Lesson 7

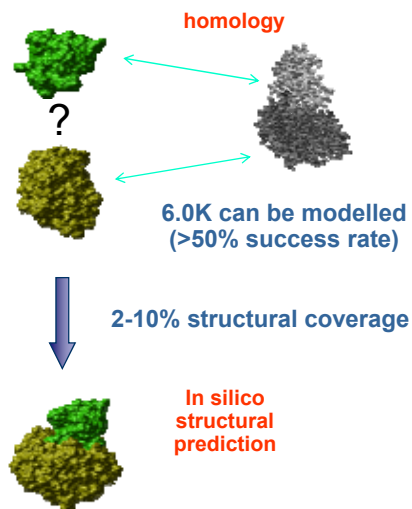# Advanced protein-protein docking:
current challenges

Juan Fernández-Recio

CSIC, BSC

juanf@bsc.es

---

Structural modeling of protein interactions

Human interactome
130-650K estimated interactions*
95K confirmed interactions*
5.8K with known structure

homology

?

**6.0K can be modelled
(>50% success rate)**

2-10% structural coverage

In silico structural prediction

* *Venkatesan et al 2009 Nature Methods 6, 83-90*
Stumpf et al 2008 PNAS 105, 6959-6964
Mosca et al. 2013 Nature Methods 10, 47

1

The challenge of flexibility
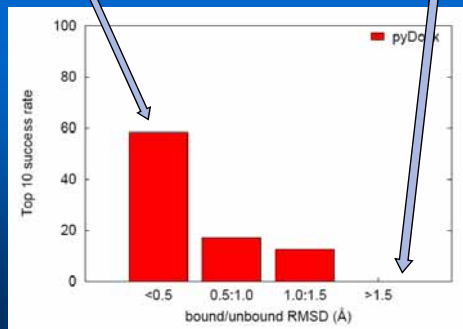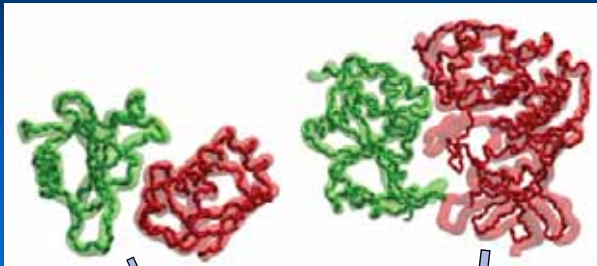
	Cases with near-native solutions (%)[a]	Number of near-native solutions[b]	Success rate top 10 (%)[c]
FTDock parameters			
ELE 0.7 Å	81	8.0	18 (22)
noELE 0.7 Å	85	8.5	16 (19)
ELE 1.2 Å	73	5.9	16 (22)
noELE 1.2 Å	77	6.4	13 (17)
Difficulty			
"Rigid-body"	89	8.8	24 (27)
Medium	62	5.6	0 (0)
Difficult	50	1.3	0 (0)
Unbound-bound RMSD (Å)			
<0.5	83	12.8	58 (70)
0.5–1.0	96	7.9	18 (19)
1.0–1.5	92	7.6	13 (14)
1.5–2.0	50	5.2	0 (0)
>2.0	38	1.3	0 (0)

Pons et al. (2010) Proteins 78, 95-108

The challenge of flexibility

Types of conformational changes in proteins

- **Large scale motions:**

Type of motion	Time Scale	Amplitude
Side chain motions (protein surface)		1-5 Å
Backbone motions in protein loop regions :		1-10 Å
Motions of the N- or C-terminus of a protein:		1-5 Å
Rigid body motions of secondary structures :		1-5 Å
Protein domain motions :		5-10 Å
(for example hinge bending motions)		
Allosteric transitions:		5-10 Å
(correlated motion of several subunits)		
Local folding and unfolding transitions		~5 Å
(helix-coil transitions, loop folding)		

(from McCammon & Harvey, Dynamics of proteins and nucleic acids, Cambridge University Press)
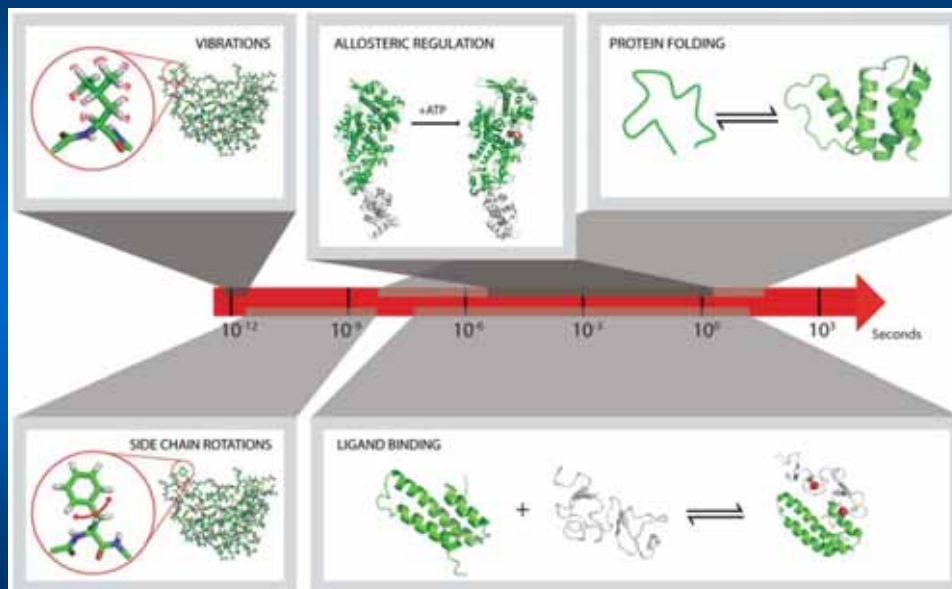
The challenge of flexibility

