

# OMICS Journals are welcoming Submissions

OMICS International welcomes submissions that are original and technically so as to serve both the developing world and developed countries in the best possible way.

OMICS Journals are poised in excellence by publishing high quality research. OMICS International follows an Editorial Manager® System peer review process and boasts of a strong and active editorial board.

Editors and reviewers are experts in their field and provide anonymous, unbiased and detailed reviews of all submissions. The journal gives the options of multiple language translations for all the articles and all archived articles are available in HTML, XML, PDF and audio formats. Also, all the published articles are archived in repositories and indexing services like DOAJ, CAS, Google Scholar, Scientific Commons, Index Copernicus, EBSCO, HINARI and GALE.

**For more details please visit our website:**  
<http://omicsonline.org/Submitmanuscript.php>

# Protein-Protein Complex Structure Prediction

**Dr Jean-Christophe Nebel**  
Associate Professor

Leader of the Bioinformatics & Genomic Signal Processing Research Group  
<http://staffnet.kingston.ac.uk/~ku33185/Bioinformatics.html>  
Faculty of Science, Engineering & Computing

[J.Nebel@Kingston.ac.uk](mailto:J.Nebel@Kingston.ac.uk)

**Kingston University London**

# Contents

## **Protein-Protein Interactions**

## **Protein Interface Prediction**

Introduction

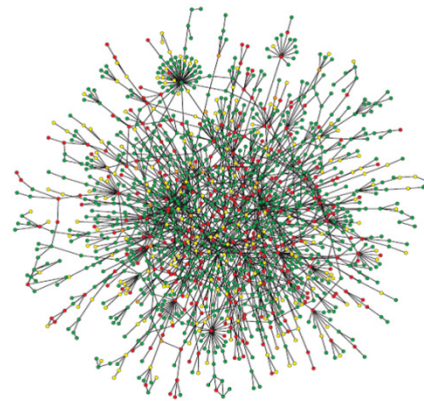
Interface Conservation & Ligand Diversity

## **Protein-Protein Complex Structure Prediction**

## **T-PioDock Software**

## **Current Research Interests**

# Protein-Protein Interactions



# Protein-Protein Interactions

**Essential** to most processes that take place within a living cell  
e.g. main signalling pathways activated by insulin\*



### Abbreviations:

IRS: Insulin receptor substrate.

SHC: Src homology 2-containing protein.

Grb2: Growth factor receptor-bound protein 2.

SOS: Son of Sevenless.

Ras: A small GTPase.

RAF: MAP kinase kinase kinase.

MEK: MAP kinase/ERK kinase, MAP kinase kinase.

ERK: Extracellular signal-regulated kinase.

P90 RSK: Ribosomal Protein S6 kinase.

PI3K: Phosphatidylinositol 3- kinase.

PIP2: Phosphatidylinositol 3,4 bisphosphate.

PIP3: Phosphatidylinositol 3,4,5 trisphosphate.

PDK: 3-phosphoinositide - dependent protein kinase.

Akt: Protein kinase B (PKB).

FOXO: Forkhead box O.

mTOR: Mammalian target of rapamycin.

GLUT4. Glucose transporter 4.

PTP1B: Protein tyrosine phosphatase 1B.

PTEN: Phosphatase and tensin homologue deleted on chromosome 10.

GSK3: Glycogen synthase kinase-3.

\*Insulin and IGF-1 receptor signalling pathways: where is the specificity? by Pierre De Meyts

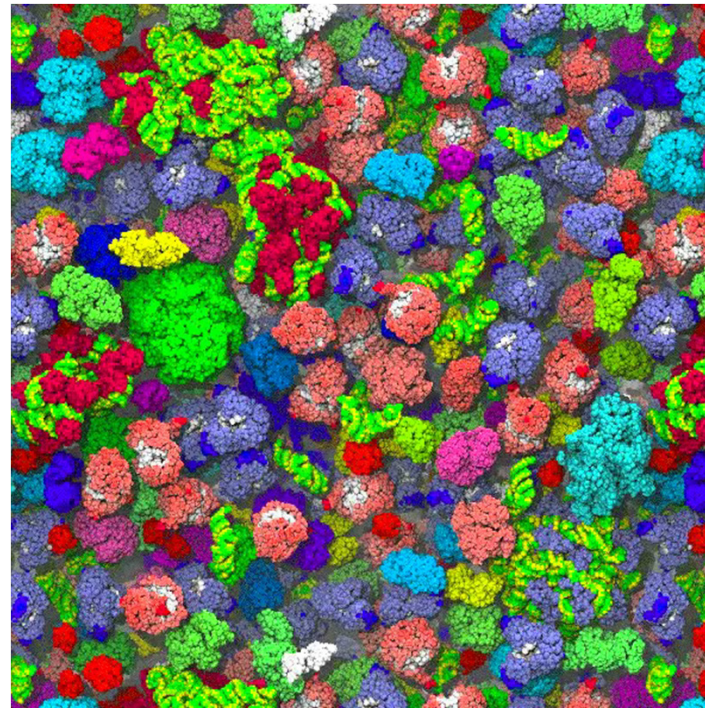
<http://www.fefchemicals.com/biopharm/scientific-information/articles/insulin-and-igf-1-receptor-signalling-pathways-where-is-the-specificity>

# Protein-Protein Interactions

The interior of cells is **crowded**

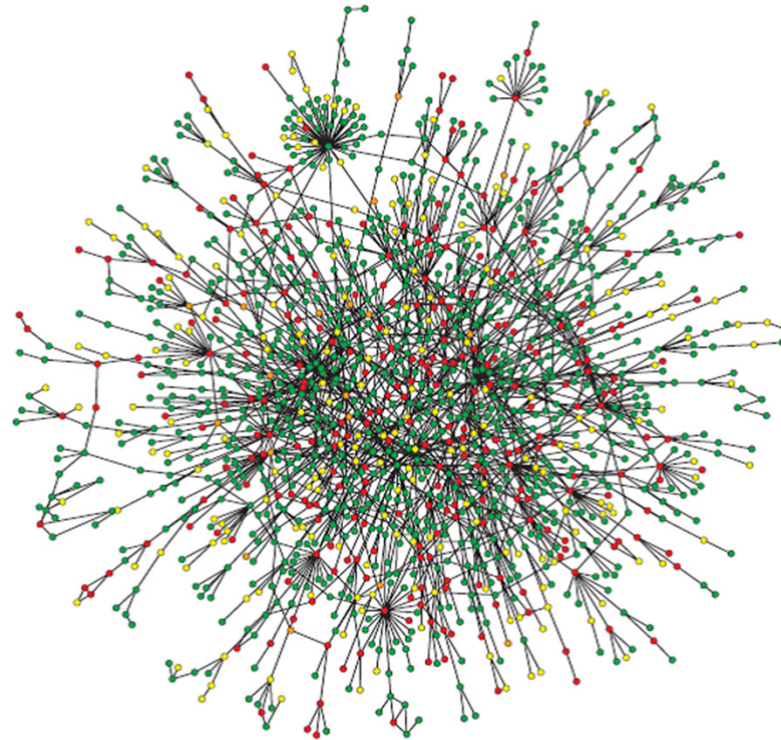
e.g. dynamic molecular model of the bacterial cytoplasm

