

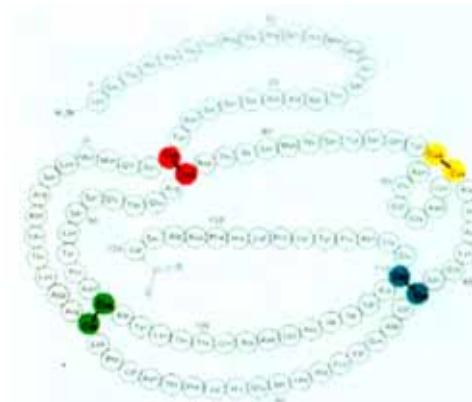
Protein interactions

- **Importance of protein interactions**
- **Study of protein interaction networks**
- **Types of complexes**
- **Computational analysis and structural modeling**
- **Protein-protein docking**
- **Recognition and prediction of interaction sites**

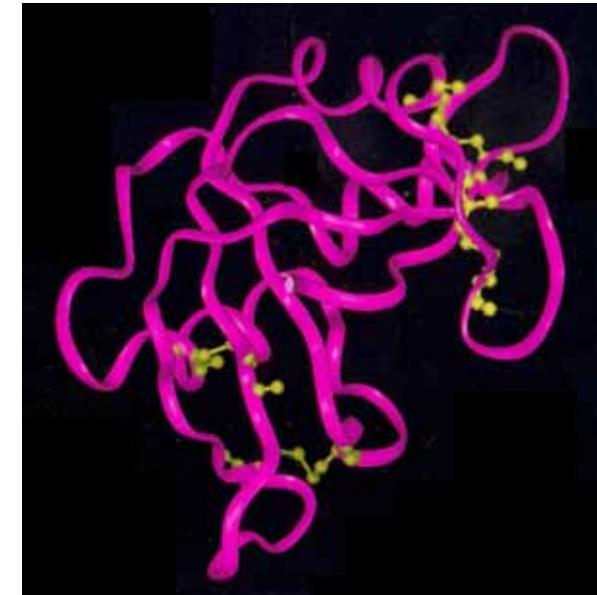
Importance of protein-protein interactions in biomedicine

```
AAGTCAAGCTGCTCTGTGGGCTGTGATCTGCCTCAAACCCACAGCCTGGGTAGCAGG  
AGGACCTTGATGCTCTGCCACAGATGAGGAGAAATCTCTTTCTCTGCTTGAAG  
GACAGACATGACTTGGATTCCCCAGGGAGTTGGCAACCAAGTCCAAAAGGCT  
GAAACCATCCCTGTCCTCATGAGATGATCCAGCAGATCTCAATCTTCAGCACA  
AAGGACTCATCTGCTGGGATGAGACCCCTCTAGACAAATTCTACACTGAAC  
TACCAAGCAGCTGAATGACCTGGAAAGCCTGTTGATAACAGGGGTGGGGTGACAGAG  
ACTCCCCTGATGAAGGAGGACTCCATTCTGGCTGTGAGGAAATACTCCAAAGAAC  
ACTCTCTATCTGAAAGAGAAGAAATACAGCCCTGTGCCTGGAGGTTGTAGAGCA  
GAAATCATGAGATCTTTCTTGTCAACAAACTTGCAGAAAGTTAAGAAGTAAG  
GAATGA, TGTGATCTGCCTCAAACCCACAGCCTGGTAGCAGGAGGACCTGATGC  
TCCTGGCACAGATGAGGAGAAATCTCTTTCTCTGCTTGAAGGACAGACATGACT  
TTGGATTTCCCAGGGAGTTGGCAACCAAGTCCAAAAGGCTGAAACCATCCCTG  
TCCTCCATGAGATGATCCAGCAGATCTCAATCTTCAGCACAAGGACTCATCTG  
CTGCTTGGATGAGACCCCTCTAGACAAATTCTACACTGAACCTACAGCAGCTGA  
ATGACCTGGAAAGCCTGTGATACAGGGGTGGGGTGACAGAGACTCCCTGATGA
```

DNA Sequence



Protein Sequence



3D Structure

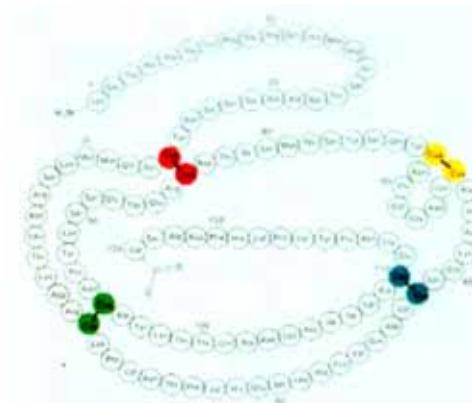
Function



Importance of protein-protein interactions in biomedicine

```
AAGTCAAGCTGCTCTGTGGGCTGTGATCTGCCTCAAACCCACAGCCTGGGTAGCAGG  
AGGACCTTGATGCTCTGCCACAGATGAGGAGAACTCTCTTTCTCTGCTTGAAG  
GACAGACATGACTTGGATTCCCCAGGAGGTTGGCAACCAAGTCCAAAAGGCT  
GAAACCATCCCTGCTCCATGAGATGATCCAGCAGATCTCAATCTTCAGCACA  
AAGGACTCATCTGCTGGGATGAGACCCCTCTAGACAAATTCTACACTGAAC  
TACCAAGCAGCTGAATGACCTGGAAAGCCTGTTGATAACAGGGGTGGGGTGACAGAG  
ACTCCCCTGATGAAGGAGGACTCCATTCTGGCTGTGAGGAAATACTCCAAGAAC  
ACTCTCTATCTGAAAGAGAAGAAATACAGCCCTGTGCTGGAGGTTGTCAGAGCA  
GAAATCATGAGATCTTTCTTGTCAACAAACTTGCAGAAAGTTAAGAAGTAAG  
GAATGA, TGTGATCTGCCTCAAACCCACAGCCTGGTAGCAGGAGACCTGATGC  
TCCTGGCACAGATGAGGAGAAATCTCTTTCTCTGCTTGAAGGACAGACATGACT  
TTGGATTTCCCAGGAGGTTGGCAACCAAGTCCAAAAGGCTGAAACCATCCCTG  
TCCTCCATGAGATGATCCAGCAGATCTCAATCTTCAGCACAAGGACTCATCTG  
CTGCTTGGATGAGACCCCTCTAGACAAATTCTACACTGAACCTACAGCAGCTGA  
ATGACCTTGGAAAGCCTGTGATACAGGGGGTGGGGTGACAGAGACTCCCTGATGA
```

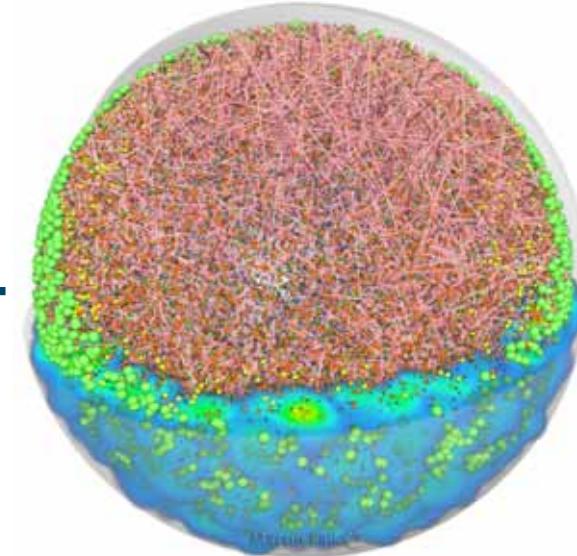
DNA Sequence



Protein Sequence



3D Structure

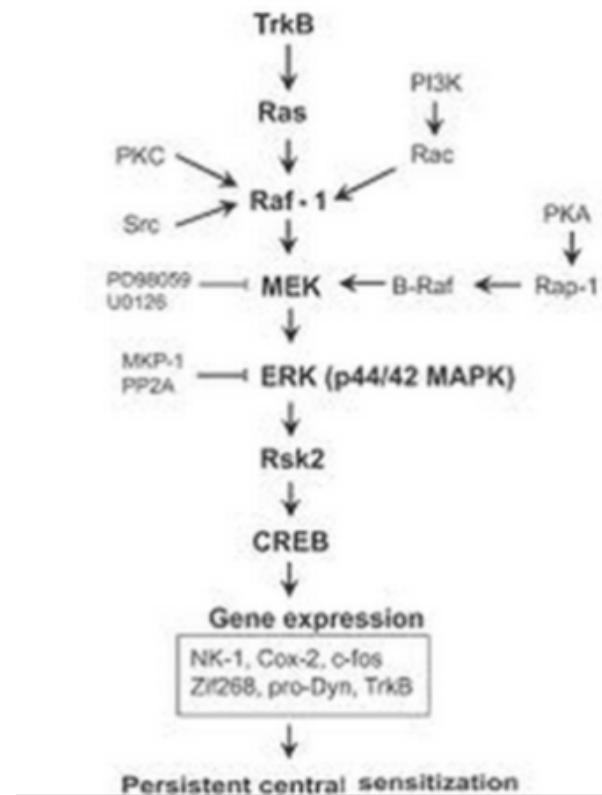


Molecular Recognition

Importance of protein-protein interactions in biomedicine

Understanding pathologies at molecular level

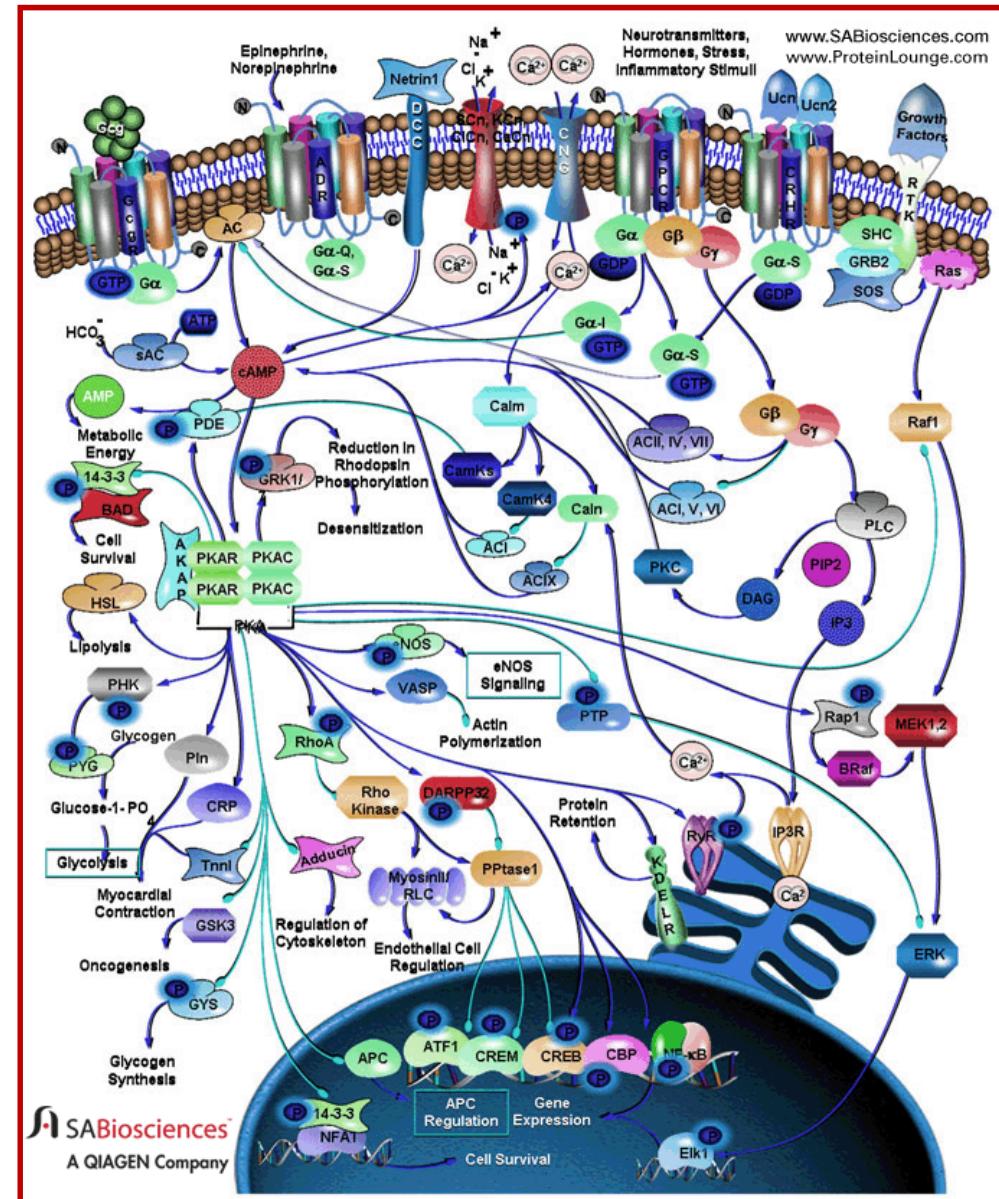
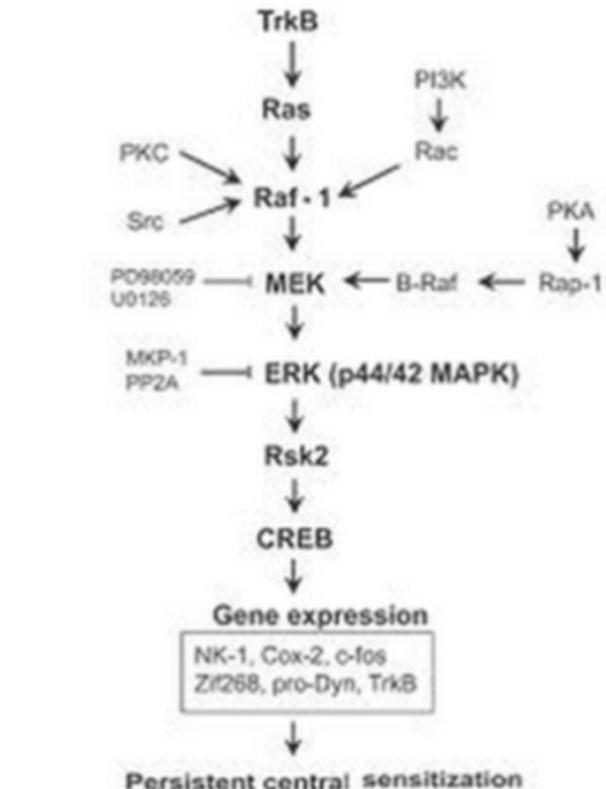
Traditional approach: characterize a pathological pathway



Importance of protein-protein interactions in biomedicine

Understanding pathologies at molecular level

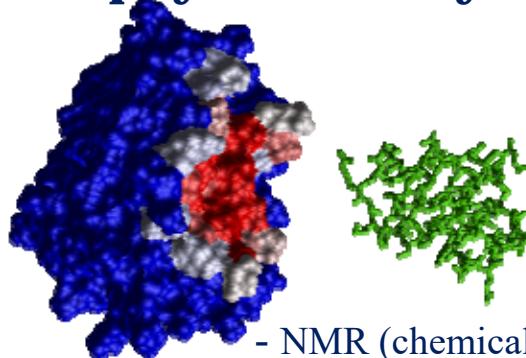
Traditional approach: characterize a pathological pathway



- Importance of protein interactions
- Study of protein interaction networks
- Types of complexes
- Computational analysis and structural modeling
- Protein-protein docking
- Recognition and prediction of interaction sites

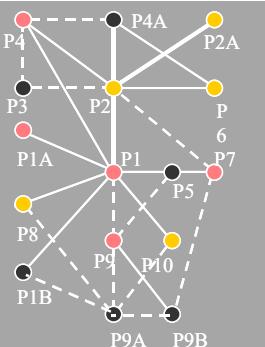
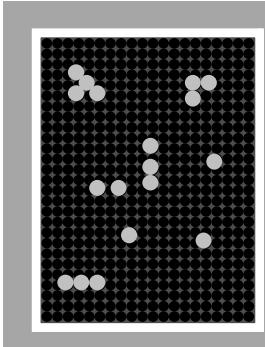
Study of Protein-Protein Interactions

Biophysical Analysis



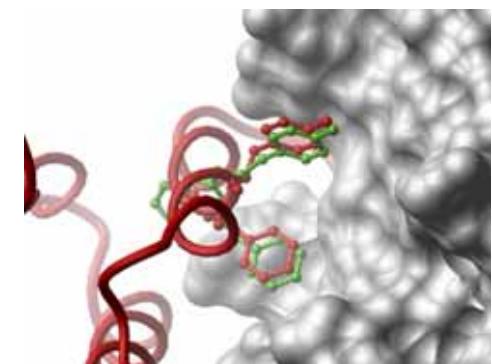
- NMR (chemical shifts)
- sequence conservation
- binding assays
- mutants & alanine-scanning

Protein Interaction Detection



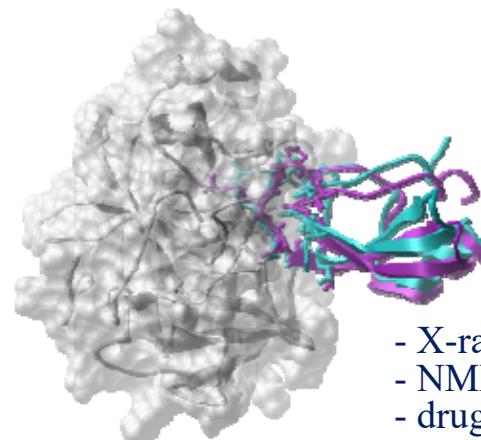
- two-hybrid test
- affinity column, gel assays...
- BIACore
- mass-spectrometry
- electron microscopy
- cross-linking
- co-immunoprecipitation
- immunofluorescence
- knock-out
- phylogenetic profiles, gene fusion events...
- ...

Applications



- protein design
- inhibitor discovery:
 - peptide mimicking
 - ligand docking
 - VLS
- association mechanism

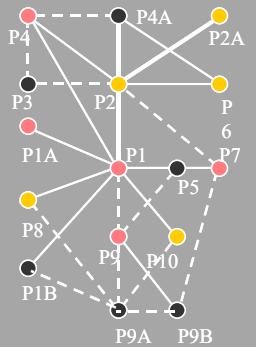
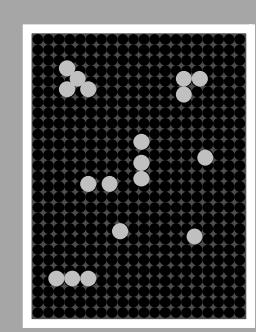
Structural Characterization



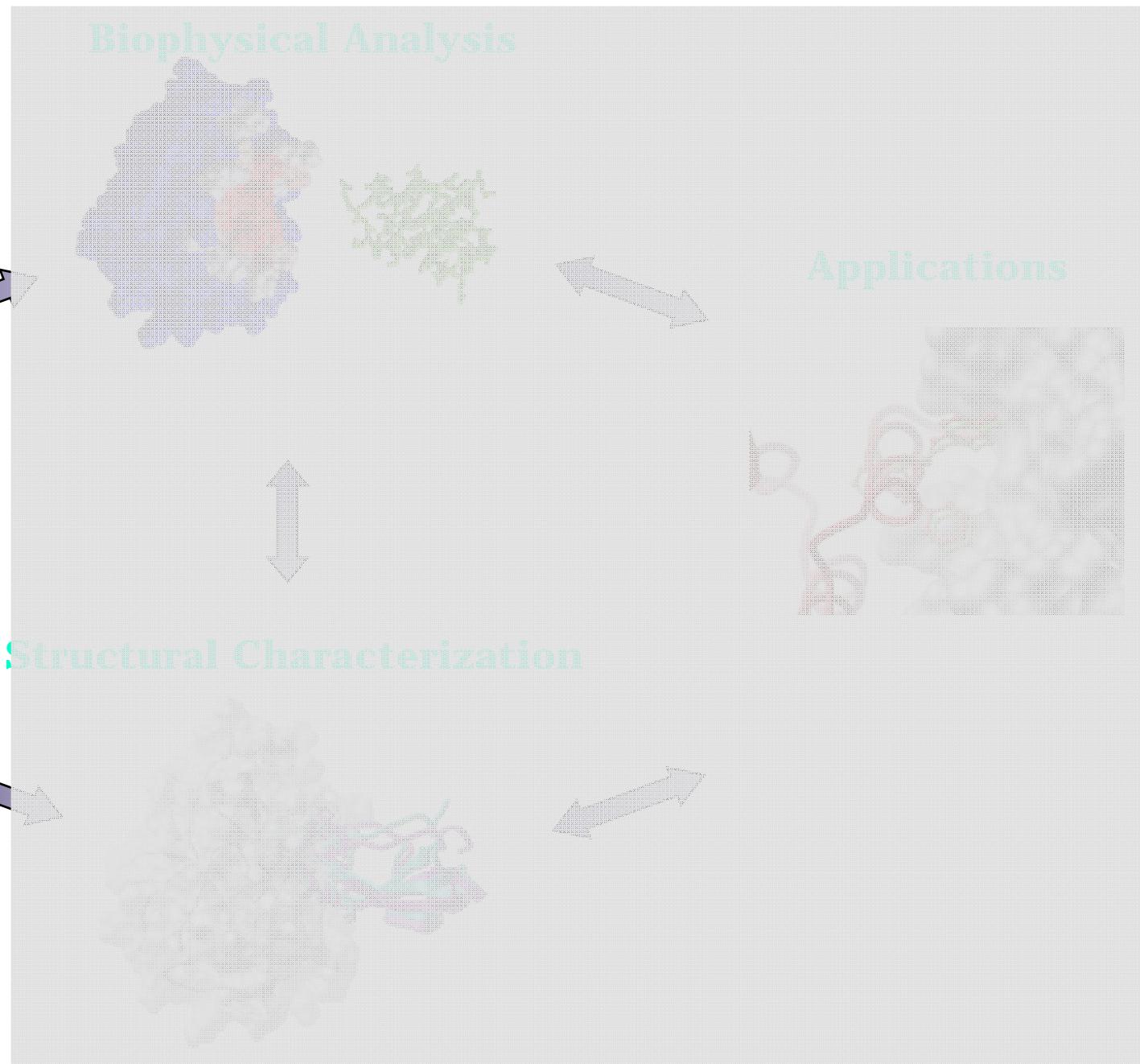
- X-ray
- NMR
- druggable pockets

Study of Protein-Protein Interactions

Protein Interaction Detection



- two-hybrid test
- affinity column, gel assays...
- BIACore
- mass-spectrometry
- electron microscopy
- cross-linking
- co-immunoprecipitation
- immunofluorescence
- knock-out
- phylogenetic profiles, gene fusion events...
- ...



Databases: protein-protein interactions

IntAct

(<http://www.ebi.ac.uk/intact/>)

BioGRID

(<http://thebiogrid.org/>)

DIP

(<http://dip.doe-mbi.ucla.edu/dip>)

MINT

(<http://mint.bio.uniroma2.it/mint>)

InnateDB

(<http://www.innatedb.com/>)

BINDTranslation

(<http://baderlab.org/BINDTranslation>)

HPRD

(www.hprd.org/)

MatrixDB

(<http://matrixdb.ibcp.fr/>)

Databases: protein-protein interactions

IMEx

<http://www.imexconsortium.org/>

Home | The International Molecular Exchange Consortium - Mozilla Firefox

File Edit View History Bookmarks Tools Help

www.imexconsortium.org Google

IMEx The International Molecular Exchange Consortium

Home Admin

About IMEx Curation Submit Your Data Training Licence Disclaimer Contact us

Home

Search the IMEx data resource Search

Use as input:

- Uniprot KB Accs
- Gene names
- Publication Ids

IMEx data

- A non-redundant set of protein-protein interaction data from a broad taxonomic range of organisms
- Expertly curated from direct submissions or peer-reviewed journals to a consistent high standard.
- Available in standard formats [MITAB](#) or [PSI-MI XML 2.5](#)
- Provided by a network of participating major public domain databases.

IMEx Partners

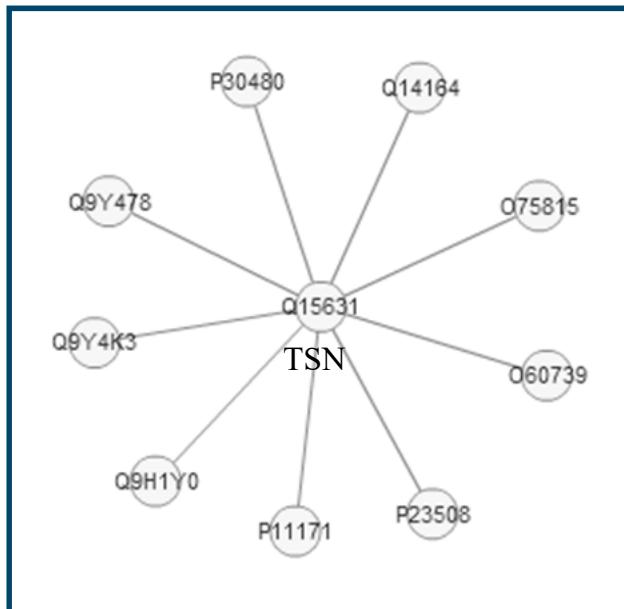
DIP	IntAct
MINT	I2D
MatrixDB	J. Craig Venter INSTITUTE
BioGRID	InnateDB
MOLECULAR CONNECTIONS	SIB

Citing IMEx

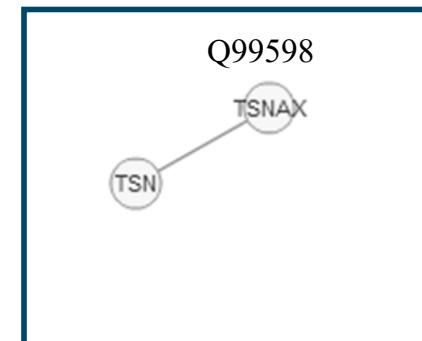
Orchard, S., et al.

A query example: translin (UniProt: Q15631)

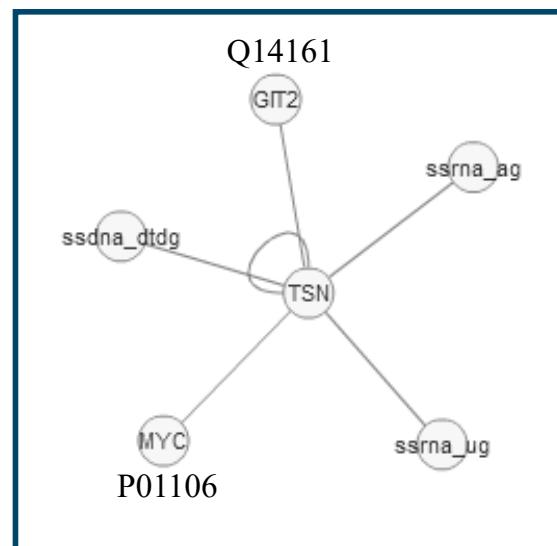
DIP



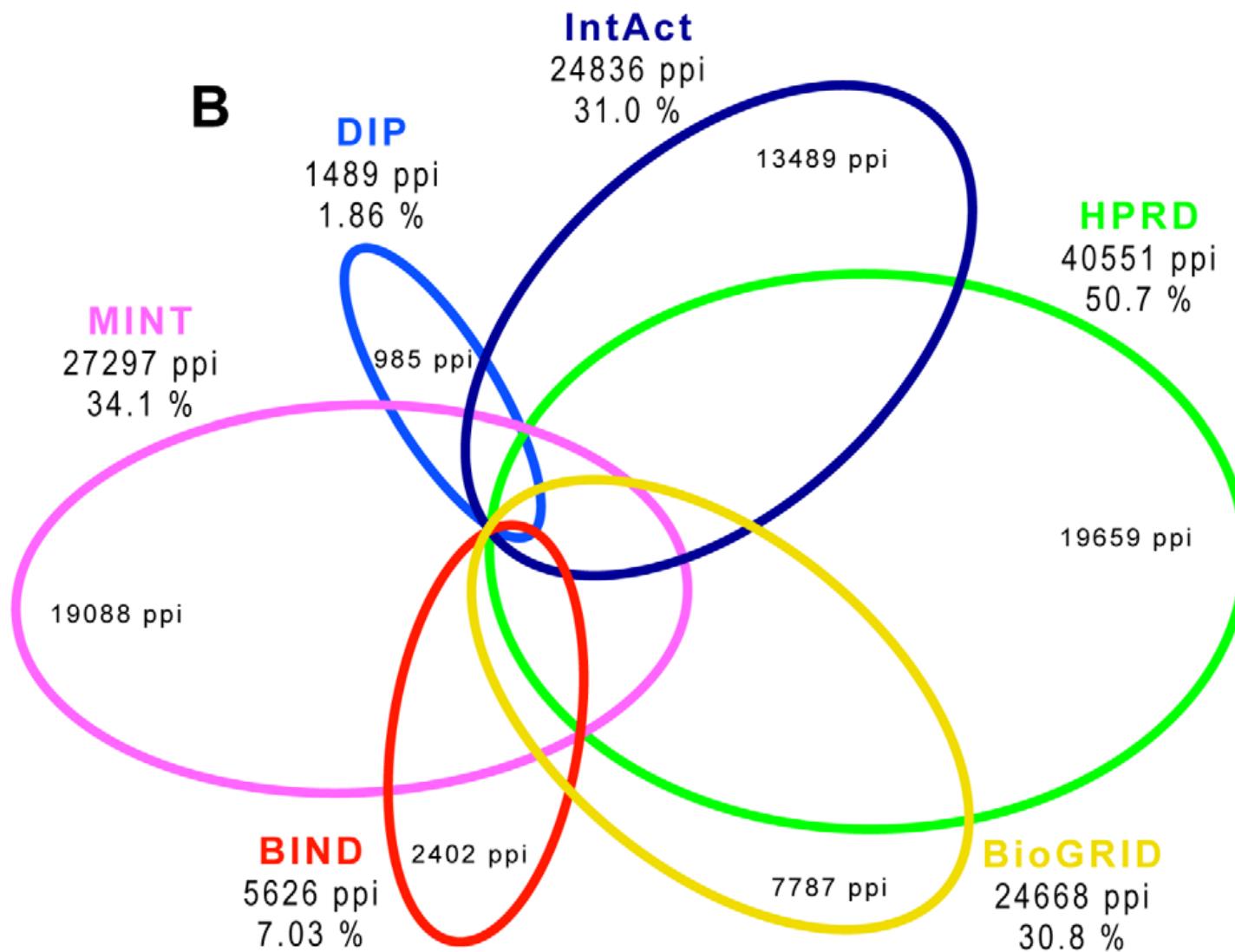
MINT



IntAct

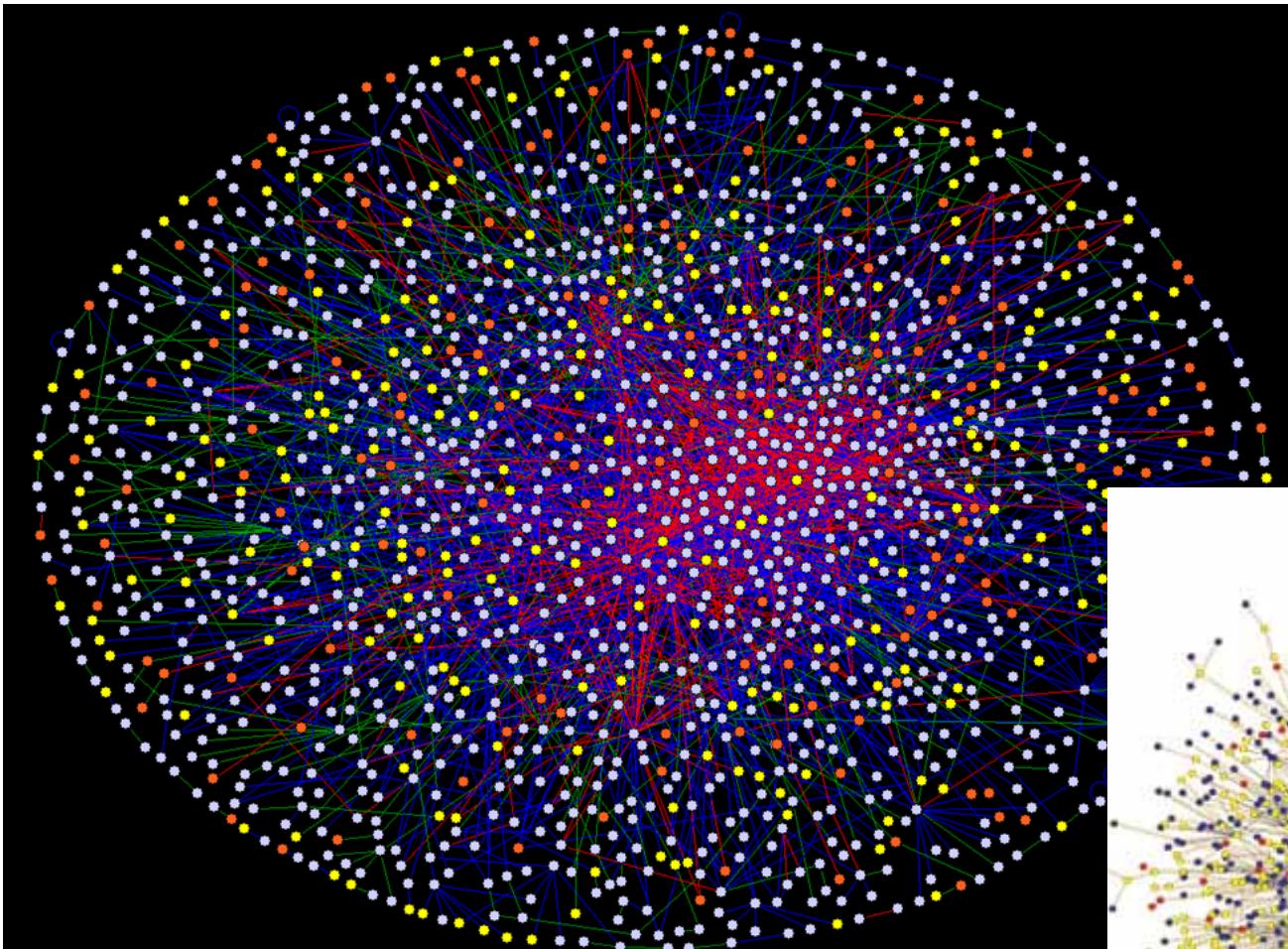


Low overlap in protein-protein interaction networks

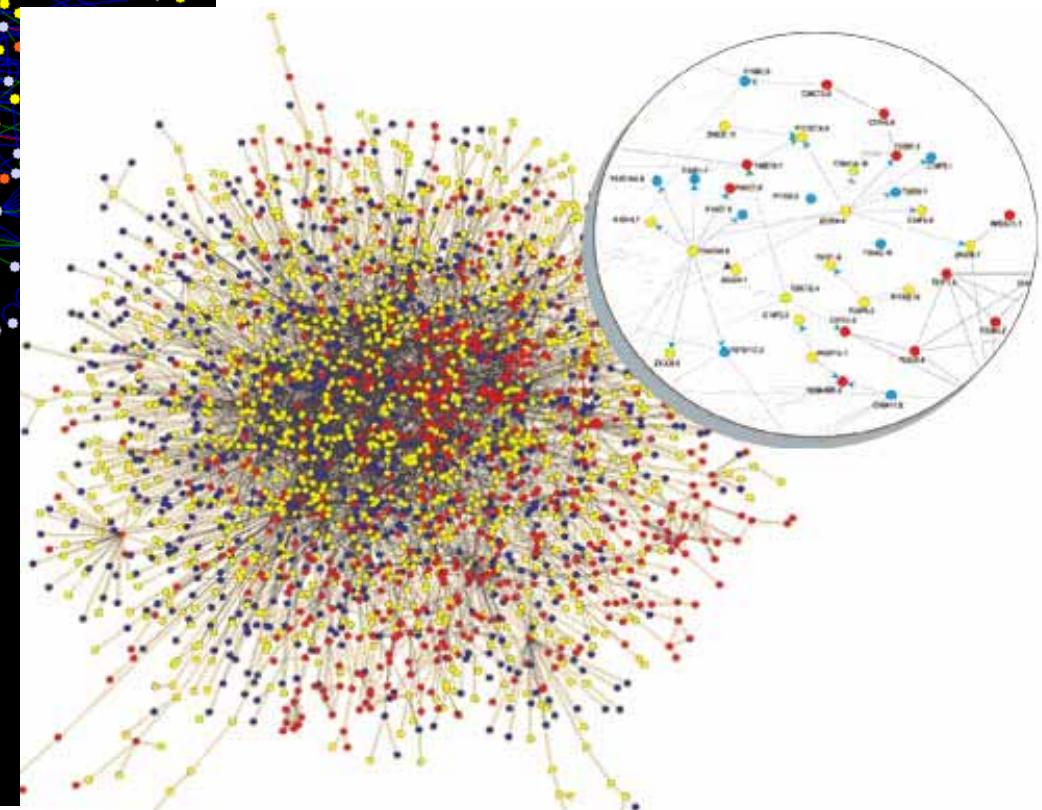


Protein-protein interaction networks

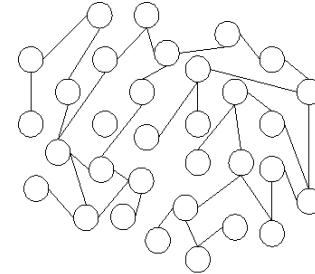
human



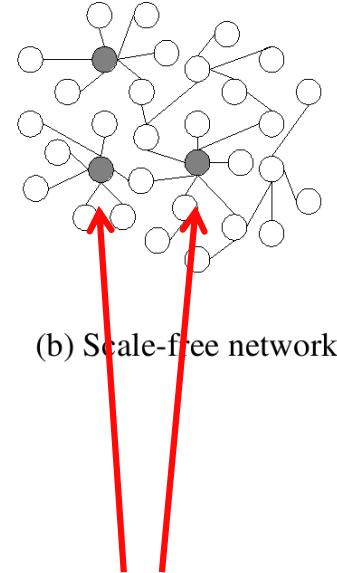
C. elegans



Graph-based network analysis



(a) Random network



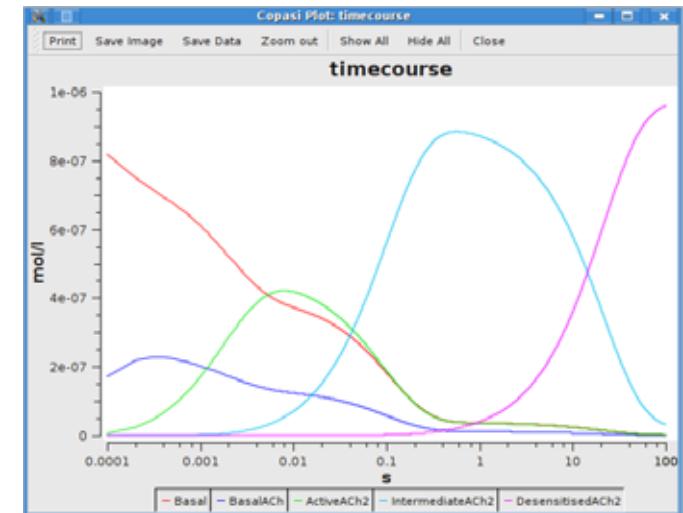
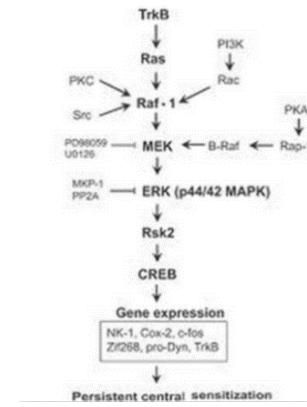
(b) Scale-free network

- **Graph theory $G = (V, E)$.**
 - Vertex represent proteins and edges represent interactions.
- **Used for Protein-protein interactions, signalling, gene regulation.**
- **Node degree: Number of edges per vertex**
 - Degree distribution defines type of network.
 - PPI networks are “scale-free”, degree of distribution is not uniform.
 - Hubs: Nodes having higher than average node degrees
- **Robustness: Scale-free networks are largely resistant to deletion of nodes (except large hubs).**
- **Networks can be compared and aligned**
 - Missing segments may reveal unknown interactions
- **Networks evolve following evolution of participating proteins.**
- **Quantitative simulation usually restricted to binary states (on/off)**

Hubs

Quantitative analysis: metabolic networks

- **Include both metabolites, and enzymes**
- **Main issue is generation of quantitative models**
 - Nodes are usually metabolites, and edges enzymes.
 - Edges contain kinetic, and flux data.
 - Require large amounts of input data!!!
- **Metabolic models are usually simulated using sets of differential equations (ODE)**
- **Metabolic control analysis**
 - Specific set of laws regarding regulation of metabolic pathways. Aim: to define key regulatory steps.
- **Databases: Kegg, BioCyc, BioModels, ...**



Quantitative analysis: metabolic networks

Firefox ▾

KEGG PATHWAY Database

www.genome.jp/kegg/pathway.html

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Select prefix map Organism Enter keywords Go Help

[New maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

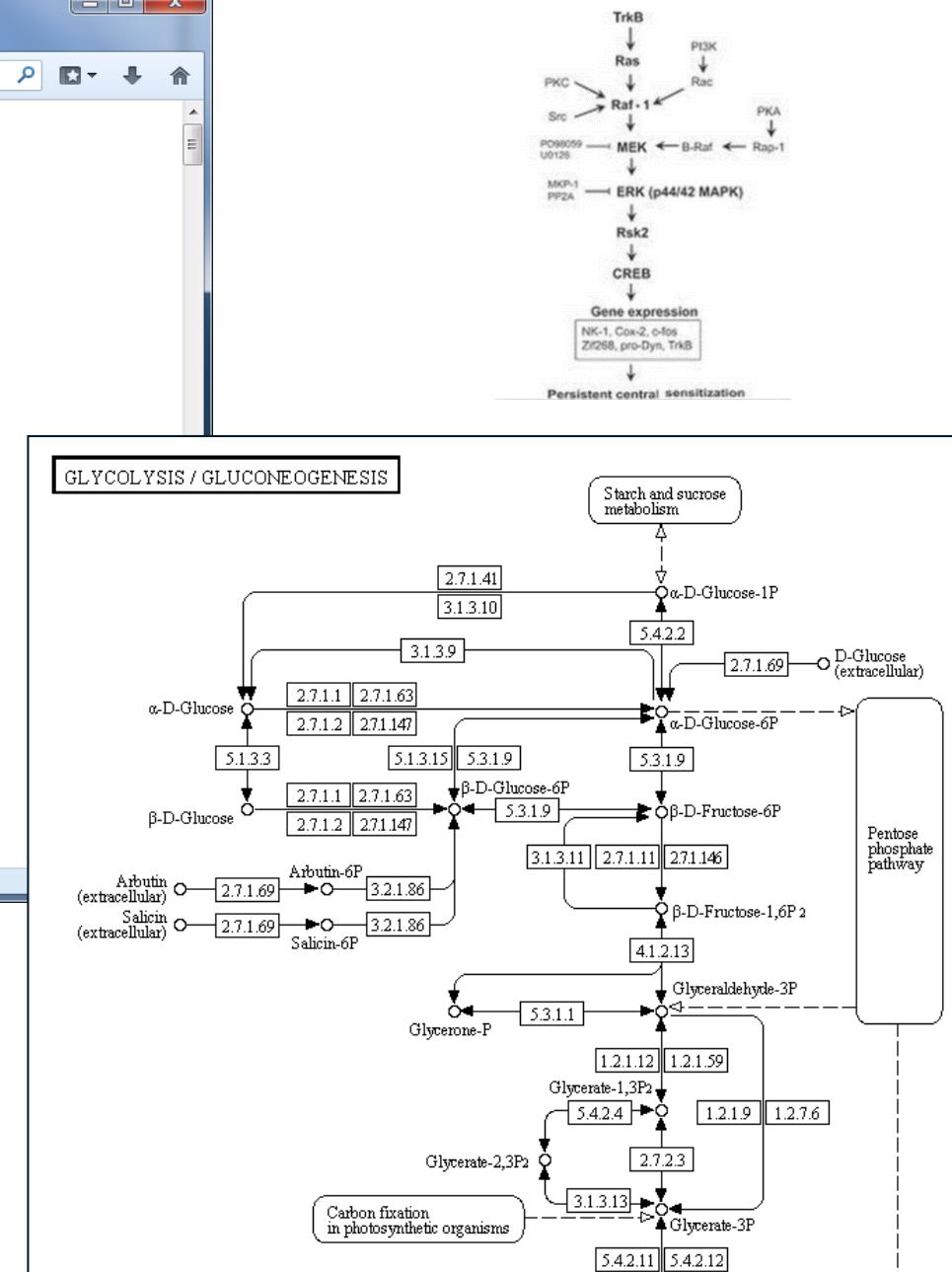
- 1. Metabolism
- 2. Genetic Information Processing
- 3. Environmental Information Processing
- 4. Cellular Processes
- 5. Organismal Systems
- 6. Human Diseases

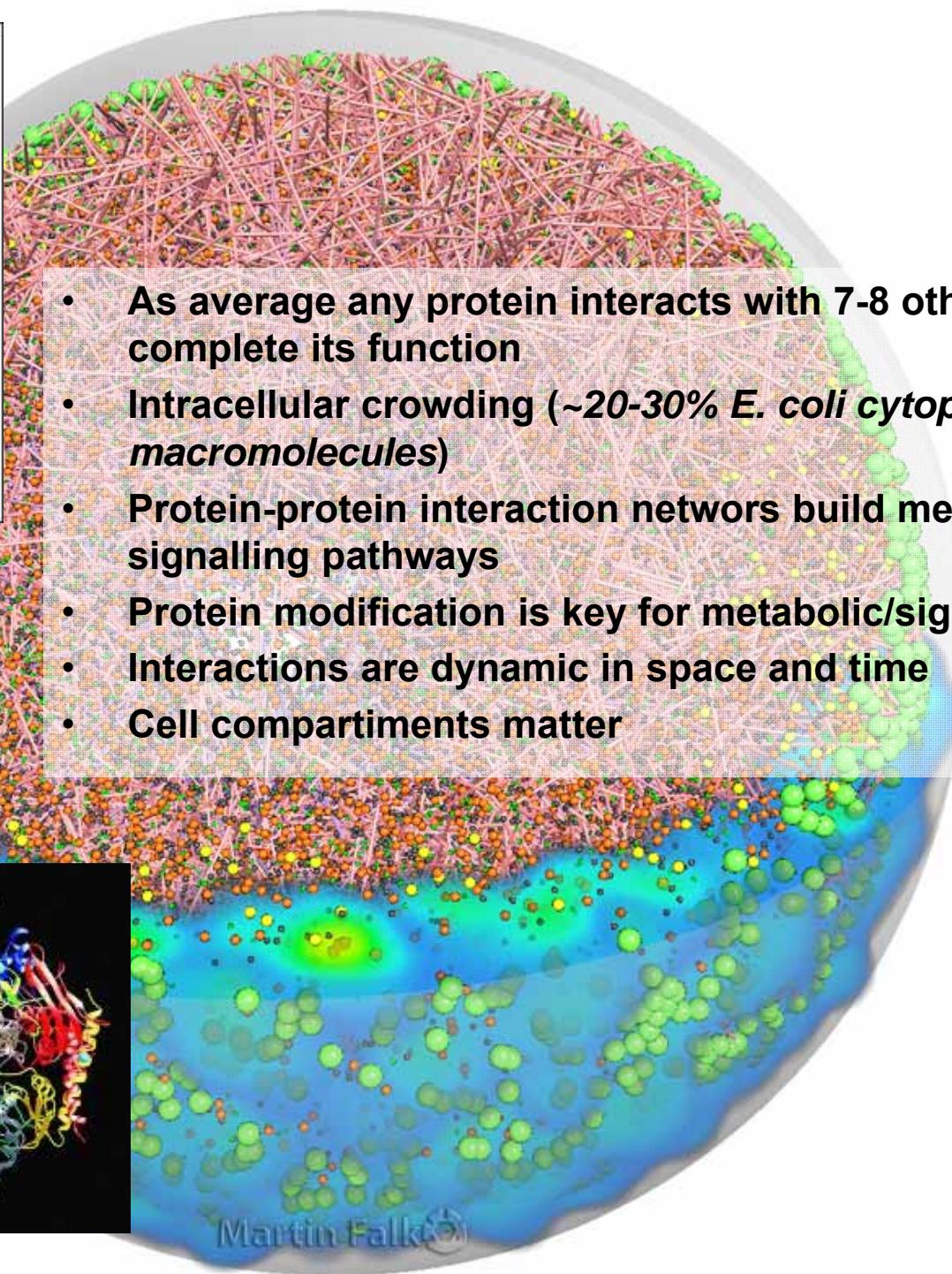
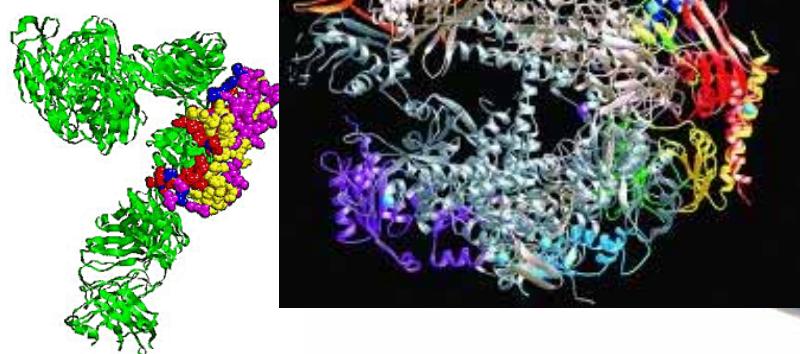
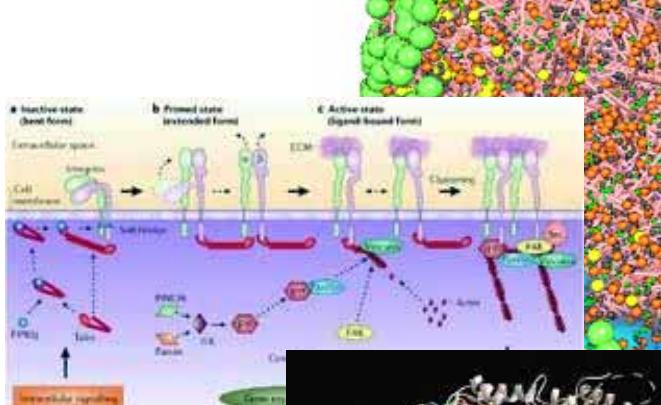
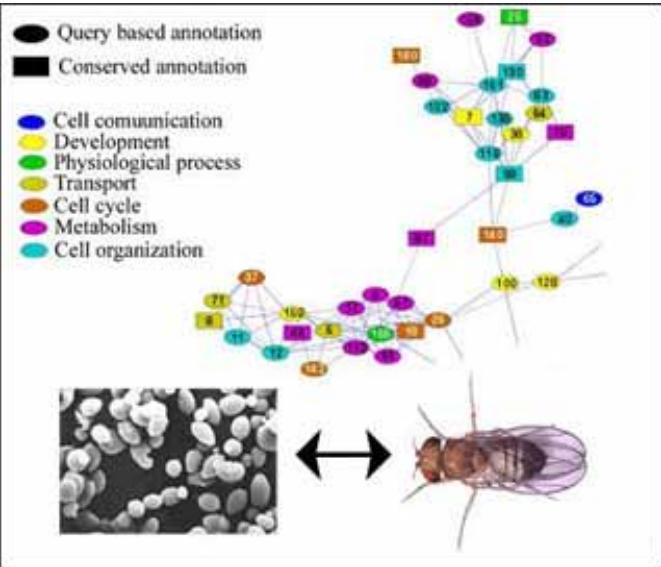
and also on the structure relationships (KEGG drug structure maps) in:

- 7. Drug Development

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

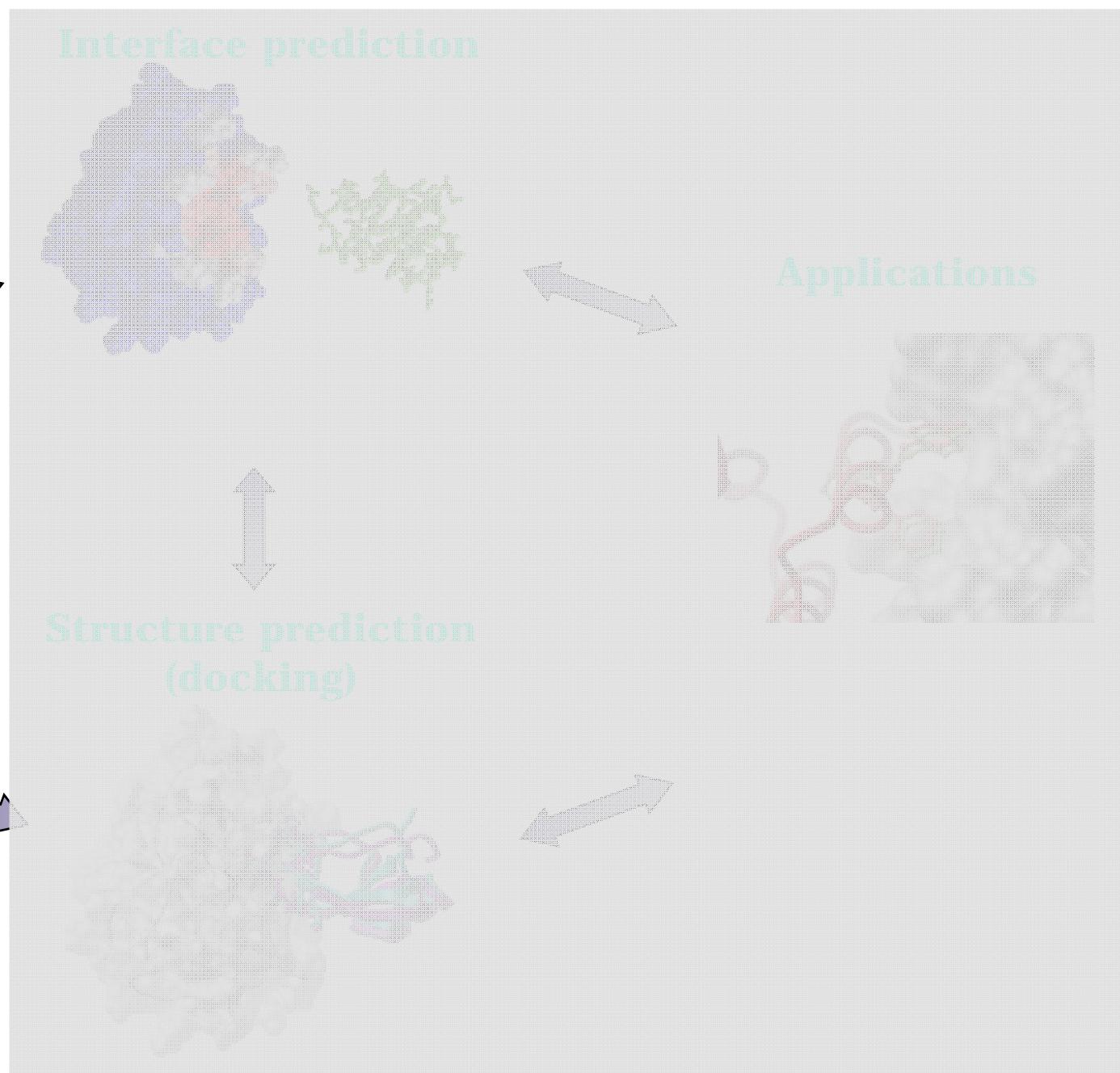
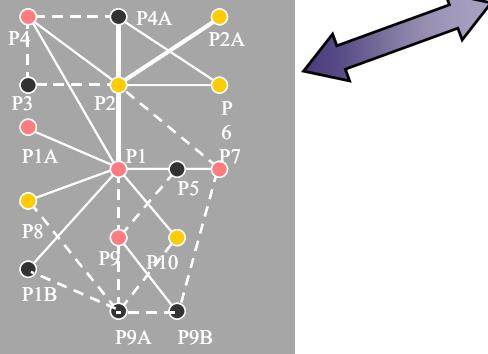
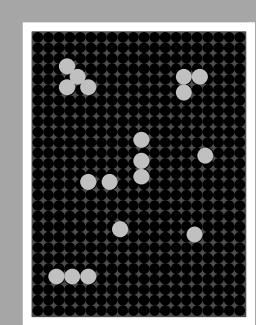




- As average any protein interacts with 7-8 other proteins to complete its function
- Intracellular crowding (~20-30% *E. coli* cytoplasm are macromolecules)
- Protein-protein interaction networks build metabolic and signalling pathways
- Protein modification is key for metabolic/signalling regulation
- Interactions are dynamic in space and time
- Cell compartments matter

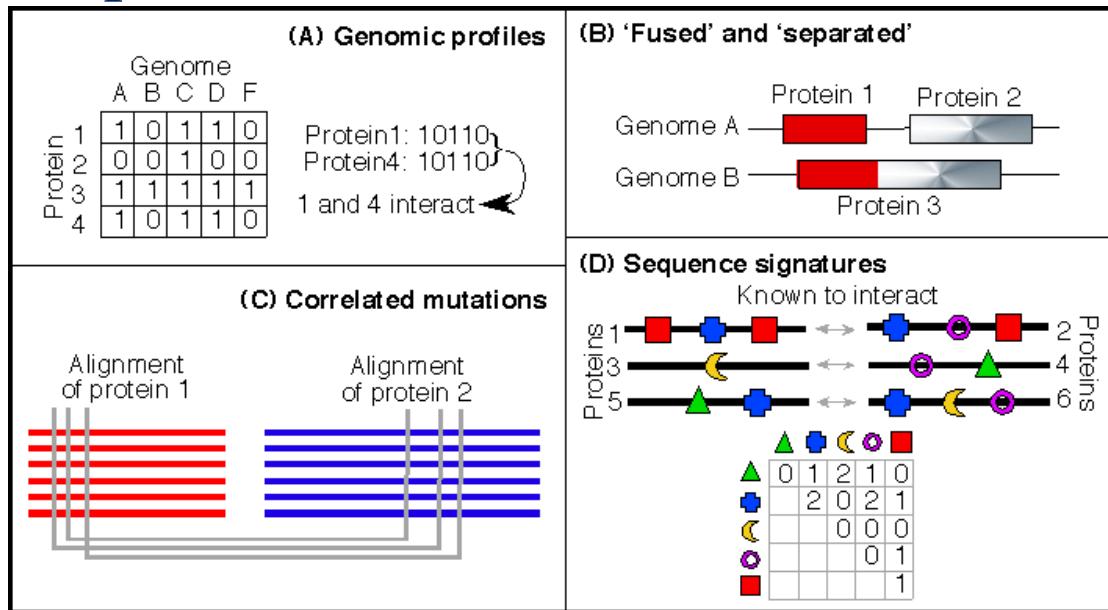
Computational studies of Protein-Protein Interactions

Protein Interaction Detection

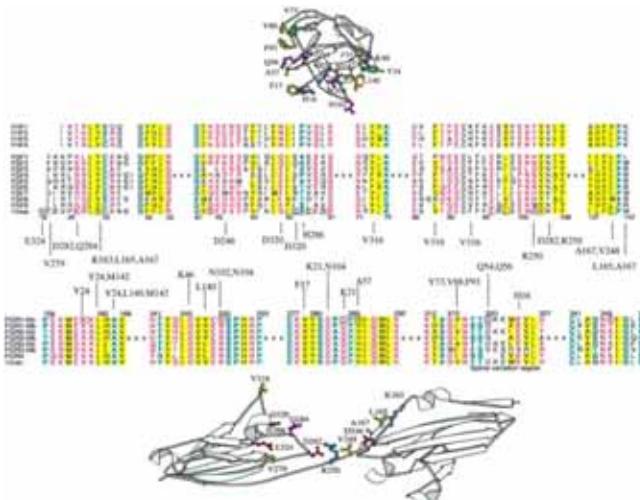


Identification of PPIs: do proteins A and B interact?

