# Conference Summary & Paper Report

Xudong Zou

17th Oct. 2014

# Outline

- National Conference on Bioinformatics and Systems biology


- Nonnatural protein-protein interaction-pair design

Nonnatural protein-protein interaction-pair design by key residues grafting

S. Liu, *et al*. *Proc. Natl. Acad. Sci. U.S.A*. 104, 5330(2007).


*Time*: 6th ~ 9th October, 2014 *Address*: NanJing

特邀报告（10 speaches）

郝柏林：
古菌分类
刘晓乐：
Genome-wide CRISPR-Cas9 Knockout screens
徐鹰：
Cancer Mechanisms through Data Mining

欧阳颀：
能量耗散与震荡系统相位稳定性的关系研究
张泽民：
Splicing Abnormalities in Tumors and Cancer Cell Lines
黄海艳：
Statistics for coexpression measures
钟声：
Cell fate studies by RNA-seq

陈润生：
长非编码RNA的系统发现及其变异检测
李浩：
System Biology of Cellular Aging
韩敬东：
Alu Repetitive Elements as Proto-Enhancers

主题报告(30)&专题报告(30)

# 金属离子调控的蛋白质折叠和功能运动

王炜-南京大学-物理学院
wangwei@nju.edu.cn

1. $Zn^{2+}$ 调控的锌指蛋白的折叠

2. $Ca^{2+}$ 结合调控的钙调蛋白局域构象运动(all-atom simulation); $Ca^{2+}$ 结合调控的钙调蛋白的别构运动

3. 钙离子调控Gelsolin蛋白折叠与功能运动

Inferring gene interactions beyond standard models using expression data

黄海艳-UC Berkeley

1. Gene coexpression measures in large heterogenous samples using count statistics
   Wang R. et al. PNAS, 2014

2. Inferring gene interactions and functional modules using sparse canonical correlation analysis(SCCA)
   Wang R. et al. AOAS

When pooling samples from different cell types, time points or other experimental conditions, a desirable coexpression measure should

- be versatile to measure a variety of functional / nonfunctional bivariate associations (e.g., *Renyi, Hoeffding's D, dCov, MI, MIC*)
- account for the fact that the associations may change or only exist across a subset of samples (e.g., *MIC (maybe), biclustering*)
- handle outliers (e.g., *rank-based methods*; MI empirically observed to be sensitive)
- fast and easy to compute (MIC slowest; Renyi, dCov, MI and MIC need permutation tests for p-values)
- lead to functionally related pairs being ranked before spuriously related ones (issue of high false positive rates)

$\Rightarrow$ match local patterns of gene expression ranks

## Detecting bivariate associations

- Data: $p$ genes with expression levels across $n$ samples
- $n$ samples from different experimental conditions, time points, cell types, etc.
- Linear relationships: Pearson's correlation
  - most widely used, fast and straightforward to compute; sensitive to outliers, high false positive rates
- Monotonic relationships: Spearman, Kendall's correlation
- General statistical dependence: Renyi correlation, Hoeffding's D, distance covariance (dCov), mutual information (MI), maximal information coefficient (MIC)

## Local interaction measure

- For a heterogeneous set of samples with potentially changing gene interactions, we can define a general coexpression measure by aggregating the interactions across all subsamples of size $k \leq n$.
- Consider expression vectors $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, define

$$W = \sum_{1 \leq i_1 < \cdots < i_k \leq n} F(x_{i_1}, \ldots, x_{i_k}; y_{i_1}, \ldots, y_{i_k}),$$

where $F(\cdot, \cdot)$ is an interaction measure on local expression profiles $(x_{i_1}, \ldots, x_{i_k})$ and $(y_{i_1}, \ldots, y_{i_k})$ from a set of $k$ samples.

- Choose $F(\cdot, \cdot)$ to be an indicator function comparing the rank patterns of the subsequences

## Matching contiguous subsequences of length $k$ ($W_1$)

- For time-course data, we preserve the order of the samples and consider only interactions within contiguous subsequences.
- Let $\psi$ be the rank function that returns the indices of elements in a vector after they have been sorted in an increasing order.
- $I_i^+$, $I_i^-$ indicators

$$I_i^+ = \begin{cases} 1 & \text{if } \psi(x_i, \ldots, x_{i+k-1}) = \psi(y_i, \ldots, y_{i+k-1}), \\ 0 & \text{otherwise;} \end{cases}$$

$$I_i^- = \begin{cases} 1 & \text{if } \psi(x_i, \ldots, x_{i+k-1}) = \psi(-y_i, \ldots, -y_{i+k-1}), \\ 0 & \text{otherwise;} \end{cases}$$

- Define $W_1 = \sum_{i=1}^{n-k+1} (I_i^+ + I_i^-)$.
- Running time $O(k(\log k)n)$

| | | | | | |
|---|---|---|---|---|---|
| $x$ : | 1 | 3 | 4 | 2 | 5 |
| $y$ : | 1 | 4 | 5 | 2 | 3 |

