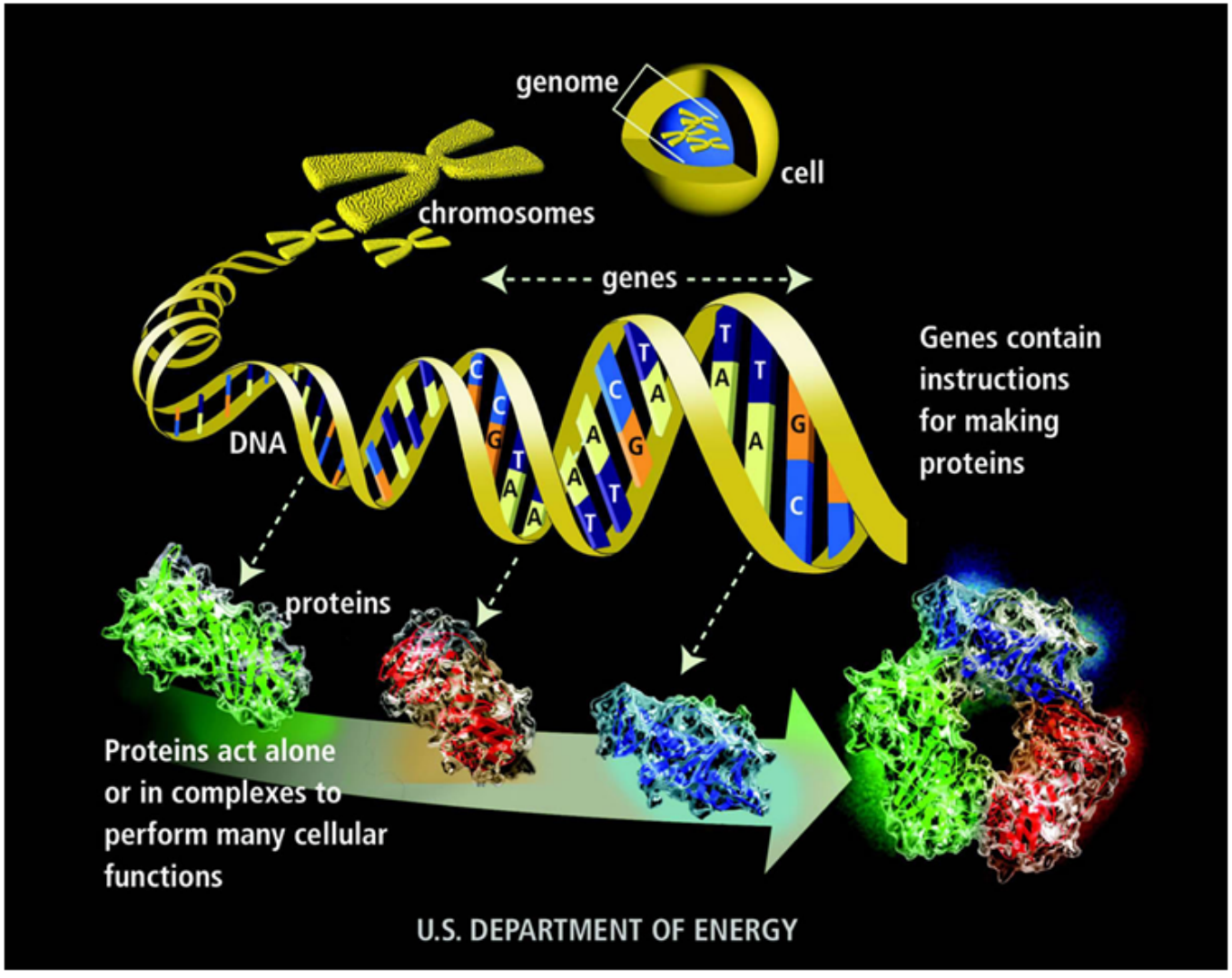
Protein-protein docking
Arrimage protéine-protéine

Dirk Stratmann

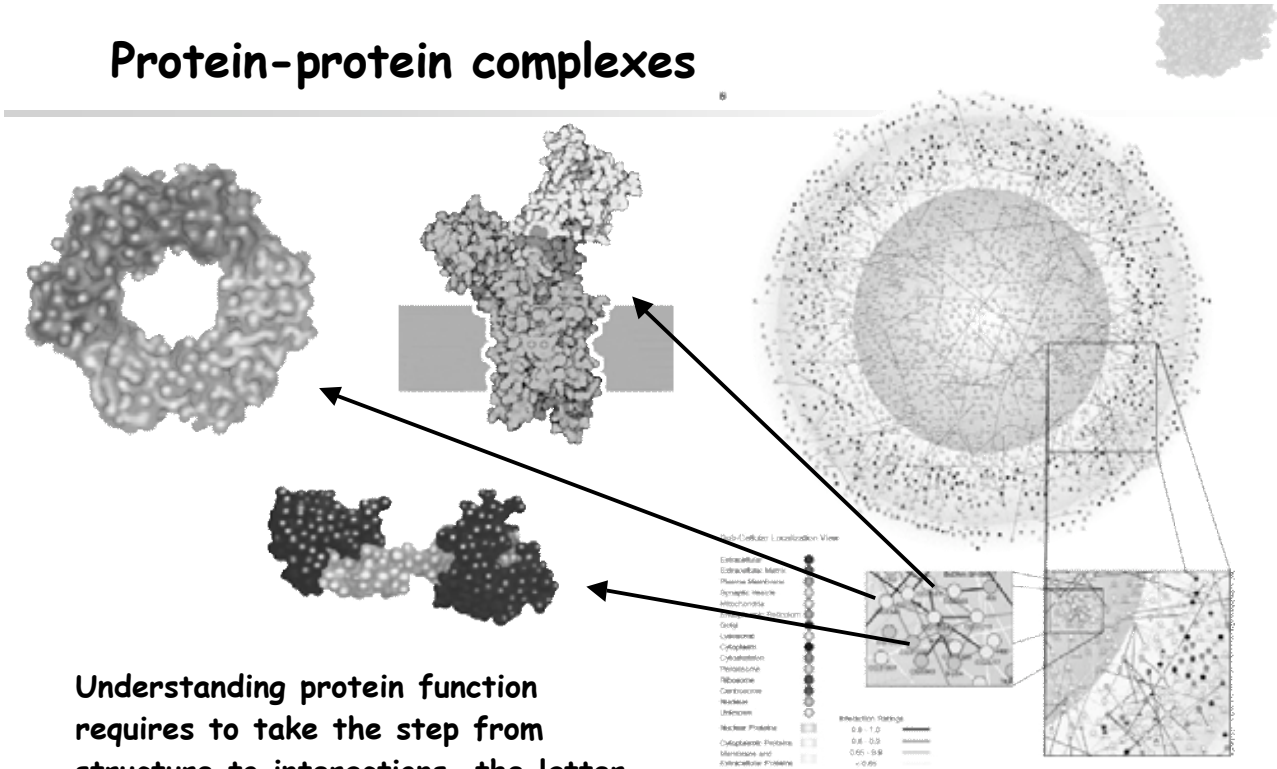
Master M2 BIM (NV 634, Biopol), Université Pierre et Marie Curie (Paris VI)

03 décembre 2012



https://public.ornl.gov/site/gallery/detail.cfm?id=403

Protein function



Source: Bonvin

Free proteins - Structural genomics

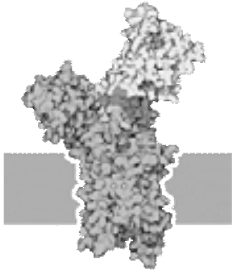
- 3D structure of a large number of unbound/free proteins solved => PDB
- Only about 1000 types of folds, almost all known.
- => Comparative modeling / Homology modeling

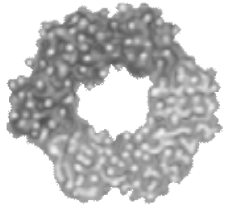
Protein-protein complexes

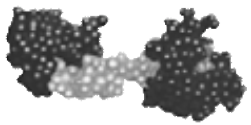
- Number of types of protein-protein interactions at least 10x times greater (> 10.000) than number of folds (1000).
- Experimental difficulties to solve protein-protein 3D structures.

Models of Protein Complexes

What can we learn from 3D structures (models) of complexes?







- **Models provide structural insight into function and mechanism of action**
- **Models can drive and guide experimental studies**
- **Models can help understand and rationalize the effect of disease-related mutations**
- **Models provide a starting point for drug design**

AB/10-07

Source: Bonvin

Protein-docking problem

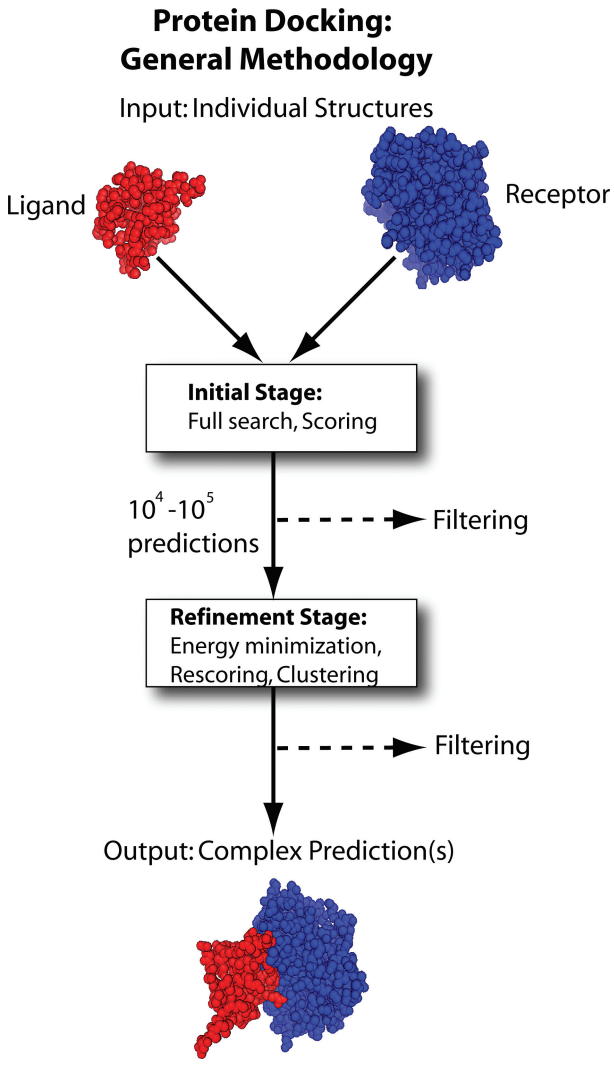
- Connolly [Connolly, 1986] has posed the protein-docking problem as: "Given the structures of any two proteins, is it possible to predict whether they associate, and if so, in what way?"
- Connolly was very optimistic at that time:
  "With a few years more development they stand a good chance of solving the protein-docking problem. If the protein-docking problem cannot be solved by a purely geometric approach, there remains the option of bringing in chemical considerations."
- The problem of docking molecules of any complexity based on the complementarity of their features has been shown to be NP-complete (Kuhl et al., 1984).

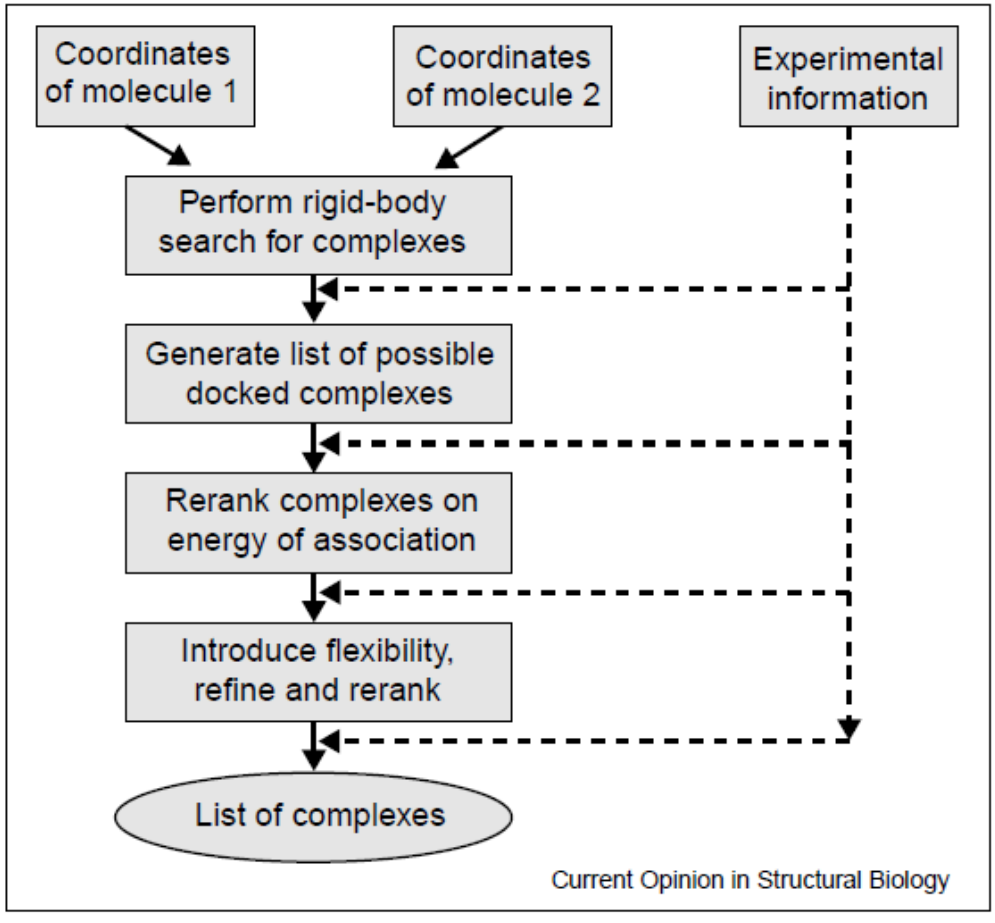
Representation, Sampling and Scoring

Three key ingredients:

- *Representation of the system*
- *Global conformational space search*
- *Reranking of top solutions based on scoring function*

Similar steps as for protein folding
Reviews: [Smith and Sternberg, 2002], [Halperin et al., 2002]