© 2005 BioStudio Visual Communications, Inc. <http://www.cellimagelibrary.org/images/28234>



# Protein-Protein Interactions

Most proteins are involved in **many interactions**  
e.g. the yeast interactome (2007)

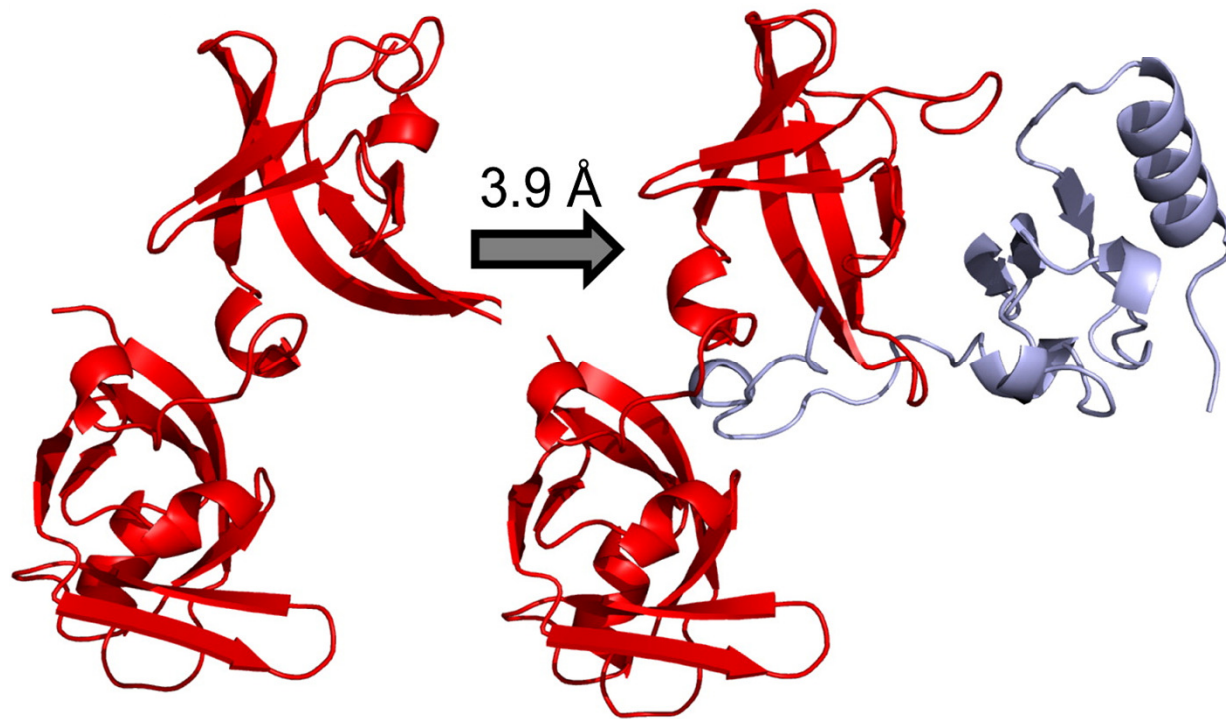


# Protein-Protein Interactions

**Conformational changes** upon binding are common

e.g. ribosome maturation protein (rimm) (2DYI)

rimm in complex with ribosomal protein S19 (3A1P)





# Protein-Protein Interactions

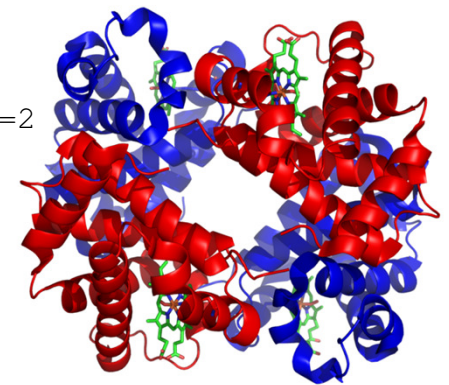
Abnormal interactions may lead to **critical diseases**

e.g. haemoglobin misfolding

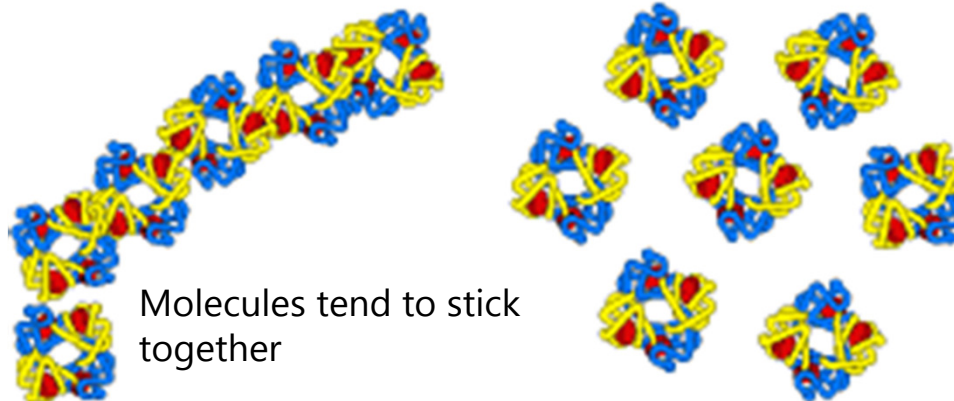
```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2
MVHLTPEEKSAVTALWGKVNVDVEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGKVLGAFSDGLAHLNLDLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH
```



Effect of single genetic mutation:  
from malaria benefit to anaemia



Red blood cells

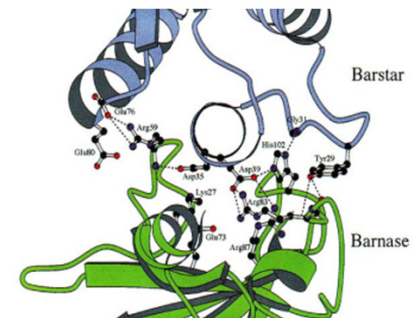
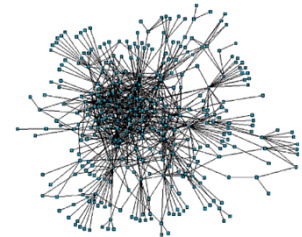


-> Sickle Cell Anaemia ! (life expectancy ~55 in the UK)

# Protein-Protein Interaction Research

## Wet lab techniques

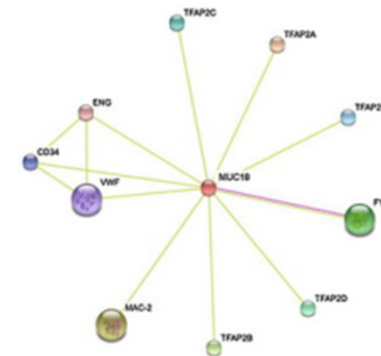
- Discovery of interactions,  
e.g. yeast two-hybrid system (Y2H)
- Mode of interaction revealed by 3D structure of protein complexes (>50% structures in PDB are complexes)
- Identification of interface residue,  
e.g. mutagenesis
- ...