Sequence-based

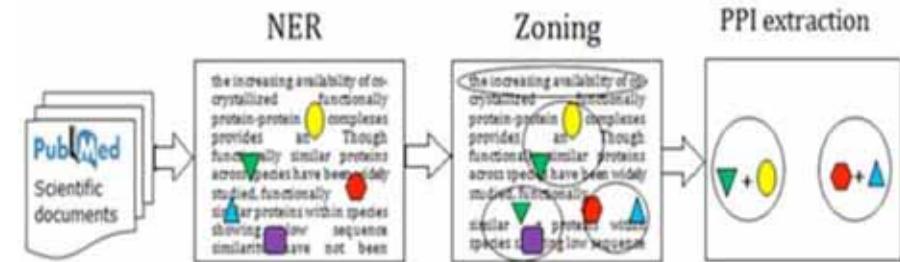


Structural bioinformatics



- Interolog search
 - Interactions are conserved among organisms (BIANA/BIPS from Baldo's)
- Gene neighborhood
- Gene coexpression profiles
 - Expressed amounts of interacting proteins should be correlated.

Data mining



Identification of PPIs: do proteins A and B interact?

STRING (<http://string-db.org/>)

IL10 protein (Homo sapiens) ×

string-db.org/newstring_cgi/show_network_section.pl

Imported From Firefox LIST OF PUBLISHERS | ... Beall's List of Predator... Portada elPeriódico.com EL PAÍS: el periódico g... BBC BBC - Homepage CAPRI HOME

confidence evidence actions interactive advanced more less save

(requires Flash player 10 or better)

Your Input:

IL10 interleukin 10; Inhibits the synthesis of a number of cytokines, including IFN-gamma, IL-2, IL-3, TNF and GM-CSF produced by activated macrophages and by helper T-cells (178 aa) (*Homo sapiens*)

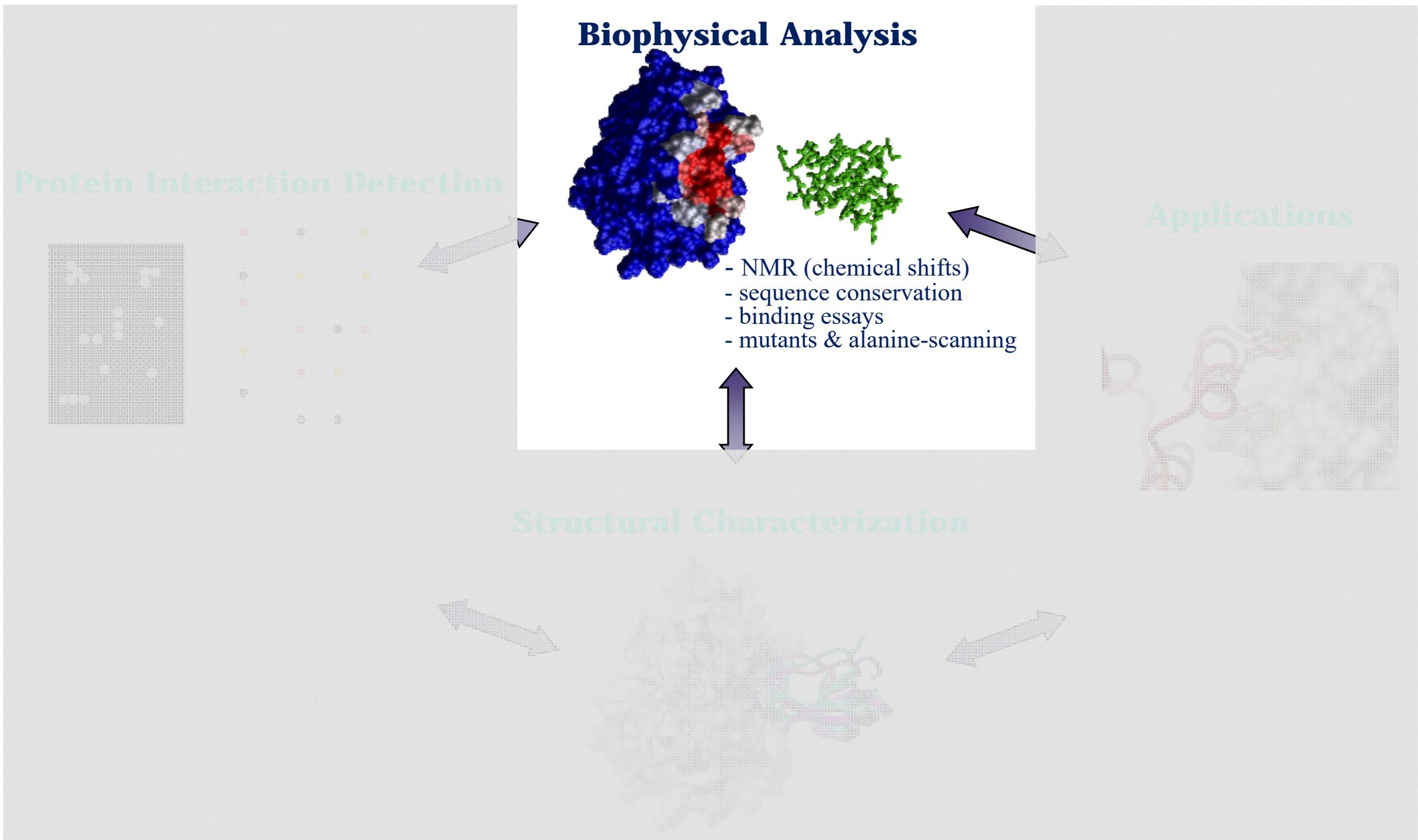
Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
IL10RA	interleukin 10 receptor, alpha; Receptor for IL10; binds IL10 with a high affinity (578 aa)	●	●	●						0.999
IL10RB	interleukin 10 receptor, beta; Receptor for IL10 and IL22. Serves as an accessory chain essenti [...] (325 aa)	●	●	●						0.986
MRC1L1	mannose receptor, C type 1-like 1; May mediate the endocytosis of glycoproteins (By similarity) (1456 aa)		●	●						0.977
DIF	Tumor necrosis factor Precursor (TNF-alpha)(Tumor necrosis factor ligand superfamily member 2) ([...] (233 aa)			●						0.976
NOS2	nitric oxide synthase 2, inducible; Produces nitric oxide (NO) which is a messenger molecule wi [...] (1153 aa)		●	●						0.973
IL6	interleukin 6 (interferon, beta 2); Cytokine with a wide variety of biological functions. It is [...] (212 aa)			●						0.969
IL4	interleukin 4; Participates in at least several B-cell activation processes as well as of other [...] (153 aa)			●						0.968
IL2	interleukin 2; Produced by T-cells in response to antigenic or mitogenic stimulation, this prot [...] (153 aa)			●						0.967
CD4	CD4 molecule; Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regul [...] (458 aa)			●						0.966
TNF	tumor necrosis factor (TNF superfamily, member 2); Cytokine that binds to TNFRSF1A/TNFR1 and TN [...] (233 aa)			●						0.958

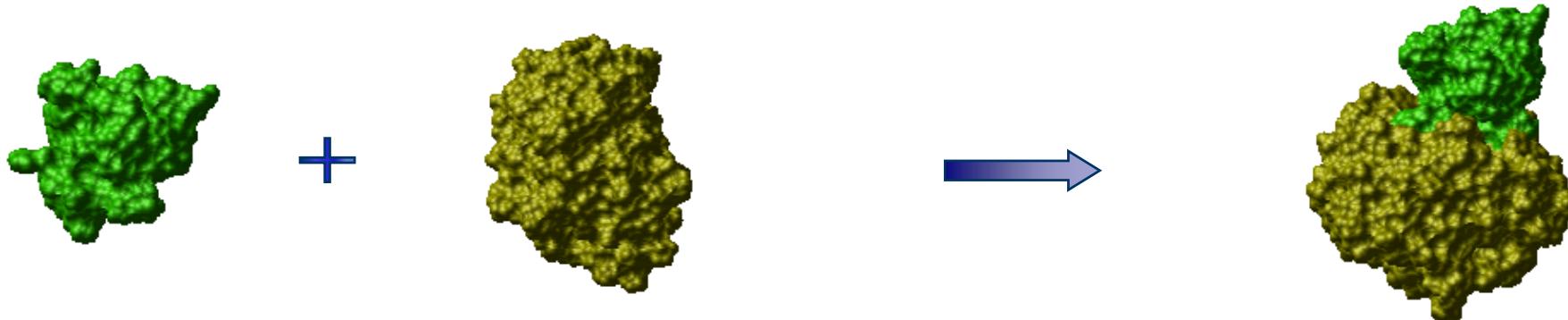
Views:

Neighborhood Fusion Occurrence Coexpression Experiments Database Textmining Summary Network

Study of Protein-Protein Interactions



Protein-protein binding



In equilibrium ($\nu_a = \nu_d$) :

$$k_a [A][B] = k_d [AB]$$

$$K_a = \frac{1}{K_d} = \frac{k_a}{k_d} = \frac{[AB]}{[A][B]}$$

$$\Delta G_a = -RT \ln K_a$$

Early protein-protein biophysical studies

Adv. Protein Chem. 9, 325 (1954)

Protein-Protein Interactions

BY DAVID F. WAUGH

Massachusetts Institute of Technology, Cambridge, Massachusetts

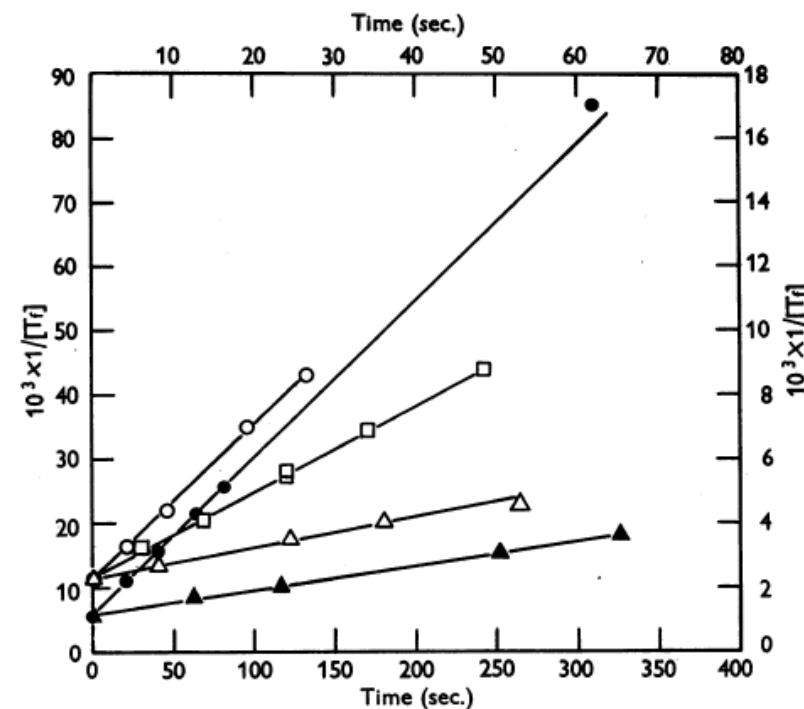
CONTENTS

	<i>Page</i>
I. Introduction.....	326
II. Group Interactions.....	328
1. Covalent Bonds.....	330
2. Charged Groups.....	331
3. Hydrogen Bonds.....	333
4. Van der Waals' Forces (Electronic Dispersion Forces).....	334
5. Use.....	338
6. Forces at Long Range.....	338
7. Urea.....	339
III. The Native Protein.....	341
1. Protein Composition.....	341
2. A View of the Protein Molecule.....	347
3. The Nature of the Internal Volume.....	354
a. Electrostriction.....	354
b. Enzymatic Hydrolysis and Denaturation.....	355
c. Side Chain Groups.....	355
d. Spectroscopic Data.....	357
e. Size, Shape, Hydration.....	357
4. The Internal Volume and Composition.....	358
IV. Specific Interactions.....	359
1. Antigen-Antibody.....	359
2. Hemocyanin.....	362
a. Urea and Other Agents.....	365
3. Plant Seed Proteins.....	366
4. Insulin.....	369
a. Readily Reversible Associations.....	370
b. Insulin Fibrils.....	374
5. Trypsin.....	377
6. Crystalline Soybean Trypsin Inhibitor.....	379
7. Pepsinogen and Pepsin.....	379
8. Chymotrypsinogen.....	380
9. Chymotrypsin.....	383
10. Serum Albumin.....	384
a. Aggregation.....	386
b. Urea.....	390
c. Other Interaction.....	392

>15K PubMed entries for «protein-protein interactions»

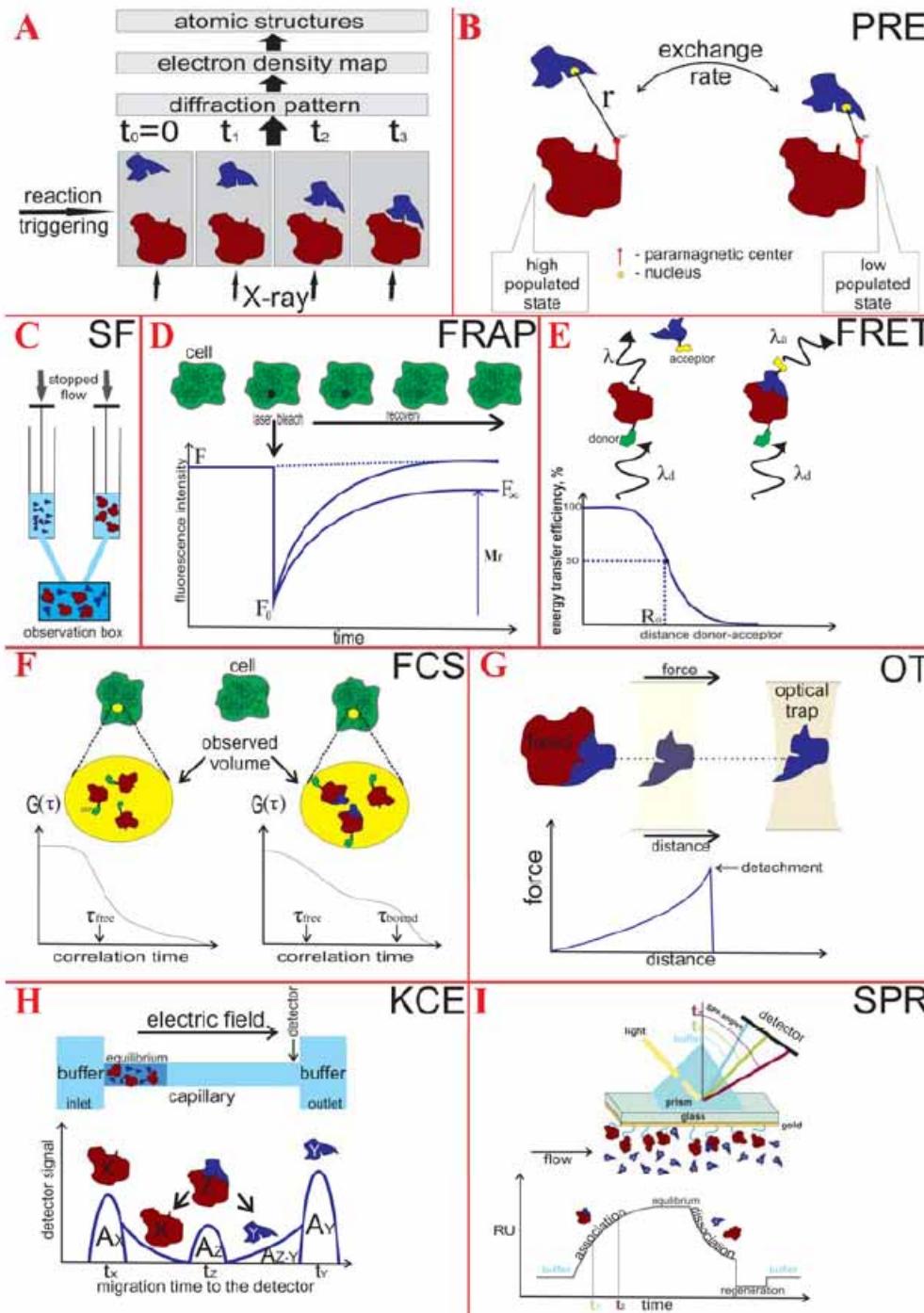


First PubMed reference on PPI



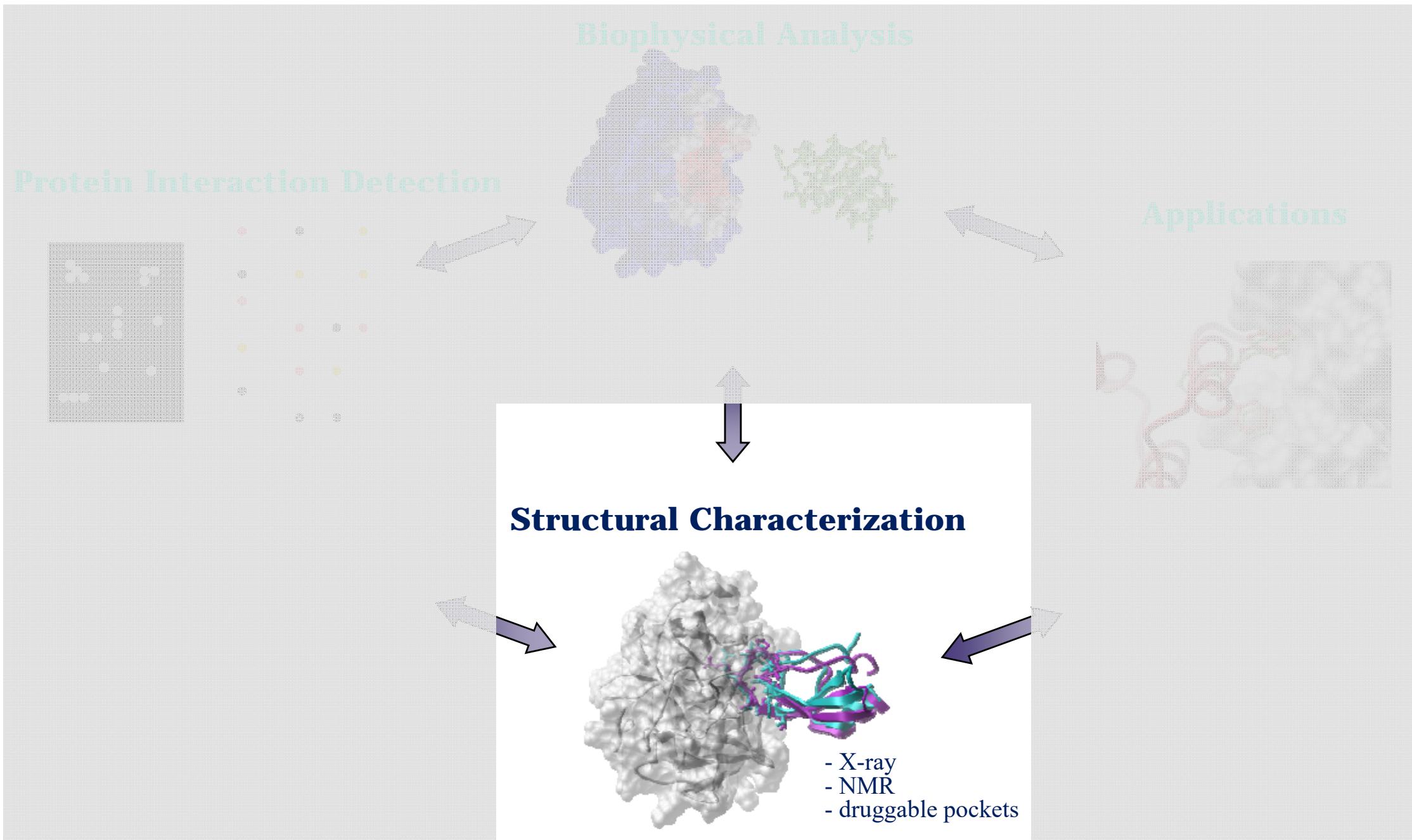
Green N.M. (1957) Kinetics of the reaction between trypsin and the pancreatic trypsin inhibitor. *Biochem J.* 66, 407

Protein-protein interactions: Biophysical methods

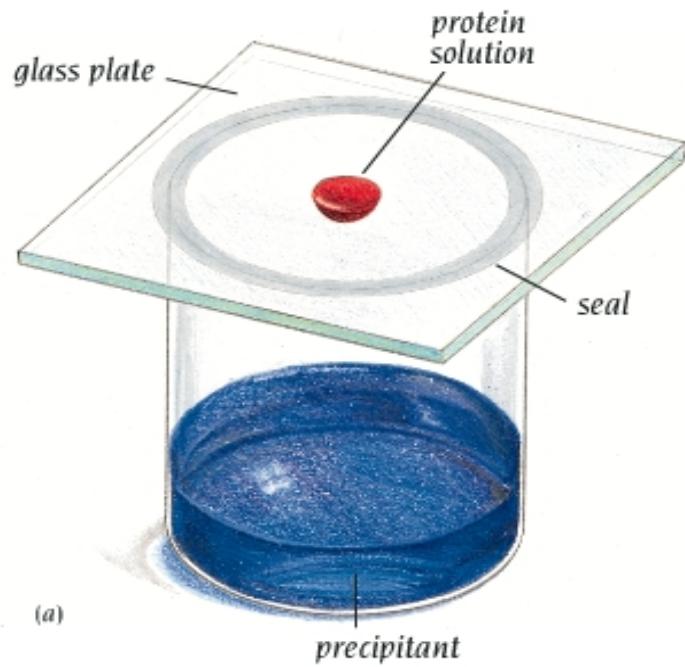


- X-ray crystallography
- NMR
- Stopped-flow methods
- Fluorescence Recovering After Photobleaching
- Fluorescence Resonance Energy Transfer
- Fluorescence Correlation Spectroscopy
- Optical Tweezers / AFM / Biocore
- Electrophoresis
- Surface Plasmon Resonance (SPR)
- Isothermal Titration Calorimetry (ITC)

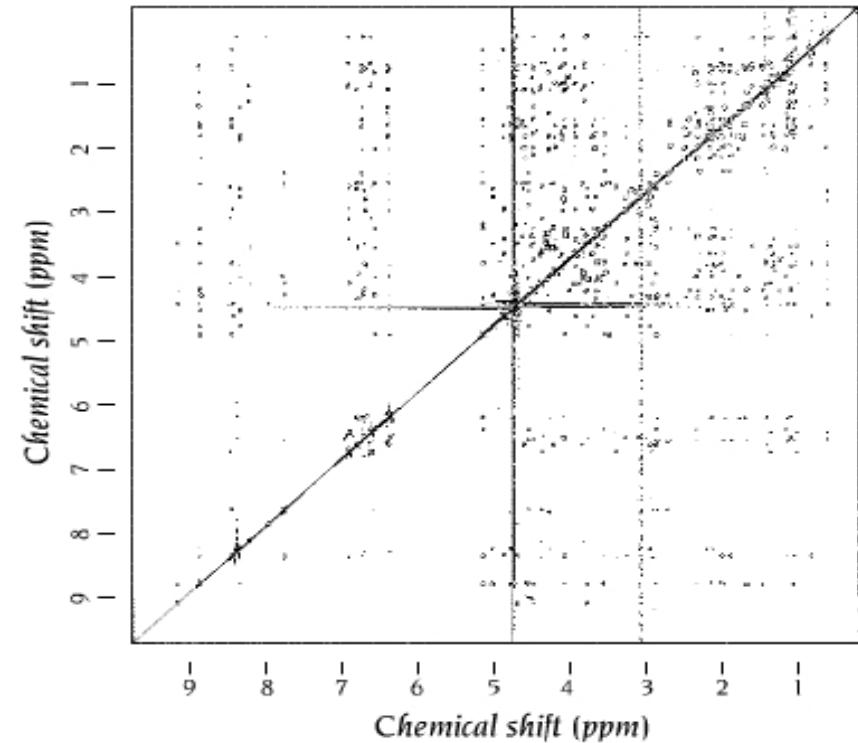
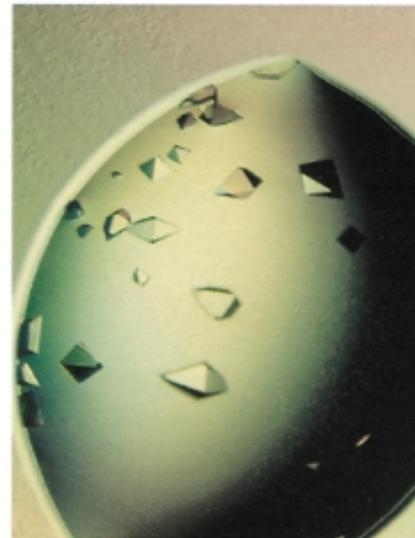
Study of Protein-Protein Interactions



3D structure - experimental determination

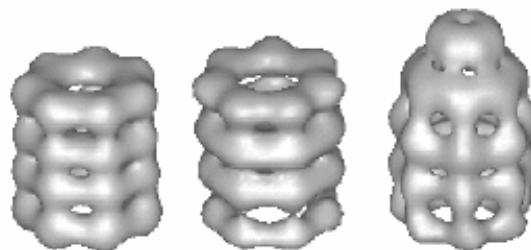


(b)



©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

X-ray



NMR

Electron microscopy, SAXS...

3D structure - experimental determination

ATOM	1	N	CYS	E	1	11.377	21.513	11.770	1.00	7.18	1CGI	131
ATOM	2	CA	CYS	E	1	12.025	21.956	13.016	1.00	5.40	1CGI	132
ATOM	3	C	CYS	E	1	11.406	21.350	14.300	1.00	6.41	1CGI	133
ATOM	4	O	CYS	E	1	10.216	21.020	14.517	1.00	5.73	1CGI	134
ATOM	5	CB	CYS	E	1	12.168	23.454	12.852	1.00	3.26	1CGI	135
ATOM	6	SG	CYS	E	1	10.913	24.625	13.296	1.00	2.00	1CGI	136
ATOM	7	N	GLY	E	2	12.379	21.161	15.213	1.00	6.48	1CGI	137
ATOM	8	CA	GLY	E	2	12.408	20.728	16.555	1.00	5.36	1CGI	138
ATOM	9	C	GLY	E	2	11.698	19.535	17.075	1.00	5.75	1CGI	139
ATOM	10	O	GLY	E	2	10.898	19.506	18.027	1.00	4.47	1CGI	140
ATOM	11	N	VAL	E	3	12.037	18.470	16.365	1.00	7.72	1CGI	141
ATOM	12	CA	VAL	E	3	11.501	17.132	16.643	1.00	9.05	1CGI	142
ATOM	13	C	VAL	E	3	12.512	16.048	16.228	1.00	9.18	1CGI	143
ATOM	14	O	VAL	E	3	12.510	15.741	15.035	1.00	9.80	1CGI	144
ATOM	15	CB	VAL	E	3	10.164	16.849	15.920	1.00	9.19	1CGI	145
ATOM	16	CG1	VAL	E	3	9.557	15.653	16.641	1.00	9.10	1CGI	146
ATOM	17	CG2	VAL	E	3	9.227	18.034	15.785	1.00	8.92	1CGI	147
ATOM	18	N	PRO	E	4	13.224	15.547	17.207	1.00	8.64	1CGI	148
ATOM	19	CA	PRO	E	4	14.215	14.528	17.062	1.00	8.95	1CGI	149
ATOM	20	C	PRO	E	4	13.596	13.198	16.728	1.00	10.49	1CGI	150
ATOM	21	O	PRO	E	4	12.678	12.712	17.362	1.00	10.03	1CGI	151
ATOM	22	CB	PRO	E	4	14.852	14.352	18.458	1.00	8.37	1CGI	152
ATOM	23	CG	PRO	E	4	14.656	15.760	18.999	1.00	8.59	1CGI	153
ATOM	24	CD	PRO	E	4	13.191	15.992	18.624	1.00	8.89	1CGI	154
ATOM	25	N	ALA	E	5	14.256	12.662	15.700	1.00	13.28	1CGI	155
ATOM	26	CA	ALA	E	5	13.919	11.330	15.124	1.00	14.12	1CGI	156
ATOM	27	C	ALA	E	5	14.040	10.433	16.352	1.00	14.67	1CGI	157
ATOM	28	O	ALA	E	5	13.228	9.554	16.573	1.00	15.43	1CGI	158

3D structure - experimental determination

RCSB Protein Data Bank

http://www.pdb.org/ Google

Apple (113) ▾ Amazon eBay Yahoo! News (165) ▾ General Bookmarks ▾

PDB
PROTEIN DATA BANK

Contact Us | Help | Print Page

All PDB ID or keyword Web Pages Author SEARCH | Advanced Search

Welcome to the RCSB PDB

The RCSB PDB provides a variety of tools and resources for studying the structures of biological macromolecules and their relationships to sequence, function, and disease.

The RCSB is a member of the [wwPDB](#) whose mission is to ensure that the PDB archive remains an international resource with uniform data.

This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

Information about compatible browsers can be found [here](#).

A [narrated tutorial](#) illustrates how to search, navigate, browse, generate reports and visualize structures using this new site. [This requires the Microsoft Silverlight plug-in.]

Comments? info@rcsb.org

Molecule of the Month: Topoisomerases

Each of your cells contains about the tiny space inside the nucleus. As you might imagine, these things very easily get tangled up in the busy environment. To make matters even more complicated, these things need to be unwound to access the genetic material.

[More ...](#)

[Previous Features](#)

Experimental Method

Method	Count
X-ray	142272
Solution NMR	12768
Electron Microscopy	4269
Hybrid	158
Electron Crystallography	145
Solid-State NMR	107
Neutron Diffraction	68
Fiber Diffraction	38
Solution Scattering	32
Other	24

Jan 2020

NEWS

- Complete News
- Newsletter
- Discussion Forum

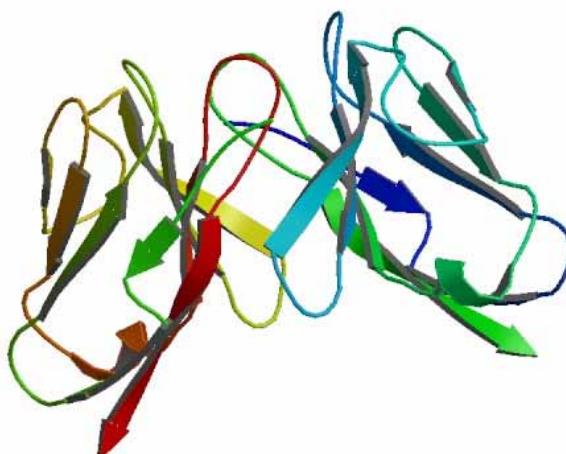
17-Jan-2006
Time-stamped Copies of PDB Archive Available via FTP

Time-stamped yearly snapshots of the PDB Archive

First 3D structures of protein-protein complexes

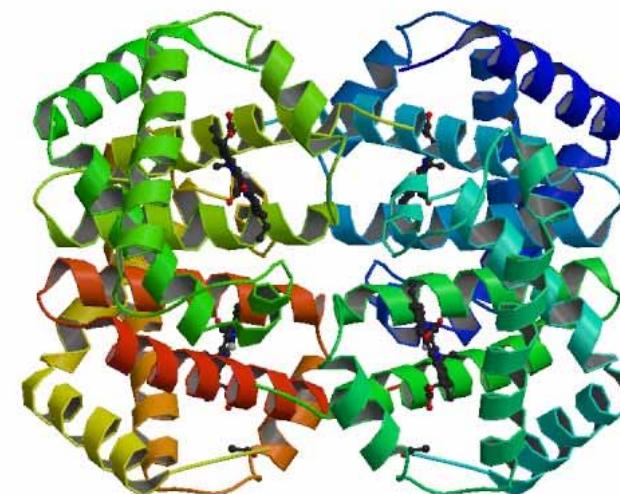
First homodimer
(1976)

Bence-Jones protein REI
1REI



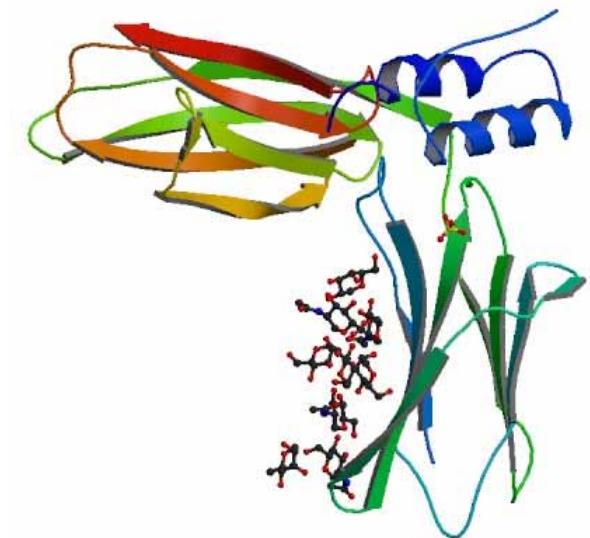
First permanent heteromer
(1976)

Deoxyhemoglobin α/γ chains
1FDH



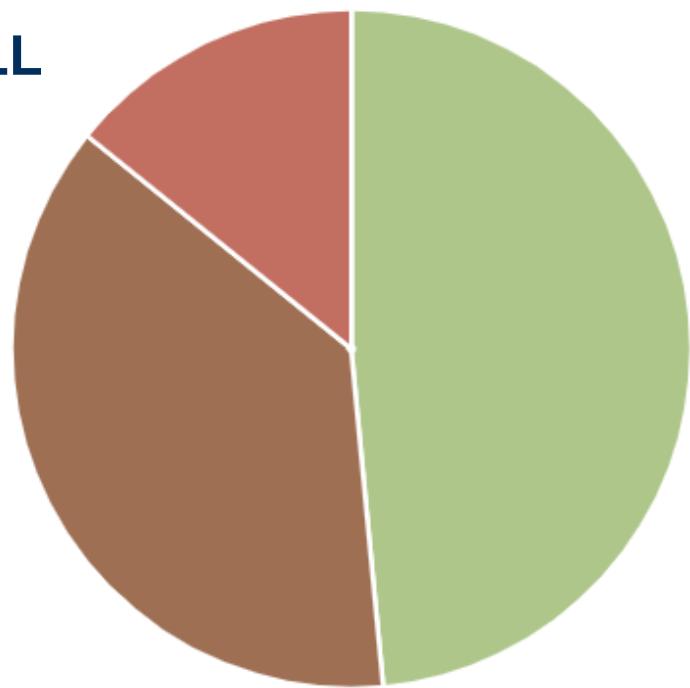
First transient heteromer
(1981)

Ab Fc / Protein A
1FC2



3D structures of protein-protein complexes

ALL



- Monomer (74922)
- Homomer (57593)
- Heteromer (21915)

Non-redundant 100%

35,369 protein monomers

29,572 protein-protein homomers

5,704 protein-protein heteromers

www.pdb.org
Jan 2020

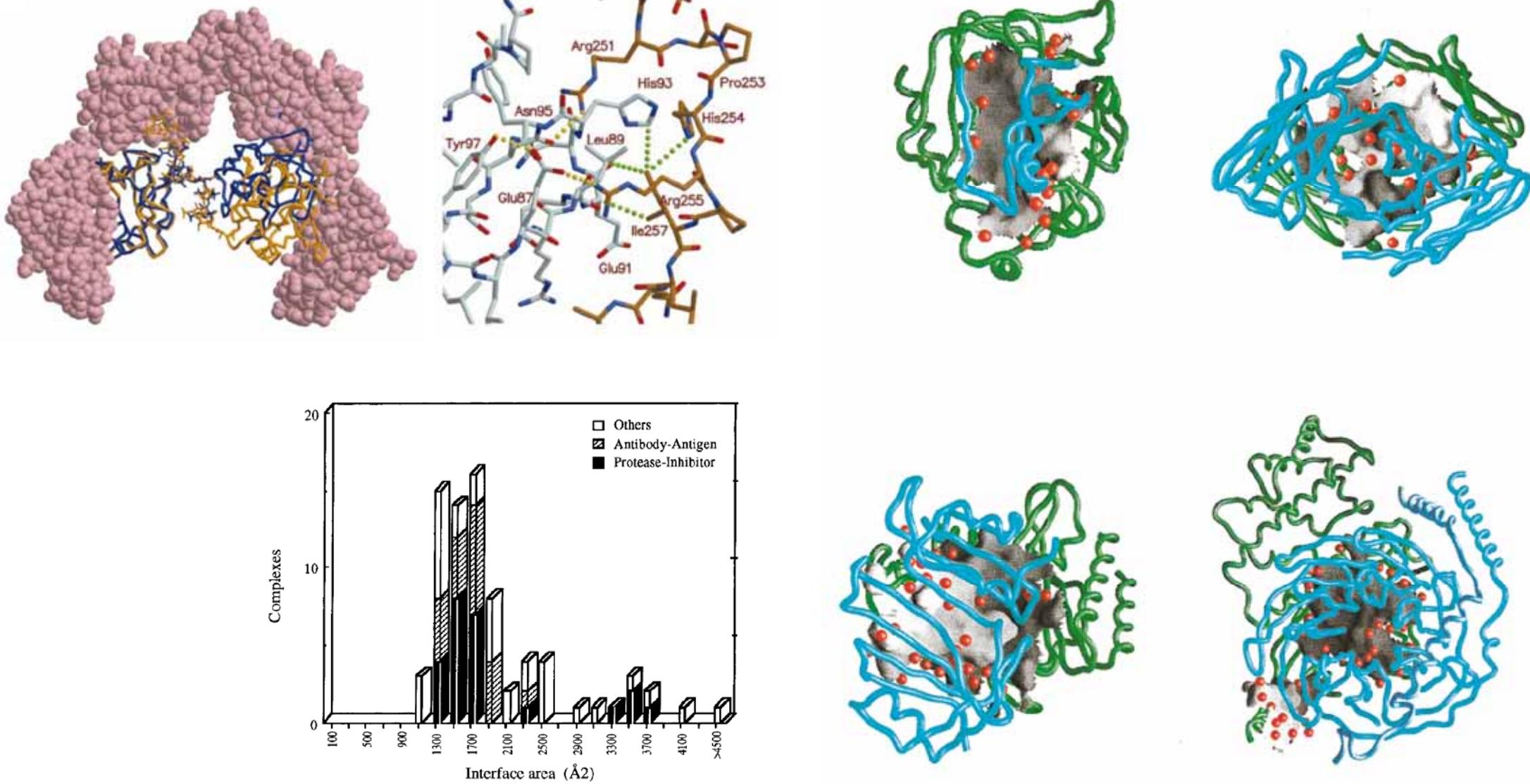
PROTEIN STOICHIOMETRY

- Homo 2-mer - A2 (18642)
- Homo 4-mer - A4 (4847)
- Homo 3-mer - A3 (2377)
- Homo 6-mer - A6 (1493)
- Homo 8-mer - A8 (495)
- Homo 5-mer - A5 (374)
- Homo 12-mer - A12 (343)
- Homo 10-mer - A10 (214)
- Other (787)

PROTEIN STOICHIOMETRY

- Hetero 2-mer - AB (3093)
- Hetero 4-mer - A2B2 (711)
- Hetero 3-mer - ABC (531)
- Hetero 3-mer - A2B (194)
- Hetero 6-mer - A3B3 (166)
- Hetero 6-mer - A2B2C2 (121)
- Hetero 8-mer - A4B4 (91)
- Hetero 4-mer - ABCD (71)
- Other (726)

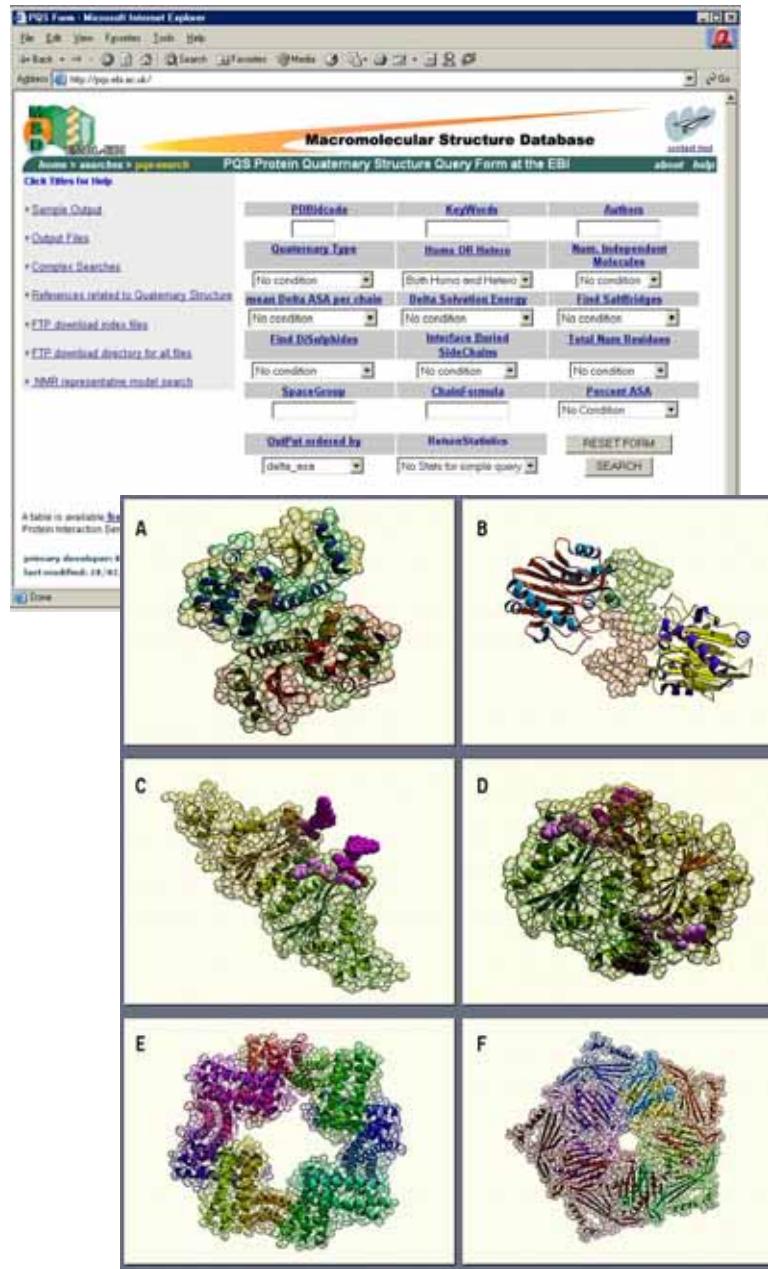
Structural Analysis at Atomic Resolution: NMR and X-ray



Databases of Protein-Protein Complexes

PQS

<http://pq.s.ebi.ac.uk>

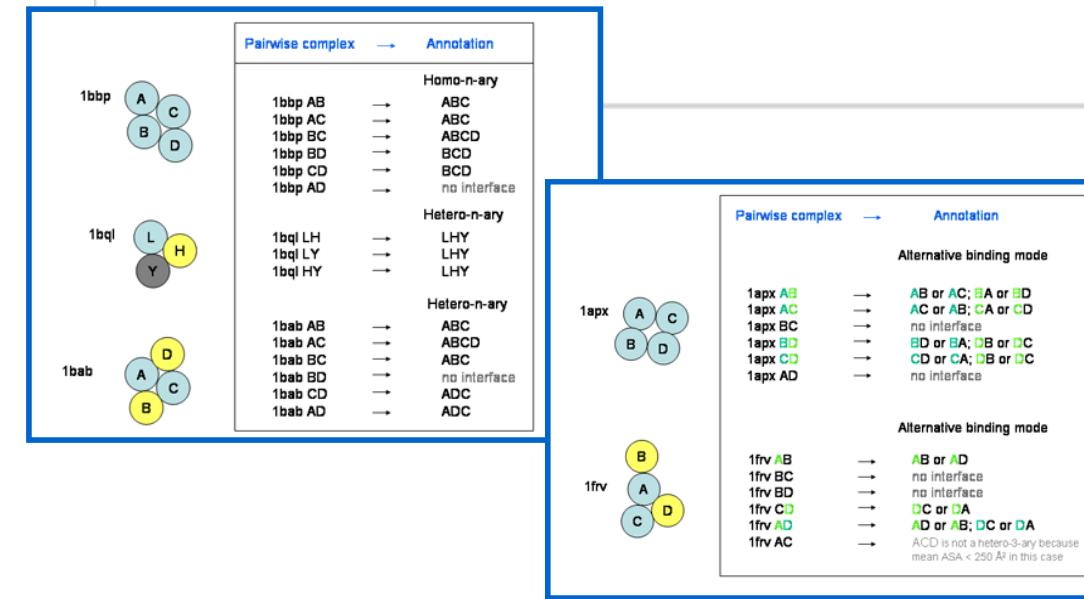


DOCKGROUND

<http://dockground.bioinformatics.ku.edu>

Screenshot of the DOCKGROUND website. It features a header with the site's name and a main menu. Below the menu, there's a section for "Protein-Protein Complexes" with tabs for Bound, Unbound, and Model. A descriptive text box explains the database's purpose: "Database of protein-protein experimentally determined & simulated unbound structures".

A screenshot of the DOCKGROUND website. The header includes the site's name and a menu with links like File, Edit, View, Go, Bookmarks, Tools, and Help. Below the header is a search bar and a Google search result. The main content area has a green header "Dockground Integrated system of databases for protein recognition studies". It features a "Protein-Protein Complexes" section with tabs for Bound, Unbound, and Model. A descriptive text box states: "Database of protein-protein experimentally determined & simulated unbound structures".



Databases of Protein-Protein Complexes

3Dcomplex

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

The screenshot shows the homepage of 3DComplex.org v2.0. It features a large image of a crystallographic structure of a protein complex composed of various colored subunits (red, yellow, green). To the right, there is a 3D graph representation of the complex, where nodes represent amino acid residues and edges represent interactions. A legend explains the symbols: identical chains (two yellow circles), homologous chains (yellow and red circles), and different chains (yellow, red, and green circles). A callout box provides details about the graph: identities, homologies, and contacts. Below the graph is a table showing the number of amino-acids in contact for each residue.