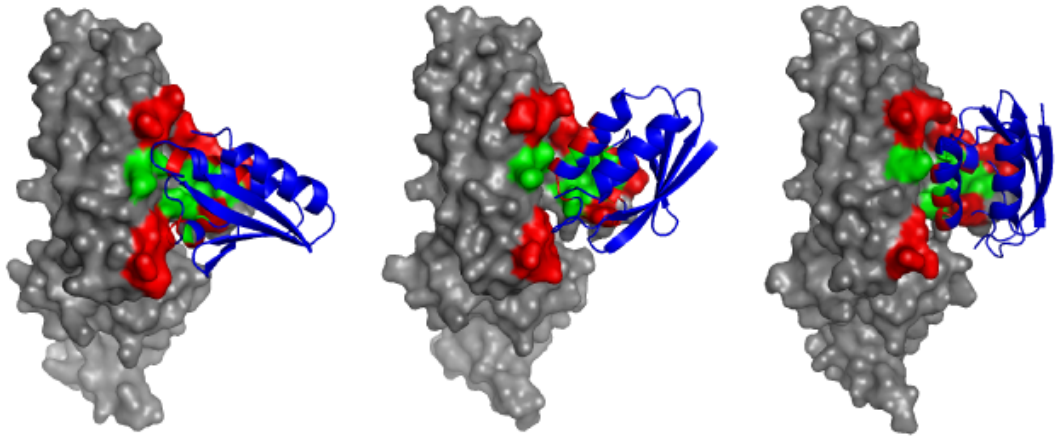


[Smith and Sternberg, 2002]

