Conclusions

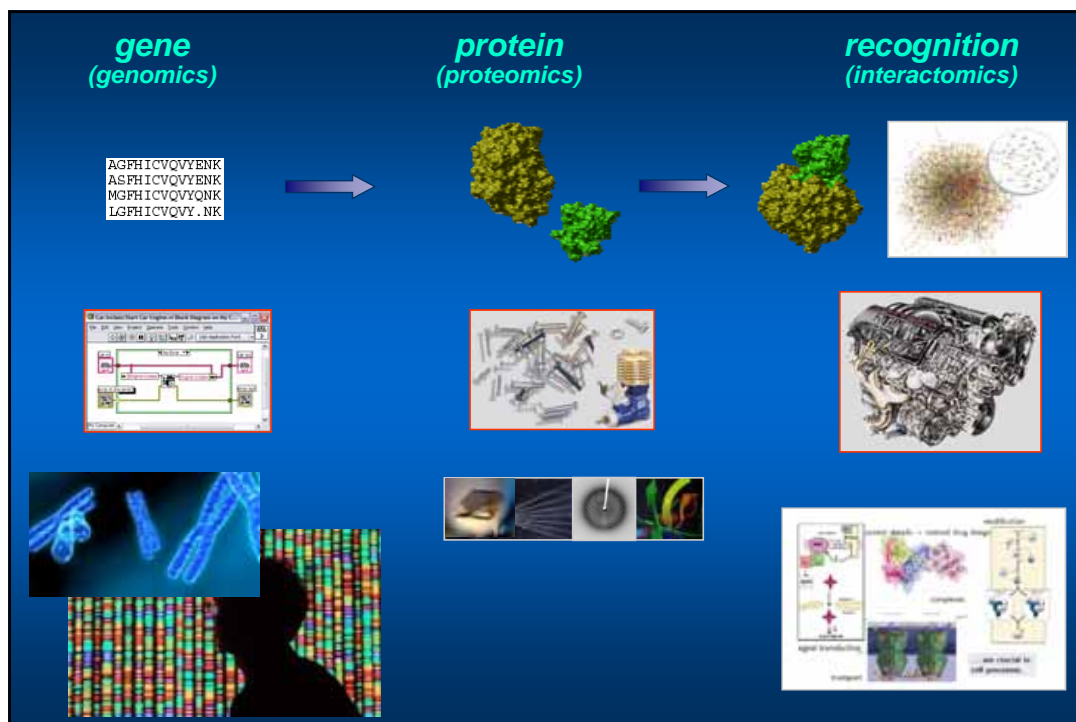# Protein-protein docking: a summary

Juan Fernández-Recio

CSIC, BSC

juanf@bsc.es

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

- **Future challenges in protein docking**

gene
(genomics)

protein
(proteomics)

recognition
(interactomics)

```
AGFHICVQVYENK
ASFHICVQVYENK
MGFHICVQVYQNK
LGFHICVQVY.NK
```

Databases of protein-protein interactions
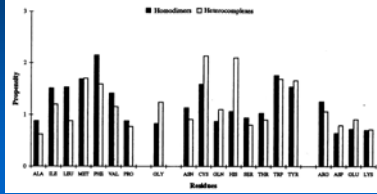
IMEx   http://www.imexconsortium.org/

Types of protein-protein interactions

PERMANENT / NON-OBLIGATORY

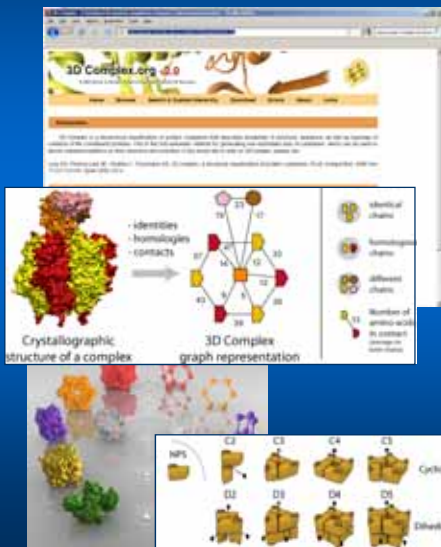*Jones & Thornton (1996) PNAS, 93, 13*

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

Databases of protein-protein complex structures

3Dcomplex

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

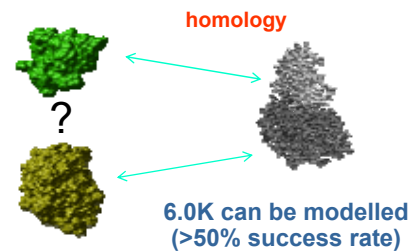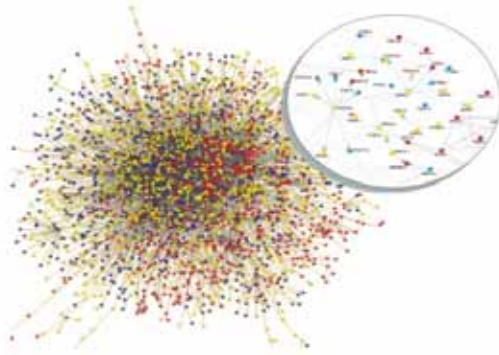### Interactome3D

http://interactome3d.irbbarcelona.org/

Structural coverage 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*

Complex structure prediction (docking)

Generation of the structure of a protein-protein complex
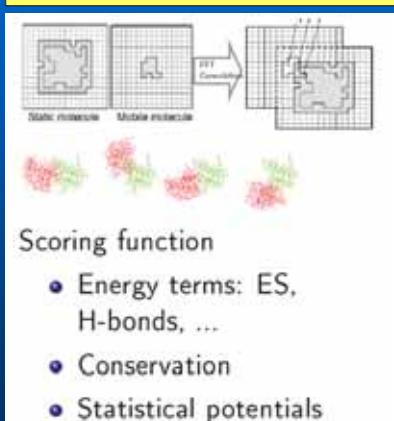*from the individual protein structures*
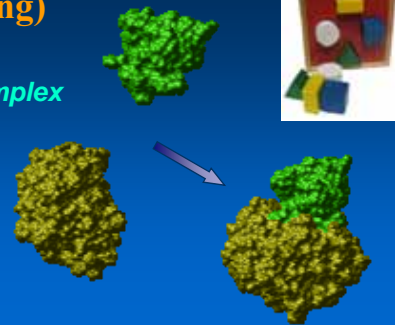
Motivation …

- *X-ray, NMR: Determination of complex structures remains difficult*

- *Low-resolution data on PPI available (cryo-EM, MS…)*

- *Understand energetics and mechanism of protein-protein association*

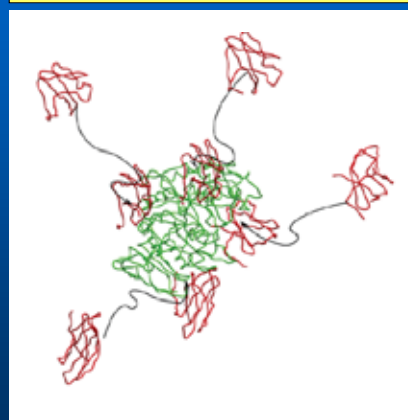- *Protein design (diagnostic, environment) and drug discovery*