Residue	Number of amino-acids in contact (average on both chains)
1	12
5	12
9	12
14	12
17	17
19	19
23	23
33	33
35	35
39	39
43	43
47	47
57	57

Crystallographic structure of a complex

3D Complex graph representation

identical chains
homologous chains
different chains
Number of amino-acids in contact (average on both chains)

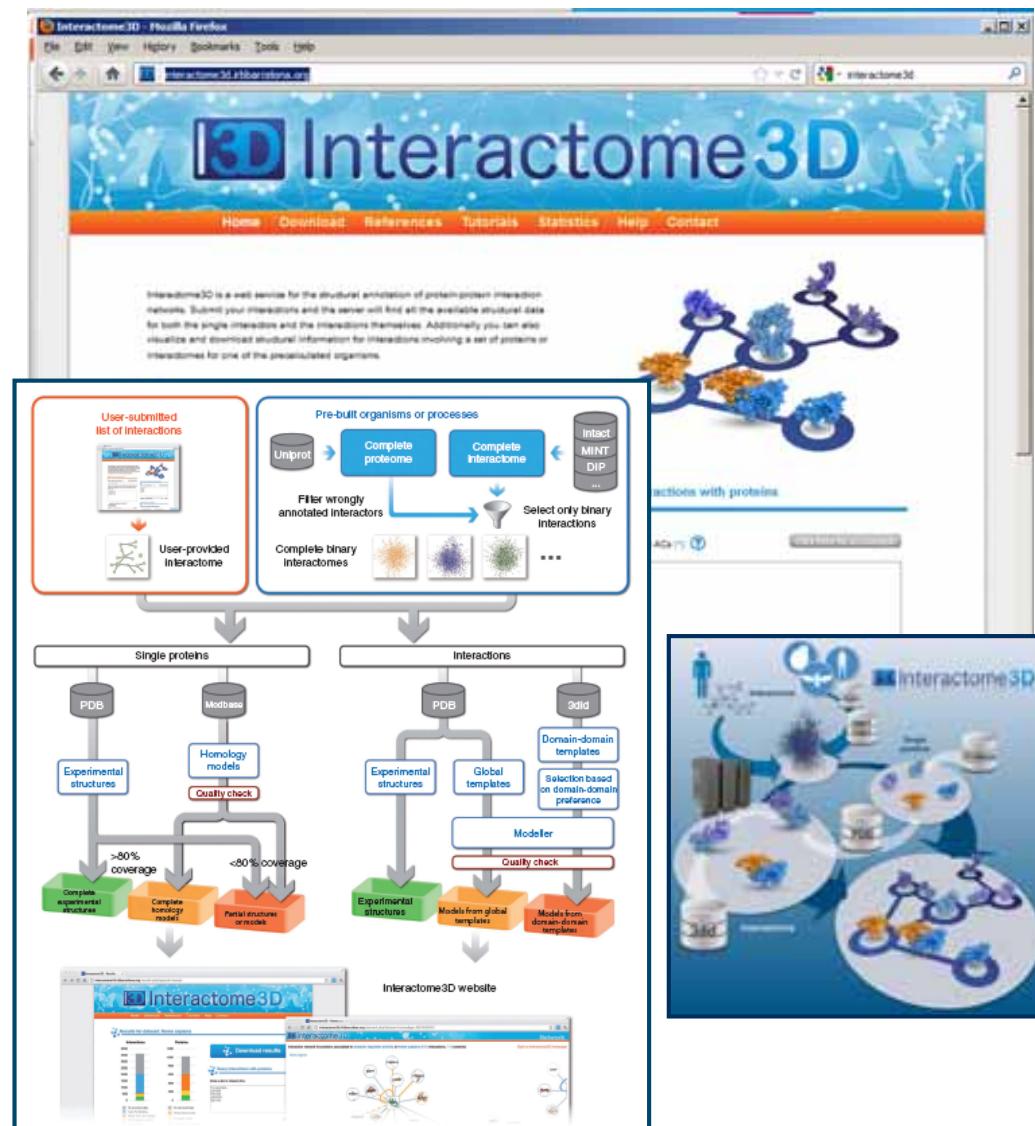
NPS

Cyclic

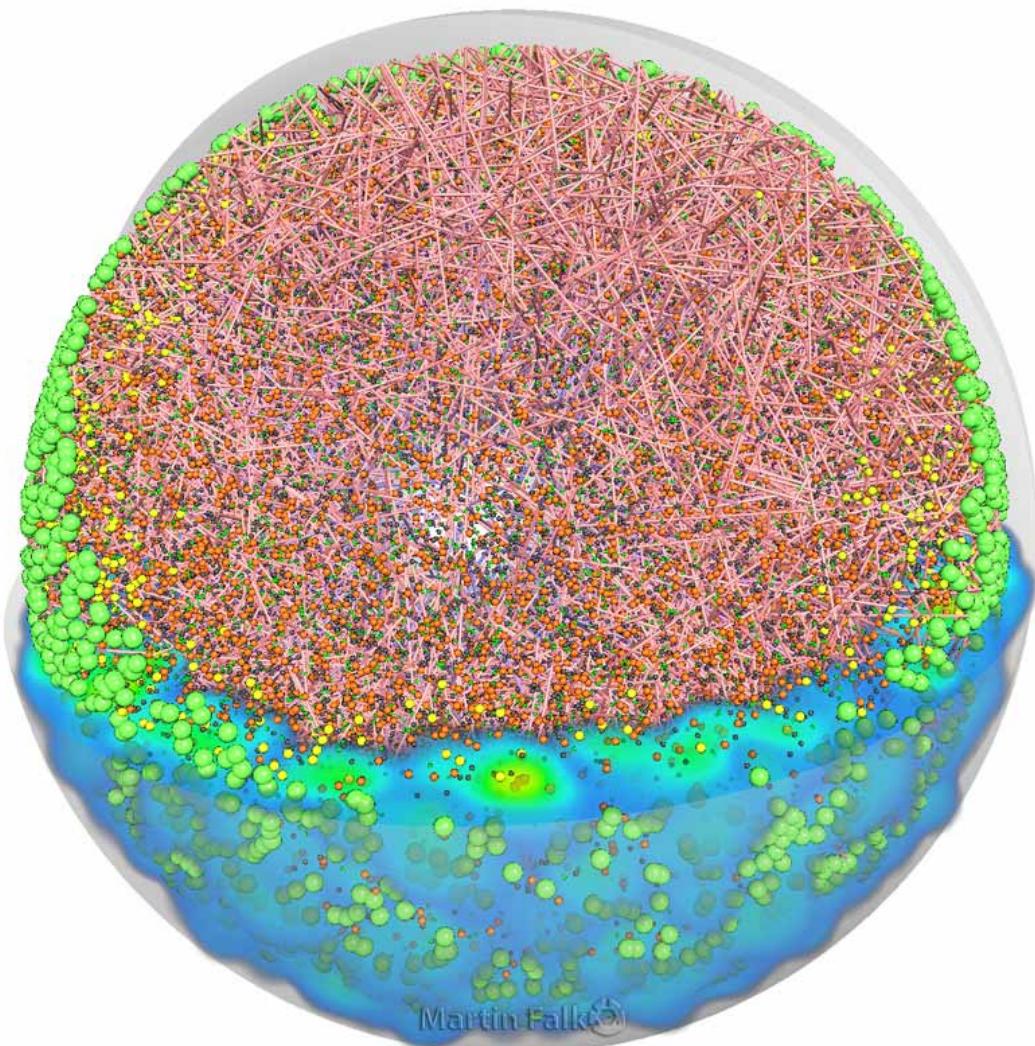
Dihedral

Interactome3D

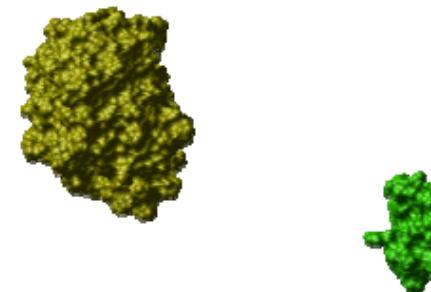
<http://interactome3d.irbbarcelona.org/>



Protein-Protein Interactions

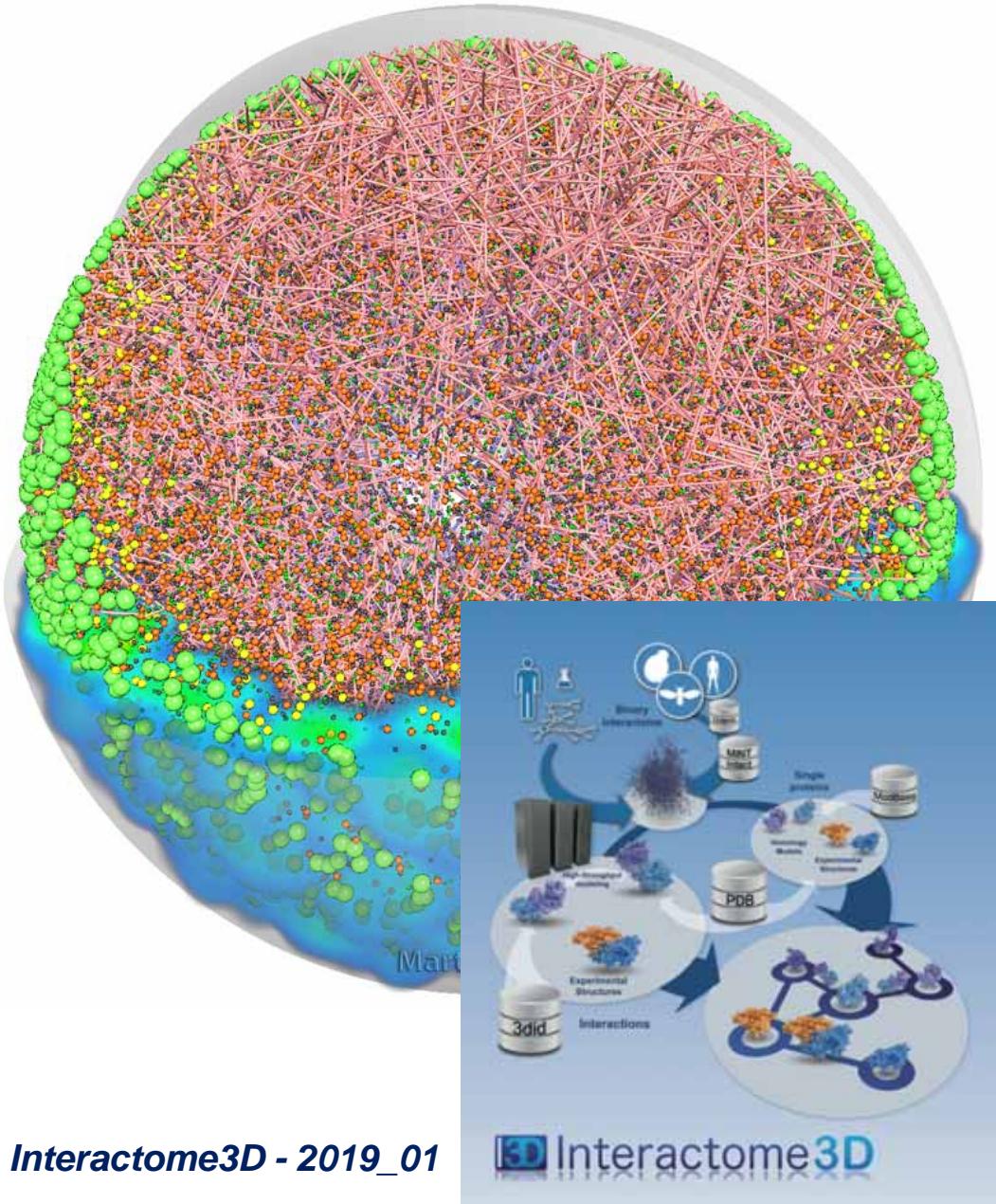


Human proteome
20K estimated proteins
4K with known structure
+ 6K homologous structure >50%ID
+4K homologous structure 30-50%ID
(50-70% structural coverage) #

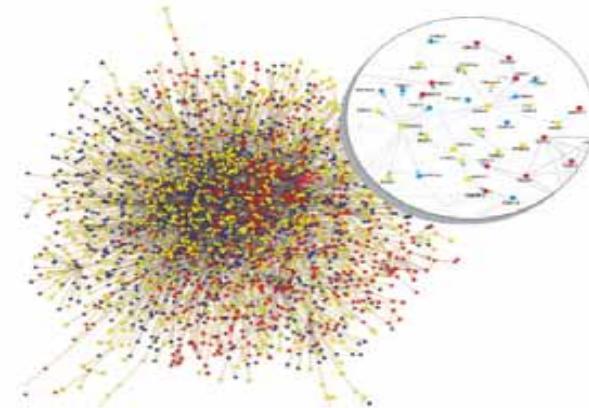


Somody et al 2017 Drug Discov. Today 22, 1792

Protein-Protein Interactions



Human proteome
20K estimated proteins
4K with known structure
+ 6K homologous structure >50%ID
+4K homologous structure 30-50%ID
(70% structural coverage) #



Human interactome
130-650K estimated interactions*
120K confirmed interactions *
7K with known structure
7K with homologous structure >30%ID
(2-10% structural coverage)

Somody et al 2017 *Drug Discov. Today* 22, 1792

* Venkatesan et al 2009 *Nature Methods* 6, 83-90

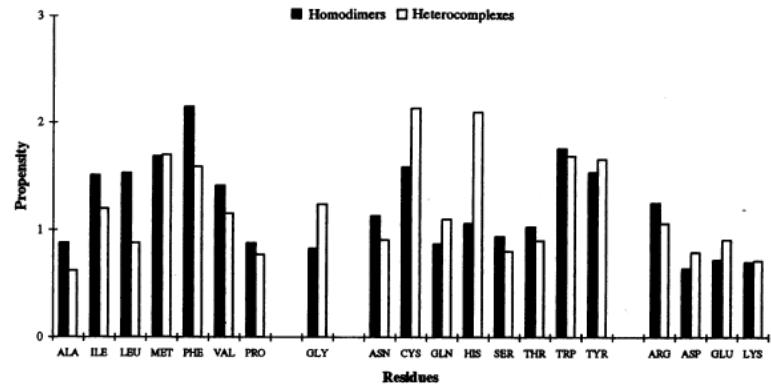
Stumpf et al 2008 *PNAS* 105, 6959-6964

Mosca et al. 2013 *Nature Methods* 10, 47

- Importance of protein interactions
- Study of protein interaction networks
- Types of complexes
- Computational analysis and structural modeling
- Protein-protein docking
- Recognition and prediction of interaction sites

Types of protein-protein interactions

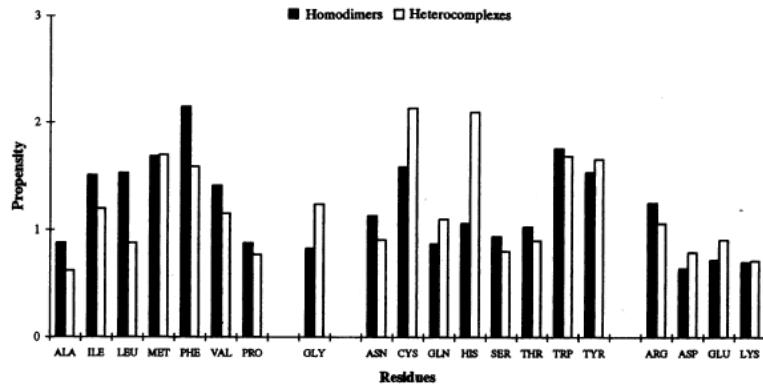
PERMANENT / NON-OBLIGATORY



Jones & Thornton (1996) PNAS, 93, 13

Types of protein-protein interactions

PERMANENT / NON-OBLIGATORY



Jones & Thornton (1996) PNAS, 93, 13

OBLIGOMERS / COMPLEXES

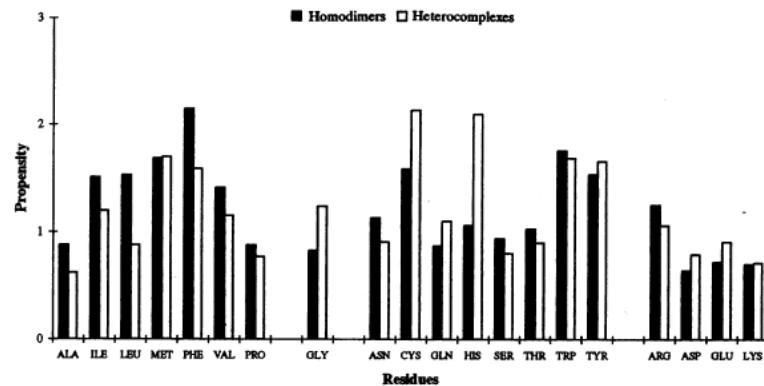
HOMO- / HETERO-

Type of interface	Number of contacts
<i>External</i>	
Homo-oligomers	218,104
Homo-complexes	3077
Hetero-oligomers	18,886
Hetero-complexes	166,412

Ofran & Rost (2003) JMB, 325, 377

Types of protein-protein interactions

PERMANENT / NON-OBLIGATORY



Jones & Thornton (1996) PNAS, 93, 13

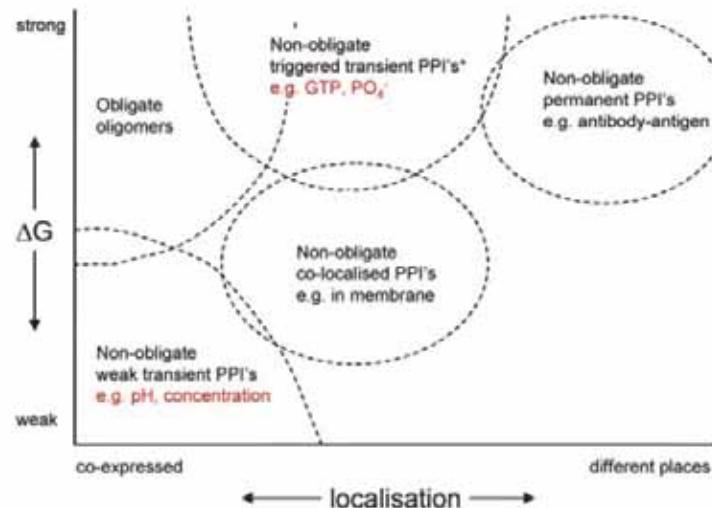
OBLIGOMERS / COMPLEXES

HOMO- / HETERO-

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

Ofran & Rost (2003) JMB, 325, 377

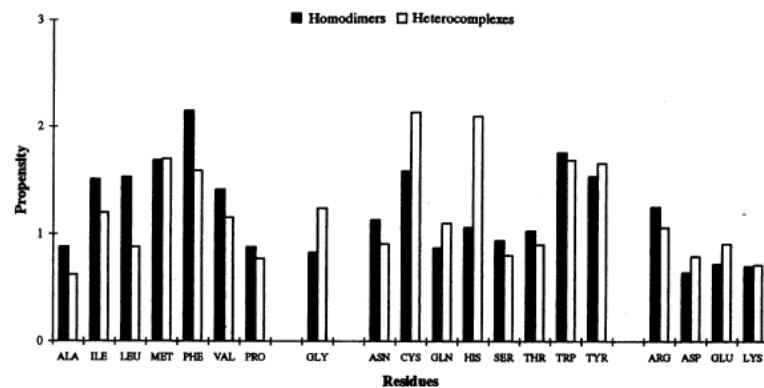
NON-OBLIGATE / OBLIGATE HOMO- / HETERO-OLIGOMERIC TRANSIENT / PERMANENT



Nooren & Thornton (2003) EMBO J, 22, 3486

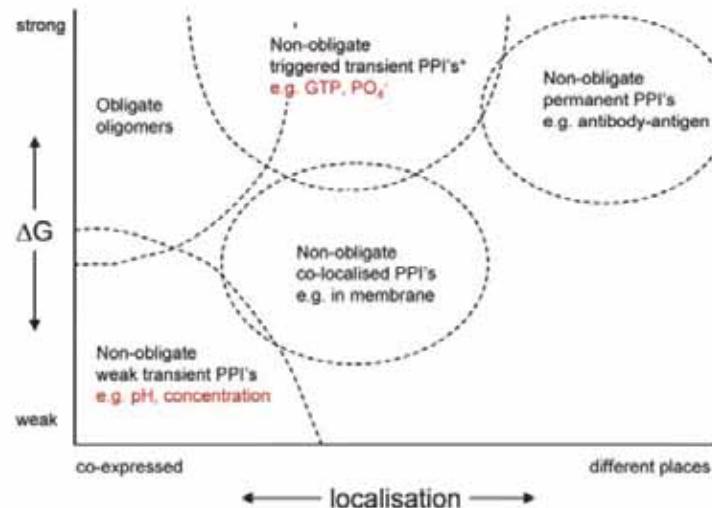
Types of protein-protein interactions

PERMANENT / NON-OBLIGATORY



Jones & Thornton (1996) PNAS, 93, 13

NON-OBLIGATE / OBLIGATE HOMO- / HETERO-OLIGOMERIC TRANSIENT / PERMANENT



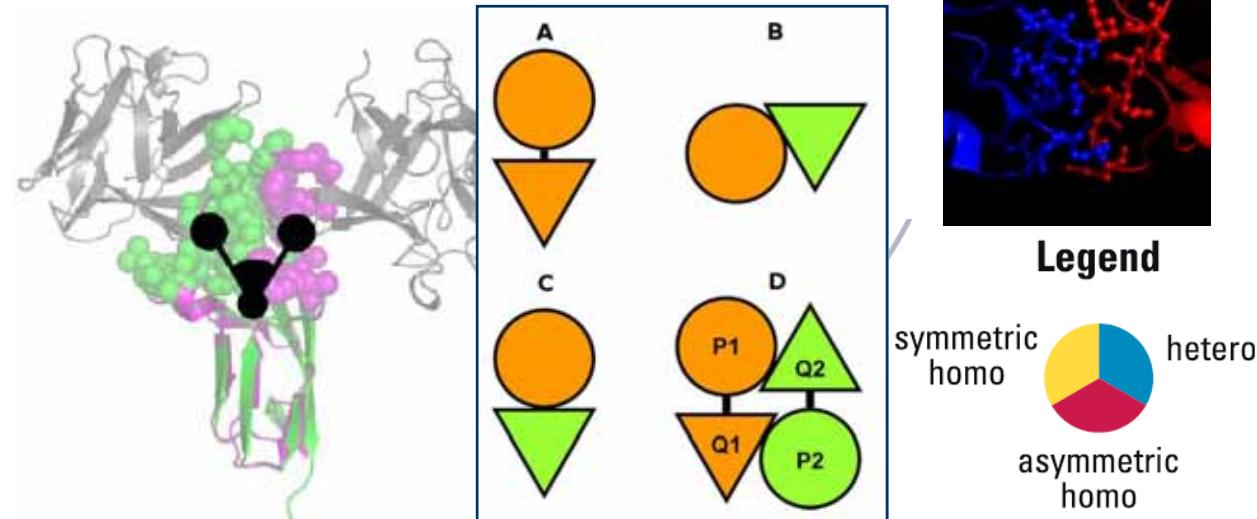
Nooren & Thornton (2003) EMBO J, 22, 3486

OBLIGOMERS / COMPLEXES HOMO- / HETERO-

Type of interface	Number of contacts
<i>External</i>	
Homo-oligomers	218,104
Homo-complexes	3077
Hetero-oligomers	18,886
Hetero-complexes	166,412

Ofran & Rost (2003) JMB, 325, 377

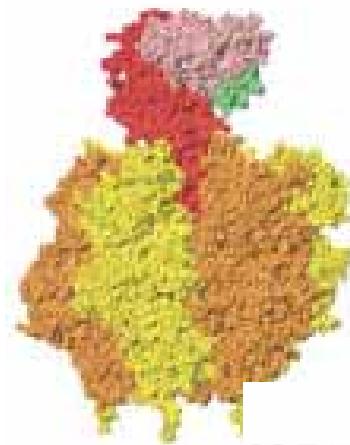
DOMAIN-DOMAIN INTERFACE CLASSIFICATION PERMANENT / TRANSIENT, SYMMETRIC...



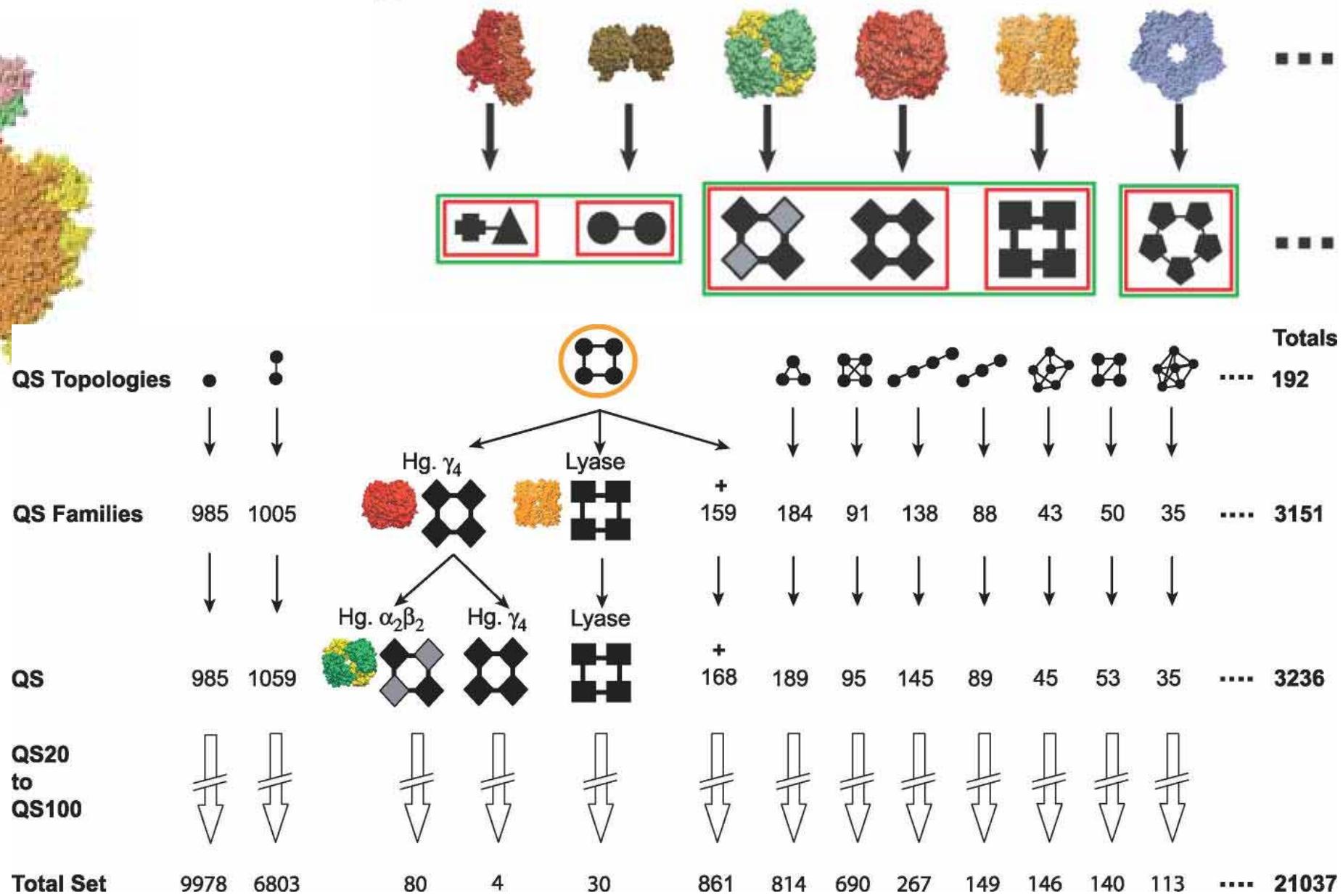
Kim et al. (2006) PLOS Comp Biol, 2, e124 (<http://www.scoppi.org>)

Types of protein-protein interactions

MULTI-MOLECULAR ASSEMBLIES

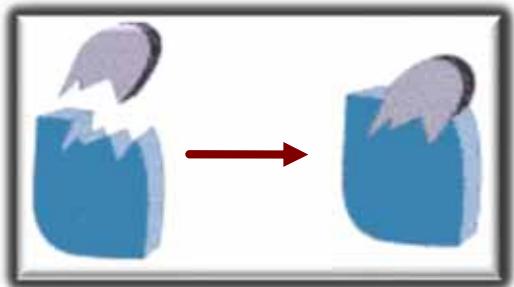


3Dcomplex, <http://supfam.mrc-lmb.cam.ac.uk/elevy/3dcomplex/Home.cgi>



Protein-protein binding mechanism

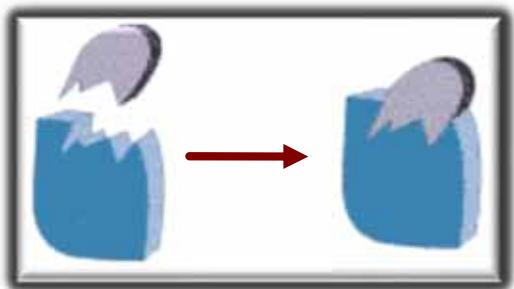
1894: LOCK AND KEY: E. FISCHER



*MOLECULES STERIC COMPLEMENTARY
ESSENTIAL FOR AFFINITY AND SPECIFICITY
IN THE BINDING.*

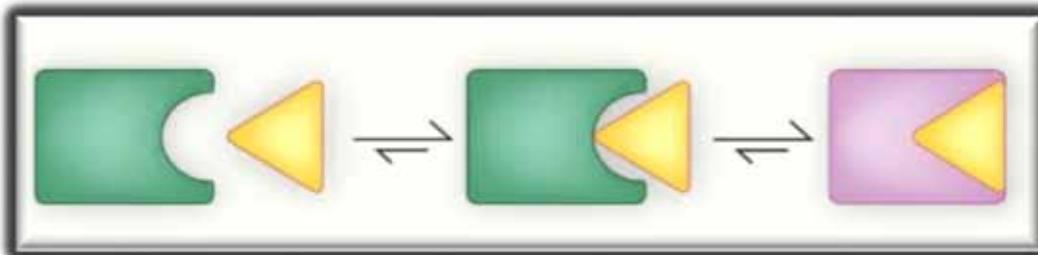
Protein-protein binding mechanism

1894: LOCK AND KEY: E. FISCHER



*MOLECULES STERIC COMPLEMENTARY
ESSENTIAL FOR AFFINITY AND SPECIFICITY
IN THE BINDING.*

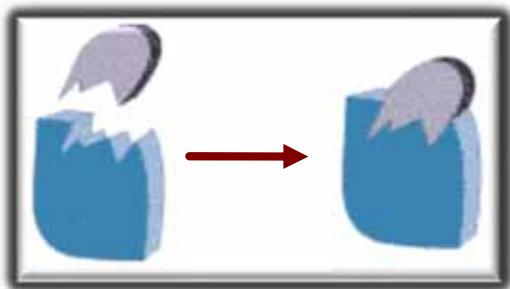
1958: INDUCED FIT: D. E. KOSHLAND



*INITIAL CONFORMATIONAL CHANGES
INDUCE THE FINAL BINDING.*

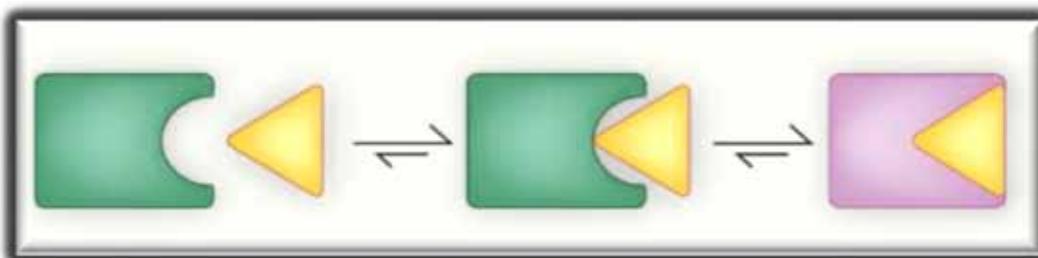
Protein-protein binding mechanism

1894: LOCK AND KEY: E. FISCHER



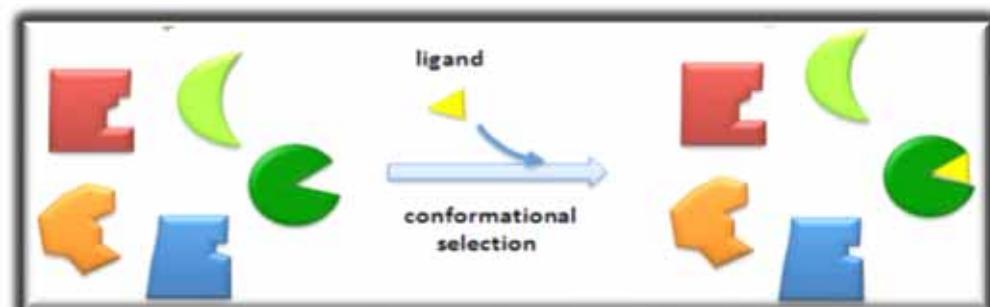
MOLECULES STERIC COMPLEMENTARITY
ESSENTIAL FOR AFFINITY AND SPECIFICITY
IN THE BINDING.

1958: INDUCED FIT: D. E. KOSHLAND

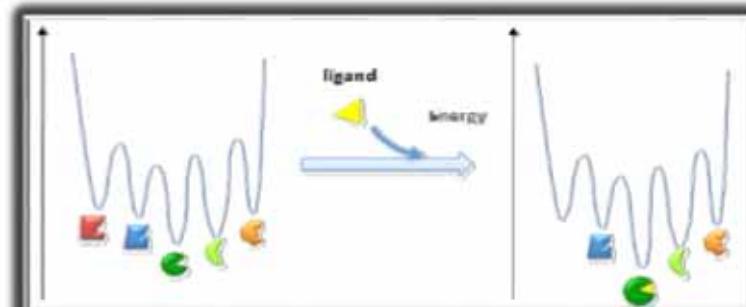


INITIAL CONFORMATIONAL CHANGES
INDUCE THE FINAL BINDING.

1999: CONFORMATIONAL SELECTION: R. NUSSINOV



PROTEIN AS CONFORMATIONAL ENSEMBLE



INITIAL INTERACTION, EQUILIBRIUM SHIFT AND FINAL
PROTEIN BINDING.

- Importance of protein interactions
- Study of protein interaction networks
- Types of complexes
- Computational analysis and structural modeling
- Protein-protein docking
- Recognition and prediction of interaction sites

First computer studies on protein-protein interactions

Proc. Nat. Acad. Sci. USA
Vol. 72, No. 4, pp. 1330-1334, April 1975

Hemoglobin Interaction in Sickle Cell Fibers I: Theoretical Approaches to the Molecular Contacts

(model building/protein interactions/computer graphics)

CYRUS LEVINTHAL, SHOSHANA J. WODAK, PETER KAHN, AND ARTEM K. DADIVANIAN

Department of Biological Sciences, Columbia University, New York, N.Y. 10027

Contributed by Cyrus Levinthal, January 10, 1975

Wodak SJ, Janin J (1978). "Computer Analysis of Protein-Protein Interactions".
Journal of Molecular Biology **124** (2): 323-42.

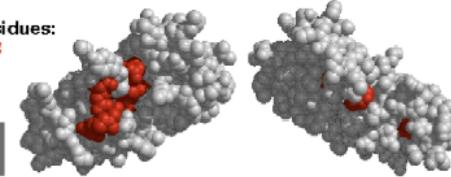
ET

Residue interface prediction

Evolutionary Trace of Protein Functional Determinants

Contact Residues:
167 83 87 88

A



Contact Residues:
167 83 87 88 128,
159 126
Novel Residues:
123

B



Contact Residues:
167 83 87 88 128 159
126 131 163 134

Novel Residues:
123 77 117,121
122 124 129

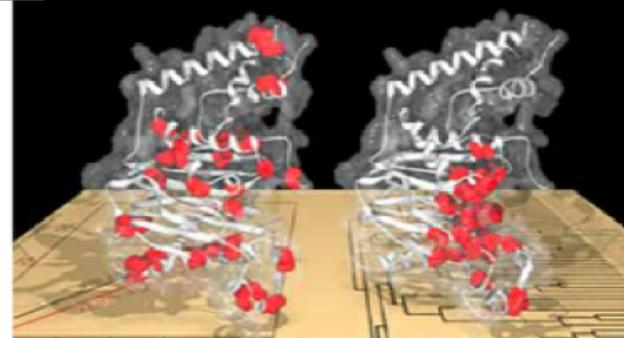
C



Contact Residues:
167 83 87 88 128 163
159 126 131 134 84

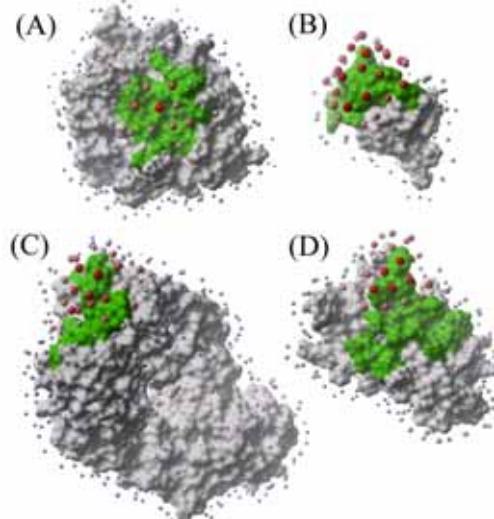
D

A ET residues cluster

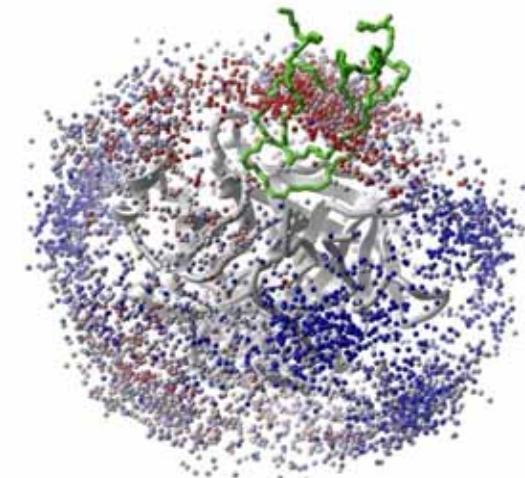


Random

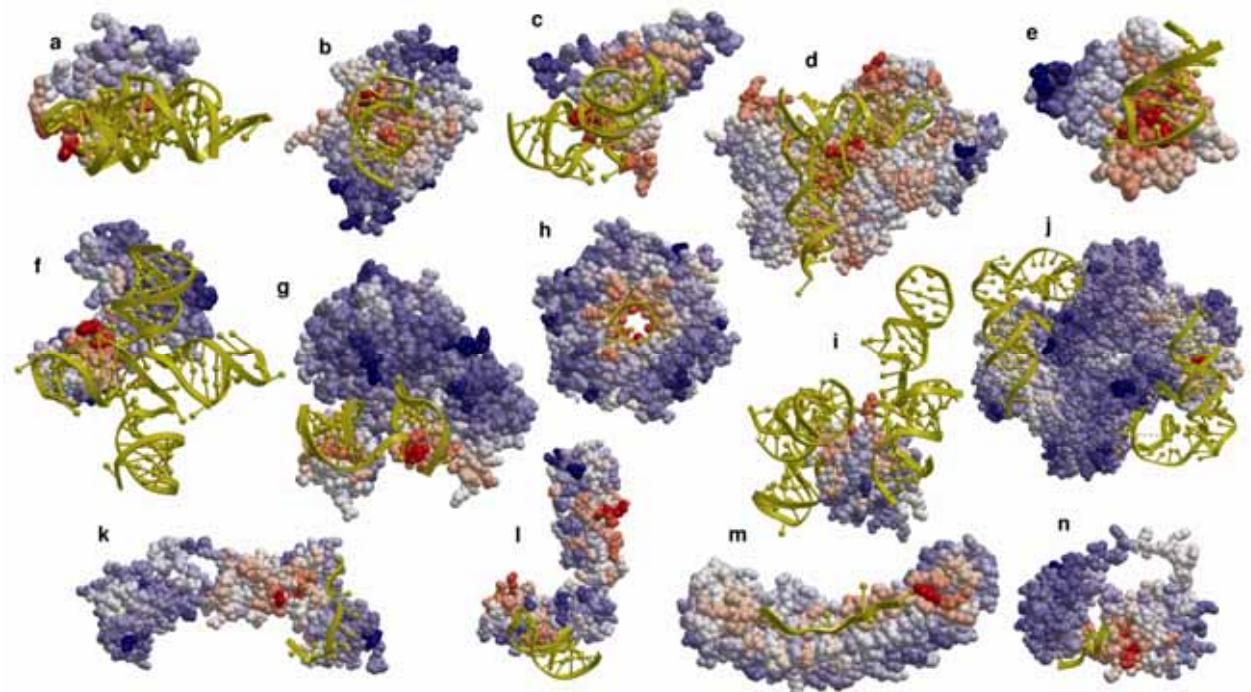
ET



ODA



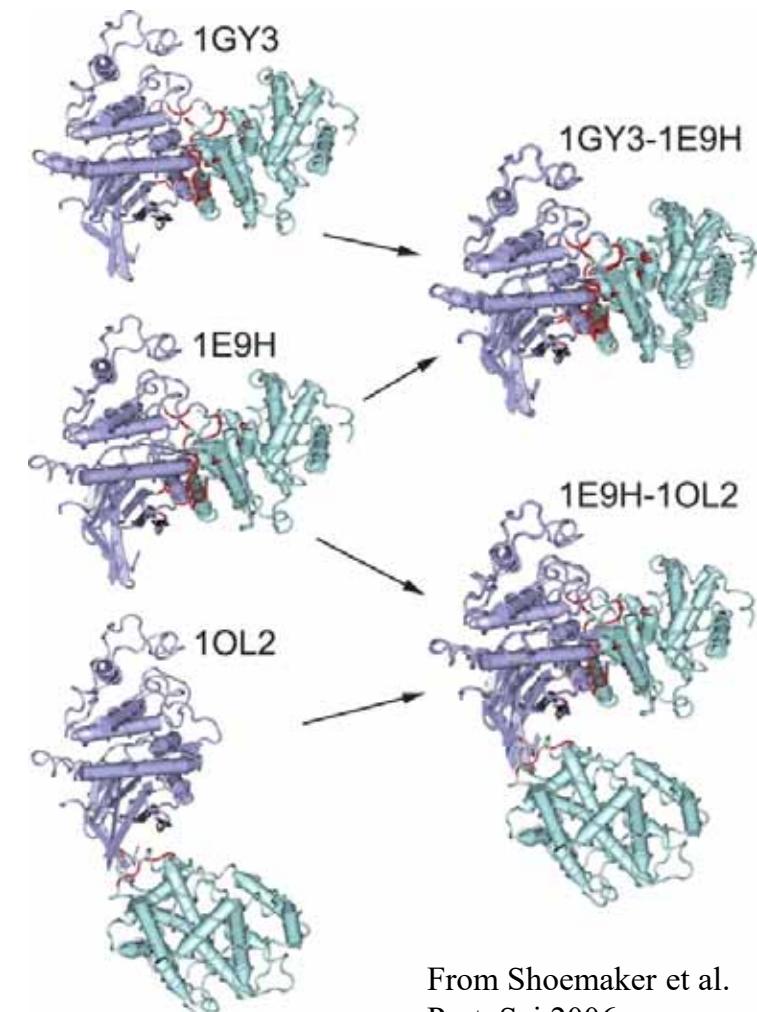
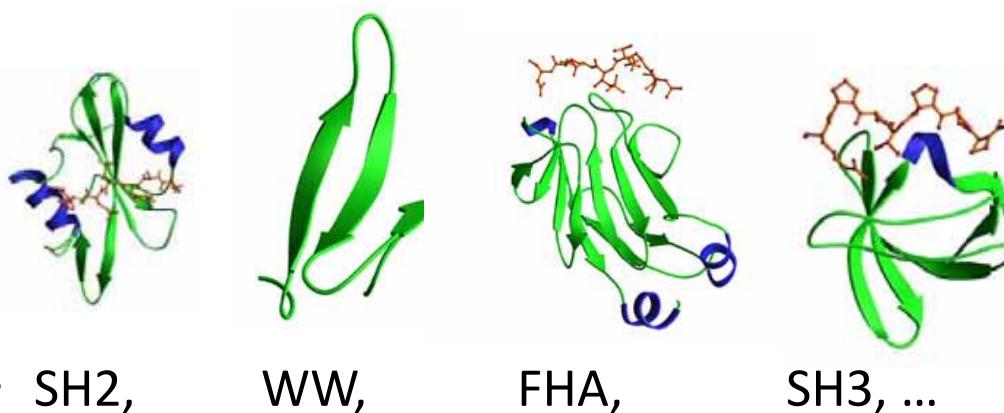
NIP



OPRA

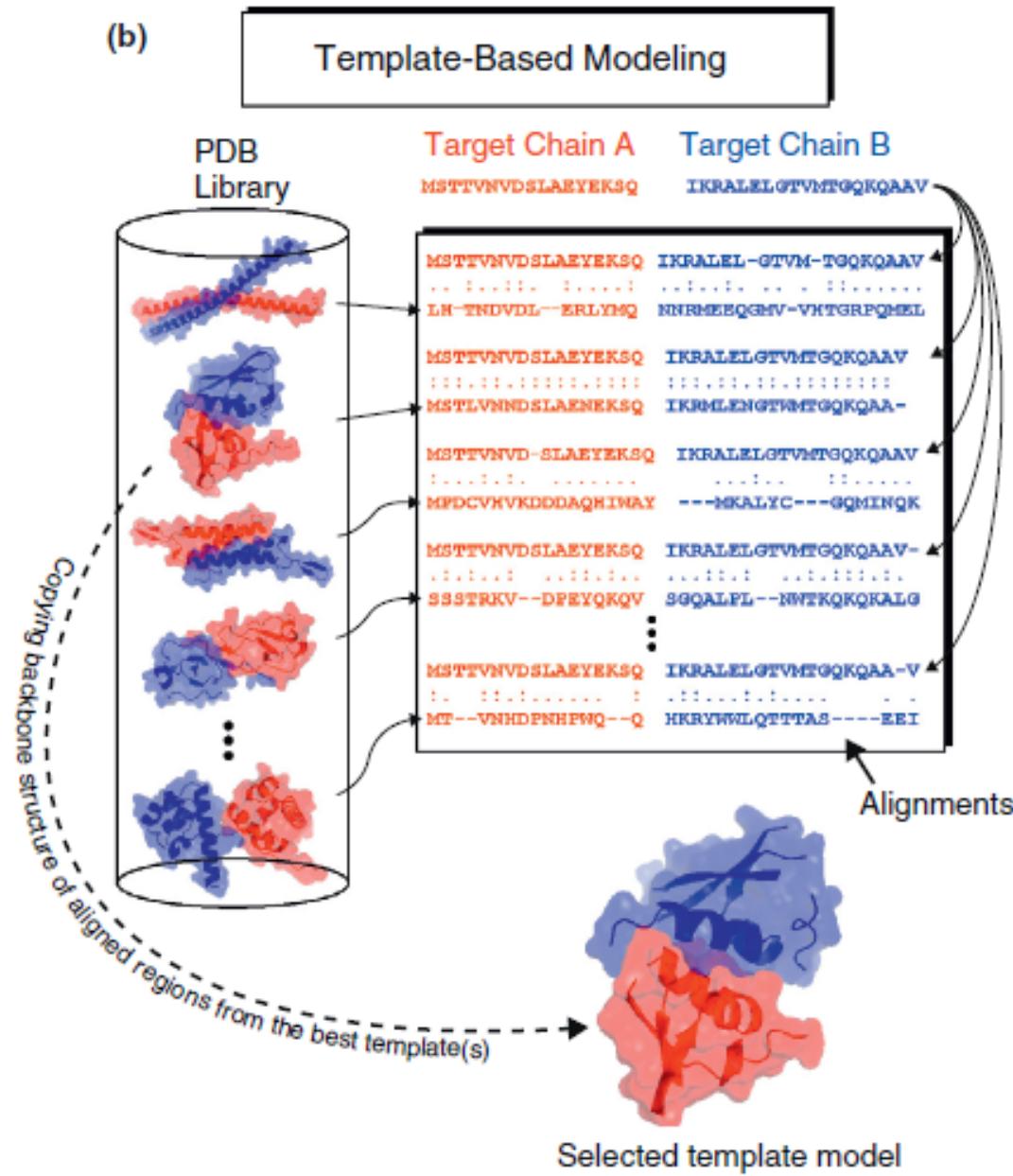
Domain-domain interactions

- Conserved interfaces/binding modes among interacting domains
- Specialized protein-protein interacting domains



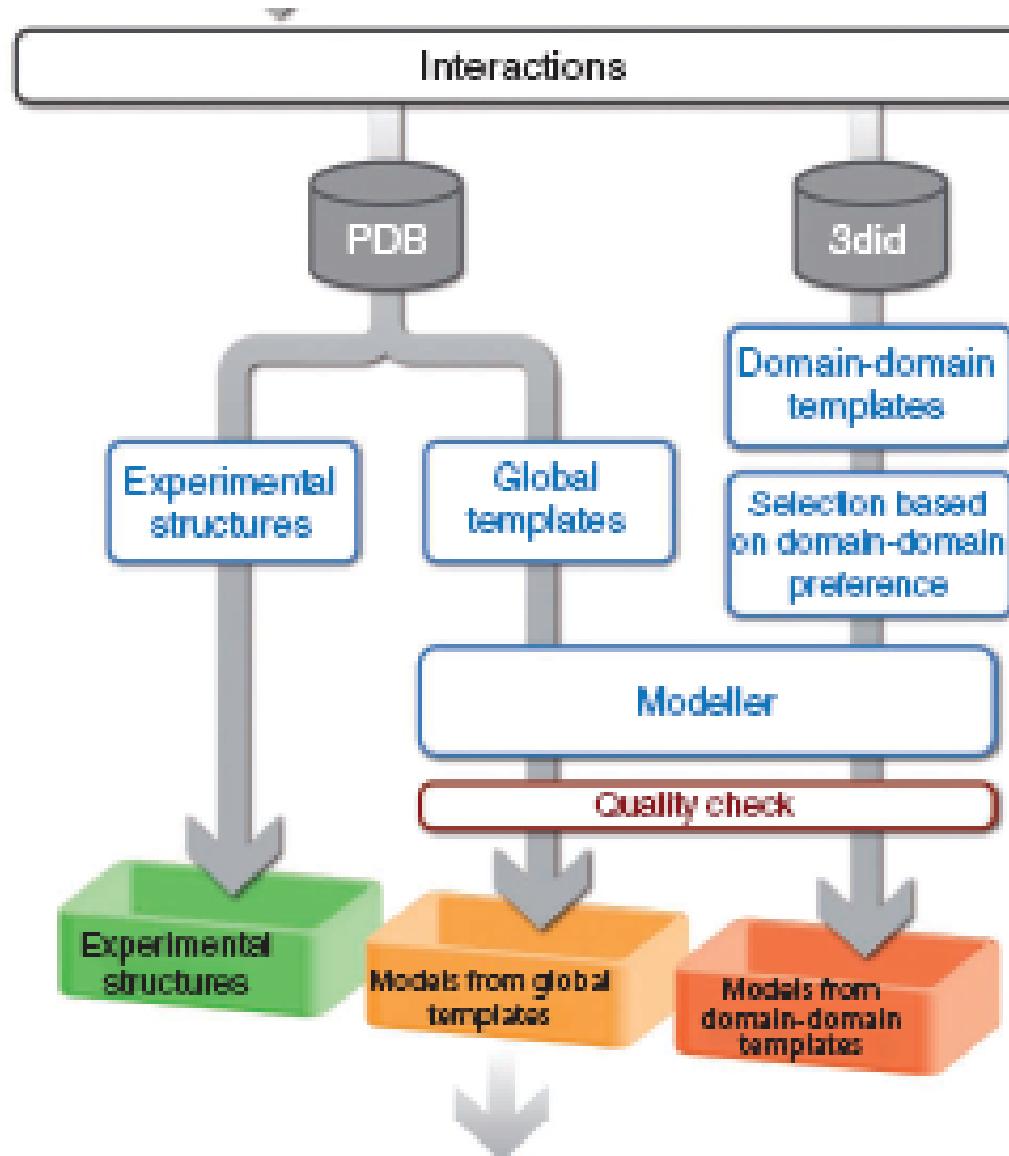
From Shoemaker et al.
Prot. Sci 2006

Complex structure prediction



template-based docking

sequence-based alignment



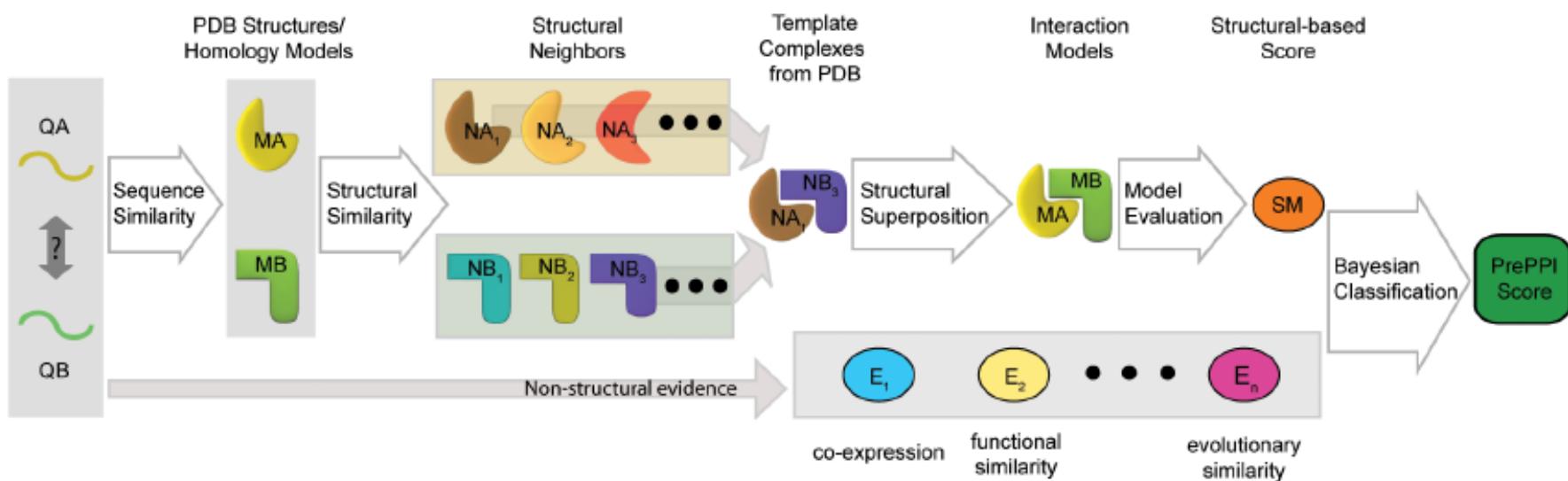
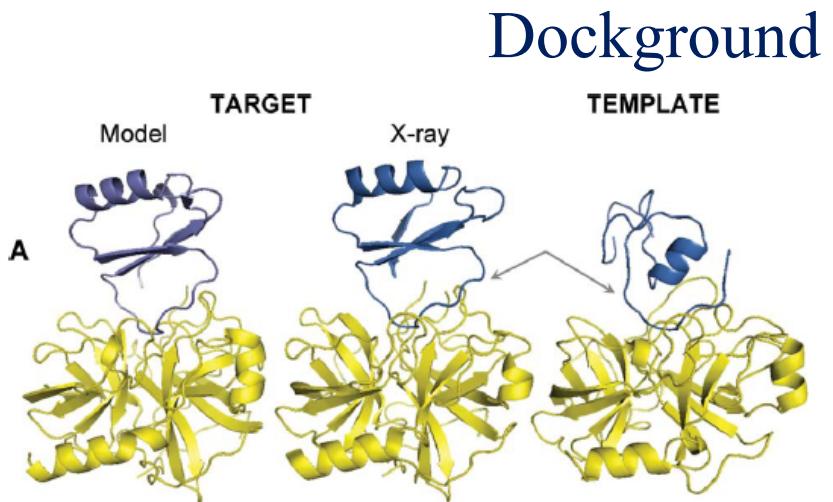
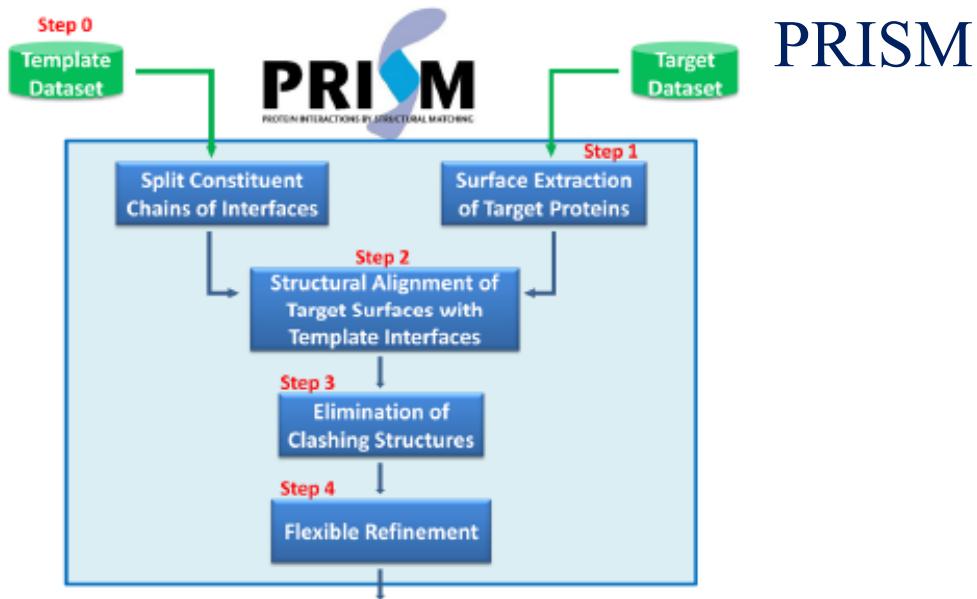
Interactome3D

7K can be homol. modelled
(seq. id > 30%)
(57% success rate)