Steps of protein-protein docking

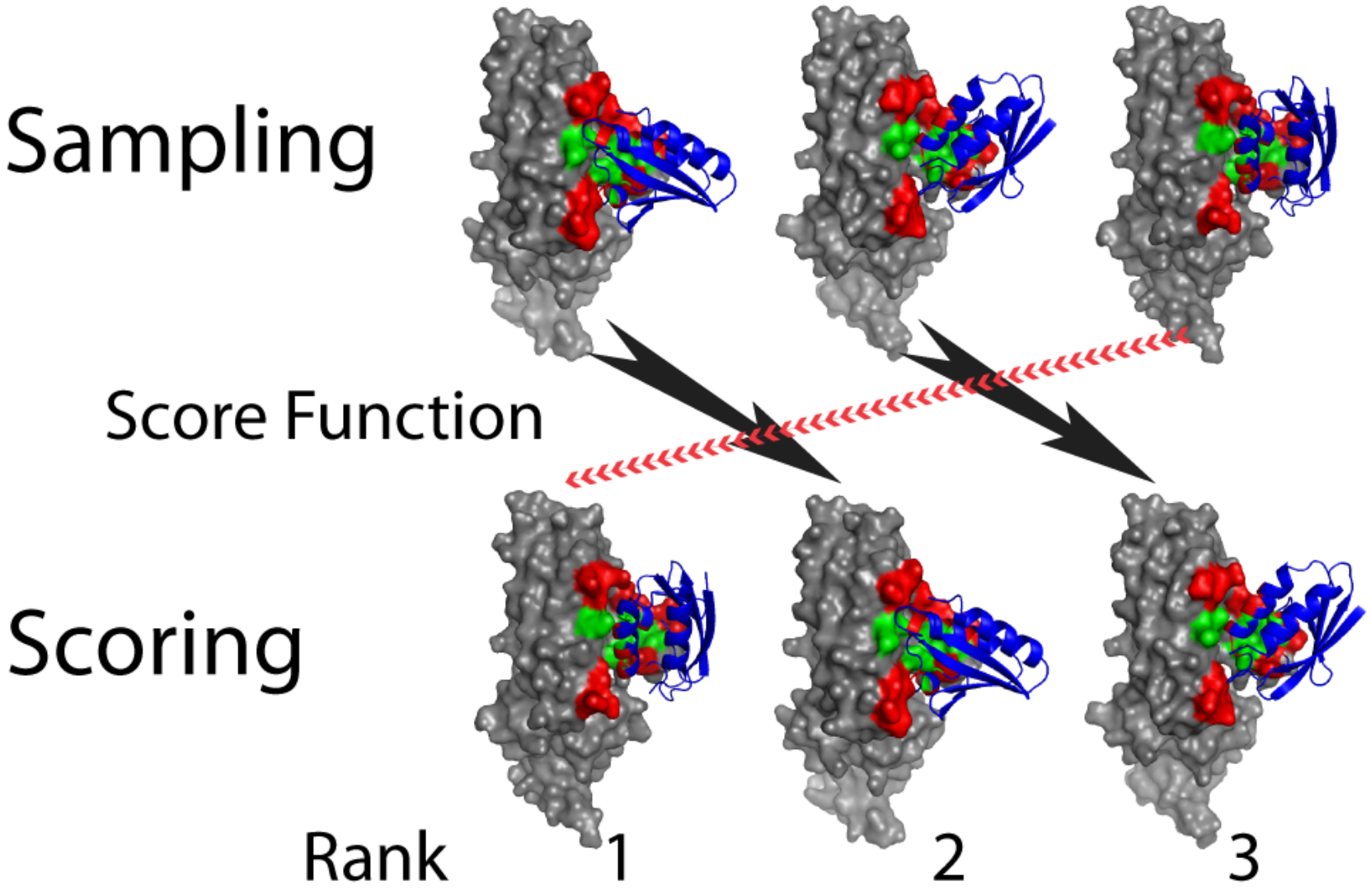
Sampling and Scoring

Sampling

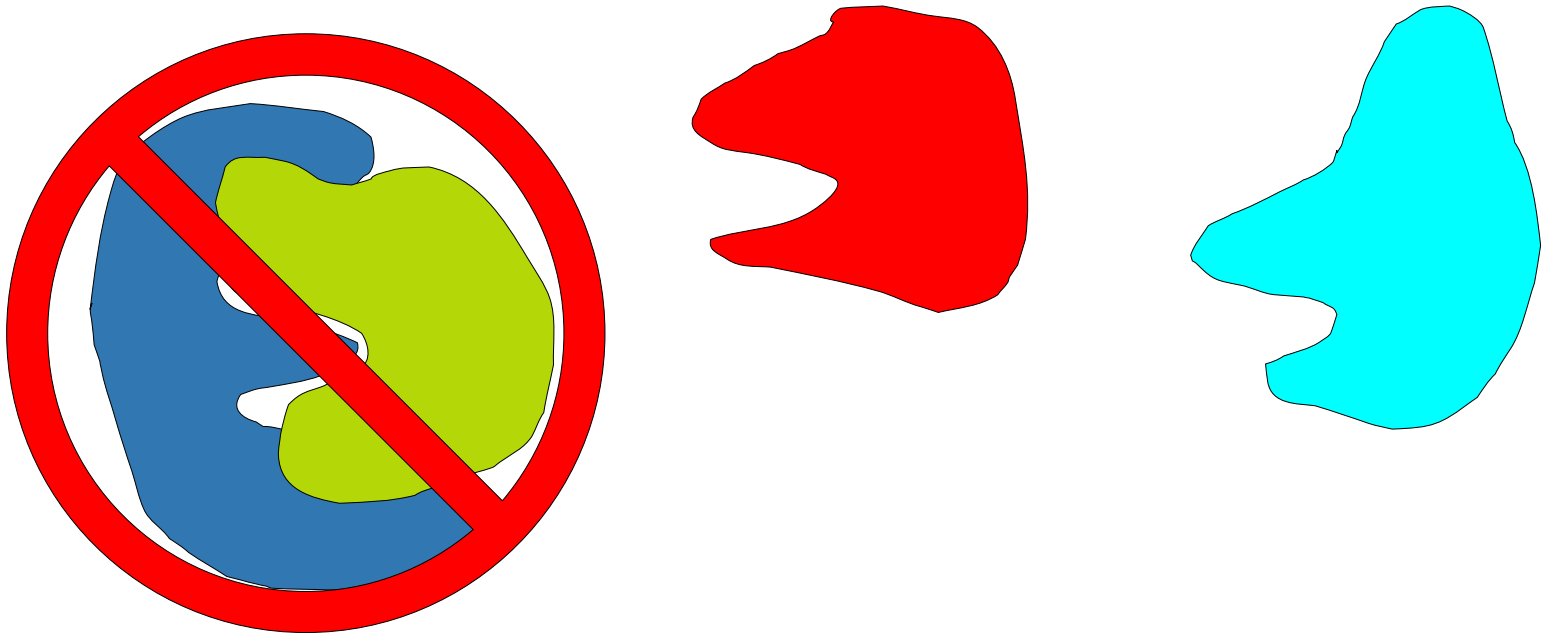


Steps of protein-protein docking

Sampling and Scoring

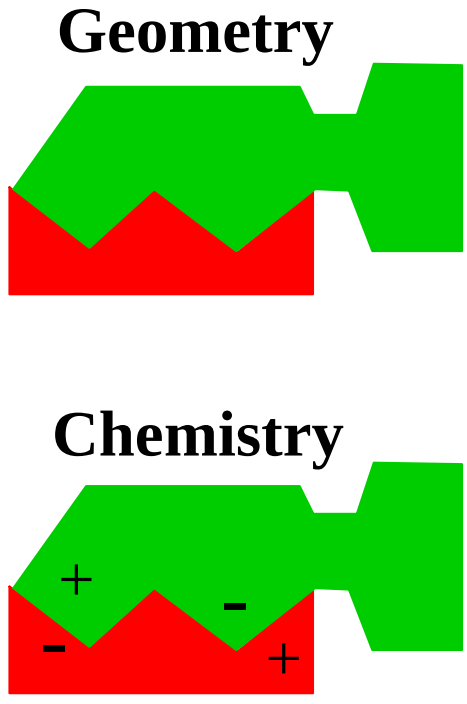


Lock and Key

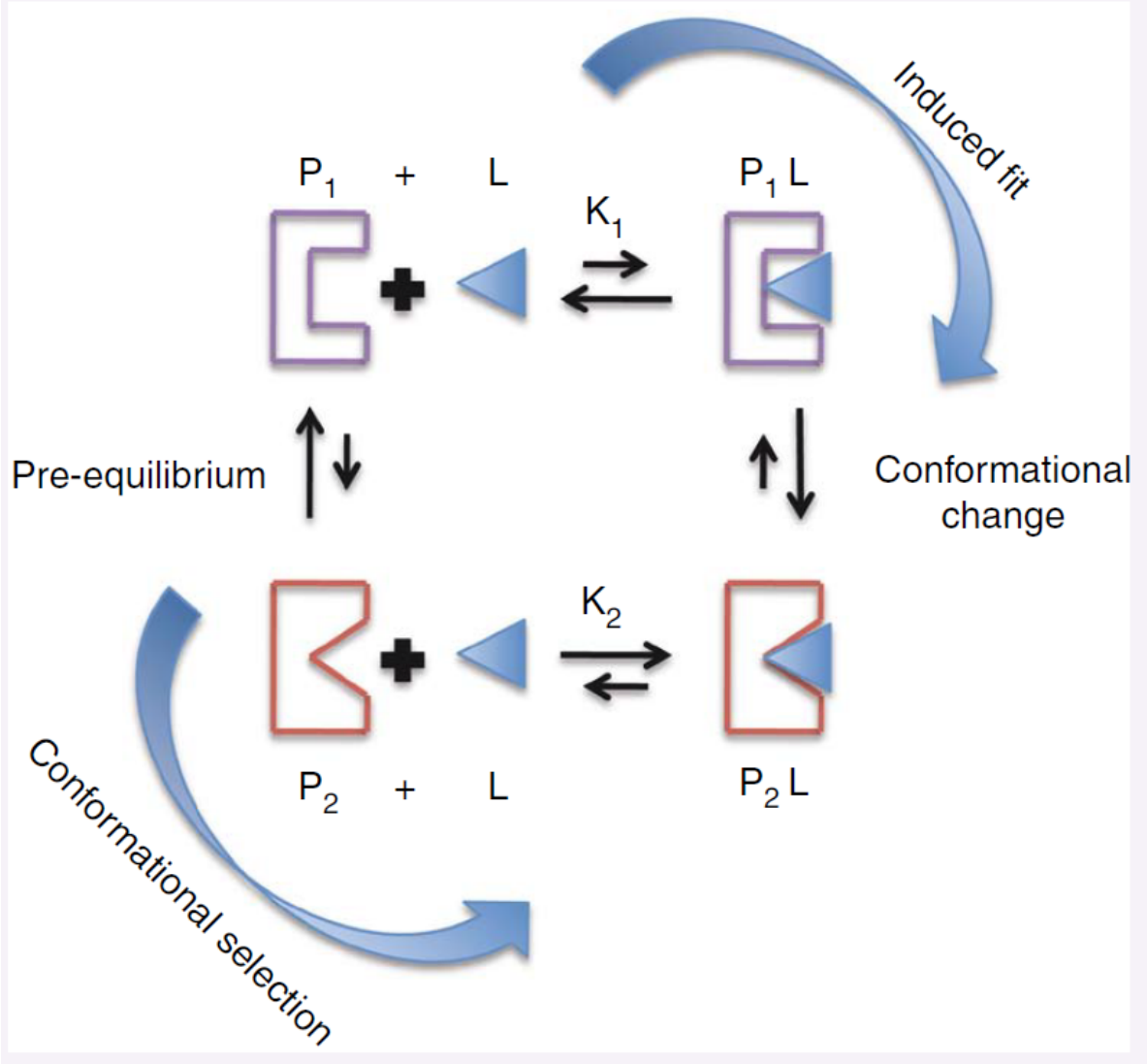


Source: Kohlbacher and Lenhof

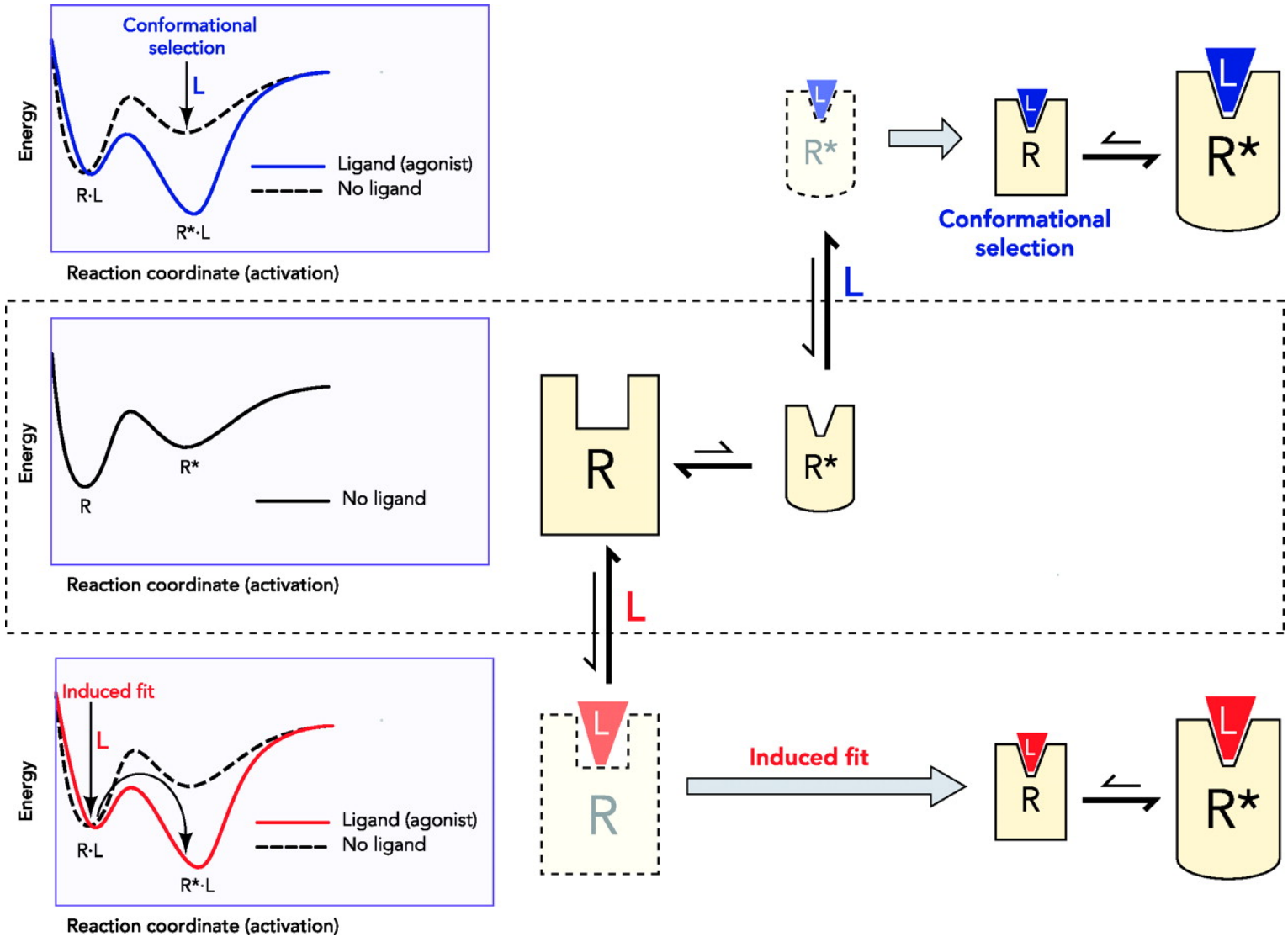
Lock and Key



Source: Kohlbacher and Lenhof



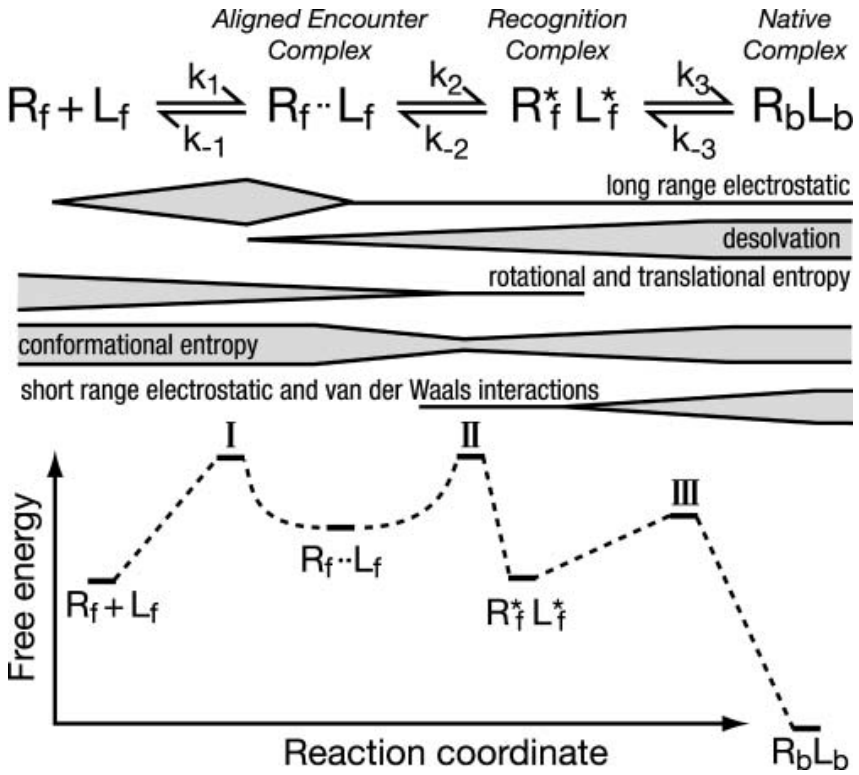
[Boehr et al., 2009]



[Deupi and Kobilka, 2010]

Flexible Protein Recognition

3-step mechanism of diffusion, free conformer selection, and refolding:



[Grünberg et al., 2004]

Enzyme / Inhibitor

Enzymes and their inhibitors have co-evolved to form an interface with a high degree of surface complementarity

Antibody / Antigen

The immune system produces many different antibodies in response to an antigen, some of which bind their respective epitopes quite well while others bind quite poorly.
Antibody => always the same binding site location Antigen => Highly variable binding site locations

Protein-Protein Docking Benchmark 4.0

http://zlab.umassmed.edu/benchmark/
PDB => 1667 complex structures with unbound structures =>
109 non-redundant complexes (according to SCOP families) =>
176 unbound-unbound cases with reference complex structure

Table II
Statistics of the Three Classes of Difficulty in the Entire Benchmark 4.0 and the New Cases (in Parentheses)

	I-RMSD	f_{nat}	$f_{non\text{-}nat}$	Number
Rigid body	0.90 (1.12)	0.79 (0.80)	0.21 (0.19)	121 (33)
Medium	1.76 (1.86)	0.63 (0.66)	0.35 (0.27)	30 (11)
Difficult	3.76 (3.45)	0.51 (0.60)	0.51 (0.41)	25 (8)

52 enzyme-inhibitor, 25 antibody-antigen, 99 other functions

[Hwang et al., Proteins 2010]

Introduction

- What distinguishes the true complex structure from "false positives"?
- *Physical chemistry:* Complex structure with the lowest binding free energy is the one observed in nature.
- *Caveat:* relies on sufficiently complete sampling of conformation space

Prediction of Binding Free Energy

- Currently very difficult
- Would need to include entropic contributions and solvent effects
- Free energy prediction is also very difficult in:
  - Protein-ligand docking
  - Protein structure prediction

Prediction of Binding Free Energy

$$\Delta G_{binding} = \Delta G_{elec} + \Delta E_{vdW} + \Delta G_{des} + \Delta E_{int} - T\Delta S_{sc} - T\Delta S_{bb} \quad (1)$$

ΔG_{elec} electrostatic, ΔE_{vdW} van der Waals, ΔG_{des} desolvation, ΔE_{int} conformational changes upon binding
$-T\Delta S_{sc}$ and $-T\Delta S_{bb}$ entropy changes from side chain and backbone, respectively.

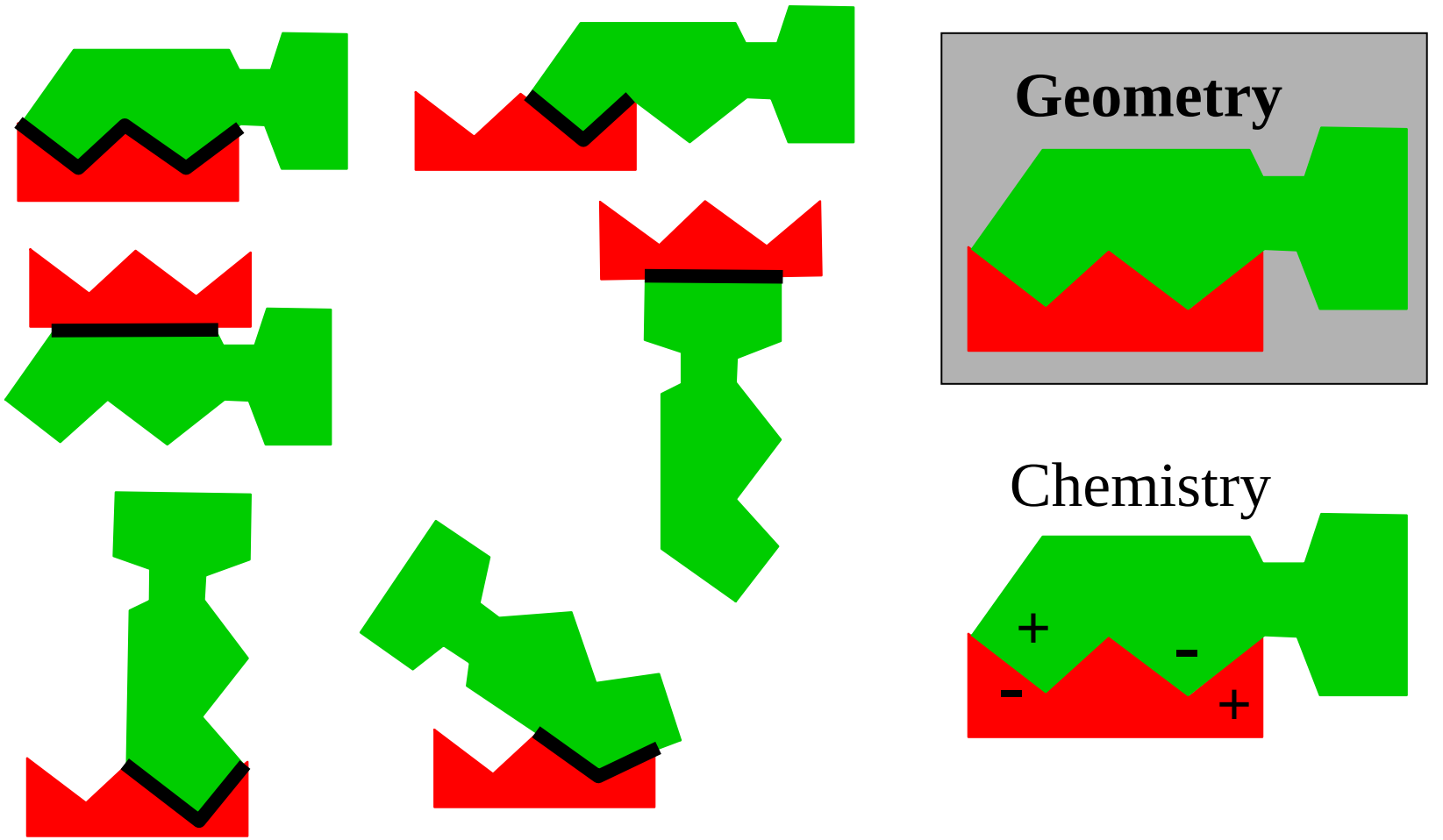
[Pierce and Weng, 2007]

Alternative: Scoring Functions

- Geometry:
  - Lock and key principle
  - Large contact areas are favorable
  - Steric clashes / overlaps should be avoided
- Chemistry:
  - Models based on physicochemistry
  - Compromise between speed and accuracy

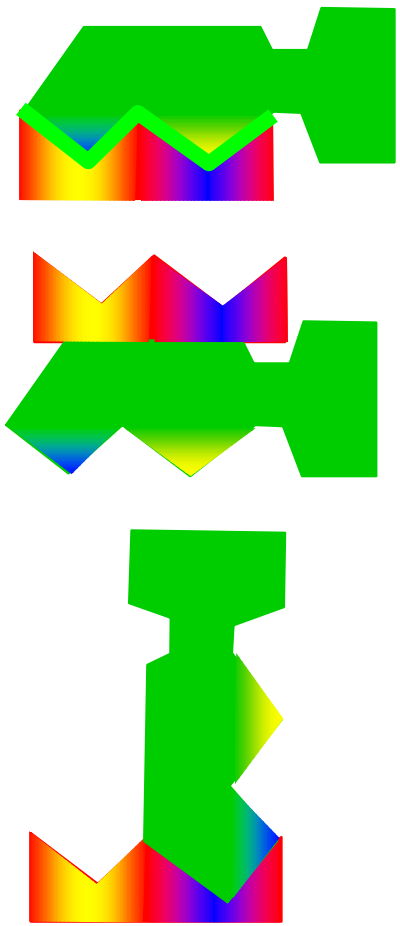
Scoring functions must be accurate and fast at the same time to evaluate serval billions of docking poses.
Scoring functions based only on geometry or only on chemistry are not successful in general.

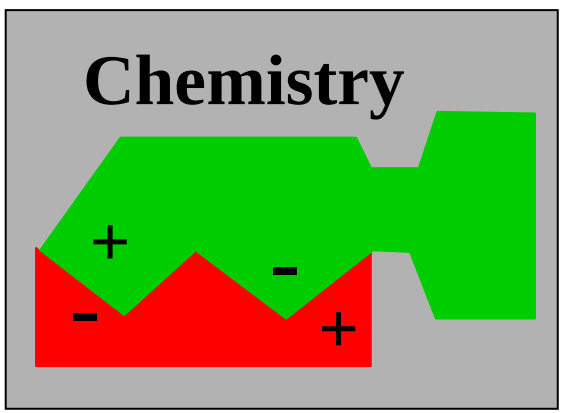
Geometry and Chemistry



Source: Kohlbacher and Lenhof

Geometry and Chemistry





Source: Kohlbacher and Lenhof

Geometry

1. Steric complementarity of shapes
2. Buried surface area (BSA) = $SAS_A + SAS_B - SAS_{AB}$, typical values for complexes: 1200-2200 Å^2

Chemistry

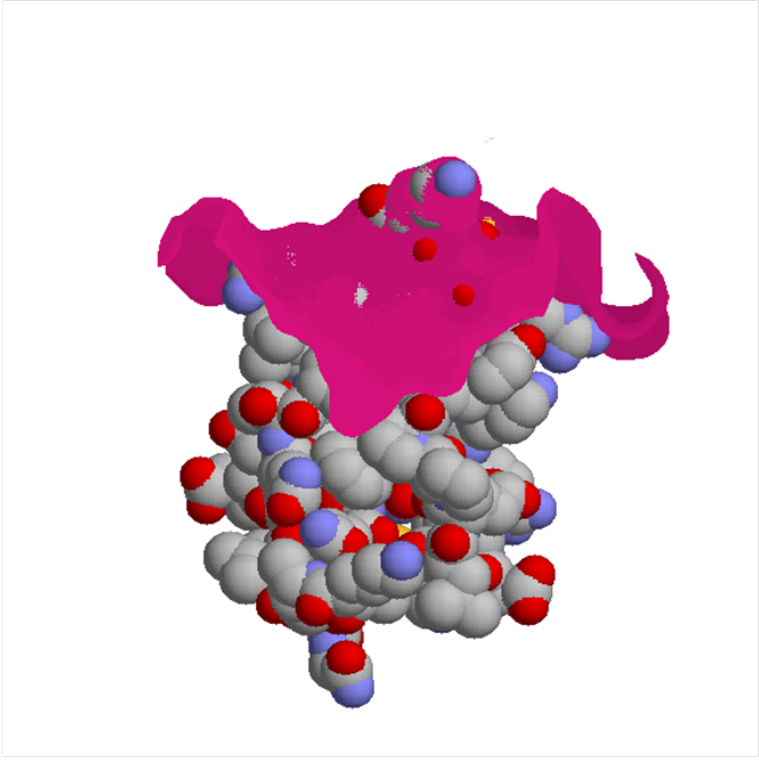
- Electrostatic interactions
- Hydrogen bonding
- *Desolvation*: Exclusion of the solvent from the interface => solvent entropy change