4

SAMPLING & SCORING

SAMPLING & SCORING
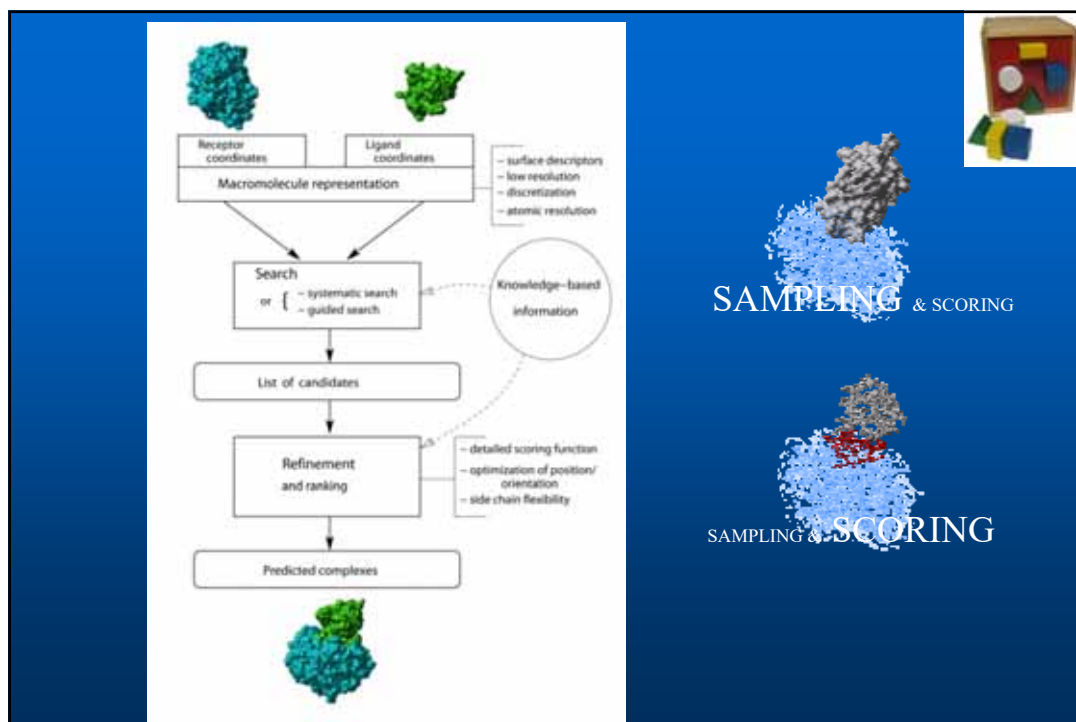
Protein-protein docking methods

Exhaustive search (FFT, surface-based)	Stochastic sampling (Monte-Carlo, minimization)

Scoring function
- Energy terms: ES, H-bonds, ...
- Conservation
- Statistical potentials

Geometry-based docking
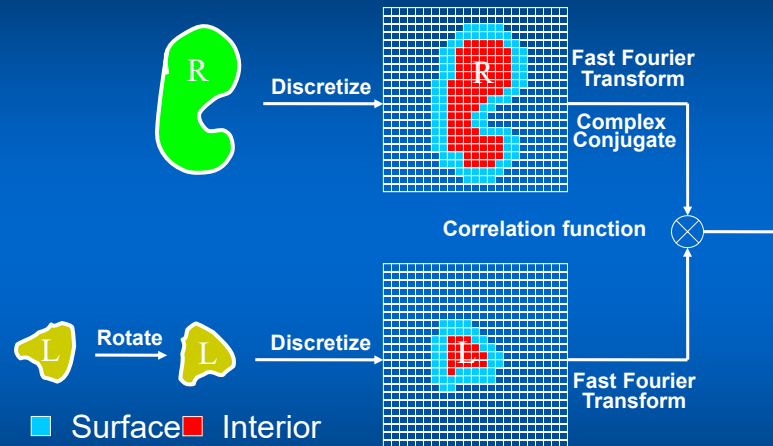
Energy-based docking

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

- **Future challenges in protein docking**

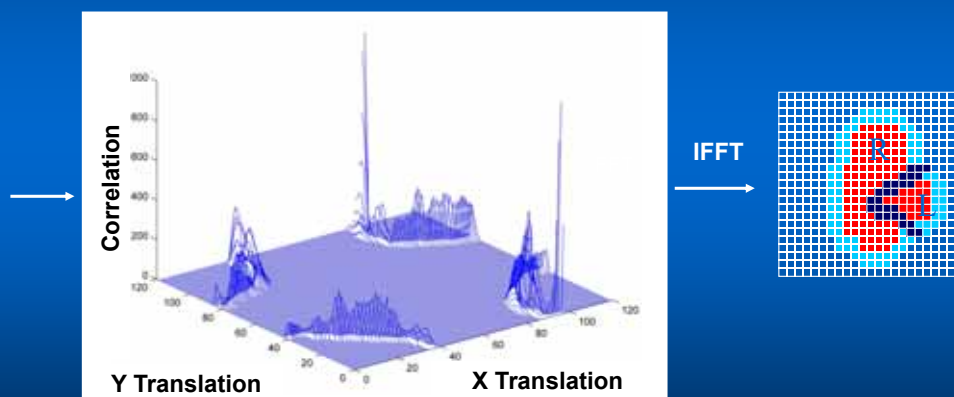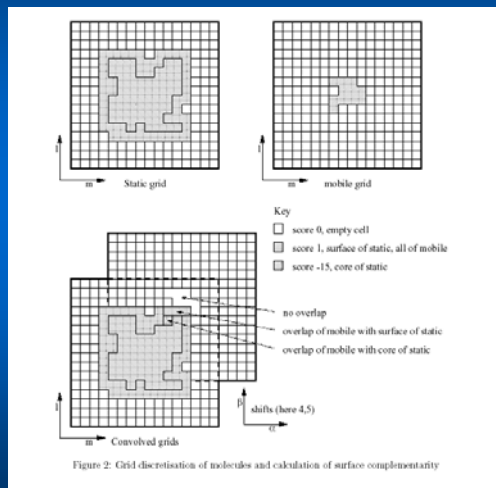# Rigid-Body Docking: Geometry Approach
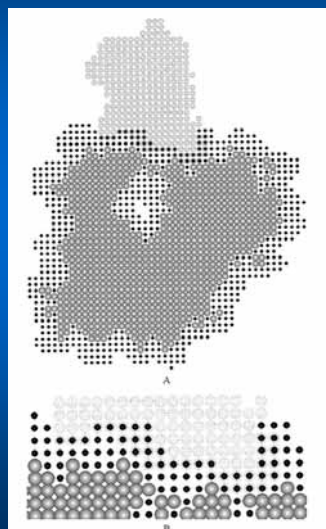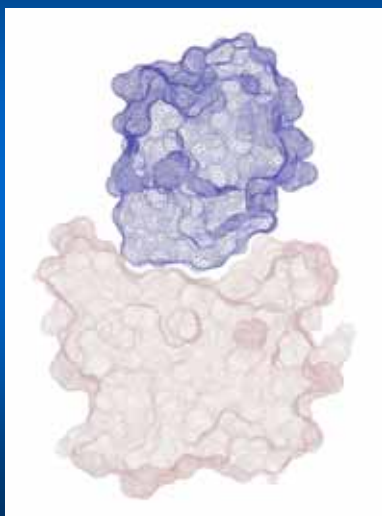
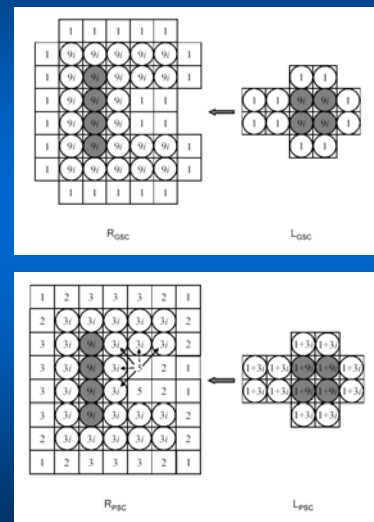# Protein Docking Using FFT

Protein Docking Using FFT

Comp. cost can decrease by >10^4 (from N^6 to $N^3 \ln N^3$)