Mosca et al. 2013 Nature Methods 10, 47

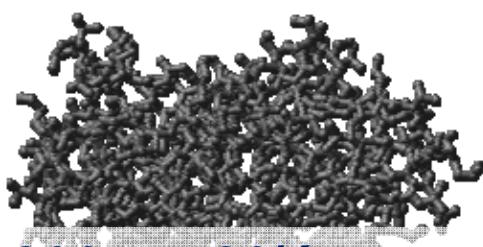
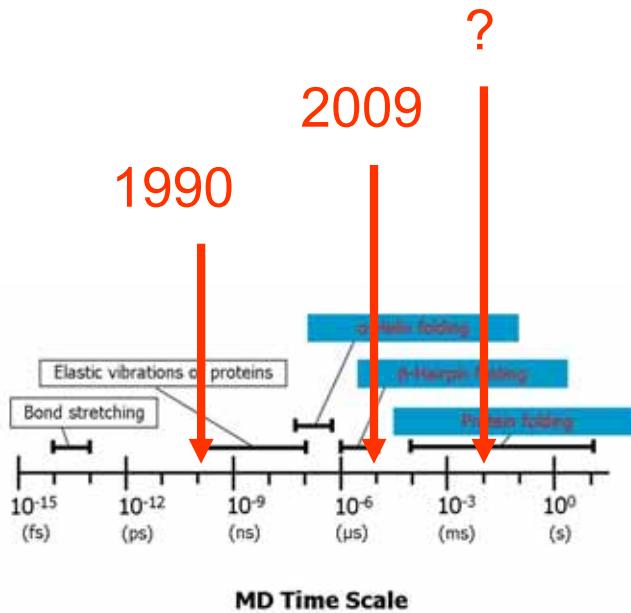
template-based docking

structure-based alignment

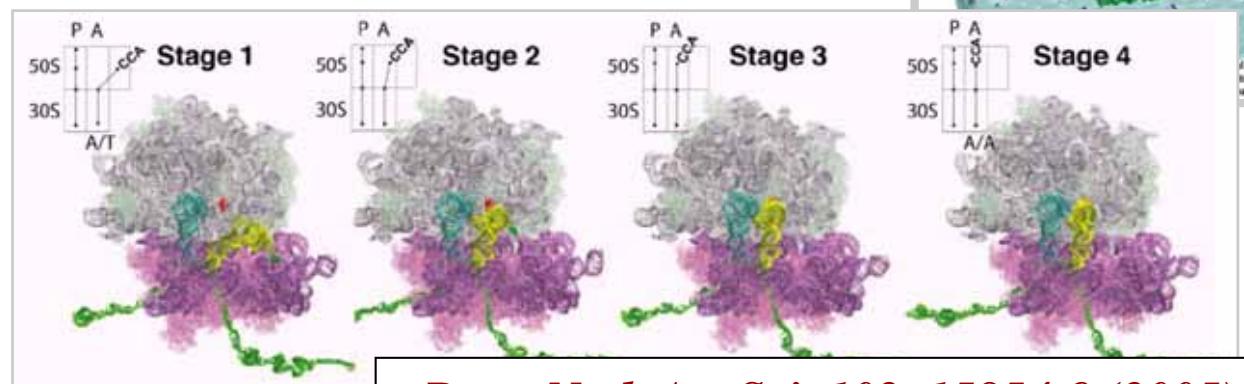
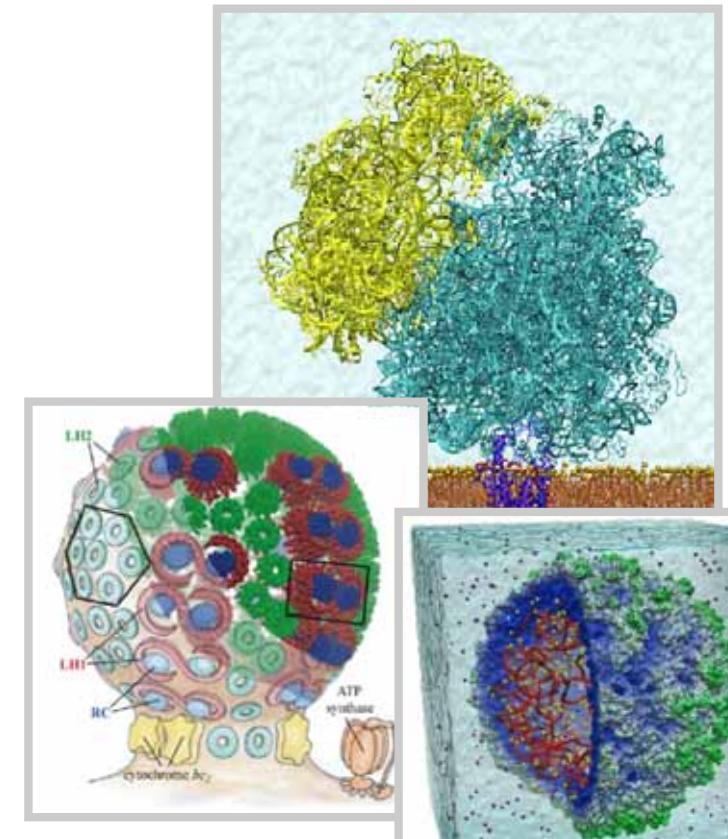


24-300K?? could be homol. modelled (low success rate)

ab initio modeling: molecular dynamics



300res, 30K atoms
→ 10ns, 200 MN CPU/days
→ 10ms 55K MN CPU/years
(50 years in entire MN)

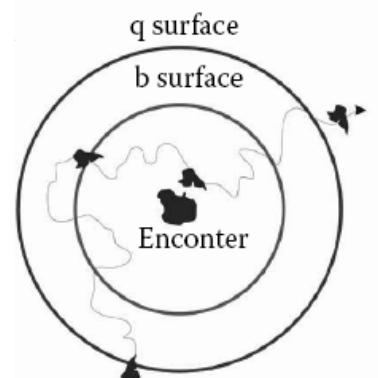


Simulations of association

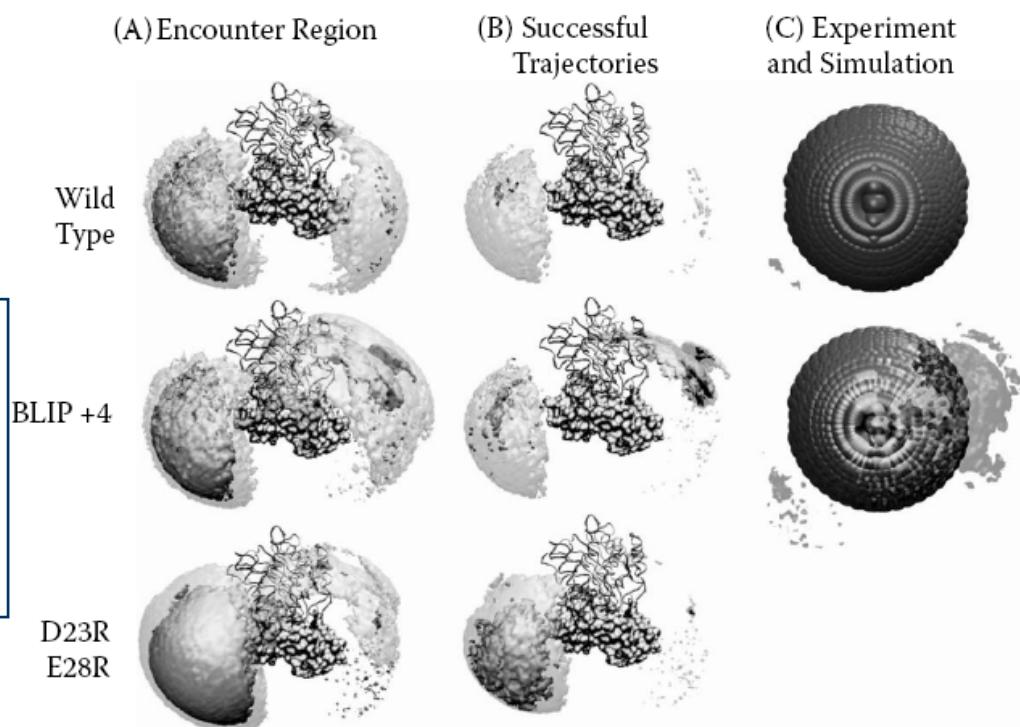
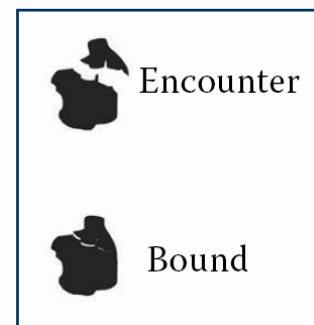
Molecular dynamics



Brownian dynamics



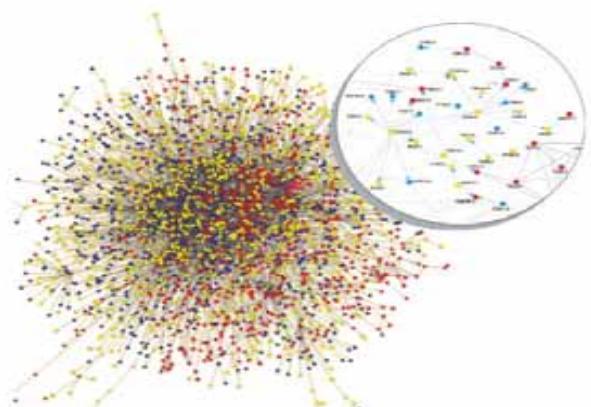
$$k_{\text{on}} = \frac{4\pi D b\beta}{1-(1-\beta)b/q}$$



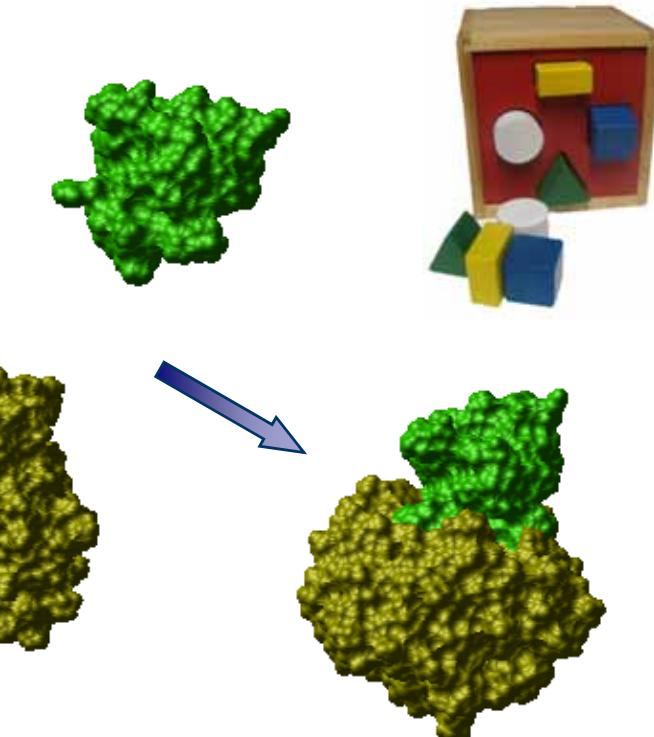
- **Importance of protein interactions**
- **Study of protein interaction networks**
- **Types of complexes**
- **Computational analysis and structural modeling**
- **Protein-protein docking**
- **Recognition and prediction of interaction sites**

Complex structure prediction (docking)

Generation of the structure of a protein-protein complex from the individual protein structures



Human interactome
130-650K estimated interactions*
120K confirmed interactions *
7K with known structure
7K with homologous structure >30%ID
(2-10% structural coverage)

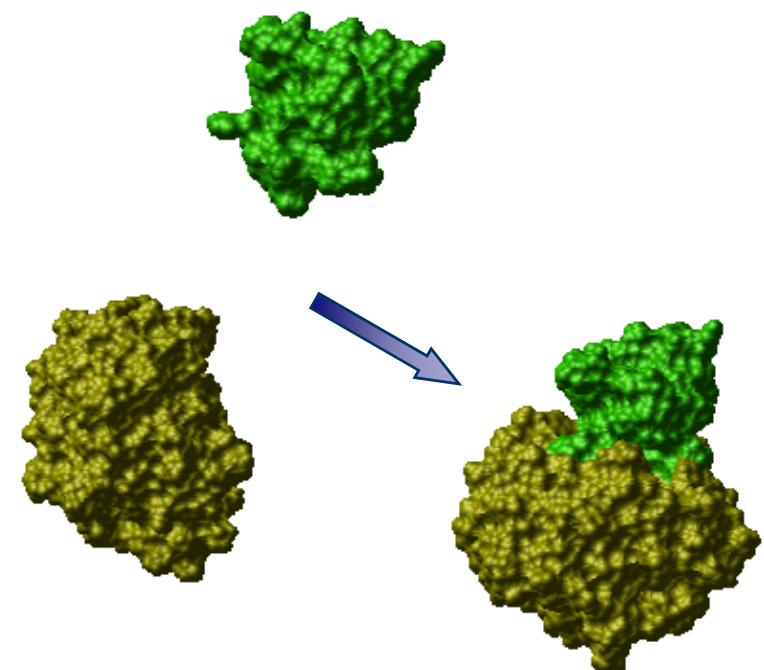


Why Is Docking A Hard Problem?

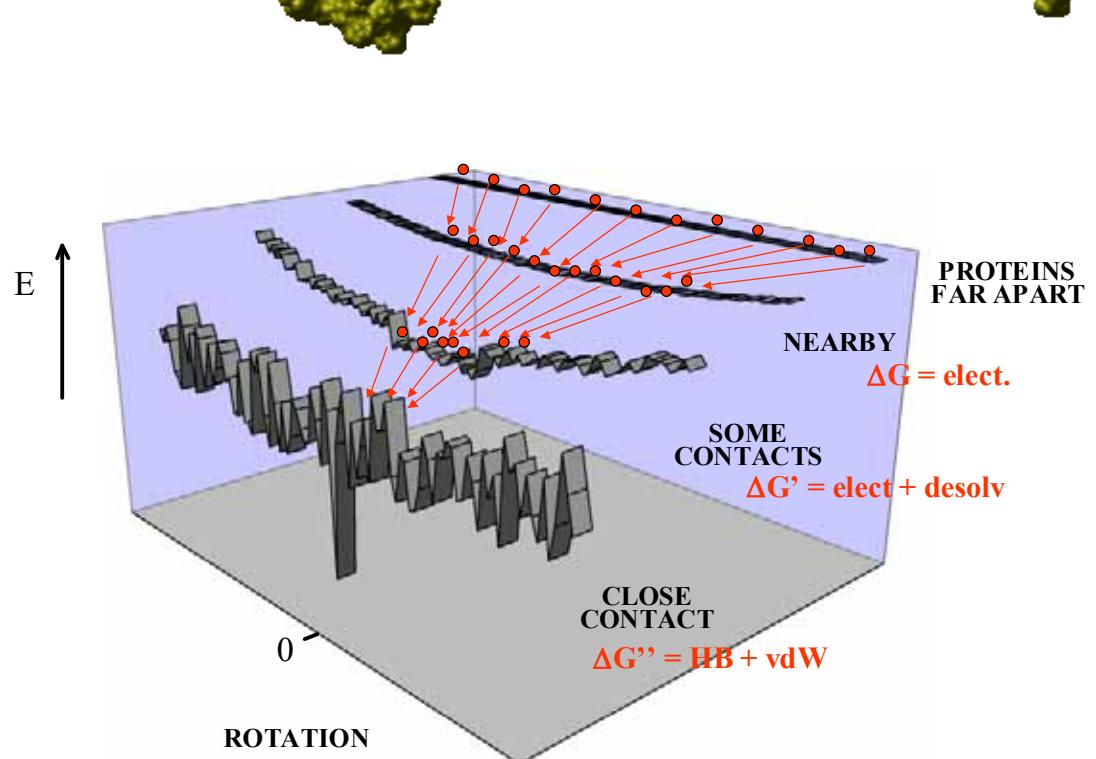
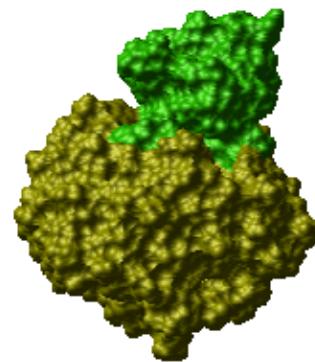
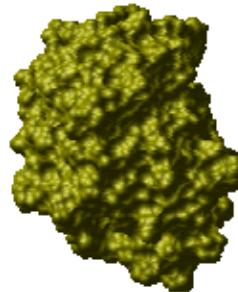
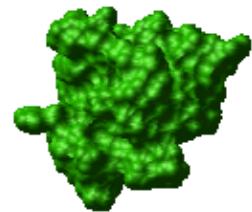
ab initio docking



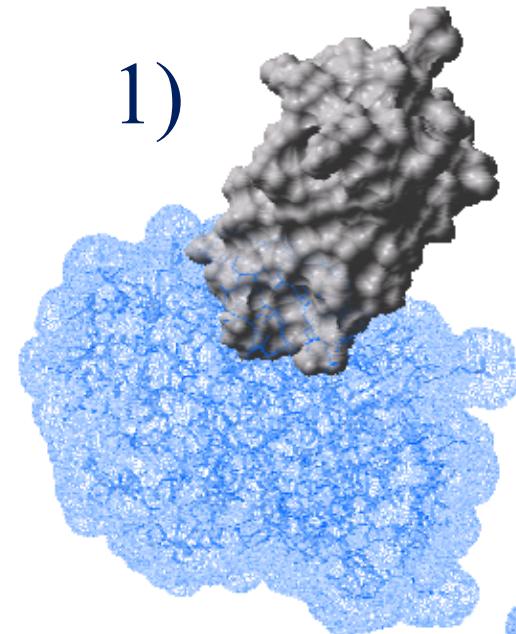
- Large Search Space → “Hard”
- “The Curse of Dimensionality”
- Sequence Alignment: 1D - Easy!
- Structure Alignment: 3-6D - Hard
- Rigid Body Docking: 6D - Hard!
- Flexible Docking: $3N$ - “Impossible!”
- Sequence Space: Discrete
- Structure Space: Continuous



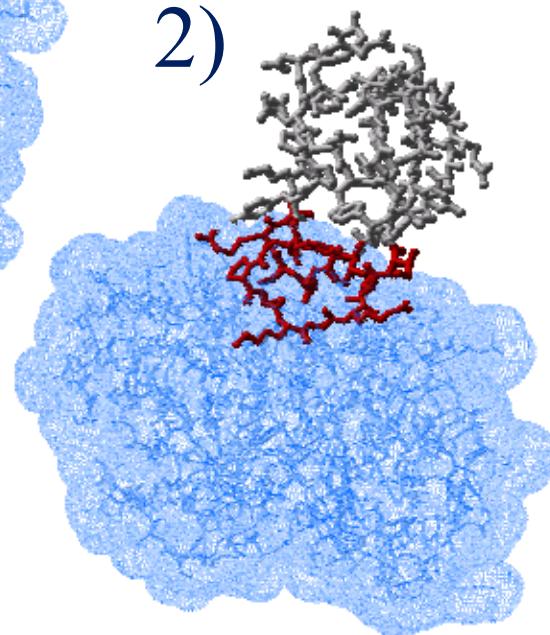
docking



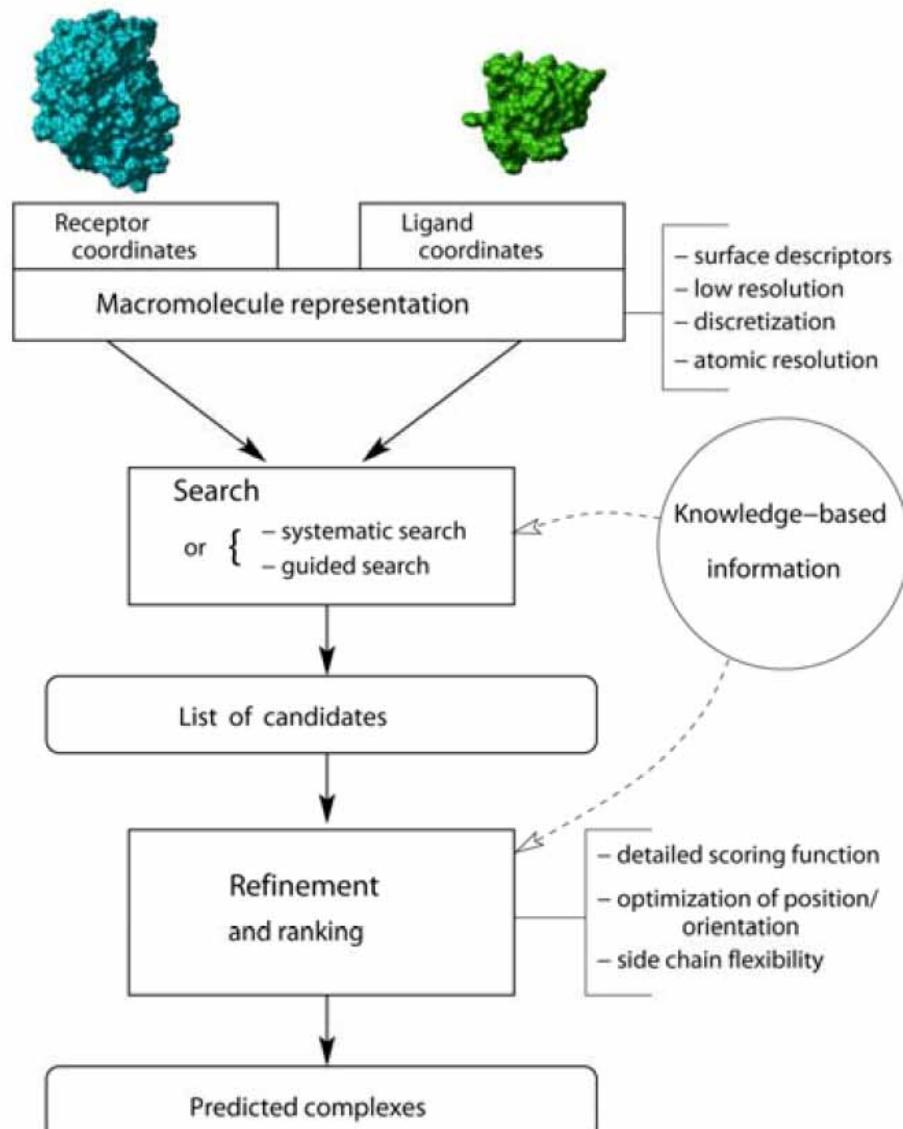
1)



2)

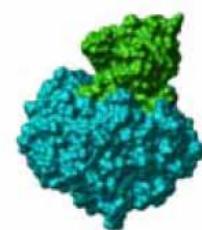


ab initio docking



SAMPLING & SCORING

SAMPLING & SCORING



Docking Search Strategies

- **Pseudo Random**

- Simulated Annealing / Monte Carlo
- Genetic Algorithms

- **Directed Search**

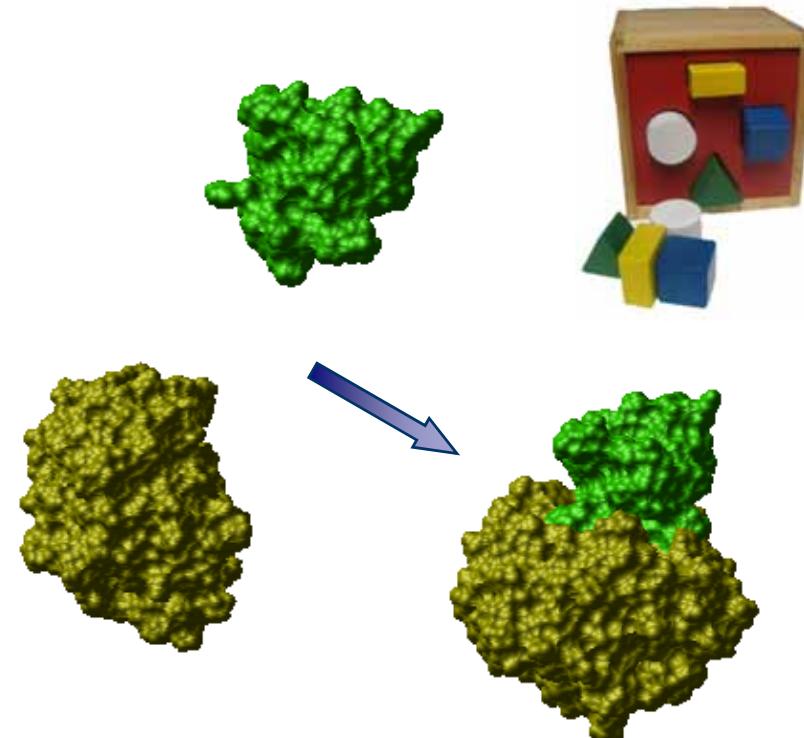
- Geometric Hashing
- Spherical Harmonic Surface Triangles

- **Brute-Force Search**

- Explicit Grid Correlations
- Fast Fourier Transform (FFT) Correlations
- Spherical Polar Fourier Correlations

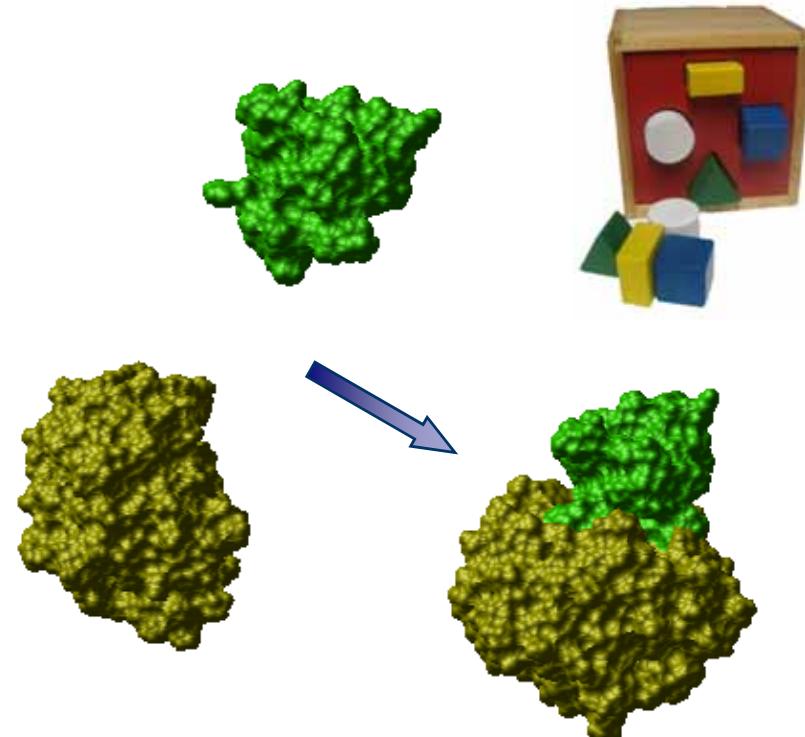
- **Refinement Phase**

- Classical or Soft Potentials (+/- Electrostatics)
- Desolvation, Solvent Dipoles...
- Visual Inspection!!



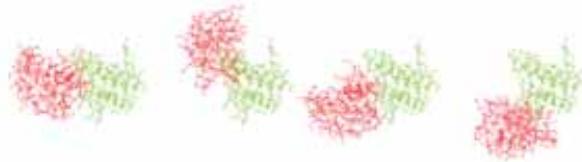
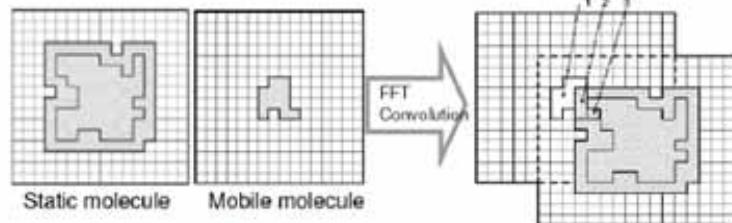
Criteria for Good Docking Orientations

- Low Free Energy (Difficult!)
- Low Pseudo-Energy (Easy) Based On...
- Large Surface Burial: $\sim 1600 \pm 400 \text{ \AA}^2$
- Small van der Waals Overlaps
- No Large Cavities in Interface
- Good H-Bonding: $\sim 1 \text{ HB}/100 \text{ \AA}^2$
- Good Charge Complementarity
- Polar/Polar Contacts Favoured
- Polar/Non-Polar Contacts Disfavoured



Protein-protein rigid-body docking methods

Exhaustive search
(FFT, surface-based)

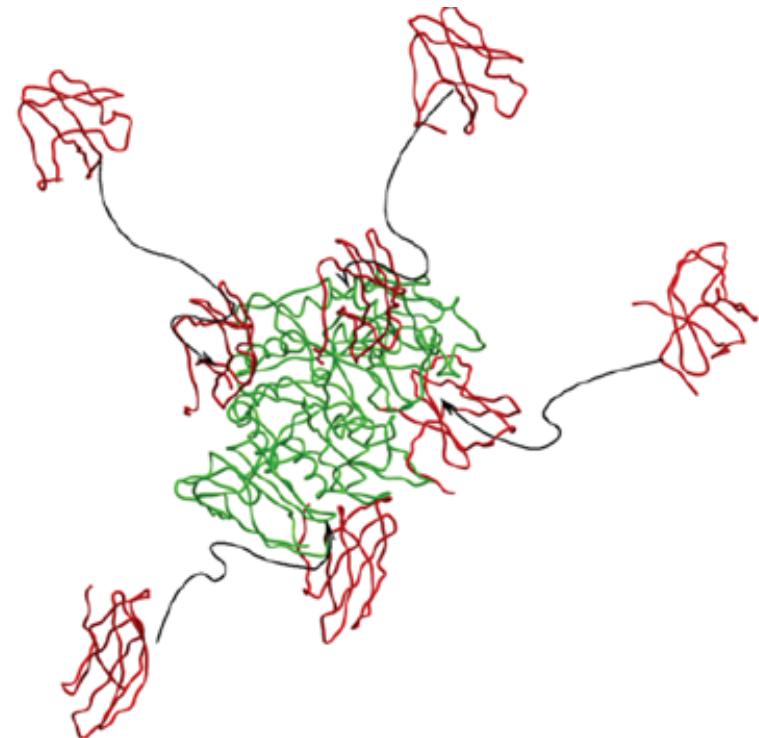


Scoring function

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

Geometry-based docking

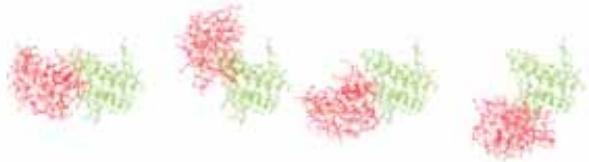
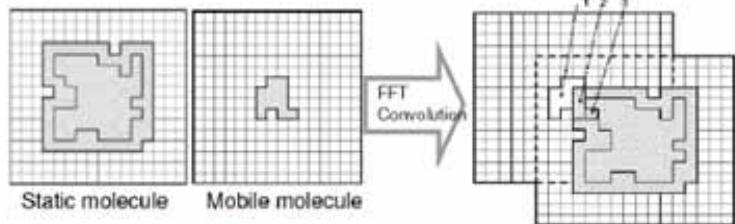
Stochastic sampling
(Monte-Carlo, minimization)



Energy-based docking

Protein-protein rigid-body docking methods

Exhaustive search (FFT, surface-based)



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

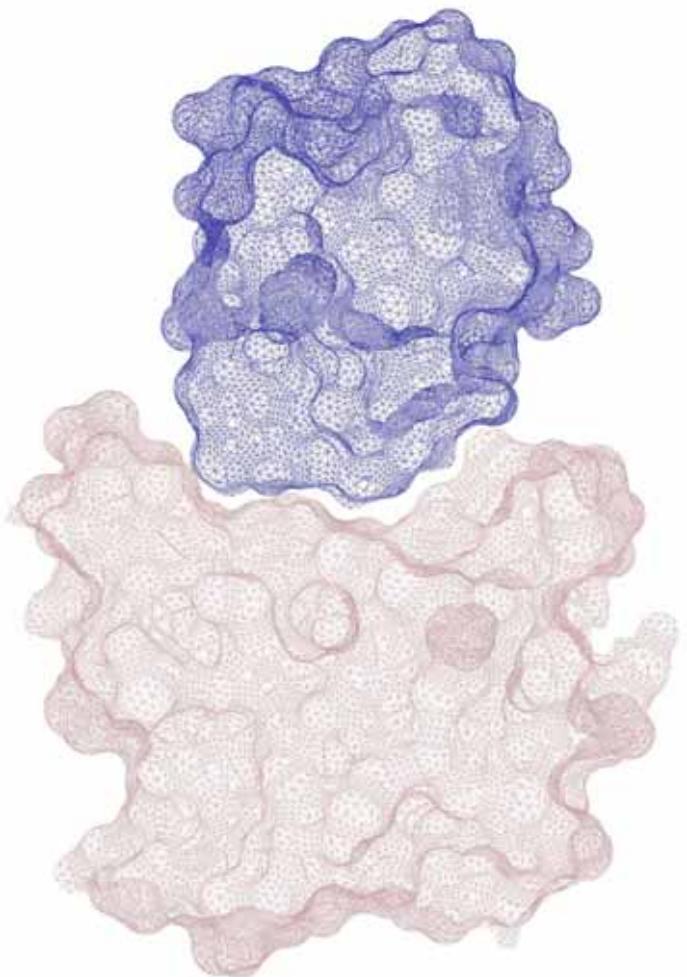
Geometry-based docking

Stochastic sampling (Monte-Carlo, minimization)

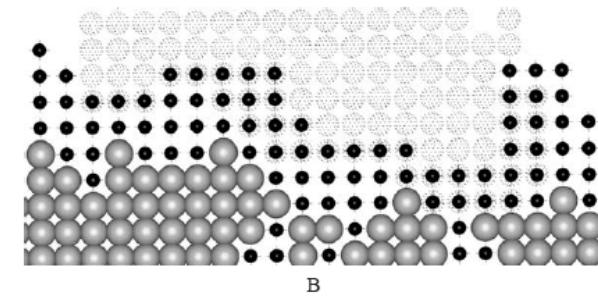
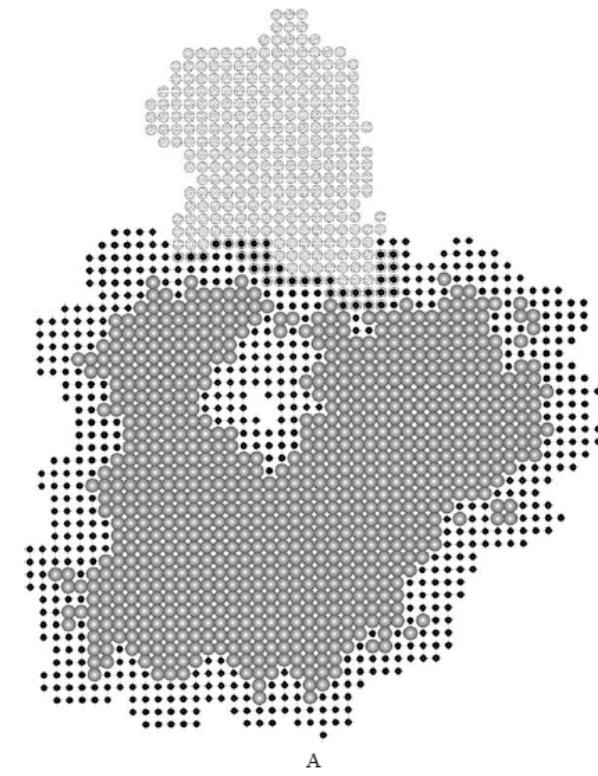
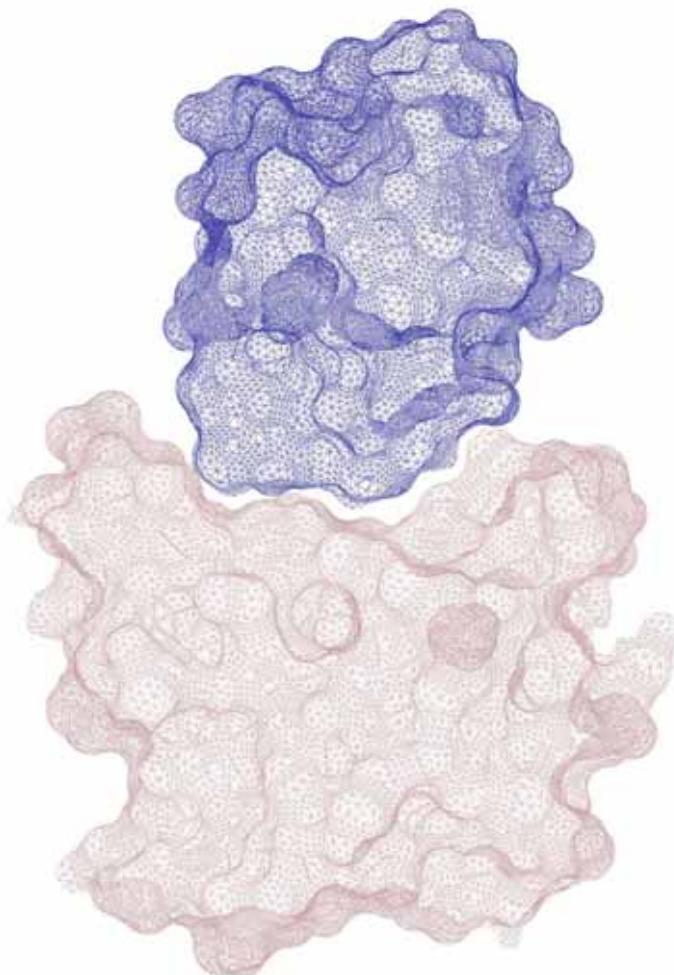


- Geometry-based docking
 - **FFT-based grid search**
 - **Geometric Hashing**
 - **Adding distance-based constraints**

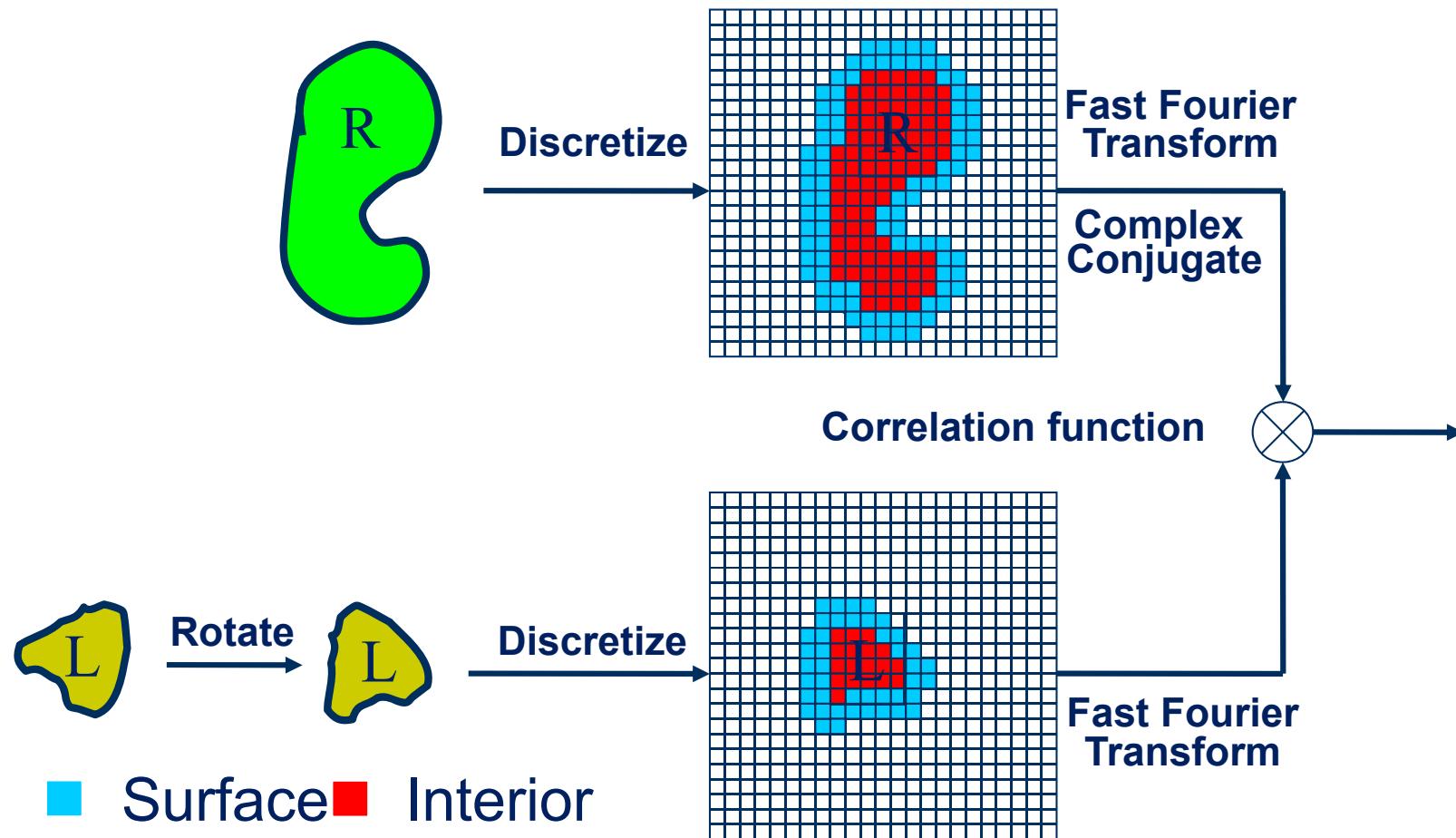
Rigid-Body Docking: Geometry Approach



Rigid-Body Docking: Geometry Approach

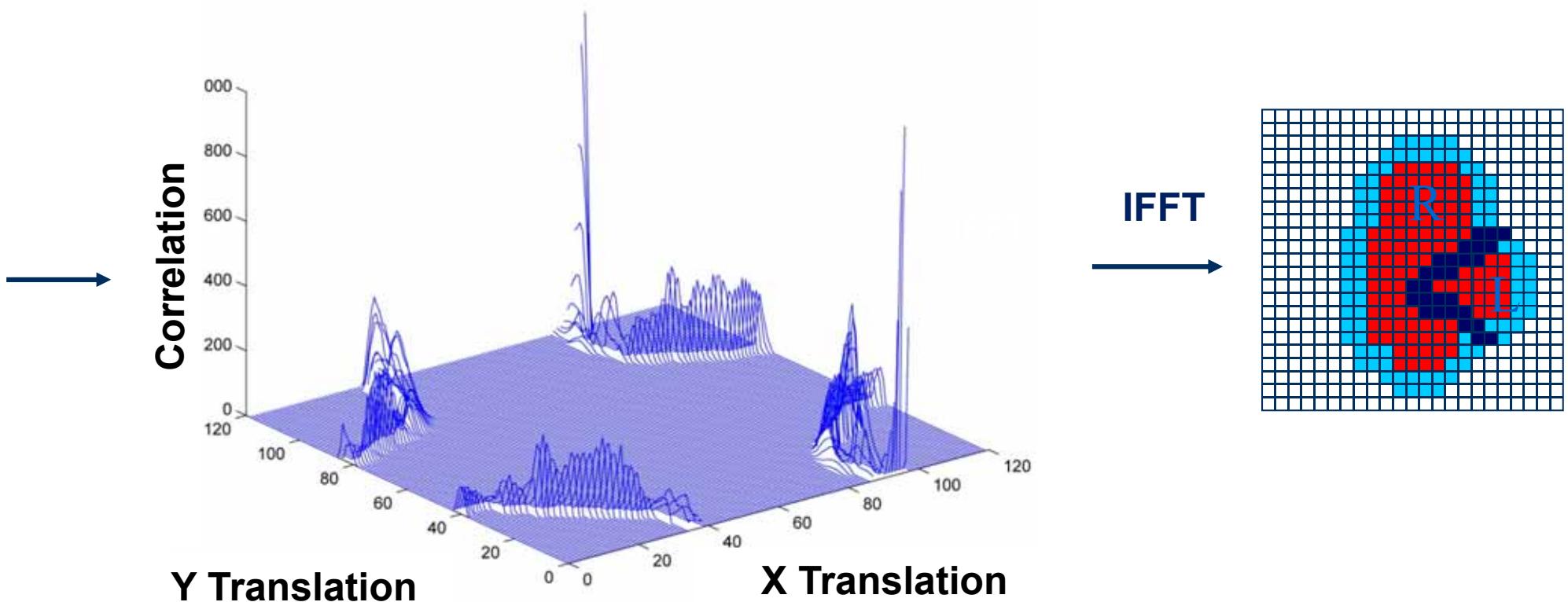


Protein Docking Using FFT



Protein Docking Using FFT

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



FTDOCK

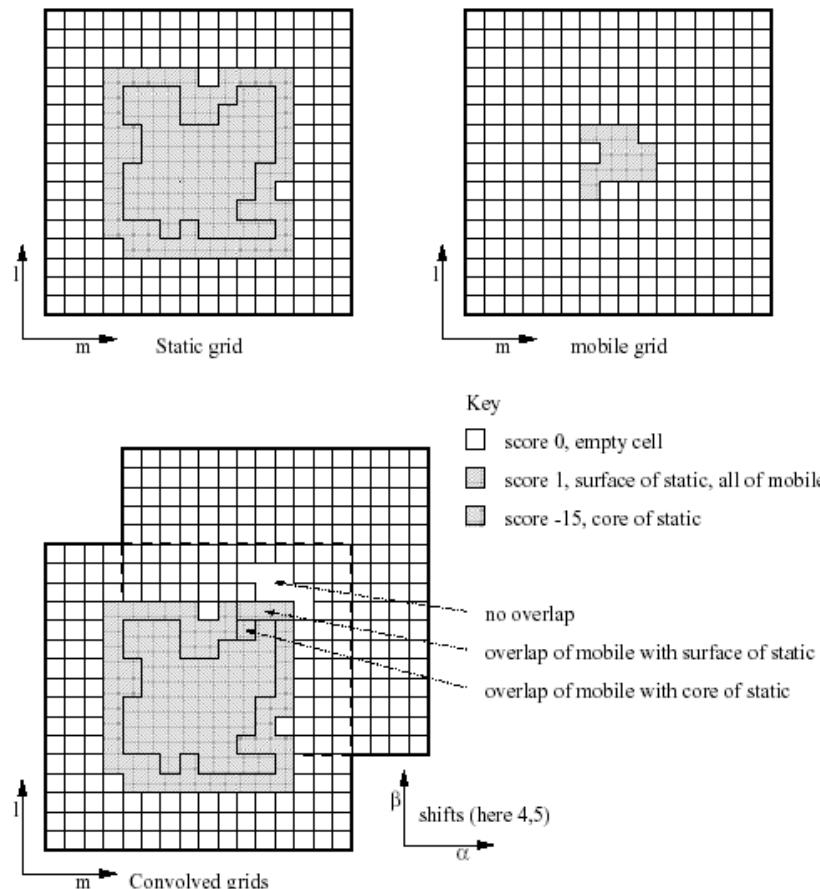
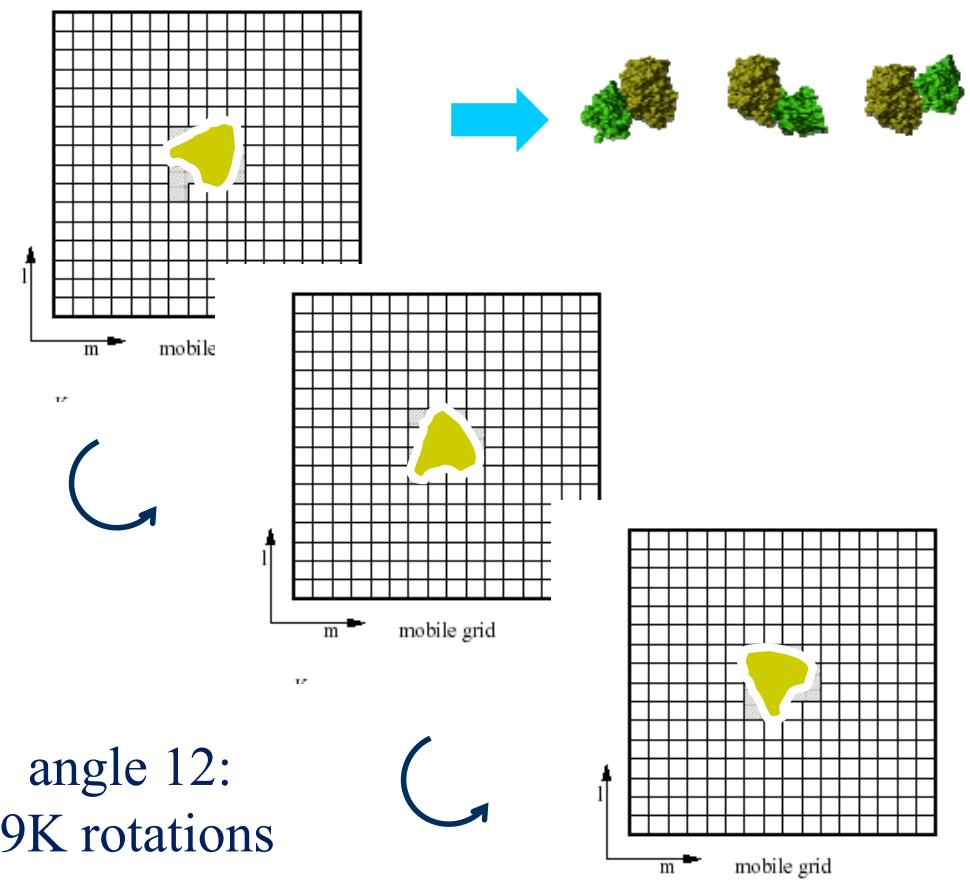
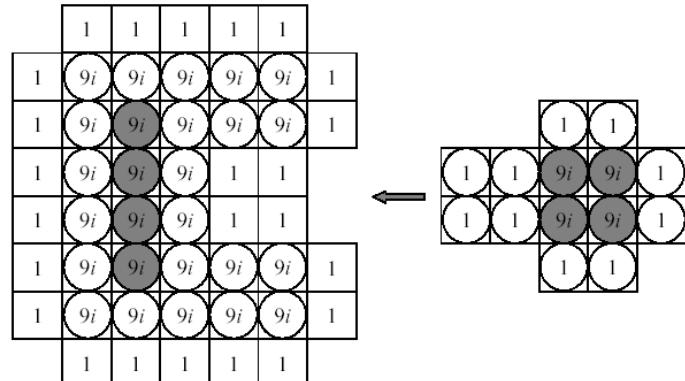


Figure 2: Grid discretisation of molecules and calculation of surface complementarity

```
ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log
```