Types of conformational changes in proteins

- **Large scale motions:**

Type of motion	Time Scale	Amplitude
Side chain motions (protein surface)	0.1 ps- 0.1 ns	1-5 Å
Backbone motions in protein loop regions :	several ns	1-10 Å
Motions of the N- or C-terminus of a protein:	several ns	1-5 Å
Rigid body motions of secondary structures :	0.05 – 1 µs	1-5 Å
Protein domain motions :	1 µs – 1 ms	5-10 Å
(for example hinge bending motions)		
Allosteric transitions:	1 µs – 100 ms	5-10 Å
(correlated motion of several subunits)		
Local folding and unfolding transitions	0.1 µs – 10 ms	~5 Å
(helix-coil transitions, loop folding)		

(from McCammon & Harvey, Dynamics of proteins and nucleic acids, Cambridge University Press)

The challenge of flexibility

The challenge of flexibility

Types of conformational changes in proteins

- **Large scale motions:**

Type of motion	Time Scale	Amplitude
Side chain motions (protein surface)	0.1 ps- 0.1 ns	1-5 Å
Backbone motions in protein loop regions :	several ns	1-10 Å
Motions of the N- or C-terminus of a protein:	several ns	1-5 Å
Rigid body motions of secondary structures :	0.05 – 1 µs	1-5 Å
Protein domain motions :	1 µs – 1 ms	5-10 Å
(for example hinge bending motions)		
Allosteric transitions:	1 µs – 100 ms	5-10 Å
(correlated motion of several subunits)		
Local folding and unfolding transitions	0.1 µs – 10 ms	~5 Å
(helix-coil transitions, loop folding)		

(from McCammon & Harvey, Dynamics of proteins and nucleic acids, Cambridge University Press)

Rigid-body docking
Refinement
MD ensembles

Flexible docking

Ab initio folding

INDUCED FIT

CONFORMATIONAL SELECTION

**Docking +
Flexible refinement**

**Precomputed
unbound ensembles +
docking**

Multiple conformers: large-scale test
Protein docking benchmark 3.0 – 124 cases

MD, Modeller...

rank 1 — rank N

N_{rec} N_{lig}

Max(N_{rec} , N_{lig}) docking runs

clustering

Ensemble

Bound

Unbound

The challenge of weak interactions

- Bound
- Unbound
- Unbound (only RB)

Top 10 success rate

ΔG

>-7.5 $-7.5:-10.0$ $-10.0:-12.5$ $-12.5:-15.0$ <-15.0
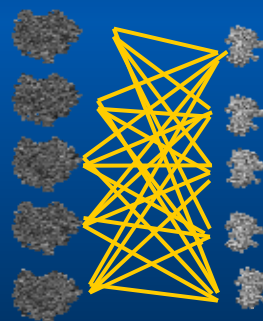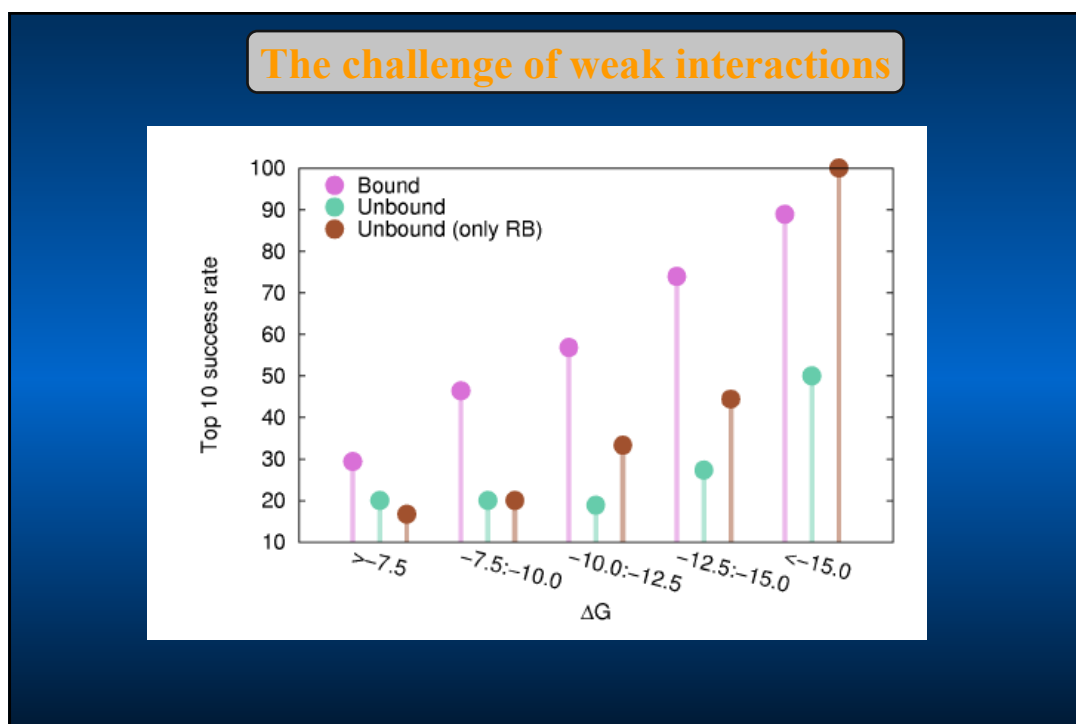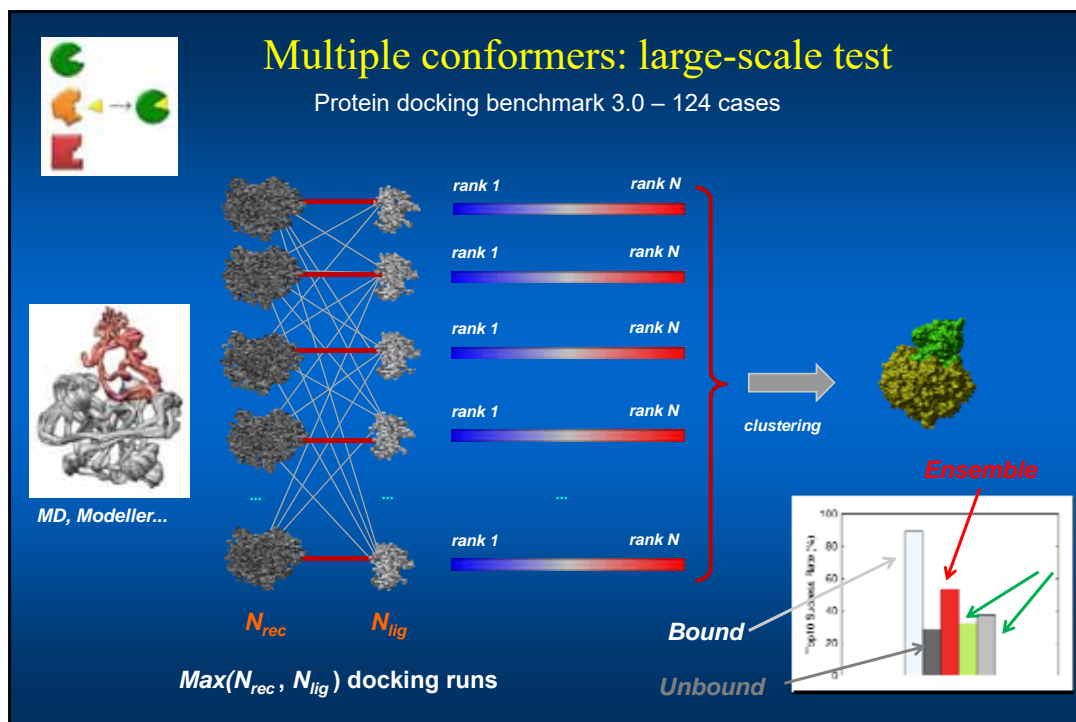
The challenge of weak interactions
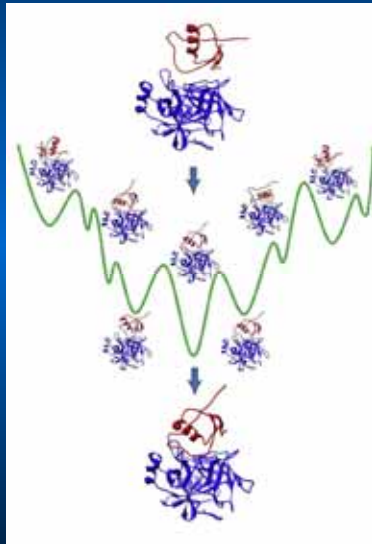
error
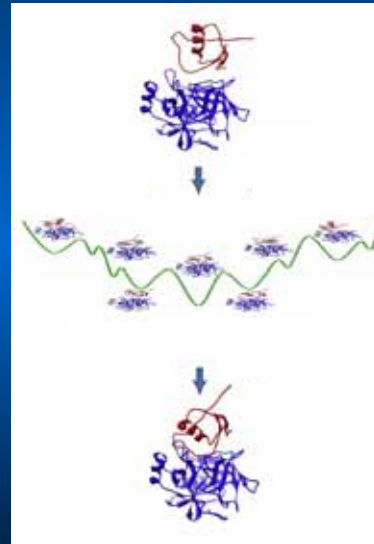
The challenge of weak interactions

