STRUCTURAL MODELING OF PROTEIN INTERACTIONS USING EVOLUTION

Raphaël GUEROIS

STRUCTURAL BIOLOGY AND RADIOBIOLOGY LAB
I2BC - CEA SACLAY
• The activation/inhibition of cellular pathways rely on **synergies and competitions** at the binding surfaces of macromolecules.

• An atomic/residue scale analysis
  - Highlight steric hindrance, co-association
  - Allow for the specific design of disruptive and compensatory mutants

Macromolecules keep encountering each other in the crowded cell.

Diffusion & crowding model

*E. coli* cytoplasm

Macromolecules

*in cellulo*: ~400 mg/ml


Protein in crystals: ~500 - 1500 mg/ml

Size: 1/12th of *E. coli* cell

80 nm

1,000 proteins

CONSERVATION OF BINDING MODES DURING THE COURSE OF EVOLUTION

Threshold for interface conservation
≈ 30% seq. id
Analysis of an interface contact

Rpb7
S. cerevisiae

Rpb4
S. cerevisiae
Analysis of an interface contact
Analysis of an interface contact

Rpb7
H. sapiens

Rpb4
H. sapiens

E35

R21

N35

Saccharomyces_cerevisiae
Schizosaccharomyces_pombe
Aspergillus_fumigatus
Neurospora_crassa
Candida_albicans
Candida_glabrata
Pichia_pastoris
Gibberella_zea
Nectria_haematococca
Homo_sapiens
Apis_mellifera
Anopheles_gambiae
Aedes_aegypti
Drosophila_melanogaster
Ixodes_iscapularis
Acrithosiphon_pismus
Schistosoma_mansoni
Trichoplax_adhaerens
Interest in structural information to interpret sequence data.

Analysis of an interface contact

<table>
<thead>
<tr>
<th>Saccharomyces_cerevisiae</th>
<th>Schizosaccharomyces_pombe</th>
<th>Aspergillus_fumigatus</th>
<th>Neurospora_crassa</th>
<th>Candida_albicans</th>
<th>Candida_glabrata</th>
<th>Pichia_pastoris</th>
<th>Gibberella_zeaee</th>
<th>Nectria_hasmatococca</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apis_mellifera</td>
<td>Anopheles_gambiae</td>
<td>Aedes_aegypti</td>
<td>Drosophila_melanogaster</td>
<td>Ixodes_scapularis</td>
<td>Acrithosiphon_pisum</td>
<td>Schistosoma_mansoni</td>
<td>Trichoplax_adhaerens</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E35</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>V</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R31</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>L</td>
<td>A</td>
<td>Q</td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>L</td>
<td>E</td>
<td>N</td>
<td>K</td>
<td>K</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R21</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>K</td>
<td>A</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>K</td>
<td>N</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>K</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N35</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>K</td>
<td>A</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>K</td>
<td>N</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>Q</td>
<td>N</td>
<td>N</td>
<td>E</td>
<td></td>
</tr>
</tbody>
</table>
INTERFACE CONSERVATION AND PLASTICITY WHILE MAINTAINING BINDING SPECIFICITY

Hsm3/S5B : Rpt1 assembly factor

ATPase from proteasome 19S

1-6 share 45% seq id

9% identity 85% identity

Yeast HSM3
Human S5B

Barrault, M. B., et al. (2012) PNAS
Overviews about the conservation and coevolution of protein-protein interactions (PPIs).

- Interplay between sequence variations and PPI properties?
- How does conservation and coevolution of sequences impact the structural dimension?
- Can sequence coevolution profiles be of any help to get further insights into structural interactomes?
GRAPH REPRESENTATIONS TO ACCOUNT FOR PPI NETWORKS TOPOLOGIES

Physical interactions

Proteins
Size of the physical interactome

Statistics on the Biogrid database
http://thebiogrid.org/

- Human: ~20,000 genes
- S. cerevisiae: 6,000 genes
- Arabidopsis: 26,000 genes
- Drosophila: 14,000 genes
Apparent simplicity hiding a variety of complex behaviour
- Direct physical interactions vs indirect
- Conditional binding
- Different time dependent expression levels
- Isoforms arising from alternative splicing
- …
Experimental structure determination
Homology modelling
Structural prediction of protein-protein interfaces
How far are PPI conserved at the global level?

What are the proper determinants to infer conserved binding and conserved interfaces?