$W_1 = 2$

# Designing Functional Proteins

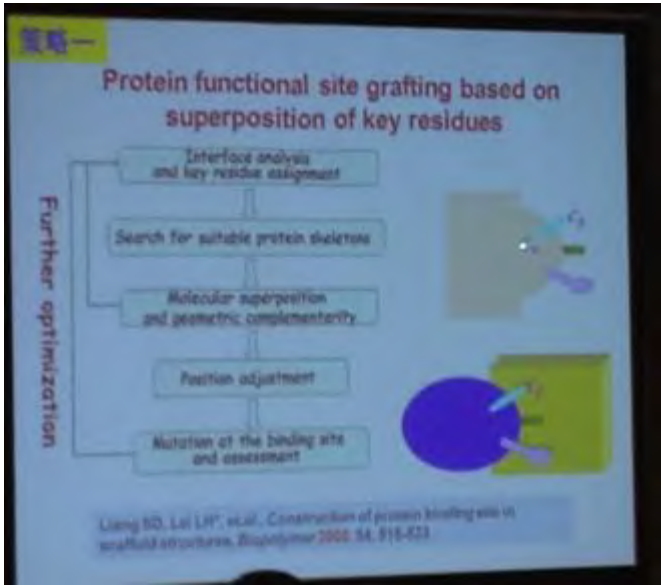
来鲁华-北京大学-化学与分子工程学院
lhlai@pku.edu.cn

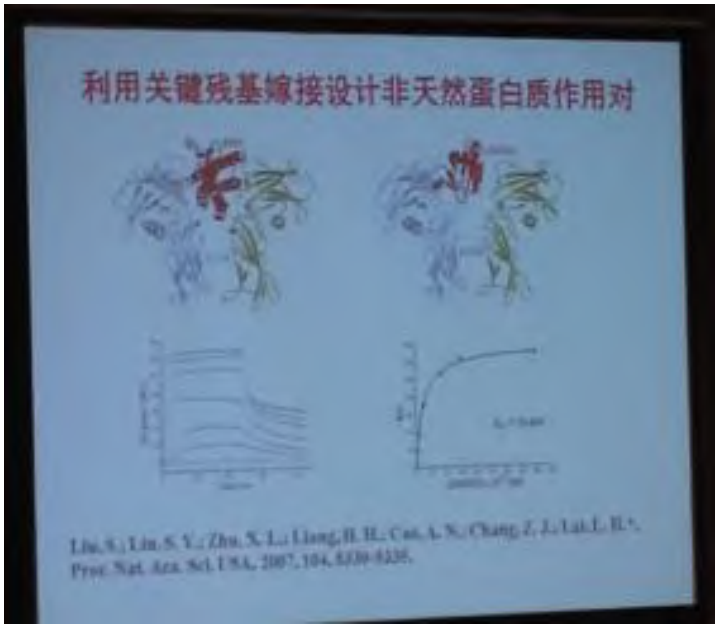
**Three different strategies:**

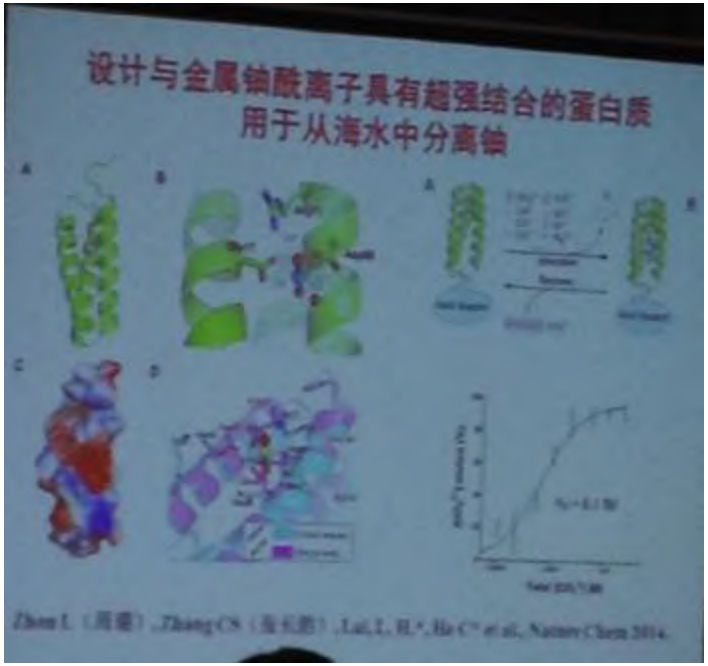
