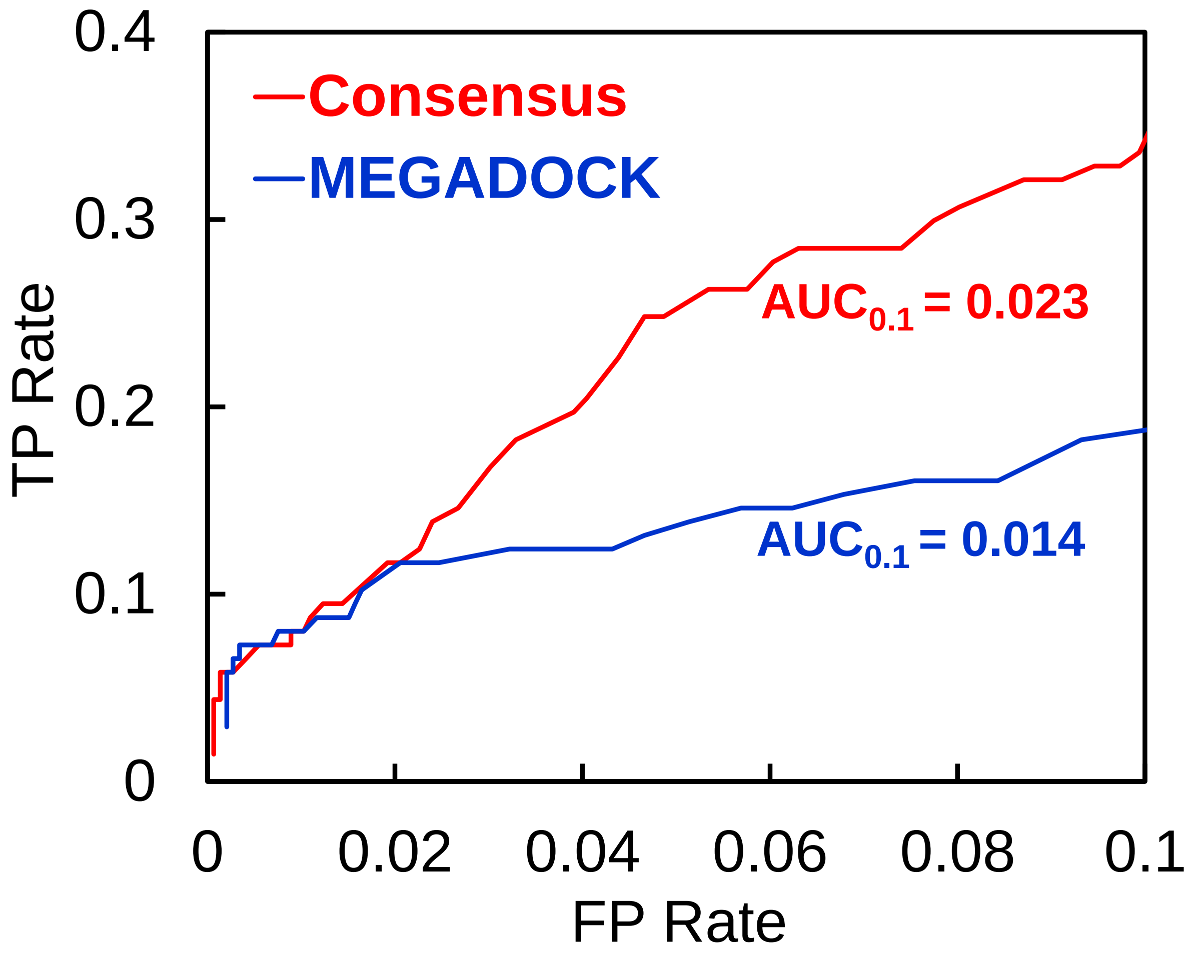
Varied sensitivity parameter



Better precision maintaining the same F-measure level

ROC$_{0.1}$ curve

Larger AUC$_{0.1}$ (Area Under the ROC$_{0.1}$ curve) value is better.



Comparison of results

- F-measure (overall accuracy)
  - **Consensus** ≒ PRISM(template) > MEGADOCK(*de novo*)
- Precision (reliability)
  - **Consensus** > PRISM(template) > MEGADOCK(*de novo*)

- **Consensus method has high reliability**
  - More accurately than the conventional single method
  - Consensus must be useful for target protein screenings

Concluding Remarks

Conclusions

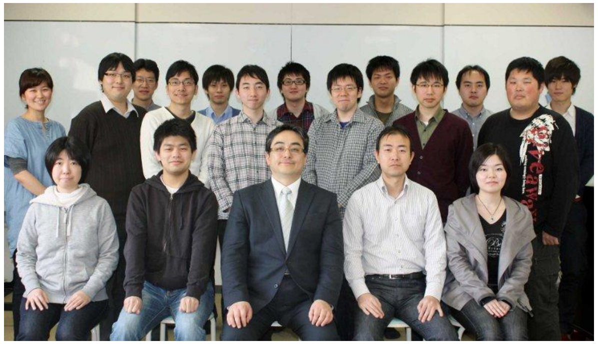
- **We proposed a new PPI prediction method**
  - Consensus between template-based and *de novo* docking
  - More accurate prediction method
  - Our method can reduce biological screening costs
  - *In recent years, template-based docking grab attention, but possibility of template-free docking should be considering.*

- Future work
  - Improve the way of combining the two techniques by using biological information
    - Biochemical function
    - Subcellular localization information

Acknowledgments

- This work was supported in part by
  - ✓ a Grant-in-Aid for JSPS Fellows 238750,
  - ✓ a Grant-in-Aid for Research and Development of The Next-Generation Integrated Life Simulation Software (ISLiM),
  - ✓ the Education Academy of Computational Life Sciences (ACLS),

  all of which were from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).
- Parts of the results were obtained by the K-computer at the RIKEN Advanced Institute for Computational Science (AICS).
  (Early access and HPCI research program hp120131)



Tokyo Tech, Akiyama Lab.

ACLS

ISLiM

RIKEN AICS, K-computer