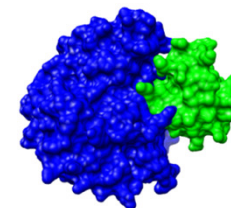
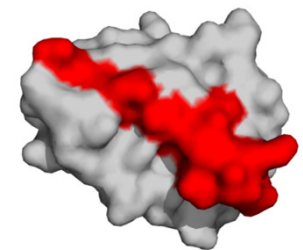
# Protein-Protein Interaction Research

## Bioinformatics techniques

- Prediction of interaction partners
- Interaction network evolution
- Literature mining
- Prediction of interaction sites
- Prediction of a complex structure
  - Protein-protein docking
  - Model ranking



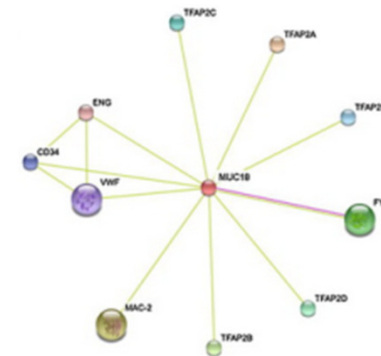
| x | | x | | ... x |



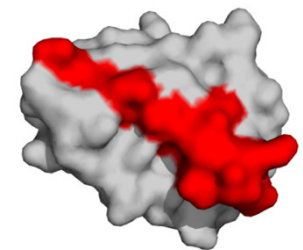
# Protein-Protein Interaction Research

## Bioinformatics techniques

- Prediction of interaction partners
- Interaction network evolution
- Literature mining

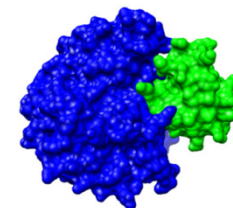


- **Prediction of interaction sites** | x | | x | | ... x |

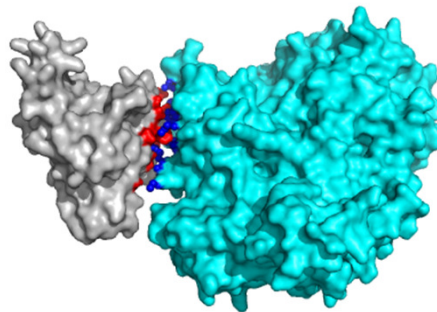


- **Prediction of a complex structure**

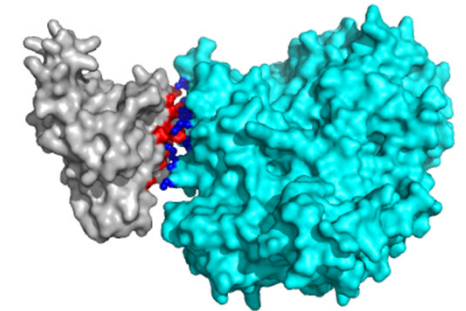
- Protein-protein docking
- **Model ranking**



# Protein Interface Prediction



## Protein Interface Prediction



### Interface residues

**CAPRI** (Critical Assessment of **P**rediction of **I**nteractions) definition

*"all residues of a protein chain that have atoms less than 5 Å apart from the interacting partner"*

### Input data

- Protein sequence
- Protein 3D structure
- Target pair

# Protein Interface Prediction

## Approaches

- **Intrinsic-based Predictors**

- Specific features such as hydrophobicity, interface propensity and solvent accessibility
- Evolutionary conservation information
- 3D Docking

- **Template-based Predictors**

- Homologous models
- Structural Neighbours

# Protein Interface Prediction

## Approaches

- **Intrinsic-based Predictors**

- Specific features such as hydrophobicity, interface propensity and solvent accessibility
- Evolutionary conservation information
- 3D Docking

- **Template-based Predictors**

- Homologous models
- Structural Neighbours

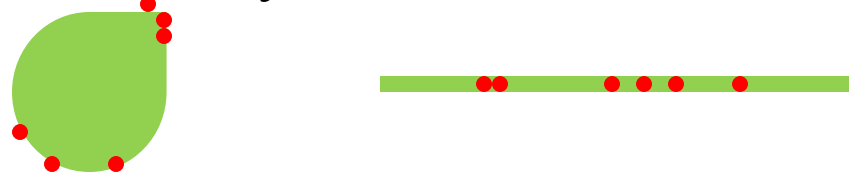


# Protein Interface Prediction

**Exploiting interface  
conservation & ligand diversity**

# Protein Interface Prediction exploiting Interface conservation & ligand diversity

**Goal:** To predict **residues** likely to be involved in interactions



**Homology-based** approach using complex **structures**:

- The more homologous to **the target**, the more informative

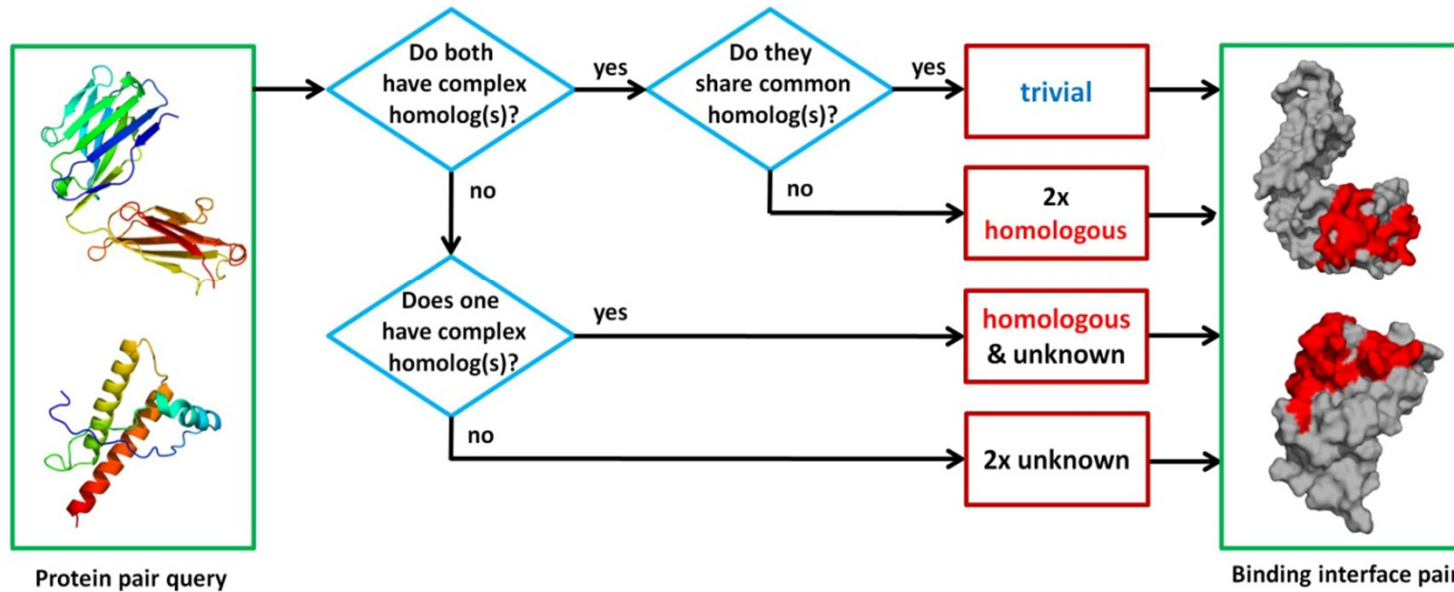
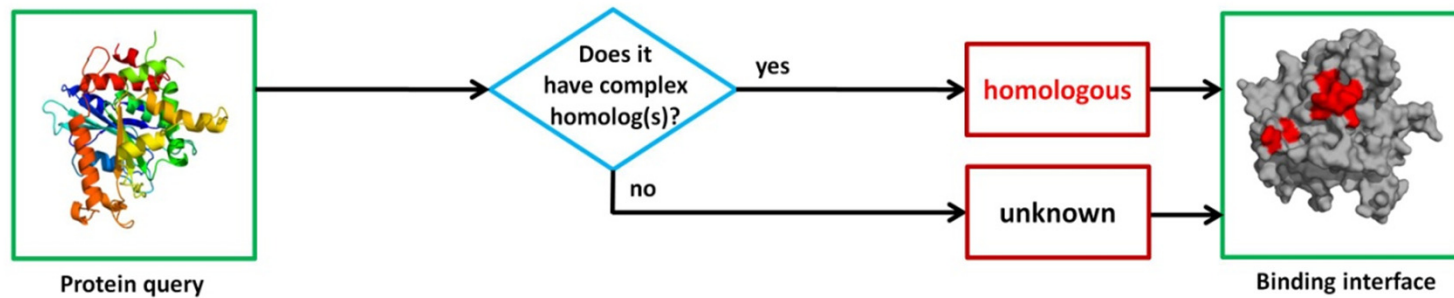