ZDOCK

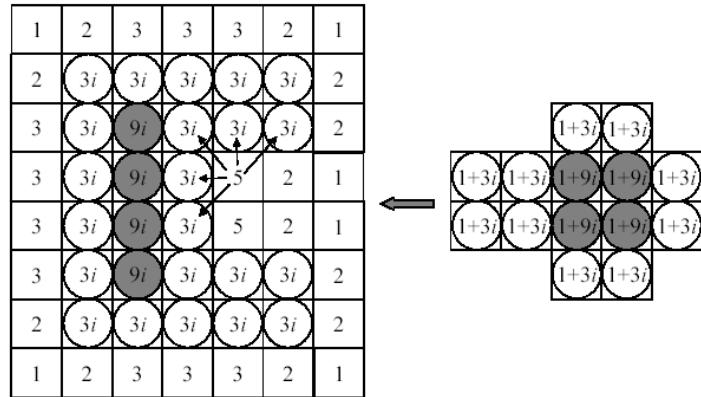


R_{GSC}



L_{GSC}

mark_sur 1A2P_A.pdb 1BRS_r.pdb
mark_sur 1A19_A.pdb 1BRS_l.pdb

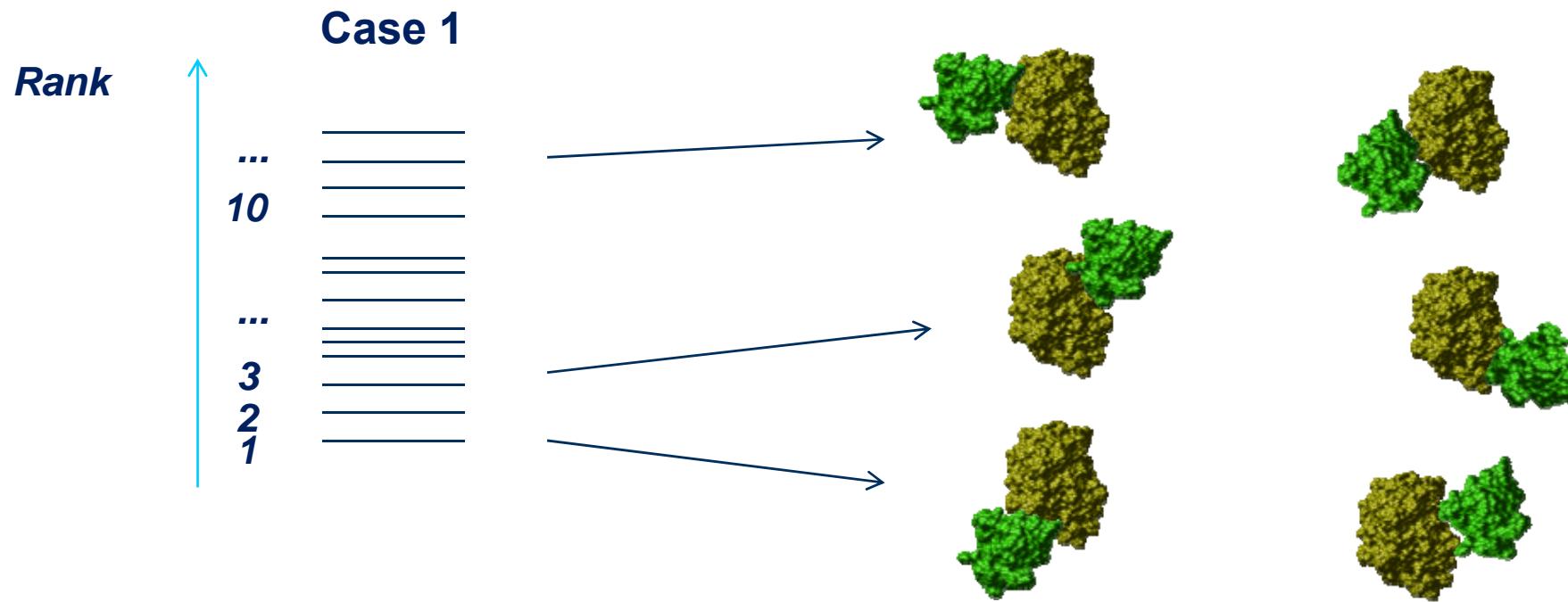


L_{PSC}

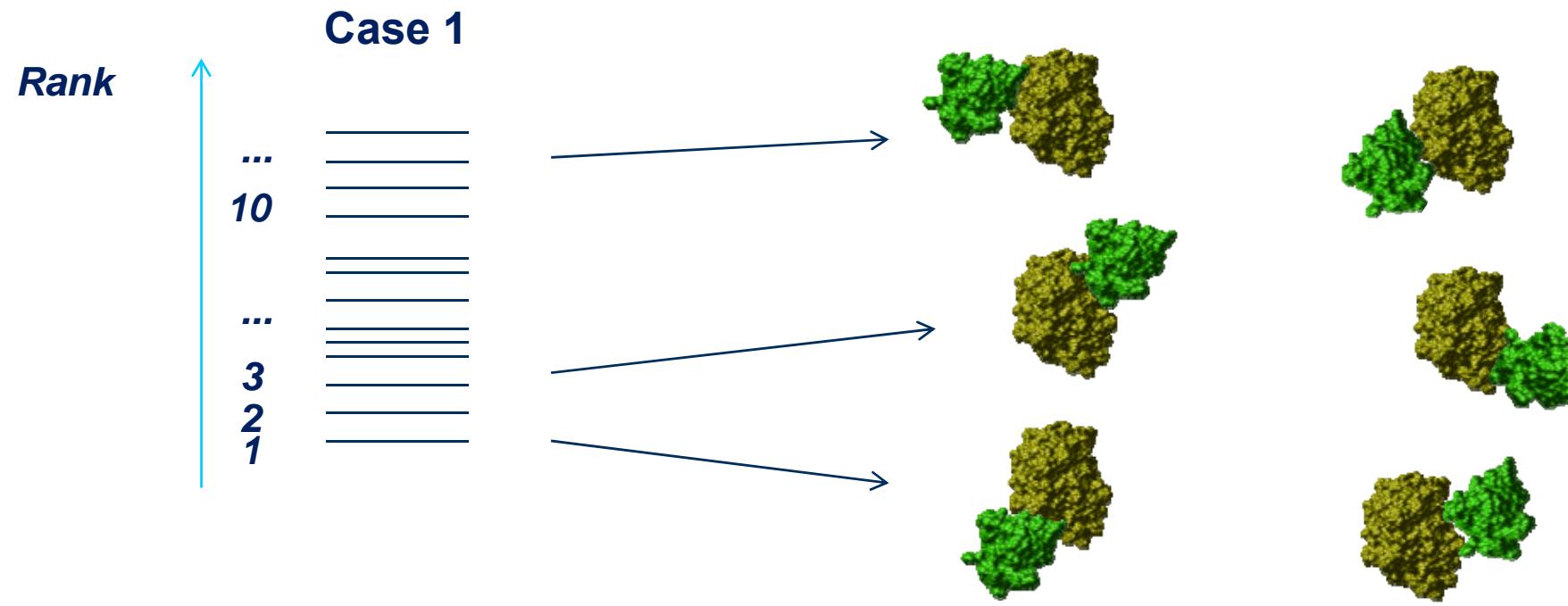


zdock -R 1BRS_r.pdb -L 1BRS_l.pdb -o 1BRS_zdock.out

Assessment of docking performance

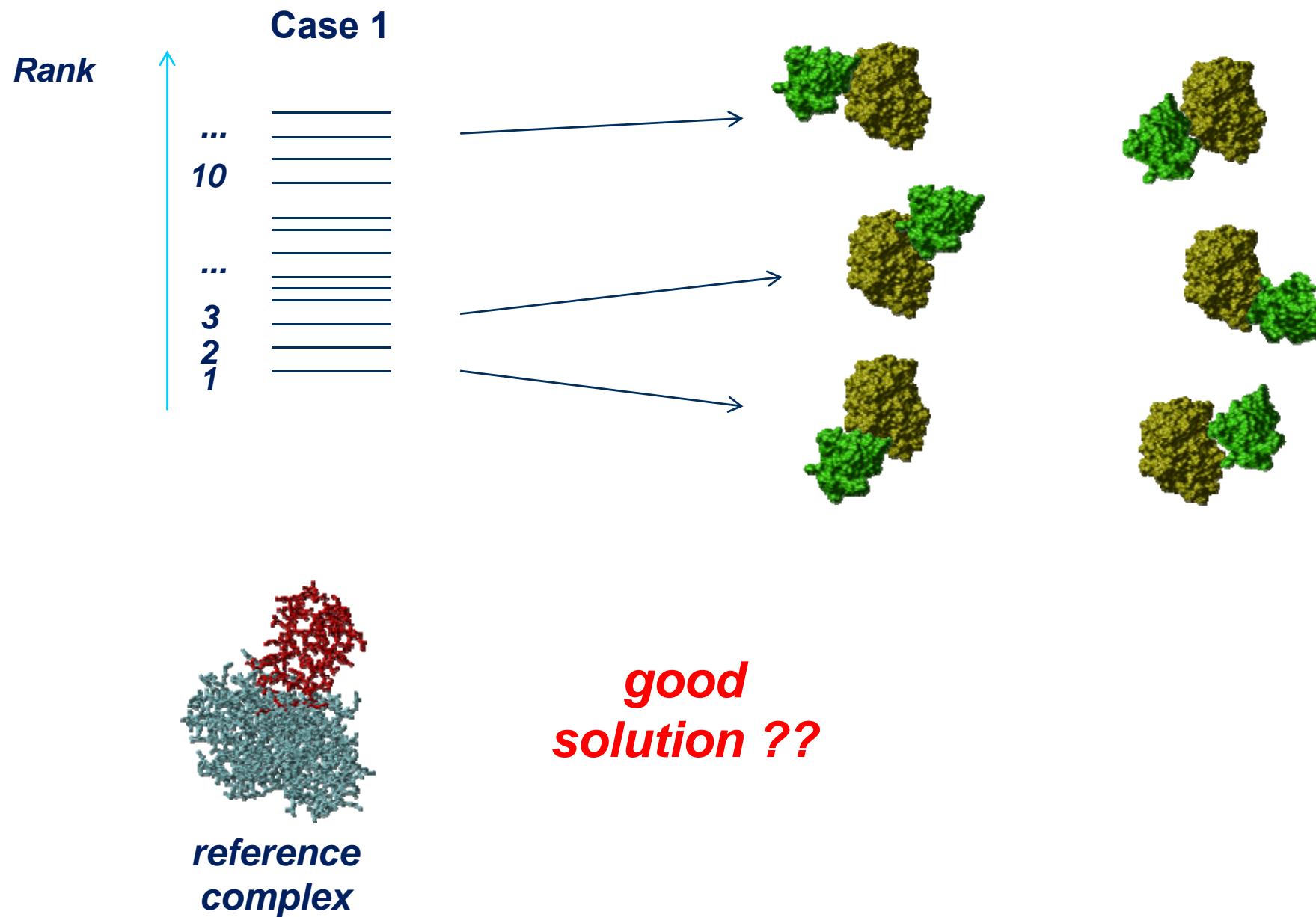


Assessment of docking performance



*good
solution ??*

Assessment of docking performance



Assessment of docking performance

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

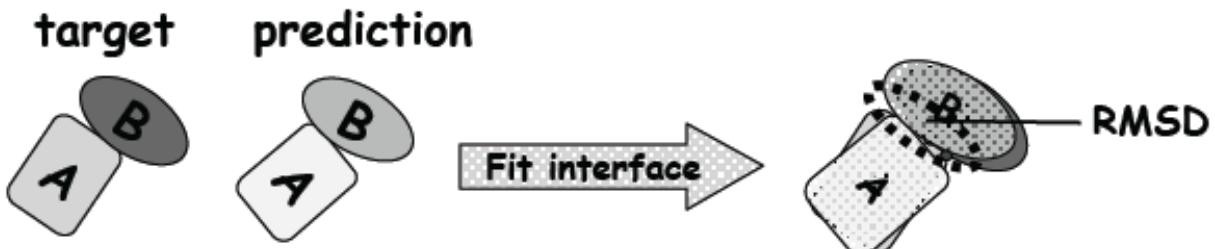
$$F_{\text{nat}} = \frac{\text{Correctly predicted contacts}}{\text{Total number of contacts in the target}}$$

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

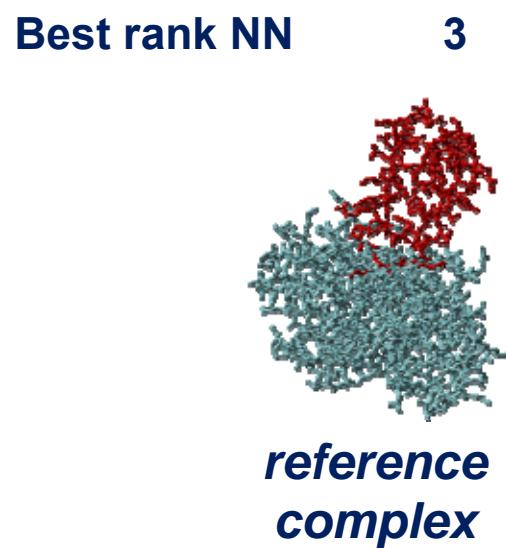
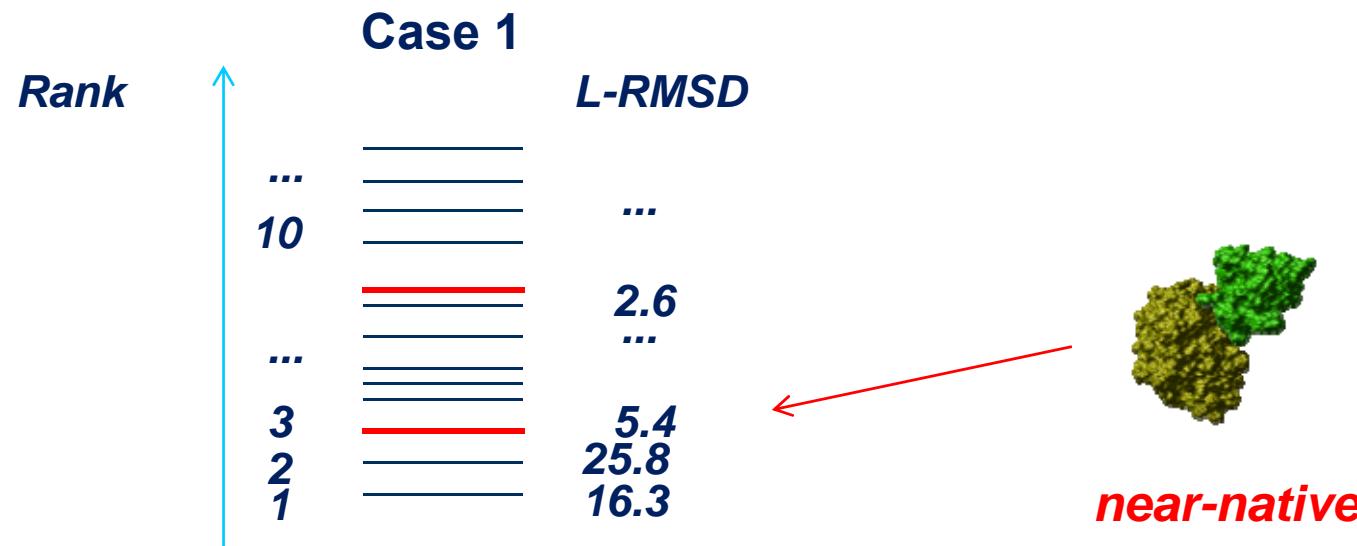
**near-native:
L-RMSD < 10 Å**



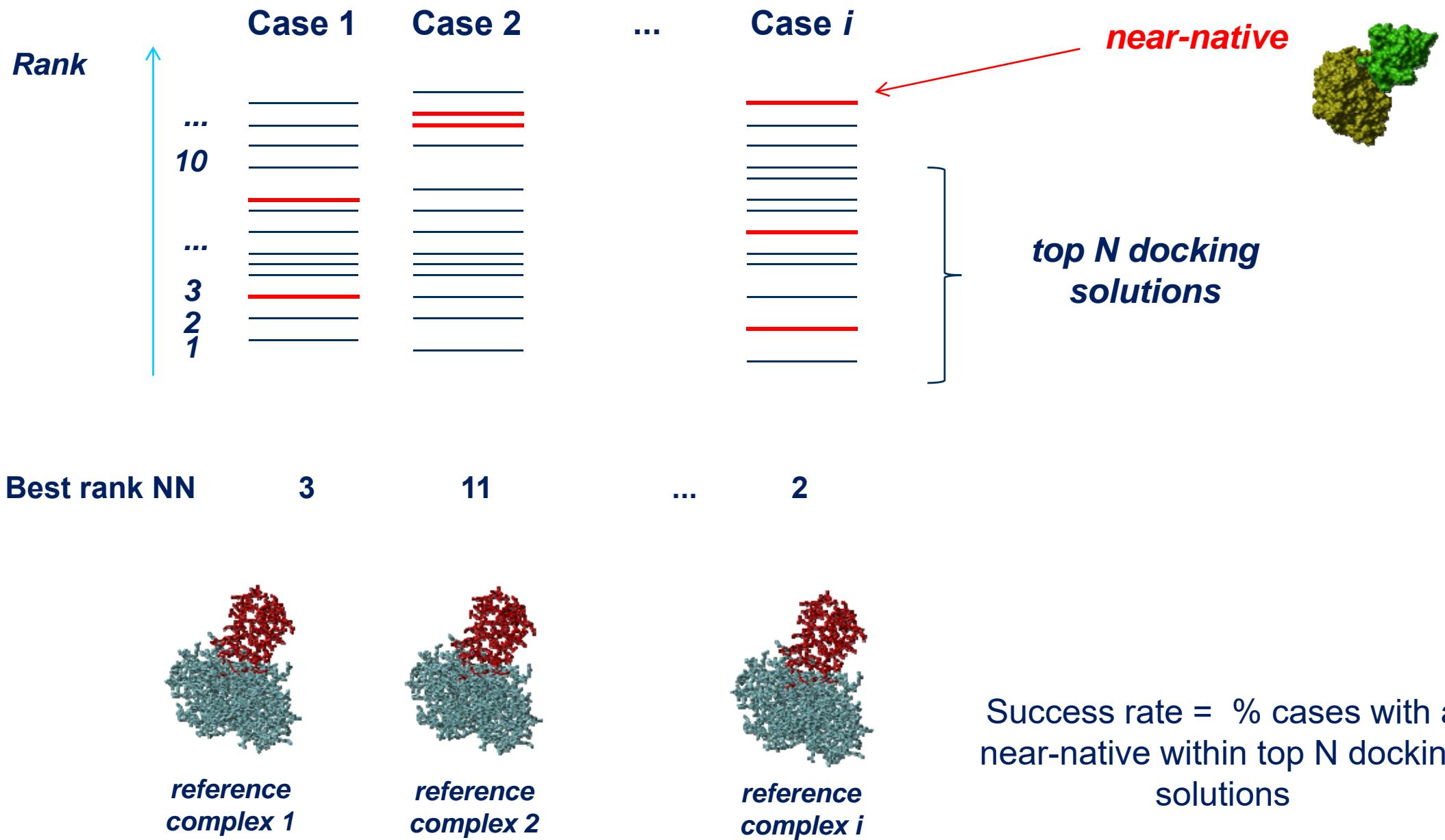
- i-RMSD: RMSD on interface residues (within 10 Å)



Assessment of docking performance



Assessment of docking performance



Protein-protein benchmarking

2002-2018 Weng's benchmark series

0.0: 54 cases

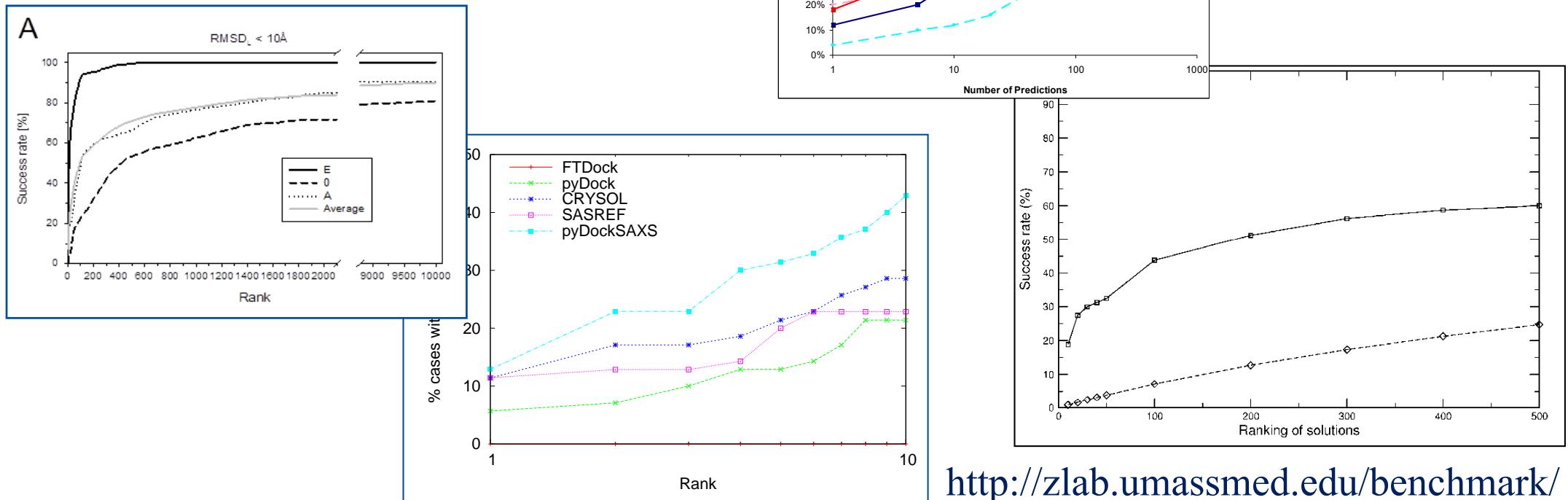
1.0: 59 cases

2.0: 84 cases

3.0: 124 cases

4.0: 176 cases

5.0: 230 cases





Supplementary Information for

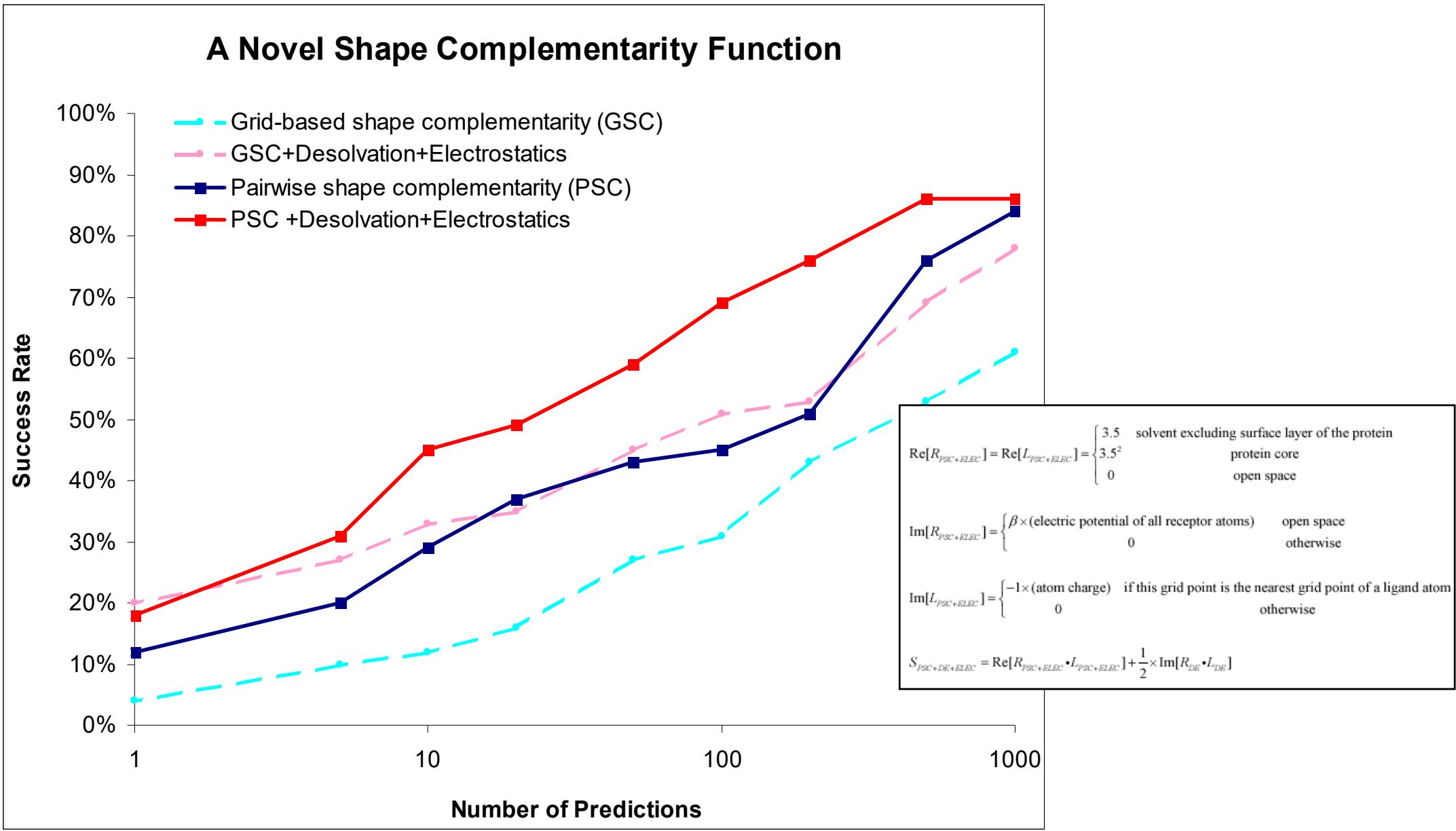
Vreven, T, Moal, IH, Vangone, A, Pierce, BG, Kastritis, PL, Torchala, M, Chaleil, R, Jimenez-Garcia, B, Bates, PA, Fernandez-Recio, J, Bonvin, AMJJ and
Updates to the integrated protein-protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2. (Journal of Molecular Biology,
[Download cleaned-up PDB files for the benchmark in a gzipped archive](#) [Download Table in excel file](#)

Follow the "info" link in the leftmost column to view bound/unbound chain BLAST alignments and potentially useful comments about individual benchmark cases

Table: Protein-Protein Docking

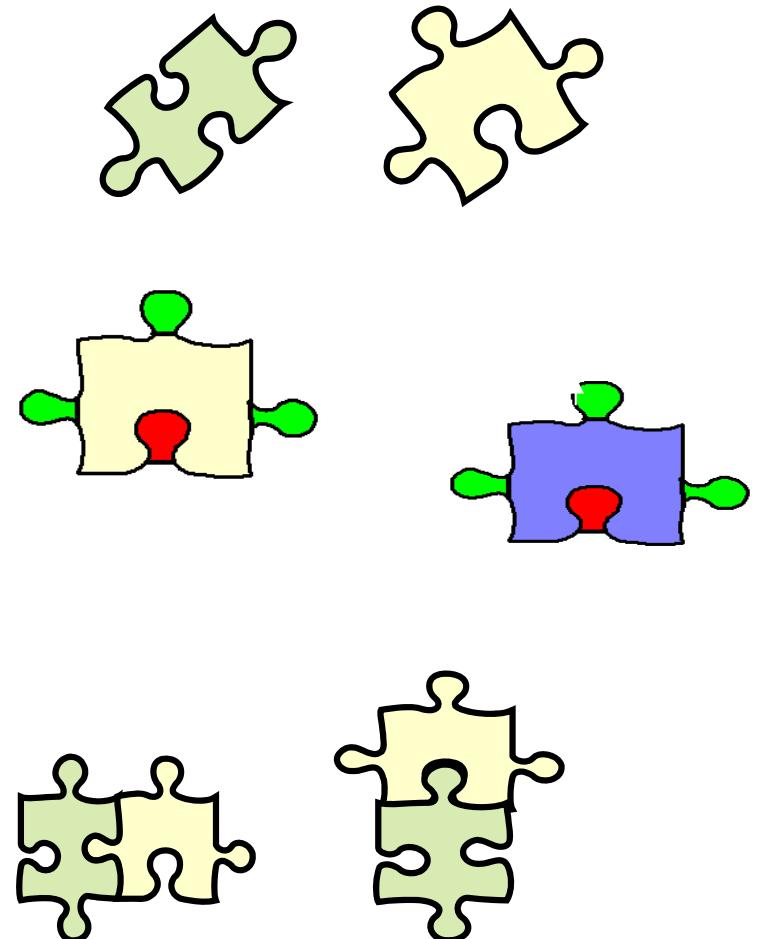
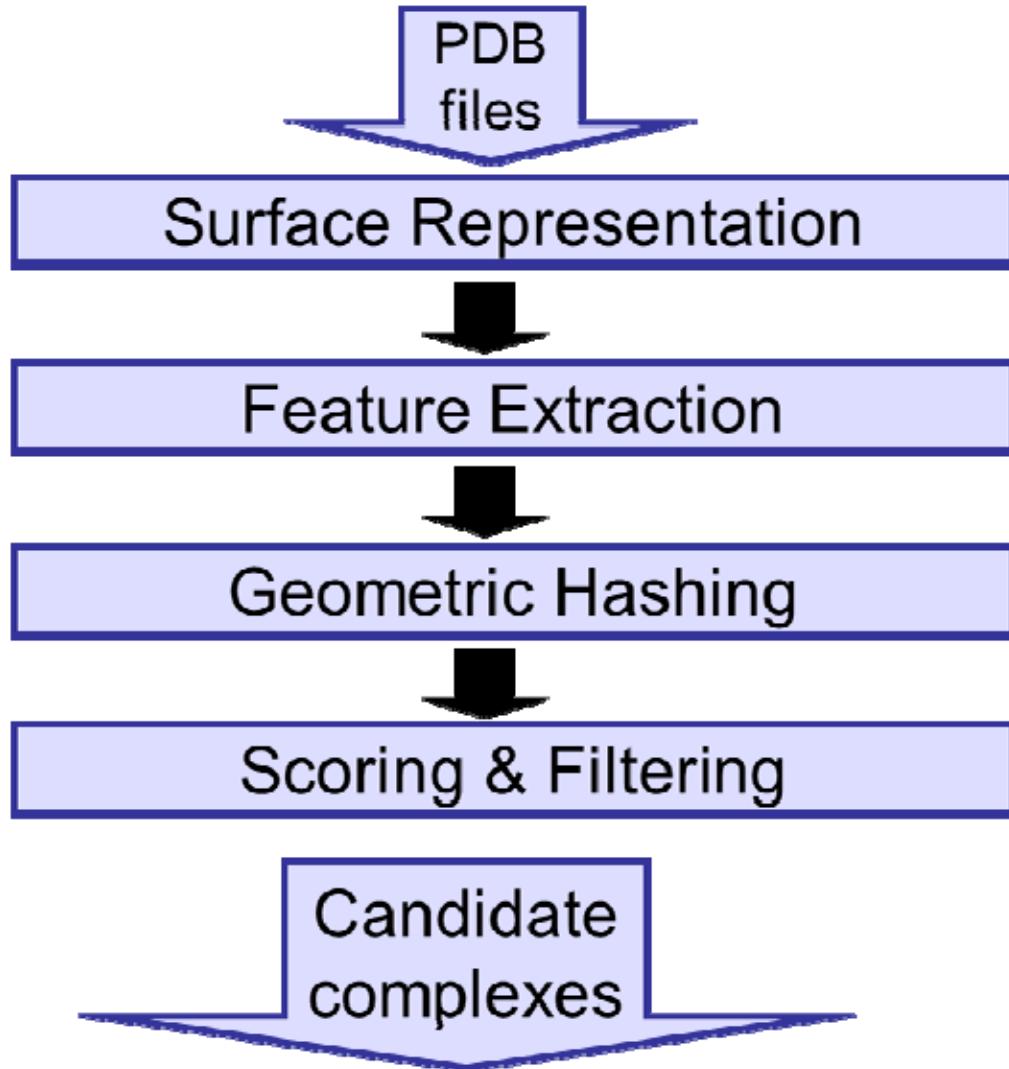
+Info	Complex	Cat. ^a	PDBid 1	Protein 1	HETATMs	PDBid 2	Protein 2	HETATMs	RMSD ^b (Å)	DASA ^c (Å ²)	Multimer
	Rigid-body (151)										
info	1AHW_AB:C	A	1FGN_LH	Fab 5g9		1TFH_A	Tissue factor		0.69	1899	
info	1BVK_DE:F	A	1BVL_BA	Fv Hulys11		3LZT	HEW lysozyme		1.24	1321	
info	1DQJ_AB:C	A	1DQQ_CD	Fab HyHEL63		3LZT	HEW lysozyme		0.75	1765	
info	1E6J_HL:P	A	1E6O_HL	Fab		1A43	HIV-1 capsid protein p24		1.05	1245	
info	1JPS_HL:T	A	1JPT_HL	Fab D3H44		1TFH_B	Tissue factor		0.51	1852	
info	1MLC_AB:E	A	1MLB_AB	Fab44.1		3LZT	HEW lysozyme		0.6	1392	
info	1VFB_AB:C	A	1VFA_AB	Fv D1.3		8LYZ	HEW lysozyme		1.02	1383	
info	1WEJ_HL:F	A	1QBL_HL	Fab E8		1HRC	Cytochrome C	HEM,ACE	0.31	1177	
info	2FD6_HL:U	A	2FAT_HL	Plasminogen receptor antibody		1YWH_A	Plasminogen activator receptor		1.07	1139	
info	2I25_N:L	A	2I24_N	Shark single domain antigen receptor		3LZT	Lysozyme		1.21	1425	
info	2VIS_AB:C	A	1GIG_LH	Fab		2VIU_ACE	Flu virus hemagglutinin		0.8	1296	Ligand
info	2VXT_HL:I	A	2VXU_HL	Murine reference antibody 125-2H		1J0S_A(6)	Interleukin-18		1.33	2163	
info	2W9E_HL:A	A	2W9D_HL	ICSM 18 FAB fragment		1QM1_A	Prion protein fragment		1.13	1677	
info	3EOA_LH:I	A	3EO9_LH	Efalizumab FAB fragment		3F74_A	Integrin alpha-L I domain		0.39	1272	
info	3HMX_LH:AB	A	3HMM_LH	Ustekinumab FAB		1F45_AB	Interleukin-12		0.73	1841	
info	3MXW_LH:A	A	3MXV_LH	Anti-Shh 5E1 chimera FAB fragment		3M1N_A	Sonic Hedgehog N-terminal domain		0.48	1696	
info	3RVW_CD:A	A	3RVT_CD	4C1 FAB		3FSV_A	DER P 1 allergen		0.5	1383	
info	4DN4_LH:M	A	4DN3_LH	CNT0888 FAB		1DOL_A	MCP-1		0.81	1317	
info	4FOI_HL:ABEFCD:A	A	4FOH_HL	CR9114 FAB		2FK0_ABCD	H5N1 influenza virus hemagglutinin		1.08	1459	Complex
info	4G6J_HL:A	A	4G5Z_HL	Canakinumab antibody fragment		4I1B_A	Interleukin-1 beta		0.61	1893	
info	4G6M_HL:A	A	4G6K_HL	Gevokizumab antibody fragment		4I1B_A	Interleukin-1 beta		0.49	1673	
info	4CYU_MN-BREFCD:N	A	4CYU_WT	1F1 antibody		1DT17_WT_TK1	191R H1 Hemagglutinin		0.78	1830	

ZDOCK performance



- Geometry-based docking
 - FFT-based grid search
 - **Geometric Hashing**
 - Adding distance-based constraints

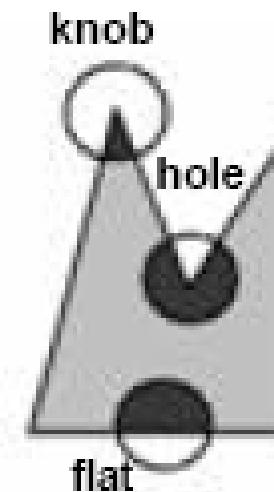
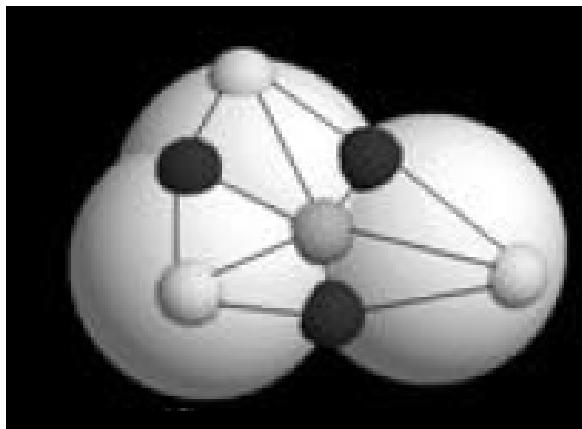
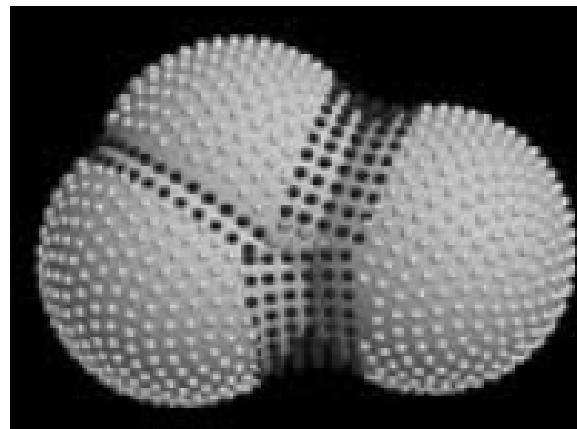
Geometrical hashing - PatchDock



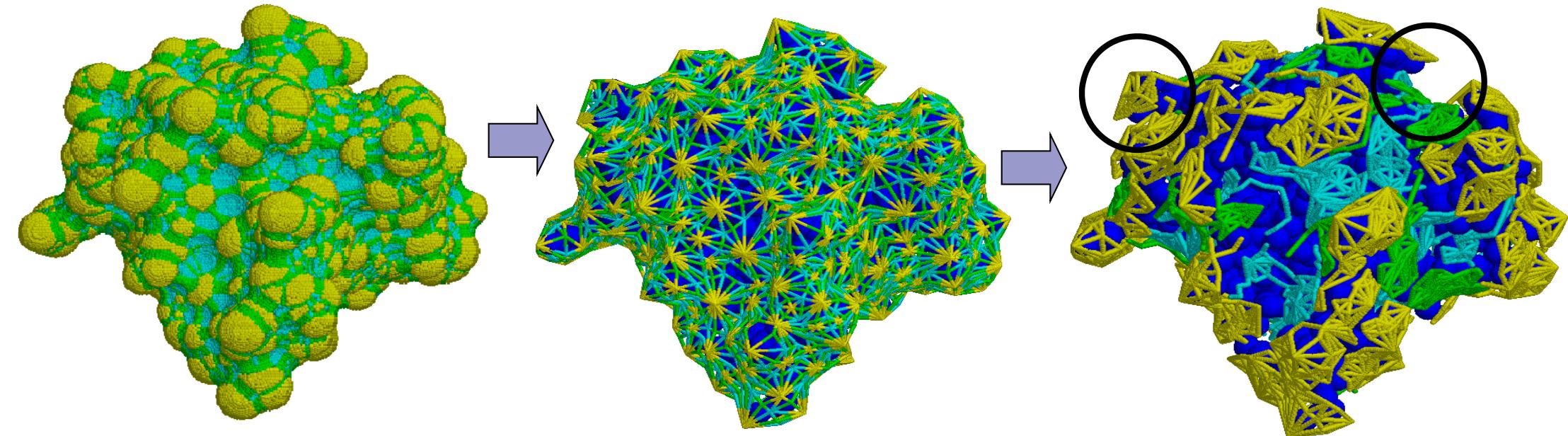
Surface Representation

Surface Representation

- Dense MS surface (Connolly)
- Sparse surface (Shuo Lin et al.)



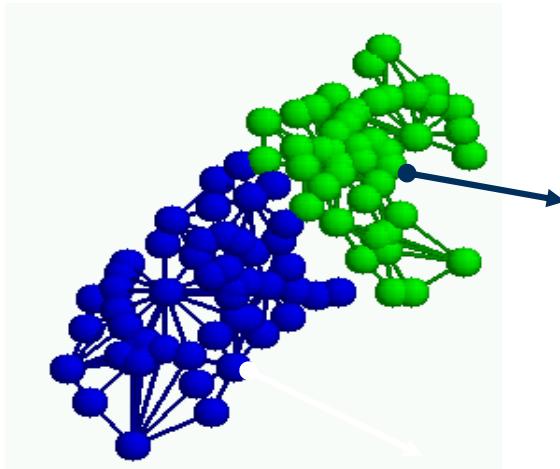
Feature Extraction



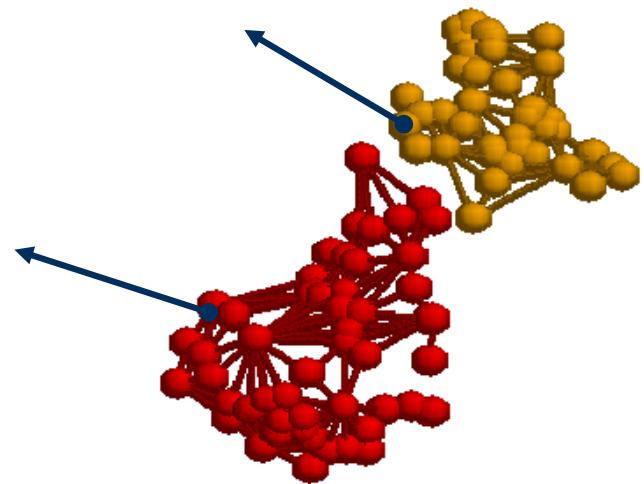
- We focus on sparse surface features, preserving the quality of shape representation.
- The sparse features reduce the complexity of the matching step.

Patch-Pair Matching

Receptor patches



Ligand patches



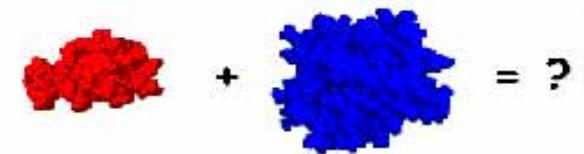
Transformation



Base: 1 critical point with its normal from one patch and
1 critical point with its normal from a neighboring patch.

- Match every base from the receptor patches with all the bases from complementary ligand patches.
- Compute the transformation for each pair of matched bases.

PATCHDOCK



Molecular Docking Algorithm Based on Shape Complementarity Principles

[About PatchDock] [Web Server] [Download] [Help] [FAQ] [References]

3 mandatory fields

Type PDB codes of receptor and ligand molecules or upload files in PDB format

Receptor Molecule:

(PDB:chainId e.g.
2kai:AB) or upload file:

 Browse...

Ligand Molecule:

(PDB:chainId e.g.
2kai:I) or upload file:

 Browse...

E-mail address:

(the results are sent to this address)

Clustering RMSD:

 4.0

Complex Type:

Be sure to give receptor and ligand in the corresponding order!

Clustering RMSD determines the number of output solutions

Complex Type adjusts parameter set for the specific complex type, such as enzyme-inhibitor, antibody-antigen, protein-small ligand

OutputPatchDock - server

Input data

Top 20 solutions

Receptor	Ligand	Complex	Type	Clustering	RMSD	User e-mail	Receptor Site	Ligand Site
A.pdb.tr	B.pdb.tr	Default		4.0		duhovka@tau.ac.il	-	-

Solution No	Score	Area	ACE	Transformation								PDB file of the complex
1	10166	1528.90	-12.07	-1.15	-0.37	-0.32	-11.03	-15.21	22.35			result.1.pdb
2	9740	1285.70	129.52	-2.07	-0.40	-1.44	-3.43	9.43	18.35			result.2.pdb
3	9380	1284.70	-8.60	0.77	0.16	-0.54	-1.20	8.06	-10.85			result.3.pdb
4	9122	1298.80	203.84	2.45	-0.57	-1.59	13.45	2.35	14.25			result.4.pdb
5	9048	1185.20	-243.82	3.11	0.11	1.04	-16.19	-29.02	2.25			result.5.pdb
6	9034	1094.70	301.73	2.09	-0.49	-1.37	6.24	3.03	10.96			result.6.pdb
7	8972	1182.40	-39.65	-2.07	-1.05	-2.85	6.41	-17.96	28.59			result.7.pdb
8	9800	1162.80	165.52	0.50	0.02	2.17	12.62	22.60	10.70			result.8.pdb
9	8886	1195.40	-31.63	0.92	0.49	3.07	18.21	-10.46	-11.65			result.9.pdb
10	8850	1240.20	-149.90	-1.43	0.72	0.57	-14.80	-28.15	-1.02			result.10.pdb
11	8768	962.70	19.42	1.12	0.24	0.71	12.24	0.72	21.55			result.11.pdb
12	8740	1110.70	2.06	0.34	0.78	2.56	2.98	3.29	-6.48			result.12.pdb
13	8700	973.80	-79.03	-1.32	-0.69	-0.86	-7.61	-8.97	26.23			result.13.pdb
14	8696	1012.60	-161.10	0.23	0.68	2.24	14.51	-28.88	-3.17			result.14.pdb
15	8644	1206.50	-46.74	-1.89	-0.36	1.91	5.66	-34.25	17.10			result.15.pdb
16	8628	1192.00	330.84	-0.75	0.74	0.78	-0.58	0.95	23.63			result.16.pdb
17	8602	1285.20	17.88	-2.71	0.40	-1.13	0.42	4.95	-2.29			result.17.pdb
18	8574	1342.30	118.56	2.27	-0.07	0.36	-10.58	13.53	29.87			result.18.pdb
19	8526	1039.50	-111.44	-0.65	0.10	-1.60	-11.41	7.24	13.05			result.19.pdb
20	8524	1098.60	-27.75	0.71	-0.70	2.28	24.54	7.03	47.47			result.20.pdb

[NEW: Jmol view](#)

[show next 20 >>](#)

Download output

DOWNLOAD best solutions as a ZIP file:

(solutions number, from 2 to 100)

GO

(this takes few seconds, please wait patiently)

Send for refinement

DOWNLOAD [solutions table](#) [transformations file](#)

REFINE best solutions with [FireDock](#): (solutions number, from 1 to 1000)

- Geometry-based docking
 - FFT-based grid search
 - Geometric Hashing
 - Adding distance-based constraints

Optional parameters: binding site

Advanced Options:

[Show] [Hide]

Receptor Binding Site: [Browse...](#) upload receptor binding site file

Ligand Binding Site: [Browse...](#) upload ligand binding site file

Distance Constraints: [Browse...](#) upload distance constraints file

[Submit Form](#)

[Clear](#)

Optional parameters: binding site

Advanced Options:

[Show] [Hide]

Receptor Binding Site: Browse... upload receptor binding site file

Ligand Binding Site: Browse... upload ligand binding site file

Distance Constraints: Browse... upload distance constraints file

Submit Form

Clear

Binding Site — list the residues of the receptor potential binding site. Only patches that include these residues will be used for matching.

Format: *[residue index] [chain ID]*

Example:

74 A

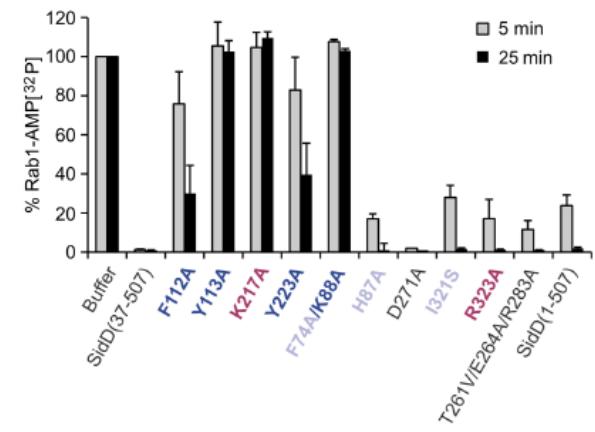
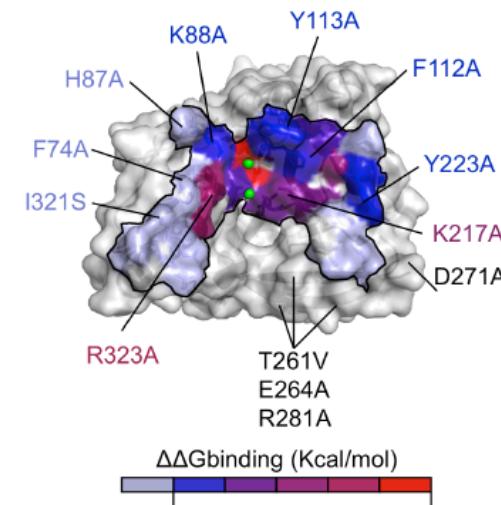
88 A

112 A

113 A

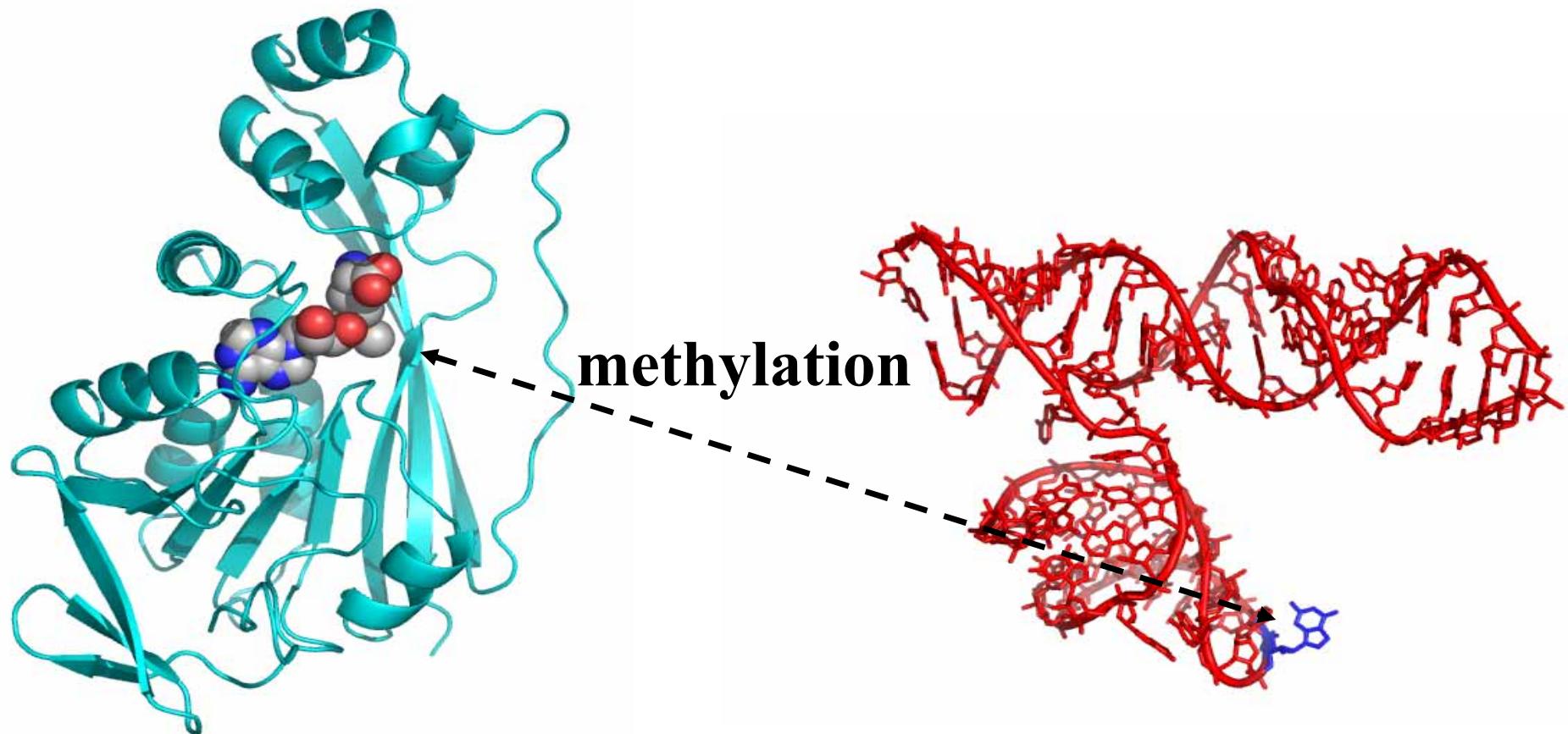
217 A

223 A



Docking with Distance Constraints

A complex of Rlma2 methyltransferase of *S. pneumoniae* and a 74 nucleotide RNA transcript



Optional parameters

Advanced Options:

[Show] [Hide]

Receptor Binding Site:	<input type="text"/>	Browse...	upload receptor binding site file
Ligand Binding Site:	<input type="text"/>	Browse...	upload ligand binding site file
Distance Constraints:	<input type="text"/>	Browse...	upload distance constraints file

Submit Form **Clear**

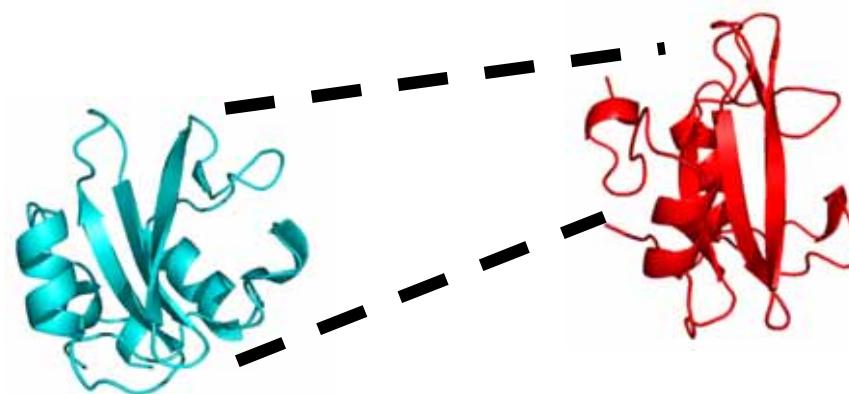
Distance Constraints –between pairs of atoms, one in the receptor and one in the ligand.