FTDOCK

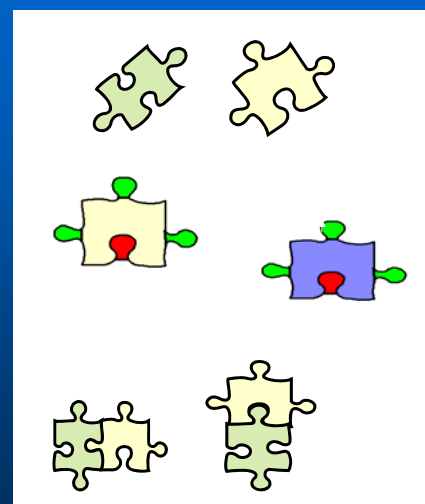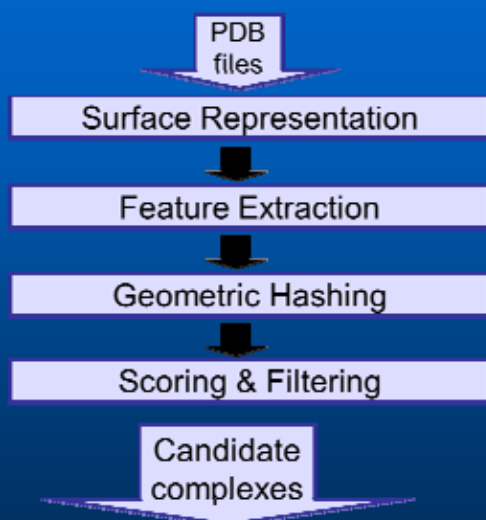Figure 2: Grid discretisation of molecules and calculation of surface complementarity
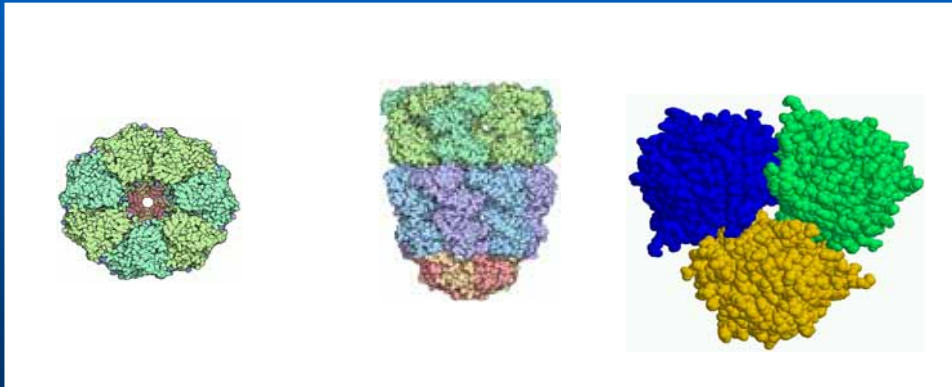
ZDOCK

Geometrical hashing - PatchDock

Schneidman-Duhovny et al. Proteins 2003
Duhovny (Schneidman), D., Nussinov, R. Wolfson,HJ.   WABI 2002

SymmDock

Symmetric Docking

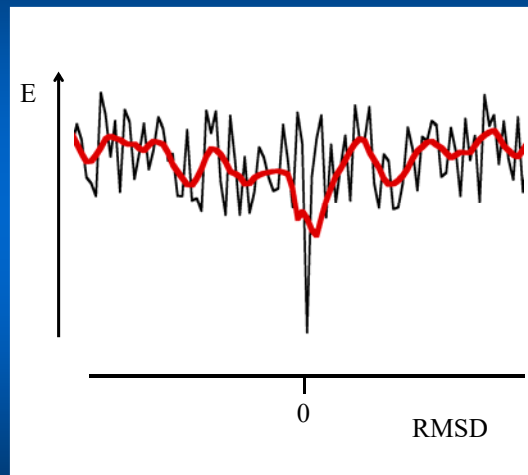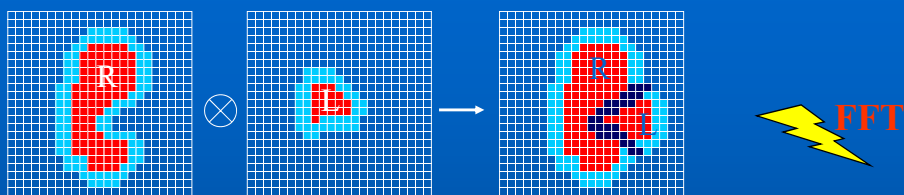Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ   Proteins 05

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

- **Future challenges in protein docking**

Protein-Protein Docking Energy

pyDock: scoring by energy

$$C(x,y,z) = \sum_{h=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(f_{h,k,l} \times g_{h+x,k+y,l+z}\right)$$

FFT
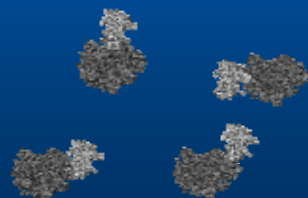
$$E = 0.1\sum_{i}^{REC}\sum_{j}^{LIG}\sqrt{e_i e_j}\left(\left(\frac{r_i+r_j}{d_{ij}}\right)^{12} - 2\left(\frac{r_i+r_j}{d_{ij}}\right)^{6}\right) +$$

$$+ 332.0\sum_{i}^{REC}\sum_{j}^{LIG}\frac{q_i q_j}{4 d_{ij}^2} +$$

$$+ \sum_{i}^{REC}ADP_i BSA_i + \sum_{j}^{LIG}ADP_j BSA_j$$

pyDock – Cheng, Blundell, Fernandez-Recio (2007) Proteins 68, 503-515

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

- **Future challenges in protein docking**

Protein-protein binding mechanism

1894: LOCK AND KEY: E. FISCHER

  - RIGID-BODY DOCKING

1958: *INDUCED FIT:* D. E. KOSHLAND

  - *small:* RIGID-BODY + REFINEMENT

  - *large:* FLEXIBLE DOCKING SEARCH

1999: CONFORMATIONAL SELECTION: R. NUSSINOV

  - PRECOMPUTED ENSEMBLES

Protein-protein binding mechanism

1894: LOCK AND KEY: E. FISCHER
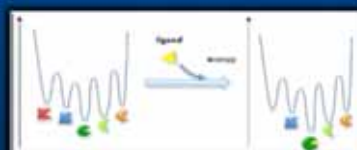
- RIGID-BODY DOCKING

1958: INDUCED FIT: D. E. KOSHLAND

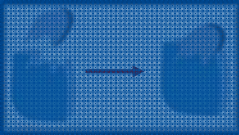- *small:* RIGID-BODY + REFINEMENT

- *large:* FLEXIBLE DOCKING SEARCH

1999: CONFORMATIONAL SELECTION: R. NUSSINOV

- PRECOMPUTED ENSEMBLES

Docking unbound trypsin/BPTI

Prot. Sci., 2002

Distribution of solutions

Real complex

Best docking solution

Docking complexed subunits

Docking unbound

Energy (kcal / mol)

-40
-50
-60
-70
-80
-90
-100

Correct solution

Refinement unbound trypsin/BPTI

Prot. Sci., 2002

Algorithm:
ICM-BP-SGO of side-chains
and positional variables

Real complex

Best docking solution

After refinement

Distribution of solutions

Docking complexed subunits

Docking unbound

Grid refinement

Energy (kcal / mol)

-40
-50
-60
-70
-80
-90
-100

Correct solution
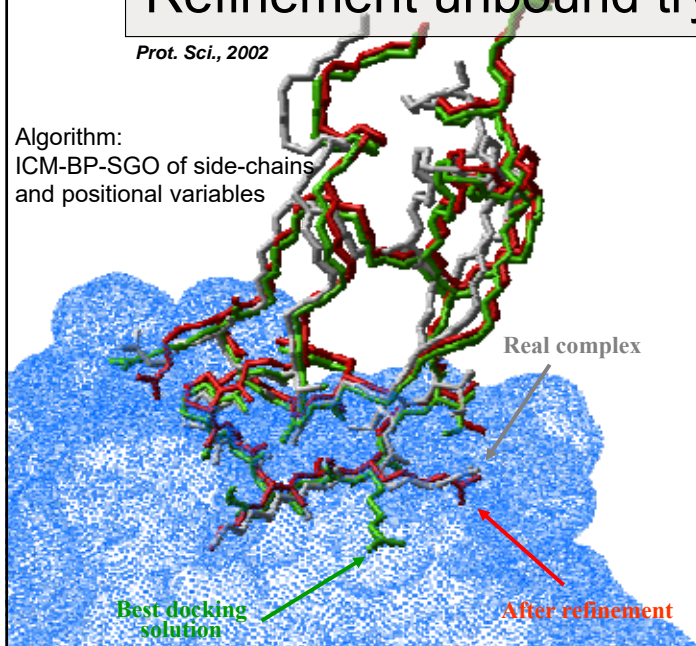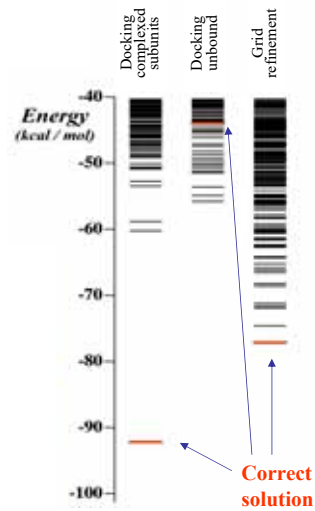
FireDock