1. Key functional site grafting

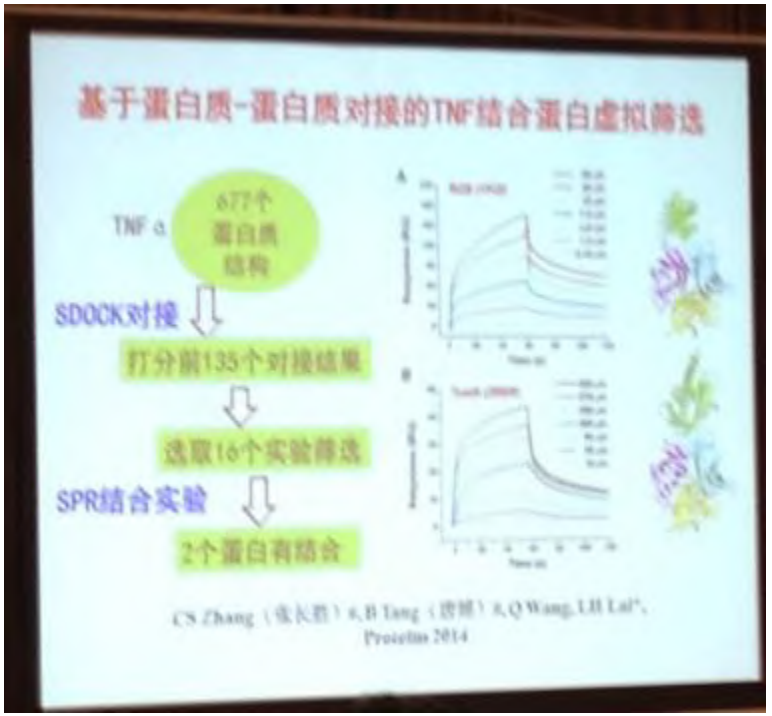
2. *De novo* binding protein design

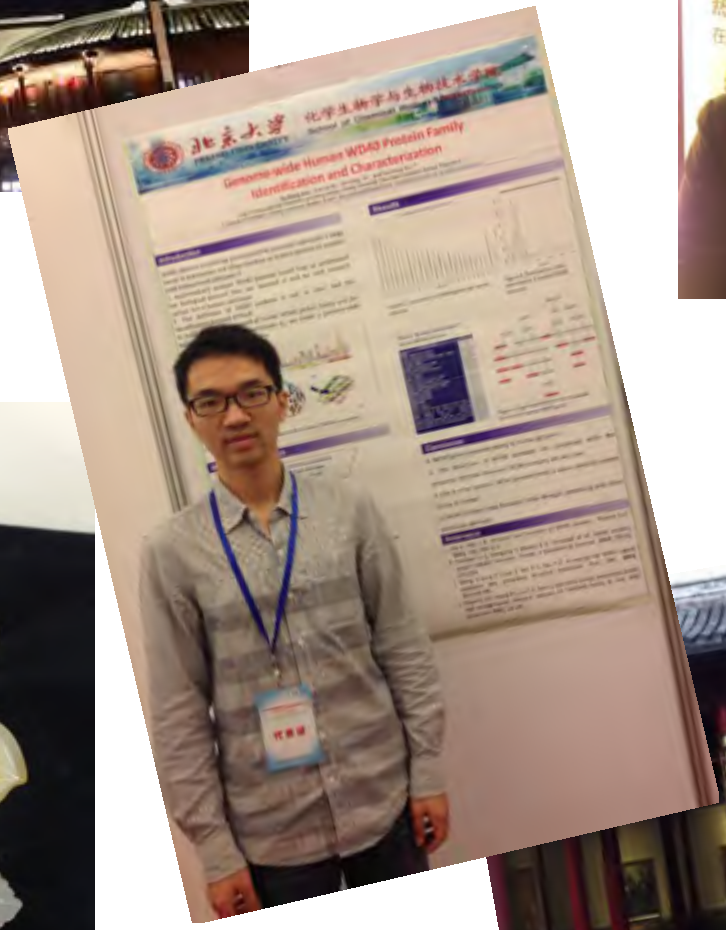
3. Binding protein virtual screening based on protein-protien docking