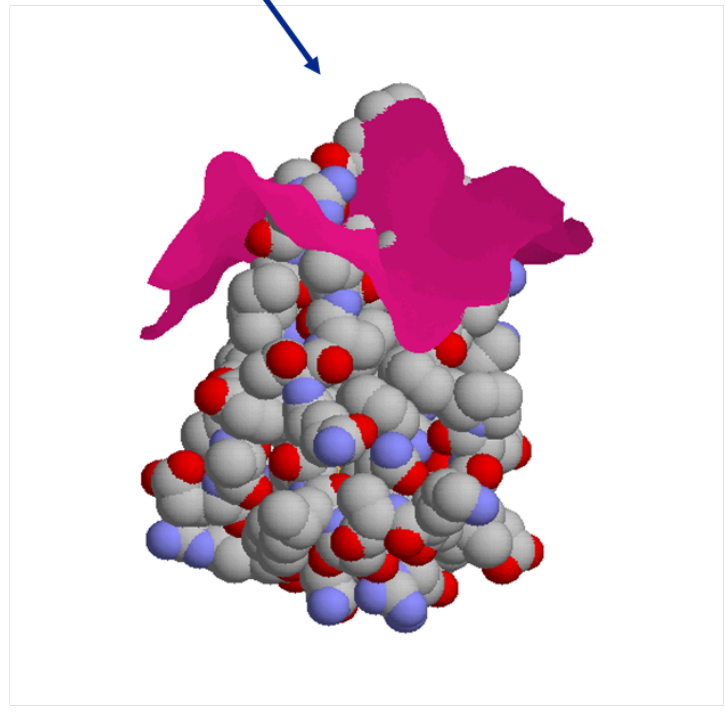
Categories of scoring functions

- *Knowledge-based*
- *Empirical*
- *Forcefield-based*

[Moreira et al., 2010]

Bound VS unbound

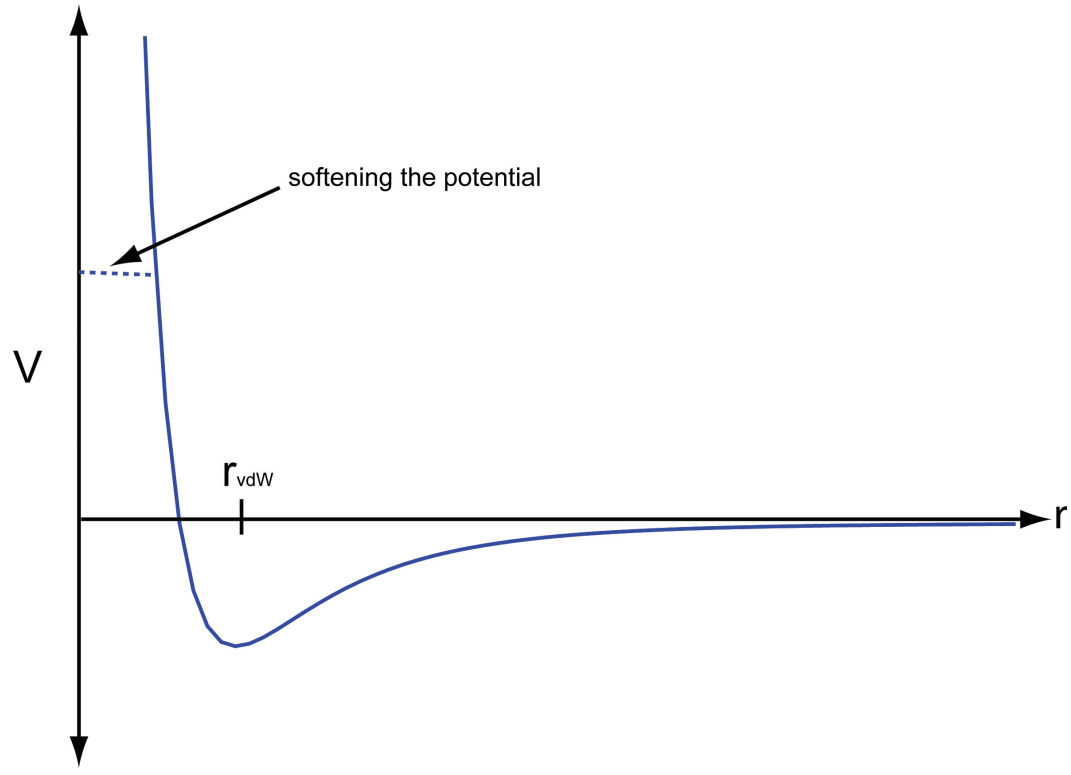




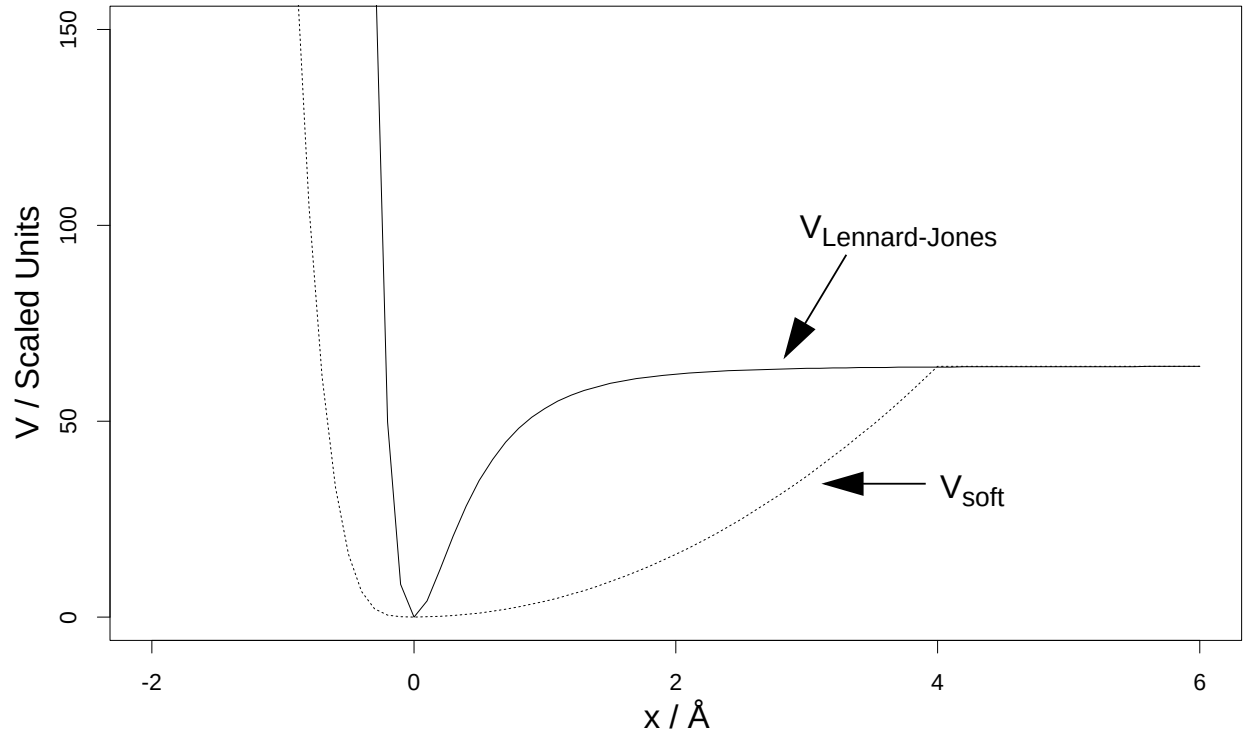
Kallikrein A/trypsin inhibitor complex (PDB codes 2KAI,6PTI)

Soft van der Waals

$$V_{L-J} = A/r^{12} - B/r^{6} \tag{2}$$



[Pierce and Weng, 2007]



$$V_{\text{Lennard-Jones}} = 4\varepsilon\left\{\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right\}$$

$$V_{\text{soft}} = \begin{cases} 256x^4 & x < 0\text{Å} \\ 4x^2 & \text{when} \quad 0\text{Å} \le x \le 4\text{Å} \\ 64 & x > 4\text{Å} \end{cases}$$

Source: Matthew James Betts, PhD thesis, 1999

Poisson-Boltzmann equation:

$$\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})) - \epsilon(\mathbf{r})\kappa^2(\mathbf{r})\sinh(\phi(\mathbf{r})) + \rho(\mathbf{r}) = 0 \tag{3}$$

ϵ dielectric term, ϕ electrical potential, ρ charge density, κ charge screening parameter for mobile ions.
Simplifications:

1. no mobile ions => $\kappa = 0$
2. dielectric term invariant inside the protein: $\epsilon(\mathbf{r}) = \epsilon$

=> Poisson's equation:

$$\nabla^2\phi(\mathbf{r})) = -\frac{\rho(\mathbf{r})}{\epsilon} \tag{4}$$

=> Coulomb force:

$$F = \frac{Q_1 Q_2}{4\pi\epsilon_0 r^2} \sim 1/r^2 \tag{5}$$

No point - point model, but point - field model, as side chain positions are not always correct

Desolvation

Desolvation in protein binding is the energy needed to change water-protein bonds with bonds between proteins.
= "Hydrophobic effect"
Atomic contact energy (ACE) [Zhang et al., 1997]:

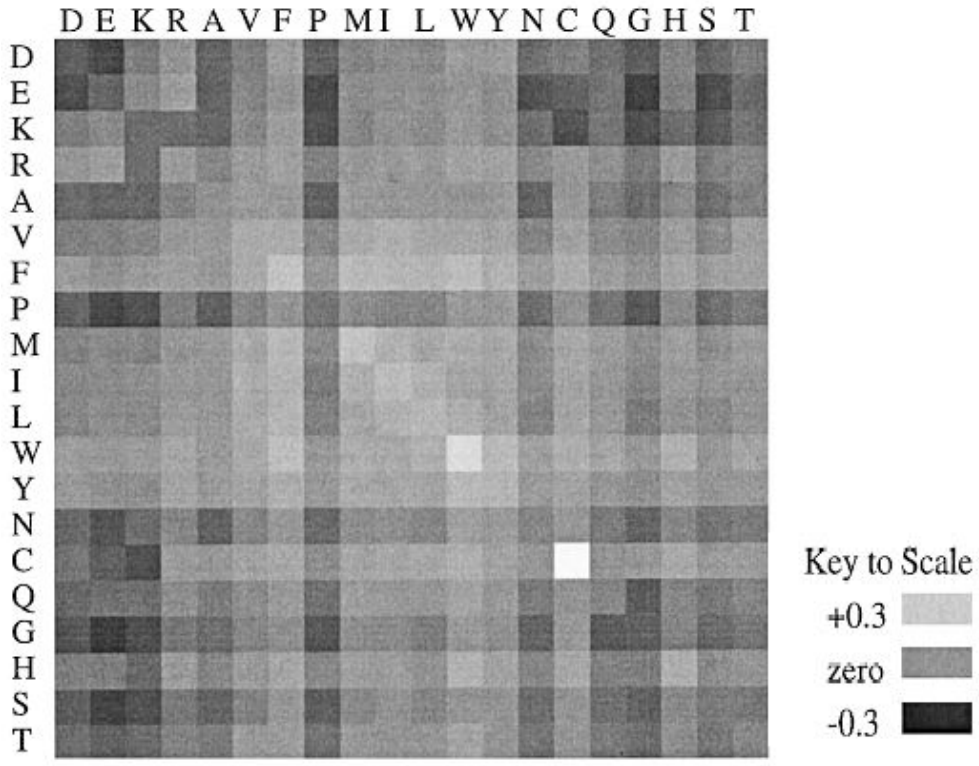
- Contact energies ΔG_i for 18 atom types obtained from known structures
- Statistical potential (like [Miyazawa and Jernigan, 1996])

$$\Delta G_{des} = \sum_i N_i \Delta G_i \tag{6}$$

N_i: number of atom pairs of type i

Statistical pairwise-potential

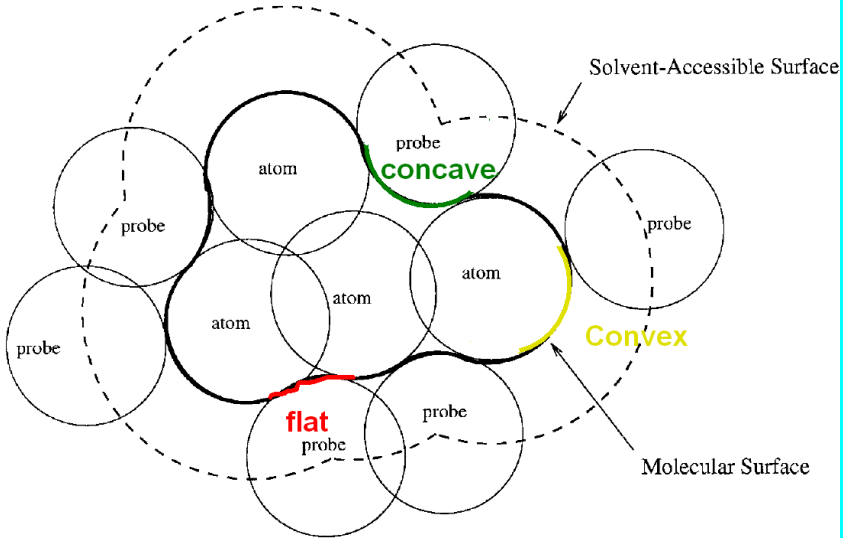
Derived from an analysis of complexes with known 3D structure, example:

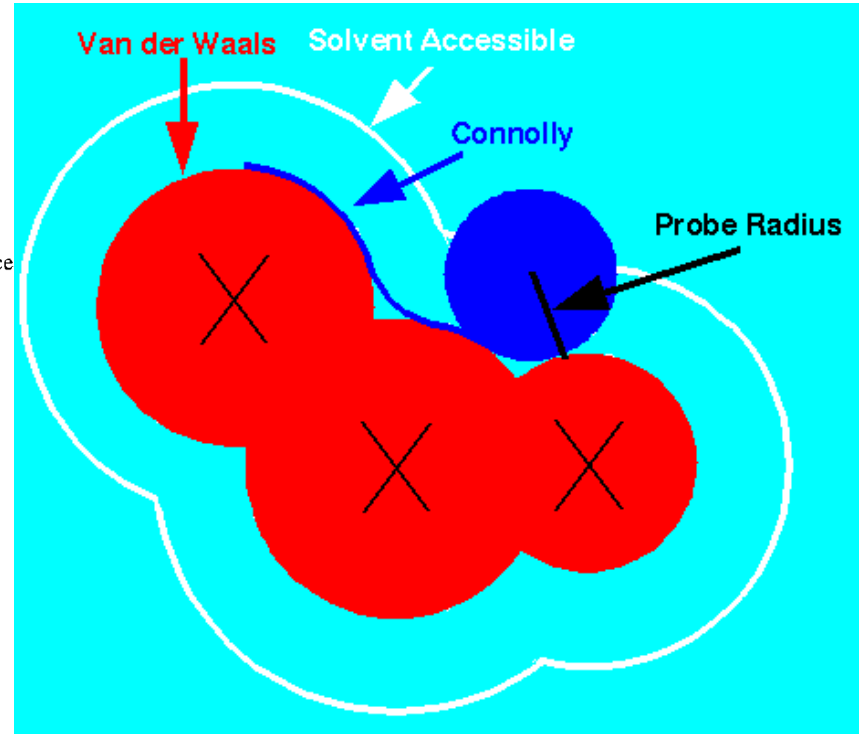


[Moont et al., 1999]