Is % sequence identity a good proxy?

Can we expect the same behavior depending on the nature of the complex?
HOW FAR ARE PPI CONSERVED AT THE GLOBAL LEVEL?

Comparison of 3 binary interactomes

Although absent in the human PPI database: 15 out of 36 could eventually be detected by targeted interaction assays on human proteins

Small overlap between large ensemble of binary interactions...

Difficult to quantify given the risks of false negatives
HOW FAR ARE PPI CONSERVED AT THE GLOBAL LEVEL?

• Cases of one-to-one orthologs (neither duplication, nor gene loss/gain), the estimated fraction of conserved PPI is rather high

• Small PPI datasets experimentally tested encompassing yeasts, worms, and mammals were used to get confidence intervals of the PPI evolutionary rate leading to:

<table>
<thead>
<tr>
<th>Organisms</th>
<th>Fraction of conserved PPIs</th>
<th>Divergence time (MY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>1.00</td>
<td>0</td>
</tr>
<tr>
<td>mouse</td>
<td>0.98</td>
<td>90</td>
</tr>
<tr>
<td>fish</td>
<td>0.89</td>
<td>430</td>
</tr>
<tr>
<td>fly</td>
<td>0.79</td>
<td>900</td>
</tr>
<tr>
<td>worm</td>
<td>0.77</td>
<td>1000</td>
</tr>
<tr>
<td>fungi</td>
<td>0.71</td>
<td>1300</td>
</tr>
<tr>
<td>plants</td>
<td>0.66</td>
<td>1600</td>
</tr>
</tbody>
</table>


rate of PPI evolution: $\sim 3 \times 10^{-10}$ per PPI per year

Several ‘recent’ duplications occurred in Arabidopsis genomes providing key insights into PPI rewiring during evolution.
OTHER FACTORS CONTRIBUTING TO REDUCE THE CONSERVATION OF PPI NETWORKS

- Isoforms mediated rewiring of PPI networks

  ➔ Significant change in PPI profiles were observed for a large majority of the isoforms pairs

DISTINGUISHING DIFFERENT CLASS OF PROTEIN FAMILIES IN THE ANALYSIS OF PPI COEVOLUTION

Interactions between globular regions of proteins
Permanent and Transient Complexes

Linear Motif Binding domains
Linear Motifs
Phosphorylations, etc...

Transcription Factors
Is sequence identity a good proxy to estimate the likelihood of PPI conservation?

How robust are complexes to variations in sequences?

Example of a large molecular machine complexes

Substitution of one specific subunit from other yeasts

- N. crassa 35%
- Y. lipolytica 45%
- C. Glabrata 85%

S. cerevisiae Exosome

% of exosome assembly

% seq id 100% 85% 45% 35%

WT phenotypes
Impaired phenotypes

PPI conservation decreases with increasingly transient interactions...

Comparison of PPI map for SH3-containing orthologs in yeast and worm

Two sources of variability:

- Shift in Binding Specificity for SH3 and other LMBDs
- Fast evolutionary rates in Linear Motifs


- Conservation of SH3-mediated interactions: no better than random
- Interactions between non-SH3 proteins: 50% conservation for orthologs worms and yeasts

For transient interactions, with disordered regions inferring PPI conservation can become problematic

⇒ Low conservation when interactions are mediated by Linear Motif Binding Domains such as SH3 modules
Conservation of interactions between Transcription Factors appears as the most versatile category of proteins

- **Well established for > 400 TFs in C. elegans**

- **in 53 bZIP TF in Humans**

Only in the case of very stable associations (<50 nM), do bZIP TFs share similar binding properties following their % of sequence identity.
CONSERVATION AND COEVOLUTION IN PPI NETWORKS

Interactions between globular proteins

Permanent and Transient Complexes

Linear Motif Binding domains
Linear Motifs
Phosphorylation

Transcription Factor

% Sequence Identity good proxy for conservation of interactions
Experimental structure determination
Homology modelling
Homology modelling
8600 exp. structures used to predict
~14000 human proteins structures

(Califano & Honig’s groups, Nature 2012)
Experimental structure determination
Homology modelling

Structural prediction of protein-protein interfaces
HOW TO PREDICT THE STRUCTURE OF AN ASSEMBLY?