- The more diverse the **ligands**, the more general the interaction patterns



- Processing depending of complexity of target (trivial, homologous or unknown)

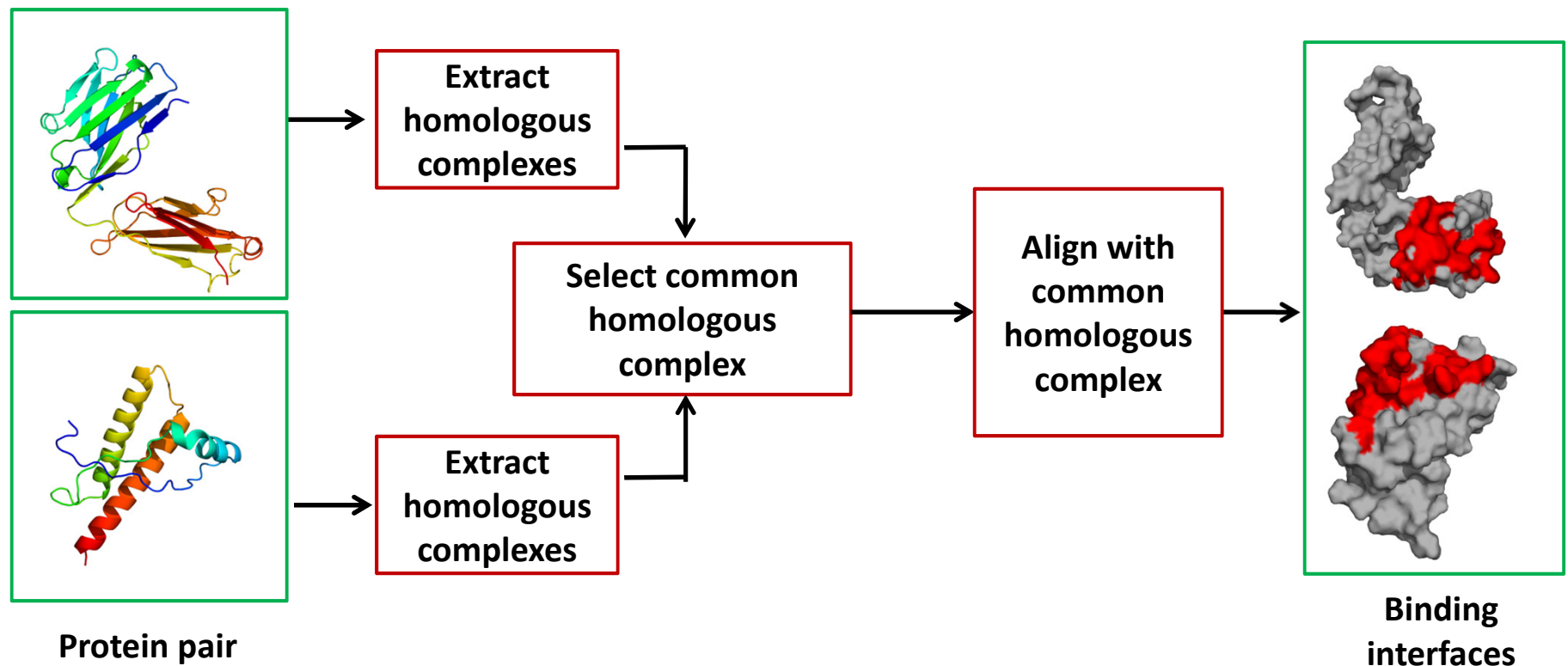
## T-PIP Framework



Unknown category: no homologous complex available -> usage of PredUs

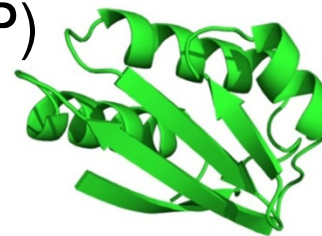
## Trivial category

1. Extract homologous complexes for each protein of the pair
2. Select complex with best combined E-value score
3. Align and map interfaces on query

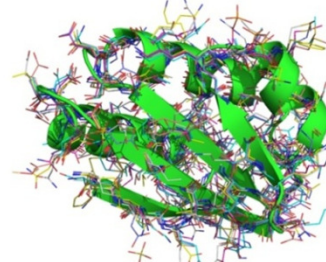


## Homologous category

1. Structurally align query protein (QP) with its structural neighbours



QP  
↓

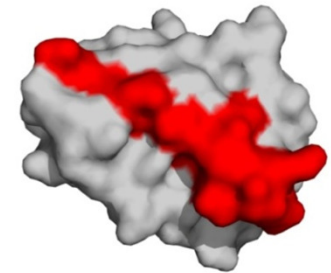


Multiple Structural Alignment  
(QP, P<sub>1</sub>, P<sub>2</sub>, ..., P<sub>n</sub>)