Fast Interaction Refinement in Molecular Docking
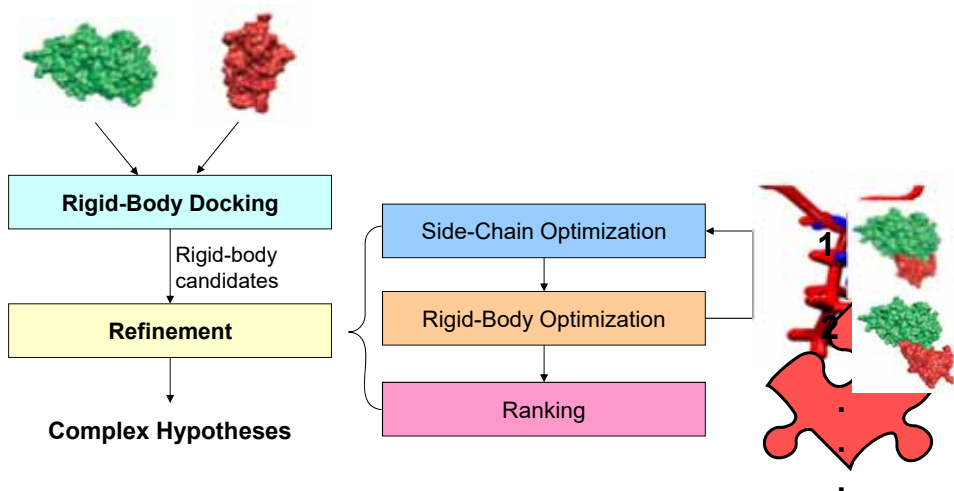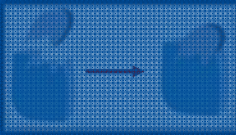
Rigid-Body Docking

Rigid-body candidates

Refinement

Complex Hypotheses

Side-Chain Optimization

Rigid-Body Optimization

Ranking

1
2
.
.
.

Protein-protein binding mechanism

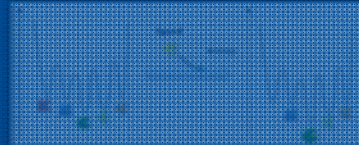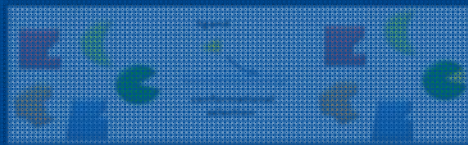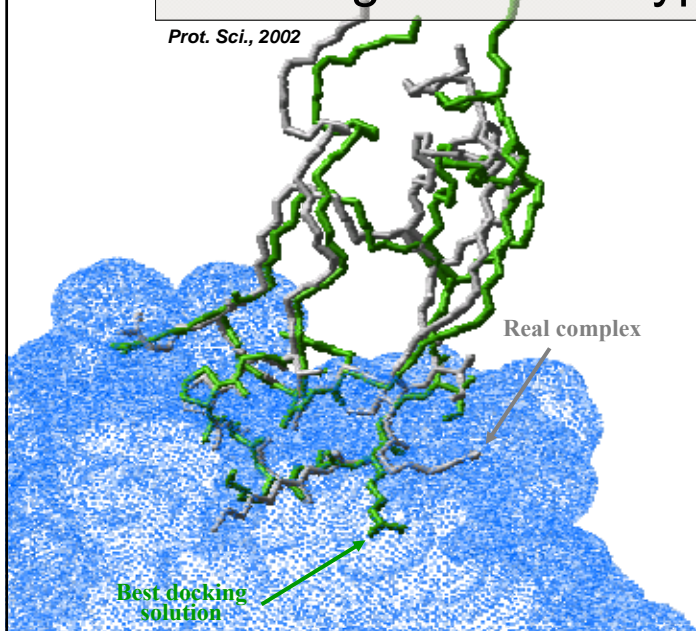*1894: LOCK AND KEY: E. FISCHER*

- RIGID-BODY DOCKING

1958: *INDUCED FIT:* D. E. KOSHLAND

- *small:* RIGID-BODY + REFINEMENT
- *large:* **FLEXIBLE DOCKING SEARCH**
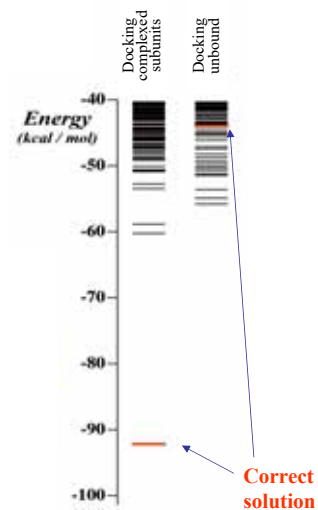
1999: CONFORMATIONAL SELECTION: R. NUSSINOV

- PRECOMPUTED ENSEMBLES

FlexDock

Detection of Hinges and Rigid Parts in the Flexible Molecule

Rigid Parts Docking via Geometric Hashing

Assembly of partial dockings into a flexible result

part1   part2   part3

part1   part3   part2   A   ...
A   part1

part1   part2   part3   A

Protein-protein binding mechanism

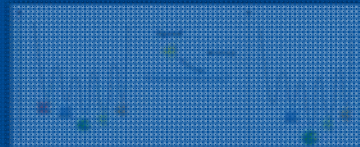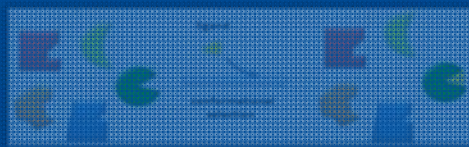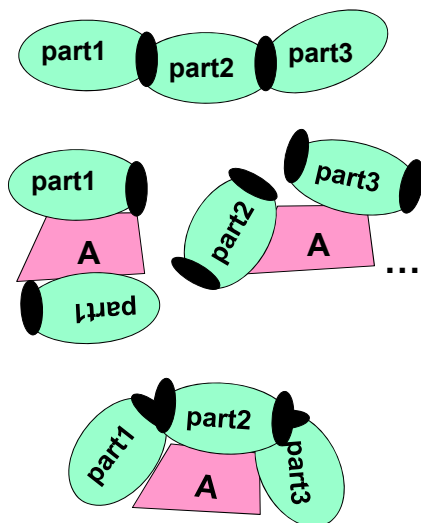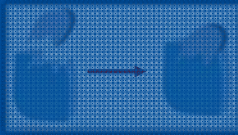*1894: LOCK AND KEY: E. FISCHER*

- RIGID-BODY DOCKING

1958: INDUCED FIT: D. E. KOSHLAND
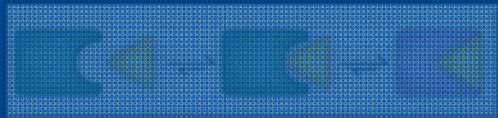
- *small:* RIGID-BODY + REFINEMENT

- *large:* FLEXIBLE DOCKING SEARCH

1999: CONFORMATIONAL SELECTION: R. NUSSINOV
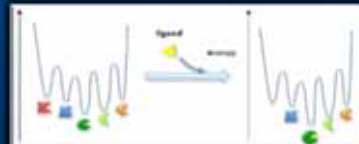
- PRECOMPUTED ENSEMBLES

Precomputed ensembles + docking

Docking with multiple side chain copies

- Surface side chains are re-presented by several sterically allowed rotamer copies.
- Selection of most favorable copies during docking using a meanfield or switching approach.