Solvent accessible surface - SAS

Connolly's MS (molecular surface) algorithm



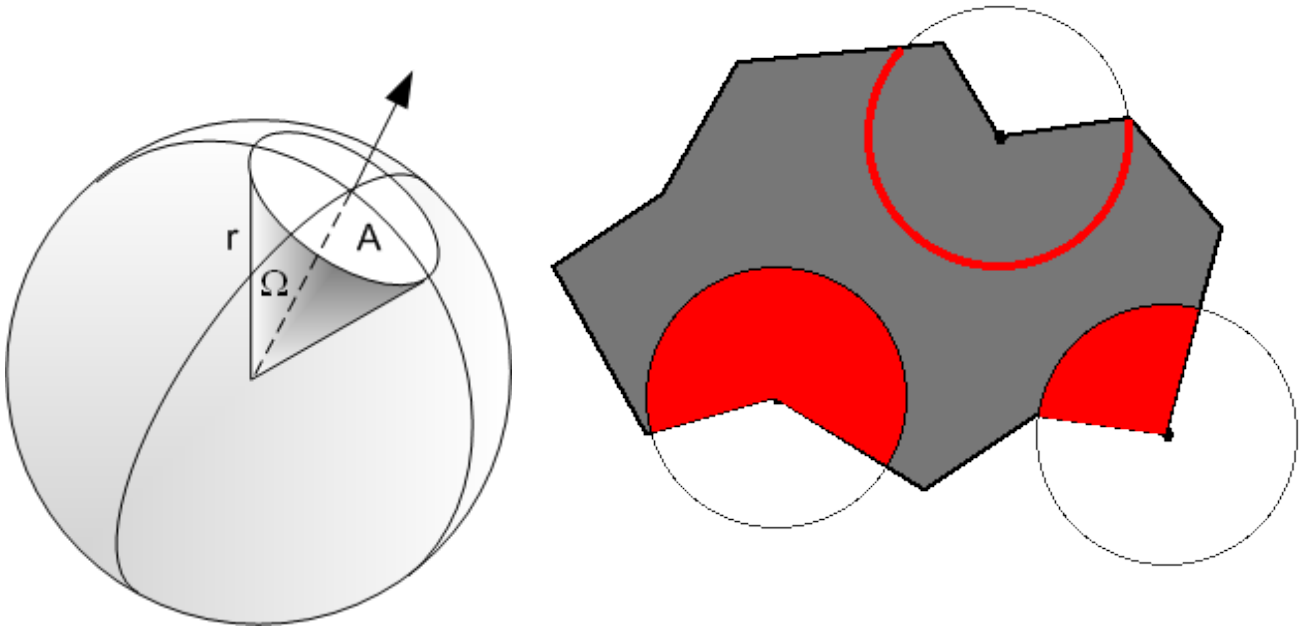


Cai 1998 / http://www.simbiosys.ca/sprout/eccc/cangaroo.html

Knobs and Holes

Solid angle

$$\Omega[sr] = A/r^2 = [0...4\pi]$$

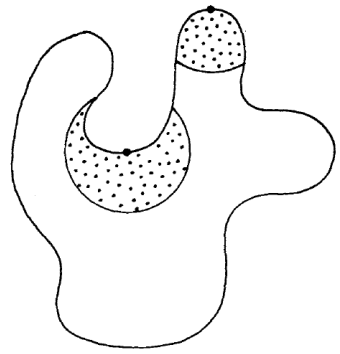


Michael L. Connolly, *Molecular Surfaces: A Review*
http://www.netsci.org/Science/Compchem/feature14.html
Connolly 1986, J Mol Graph

Knobs and Holes

Sphere volume inside the protein

[Connolly, 1986]:



shape function => sphere volume:

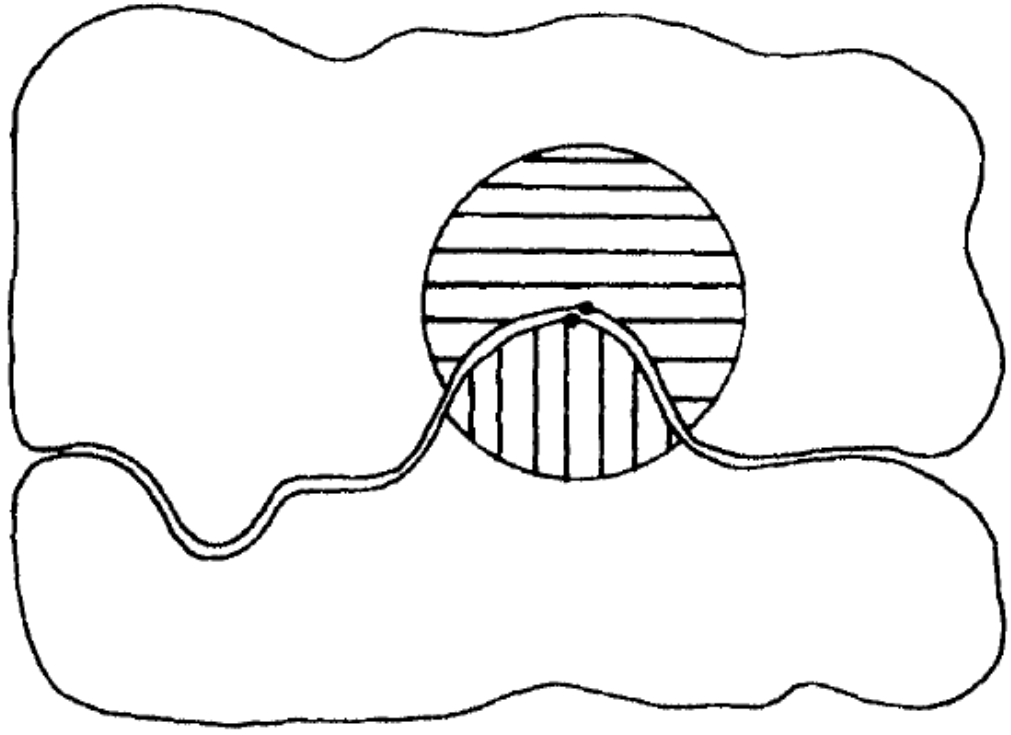
- concave/Hole = larger sphere volume = local maximum of shape function
- convex/Knob = smaller sphere volume = local minimum of shape function

here: sphere radius = 6Å (approximation of the radius of an amino acid)

Knobs and Holes

Matching

[Connolly, 1986]:

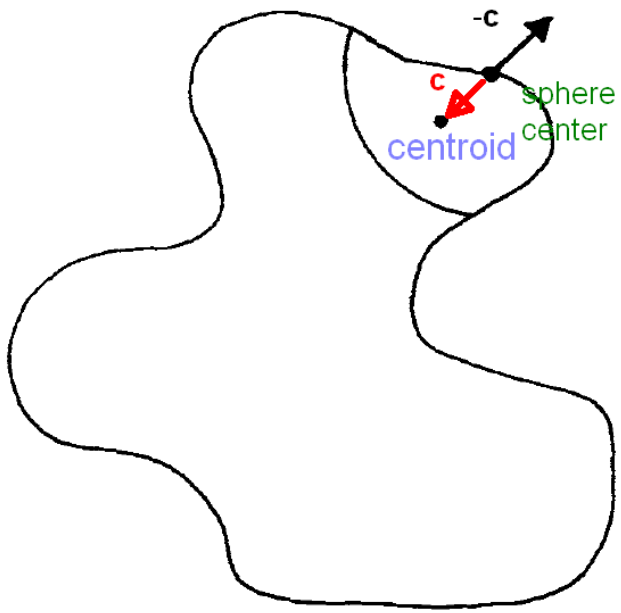


sum of sphere volumes should give a whole sphere

Knobs and Holes

Matching

[Connolly, 1986]:
One outward pointing vector **-c** at each sphere center => vector field
good shape match = anti-parallel vectors



centroid = barycentre

Knobs and Holes

Matching

[Connolly, 1986]:
Criteria for a good surface shape measure for docking:

1. local, i.e. not dependent on distant parts of the protein (the protein-protein interface is only a local part of the whole surface)
2. independent of the coordinate system (otherwise the complementarity is difficult to find, as proteins
3. fast way to identify complementary shapes

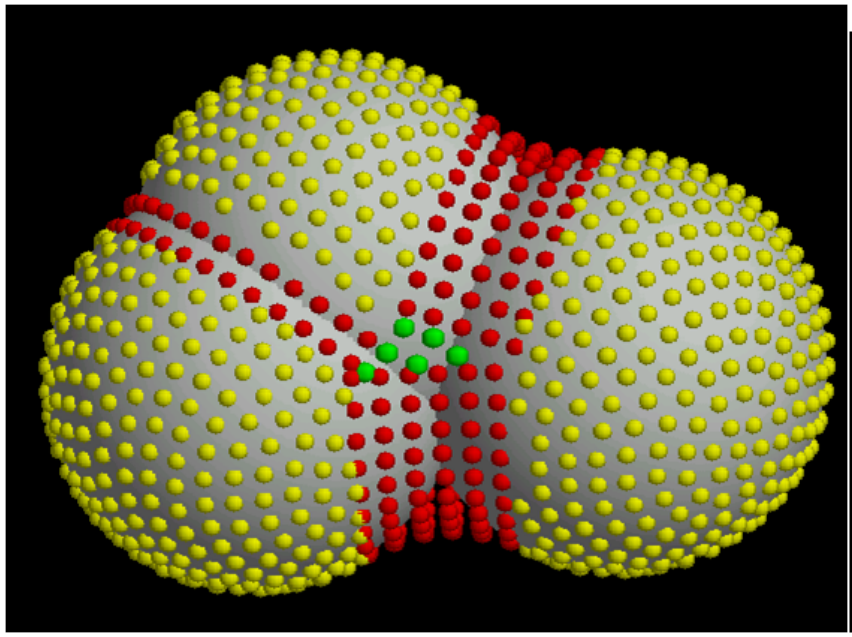
Critical Points

Critical points = Local extrema of shape function = knob and holes
Find critical points:

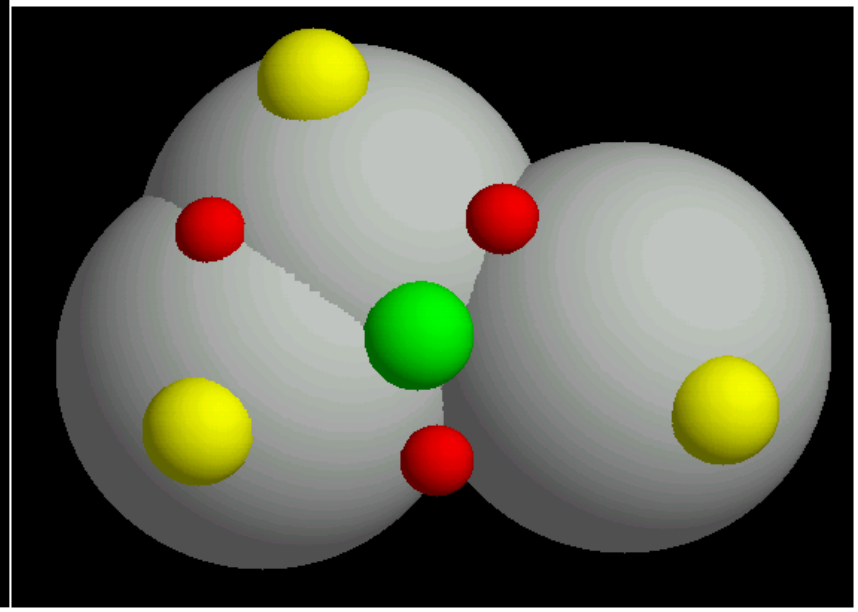
1. triangulate the solvent-accessible surface => polyhedron with triangular faces (better than dot surface representation, as it defines which vertices are neighbors)
2. calculate shape function at each vertex of the polyhedron
3. compare values with neighboring vertices

Knob = value lower than any of the neighboring vertices
Hole = value higher than any of the neighboring vertices
Shape function = sphere volume inside the protein
Tested on one complex: about 160 knobs and holes per protein
[Connolly, 1986]