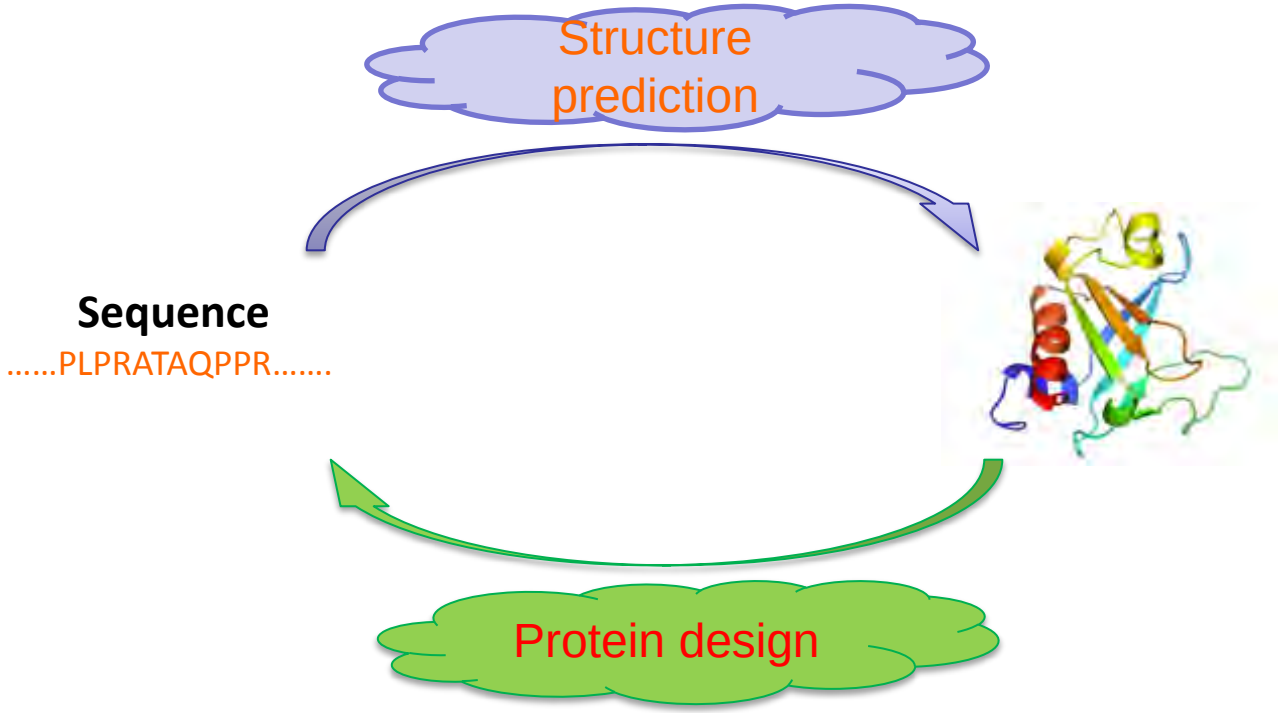

策略一
Protein functional site grafting based on superposition of key residues
Interface analysis and key residue assignment
Search for suitable protein skeletons
Molecular superposition and geometric complementarity
Position adjustment
Mutation at the binding site and assessment
Further optimization


利用关键残基嫁接设计非天然蛋白质作用对


设计与金属铀酰离子具有超强结合的蛋白质
用于从海水中分离铀


基于蛋白质-蛋白质对接的TNF结合蛋白虚拟筛选
TNF α
677个蛋白质结构
SDOCK对接
打分前135个对接结果
选取16个实验筛选
SPR结合实验
2个蛋白有结合

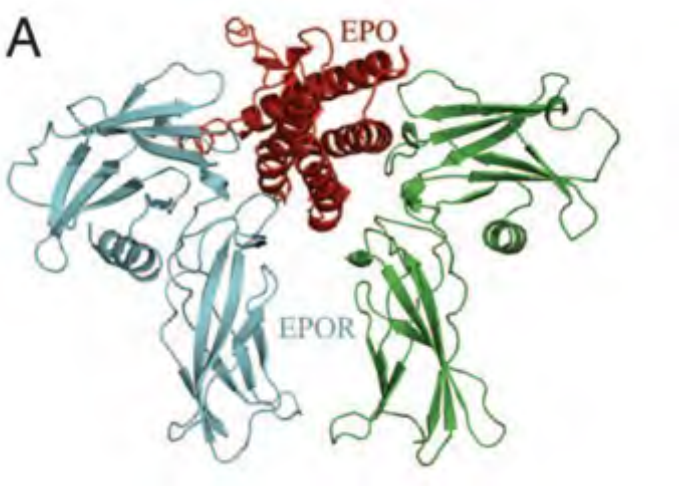

Genome-Wide Human WD40 Protein Family
Identification and Characterization


Structure prediction
Sequence
……PLPRATAQPPR…….
Protein design

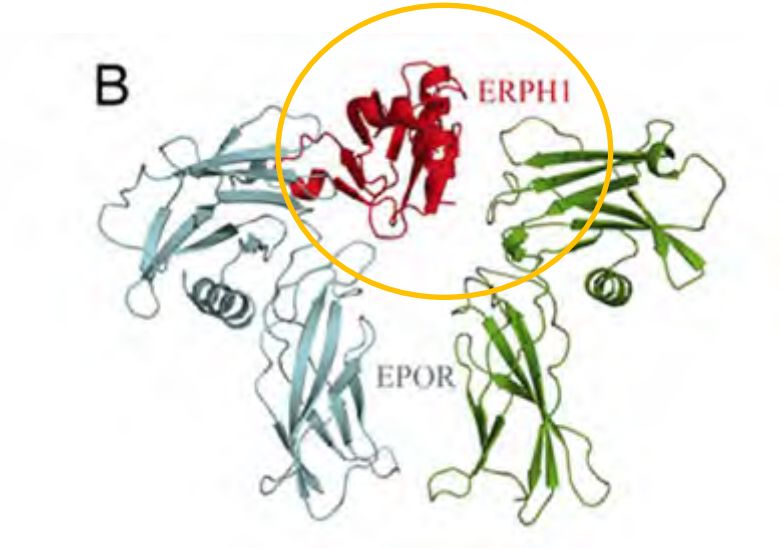
# Nonnatural protein-protein Interaction-pair Design

## Main idea

Designed a unrelated protein that can bind with EPOR.




EPO(红细胞生成素) is the crucial regulator of red blood-cell production and delivers essential growth, differentiation and survival signals to erythroid progenitors.




ERPH1 is one of the designed proteins that mutated from a rat protein PLCδ1-PH.



Richmond TD, *et al*. *Trends Cell Biol*, 2005, 15:146-155.

# Nonnatural protein-protein Interaction-pair Design

## Methods

Key residues grafting

Transfering functional epitopes from one protein to another.