	- copies		+ copies	
Complex	Rank	Lrmsd$_A$	Rank	Lrmsd$_A$
Trypsin-BPTI	>120	7.1	11	1.4
Kallikrein-BPTI	>160	3.4	31	2.4
Chymo-BPTI	>150	3.4	29	3.2
SubtiN/Chy-Inh	>1100	3.8	59	3.0

ligand protein (atomic model)

multiple copies

ligand protein (reduced model)

receptor protein
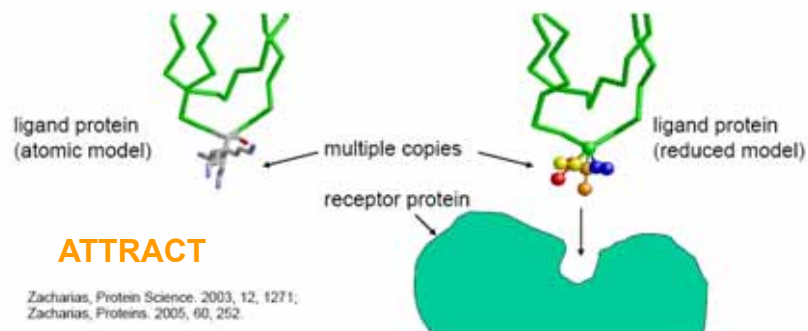
ATTRACT

Zacharias, Protein Science. 2003, 12, 1271;
Zacharias, Proteins. 2005, 60, 252.

15
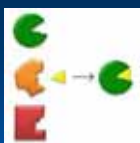
Multiple conformers: ubiquitine ensemble

RDC-based ensembles
(Fenwick et al. JACS 2011, 133, 10366)

11 complexes
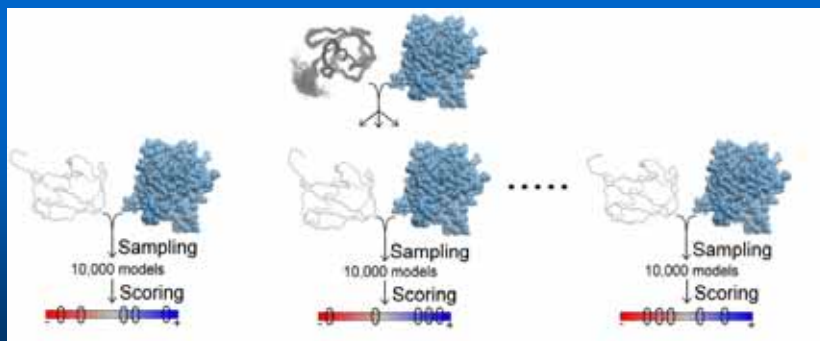
x 320

Pons et al. JCTC 2013

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

- **Future challenges in protein docking**

Protein-protein benchmarking

2002-2018 **Weng's benchmark series**

N test cases

0.0: 54 cases
1.0: 59 cases
2.0: 84 cases
3.0: 124 cases
4.0: 176 cases
5.0: 230 cases
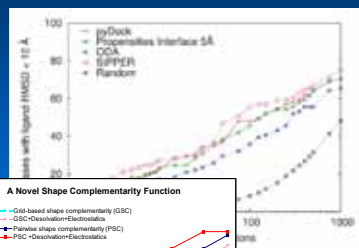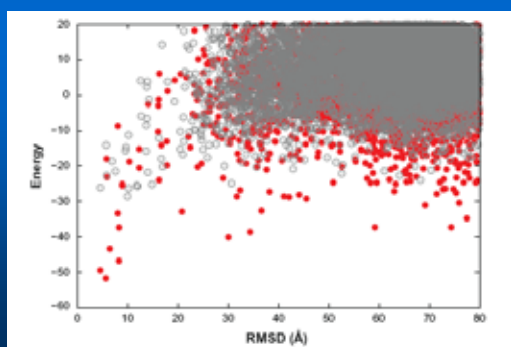
1 test case

http://zlab.umassmed.edu/benchmark/

DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

T01
Hpr (*unbound*)
HPr kinase (*unbound*)

T02
VP6 (*unbound*)
Fab (*bound*)

T03
Hemagglutinin (*unbound*)
Fab (*bound*)

T04, T05, T06
α-amylase (*unbound*)
VHH (*bound*)

T07
TCRβ (*unbound*)
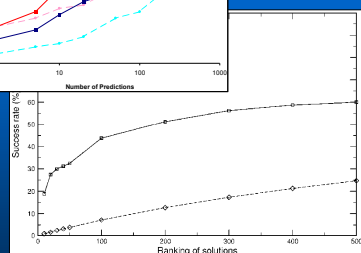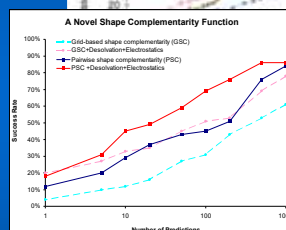speA (*unbound*)

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

1st CAPRI – Sep02 La Londe (France)
*PROTEINS: Structure, Function, and Genetics 52
(July 2003)*

2nd CAPRI – Dec04 Gaeta (Italy)
*PROTEINS: Structure, Function, and Bioinformatics 60
(July 2005)*

3rd CAPRI – Apr07 Toronto (Canada)
*PROTEINS: Structure, Function, and Bioinformatics 69
(December 2007)*

4th CAPRI – Dec10 Barcelona (Spain)
*PROTEINS: Structure, Function, and Bioinformatics 78
(November 2010)*

5th CAPRI – Apr13 Utrecht (Netherlands)
*PROTEINS: Structure, Function, and Bioinformatics 81
(December 2013)*

6th CAPRI – Apr16 Tel Aviv (Israel)

DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

- F_{nat}: fraction of native contacts (within 5Å)

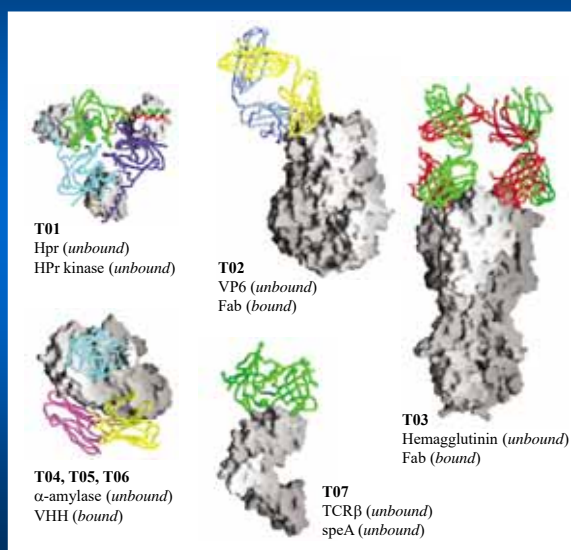$$F_{nat} = \frac{\text{Correctly predicted contacts}}{\text{Total number of contacts in the target}}$$

- I-RMSD: RMSD on second protein after superposition on first

- i-RMSD: RMSD on interface residues (within 10 Å)
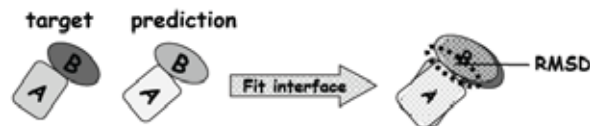
DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

Rank	F_{native}	AND (I-RMSD [Å]	OR i-RMSD [Å])
High ***	> 0.5	x < 1.0	or x < 1.0
Good **	> 0.3	1 < x < 5	or 1 < x < 2
Acceptable *	> 0.1	5 < x < 10	or 2 < x < 4
Incorrect	< 0.1		

- F_{native}: fraction of native contacts (within 5Å)
- I-RMSD: rmsd on second protein after superposition on first
- i-RMSD: rmsd on interface residues (within 10Å)

18

6th CAPRI

*Courtesy of
Marc Lensink*

| Predictor	Participation	Rank
Guerois	18	10/1***/8**
Zacharias	16	10/3***/2**
Vajda/Kozakov, Seok	17, 18	8/3***/2**
Weng	17	6/1***/4**
Fernandez-Recio	18	7/1***/3**
Vakser	18	6/2***/2**
Eisenstein	9	4/2***/2**
Zou	18	7/1***/2**
Bates	18	6/3**
Huang	18	5/3***
Zhou	16	4/2***/1**
Grudinin	18	4/3**
Bradley	5	3***
Shen	17	6/1***/1**
Baker	7	5/2**
Bonvin	18	4/1***/1**
Gray, Kihara	12, 18	3/2**
Furman	6	3/1***/1**
Takeda-Shitaka	2	2***
S_Liang, Di Maio, Moal, Negi	2, 2, 3, 10	2/1***/1**
... **Total 62 groups**		