Fd

FNR
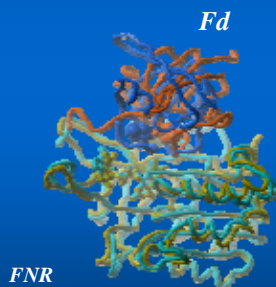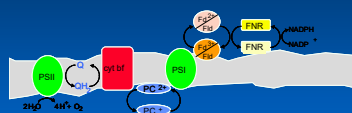
X-ray   *Anabaena (1ewy)*
             Maize (1gaq)
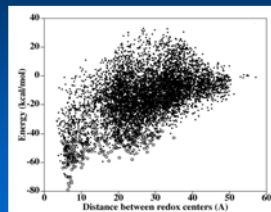
X-ray structure:
explains 82 % of mutant data

Docking model:
explains 100 % of mutant data
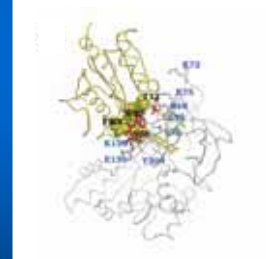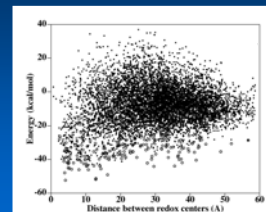
No X-ray structure

Docking model:
explains 96 % of mutant data

Medina, Abagyan, Gómez-Moreno, Fernández-Recio (2008) Proteins 72, 848

Understanding the binding mechanism

HPr / EIN
PRE data
*Tang et al. (2006)
Nature 444, 383*
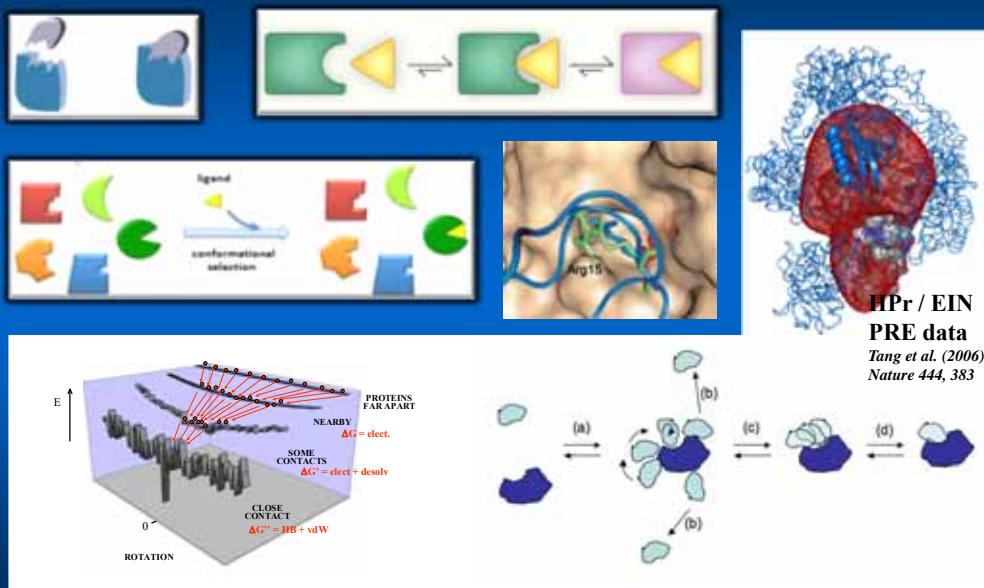
When there is no structure for interacting proteins
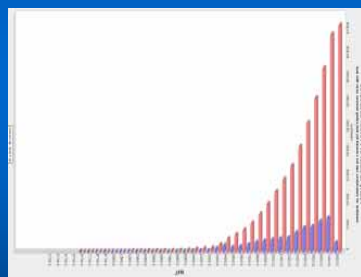
Human proteome
**20K estimated proteins
4K with known structure
+ 6K homologous structure >50%ID
+4K homologous structure 30-50%ID
(50-70% structural coverage)#**