2. Produce Structure based Multiple Sequence Alignment (S-MSA)\*  
X: non-interface, I: interface

3. Rank residues according to their *interaction score*

4. Select the top ***T*** residues as interface



QP Interface  
↑

P <sub>1</sub>	x	x	I	x	x	x	I	...	x	I
P <sub>2</sub>	x	x	I	x	x	I	x	...	x	x
P <sub>3</sub>	I	x	I	x	x	I	x	...	x	x
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P <sub>n-1</sub>	x	x	I	x	x	x	I	...	x	I
P <sub>n</sub>	x	x	I	x	I	x	I	...	x	x
QP	I	x	I	I	x	I	I	...	x	I
QP Interface	I	x	I	I	x	I	I	...	x	I

S-MSA

\*MSA (or partial S-MSA) can be used if QP does not have a known structure

## Interaction estimation

It relies on **3** elements:

1. **Number**,  $N$ , of homologous proteins suggesting interaction
2. **Query weight**: The degree of homology between the QP sequence and homologous protein,  $k$ , in complex

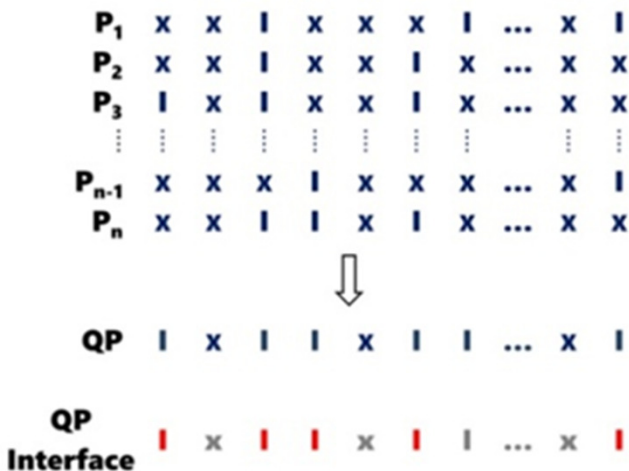
$$x_k = \begin{cases} 1 - 10^{-200}, & \text{if } E_k < 10^{-200} \\ 1 - E_k, & \text{if } 10^{-200} \leq E_k \leq 10^{-2} \\ 0, & \text{if } E_k > 10^{-2} \end{cases}$$

3. **Ligand weight**: The nature of the ligand involved in the interaction with the homologous protein.  $k$

$$y_k = \begin{cases} \frac{\sum_{j=1, j \neq k}^N E_{(L_k, L_j)}}{N - 1}, & \text{if } N > 1 \\ 1, & \text{if } N = 1 \end{cases}$$

## Residue Interaction Score

The interaction score,  $S_i$ , of residue,  $i$ , is the weighted sum of interface residue scores in the homologs over all corresponding residues scores



$$S_i = \frac{\sum_{k=1}^N W_{kj}}{\sum_{k=1}^N X_j Y_j}$$

where  $W_{ik} = \begin{cases} X_k Y_k, & \text{if } i \text{ interacts with } L_k \\ 0, & \text{otherwise} \end{cases}$

The number of interface residues,  $T$ :  
 a weighted average of interface size in homologs

## T-PIP Performance

Standard benchmark dataset: Ds56unbound (CAPRI)

56 unbound chains homologous to known complexes

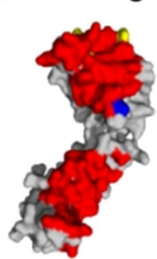
Predictor	Accuracy
T-PIP trivial (DS27unbound)	87.0
T-PIP homologous (DS24unbound)	82.3
PredUs (DS5unbound)	75.8
<b>T-PIP framework (DS56unbound)</b>	<b>84.0</b>

⇒ Exploitation of homology improves interface prediction

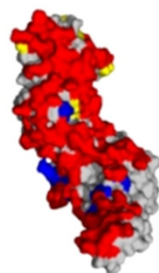


## T-PIP Performance

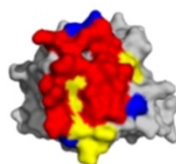
### a) Homologous



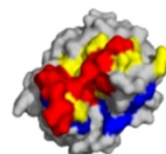
1YNT-A  
F1=95.4



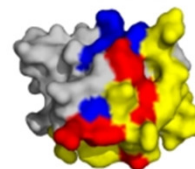
1QHD-A  
F1=84.5



2J59-A  
F1=77.3



1TE1-B  
F1=65.3



1KXQ-H  
F1=47.6

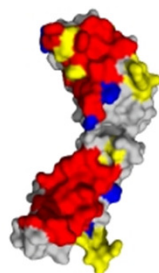


1S70-B  
F1=22.2

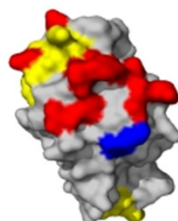
### b) Trivial



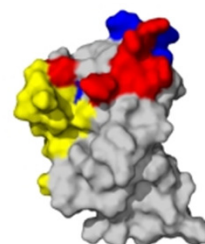
1V74-A  
F1=97.9



1KEN-L  
F1=80.0

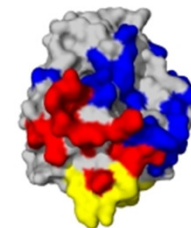


1ZHI-B  
F1=66.6



1TPX-A  
F1=35.3

### c) Unknown



1TA3-A  
F1=42.2



## Evaluation of TPIP's weights

<i>T-PIP homologous (DS24unbound)</i>					
Query weight	Ligand weight	Precision	Recall	F1	Accuracy
1	1	40.6	32.5	37.0	81.6
$x_k$	1	40.8	32.7	38.7	82.4
1	$y_k$	42.4	34.6	37.3	<b>82.6</b>
$x_k$	$y_k$	<b>43.3</b>	<b>35.6</b>	<b>41.7</b>	82.3

- Query weight ( $x_k$ ): modest improvements
- Ligand weight ( $y_k$ ): significant increase of performance
- Combined weights: further improvements

## T-PIP Comparative Study

Predictor (DS56unbound)	Precision	Recall	F1	Accuracy	MCC
Promate	28.7	27.3	28.0	76.6	14.0
PINUP	30.4	30.1	30.2	76.9	16.4
Cons-PPISP	37.4	34.5	35.9	79.5	23.8
Meta-PPISP	38.9	24.0	29.7	81.1	20.2
IBIS	48.2	29.3	34.4	82.5	27.9
PrISE	43.7	44.0	43.8	81.2	32.6
PredUs	43.3	53.6	47.9	73.2	30.4
T-PIP framework	53.9	48.5	49.6	84.0	41.1

## T-PIP Comparative Study

Predictor	Precision	Recall	F1	Accuracy	MCC
<b>T-PIP DS120</b>	52.6	56.1	52.5	85.4	45.1
<b>PredUs DS120</b>	47.3	58.2	48.5	69.4	24.4
<b>PrISE DS120</b>	38.5	48.9	40.9	80.7	31.2
<b>IBIS DS120</b>	40.9	36.9	36.2	83.6	28.8
<b>T-PIP DS236</b>	53.2	55.3	52.1	85.3	44.8
<b>PrISE DS236</b>	41.2	47.5	41.5	81.0	32.0
<b>IBIS DS236</b>	42.6	37.4	37.4	83.8	29.9