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

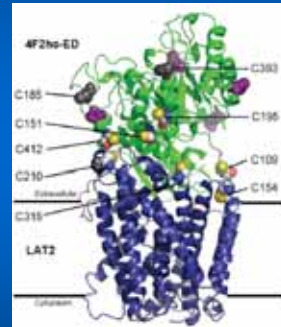- **Future challenges in protein docking**

Complex Structure Prediction by Docking: Examples

SidD / Rab1 complex
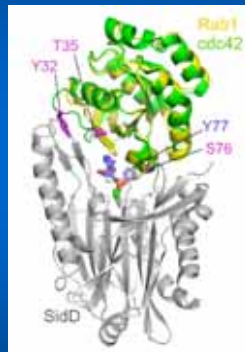
Chen et al. (2013) PLOS Pathogens
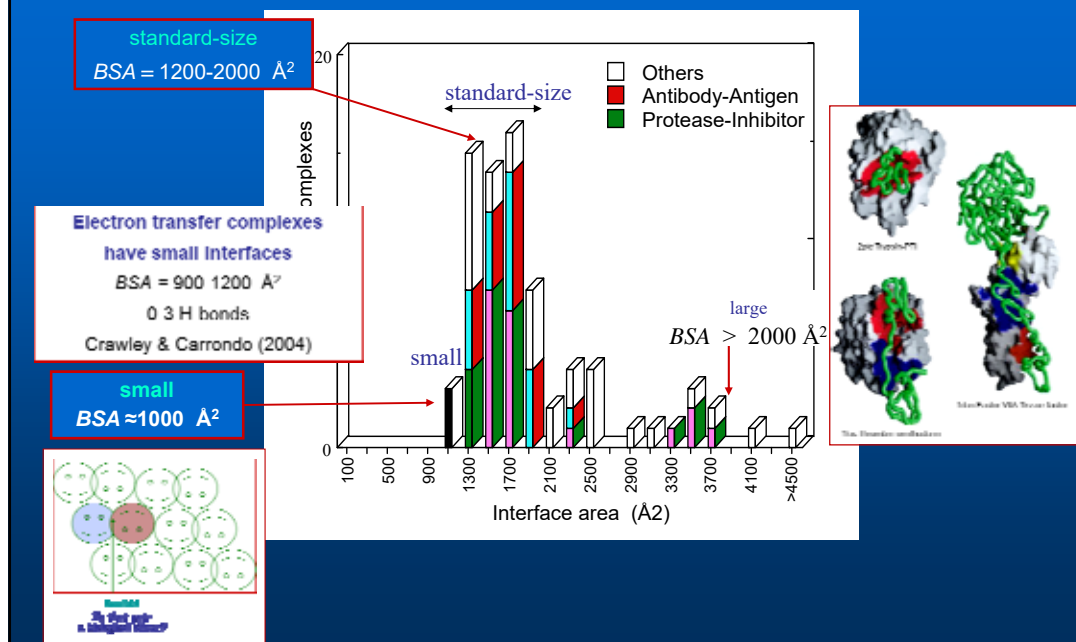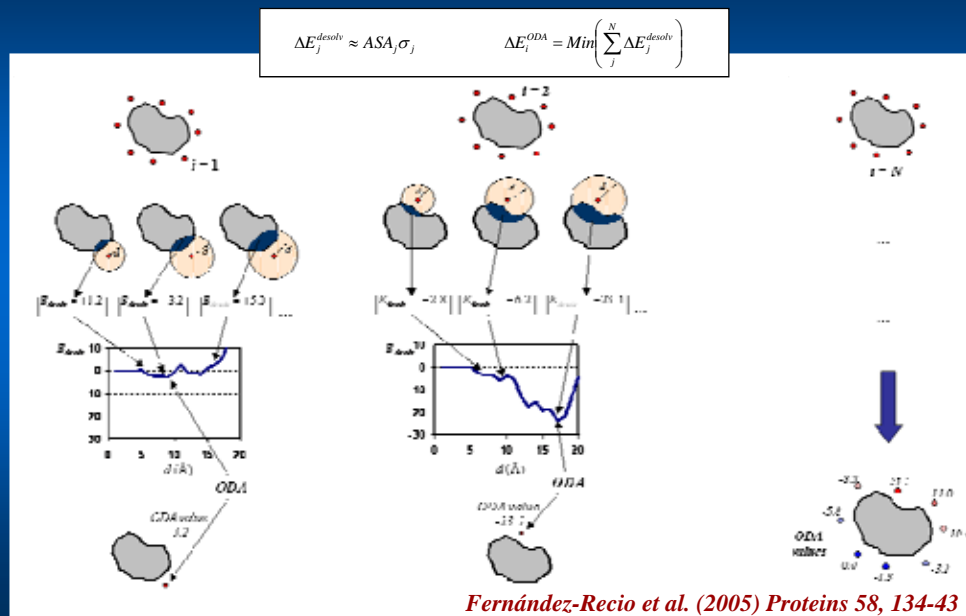
LAT2 / 4F2hc complex

Rosell et al. (2014) PNAS

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

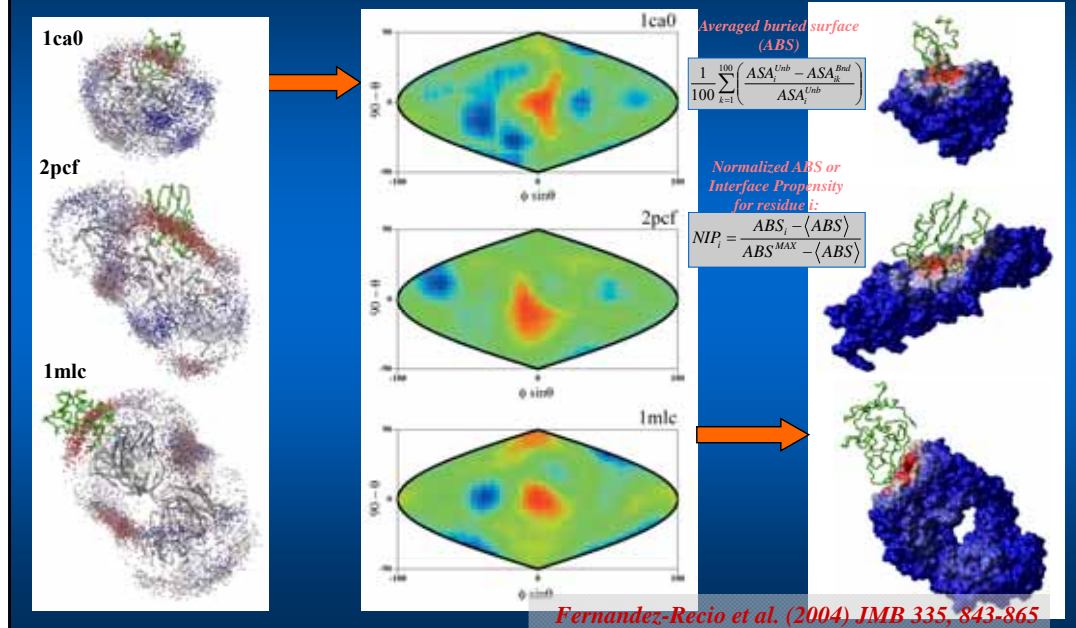- **Hot-spot identification**

- **Future challenges in protein docking**

Protein-protein interface size

standard-size
$BSA = 1200\text{-}2000\ \text{Å}^2$

Electron transfer complexes have small interfaces
$BSA = 900\ 1200\ \text{Å}^2$
0 3 H bonds
Crawley & Carrondo (2004)

small
$BSA \approx 1000\ \text{Å}^2$

Others
Antibody-Antigen
Protease-Inhibitor

standard-size

small

large
$BSA > 2000\ \text{Å}^2$

Interface area (Å2)

Optimal Docking Areas (ODA)

$$\Delta E_j^{desolv} \approx ASA_j \sigma_j \qquad \Delta E_i^{ODA} = Min\left(\sum_j^N \Delta E_j^{desolv}\right)$$

Fernández-Recio et al. (2005) Proteins 58, 134-43

Interface predictions from docking results

1ca0

2pcf

1mlc

Averaged buried surface (ABS)

$$\frac{1}{100}\sum_{k=1}^{100}\left(\frac{ASA_i^{Unb}-ASA_{ik}^{Bnd}}{ASA_i^{Unb}}\right)$$