## Pipeline

Select key residues of EPO-EPOR complex

↓

Searching for scaffold proteins

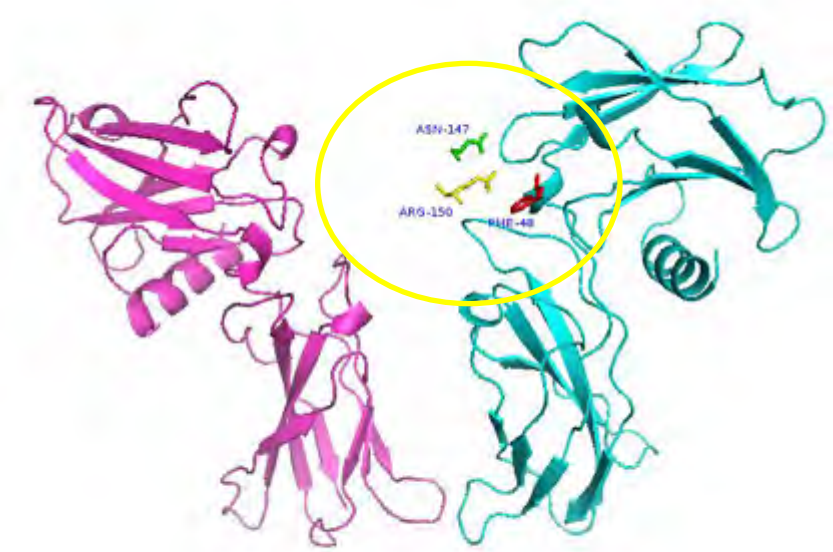
↓

Design the protein

↓

Validation the interactions

Three key residues in EPO, Phe-48, Asn-147, and Arg-150 were selected and used in next step.






Liang S, et al. *J Biomol Struct Dyn*, 2000, 17:821-828.

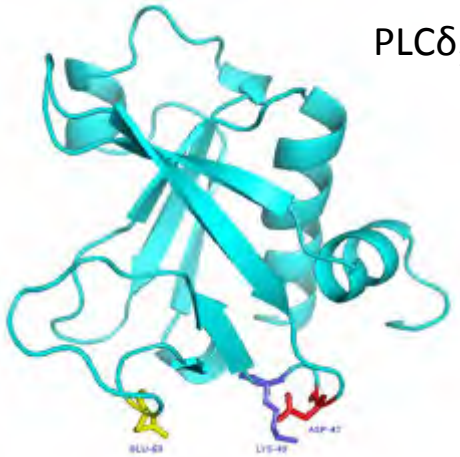
# Nonnatural protein-protein Interaction-pair Design

## Searching for scaffold proteins

1. Choosing proteins with 100-200 residues in PDB as candidate scaffold proteins.
2. Defining three key residues on scaffold proteins
   Three residues should satisfy the geometric relationships of $C_\alpha$-$C_\beta$ vectors of the three key residues in EPO.
3. Defining the interface
   A set of atoms on a protomer that loses at least 0.1 $Å^2$ of ASA upon binding to a partner and has a ASA of < 15$Å^2$.
4. Calculating buried surface area and filtering according to buried surface area
   Buried surface area of the computed protein-protein complex >= 1200$Å^2$.

**Other rules:**

a) Expression in *E. coli*.
b) Without metal ions
c) Have less than four cysteine residues or two pairs disulfide bonds
d) Human origin or highly homologous



PLC$\delta_1$-PH

# Nonnatural protein-protein Interaction-pair Design

## Design the protein

Mutate the three key residues in PLCδ1-PH

| | |
|---|---|
| ERPH1 | E63F, D47N, K49R |
| ERPH2 | E63F, D47N, K49R, E46A |
| ERPH3 | E63F |
| ERPH4 | E63F, D47N |
| ERPH5 | E63F, K49R |
| ERPH6 | D47N, K49R |
| ERPH7 | D47N |
| ERPH8 | K49R |

# Nonnatural protein-protein Interaction-pair Design

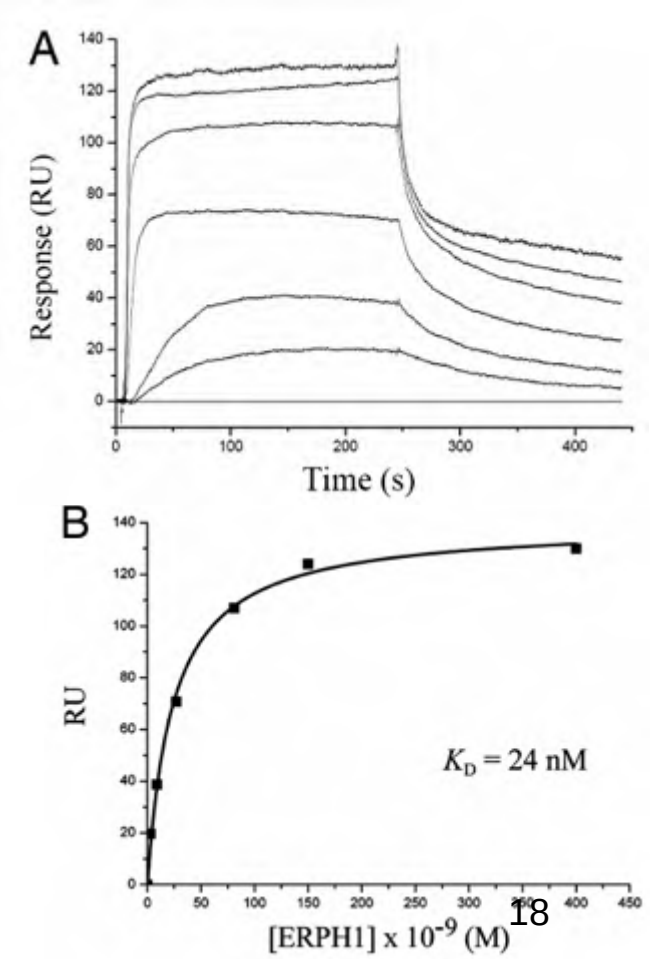
## Validation the interactions

The rank of binding free energies is consistent with that of binding constants.

