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

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
Protein-Protein Interactions

The interior of cells is crowded
e.g. dynamic molecular model of the bacterial cytoplasm
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

Effect of single genetic mutation:
from malaria benefit to anaemia

\[
>\text{sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2}
MVHLTTVEKSAVTALWGKVNVDEVGEALGRLLVVYPWTTQRFESFGDLSTPDAMGNPK
VKAHLKVLGAFSDGLAHLDNLKGTFLSELHCDKLHVDPRFLLGNVLCVLHHF
KEFTPPVAAYQKVAGVANALAHKYY
\]

-> 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
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
Protein Interface Prediction
Protein Interface Prediction

Interface residues

CAPRI (Critical Assessment of PRediction of Interactions) 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

Kingston University London
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

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

*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, $k$

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

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-PIP trivial</td>
<td>87.0</td>
</tr>
<tr>
<td>T-PIP homologous</td>
<td>82.3</td>
</tr>
<tr>
<td>PredUs</td>
<td>75.8</td>
</tr>
<tr>
<td>T-PIP framework</td>
<td>84.0</td>
</tr>
</tbody>
</table>

⇒ 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

<table>
<thead>
<tr>
<th>T-PIP homologous (DS24unbound)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query weight</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>$x_k$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>$x_k$</td>
</tr>
</tbody>
</table>

- Query weight ($x_k$): modest improvements
- Ligand weight ($y_k$): significant increase of performance
- Combined weights: further improvements
# T-PIP Comparative Study

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

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

<table>
<thead>
<tr>
<th>Predictor (DS56unbound)</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Accuracy</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-PIP</td>
<td>53.9</td>
<td>48.5</td>
<td>49.6</td>
<td>84.0</td>
<td>41.1</td>
</tr>
<tr>
<td>T-PIP$_{QPseq+S-MSA}$</td>
<td>53.4</td>
<td>48.1</td>
<td>49.2</td>
<td>83.9</td>
<td>40.7</td>
</tr>
</tbody>
</table>

- 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!

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Accuracy</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>~55%</td>
<td>~60%</td>
<td>~55%</td>
<td>~85%</td>
<td>~45%</td>
</tr>
</tbody>
</table>

‘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)

Protein-Protein Complex Structure Prediction

Kingston University London
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**

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

\[
\text{overlap}_A = \frac{\text{interface } A_{\text{Docked}} \cap \text{interface } A_{T-\text{PIP}}}{\sqrt{\text{interfaces } A_{\text{Docked}} \cdot \text{interfaces } A_{T-\text{PIP}}}}
\]
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

$$\chi^2 = \sum_{k=1}^{n} \frac{(observed_k - expected_k)^2}{expected_k}$$

Perfect ranking: 0

<table>
<thead>
<tr>
<th>Ground truth (CAPRI)</th>
<th>Ranking method applied to DS93 (‘homologous’ models)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x-rmsd</td>
</tr>
<tr>
<td>i-rmsd</td>
<td>5.2</td>
</tr>
<tr>
<td>l-rmsd</td>
<td>6.0</td>
</tr>
</tbody>
</table>

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/

Kingston University London
T-PioDock Software

- T-PioDock software available to download:
  manorey.net/bioinformatics/wepip/

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

Current Research Interests (1/2)

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

Protein annotation from either sequence or 3D structure


Protein 3D structure prediction

- **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**

**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
- 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
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