Normalized ABS or Interface Propensity for residue i:

$$NIP_i = \frac{ABS_i - \langle ABS \rangle}{ABS^{MAX} - \langle ABS \rangle}$$

Fernandez-Recio et al. (2004) JMB 335, 843-865

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

- **Future challenges in protein docking**

Hot-Spots

The **O-ring model**: alanine-scanning <u>hotspots</u> tend to be located at the center of the interface and surrounded by energetically unimportant residues that occlude solvent from them.
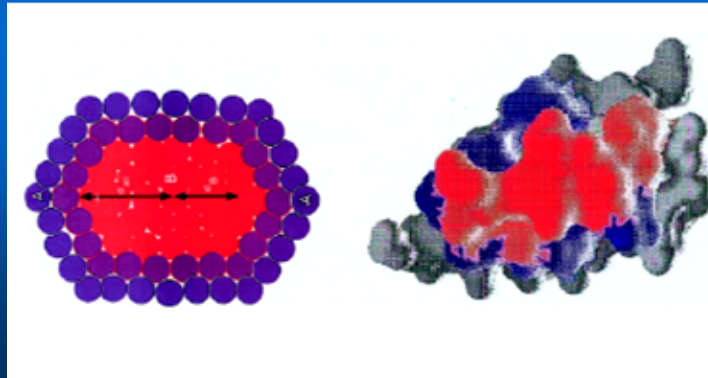
Bogan & Thorn (1998)

A sketch of the **solvent accessible** and **fully buried** atoms in a standard-size protein-protein interface. Each protein contributes about 70 atoms, and 1/3 are buried.
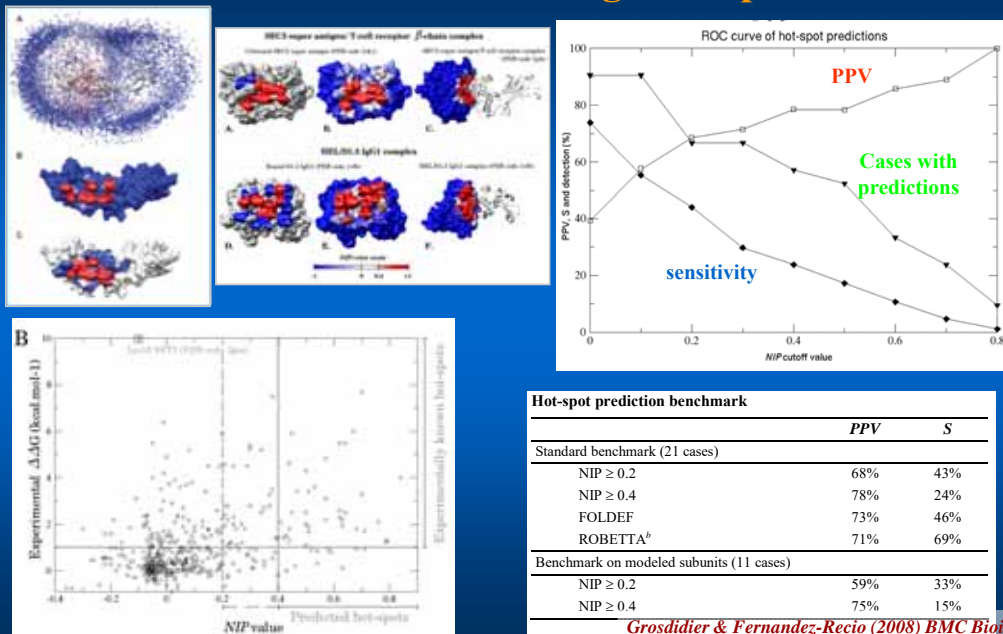
Lo Conte, Chothia & Janin (1999)

The **core/rim model** is a realistic implementation of that sketch. (Chakrabarti & Janin 2002)

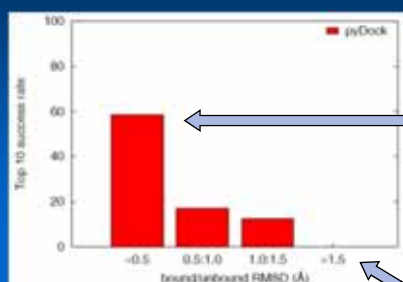The **core** and rim of the CI2 inhibitor interface with subtilisin (1sni)
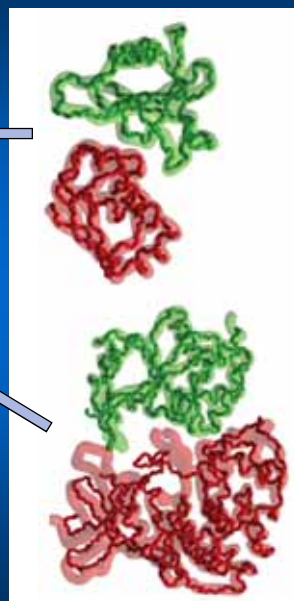

Prediction of Binding Hot-Spots

PPV

Cases with predictions

sensitivity

Hot-spot prediction benchmark

	PPV	S
Standard benchmark (21 cases)		
NIP ≥ 0.2	68%	43%
NIP ≥ 0.4	78%	24%
FOLDEF	73%	46%
ROBETTA[b]	71%	69%
Benchmark on modeled subunits (11 cases)		
NIP ≥ 0.2	59%	33%
NIP ≥ 0.4	75%	15%

Grosdidier & Fernandez-Recio (2008) BMC Bioinf

- **Introduction to protein interactions and docking**

- **Geometry-based docking**

- **Energy-based docking and scoring**

- **Flexibility in docking**

- **Benchmarking**

- **Applications to cases of biomedical interest**

- **Interface prediction**

- **Hot-spot identification**

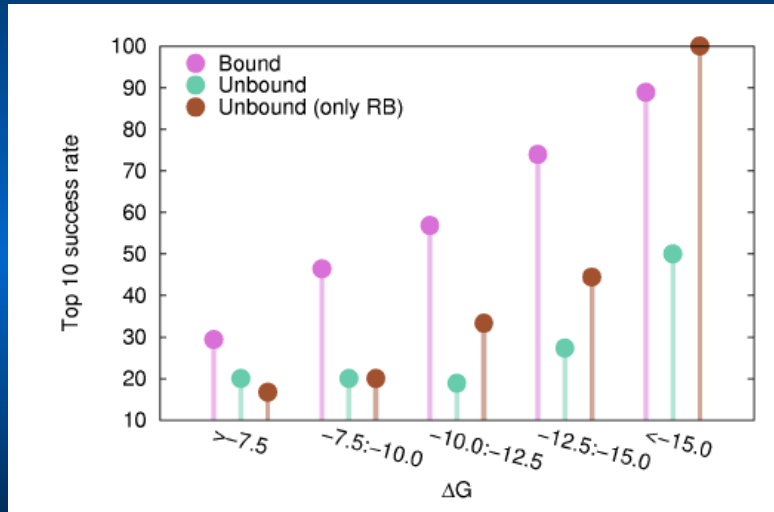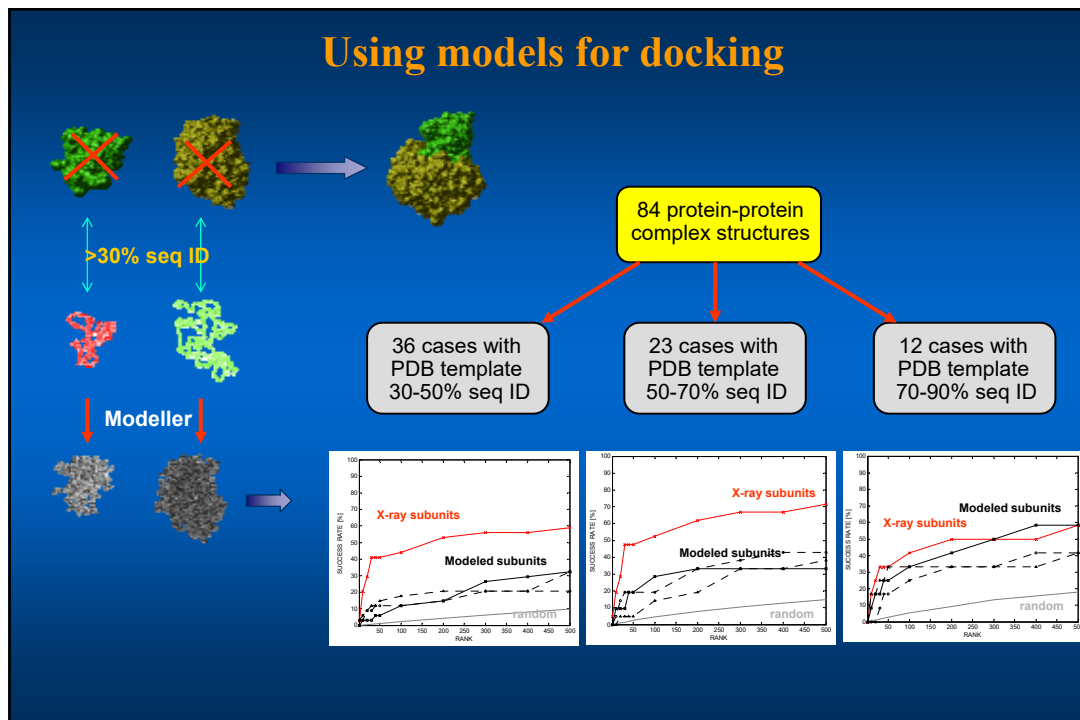- **Future challenges in protein docking**