Format: *[receptor_atom_index] [ligand_atom_index] [max_dist]*

For example:

25 377 5.0

340 5603 10.0

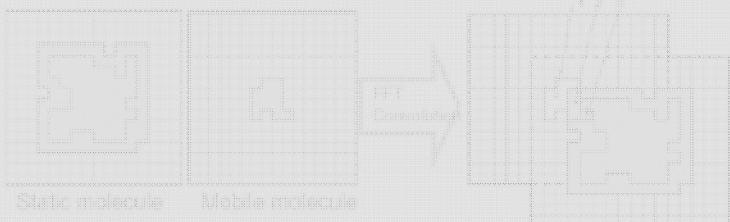


Geometry-based docking: conclusions

- *Many available methods*
- *Very fast!*
- *In most of the cases, reasonable success rates*
- *Unbound much worse than bound*
- *Some times, missing native (even with bound subunits)*
- *Good as first docking step before refinement*

Protein-protein docking methods

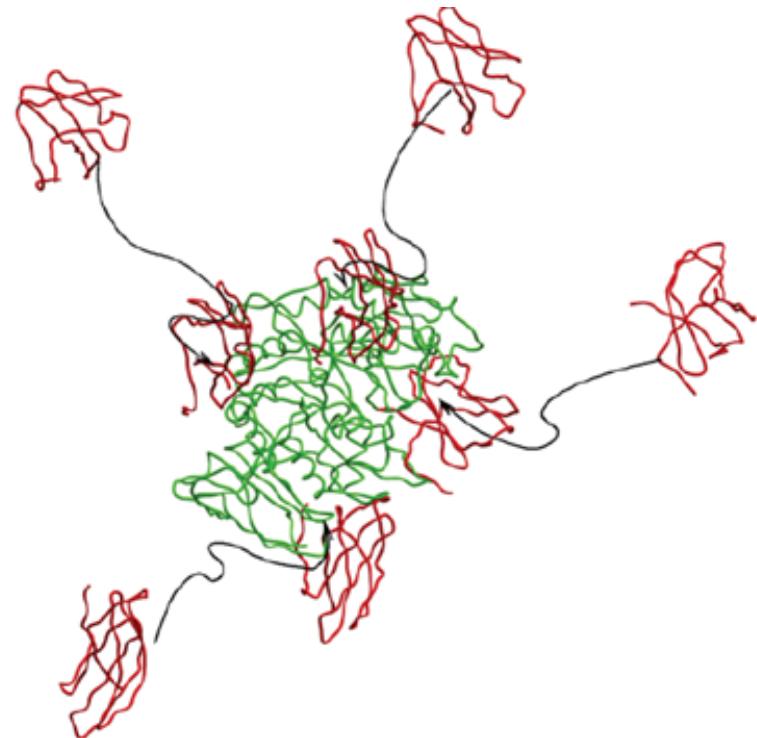
Exhaustive search
(FFT, surface-based)



Scoring function

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

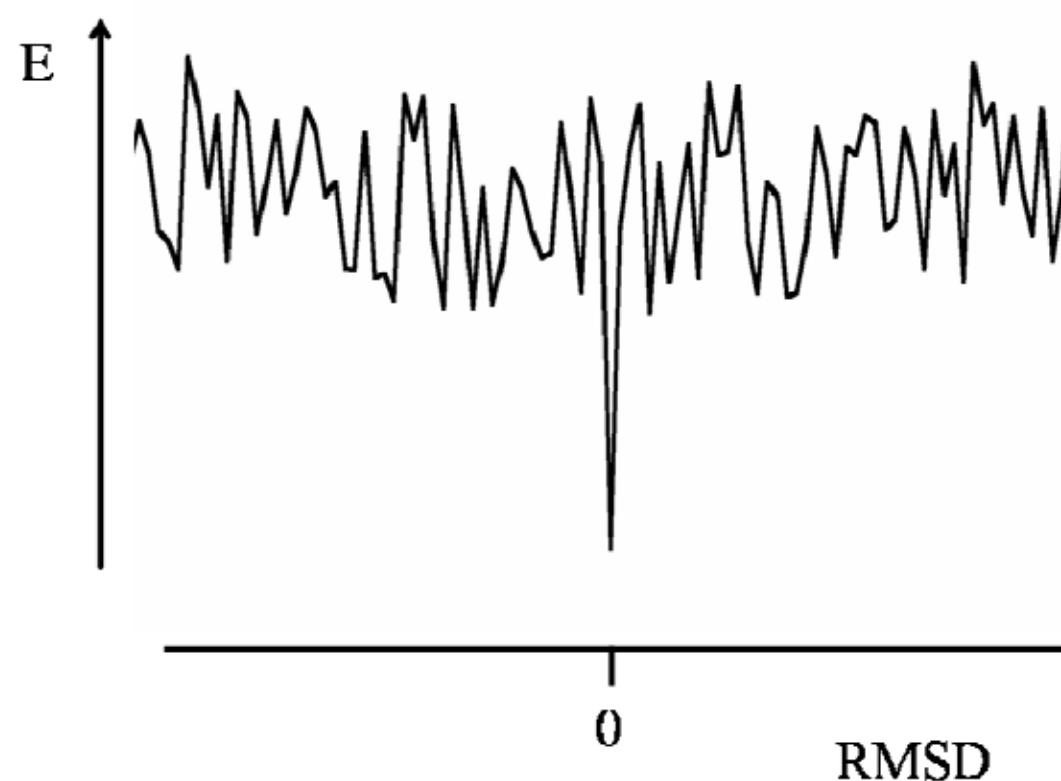
Stochastic sampling
(Monte-Carlo, minimization)



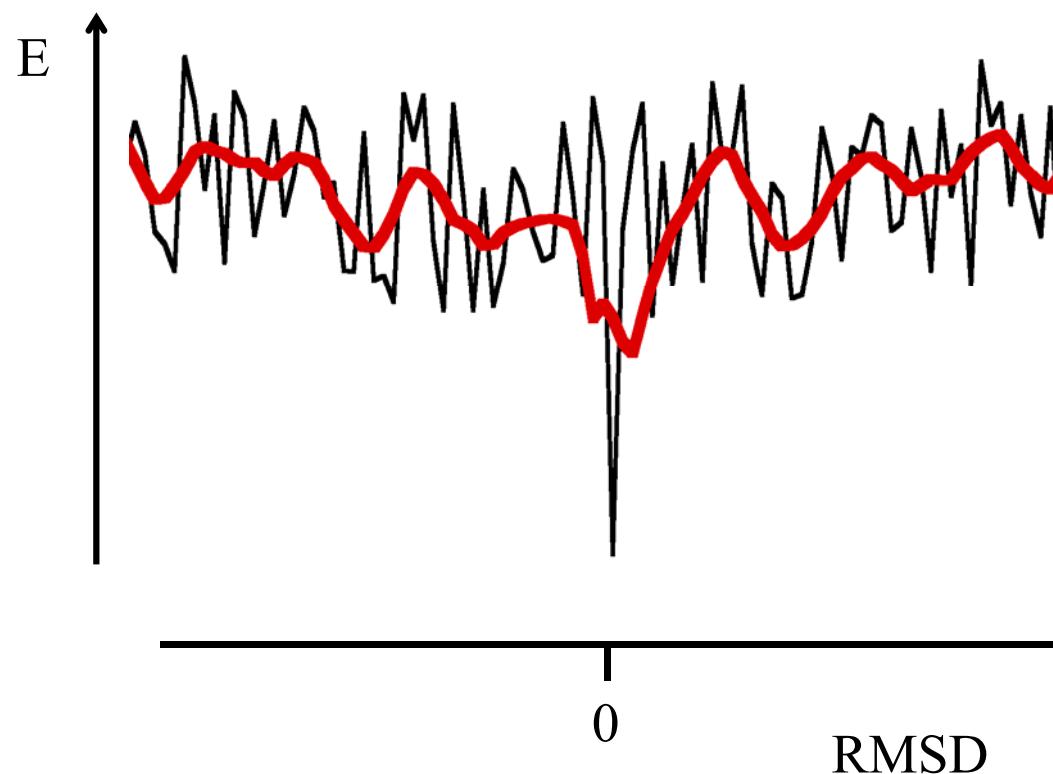
Energy-based docking

- Energy-based docking
 - **Energy minimization docking**
 - **HADDOCK: Docking with restraints**
 - **pyDock: Scoring by energy**

Protein-Protein Docking Energy

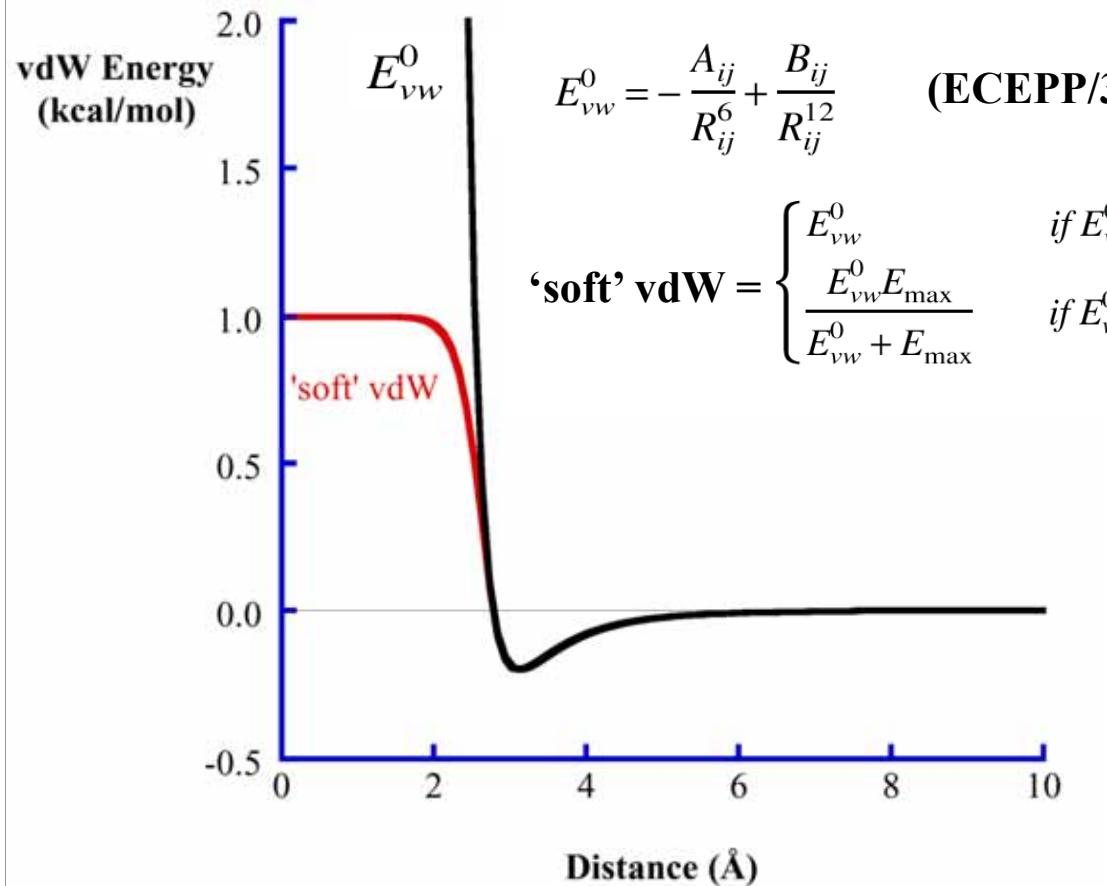


Protein-Protein Docking Energy



Protein-Protein Docking Energy

$$E = E_{\text{vw}} + E_{\text{el}} + E_{\text{hb}} + E_{\text{hp}}$$



$$E_{\text{el}} = 332.0 \frac{q_i^s q_j}{4d_{ij}^2}$$

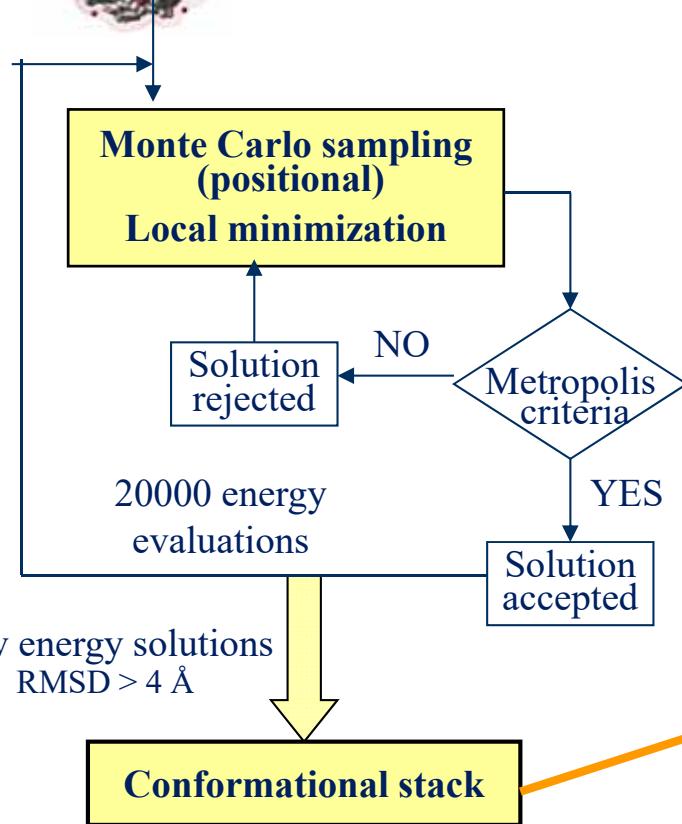
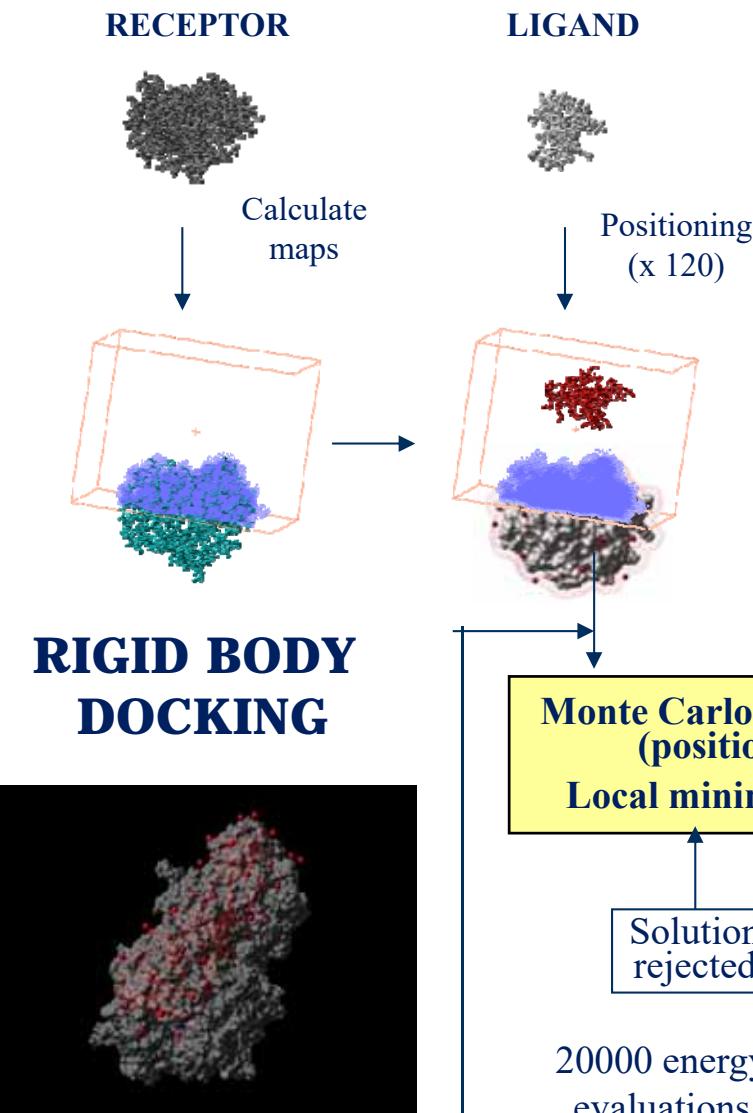
Max $E_{\text{el}} = 20$ kcal/mole
Min $E_{\text{el}} = -20$ kcal/mole

$$E_{\text{hb}} = E_{\text{hb}}^0 e^{-[(r-r_{ep})^2/d_{\text{hb}}^2]}$$

$E_{\text{hb}}^0 = 2.5$ kcal/mole
 $d_{\text{hb}} = 1.4$ Å

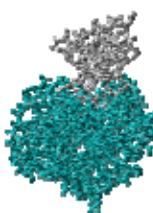
$$E_{\text{hp}} = 0.03 \text{ kcal/mole} * \text{ASA(apolar)}$$

ICM-DISCO protein docking



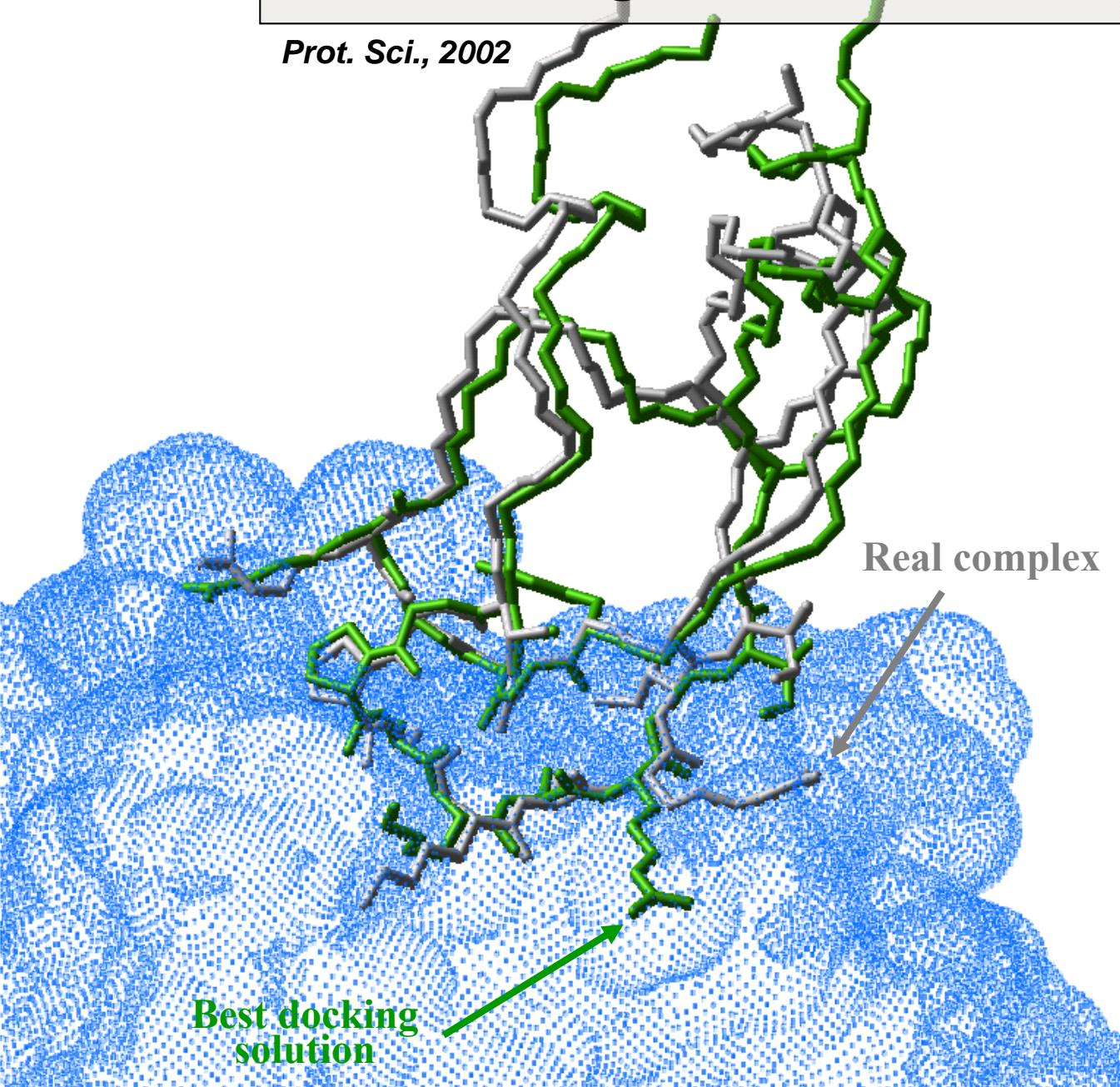
**Pseudo-Brownian
Monte-Carlo Minimization**

**Energy funct =
vdw + el + hb + desolv**

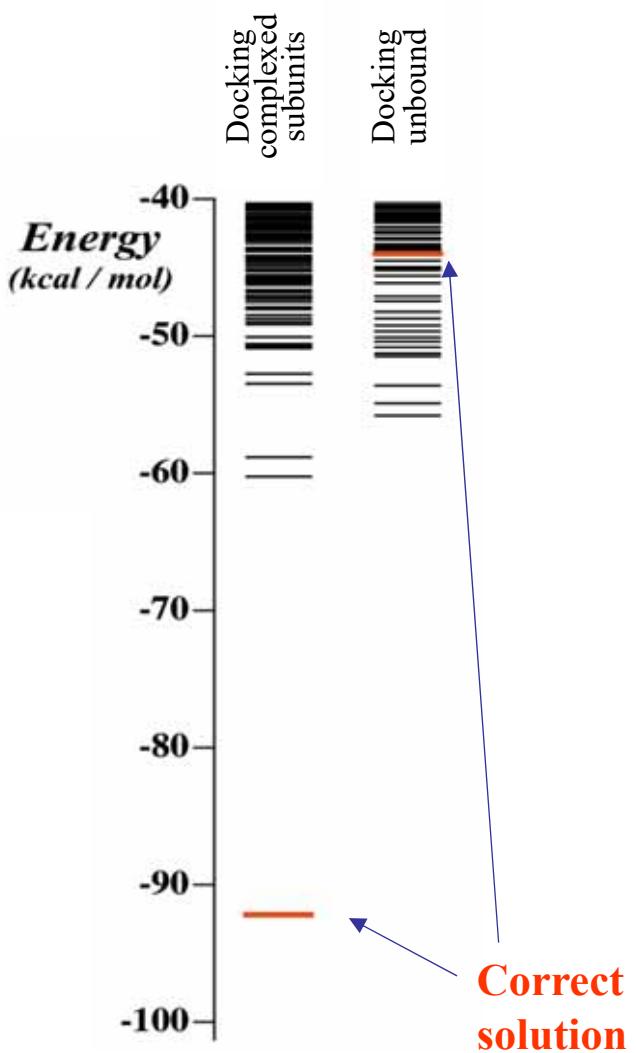


Docking unbound trypsin/BPTI

Prot. Sci., 2002

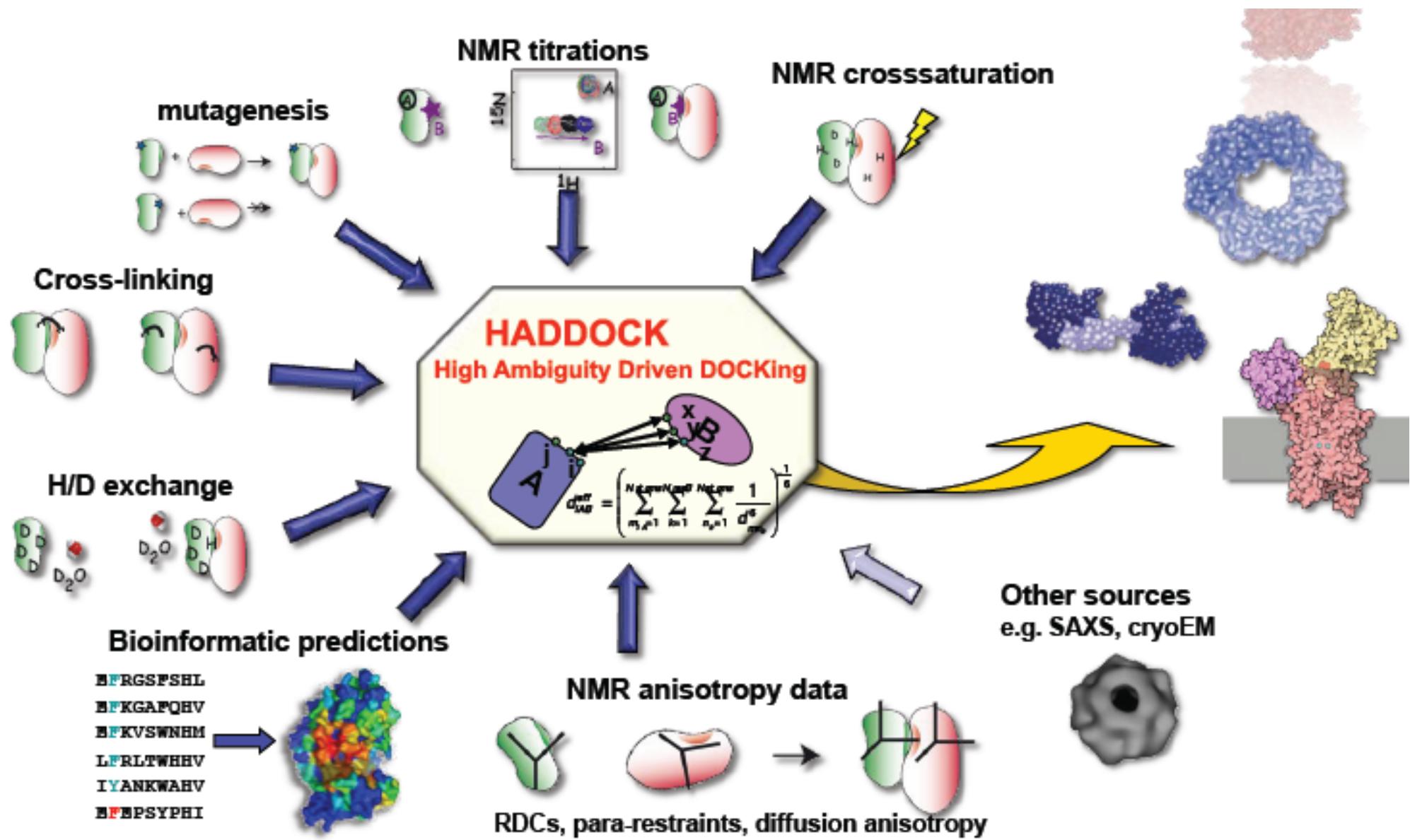


Distribution of solutions

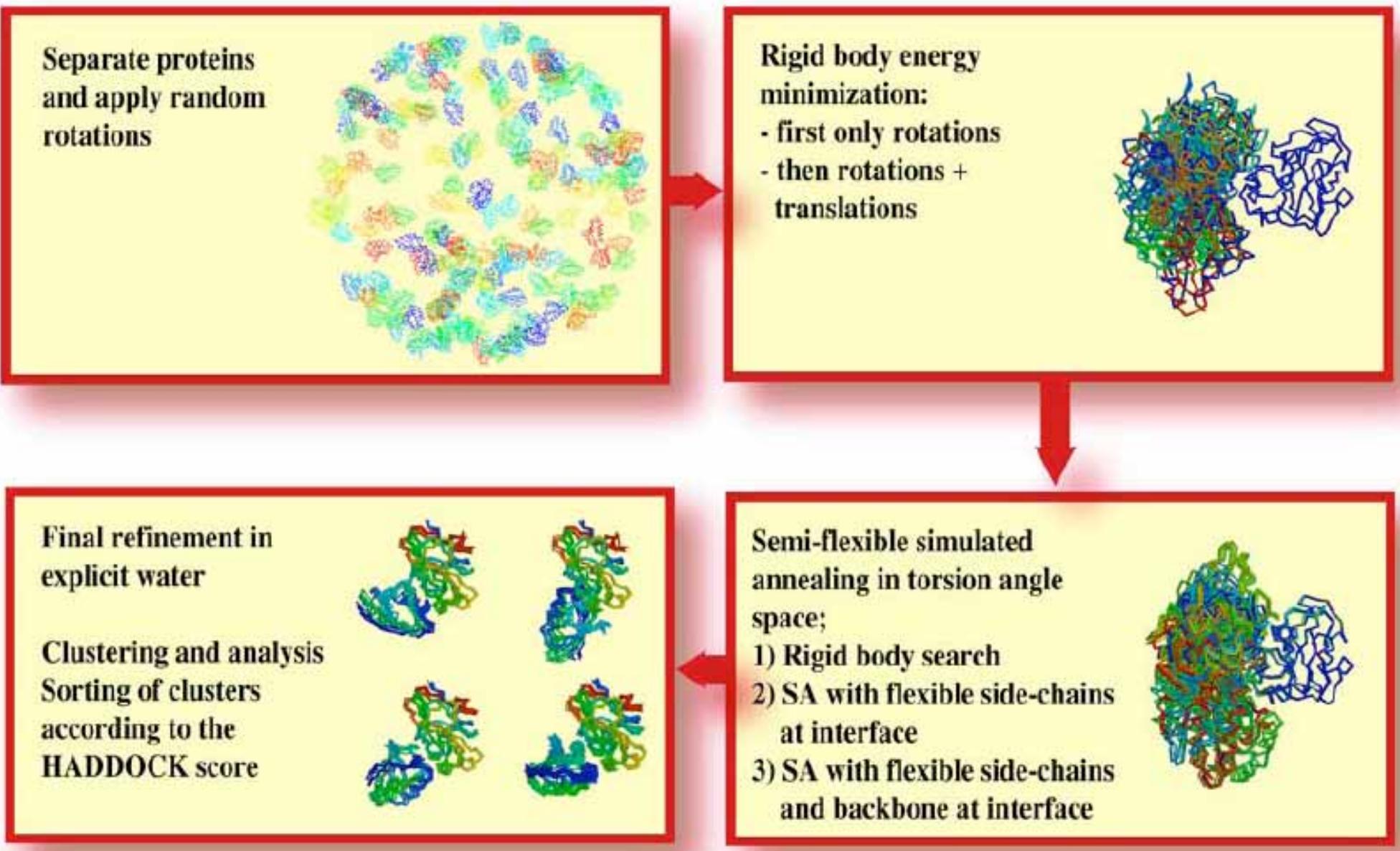


- Energy-based docking
 - **Energy minimization docking**
 - **HADDOCK: Docking with restraints**
 - **pyDock: Scoring by energy**

HADDOCK



HADDOCK



HADDOCK

The HADDOCK web portal

HADDOCK
Software web portal

WELCOME TO THE Utrecht BIOMOLECULAR INTERACTION WEB PORTAL >>

HADDOCK (High Ambiguity Driven protein-protein DOCKing) is an information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modeling problems including protein-protein, protein-nucleic acids and protein-ligand complexes.

More information about HADDOCK can be found on the HADDOCK website

HADDOCK WEB SERVER

To use the HADDOCK docking server you must have registered for an account. If you do not have a account yet you can [register here](#).

- HADDOCK server: the Easy Interface
- HADDOCK server: the Prediction Interface
- HADDOCK server: the Expert Interface (requires Expert level access)
- HADDOCK server: the Refinement interface (requires Expert level access)
- HADDOCK server: the Guru Interface (requires Guru level access)
- HADDOCK server: the Multi-body interface (requires Guru level access)
- HADDOCK server: the File upload interface

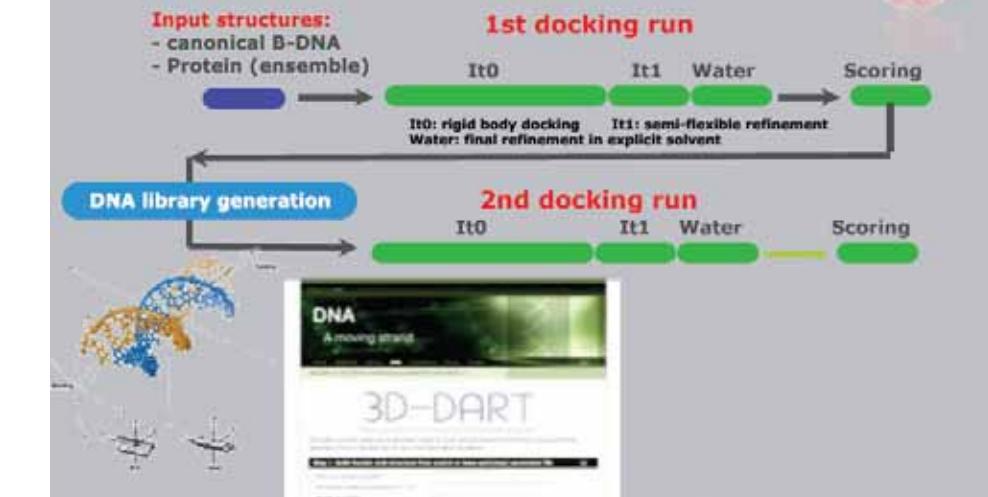
~12000 served requests as of June 2008 (~5% on the eNMR grid)

registered users: 1139
(832 easy / 110 expert / 197 guru)

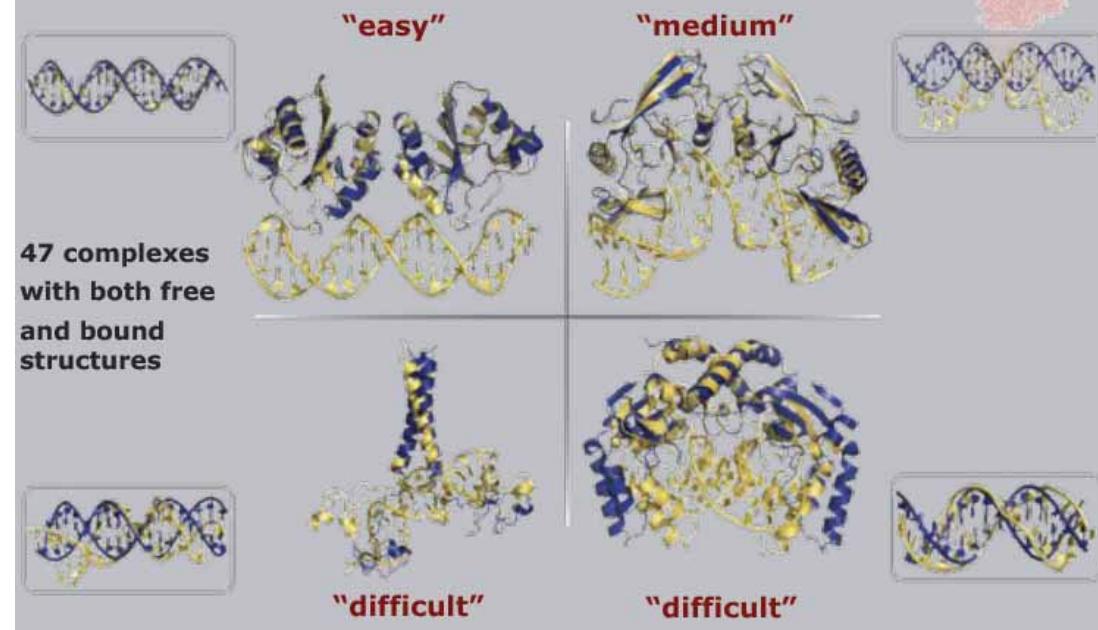
The HADDOCK PDB structure gallery



A two-stage docking procedure

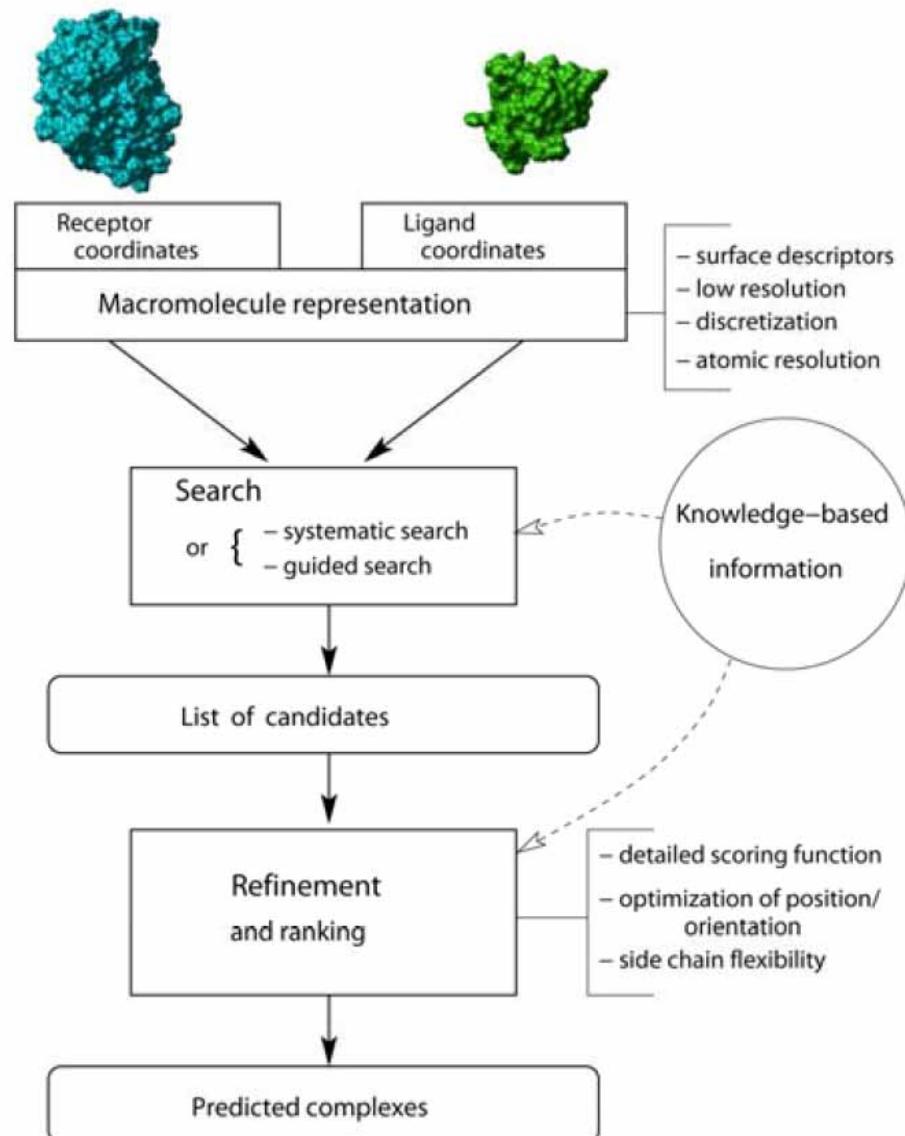


Protein-DNA benchmark



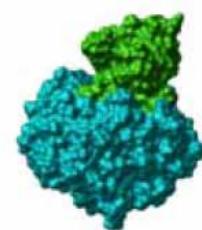
- Energy-based docking
 - **Energy minimization docking**
 - **HADDOCK: Docking and restraints**
 - **pyDock: Scoring by energy**

docking

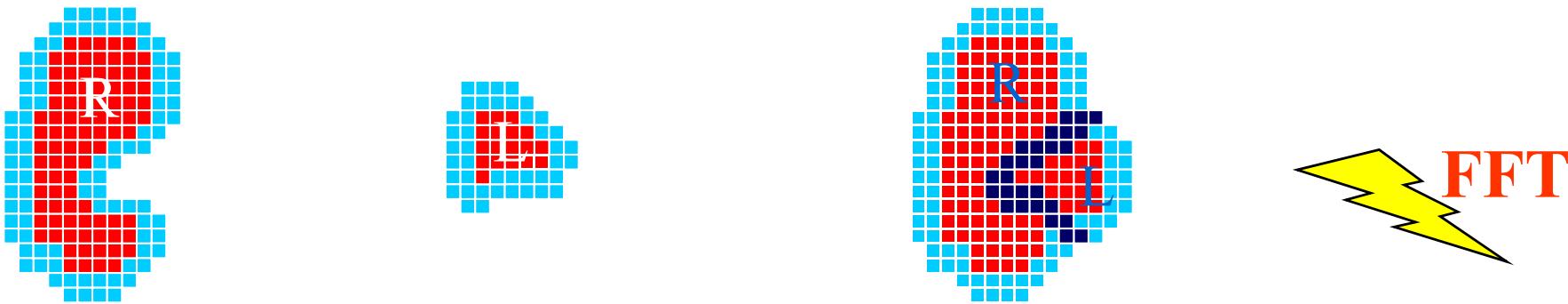


SAMPLING & SCORING

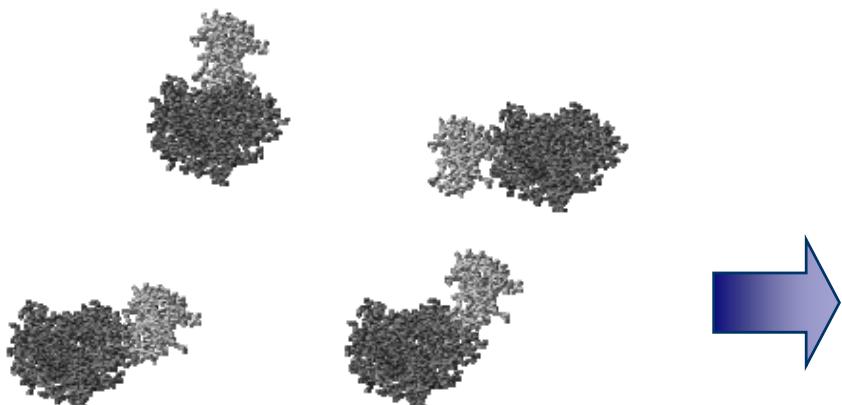
SAMPLING & SCORING



pyDock: FFT sampling + energy scoring



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

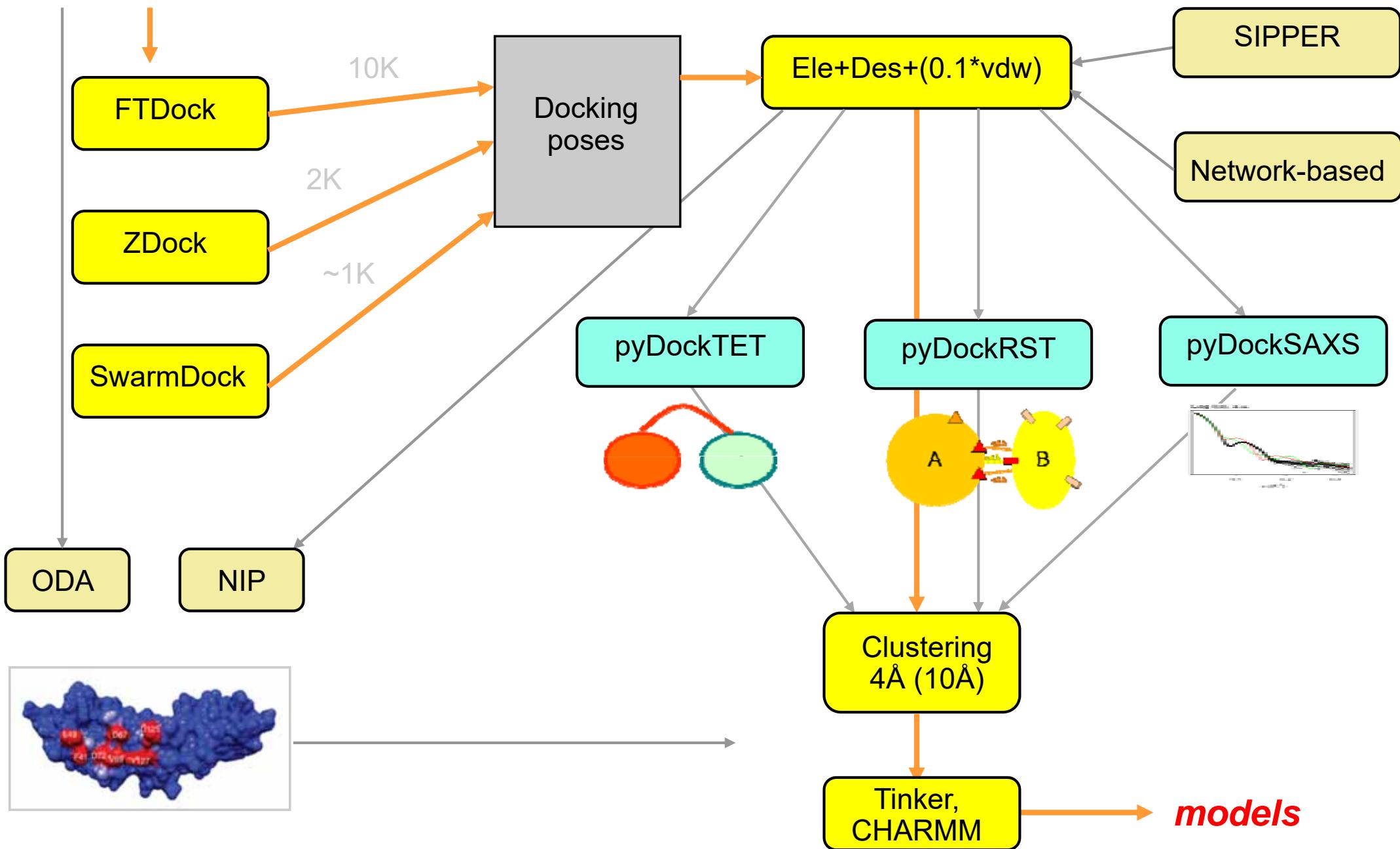


$$\begin{aligned} E = & 0.1 \sum_i^{REC} \sum_j^{LIG} \sqrt{e_i e_j} \left(\left(\frac{r_i + r_j}{d_{ij}} \right)^{12} - 2 \left(\frac{r_i + r_j}{d_{ij}} \right)^6 \right) + \\ & + 332.0 \sum_i^{REC} \sum_j^{LIG} \frac{q_i q_j}{4 d_{ij}^2} + \\ & + \sum_i^{REC} ADP_i BSA_i + \sum_j^{LIG} ADP_j BSA_j \end{aligned}$$

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

Unb proteins

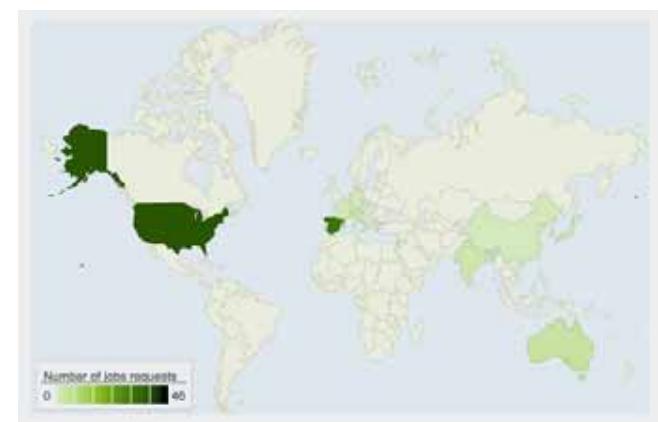
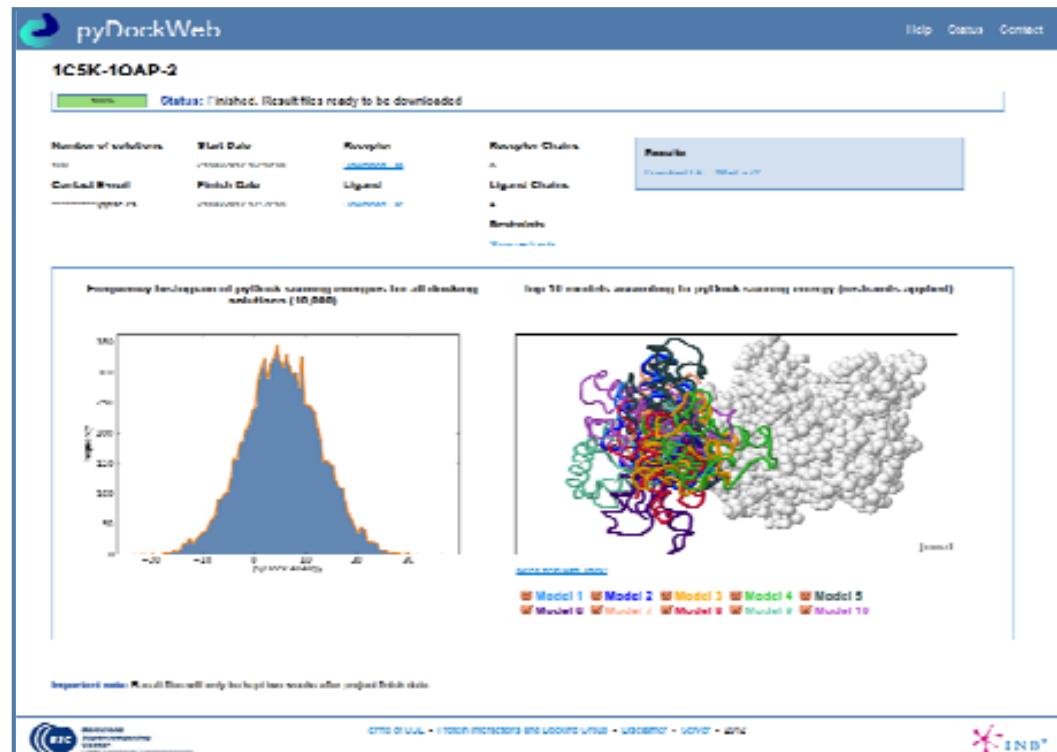
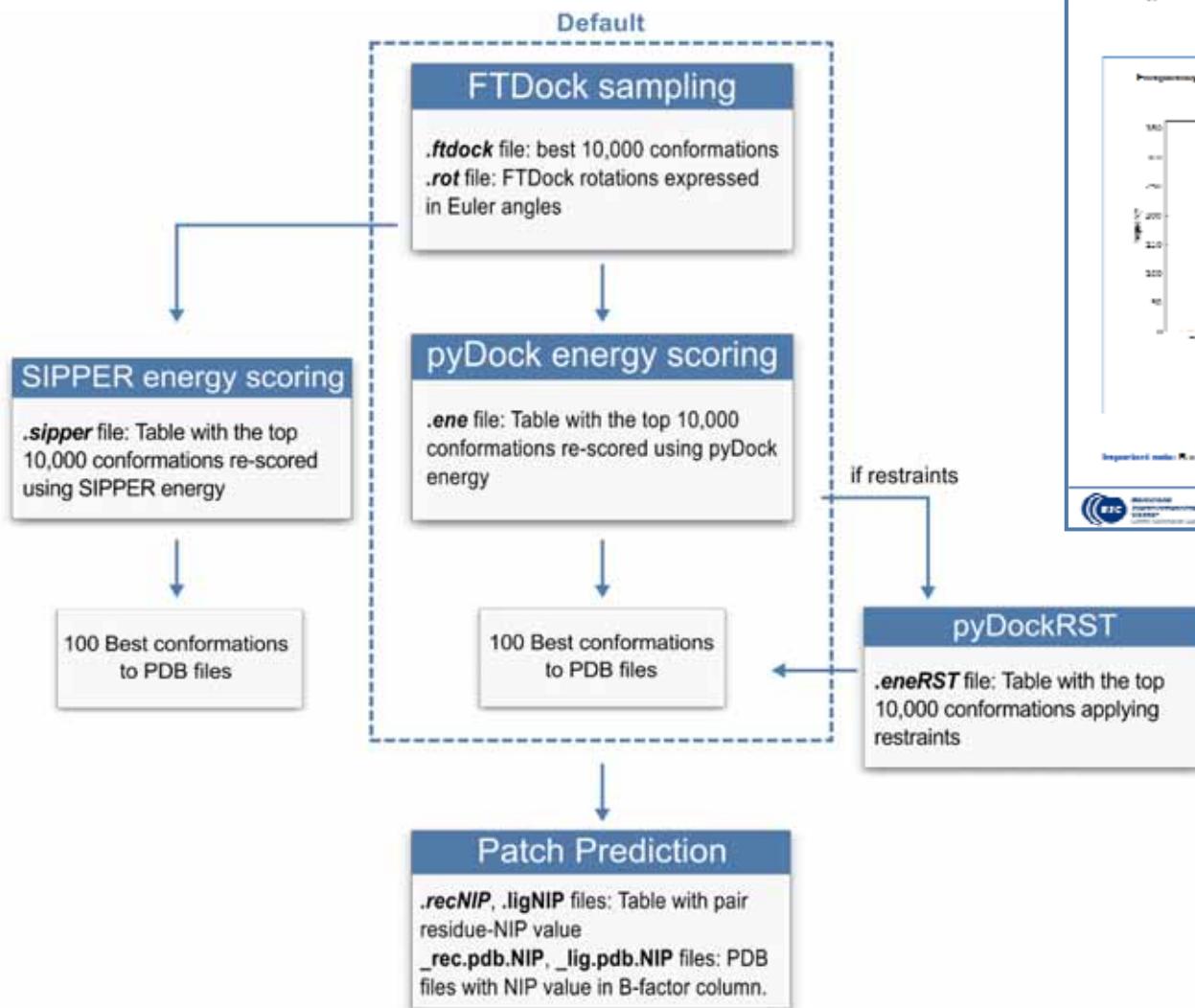
pyDock protocol





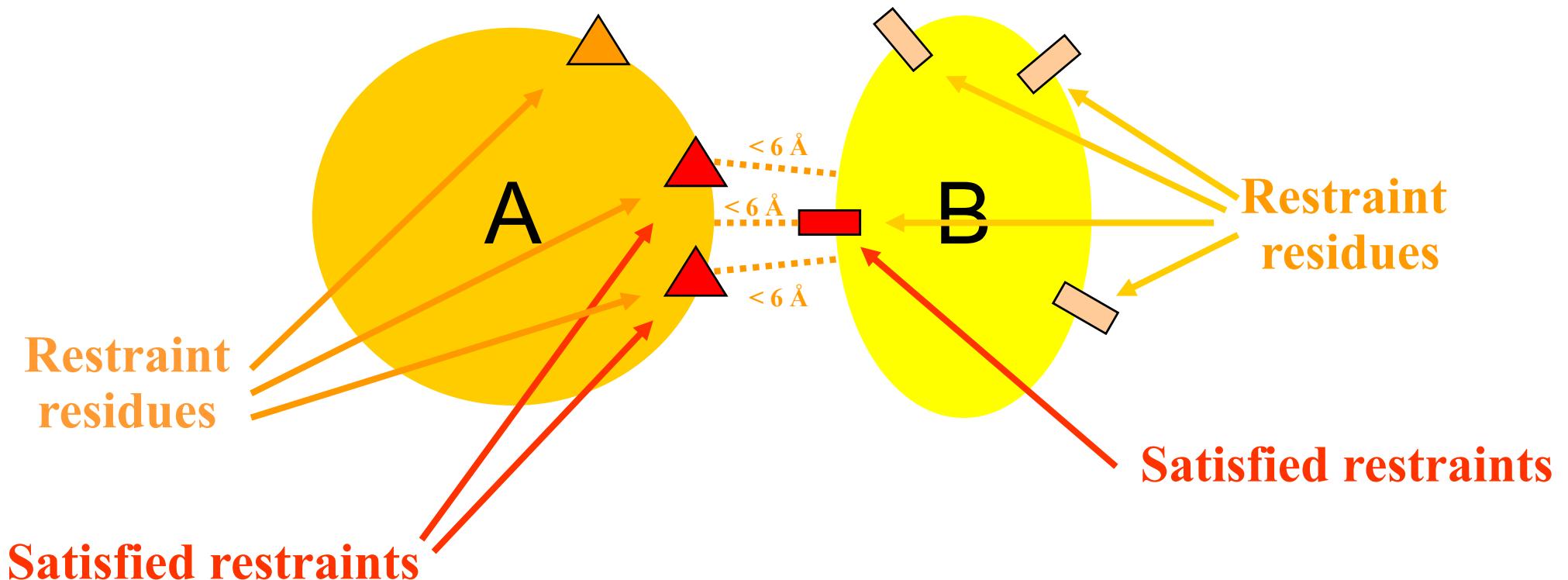
pyDock Web Server

<http://life.bsc.es/servlet/pydock>



pyDockRST: use of restraints to filter docking solutions

Docking solution *i*

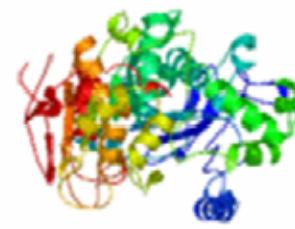
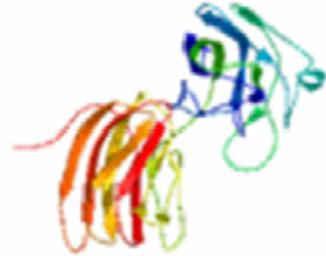
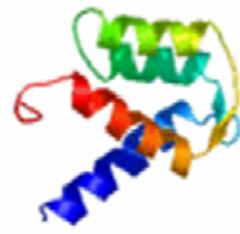
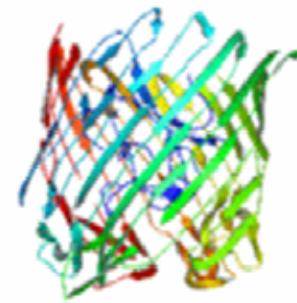
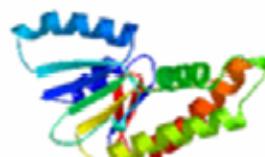


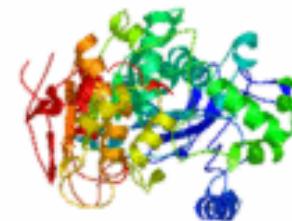
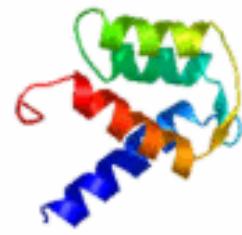
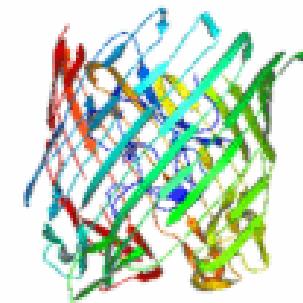
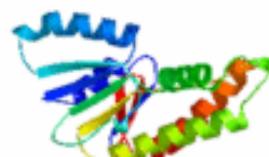
$$\text{Restraint pseudo-energy} = -1.0 * (\% \text{ satisfied restraints})$$

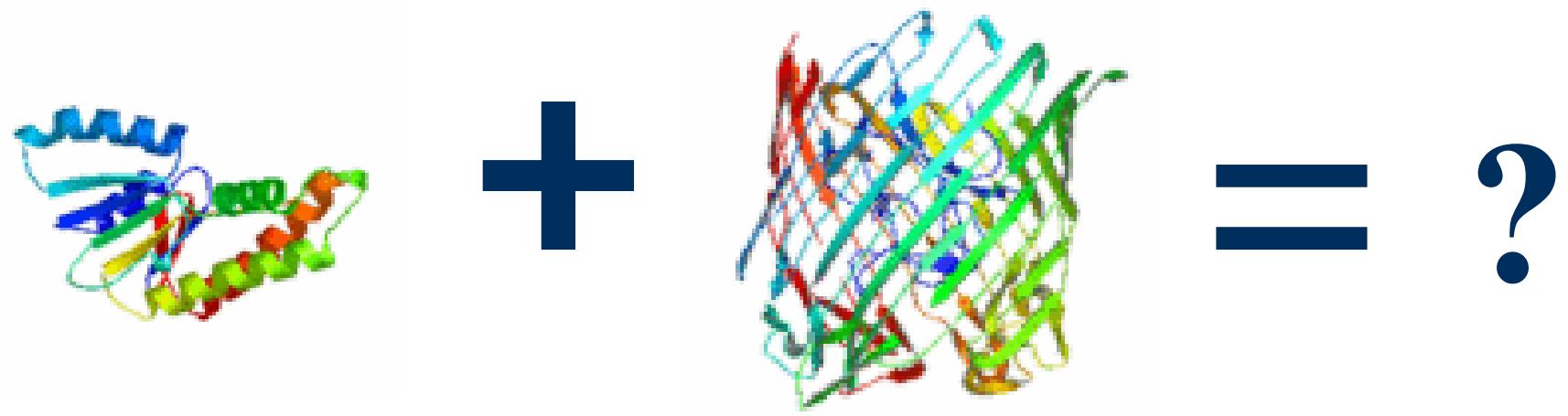
Energy-based docking: conclusions

- *Fewer available methods*
- *Slower than geometry-based*
- *Better success rates*
- *Better geometries*
- *Better understanding of binding mechanism*
- *Desolvation is very important in docking*