3D models

Database of Interface Structural Templates

Sequence 1

Sequence 2

If NO…

Systematic Geometric Search: Molecular Docking

If YES

Scoring Interface Likelihood
FINDING A CORRECT INTERFACE IN THE HAYSTACK OF POSSIBLE ASSEMBLIES

Protein A $\leftrightarrow ? \rightarrow $ Protein B

PRINCIPLES OF MOLECULAR DOCKING

STEP 1: Low resolution step. Coarse-grained rigid body docking

STEP 2: refinement step. Flexibility, atomic details

> $10^5$ decoys

Filters

~ 10 clusters

1 most likely model
WHAT INFORMATION IS AVAILABLE ABOUT PROTEIN-PROTEIN INTERFACES?

- General goal = structural prediction of interfaces
- Many levels of complementarity between proteins lead to specific recognition

**Shape / surface complementarity**

**Physico-chemical complementarity**

(basic) electrostatic potential (acidic)

MODULAR ORGANIZATION OF THE BINDING INTERFACE

Independent modules

Inside modules

Between modules

Reichmann Curr Opin Struct Biol. 2007
Interface conservation translates into selection pressures on protein surfaces

Teichmann, *JMB* 2002

CAN CONSERVATION HELP TO IDENTIFY PROTEIN ASSEMBLIES?

Potential lack of specificity for the prediction of conserved surface residues located in a given interface.

=> CO-EVOLUTION

Pairwise covariation analysis does not give a strong signal (high noise).

Direct contact analysis can yield significant signal (requires large co-alignments).

Mintseris & Weng PNAS 2005
Weigt et al PNAS 2009
Morcos et al PNAS 2011
CAN WE GET USEFUL INSIGHTS ON HOW PROTEIN COMPLEXES COEVOLVED BY ANALYSING THE PROPERTIES OF INTEROLOGS?

=> Quantitative analysis of interface plasticity in evolution
« Structural interactome »

~17,000 non-redundant interfaces among which
~3,700 heteromeric interfaces

From a Keyword or the PDB entry of a complex, you can browse:
- structural homologs for every chain in other complexes
- structural interologs for every interface
- retrieve pre-computed sequence alignments in diverse species


http://biodev.cea.fr/interevol
InterEvolAlign

From a Keyword or the PDB entry of a complex, you can browse:
- structural homologs for every chain in other complexes
- structural interologs for every interface
- retrieve pre-computed sequence alignments in diverse species

Browse the InterEvol Database

From 1 or 2 sequences of interacting partners:
- build 2 multiple sequence alignments with the same species ordered in each
- query the InterEvol database with alignments using profile-profile comparison method

http://biodev.cea.fr/interevol
Can we get useful insights on how protein complexes coevolved by analysing the properties of interologs?

> 1000 pairs of structures of interologous complexes.

=> Quantitative analysis of interface plasticity in evolution
InterEvol database: Structural interologs
A dataset to study interface plasticity in evolution

Proportion of interolog pairs in this range of iRMSD

interface RMSD between the 2 interologs
What happens to the interface positions in contact?

Analysis over 1,000 pairs of structural interologs

Conserved contacts: 59%
Non-conserved contacts: 41%

High versatility and rewiring of interface contacts
SCENARIOS OF VERSATILITY FOR CHARGED CONTACTS

Angiogenin $\leftarrow$ 38% $\rightarrow$ Ribonuclease I $\leftarrow$ 38% $\rightarrow$ Neurotoxin

Ribonuclease inhibitor (conserved partner)

Conserved Lys involved in
- Salt bridge
- H bond network
- Water mediated contact

$=>$ Quantification of these scenarios for recovery of lost complementarity

**SCENARIOS OF VERSATILITY FOR CHARGED CONTACTS**

**Salt bridges**

- seq A: CKDIN
- seq A': CKNQIN
- seq B: WVKSC
- seq B': WVRSN

What’s the structural fate of the remaining charged residue?

- Mutated
- Salt bridge
- Salt-bridges
- Uncharged contacts
- Solvent
- Switching out
- Other plasticity events

Anchor residues (Rajamani et al, PNAS 2004)

- A few residues per interface
- Important for the binding
- Protruding residues, burying a large surface upon binding
Pairwise contacts involving at least one anchor residue are very well conserved.