 # Somody et al 2017 Drug Discov. Today 22, 1792
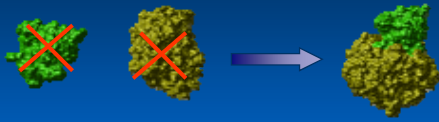
When there is no structure for interacting proteins

When there is no structure for interacting proteins

>30% seq ID

When there is no structure for interacting proteins

>30% seq ID

Modeller
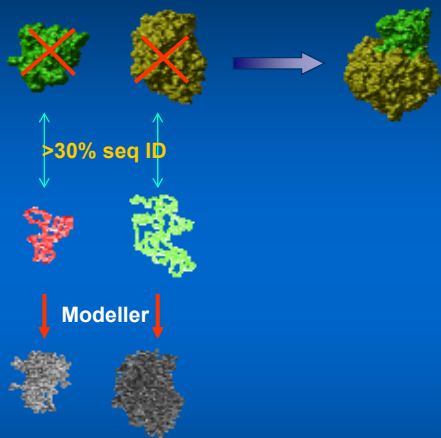
When there is no structure for interacting proteins

- Possible errors in target-template alignment
- Structural inaccuracies in segments with low sequence similarity
- Possible errors in modeled surface loops and side chains
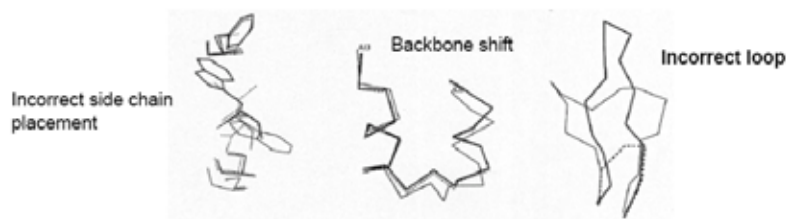
Incorrect side chain placement

Backbone shift

Incorrect loop

When there is no structure for interacting proteins

>30% seq ID

Modeller
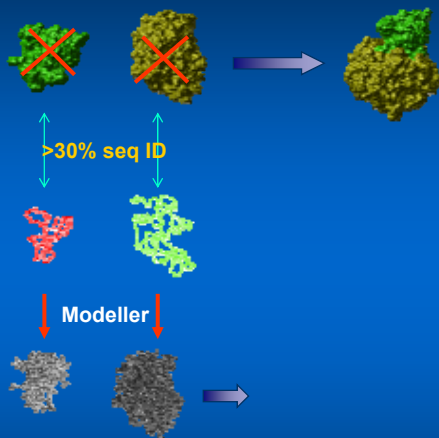
84 protein-protein complex structures

12 cases with PDB template 70-90% seq ID
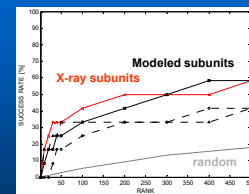
Modeled subunits

X-ray subunits

random

SUCCESS RATE [%]

RANK

When there is no structure for interacting proteins

>30% seq ID

Modeller

84 protein-protein complex structures

23 cases with PDB template 50-70% seq ID
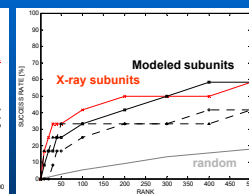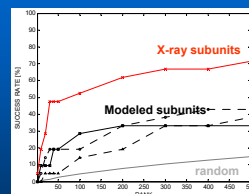
12 cases with PDB template 70-90% seq ID

X-ray subunits

Modeled subunits

random

Modeled subunits

X-ray subunits

random

When there is no structure for interacting proteins

>30% seq ID

Modeller

84 protein-protein complex structures

36 cases with PDB template 30-50% seq ID

23 cases with PDB template 50-70% seq ID

12 cases with PDB template 70-90% seq ID

Multi-protein complexes

Docking 1:1

11

Types of protein-protein interactions

PERMANENT / NON-OBLIGATORY
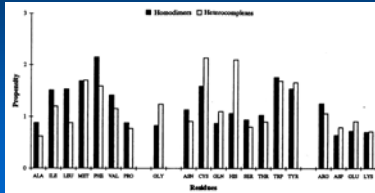
Jones & Thornton (1996) PNAS, 93, 13

HOMO- / HETERO-OBLIGOMERS / COMPLEXES

Type of interface	Number of contacts
External	
Homo-obligomers	218,104
Homo-complexes	3077
Hetero-obligomers	18,886
Hetero-complexes	166,412

Ofran & Rost (2003) JMB, 325, 377

HOMO- / HETERO-OLIGOMERIC NON-OBLIGATE / OBLIGATE TRANSIENT / PERMANENT
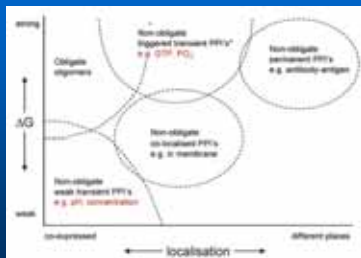
Nooren & Thornton (2003) EMBO J, 22, 3486

DOMAIN-DOMAIN INTERFACE CLASSIFICATION PERMANENT / TRANSIENT, SYMMETRIC...