Dot surface VS critical points



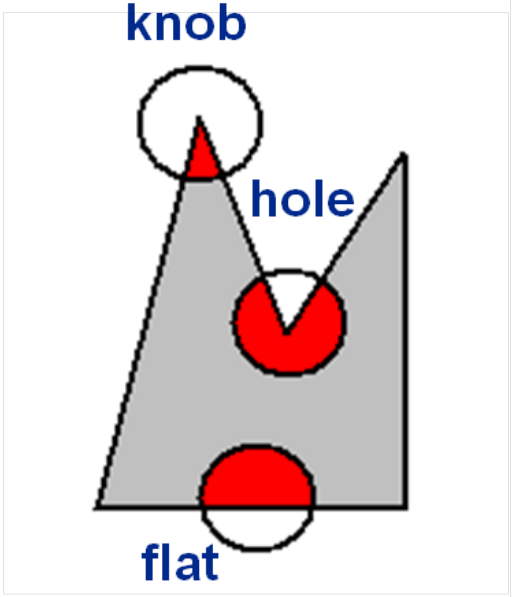
(a) dense, Connolly

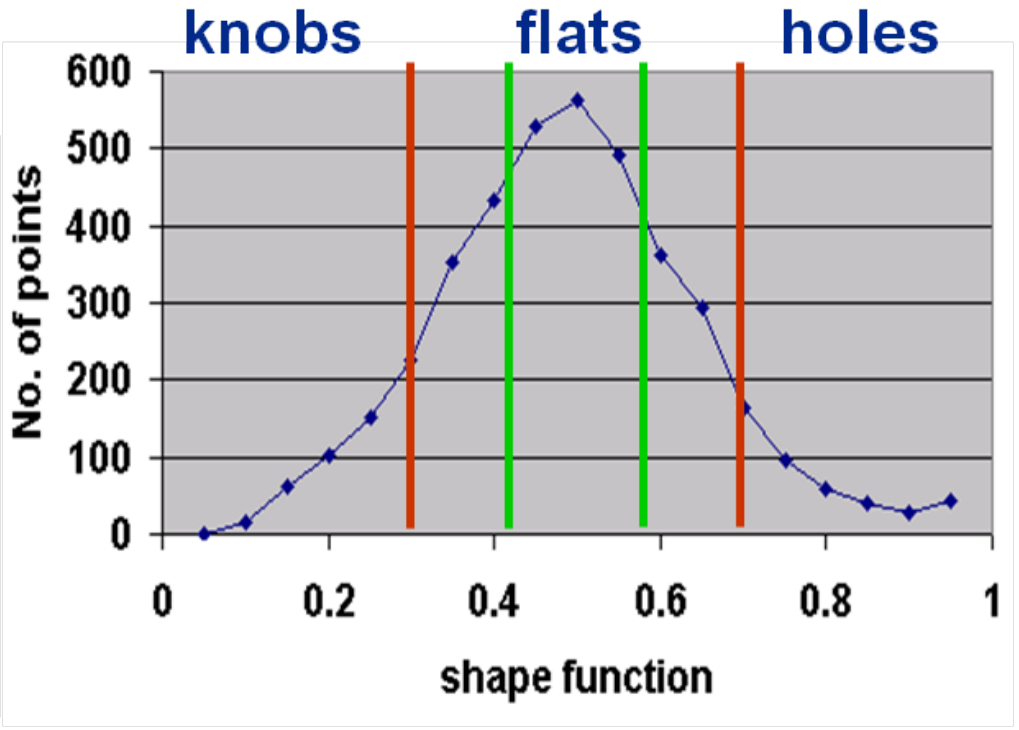


(b) sparse, Lin et al. 1994

green = concave, yellow = convex, red = flat

Critical points - Histogram

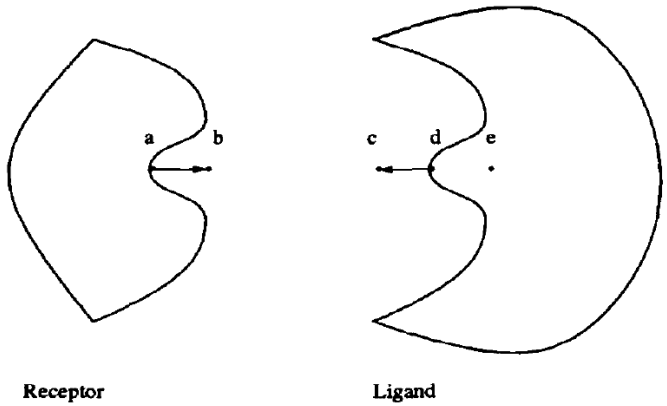




Only 30% are knobs or holes.
Max Shatsky
http://bioinfo3d.cs.tau.ac.il/Education/Workshop02a/

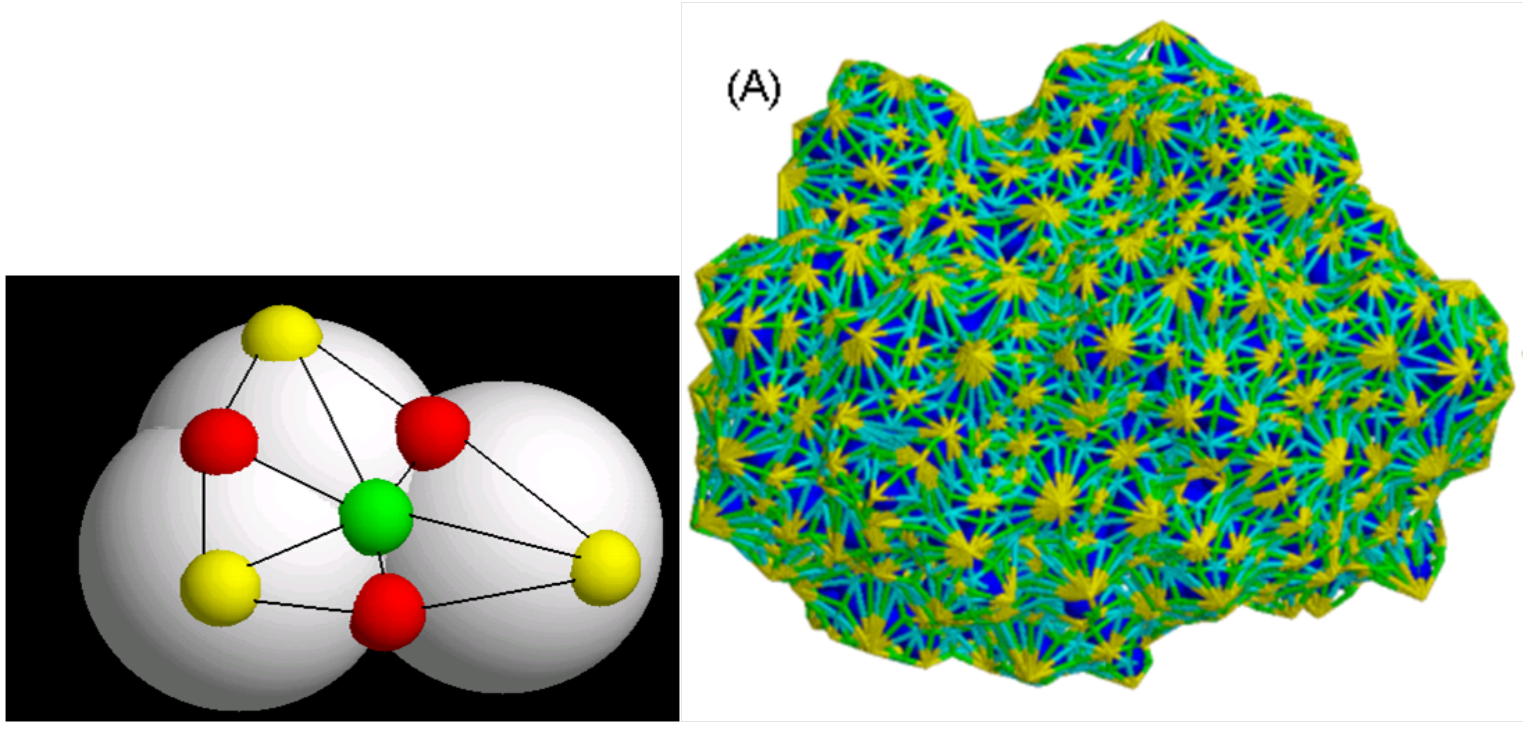
Matching with critical points

At least four points of each protein must be matched together to define one assembly unambiguously.
First try [Connolly, 1986]: Four knobs and holes pairs (Problem: difficulties to find four pairs, especially for flat interfaces, ex: trypsin + inhibitor)
Second try [Norel et al., 1994]: Two knobs and holes pairs plus points defined by their surface normals:



matchings: a <-> d, b <-> e

Topological graph G_{top}



Color code of the right figure: yellow = knob, cyan = hole, green = flat, dark blue = protein surface
http://bioinfo3d.cs.tau.ac.il/Education/Workshop02a/

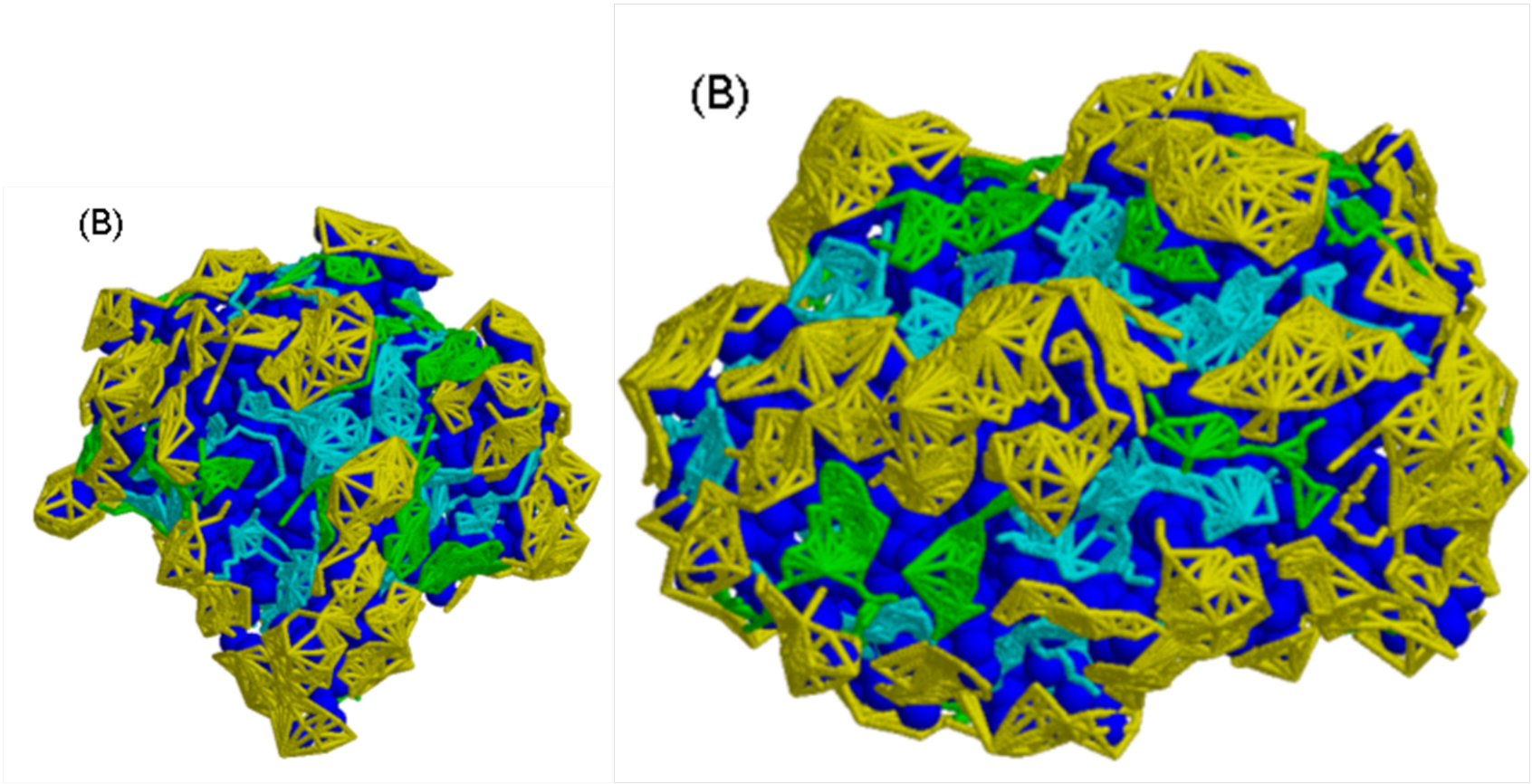
Group critical points as patches

Goal: divide the surface into connected, non-intersecting, equal sized patches of critical points with similar curvature.

- *connected* the points of the patch correspond to a connected sub-graph of G_{top}.
- *similar curvature* all the points of the patch correspond to only one type: knobs, flats or holes.
- *equal sized* to assure better matching we want shape features of almost the same size.

http://bioinfo3d.cs.tau.ac.il/Education/Workshop02a/

Group critical points as patches



yellow = knob, cyan = hole, green = flat, dark blue = protein surface

http://bioinfo3d.cs.tau.ac.il/Education/Workshop02a/

Surface Patch Matching

Knob <-> hole patches and flat patches <-> any patch

1. *Single Patch Matching*: One patch of the receptor with one patch of the ligand, for small ligands
2. *Patch-Pair Matching*: Two patches of the receptor with two patches of the ligand, for protein-protein complexes

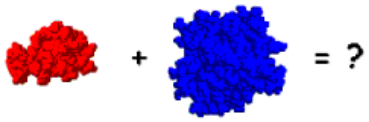
Match critical points within patches by computer vision techniques:

- Geometric Hashing
- Pose Clustering

[Duhovny et al., 2002]

Surface Patch Matching

PATCHDOCK



Molecular Docking Algorithm Based on Shape Complementarity Principles

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

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:	Parcourir...
Ligand Molecule:		(PDB:chainId e.g. 2kai:I) **or** upload file:	Parcourir...
e-mail address:		(the results are sent to this address)	
Clustering RMSD:	4.0		
Complex Type:	Default	Be sure to give receptor and ligand in the corresponding order!	

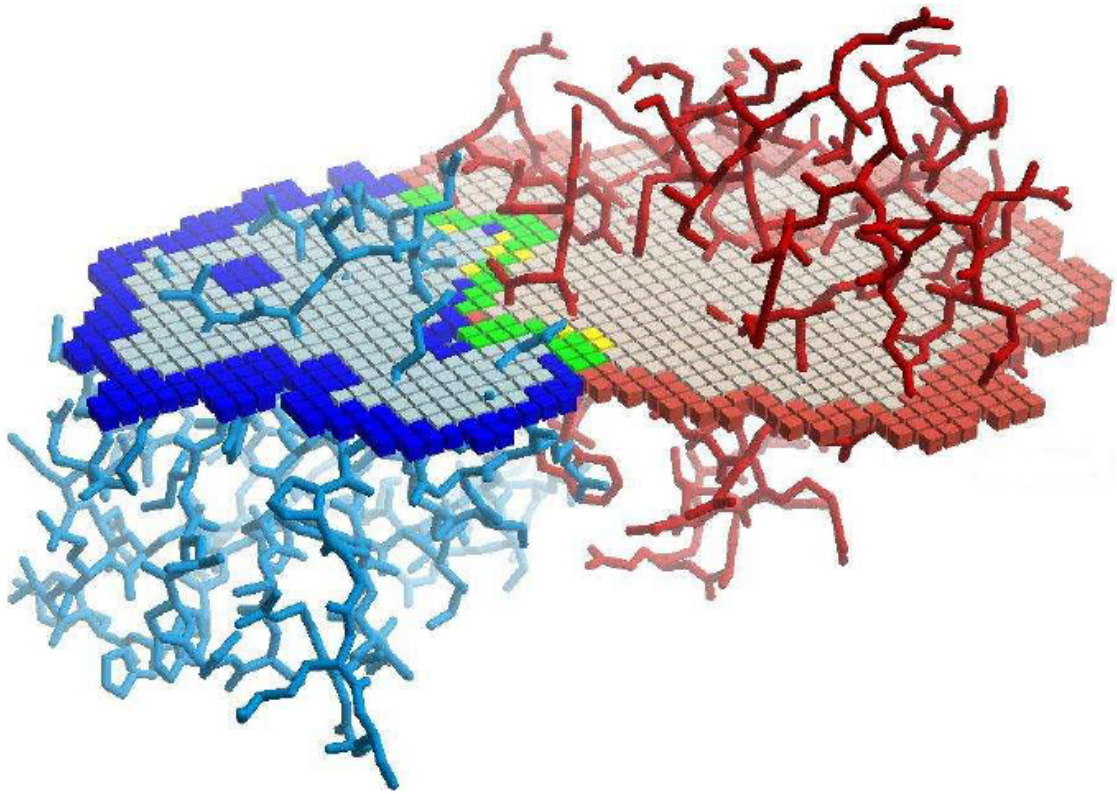
Submit Form Clear

Advanced Options:
[Show][Hide]