## T-PIP: discussion

- **State-of-the art, only PredUs performs better on Recall**
- Both interface conservation & **ligand diversity** are important
- Structure of the target is NOT required

Predictor (DS56unbound)	Precision	Recall	F1	Accuracy	MCC
T-PIP	53.9	48.5	49.6	84.0	41.1
<b>T-PIP</b> <sub>QPseq+S-MSA</sub>	53.4	48.1	49.2	83.9	40.7

- Interface residues are selected independently from each other  
-> filtering interface according to intrinsic features could be useful

# Protein Interface Prediction: conclusions

- **Homologous complexes are usually available**
- **3D structure of the target is NOT necessary**
- **Protein Interface Prediction remains an unsolved problem!**

Precision	Recall	F1	Accuracy	MCC
~55%	~60%	~55%	~85%	~45%

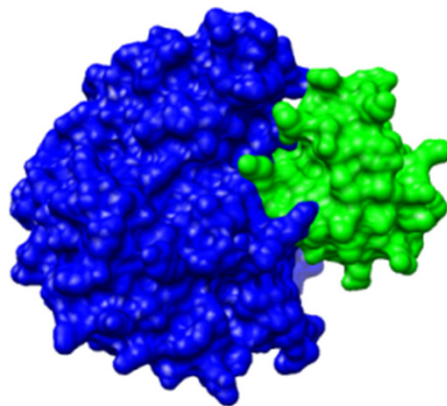
**'accuracy'** biased by the low ratio between interface & non-interface residues

- **Still, predictions can be useful...**

**PrePPI: a database of predicted and experimentally determined protein-protein interactions for yeast (31,402) and human (317,813)**

Structure-based prediction of protein-protein interactions on a genome-wide scale, Nature 490, 556–560, 2012, Barry Honig et al. (PredUs team)

# Protein-Protein Complex Structure Prediction



## Prediction of a complex structure

### Protein-protein docking (template free)

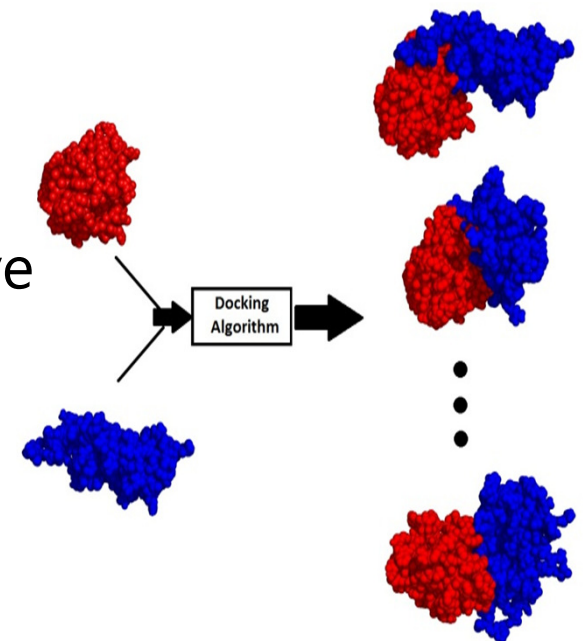
Explore conformation space using scoring functions based on energy potentials and shape complementarity

- Rigid docking + Side-Chain and Back-bone Flexibility
- Soft docking (coarse)

-> generate many docked poses

-> scoring function fail to detect near-native configurations

-> **post-processing: model ranking**





# Prediction of a complex structure

## Model ranking

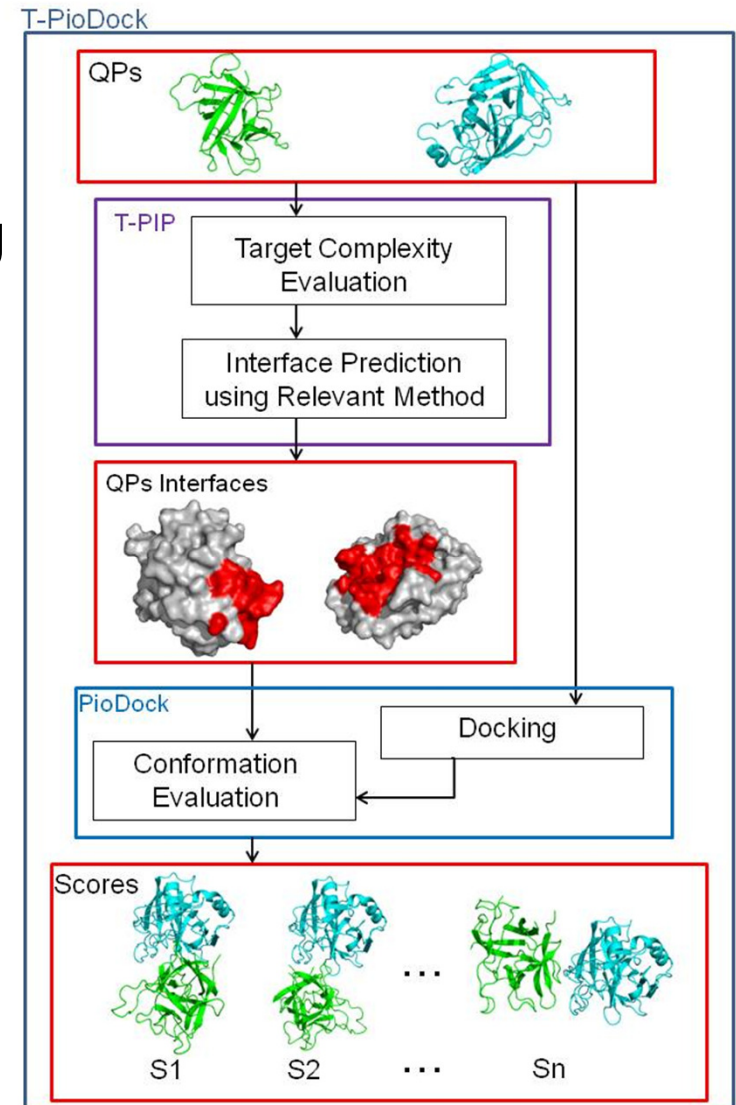
- Model clustering
- Empirical Energy Functions
- Statistical and Machine Learning Functions
- Knowledge of Predicted Interfaces

# Model ranking using Predicted Interfaces

## PioDock: Protein Interface Overlap for Docking model scoring

$$\text{complexOverlap}_{A-B} = \frac{\text{overlap}_A + \text{overlap}_B}{2}$$

$$\text{overlap}_A = \frac{\text{interface } A_{\text{Docked}} \cap \text{interface } A_{\text{T-PIP}}}{\sqrt{(\text{intrfaces } A_{\text{Docked}}) \cdot (\text{interfaces } A_{\text{T-PIP}})}}$$