The challenge of flexibility
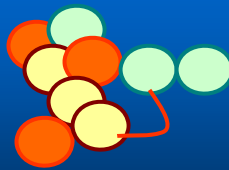
The challenge of weak interactions
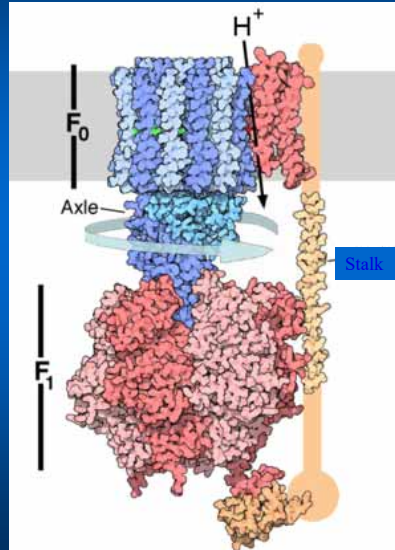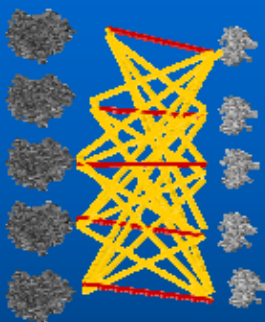

Using models for docking
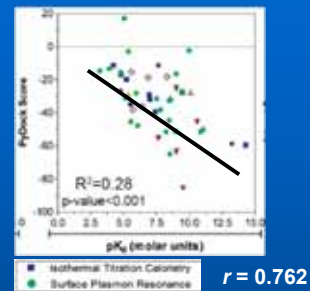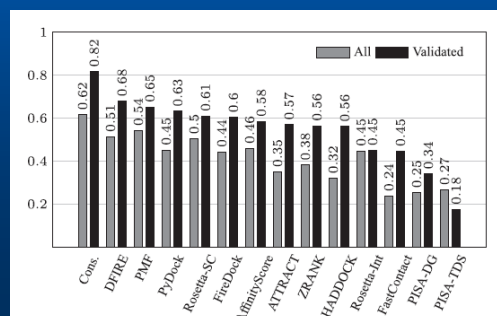
Multi-protein complexes

Docking 1:1

Multi-protein
docking ??

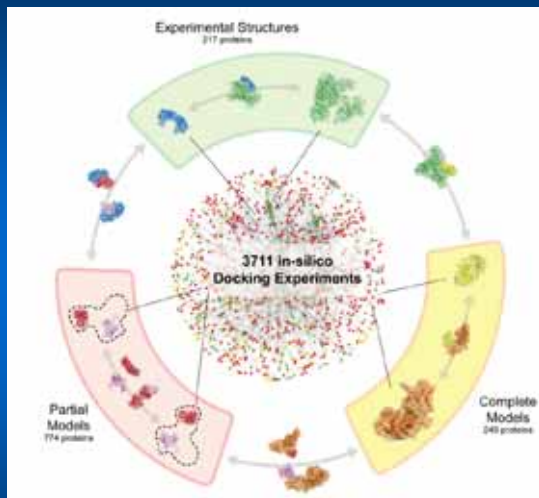**Identifying interactions**

?

r = 0.762

Significant correlation of
docking scoring functions with
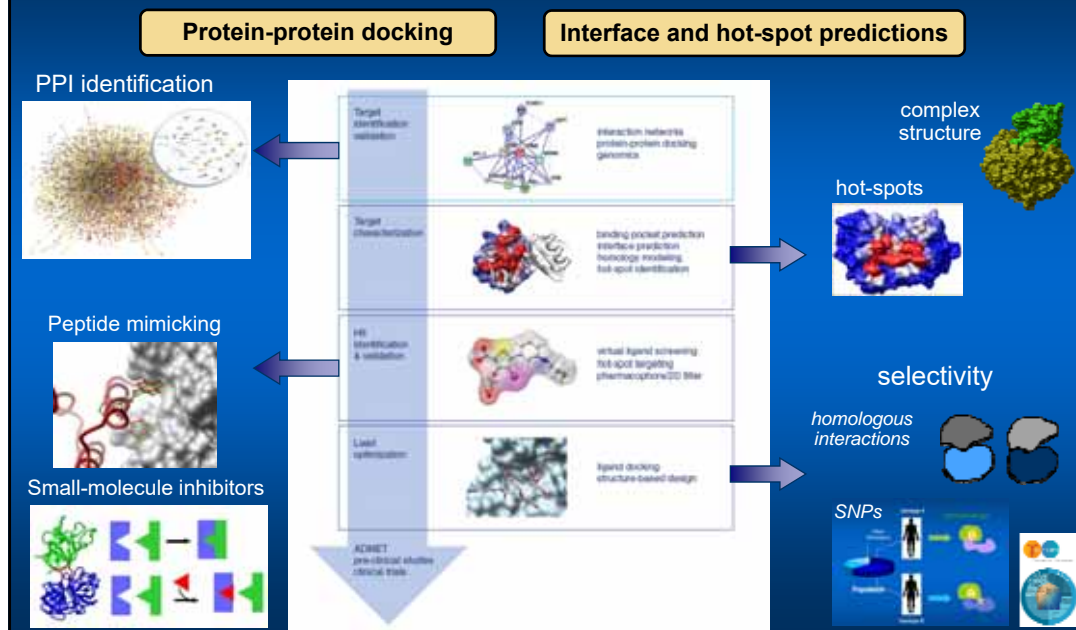experimental binding energies

Large scale docking

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

The ultimate challenge: discovery of PPI inhibitors

Protein-protein docking   **Interface and hot-spot predictions**

PPI identification

complex structure

hot-spots

Peptide mimicking

selectivity

homologous interactions

Small-molecule inhibitors

SNPs

Bibliography

Protein-Protein Interaction - General
- *Protein-Protein Complexes, M. Zacharias ed., Imperial College Press*
- *Protein-Protein Recognition, C. Kleanthous ed., Oxford University Press*
- Conte *et al.* (1999) *J. Mol. Biol* 285, 2177-2198

Docking Simulations
- *Computational Protein-Protein Interactions, R. Nussinov, G. Schreiber ed., CRC Press*
- Katchalski-Katzir *et al.* (1992) *PNAS* 89, 2195-2199
- Halperin *et al.* (2002) *Proteins* 47, 409-443
- Smith & Sternberg (2002) *Curr. Opin. Struct. Biol.* 12, 28-35
- Bonvin (2006) *Curr. Opin. Struct. Biol.* 16, 194
- Gray (2006) *Curr. Opin. Struct. Biol.* 16, 1
- Ritchie (2008) *Curr. Protein Pept. Sci.* 9, 1-15
- Andrusier et al. (2008) *Proteins* 73, 271
- May et al. (2008) *Curr. Comput. Aided Drug Des.* 4, 143-153

CAPRI
- *Proteins Special Issues* (July 2003, July 2005, December 2007, November 2010)

"Protein Interactions and Docking" group
Consejo Superior de Investigaciones Científicas
juanf@bsc.es