**Table 1. Equilibrium constants ($K_D$) of hEPO and ERPHs binding to hEPOsR as determined by SPR**

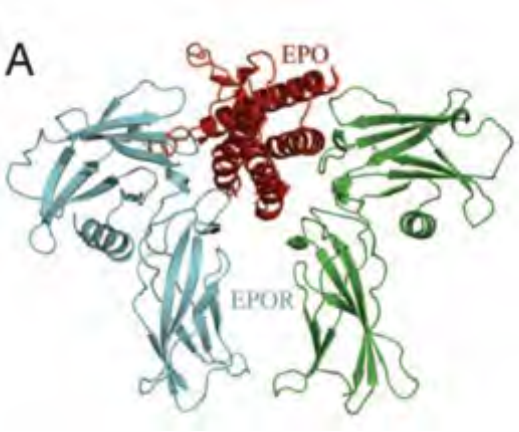
| Name | Mutation(s) | $K_D$, nM | ΔG,* kcal/mol |
|---|---|---|---|
| hEPO | — | 0.13 ± 0.05 | −21.08 |
| PLC$\delta_1$-PH | — | ND | −11.63 |
| ERPH1[a] | E63F, D47N, K49R | 24 ± 3 | −13.42 |
| ERPH2 | E63F, D47N, K49R, E46A | 26 ± 3 | −13.26 |
| ERPH3 | E63F | 240 ± 60 | −12.96 |
| ERPH4 | E63F, D47N | 29 ± 2 | −13.54 |
| ERPH5 | E63F, K49R | 69 ± 1 | −12.91 |
| ERPH6 | D47N, K49R | 500 ± 270 | −12.10 |
| ERPH7 | D47N | ND | −12.21 |
| ERPH8 | K49R | ND | −11.59 |

ND, could not be determined under the experimental conditions because of very weak or undetectable binding.
*The binding free energies calculated by using PMFScore (24).



# Nonnatural protein-protein Interaction-pair Design

## Validation the interactions

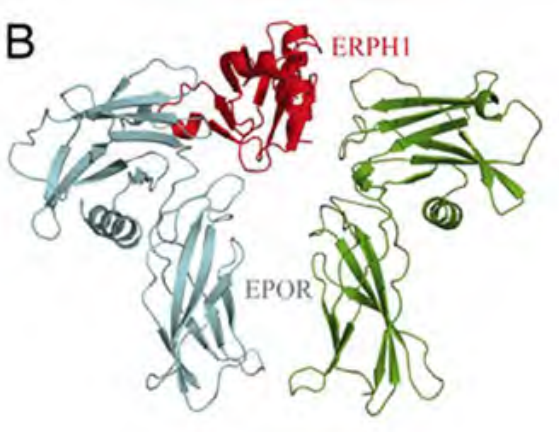


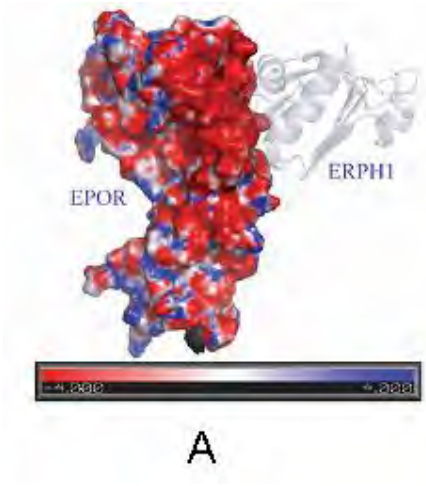

Buried surface area

| | |
|---|---|
| 1946.8Å$^2$ | 1825.5 Å$^2$ |

Buried hydrophobic surface area

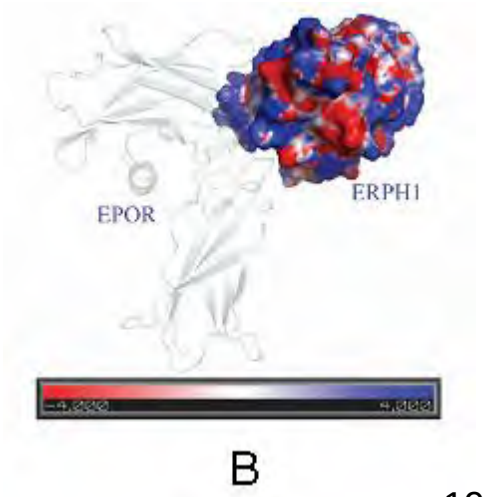
| | |
|---|---|
| 1052.2Å$^2$ | 976.7 Å$^2$ |