Anchors concentrate many interface contacts.
HIGHLY VERSATILE REWIRING OF CONTACTS...
⇒ EXCEPT FOR APOLAR PATCHES
HIGHLY VERSATILE REWIRING OF CONTACTS… → EXCEPT FOR APOLAR PATCHES

Two equivalent apolar patches in a pair of interologs sharing about 25% sequence identity.

86% of apolar contacts between patches are conserved.

What have we learnt from the analysis of interface plasticity in evolution?

- Structural interologs have very similar overall interface structure.

- However, there is a high level of non-conservation in interface contacts.

- We can analyze and quantify the different scenarios for recovery of lost complementarity.

- Anchors and apolar patches display more conserved contacts.
DEVELOPMENT OF INTEREVSCORE, A DOCKING SCORE TAKING EVOLUTION INTO ACCOUNT

Protein A $\leftrightarrow$ Protein B

10⁴ decoys

Filters

$\sim$ 10 decoys

1 most likely model

2-body + 3-body contacts

Evolutionary information

Apolar patches

Andreani et al, Bioinformatics 2013
Development of InterEvScore, a docking score taking evolution into account

Inter-molecular interface contacts

Contact propensities derived from InterEvol
(statistics on 1,289 interfaces)

<table>
<thead>
<tr>
<th>Contact Propensity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o(0.53)</td>
<td>0.53</td>
</tr>
<tr>
<td>o(-0.05)</td>
<td>0.04</td>
</tr>
<tr>
<td>o(-0.11)</td>
<td>-0.26</td>
</tr>
</tbody>
</table>

\[ o(i,j) = \frac{N_{ij}}{N} \]

\[ p(i,j) = \frac{o(i,j)}{e(i,j)} \]

\[ 2B = \sum_{(i,j)} [\ln p(i,j)] \]

\[ 3B = \sum_{(i,j,k) \in (g_i, g_j, g_k)} [\ln p(i,j) + \ln p(j,k) + \ln p(i,k) + \ln c(g_i, g_j, g_k)] \]

\[ c(g_i, g_j, g_k) = \frac{p(g_i, g_j, g_k)}{p(g_i, g_j)p(g_j, g_k)p(g_i, g_k)} \]

N_{ij} : number count of atomic contacts between residue types i and j,
N : total number count of atomic contacts participating in all two-body interactions
Development of InterEvScore, a docking score taking evolution into account

Contact propensities derived from InterEvol
(statistics on 1,289 interfaces)

<table>
<thead>
<tr>
<th>Contact</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>-0.11</td>
</tr>
</tbody>
</table>

Inter-molecular interface contacts

Keep only the best contact for each residue

Residue
Development of InterEvScore, a docking score taking evolution into account

Inter-molecular interface contacts

Evolutionary score term for \( i, j \)

<table>
<thead>
<tr>
<th>Alignment for A</th>
<th>Alignment for B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H. \text{sapiens} )</td>
<td>( M. \text{musculus} )</td>
</tr>
<tr>
<td>( D. \text{rerio} )</td>
<td>( S. \text{cerevisiae} )</td>
</tr>
</tbody>
</table>

Contact propensities derived from InterEvol (statistics on 1,289 interfaces)

<table>
<thead>
<tr>
<th>Contact</th>
<th>0.53</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.05</td>
<td>-0.26</td>
</tr>
</tbody>
</table>
Development of InterEvScore, a docking score taking evolution into account

Inter-molecular interface contacts

Sum of evolutionary score terms over all interface residues

<table>
<thead>
<tr>
<th>Residue</th>
<th>Contact Propensities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>+ 1.95</td>
<td></td>
</tr>
<tr>
<td>+ -0.32</td>
<td></td>
</tr>
<tr>
<td>+ 0.65</td>
<td></td>
</tr>
</tbody>
</table>

Keep only the best contact for each residue

Contact propensities derived from InterEvol (statistics on 1,289 interfaces)

<table>
<thead>
<tr>
<th>0.53</th>
<th>-0.05</th>
<th>-0.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>0.04</td>
<td>-0.26</td>
</tr>
</tbody>
</table>
Generate 54,000 decoys with ZDOCK

Score with InterEvScore

Rank and analyze decoys

Test procedure

Native structure

Near-native decoy
Metrics used for testing

54,000 models ranked by InterEvScore value

Decoy with best score
Top1

Decoy with worst score

Decoy with best score
Top1

Decoy with worst score

Near-native decoys

Predictive capacity = measure of the ‘concentration’ of near-native decoys in best ranked decoys

“Integrated Hit Rate”
54 complexes with known bound and unbound structures used as test.