FireDock - Fast Interaction Refinement in Molecular Docking
SymmDock - An Algorithm for Prediction of Complexes with C_n Symmetry

Beta 1.3 Version, Contact: duhovka@gmail.com

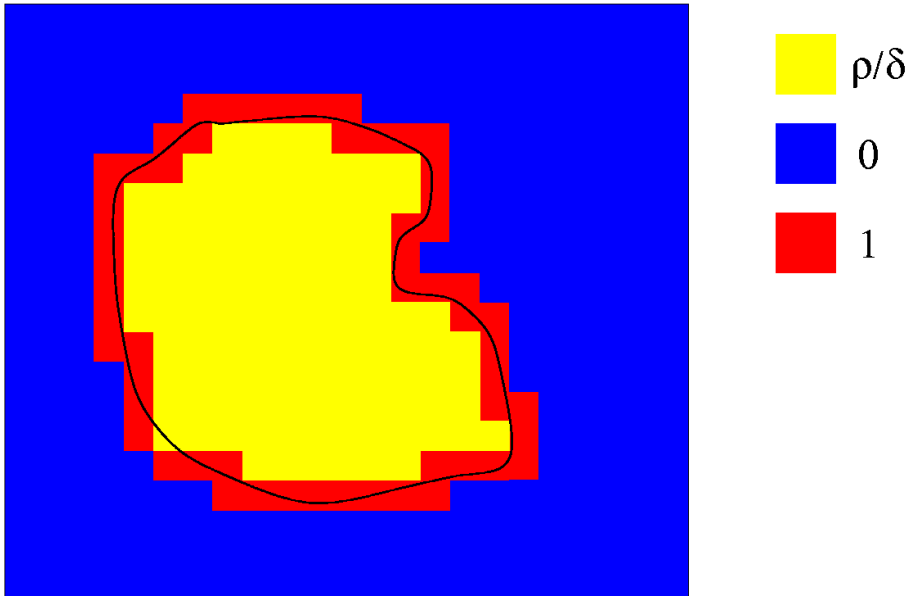
3D grid



[Palma et al., 2000] [Krippahl et al., 2003]

Katchalski-Katzir et al., PNAS 1992

- Protein on grid
- Assign values
  - $a_{i,j,k}$ =
    - 1 at the surface of A
    - ρ << 0 inside A
    - 0 outside
  - $b_{i,j,k}$ =
    - 1 at the surface of B
    - δ > 0 inside B
    - 0 outside B



A \ B	inside	surface	outside
inside	ρ* δ < 0	ρ < 0	0
surface	δ > 0	1	0
outside	0	0	0

Source: Kohlbacher and Lenhof

Correlation $c_{\alpha,\beta,\gamma}$

For all translation vectors (α, β, γ) calculate:

- surface-surface contacts $a_{i,j,k} \cdot b_{i+\alpha,j+\beta,k+\gamma} > 0$
- inside-inside contacts $a_{i,j,k} \cdot b_{i+\alpha,j+\beta,k+\gamma} < 0$

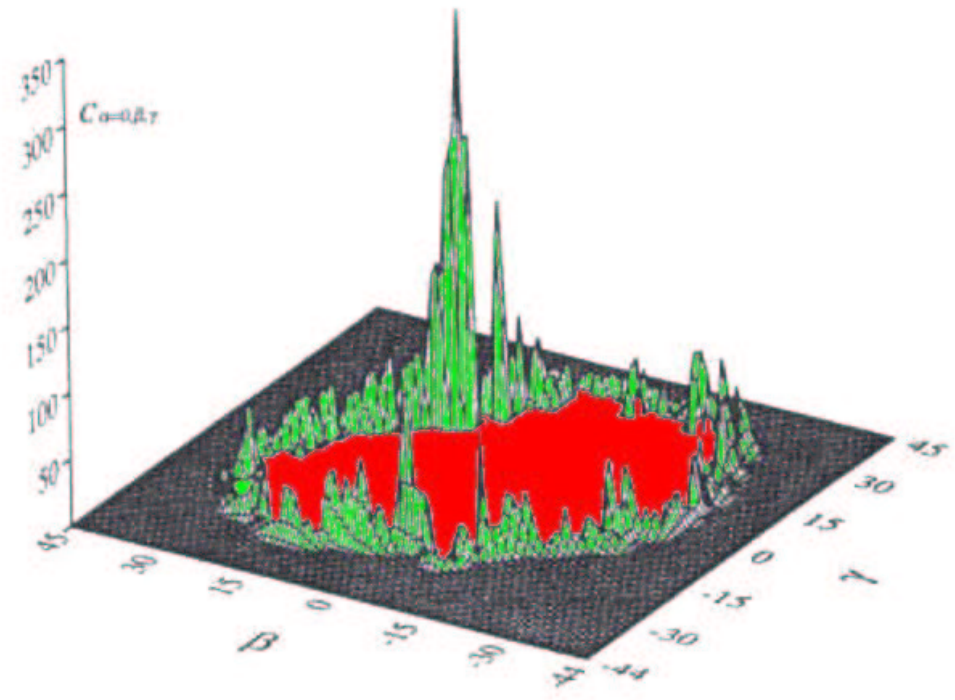
$$c_{\alpha,\beta,\gamma} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{i,j,k} \cdot b_{i+\alpha,j+\beta,k+\gamma} \tag{7}$$

Run time $O(N^6)$!

[Katchalski-Katzir et al., 1992]

Correlation $c_{\alpha,\beta,\gamma}$

Cross Section $c_{\alpha=0,\beta,\gamma}$



From: Katchalski-Katzir et al., PNAS 1992, 2195

Source: Kohlbacher and Lenhof

Fast Fourier Transform (FFT)

Discrete Fourier Transform (DFT):

$$X_{o,p,q} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} x_{i,j,k} \cdot exp[-2\pi i(oi + pj + qk)/N] \quad (8)$$

Inverse Fourier Transform (IFT):

$$c_{\alpha,\beta,\gamma} = \frac{1}{N^3}\sum_{o=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N} C_{o,p,q} \cdot exp[-2\pi i(o\alpha + p\beta + q\gamma)/N] \quad (9)$$

Fast Fourier Transform (FFT)

$$c_{\alpha,\beta,\gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{i,j,k} \cdot b_{i+\alpha,j+\beta,k+\gamma} \tag{10}$$

Cross-correlation:

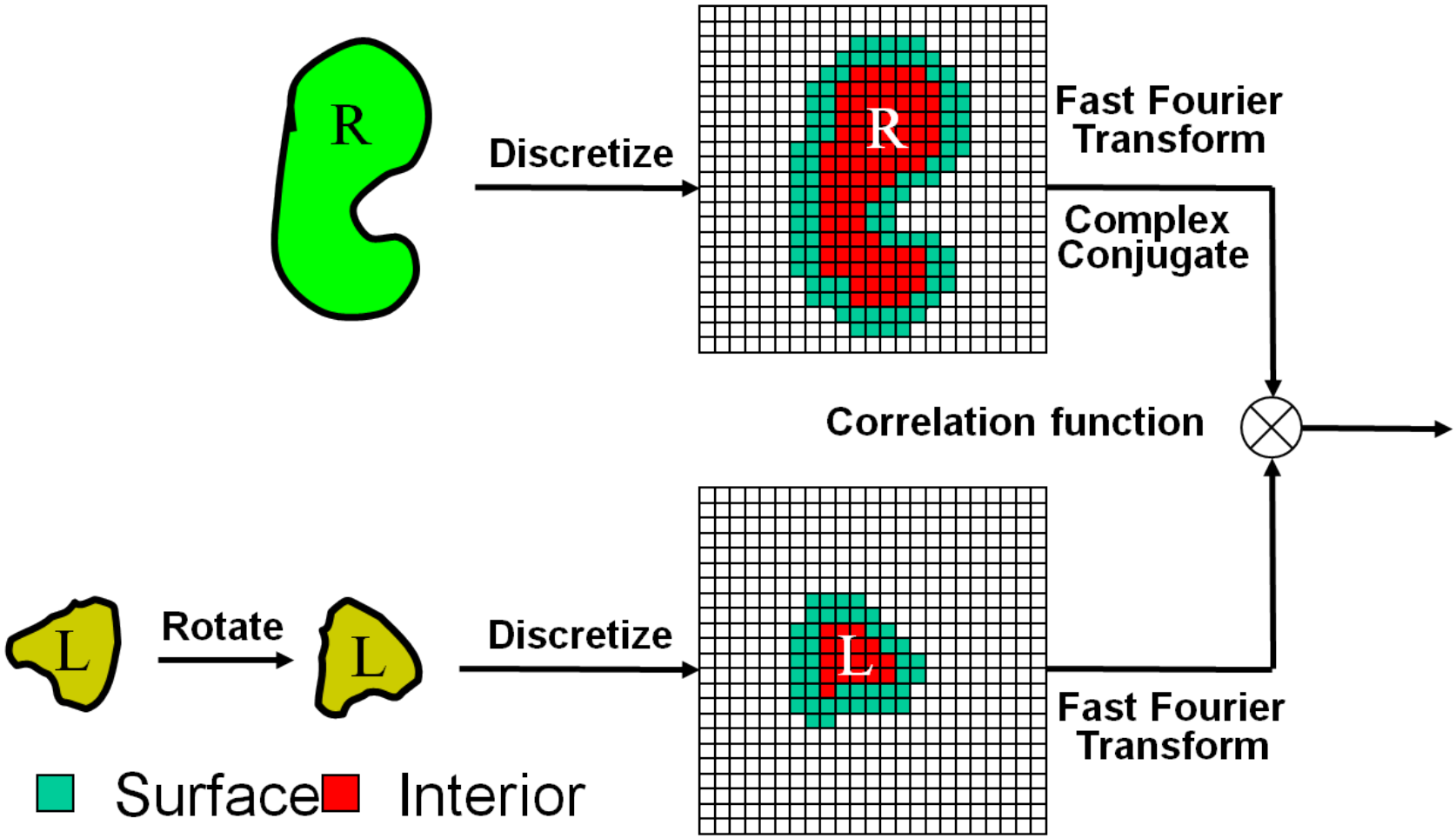
$$(f \bigstar g)[n] = \sum_{m=-\infty}^{\infty} f^*[m]g[n+m] \tag{11}$$

$$DFT(f \bigstar g) = (DFT(f))^* \cdot DFT(g) \tag{12}$$

$$C_{o,p,q} = A^*_{o,p,q} \cdot B_{o,p,q} \tag{13}$$

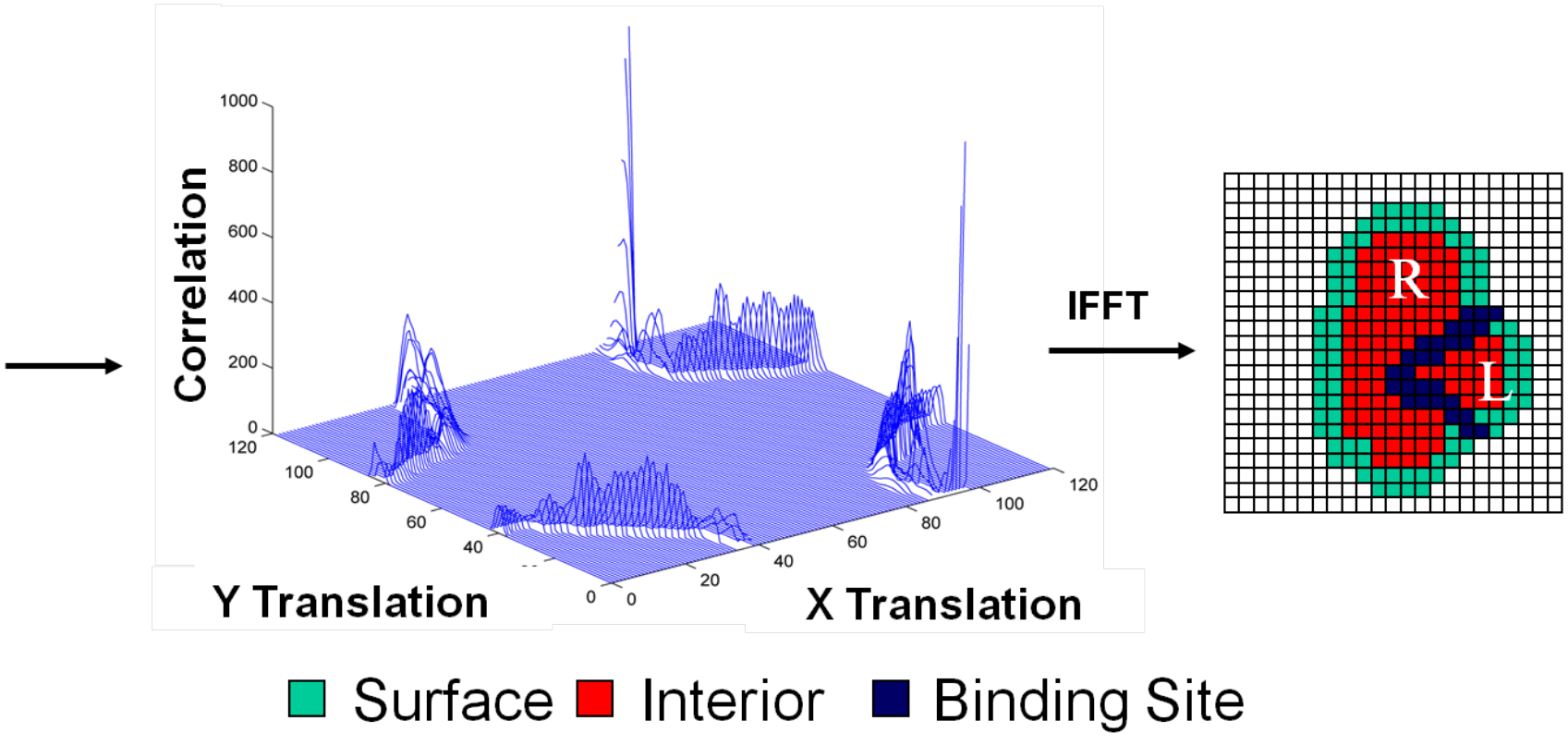
$$c_{\alpha,\beta,\gamma} = IFT(C_{o,p,q}) \tag{14}$$