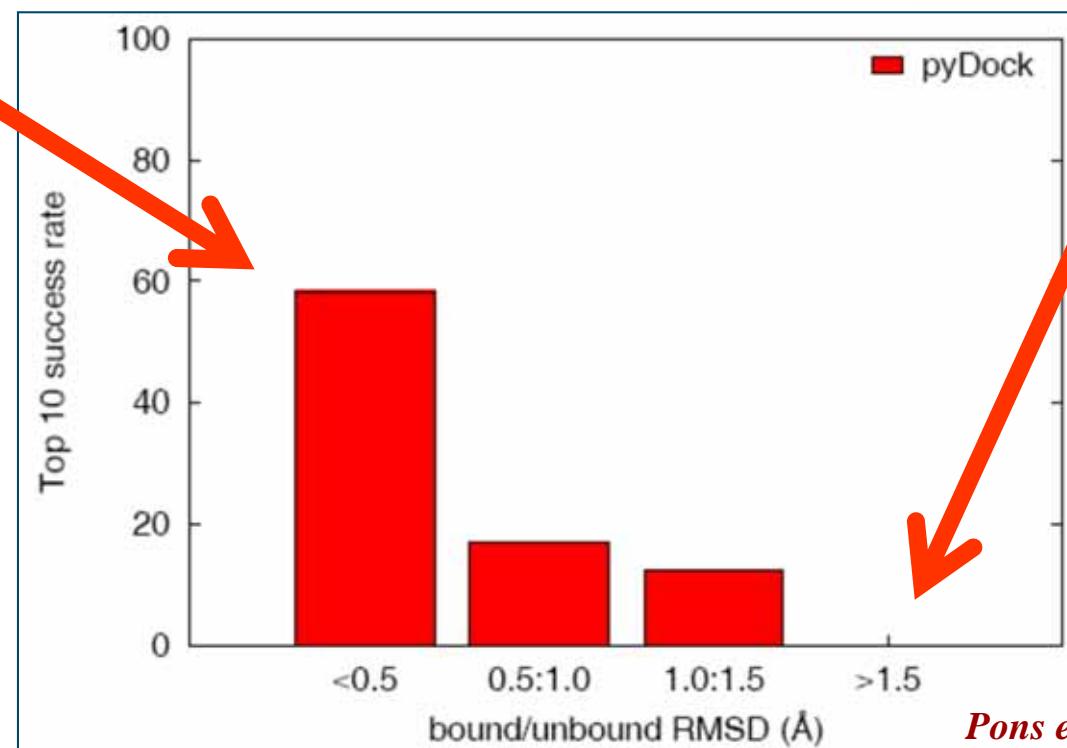
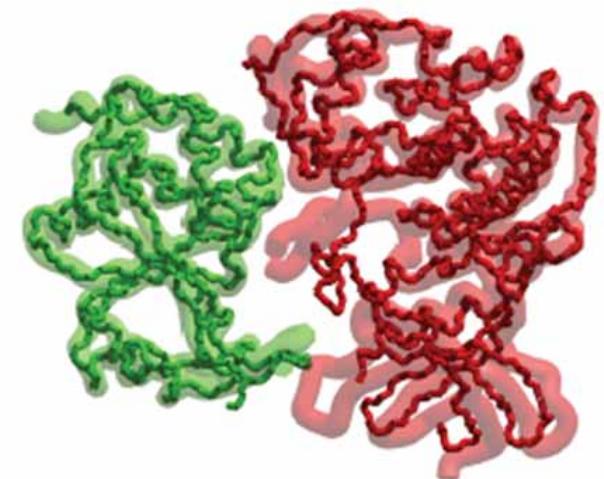
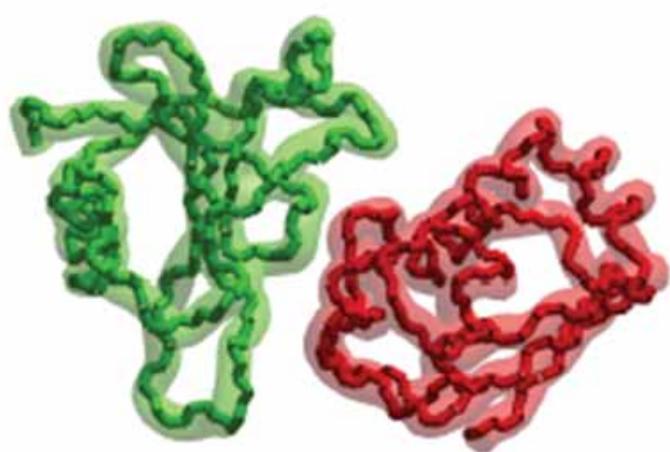
- Flexibility in docking
 - **Introduction**
 - **Rigid-body + Refinement**







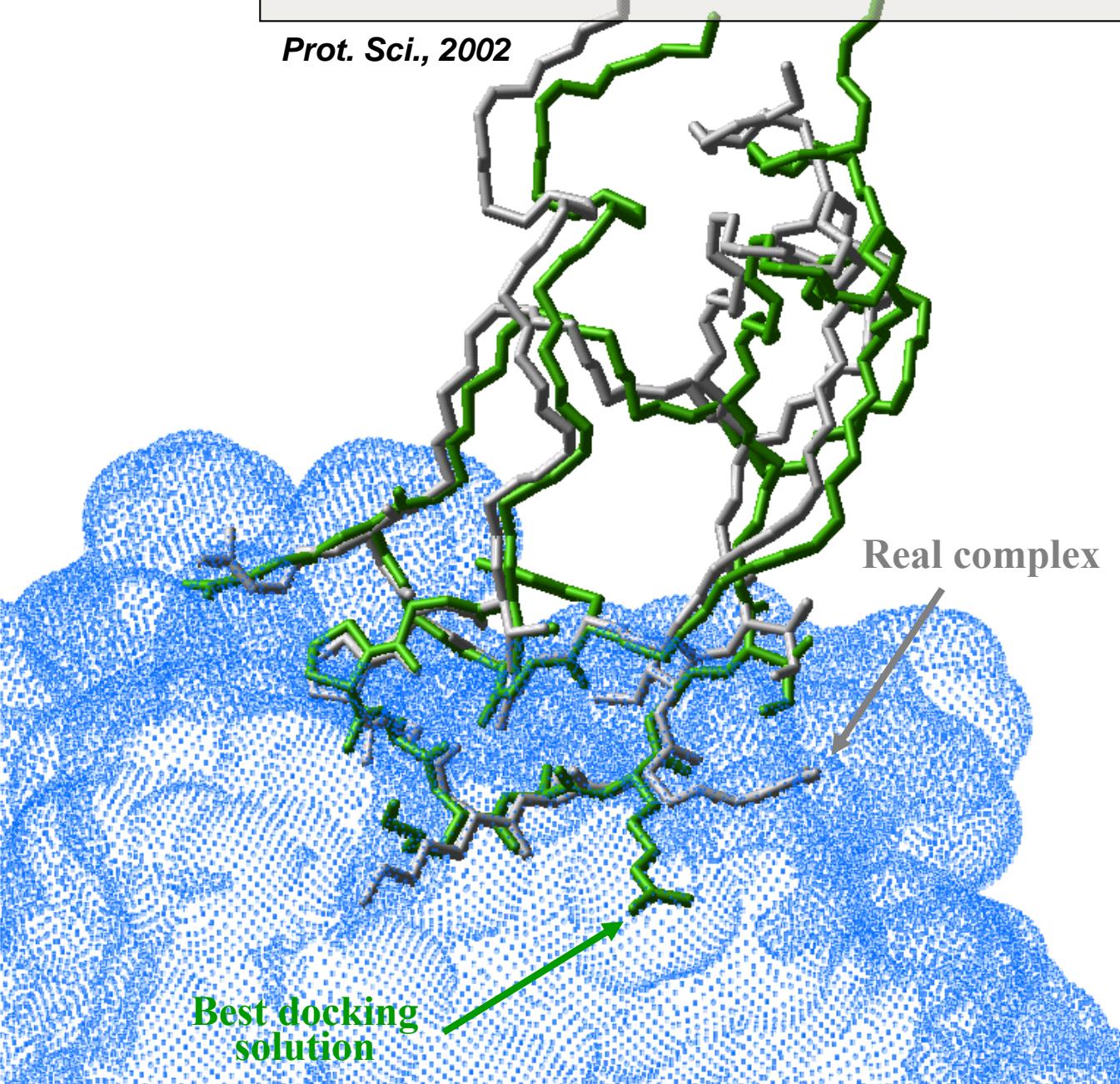
Rigid-body limitations: flexible cases



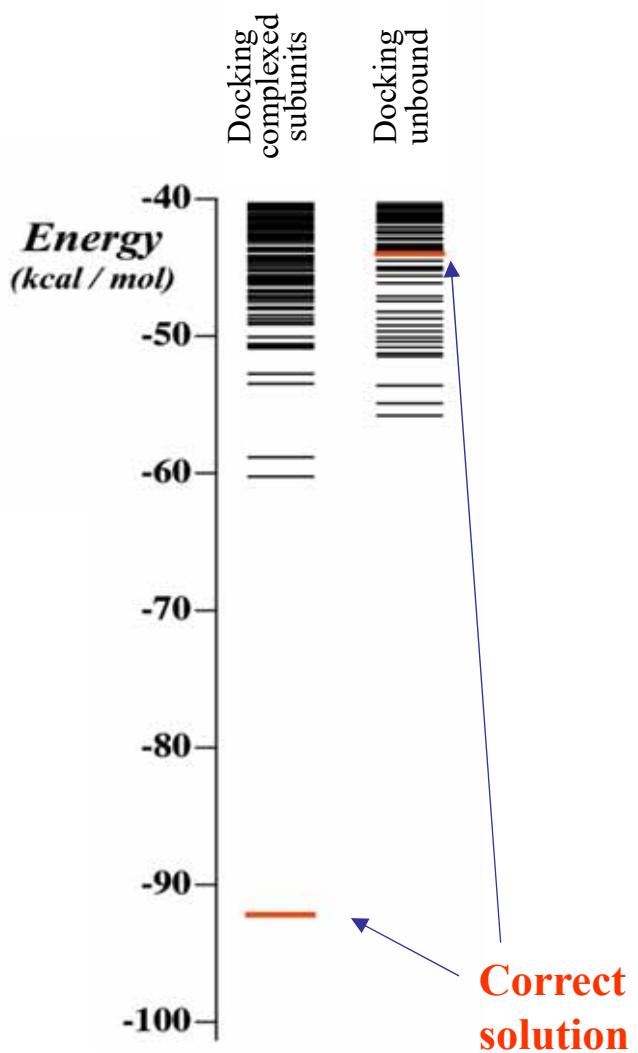
Pons et al. (2010) Proteins 78, 95-108

Docking unbound trypsin/BPTI

Prot. Sci., 2002

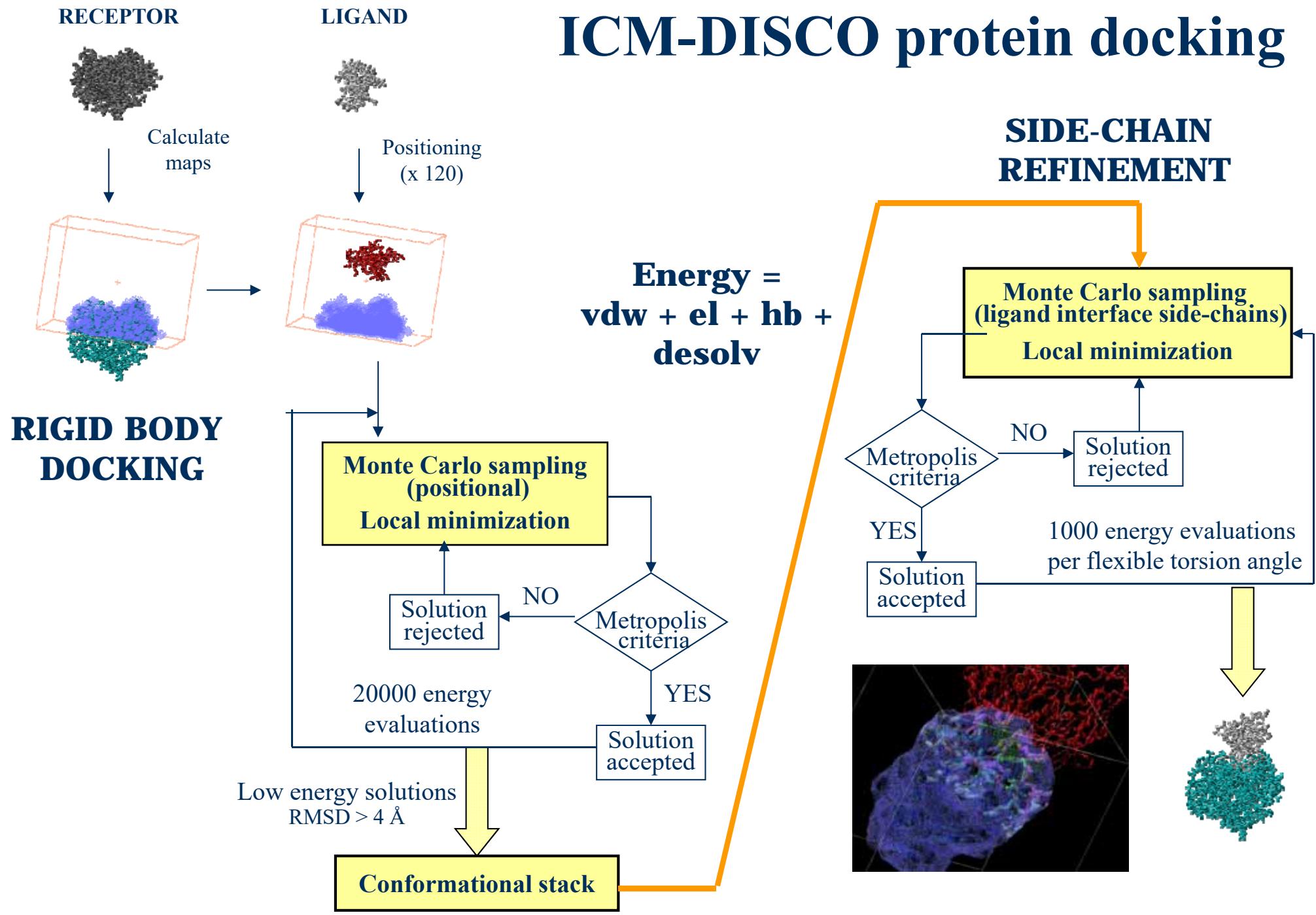


Distribution of solutions



- Flexibility in docking
 - Introduction
 - Rigid-body + Refinement

ICM-DISCO protein docking

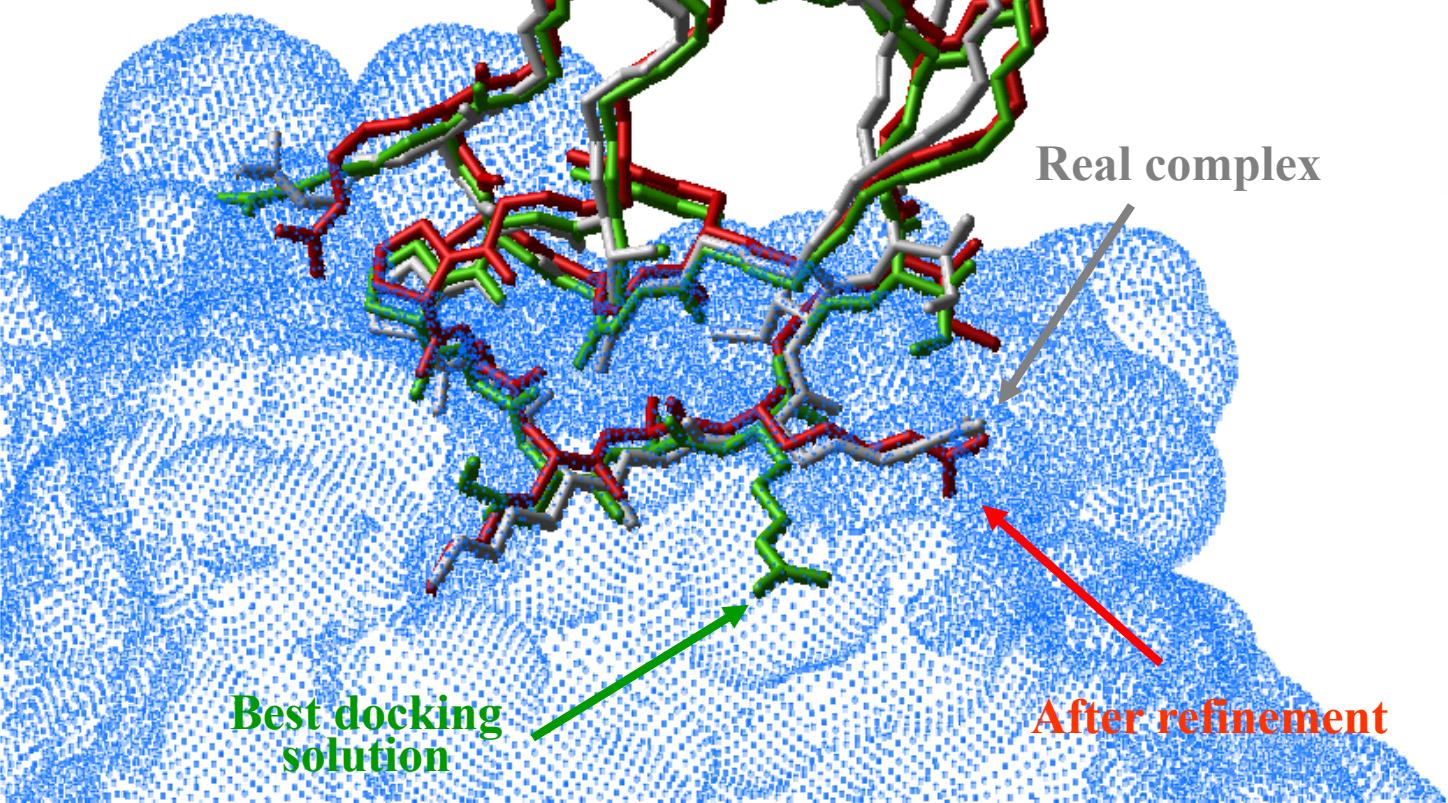


Fernández-Recio et al. 2002 Protein Sci. 11, 280-291

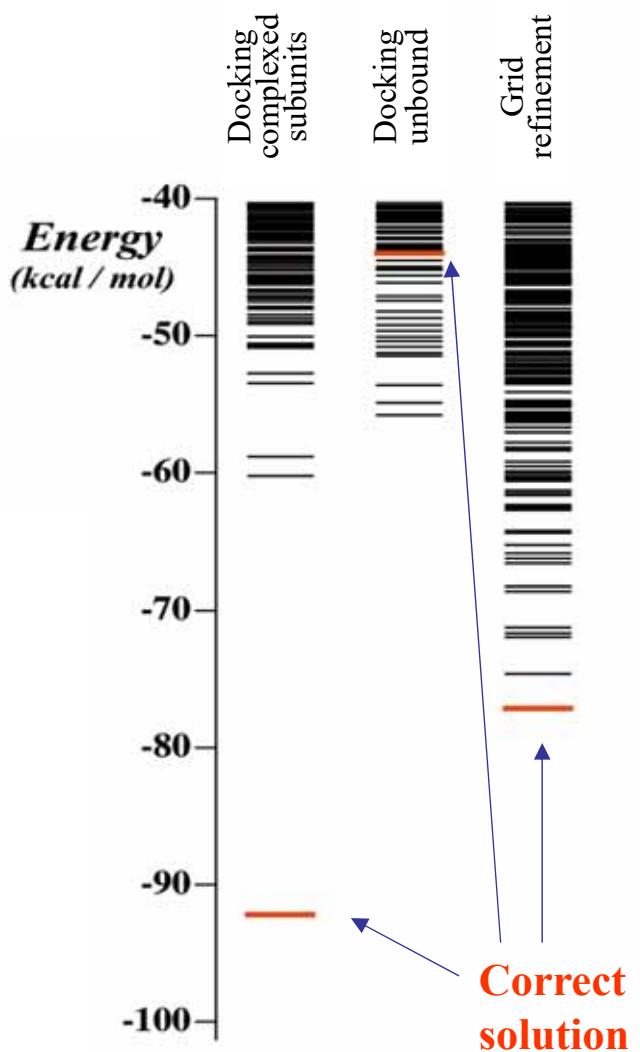
ICM Interface Side-Chain Optimization

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

Prot. Sci., 2002



Distribution of solutions



Rigid-Body Docking + Side-Chain Refinement

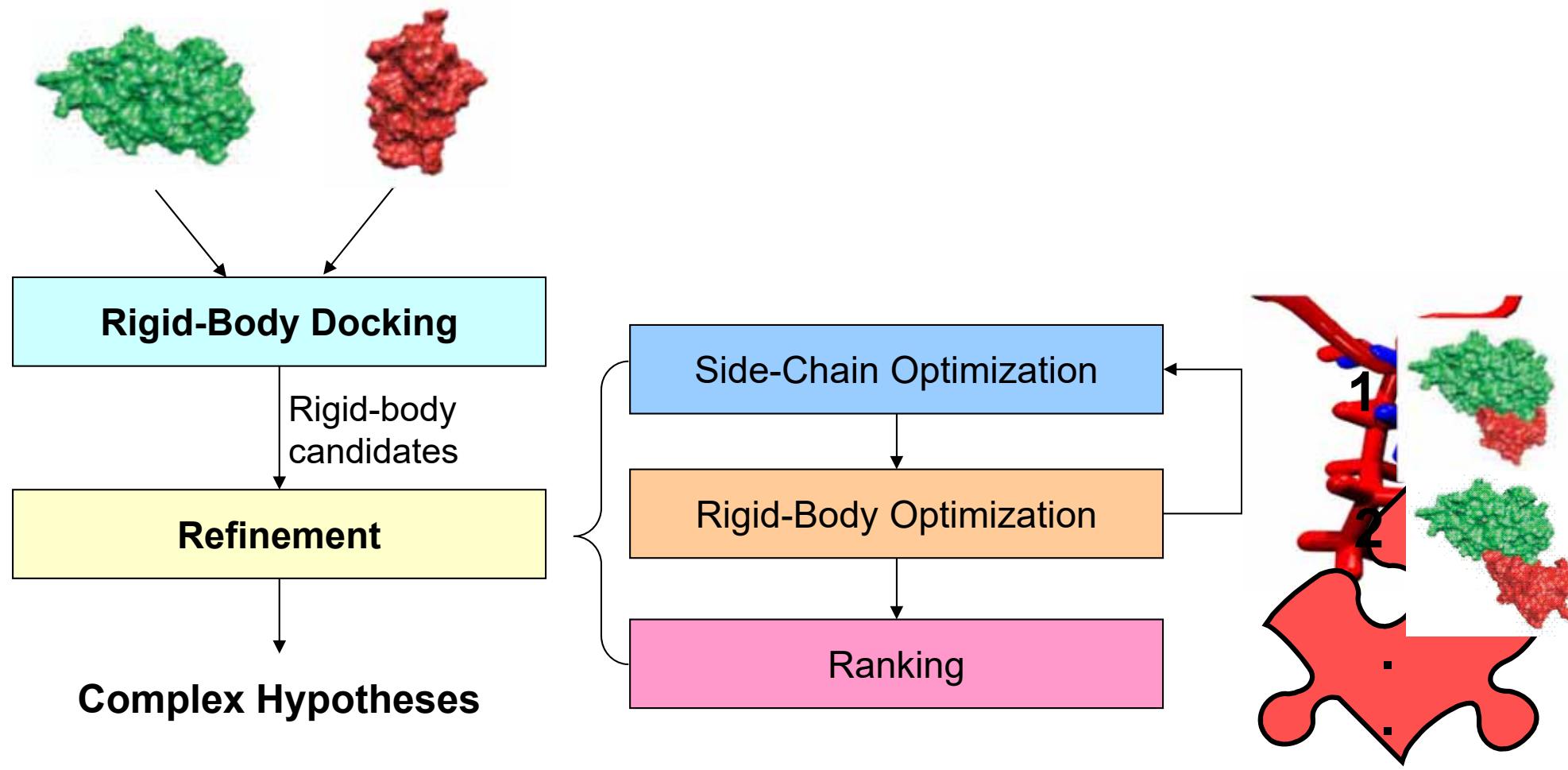
PROTEIN-PROTEIN DOCKING

*predicting hemagglutinin/Fab
complex for the CAPRI
competition 2002*

Juan Fernandez-Recio, Max Totrov & Ruben Abagyan

FireDock

Fast Interaction Refinement in Molecular Docking



Interface Side-Chain Optimization

1. Selection of movable residues

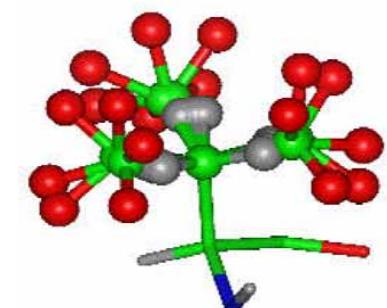
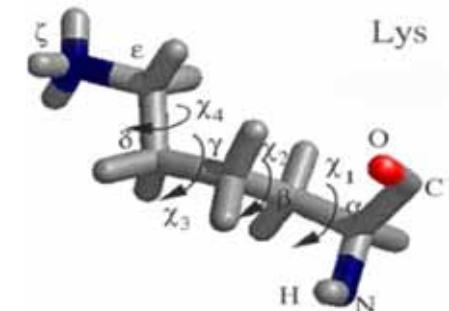
2 versions of Interface Side-Chain Optimization:

- Full (FISCO)
- Restricted (RISCO)

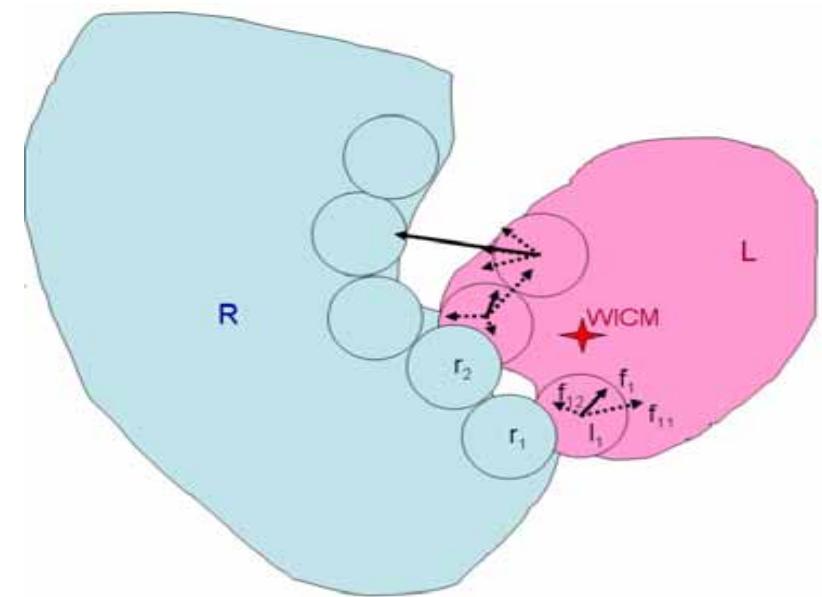
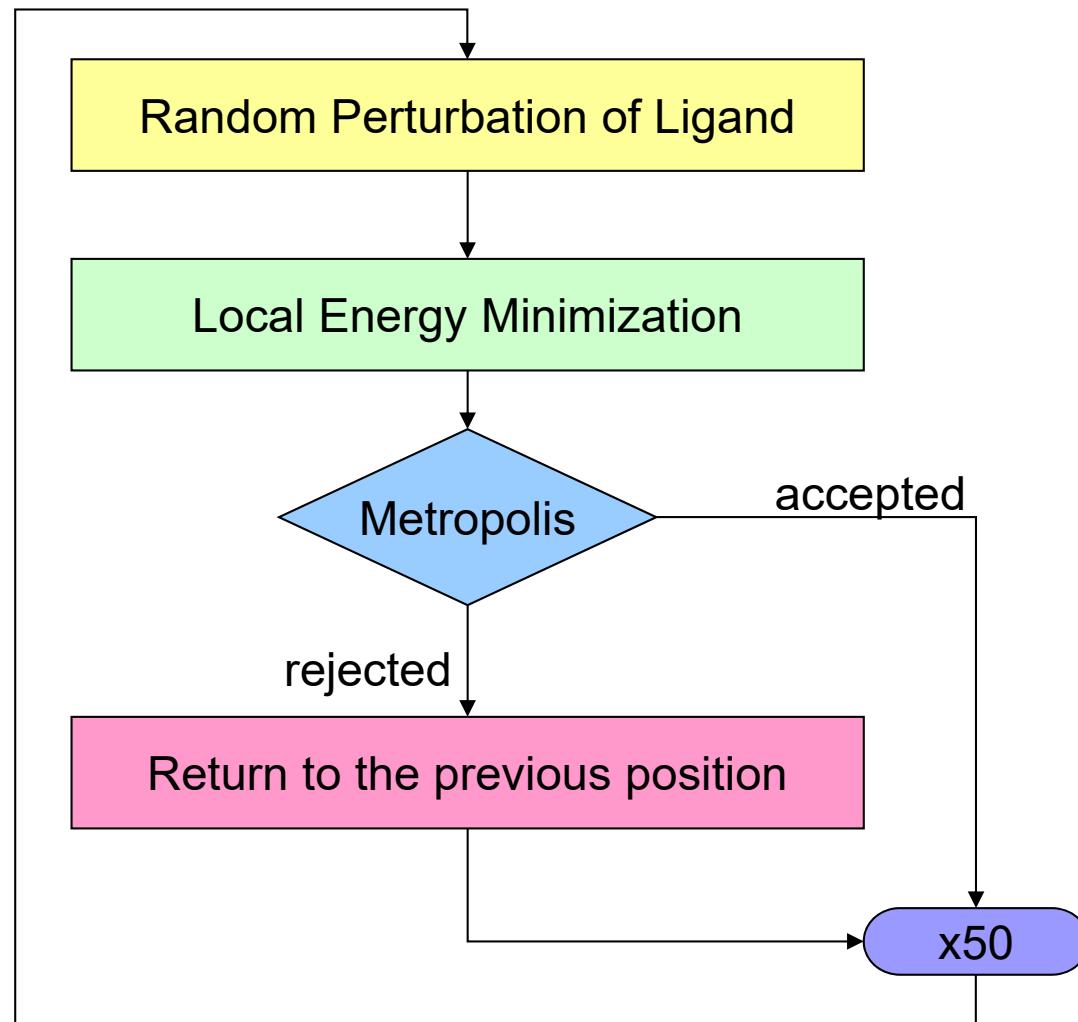
only clashing residues are movable

2. Energy calculation for backbone dependent rotamers

3. Solving the optimization using ILP



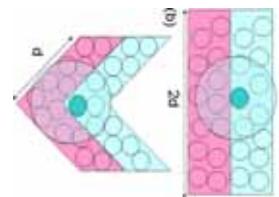
Rigid-Body Optimization



Ranking

■ According to binding score:

- ACE
- electrostatics
- π-stacking & aliphatic interactions
- van der Waals
- “insideness”



$$\begin{aligned} E_{EI} &= E_{s_attrVdW} + 0.95E_{s_repVdW} + 1.6E_{ACE} + \\ &\quad 0.07E_{attEl} + 0.12E_{repEl} + 0.3E_{l_repEl} + \\ &\quad 1.32E_{HB+SS} + E_{pipi} + 0.8E_{catpi} + 0.5E_{aliph} + 1.55E_{insideness} \\ E_{AA} &= 1.5E_{s_attrVdW} + 0.6E_{s_repVdW} + 1.6E_{ACE} + \\ &\quad 0.21E_{attEl} + 0.21E_{repEl} + 0.46E_{l_attrEl} + 0.69E_{l_repEl} + \\ &\quad 1.2E_{HB+SS} + E_{pipi} + 0.7E_{catpi} + 2.5E_{aliph} \end{aligned}$$

FireDock

Fast Interaction REfinement in molecular DOCKing

[Web Server] [About] [Download] [FAQ] [Help] [References] Contact: ppdock@tau.ac.il

FireDock is an efficient method for refinement and re-scoring of rigid-body protein-protein docking solutions.

Please choose one of the options below

Option 1 (use transformation file, a faster option)

Receptor Molecule:

(PDB:chainId e.g. 2kai:AB)

or

upload file:

[Browse...](#)

Ligand Molecule:

(PDB:chainId e.g.

Transformations File:

[Browse...](#)

(up to 100)

(Example: Receptor Molecule: 2kai:AB, Ligand Molecule: 2kai:I, Transformations File: transformations_file.txt)

Option 2 (use models file, a much slower option)

Models File:

[Browse...](#)

(up to 100)

(Example: Models File: [models_example.ent](#), Receptor chain: E, Ligand chain: I)

Number of output structures:

(up to 100)

Your e-mail address:

Transformations File:

A file containing the ligand's transformations in the following format:

index X-rotation Y-rotation Z-rotation

X-translation Y-translation Z-translation

For example:

1 2.11363 0.153389 2.82412 10.8802 -4.53751 -7.76723

2 1.3353 -0.0999924 -2.60445 44.0002 4.52072 8.73591

3 1.99255 0.159889 2.81949 12.0745 -2.47299 -9.35544

4 1.3353 -0.0999924 -2.49973 43.5566 7.54644 8.73591

5 1.32362 0.0314778 -2.64412 42.0512 3.61016 5.47968

Advanced Options: (optional)

[\[show\]](#) [\[hide\]](#)

E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov, H. J. Wolfson. NAR 08

FireDock

Fast Interaction REfinement in molecular DOCKing

[Web Server] [About] [Download] [FAQ] [Help] [References] Contact: ppdock@tau.ac.il

Advanced Options: (optional)

[\[show\]](#) [\[hide\]](#)

Complex Type: Default

Refinement Level: Restricted

Number of RBO Cycles: 50

Atomic Radius Scale: 0.8

Bound/Unbound: Receptor: Unbound

Ligand: Unbound

Fixed Residues Files: Receptor: [Browse...](#)

Ligand: [Browse...](#)

Flexible Residues Files: Receptor: [Browse...](#)

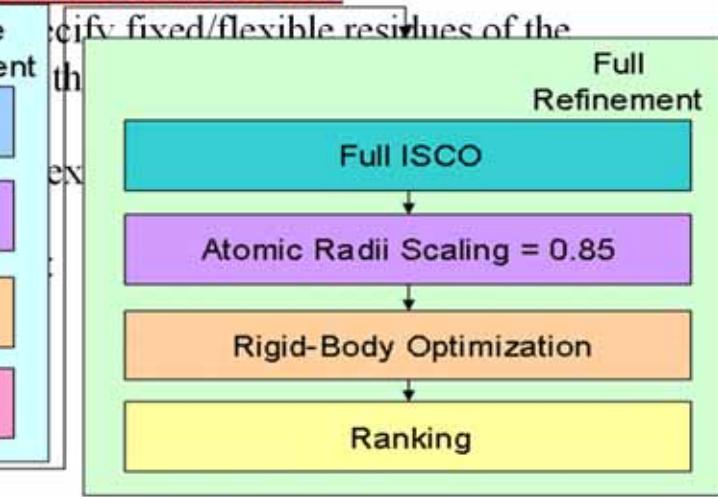
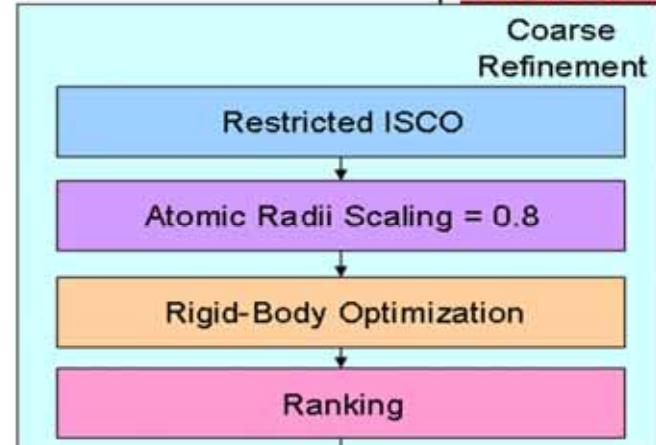
Ligand: [Browse...](#)

[Submit Query](#)

[Clear](#)

25 top-ranked candidates

Fixed/Flexible Residues Files:



FireDock

[Web Server] [About] [Download] [FAQ] [Help] [References]

Receptor

A.pdb

Ligand

C.pdb

TransFile

trans.txt

User e-mail

efratmas@gmail.com

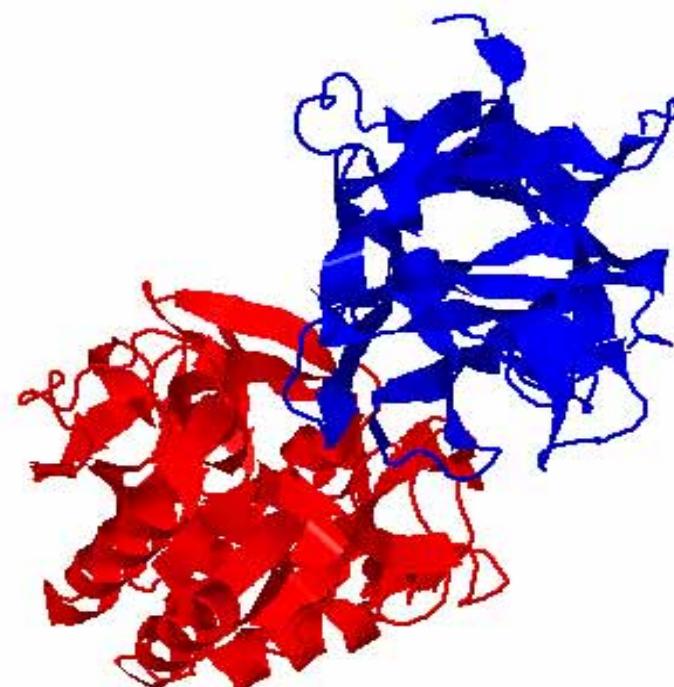
Rank	Solution Number	Global Energy	Attractive VdW	Repulsive VdW	ACE	HB	Structure show/hide
↓							
1	4	-22.88	-27.16	16.96	-11.83	-4.93	<input checked="" type="checkbox"/>
2	64	-21.95	-21.95	7.16	-11.61	-3.60	<input type="checkbox"/>
3	26	-21.50	-22.47	6.91	-11.01	-2.63	<input type="checkbox"/>
4	112	-21.40	-25.74	16.40	-11.18	-3.96	<input type="checkbox"/>
5	27	-20.67	-18.88	6.81	-9.82	-2.82	<input type="checkbox"/>
6	158	-20.00	-26.22	9.17	-10.38	-2.59	<input type="checkbox"/>
7	117	-19.27	-27.51	12.51	-11.22	-2.54	<input type="checkbox"/>
8	131	-17.37	-21.60	8.38	-13.76	-1.44	<input type="checkbox"/>
9	128	-17.01	-25.25	14.27	-9.50	-3.64	<input type="checkbox"/>
10	97	-16.46	-17.62	5.80	-10.33	-4.85	<input type="checkbox"/>
11	118	-16.40	-24.34	15.17	-9.13	-3.45	<input type="checkbox"/>
12	154	-15.38	-23.17	16.41	-12.05	-3.80	<input type="checkbox"/>
13	46	-15.22	-28.68	15.75	-10.95	-3.17	<input type="checkbox"/>
14	36	-15.21	-19.25	8.58	-8.62	-4.05	<input type="checkbox"/>
15	38	-14.84	-17.85	2.36	-8.15	-5.32	<input type="checkbox"/>
16	156	-13.58	-22.35	15.07	-9.46	-3.92	<input type="checkbox"/>
17	139	-12.25	-17.74	7.82	-11.66	-2.97	<input type="checkbox"/>
18	120	-10.88	-25.93	20.57	-13.12	-1.73	<input type="checkbox"/>
19	52	-9.29	-19.63	6.51	-8.62	-3.60	<input type="checkbox"/>
20	111	-9.15	-21.51	10.50	-7.28	-2.76	<input type="checkbox"/>

[show next](#)

[20 »»](#)

[download solutions table](#) [download best structures](#)

[show all/hide all](#)



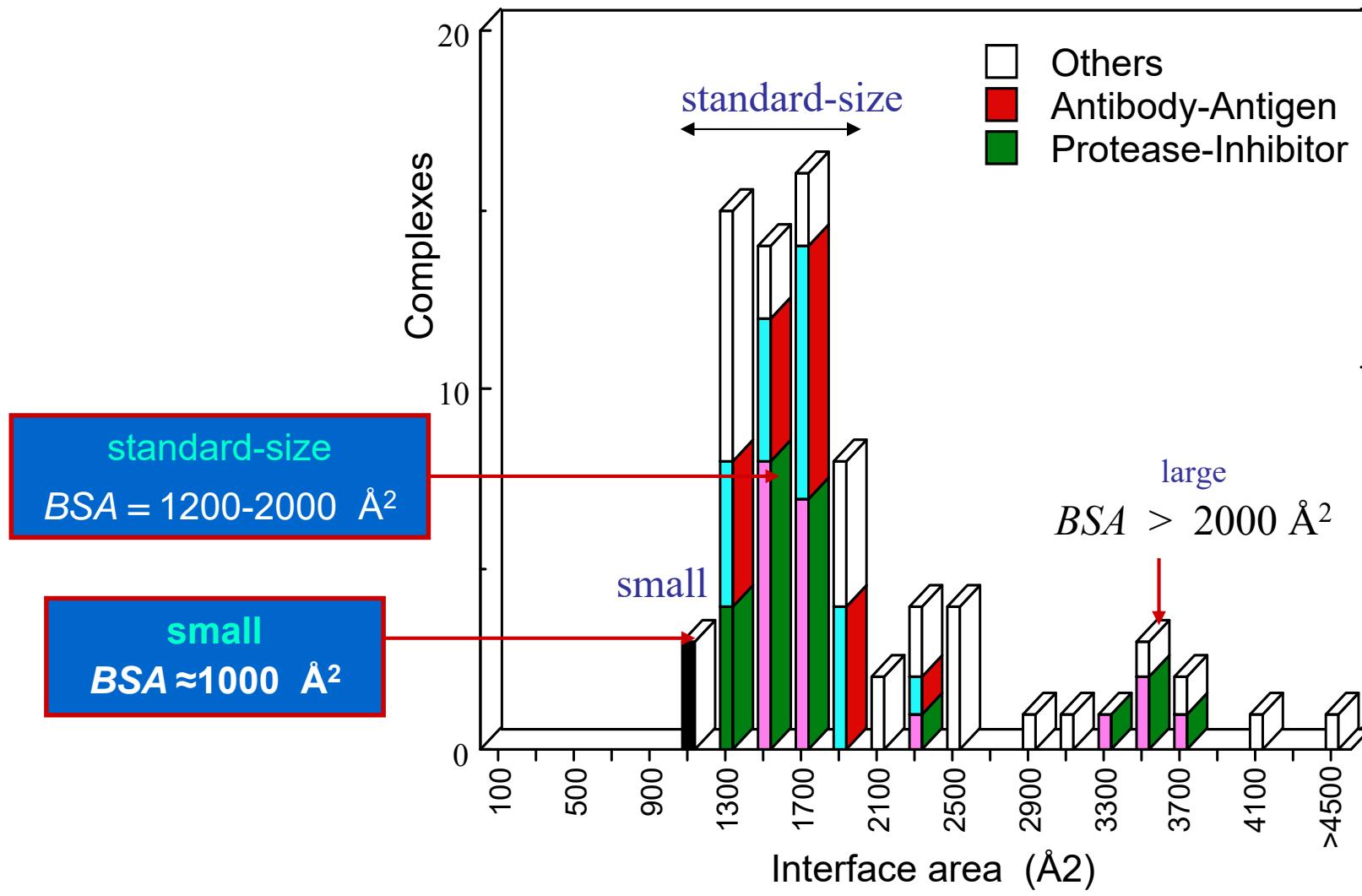
Jmol

Flexibility in docking: conclusions

- *Flexibility is the main challenge*
- *Very costly computationally*
- *Refinement useful in small-conformational changes*
- *Flexibility during docking not yet practical*

- Recognition and prediction of interaction sites
 - **Residue interface prediction**
 - **Hot-spot identification**
 - **Applications to biomedicine**

Protein-protein interface size



Protein-protein interface prediction

Common Descriptors Used for Prediction of Interaction Sites

Descriptor	Comments
Features Derived from 3-D Structure	
Neighbor list: Residues in spatial vicinity to the residue in question	9–20 residues
B-factor	A crystallographic measure that approximates the flexibility of a residue.
Solvent accessibility (ASA)	Measured in Å ² .
Relative solvent accessibility	Measured as a fraction of the overall surface of the residue that is exposed to solvent.
Shape index/curvedness	
Secondary structure	Three state (helix, strand, loop) or more.
Sequence distance	The separation in sequence between residues within the same patch. Some results indicate that structurally contiguous residues that are not adjacent in sequence are more likely to form interaction sites.
Planarity	
Predicted/Approximate Structural Features	
Predicted secondary structure	Methods relying only on sequence can use computational tools to generate predicted solvent accessibility/secondary structure.
Predicted solvent accessibility	This improves performance without limiting applicability to proteins with known 3-D structure.
Sequence neighbor list	Can be used instead of neighbor list to approximate the environment of the analyzed residue. Nine to fifteen residues around the residue in question. Four to seven on each side of the residue. Some structure-based methods use this in addition to neighbor list.

Protein-protein interface prediction

Common Descriptors Used for Prediction of Interaction Sites

Descriptor	Comments
	Evolutionary Features
Sequence profile	Extracted from a multiple sequence alignment, a profile reveals patterns of evolutionary conservation.
Conservation score	A quantification of the level of conservation of an individual position.
Conservation of physicochemical traits	If the position is not conserved, scoring conservation of traits such as charge, hydrophobicity, or size may improve prediction.
	Physicochemical Features
Hydrophobicity	Several different scales are available.
Electrostatic potential	Measured for individual residue or for a patch. Requires 3-D structure.
Atom propensities	Serves as a way to sum physicochemical properties across residues in the patch.
Desolvation energy	Used mostly in predictions for rigid-body docking.
	External Knowledge
Protein–protein interaction	Can be used to: (1) identify sequence or structural elements that are significantly overrepresented in interacting pairs, and (2) to assess coevolution of positions in interacting pairs.
Functional annotation of the protein	Enzyme–inhibitor and antigen–antibody have different types of interfaces than other complexes. Adding this information may improve prediction.

ProMate

<http://bioinfo41.weizmann.ac.il/promate/>

bioinfo41.weizmann.ac.il/promate/

Apps Imported From Firefox LIST OF PUBLISHERS | ... Beall's List of Predator... Portada elPeriódico.com EL PAÍS: el periódico g... BBC BBC - Homepage

ProMate 2.0
MultiProMate
ProMateus

ProMate's help
ProMateus' help
ProMateus' FAQ

Check out
our visitors
around the world!

Memorial

ProMate

*Predicting the location of potential protein-protein binding sites
for unbound proteins*

Upload your pdb file:
Choose File No file chosen

And/Or enter a 4 letters PDB-ID: [Search the PDB...](#)

Which chain to use?

Please choose the scores configuration: use default let me choose

Single amino acids distribution:	<input type="radio"/> use <input checked="" type="radio"/> don't use
Atoms distribution:	<input checked="" type="radio"/> use <input type="radio"/> don't use
Chemical character:	<input checked="" type="radio"/> use <input type="radio"/> don't use
Amino acid pairs distribution:	<input checked="" type="radio"/> use <input type="radio"/> don't use
Evolutionary conserved positions:	<input checked="" type="radio"/> use <input type="radio"/> don't use

Promate

<http://bioinfo41.weizmann.ac.il/promate/>

Please choose the scores configuration: use default let me choose

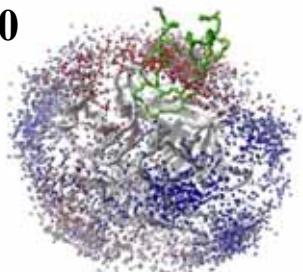
- | | |
|---|--|
| Single amino acids distribution: | <input type="radio"/> use <input checked="" type="radio"/> don't use |
| Atoms distribution: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Chemical character: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Amino acid pairs distribution: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Evolutionary conserved positions: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Non-regular secondary structure length: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Sequence distances within a circle: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Secondary structure: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Domains: | <input type="radio"/> use <input checked="" type="radio"/> don't use |
| Hydrophobic patch rank: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| Hydrophobic patch size: | <input type="radio"/> use <input checked="" type="radio"/> don't use |
| Temperature factor (B-factor): | <input type="radio"/> use <input checked="" type="radio"/> don't use |
| Water molecules: | <input checked="" type="radio"/> use <input type="radio"/> don't use |
| | |
| Initial probabilities file: | <input type="radio"/> use <input checked="" type="radio"/> don't use |

Please select the output files to be produced:

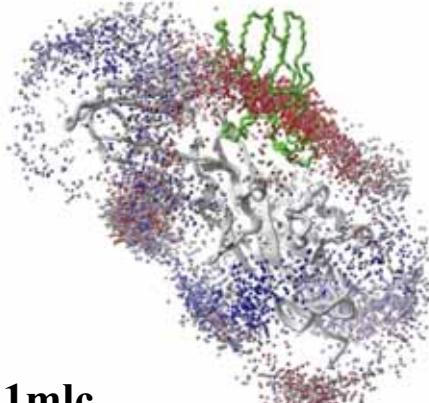
- A text file with the surface dots probabilities
- A pdb file with the AAs colored by their interface probability (full range of colors)
- A pdb file with the surface atoms colored by their interface probability (full range of colors)
- A pdb file with the predicted interface patch colored in red

Interface predictions from docking results

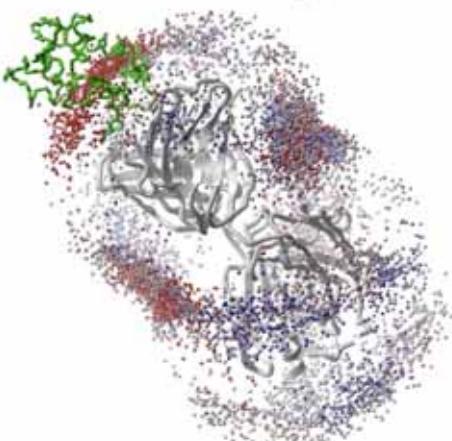
1ca0



2pcf

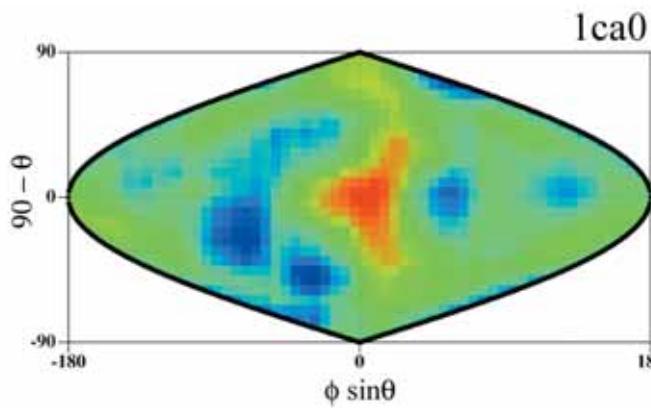
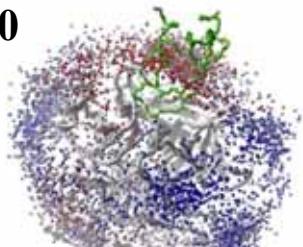


1mlc

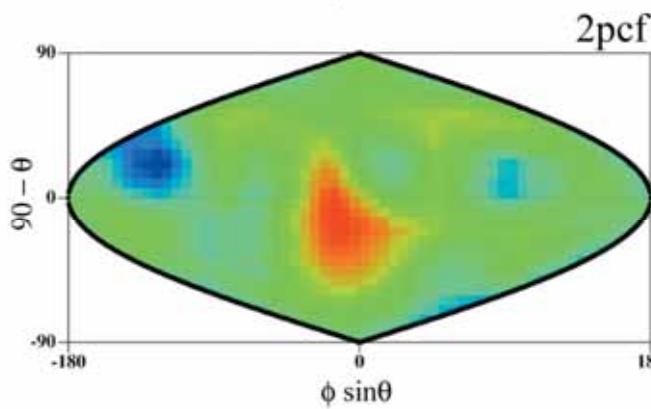
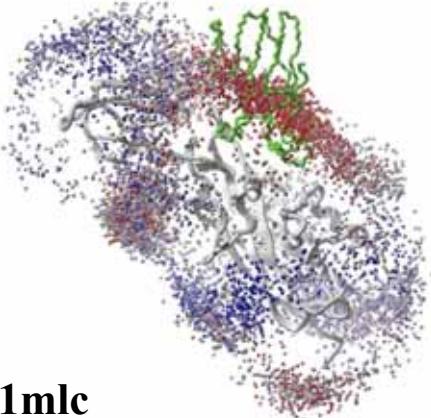


Interface predictions from docking results

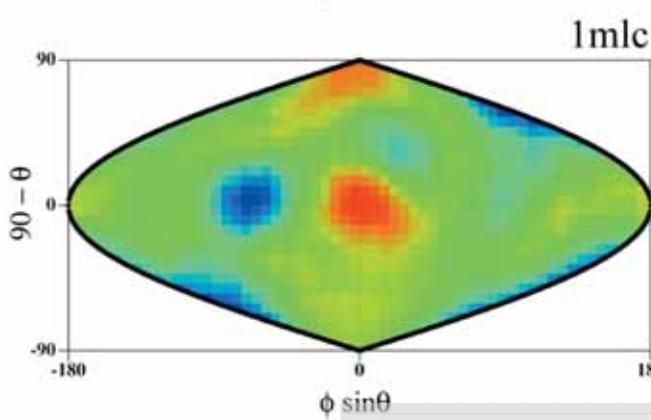
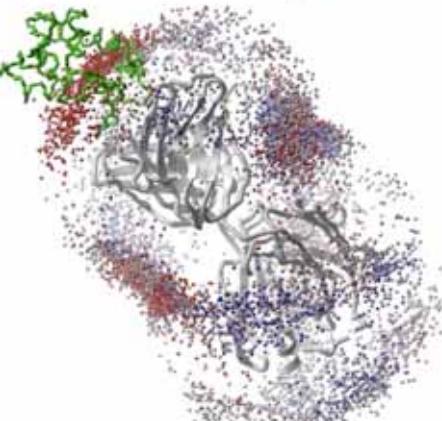
1ca0



2pcf

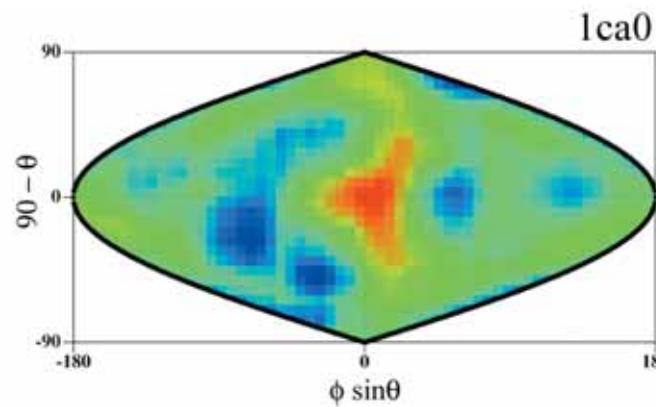
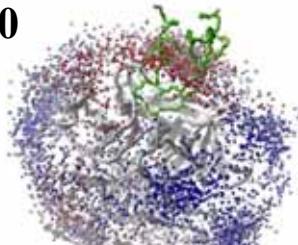


1mlc



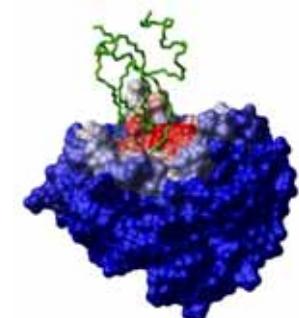
Interface predictions from docking results