## Evaluation

Docking predictions produced using the **ClusPro 2.0** docking server (performed best at CAPRI 2009)

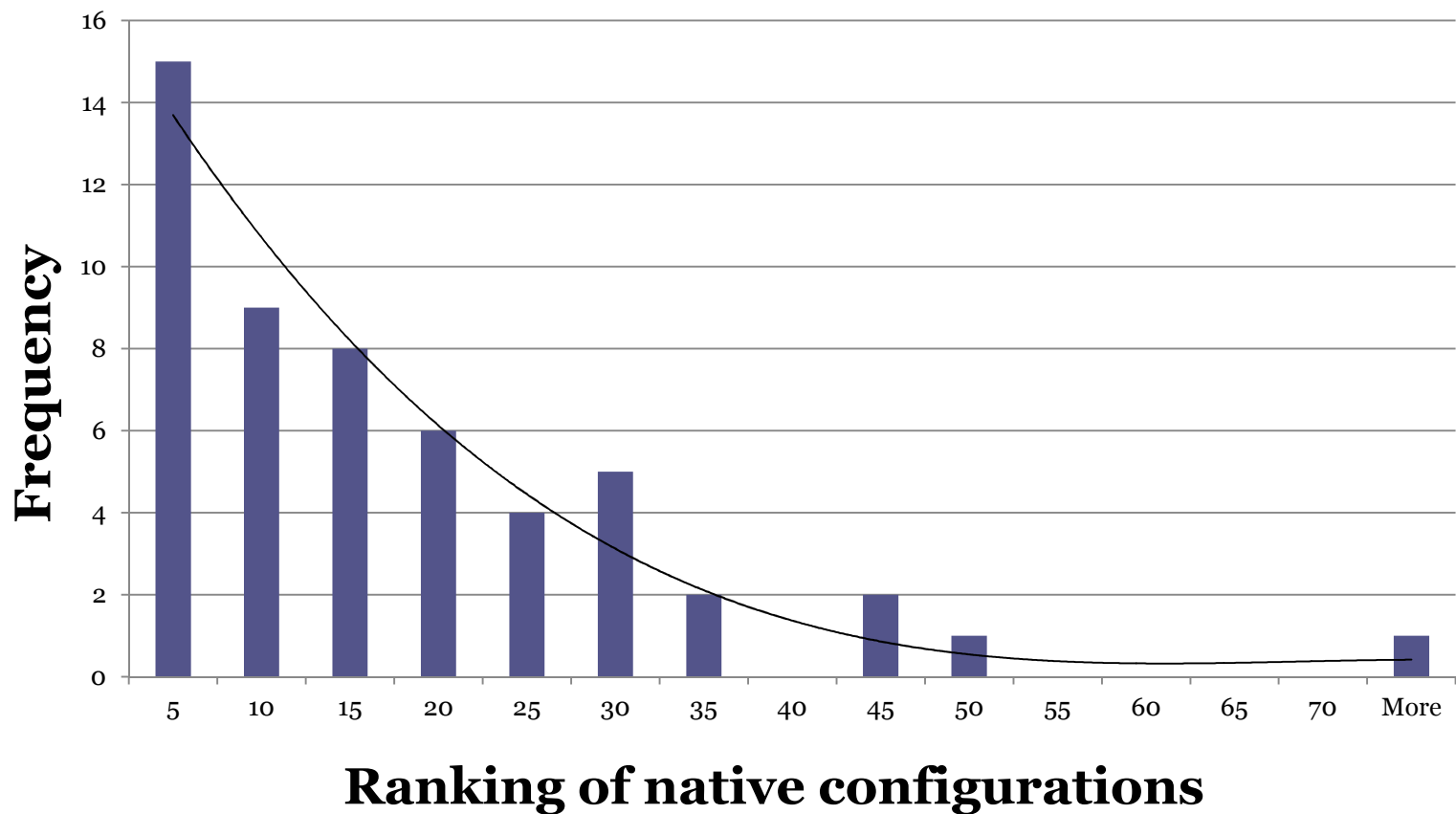
Ranking list comparison using chi-squared statistic ( $\chi^2$ )  
 -> higher weights to the models that are ranked higher

$$\chi^2 = \sum_{k=1}^n \frac{(\text{observed}_k - \text{expected}_k)^2}{\text{expected}_k} \quad \text{Perfect ranking: 0}$$

Ground truth (CAPRI)	Ranking method applied to DS93 ('homologous' models)								
	x-rmsd	Interfaces + PioDock	T- PioDock	IRAD	ZRANK	SPIDER	SVM	TSVM	MI
i-rmsd	5.2	11.6	30.0	39.5	43.3	49.1	60.7	61.4	67.8
l-rmsd	6.0	12.5	29.7	39.5	44.2	50.6	63.9	64.5	70.9

PioDock treats docking interfaces as *patches*

## Evaluation

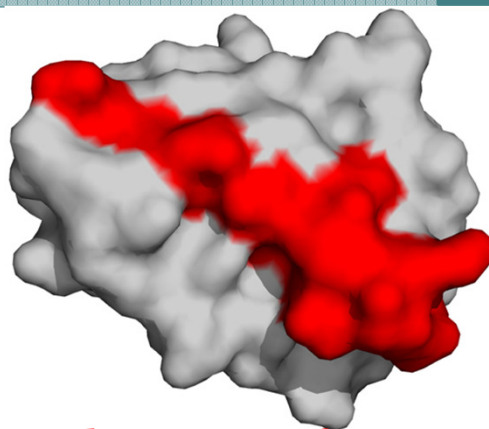


Native pose tends to be present in the top of the ranking lists

# Complex structure prediction: conclusions

- **Docking software are still not able to produce native like models for every target**
- **Complex structure prediction remains an unsolved problem!**
- **Since interface predictors do not explicitly refer to binary residue interactions, model evaluation is coarse**
  - **Energy based model could be used to reject incompatible configurations**

# T-PioDock Software



[manorey.net/bioinformatics/wepip/](http://manorey.net/bioinformatics/wepip/)

Kingston University London

# T-PioDock Software

- T-PioDock software available to download:

[manorey.net/bioinformatics/wepip/](http://manorey.net/bioinformatics/wepip/)

- Participation in the latest Critical Assessment of PRedicted Interactions (CAPRI) competition

T-PioDock Software:  
Template based Protein Interface prediction and  
protein interface Overlap for Docking model  
scoring

[Introduction](#) [Reference](#) [Contact US](#)

## Download Software

Two software are available for download:

1. **T-PIP:** For Protein Interface Prediction

This program is written in C# as a **Windows** command line application.

Download package contains: Executable application, Introduction document, Examples

**Requirements:** Blast and ClustalW are required. For more detailed Information on how to install these tools see Introduction document available in the download folder.

Please select a version:

◊ Version 1.0 - [Download](#)

2. **PioDock:** For Ranking Docking Models

This program is written in C# as a **Windows** form application.