Legend
symmetric homo — hetero — asymmetric homo
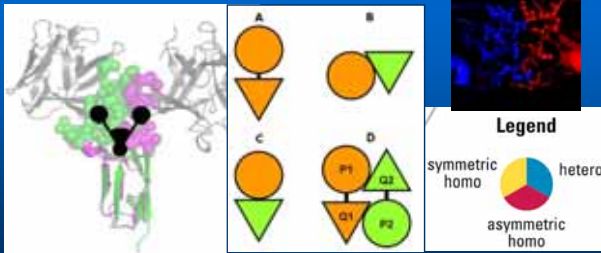
Kim et al. (2006) PLOS Comp Biol, 2, e124 (http://www.scoppi.org)

Types of protein-protein interactions

MULTI-MOLECULAR ASSEMBLIES

3Dcomplex, http://supfam.mrc-lmb.cam.ac.uk/elevy/3dcomplex/Home.cgi

Types of protein-protein interactions

PERIODIC TABLE OF PROTEIN COMPLEXES
Science 350, aaa2245 (2015)

Multi-protein complexes

Docking 1:1

Multi-protein
docking ??

Multi-protein complexes

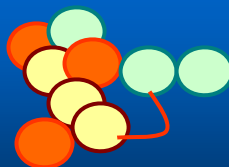PERIODIC TABLE OF PROTEIN COMPLEXES
Science 350, aaa2245 (2015)

Multi-protein complexes

Docking 1:1

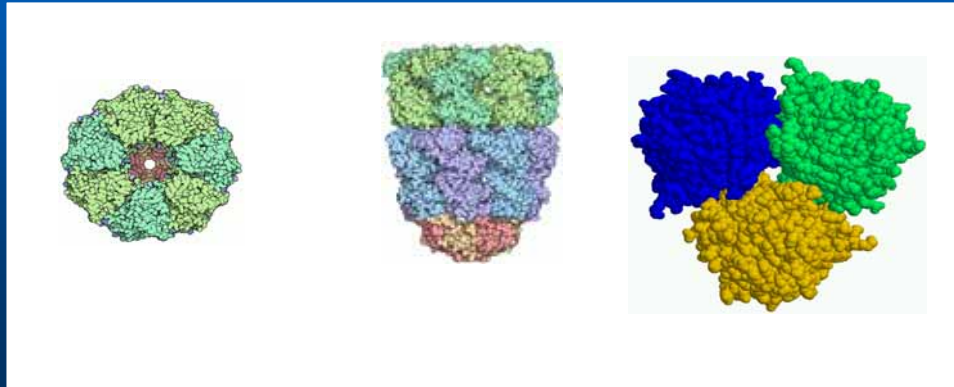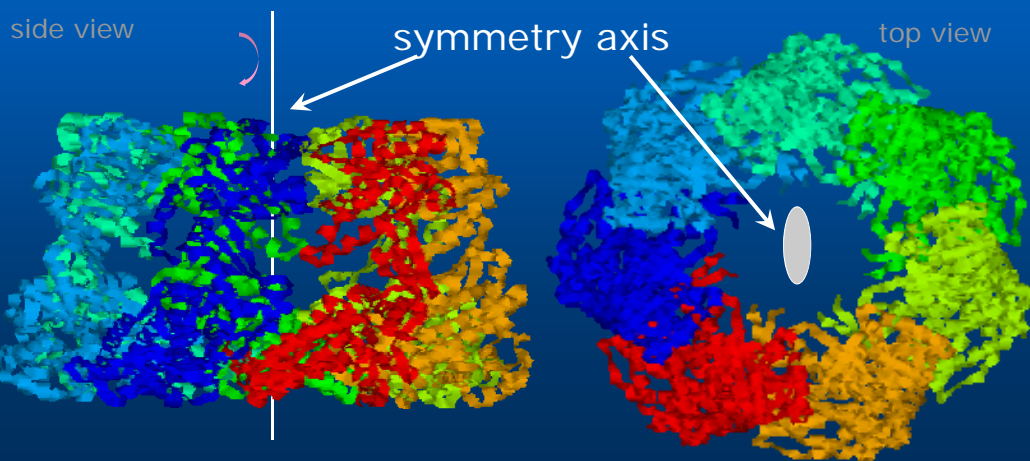**Homo-oligomerization**

Multi-protein docking ??

14

SymmDock

Cyclic Symmetry

- Cyclic symmetry is defined by rotation of a single unit around an **axis**.
- The angle is determined by a number of units **n**.

side view   symmetry axis   top view

CASP 12 - CAPRI Round 37

	available template	Predictors		Scorers	
		PyDock	Successful groups	PyDock	Successful groups
T110 Homotrimer	yes **19%**	**	22/25	**	16/16
T111 Homotrimer	yes **54%**	***	27/28	**	16/16
T112 Homotrimer	yes **23%**	*	21/26	*	14/16
T113 Heterodimer	no	0	6/27	0	2/16
T114 Homodimer	no	0	0/29	0	0/16
T116 Homodimer	yes **30%**	0	0/30	0	0/16
T117 Dimer of Heterodimer	no	0	1/27	*	2/16
T118 Homo octamer	yes **46%**	**	24/28	***	16/16
T119 Homodimer	yes **32%**	*	22/28	0	14/17
T120 Heterodimer	yes **35%**	*	20/28	**	5/15

Preliminary Evaluation
Courtesy of Marc Lensink

Multi-protein complexes

Docking 1:1

Homo-oligomerization

Multi-protein docking ??