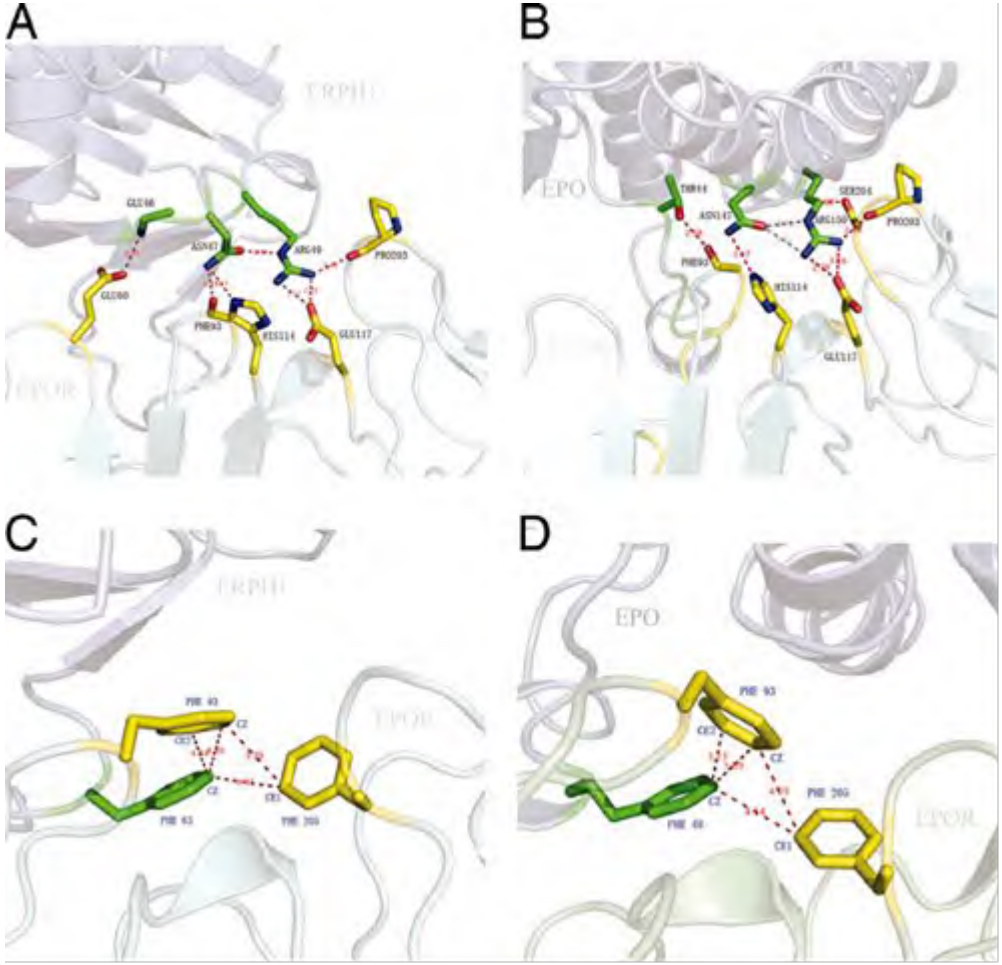

Strong negative electrostatic potential

Strong positive electrostatic potential



# Nonnatural protein-protein Interaction-pair Design

## Validation the interactions



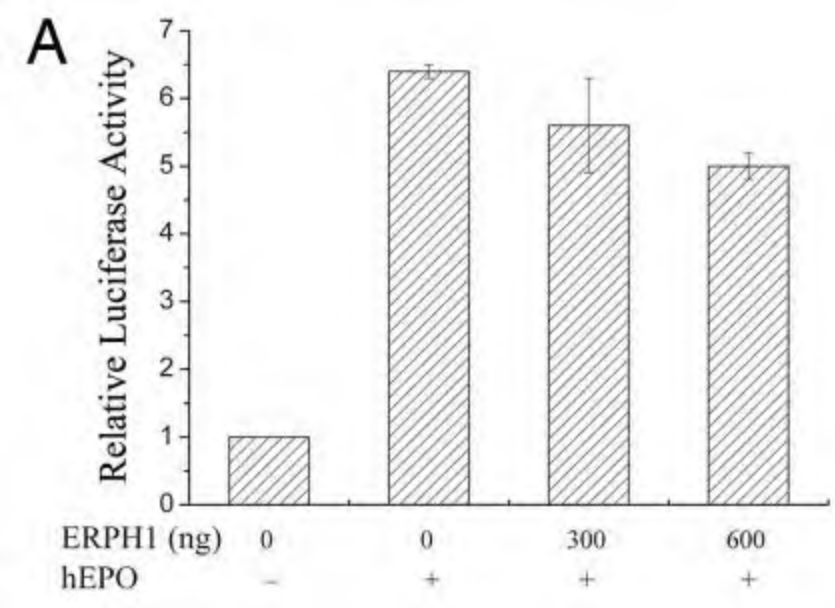

Phe-63 in ERPH1 forms a hydrophobic cluster with Phe-93 and Phe-205 in EPOR

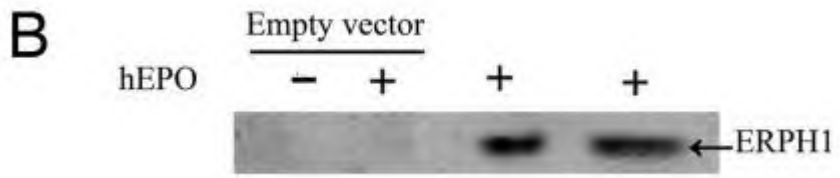
The side chain amide of Asn-47 in ERPH1 forms H-bonds with the side chain of His-114 in EPOR