Signaling and G-proteins

Enzyme-substrate or inhibitor

Transport & Chaperones

Transcription/Translation

Receptor-ligand and subunits

Hwang et al, Proteins 2010
Benchmark results: Capacity of different predictors to recognize the correct models

Integrative Success Rate over 54 cases

- 2-body statistical potential
- Residue conservation
- 2-body statistical potential + residue conservation
- InterEvScore (2-body)
- InterEvScore (2+3-body)
- Apolar patches only
- InterEvScore (2+3-body) + Apolar patches
INTEREvDock: A Server for Protein Rigid-Body Docking and Scoring Using Evolution

Yu et al, NAR 2016

http://bioserv.rpbs.univ-paris-diderot.fr/services/InterEvDock/
• Blind test for the different predictive strategies

http://www.ebi.ac.uk/msd-srv/capri/

About every six months structures of individual binding partners are released to the community.

→ X-ray structure of every query complex is kept secret.

• Up to 10 models can be submitted
• Models are ranked with respect to their precision
• Community discussions to extract the most efficient strategies
**CAPRI Evaluation Criteria**

- **$f_{(nat)}$:** Fraction of native residue contacts
- **L-rms:** Ligand residues
- **I-rms:** Interface residues

---

Lensink et al. 2013
**INTEREvDock Docking pipeline**

- **Free Docking**
  - > $10^5$ decoys
  - **Rigid-body (Zdock) Scoring (InterEvDock)**
    - ~ 10 decoys

- **Model Refinement (Rosetta)**
MONOUBIQUITINATION OF HISTONE H2A ON K119 BY POLYCOMB REPRESSIVE COMPLEX 1 (PRC1) ?

A recent CAPRI target (T95) …
A challenging multimeric assembly
Docking of the nucleosome (decamer) against a ubiquitin ligase (trimer) ?

Model proposed in 2011 by the authors based on mutagenesis analysis

Xray structure of the PRC1 Ub ligase

E2 ligase (UbcH5c)  E3 ligase (Bmi1 + Ring1b)

Bentley et al, EMBO J (2011) 30,3285-3297
OUR BEST MODELS DIFFERED SUBSTANTIALLY FROM THE ORIGINAL ONE…

“Blind” sampling with ZDOCK
Generation of 54,000 decoys
Filtering with InterEvScore
(rigid-body)
Perturbations / refinement
(flexibility)

Our most likely model…

… rotated by more than 90°

Bentley et al, EMBO J (2011)
OUR THIRD CAPRI MODEL CAME REASONNABLY CLOSED TO THE REAL STRUCTURE

Acceptable:
CAPRI model #1

Medium:
CAPRI model #3

X-ray structure
(McGinty et al, Nature 2014)
A REMARKABLE CONVERGENT EVOLUTION IN THE STRATEGY FOR BINDING NUCLEOSOMES

Sir3 ↔ Nucleosome
(Wang et al, PNAS 2013, PDB:4JJN)

Rcc1 ↔ Nucleosome
(Makde et al, Nature 2010, PDB:3MVD)

Cenp-C ↔ Nucleosome
(Kato et al, Science 2013, PDB:4INM)

⇒ A key Arginine acts as a binding anchor in all four partners of the nucleosome
FURTHER INSIGHTS FROM THE MODEL:
MOLECULAR DETERMINANTS FOR BRCA1 INTERACTION

BRCA1 monoubiquinates H2A K127/K129
→ The key Arginine seems to be conserved and likely mediates part of the recognition of the nucleosome

BRCA1 monoubiquinates H2A K127/K129

BRCA1 (Ring domain region)

BARD1

Bim1

Arg anchor

Ring1B

UbcH5

RING1B (PRC1)

B. TAUROS

PANDA

AMIBIA (Acanthamoeba)

WORM

A. THALIANA

SOYBEAN

PROTOZOAN (Babesia)

PROTOZOAN (Theileria)

PROTOZOAN (Paramecia)

**NDITSKRSLQ - - - - -**

*
OVERALL PERFORMANCE IN 6TH CAPRI

(***) 1 High – (**) 8 Medium - (*) 1 Acceptable  7 Incorrect

Protein

Protein

T59

T103

T97

T98#

T99#

T100#

T101#

T107#

T60-T64 5(**)