Multi-protein complexes

PERIODIC TABLE OF PROTEIN COMPLEXES
Science 350, aaa2245 (2015)

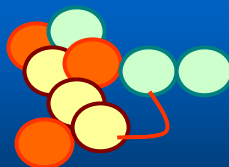# Multi-protein complexes
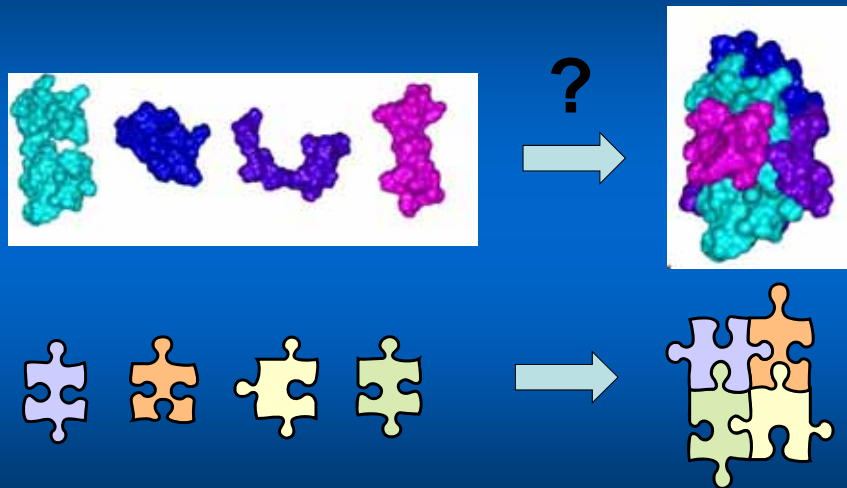

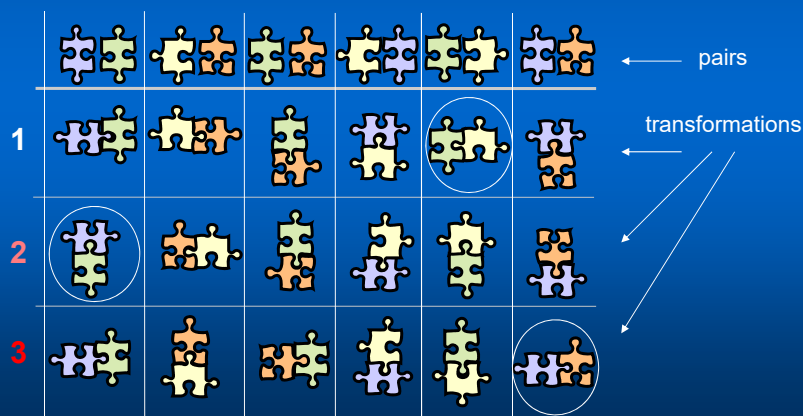Docking 1:1

Multi-protein assembly

A B C

Multi-protein
docking ??

Multi-protein docking: CombDock
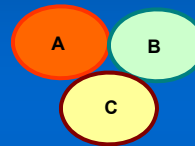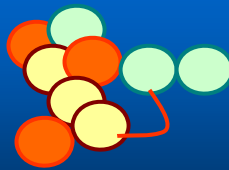
?

---

1. All Pairs Docking:

This stage outputs K possible orientations (transformations) between each pairs of building blocks. This is done by applying a geometric docking algorithm, and keeping the best K scoring solutions.
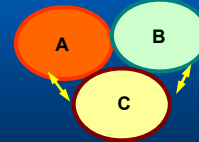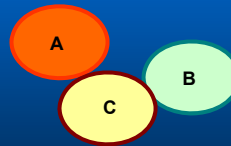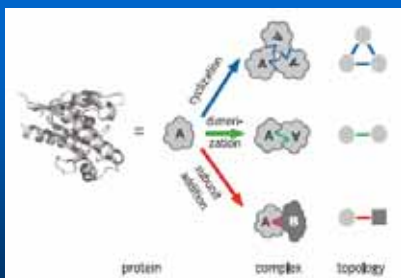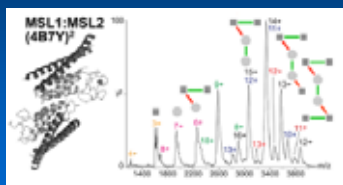

pairs

transformations

1

2

3

Multi-protein complexes

Docking 1:1

Multi-protein assembly

Multi-protein docking ??

Multi-protein complexes

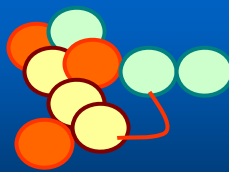*MULTI-MOLECULAR ASSEMBLIES*

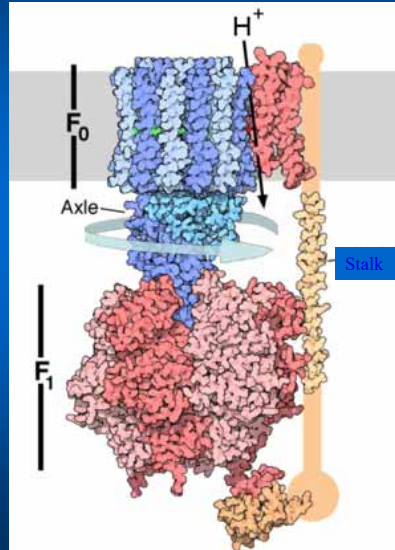PERIODIC TABLE OF PROTEIN COMPLEXES
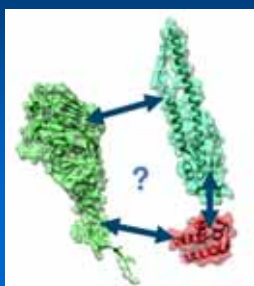Science 350, aaa2245 (2015)

Multi-protein complexes

Docking 1:1

Multi-protein docking ??

H+

F₀

Axle
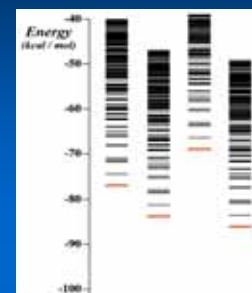
Stalk

F₁

Multi-protein complexes

Chemical X-linking
MS

A-A
B-B
A-B
B-C
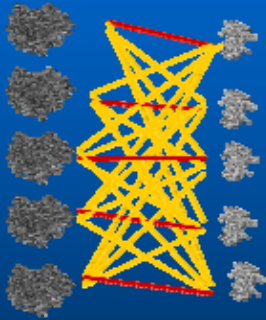Pairing and Stoichiometry