◊ Version 1.0 - [Download](#)

## Contact Information

Reyhaneh Esmailbeiki, email: [r.esmailbeiki \[at\] kingston \[dot\] ac \[dot\] uk](mailto:r.esmailbeiki@kingston.ac.uk)

R. Esmailbeiki, J.-C. Nebel, "Scoring docking conformations using predicted protein interfaces", *BMC Bioinformatics*, 15:171, 2014.

# Current Research Interests (1/2)

<http://staffnet.kingston.ac.uk/~ku33185/Bioinformatics.html>

## Protein annotation from either sequence or 3D structure

- **Primary sequence contribution to the optical function of the eye lens**, K. Mahendiran, C. Elie, J.-C. Nebel, A. Ryan & B.K. Pierscionek, *Scientific Reports*, 4, 5195, 2014
- **A Stochastic Context Free Grammar based Framework for Analysis of Protein Sequences**, W. Dyrka & J.-C. Nebel, *BMC Bioinformatics*, 10:323, 2009
- **Automatic generation of 3D motifs for classification of protein binding sites**, J.-C. Nebel, P. Herzyk & D. R. Gilbert, *BMC Bioinformatics*, 8:32, 2007
- **Generation of 3D templates of active sites of proteins with rigid prosthetic groups**, J.-C. Nebel, *Bioinformatics*, 22(10): 1183-1189, 2006

## Protein 3D structure prediction

- **Ab initio Protein Structure Prediction: Methods and Challenges**, J. Abbass, J.-C. Nebel & N. Mansour, in *Biological Knowledge Discovery Handbook: Preprocessing, Mining & Postprocessing of Biological Data*, M. Elloumi & A. Y. Zomaya (Editors), Wiley Book Series on Bioinformatics: Computational Techniques & Engineering, 32, pp. 703-724, January 2014
- **Probabilistic grammatical model of protein language and its application to helix-helix contact site classification**, W. Dyrka, J.-C. Nebel & M. Kotulska, *Algorithms for Molecular Biology*, 8:31, 2013
- **Quality assessment of protein model-structures based on structural and functional similarities**, B. Konopka, J.-C. Nebel & M. Kotulska, *BMC Bioinformatics* 2012, 13:242
- **Accuracy in predicting secondary structure of ionic channels**, B. Konopka, W. Dyrka, J.-C. Nebel & M. Kotulska, In 'New Challenges in Computational Collective Intelligence', Springer-Verlag, 244, pp. 315-326, 2009



# Current Research Interests (2/2)

<http://staffnet.kingston.ac.uk/~ku33185/Bioinformatics.html>

## 3D structure prediction of protein complexes

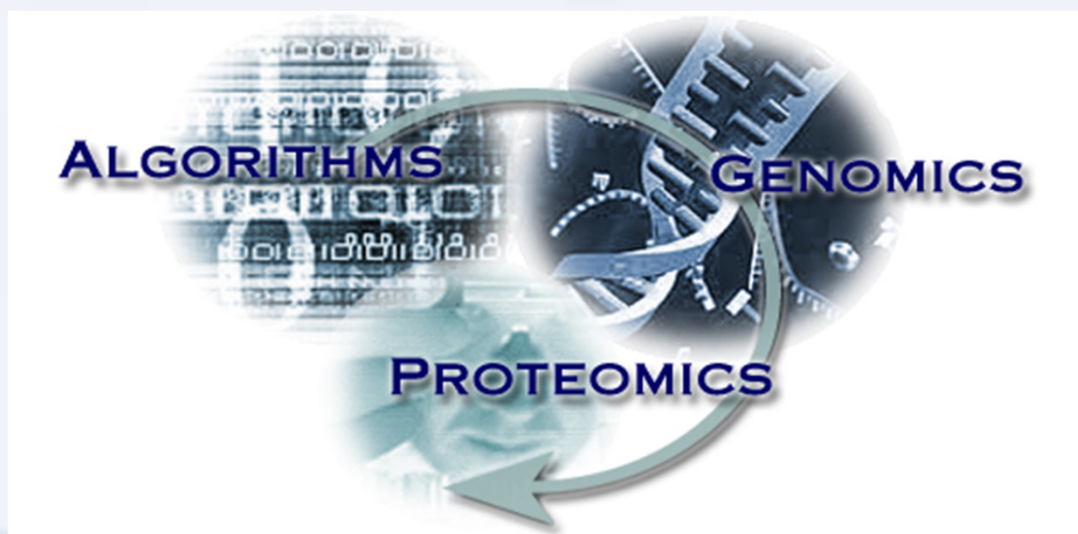
- **Scoring docking conformations using predicted protein interfaces**, R. Esmailbeiki & J.-C. Nebel, BMC Bioinformatics, 15:171, 2014
- **Proteomics and Bioinformatics Soon to Resolve the Human Structural Interactome**, J.-C. Nebel, Journal of Proteomics & Bioinformatics, 5(10): xi-xii, 2012
- **Structure prediction of LDLR-HNP1 complex based on docking enhanced by LDLR binding 3D motif**, R. Esmailbeiki, D. Naughton & J.-C. Nebel, Protein & Peptide Letters, 19(4): 458-67, 2012
- **Unbiased Protein Interface Prediction Based on Ligand Diversity Quantification**, R. Esmailbeiki & J.-C. Nebel, Open Access Series in Informatics (OASICS), Vol. 26, GCB, 2012

## Other interests

- **Identification of NAD(P)H Quinone Oxidoreductase Activity in Azoreductases from *P. aeruginosa*: Azoreductases and NAD(P)H Quinone Oxidoreductases Belong to the Same FMN-Dependent Superfamily of Enzymes**, A. Ryan, E. Kaplan, J.-C. Nebel, E. Polycarpou, V. Crescente, E. Lowe, G. Preston & E. Sim, PLOS ONE, 2014
- **Comparative study and meta-analysis of meta-analysis studies for the correlation of genomic markers with early cancer detection**, Z. Lanara, E. Giannopoulou, M. Fullen, E. Kostantinopoulos, J.-C. Nebel, H. Kalofonos, G. Patrinos & C. Pavlidis, Human Genomics, 7:14, 2013
- **Comparative Analysis of Genomic Signal Processing for Microarray data Clustering**, R. Istepanian, A. Sungoor & J.-C. Nebel, IEEE Trans. on NanoBioscience, 10(4): 225-238, 2011
- **Why inverse proteins are relatively abundant**, J.-C. Nebel & C. Walawage, Protein & Peptide Letters, 17(7): 854-860, 2010

# Journal of Proteomics & Bioinformatics Related Journals


- [Transcriptomics: Open Access](#)
- [Journal of Pharmacogenomics & Pharmacoproteomics](#)
- [Journal of Data Mining in Genomics & Proteomics](#)





For more details on Conferences Related  
Journal of Proteomics & Bioinformatics  
please visit:

[http://www.conferenceseries.com/  
biochemistry-meetings](http://www.conferenceseries.com/biochemistry-meetings)



Open Access Membership with OMICS International enables academicians, research institutions, funders and corporations to actively encourage open access in scholarly communication and the dissemination of research published by their authors.

For more details and benefits, click on the link below:

<http://omicsonline.org/membership.php>