1ca0

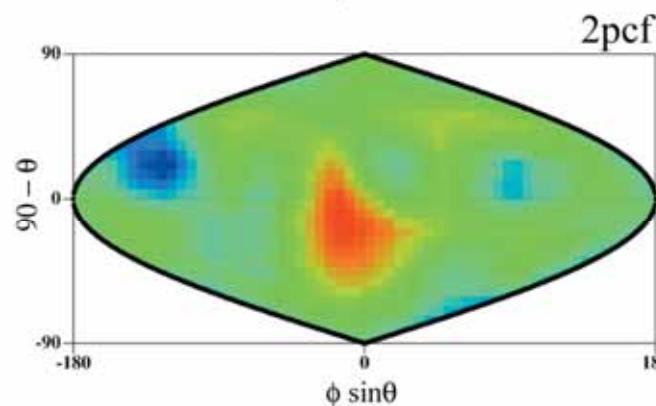
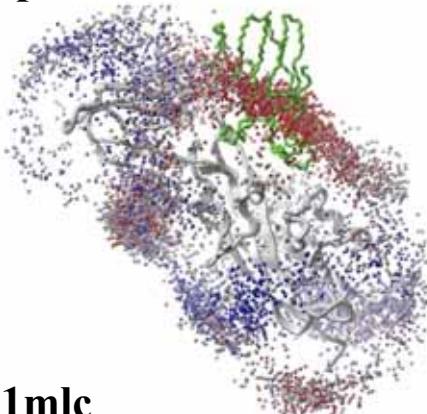


Averaged buried surface
(ABS)

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

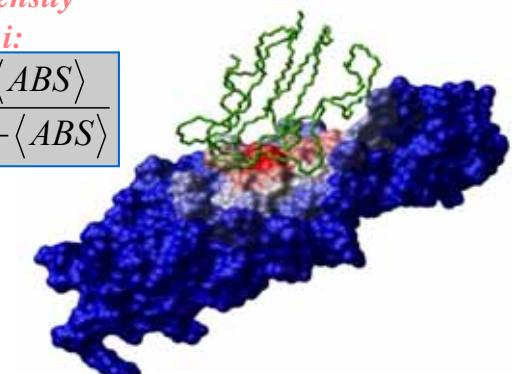


2pcf

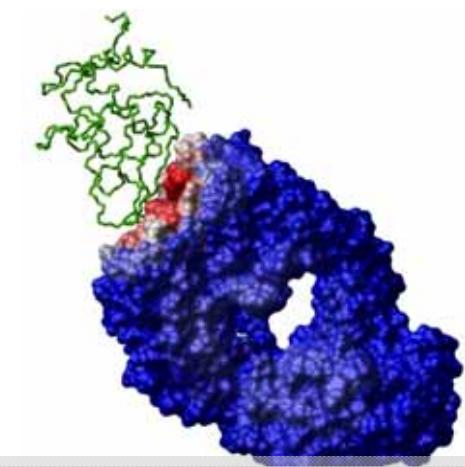
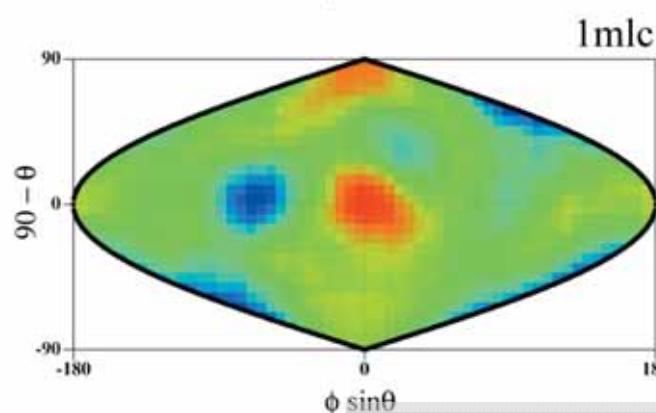
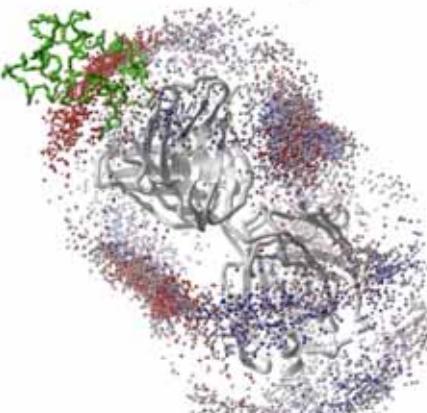


Normalized ABS or
Interface Propensity
for residue i:

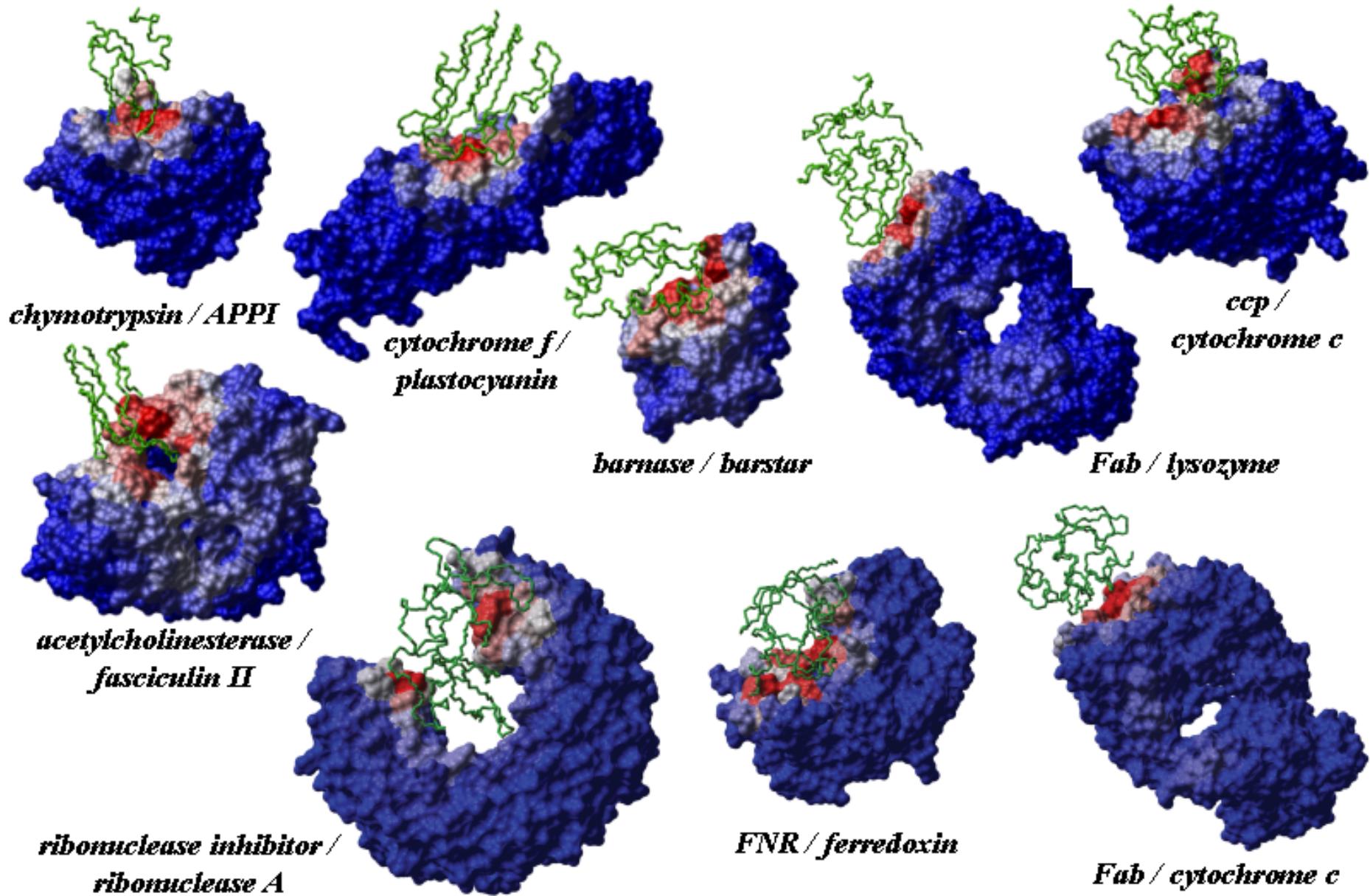
$$NIP_i = \frac{ABS_i - \langle ABS \rangle}{ABS^{MAX} - \langle ABS \rangle}$$



1mlc



Interface predictions from docking results



Protein-protein interface prediction

Method name	Input data	Method	Details	Web
ISIS ⁵⁸	sequence	Neural network	Predicted structural features, evolutionary information	http://cubic.bioc.columbia.edu/services/isis/
TreeDet ⁷¹	sequence, structure	Scoring function	sequence and structural alignments	http://treedetv2.bioinfo.cnio.es/treedet/index.html
Promate ⁷³	structure	Scoring function	Secondary structure, sequence conservation, residue type	http://bioinfo41.weizmann.ac.il/promate/
PINUP ⁷⁶	structure	Scoring function	side-chain energy score, propensity, sequence conservation	http://sparks.informatics.iupui.edu/PINUP/
InterProSurf ⁵⁹	structure	Scoring function	solvent accessibility, propensities	http://curie.utmb.edu/
PRISM ⁷⁷	structure	Scoring function	geometric complementarity, conservation	http://prism.ccbb.ku.edu.tr/prism/
ConSurf ⁶⁸	structure	Scoring function	conservation	http://consurf.tau.ac.il/
ET ⁶⁶	structure	Scoring function	multiple sequence alignments	http://mammoth.bcm.tmc.edu/traceview/
JET ⁷⁰	structure	Scoring function	structural and functional conservation	http://www.ihes.fr/~carbone/data.htm
WHISCY ⁷⁹	structure	Scoring function	conservation, surface properties	http://www.nmr.chem.uu.nl/Software/whiscy/startpage.htm
PIER ⁶¹	structure	Scoring function	atomic statistical propensities	http://abagyan.ucsd.edu/PIER/

Protein-protein interface prediction

Method name	Input data	Method	Details	Web
SiteEngines ⁶⁰	structure	Hierarchical scoring function	structural matching, physico-chemical properties	http://bioinfo3d.cs.tau.ac.il/SiteEngine/
PPI-Pred ⁸⁴	structure	SVM	surface shape, electrostatic potential	http://bioinformatics.leeds.ac.uk/ppi-pred
cons-PPISP ^{80,81}	structure	Neural network	PSI-Blast sequence profile and solvent accessibility	http://pipe.scs.fsu.edu/ppisp.html
SPPIDER ⁸⁵	structure	Neural Network	solvent accessibility and other features	http://sppider.cchmc.org/
Patch Finder Plus ⁸²	structure	Neural Network	conservation, concavity, area, H-bond, residue frequency	http://pfp.technion.ac.il/
meta-PPISP ⁸⁷	structure	Meta web server	cons-PPISP, Promate and PINUP	http://pipe.scs.fsu.edu/meta-ppisp.html
PI ² PE ⁸⁸	structure	Meta web server	cons-PPISP, WESA, DISPLAR	http://pipe.scs.fsu.edu/
SHARP ²⁹⁰	structure	Energy-based, scoring function	Desolvation, hydrophobicity, ASA, propensity, surface shape	http://www.bioinformatics.sussex.ac.uk/SHARP2/sharp2.html
ODA ⁶²	structure	Energy-based	Desolvation energy	http://www.molsoft.com/oda.html
NIP ¹⁰⁰	structure	Energy-based	Docking simulations	https://life.bsc.es/pid/pydock/

- Recognition and prediction of interaction sites
 - **Interface prediction**
 - **Hot-spot identification**
 - **Applications to biomedicine**

Hot-Spots

The **O-ring model**: alanine-scanning **hotspots** tend to be located at the center of the interface and surrounded by energetically unimportant residues that occlude solvent from them.

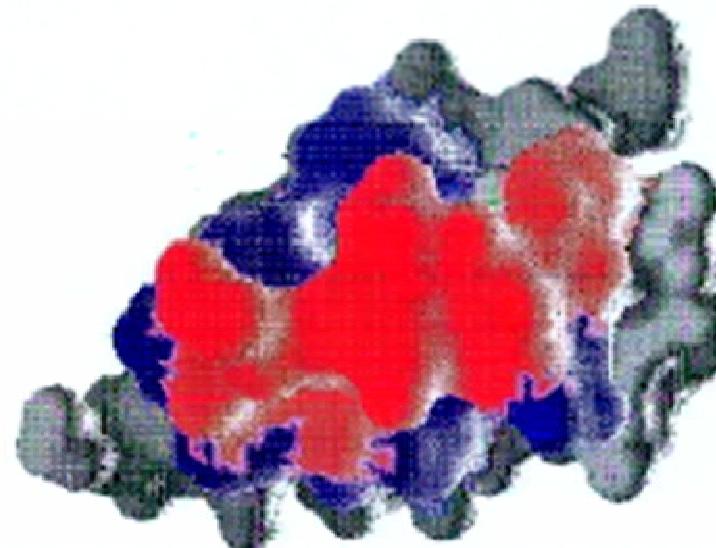
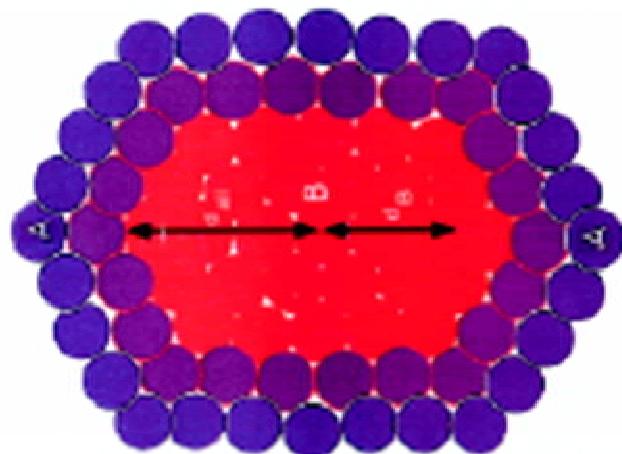
Bogan & Thorn (1998)

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.

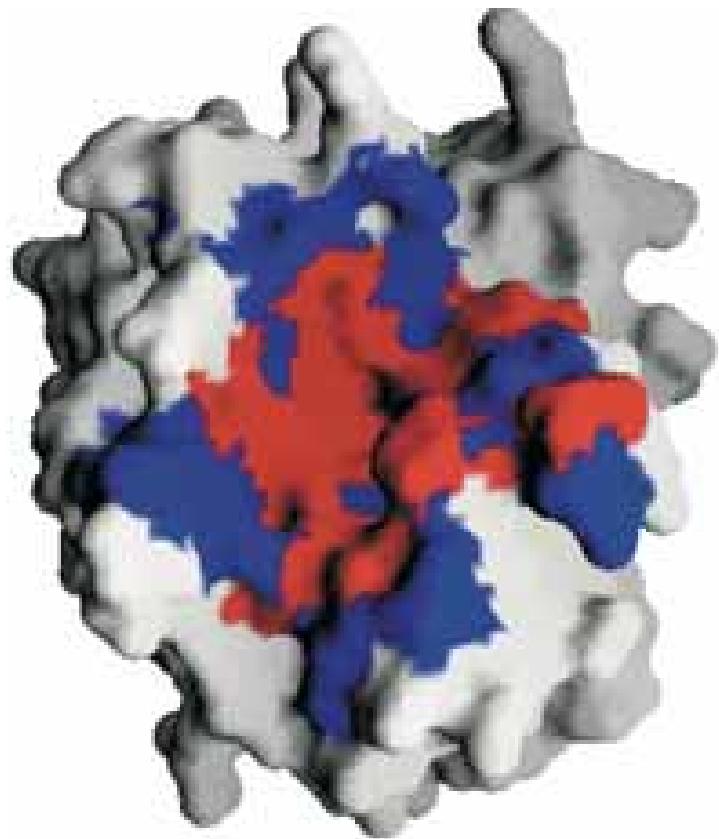
Lo Conte, Chothia & Janin (1999)

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

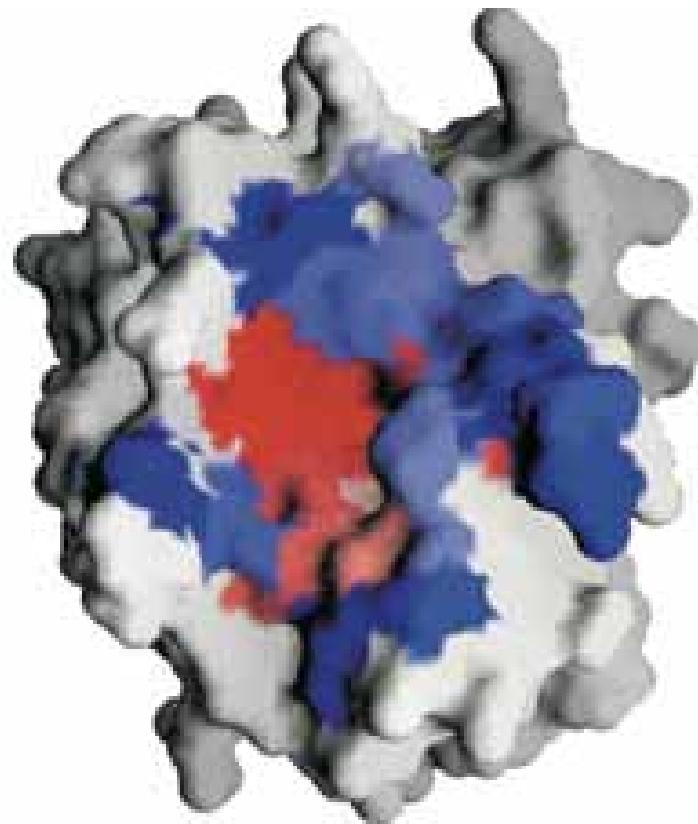
The **core** and **rim** of the C12 inhibitor interface with subtilisin (1sni)



Hot-Spots

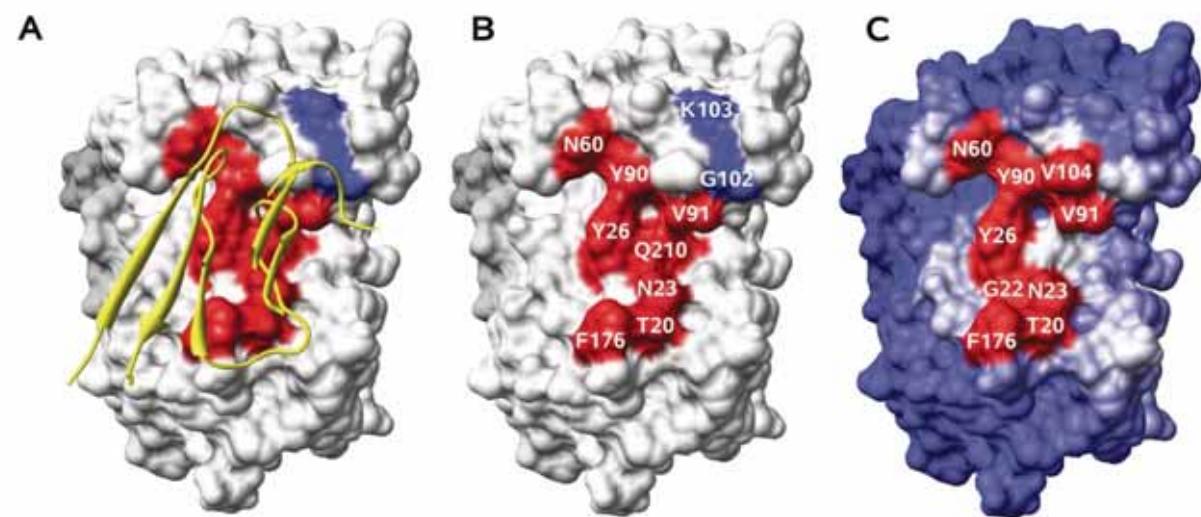
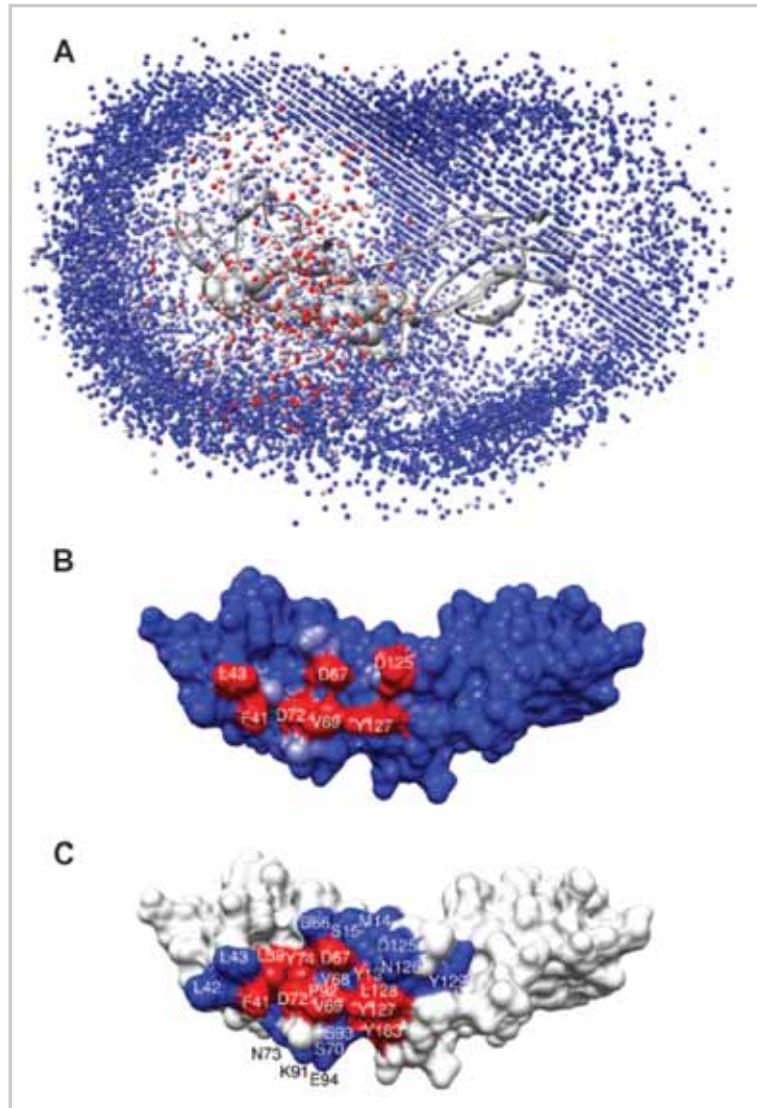


Core and **rim** in the recognition patch of the ovomucoid inhibitor
(1cho)



S=0 (red) fully conserved residues.
S>1 (purple) variable residues

Prediction of Binding Hot-Spots

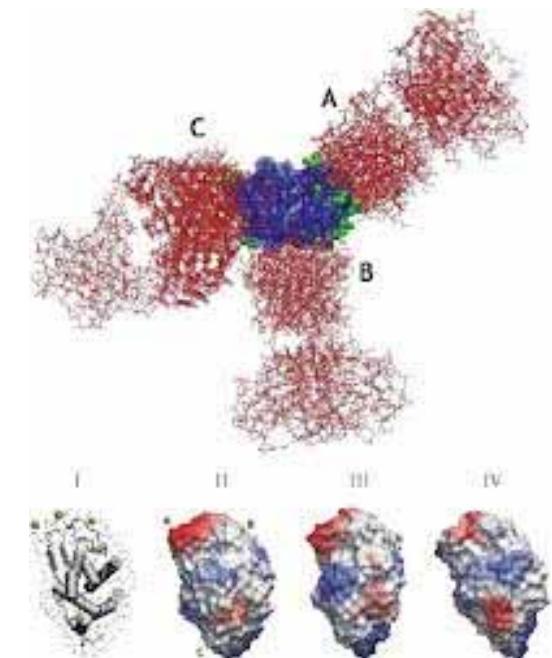
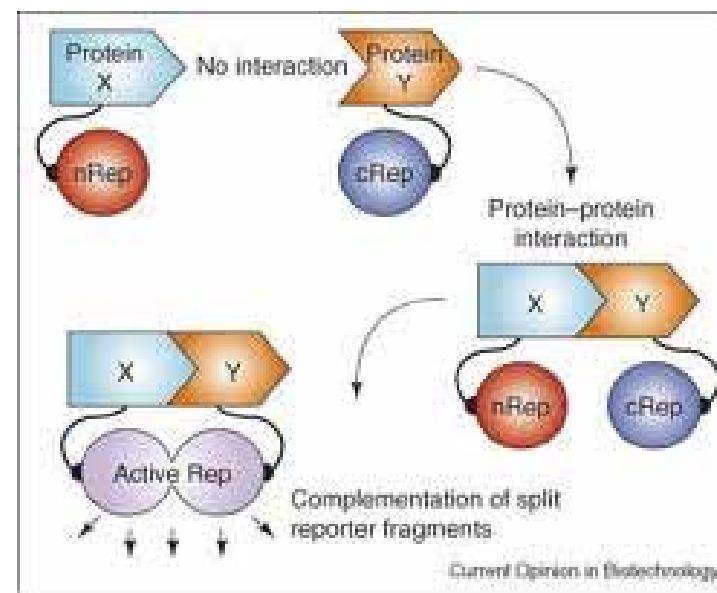
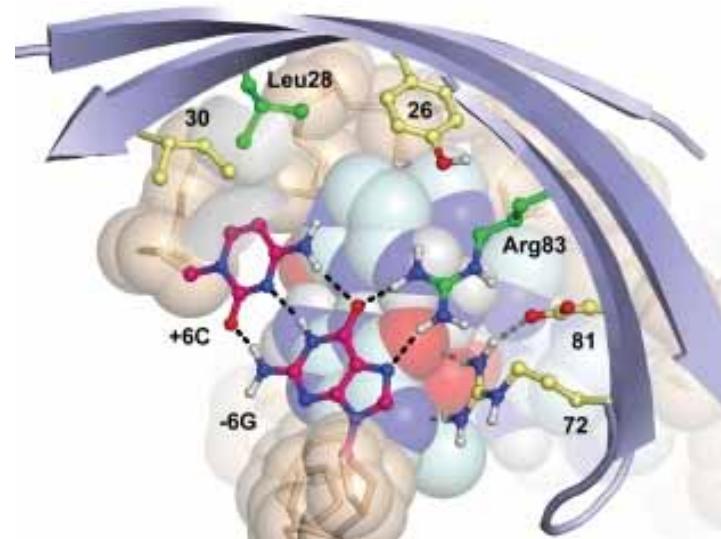
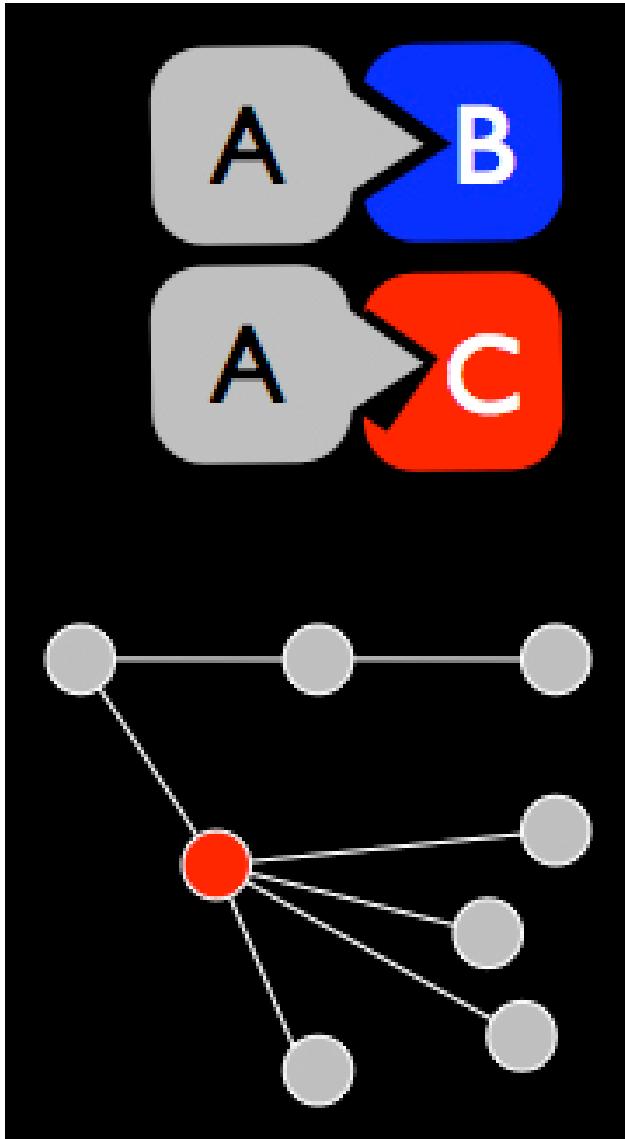


Prediction of Binding Hot-Spots

Method name	Input data	Method	Details	Sensitivity	PPV	availability
ISIS ⁵⁸	sequence	Neural network	Predicted structural features, evolutionary information	15%	89%	http://cubic.bioc.columbia.edu/services/isis/
FOLDEF ¹²⁶	complex structure	Energy-based	Alanine scanning	45-72% ^a	61-73% ^a	http://foldx.crg.es/
ROBETTA ¹²²	complex structure	Energy-based	Alanine scanning	28-69% ^b	60-71% ^b	http://robetta.org/submit.jsp
K-FADE ¹²⁵ / K-CON/ ROBETTA	complex structure	Machine learning algorithm	Physical-biochemical features	48%	53%	http://kfc.mitchell-lab.org
MAPPIS ¹¹⁹	complex structure	Evolutionary conservation	Multiple alignments, 3D clustering	66%	63%	http://bioinfo3d.cs.tau.ac.il/MAPPIS
HotPoint ¹²¹	complex structure	Empirical model	Accessibility, knowledge-based potentials	59%	70%	http://prism.ccbb.ku.edu.tr/hotpoint
pyDockNIP ¹²⁷	unbound protein structure	Energy-based	Docking simulations	42-43%	68-75%	http://mmb.pcb.ub.es/PyDock

- Recognition and prediction of interaction sites
 - **Interface prediction**
 - **Hot-spot identification**
 - **Applications to biomedicine**

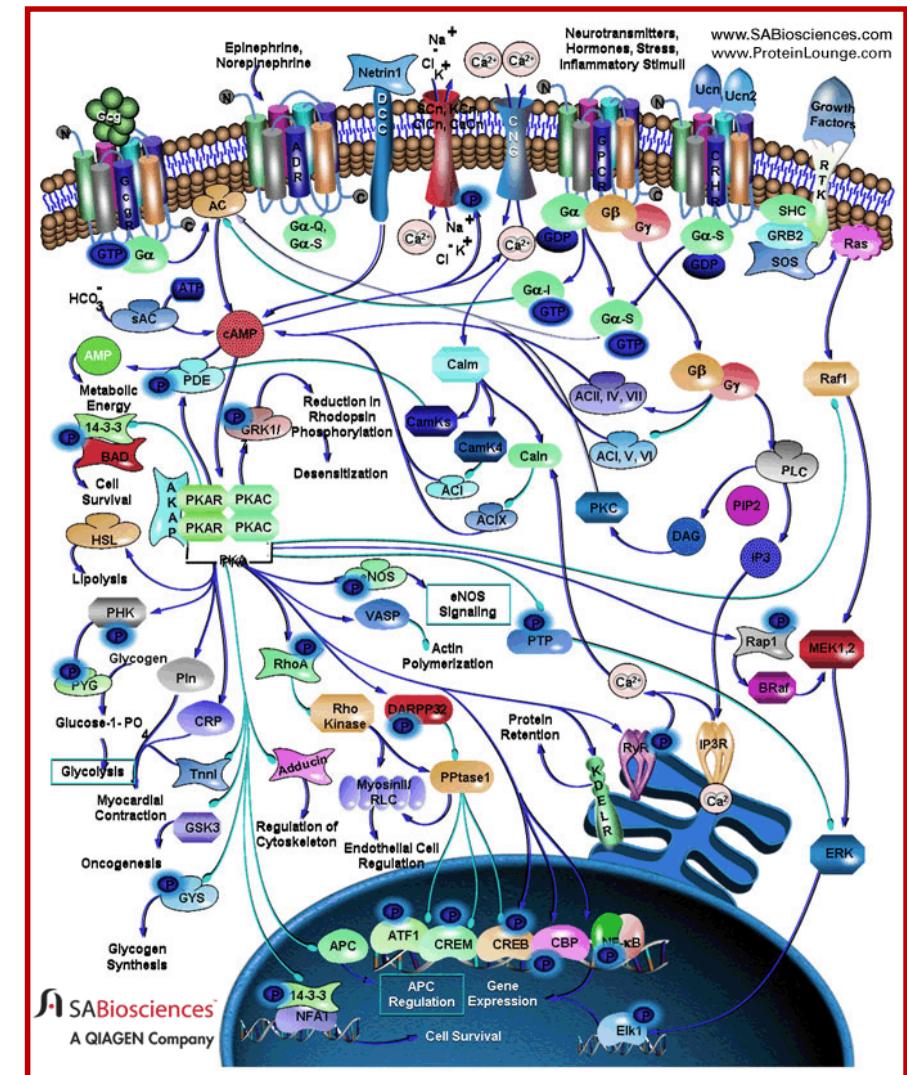
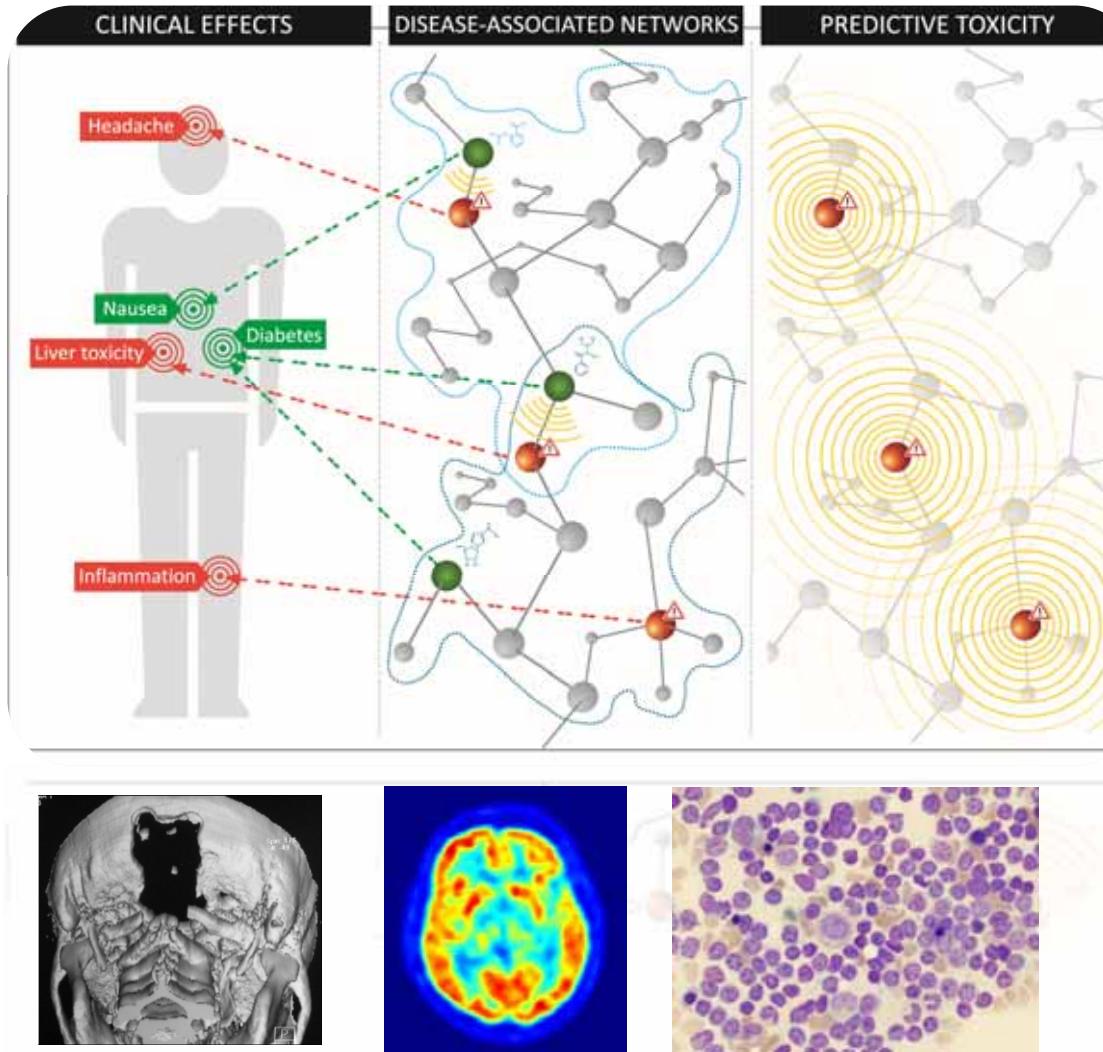
Applications in Biotechnology: Engineering of protein-protein interfaces



Current Opinion in Biotechnology

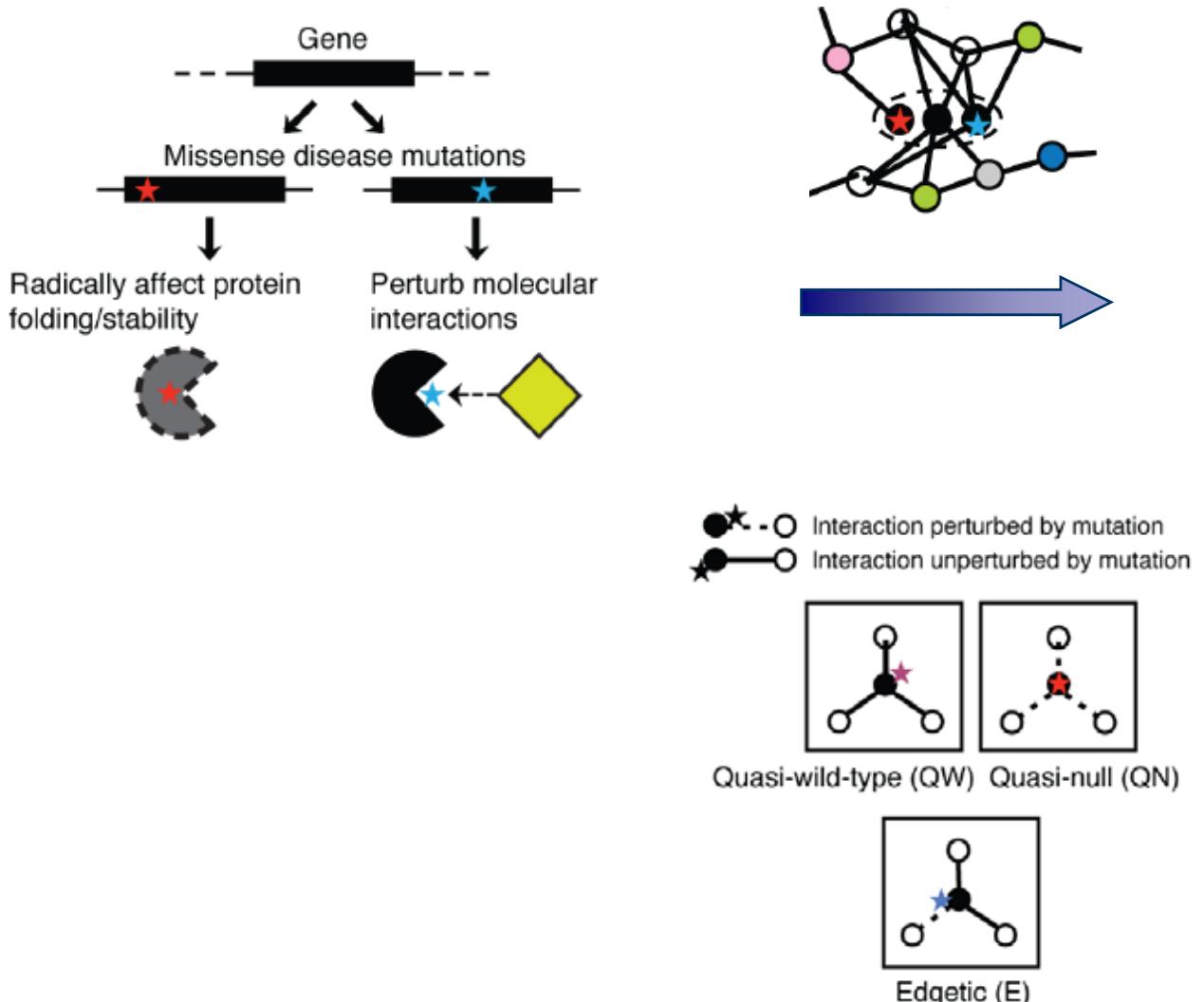
Applications in Biomedicine: Prediction and interpretation of pathological mutations

Effect of pathological mutations on PPIs and networks

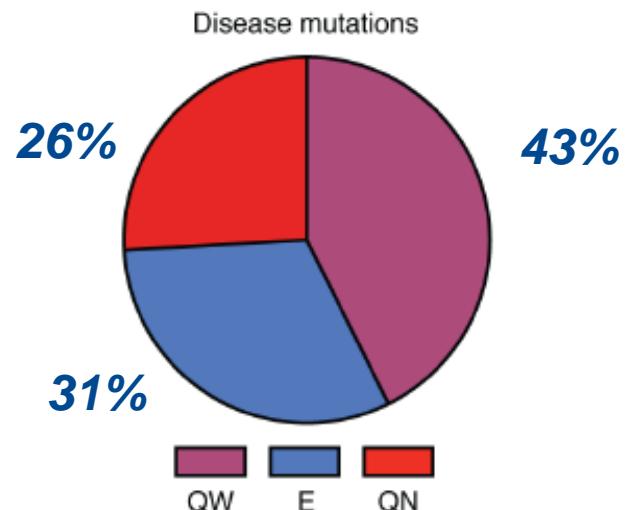


Applications in Biomedicine: Prediction and interpretation of pathological mutations

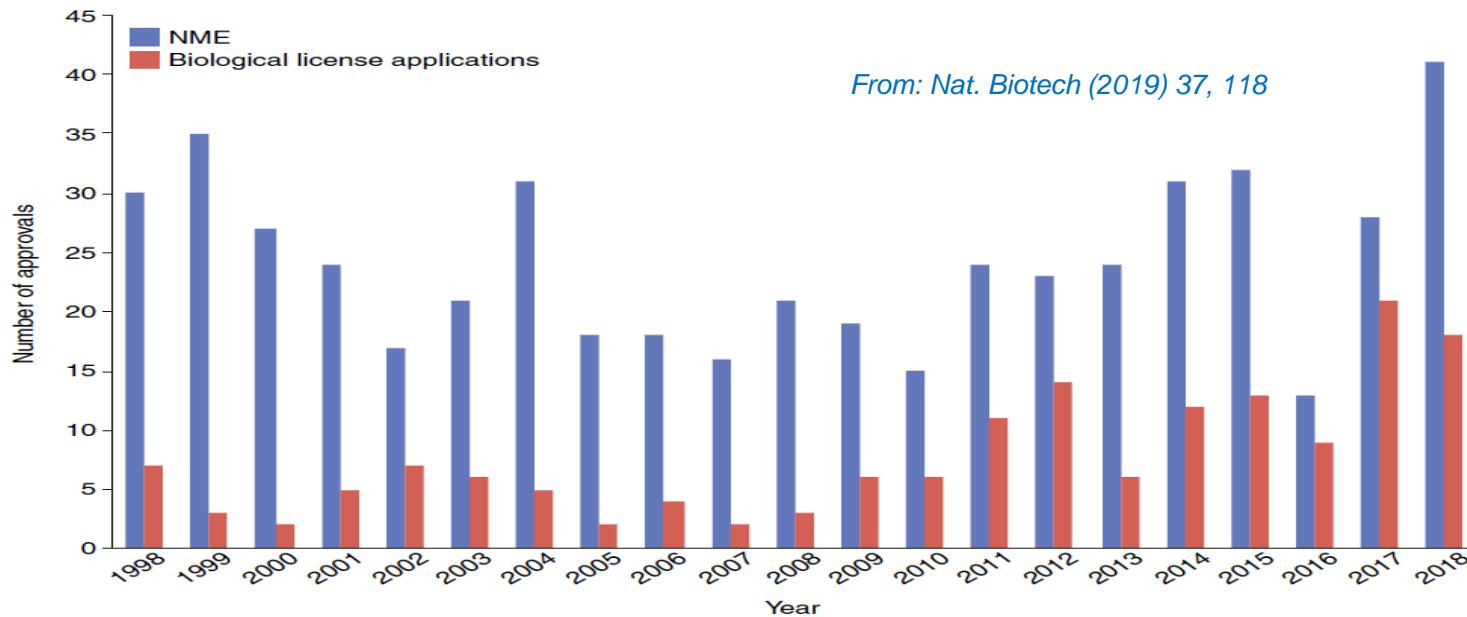
Effect of pathological mutations on PPIs and networks



n=197 mutations
89 proteins with > 1 interaction



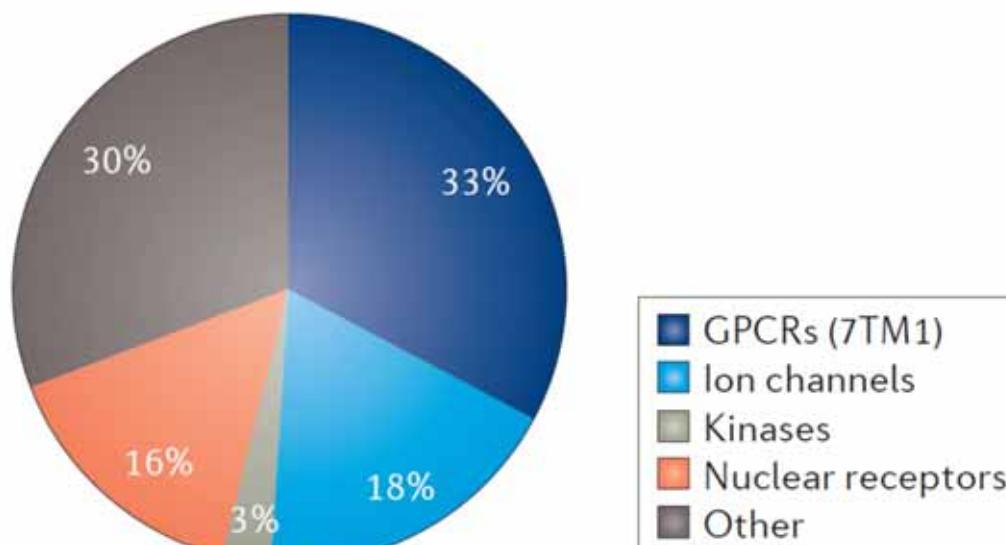
Drug Discovery



From: *Nat. Biotech* (2019) 37, 118

Drug discovery: new targets are strongly needed!

Proportion of small-molecule drugs that target major families



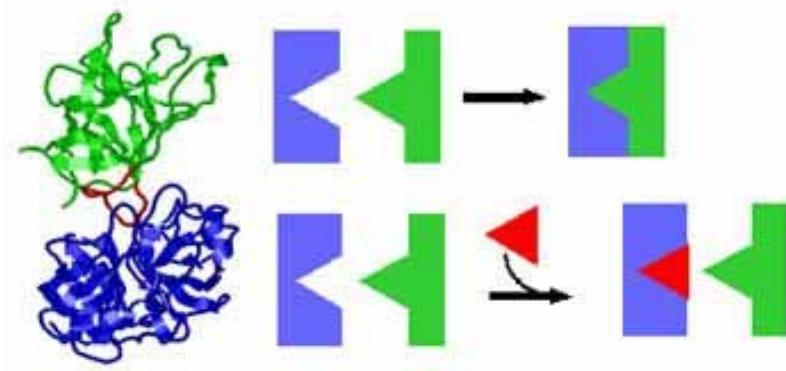
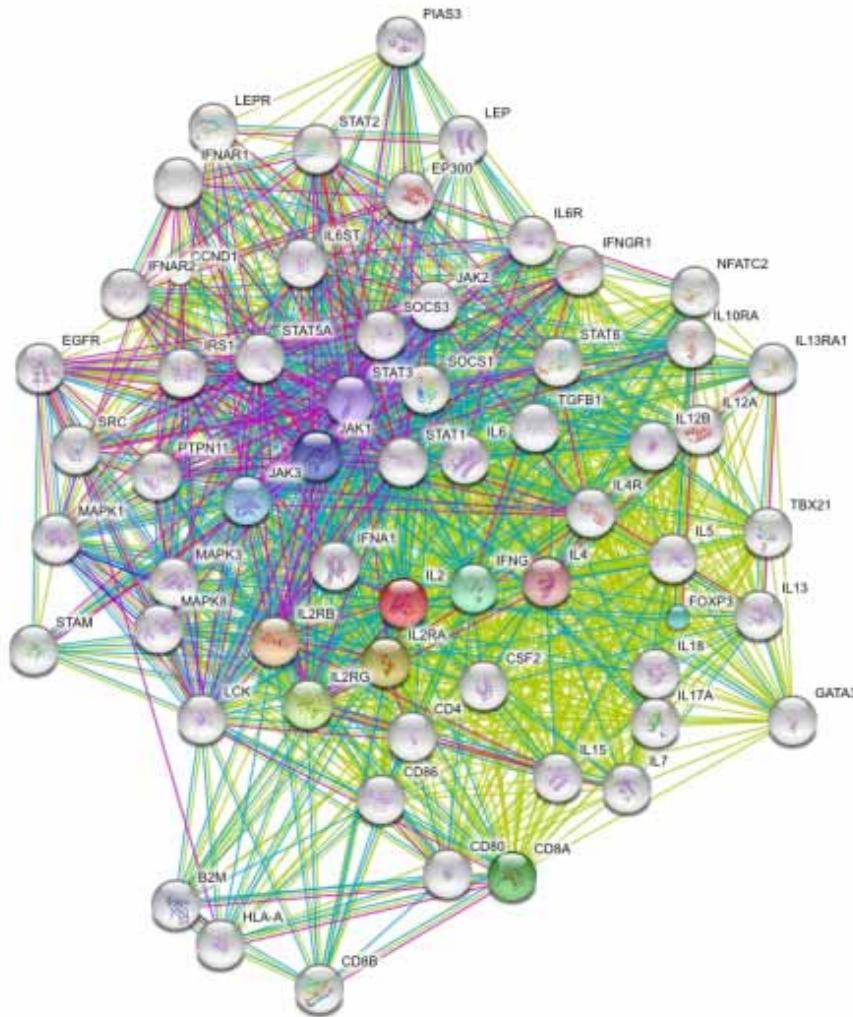
From: *Nat. Rev. Drug Disc.* (2017) 16, 19

>70% FDA-approved drugs target only 4 gene families, basically cell-surface receptors and enzymes

Most of the available drugs target individual proteins

Only 10–15% of all human proteins are currently «druggable»

Applications in Biomedicine: Drug discovery targeting protein-protein interactions

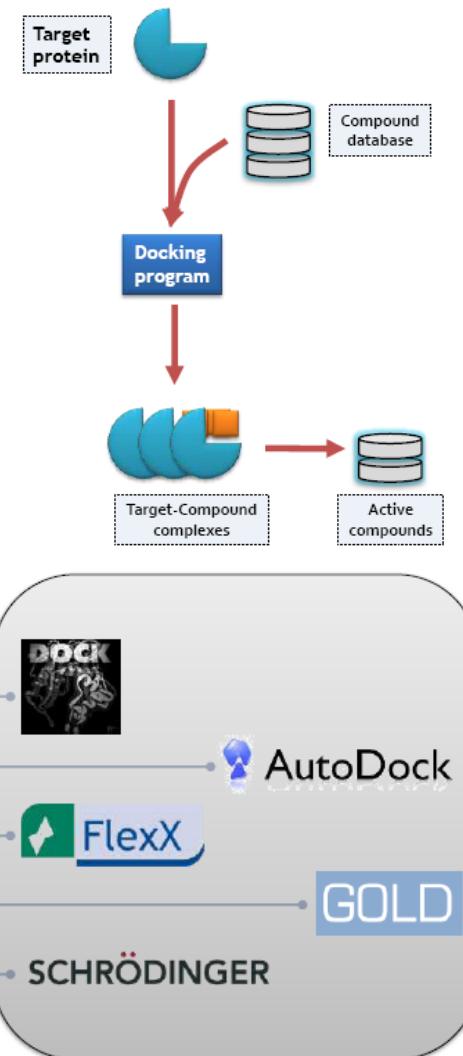
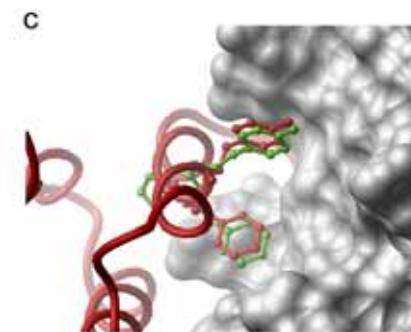
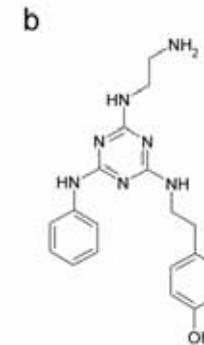
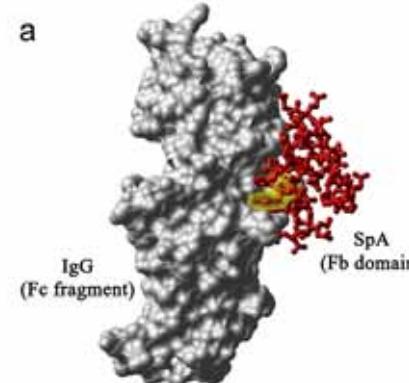
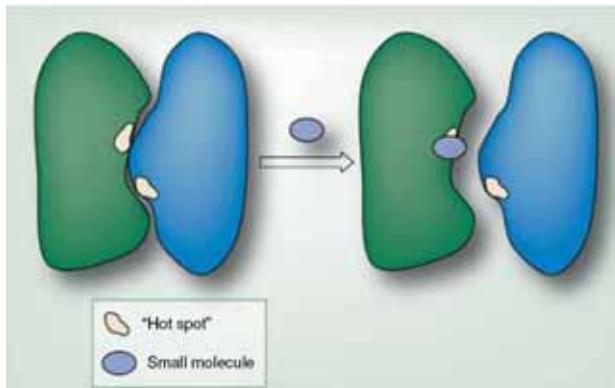
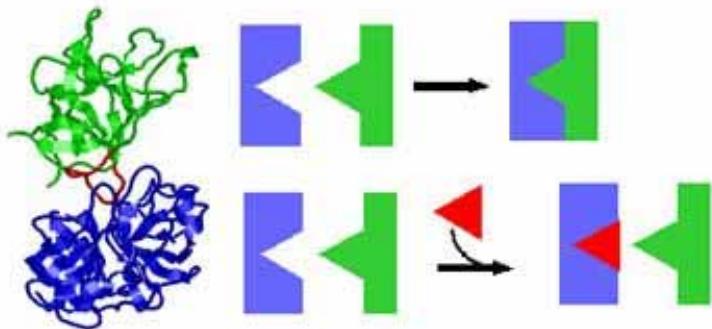


New approaches:
drugs targeting protein-protein
interactions

Applications in Biomedicine: Drug discovery targeting protein-protein interactions

Structure-based drug design, PPI inhibitors

Inhibition of protein-protein interactions



Applications in Biomedicine: Drug discovery targeting protein-protein interactions

Protein-protein docking

Interface and hot-spot predictions