Docking

CombDock
SymmDock
M-Zdock

Mutational data
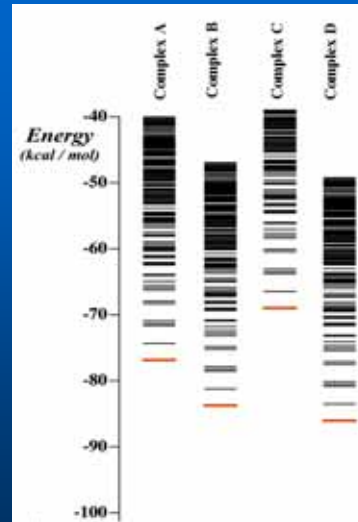
A   B

SAXS data

A   B   C
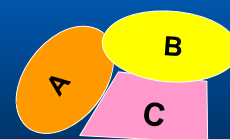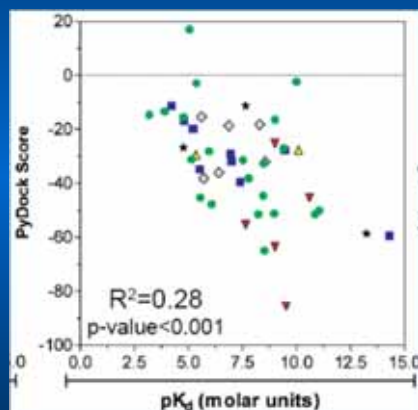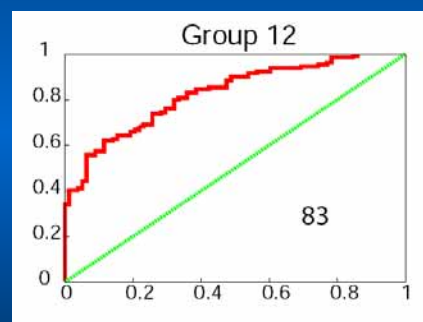
Identifying interactions

Identifying interactions

pyDock: best correlation with experimental affinities
Data from Kastritis & Bonvin 2010
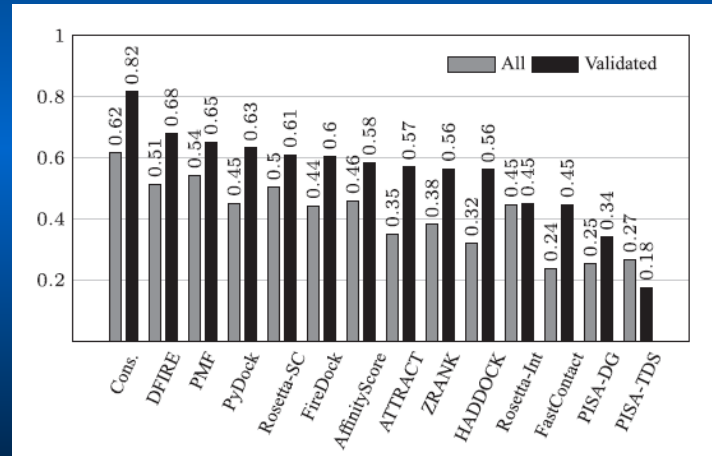*Benchmark 1.0 – 46 complexes*

CAPRI round 21

Fleishman et al. (JMB, 2011)

$R^2 = 0.28$
p-value<0.001

r = 0.762

Group 12
83

Isothermal Titration Calorietry
Surface Plasmon Resonance

Identifying interactions

Kastritis et al. 2011
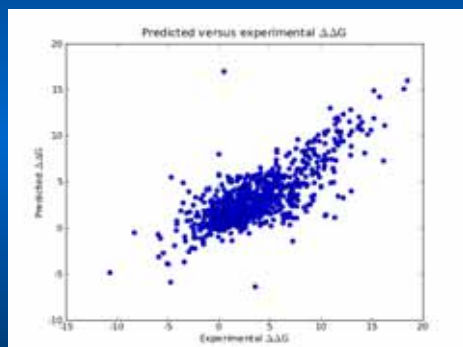*Benchmark 2.0 – 144 complexes*

57 validated experimental data
Moal et al. 2011 Bioinformatics

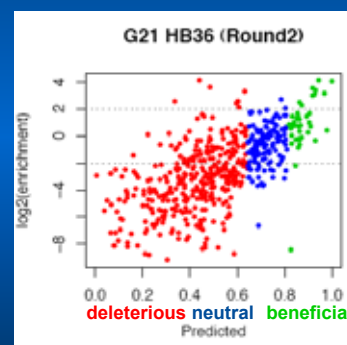*(Only from data published in Kastritis & Bonvin, 2010)*

Identifying interactions

Multi-parametric prediction of ΔΔG
Data from SKEMPI (http://life.bsc.es/pid/mutation_database/)
930 mutants

CAPRI T55-T56

r = 0.78

Top 1 - Top 1 - Top 3 - Top 3
(out of 22 groups)

Moretti et al. (2013) Proteins

Large scale docking

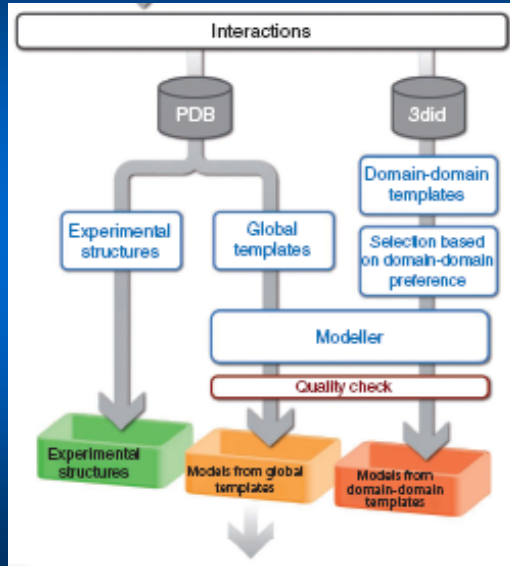**- Homology docking (template-based)**

- Ab initio docking (template-free)
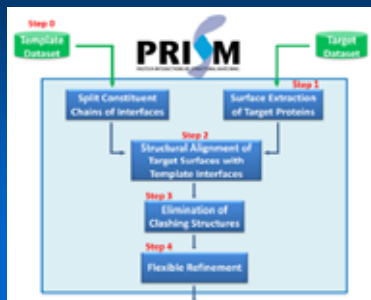
homology docking — sequence-based alignment