T95

T96

T67

T105

T104

T66

T67

# No success for any group
### PREDICTOR Ranking
(courtesy M. Lensink & S. Wodak)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Ntargets</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guerois</td>
<td>18</td>
<td>10/1***/8**</td>
</tr>
<tr>
<td>Zacharias</td>
<td>16</td>
<td>10/3***/2**</td>
</tr>
<tr>
<td>Vajda/Kozakov, Seok</td>
<td>17, 18</td>
<td>8/3***/2**</td>
</tr>
<tr>
<td>Weng</td>
<td>17</td>
<td>6/1***/4**</td>
</tr>
<tr>
<td>Fernandez-Recio</td>
<td>18</td>
<td>7/1***/3**</td>
</tr>
<tr>
<td>Vakser</td>
<td>18</td>
<td>6/2***/2**</td>
</tr>
<tr>
<td>Eisenstein</td>
<td>9</td>
<td>4/2***/2**</td>
</tr>
<tr>
<td>Zou</td>
<td>18</td>
<td>7/1***/2**</td>
</tr>
<tr>
<td>Bates</td>
<td>18</td>
<td>6/3**</td>
</tr>
<tr>
<td>Huang</td>
<td>18</td>
<td>5/3***</td>
</tr>
<tr>
<td>Zhou</td>
<td>16</td>
<td>4/2***/1**</td>
</tr>
<tr>
<td>Grudinin</td>
<td>18</td>
<td>4/3**</td>
</tr>
<tr>
<td>Bradley</td>
<td>5</td>
<td>3***</td>
</tr>
<tr>
<td>Shen</td>
<td>17</td>
<td>6/1***/1**</td>
</tr>
<tr>
<td>Baker</td>
<td>7</td>
<td>5/2**</td>
</tr>
<tr>
<td>Bonvin</td>
<td>18</td>
<td>4/1***/1**</td>
</tr>
<tr>
<td>Gray, Kihara</td>
<td>12, 18</td>
<td>3/2**</td>
</tr>
<tr>
<td>Furman</td>
<td>6</td>
<td>3/1***/1**</td>
</tr>
<tr>
<td>Takeda-Shitaka</td>
<td>2</td>
<td>2***</td>
</tr>
<tr>
<td>S_Liang, Di Maio, Moal, Negi</td>
<td>2, 2, 3, 10</td>
<td>2/1***/1**</td>
</tr>
</tbody>
</table>

Ranking on the number of medium- and high-quality models

---

... continued

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Parti.</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiorucci, Wolfson</td>
<td>2, 9</td>
<td>2**</td>
</tr>
<tr>
<td>J_Lee</td>
<td>7</td>
<td>6/1**</td>
</tr>
<tr>
<td>Niv</td>
<td>3</td>
<td>3/1**</td>
</tr>
<tr>
<td>Mitchell</td>
<td>6</td>
<td>2/1***</td>
</tr>
<tr>
<td>Ritchie</td>
<td>3</td>
<td>2/1**</td>
</tr>
<tr>
<td>Sali/Schneidman</td>
<td>5</td>
<td>1***</td>
</tr>
<tr>
<td>Fleishman</td>
<td>1</td>
<td>1**</td>
</tr>
<tr>
<td>Wallner, Totrov</td>
<td>3</td>
<td>1**</td>
</tr>
<tr>
<td>Sowdhamini</td>
<td>10</td>
<td>1**</td>
</tr>
<tr>
<td>del Carpio</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Yip, Xiao, Chang, Fernandez-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fuentes, Akiyama</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
General conclusion

1. **Evolutionary traces at complex interfaces**
   Interface contacts are highly versatile (interface plasticity). More conserved descriptors can be identified: **anchors** and **apolar patches**.

2. **Structural prediction of protein-protein interfaces**
   (Co)evolutionary information can be used to improve the structural prediction of protein interfaces, through **InterEvScore**.

3. **Tracks for improving the discrimination of coevolved interfaces**
   Better account for phylogeny
   Integration of a DCA-like method less demanding in terms of sequences…
Acknowledgements

Jinchao YU
Docking
CAPRI
experiment

Jessica
ANDREANI
InterEvScore

Guilhem
FAURE
InterEvol

Structural Biology
& Radiobiology Lab

Support
CEA Saclay - GIPSI
Arnaud Martel