The side chain of Arg-49 in ERPH1 forms H-bonds with the side chain of Glu-117 and the backbone carbonyl oxygen of Pro-203 in EPOR

# Nonnatural protein-protein Interaction-pair Design

## Validation the interactions



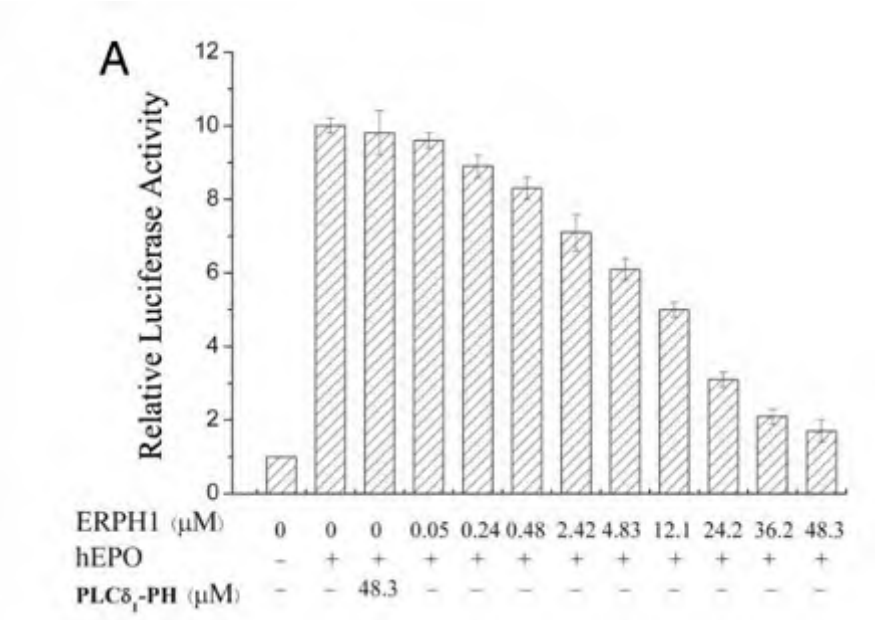

A. Testing the activity of ERPH1 in EPO-EPOR signal cascade, and the JAK2/STAT5 reporter system was used.
B. Expression level of ERPH1 detected by western blot

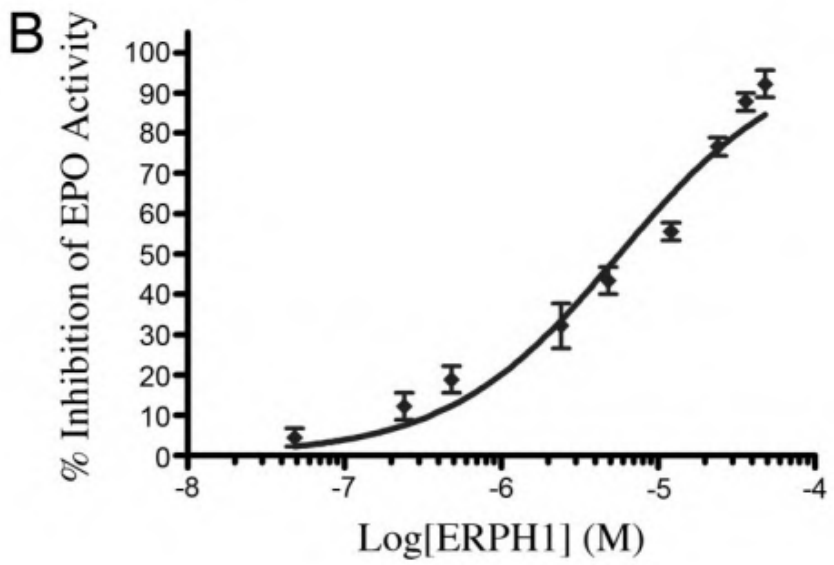
In Vivo ERPH1 activity measurement:

The activity of hEPO decreased while the expression of ERPH1 increasing

# Nonnatural protein-protein Interaction-pair Design

## Validation the interactions




- Luciferase activity decreased sharply as the concentration of ERPH1 increased from 0.05 to 48.3 μM
- The $IC_{50}$ of ERPH1 on EPO activity was 5.7 μM

# Nonnatural protein-protein Interaction-pair Design

## Summary

- This work is based on that similar functions can be performed by different proteins.
- PLCδ1-PH is ideal for grafting the three key residues from EPO, which tolerated the mutations and maintained its overall structure.
- Why only a few mutations on PLCδ1-PH made it bind to its nonnatural partner need further studies.
- It is very important to choose a better scaffold protein for designing.

***Thanks for your attention!***