Interactome3D

3.2K can be homol. modelled
(seq. id > 30%)
(57% success rate)

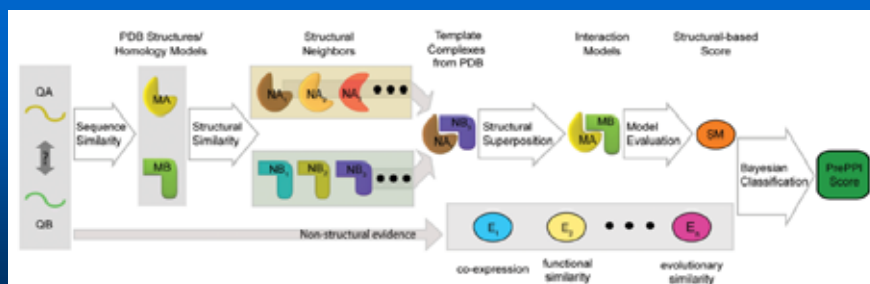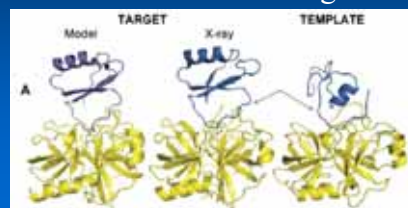**homology docking** — structure-based alignment
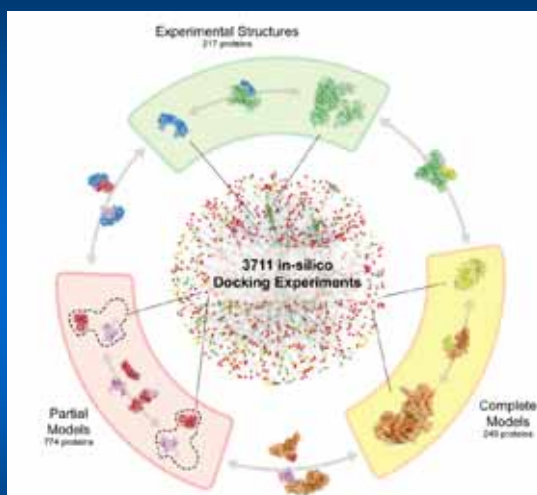
PRISM

Dockground

prePPI

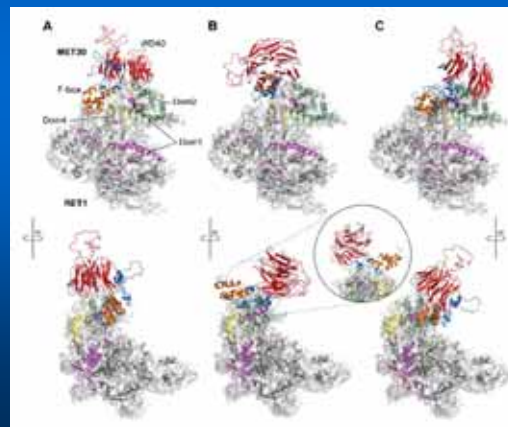24-300K?? could be homol. modelled (low success rate)

Large scale docking

- Homology docking (template-based)

- Ab initio docking (template-free)

Large scale docking

Mosca, Pons, Fernandez-Recio, Aloy (2009)
PLOS Comput. Biol. 5, e1000490

Large scale docking

- *11.1 Petaflops*
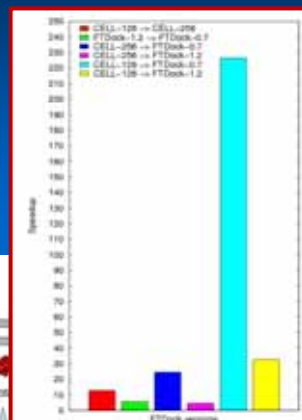- *>166,000 cores*
- *390 TB main memory*
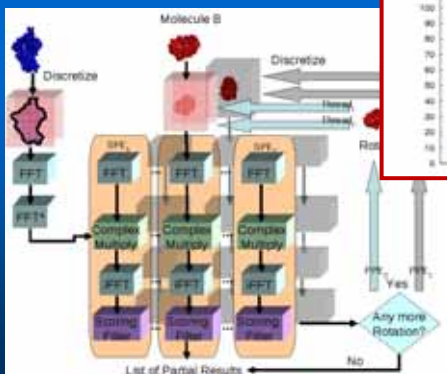
13th World
3rd Europe

Large scale docking

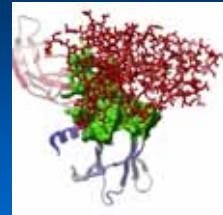FTDock on
Cell BE
~30 times
faster

FTDock on GPUs
~100 times faster

The ultimate challenge: discovery of PPI inhibitors

Major difficulties:

Small-molecule vs. Protein-protein interface

- Strategy: Targeting hot-spots.

- Hot-spot experimental identification:
  alanine-scanning (costly)

- Hot-spot computational prediction:
  ROBETTA, FOLDEF…
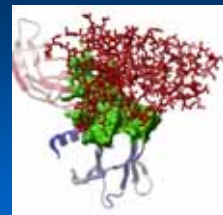 **(need 3D structure of complex)**

Absence of binding pockets in natural interfaces

- In standard drug design:
  3D structure with bound ligands,
  known active site

- In PPI inhibition:
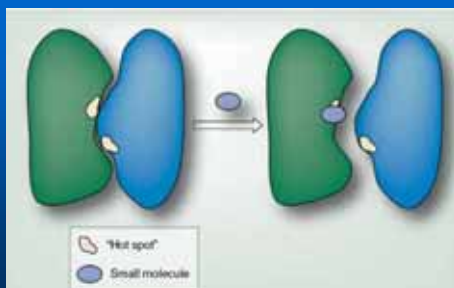 pocket finder, ligand docking and
  VLS
  (need 3D structure of complex)

The ultimate challenge: discovery of PPI inhibitors

Major difficulties:

Small-molecule vs. Protein-protein interface

Absence of binding pockets in natural interfaces

The ultimate challenge: discovery of PPI inhibitors

Current challenges in docking: a summary