FFT for DFT



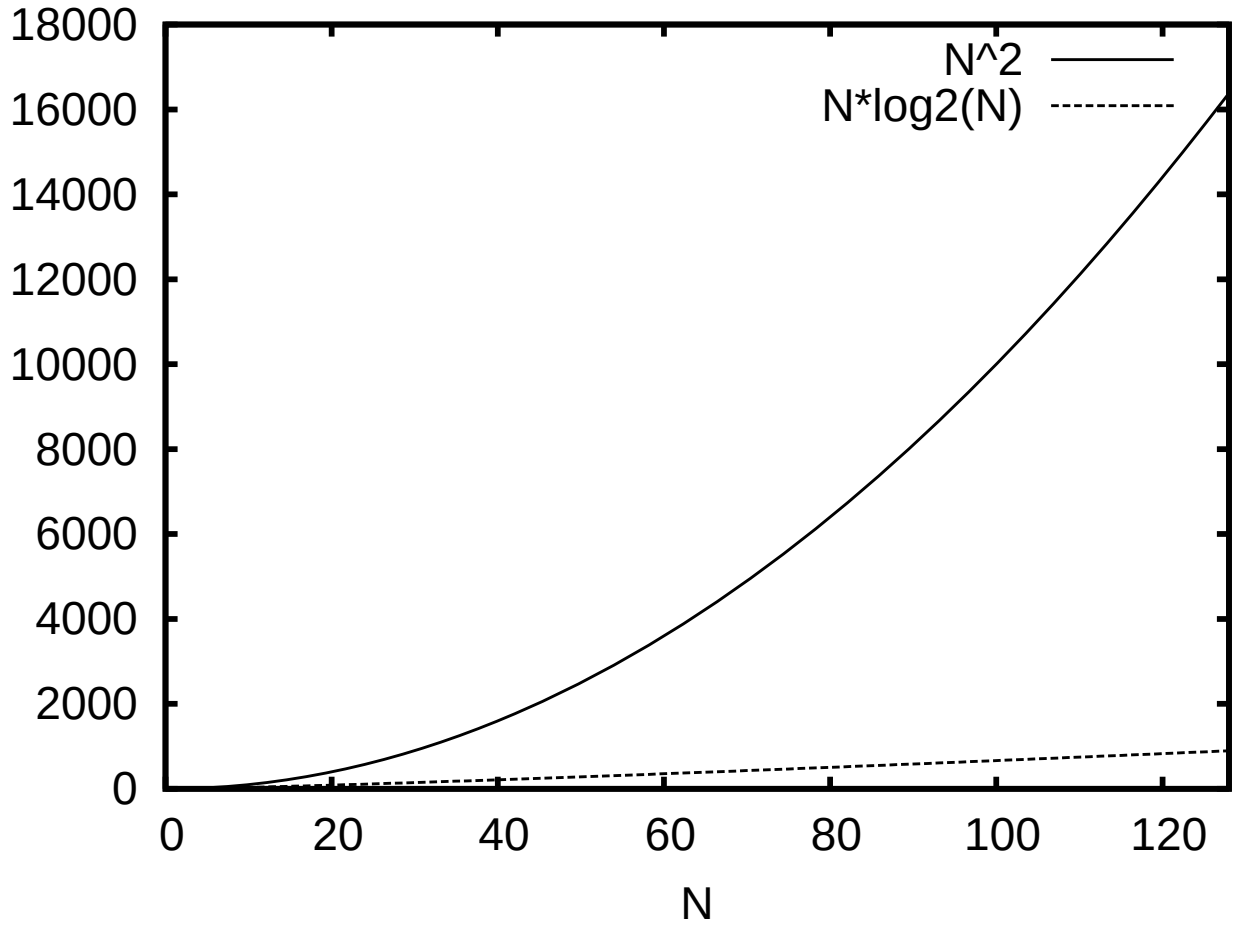
Source: Rong Chen

FFT for IFT

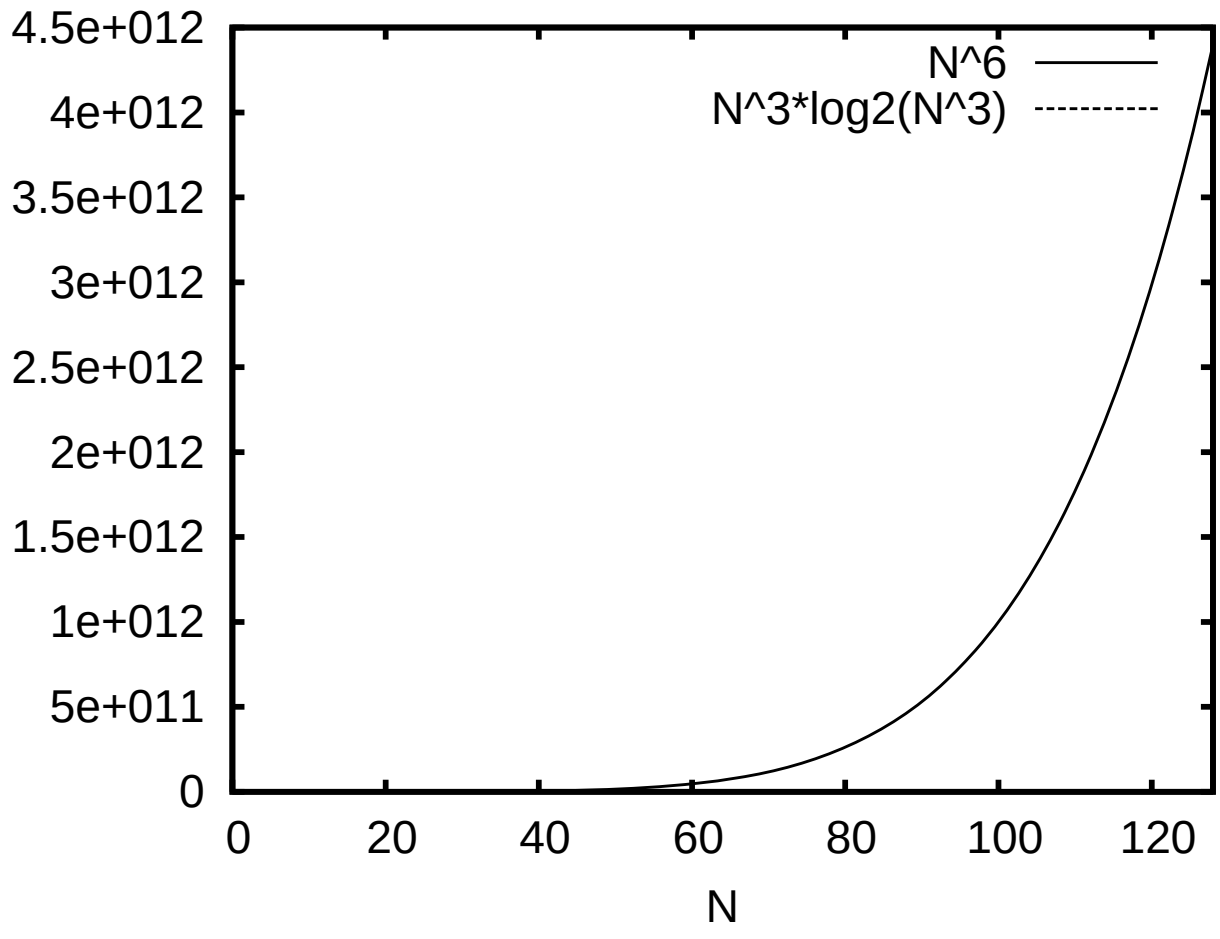


Source: Rong Chen

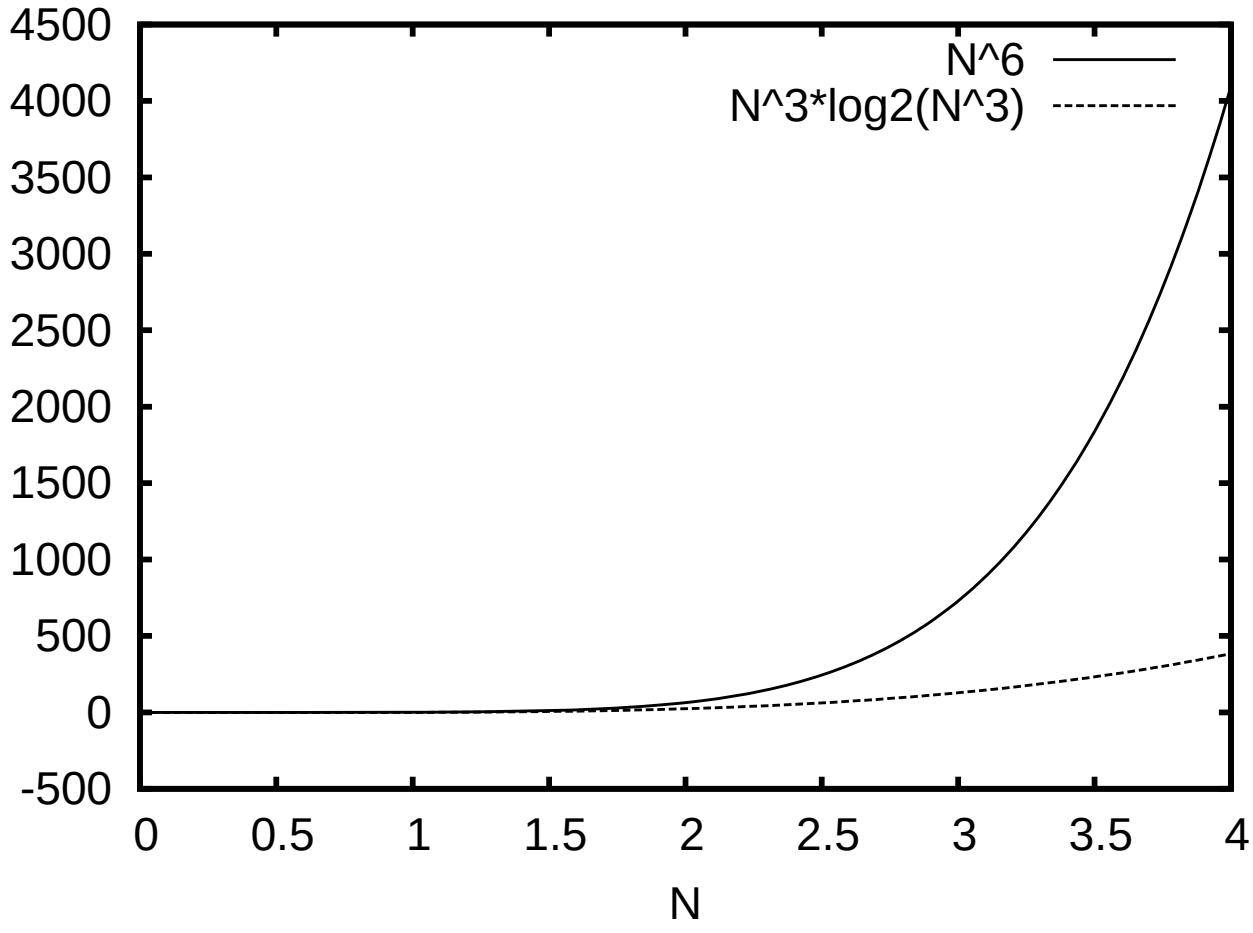
FFT 1D


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
N
N^2
N*log2(N)

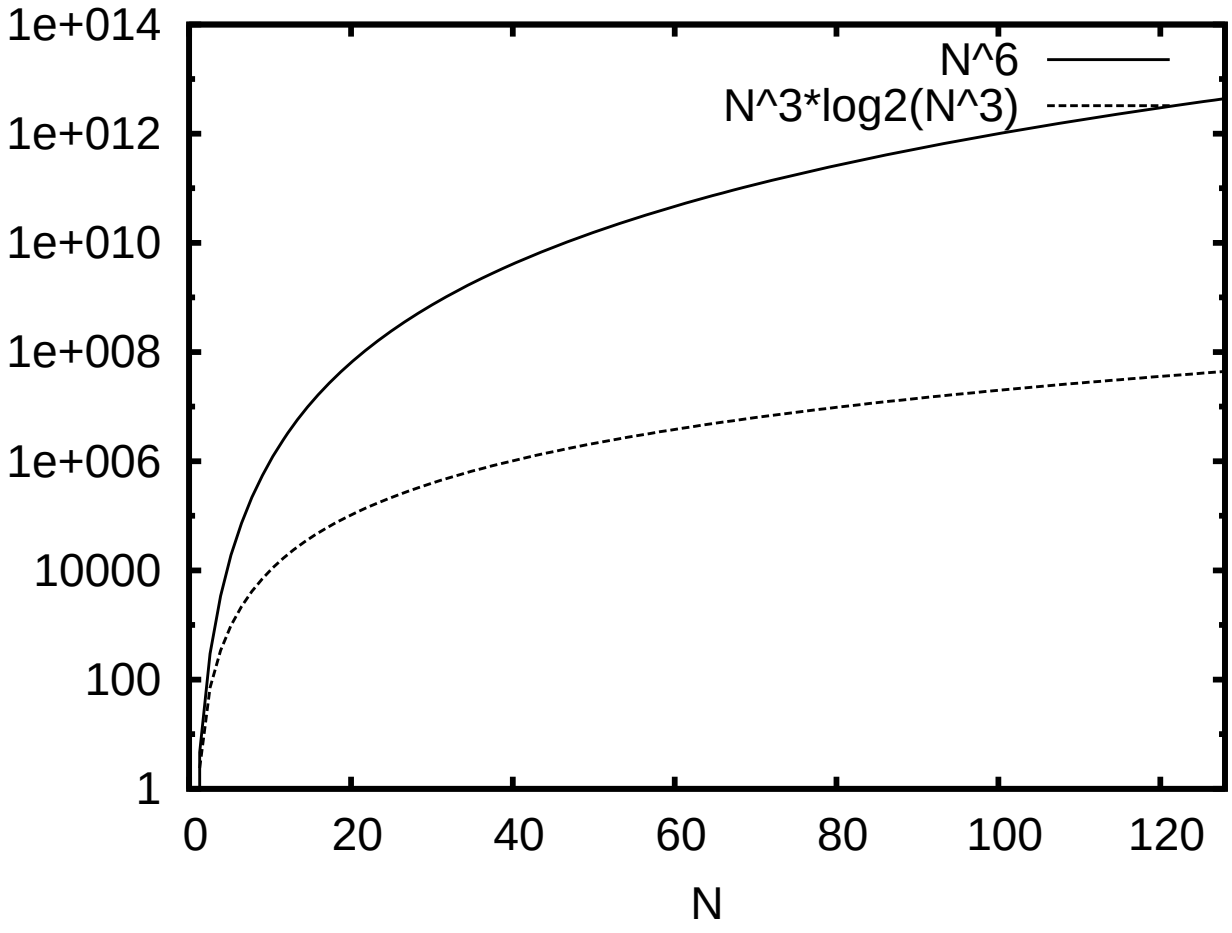
FFT 3D


4.5e+012
4e+012
3.5e+012
3e+012
2.5e+012
2e+012
1.5e+012
1e+012
5e+011
0
0
20
40
60
80
100
120
N
N^6
N^3*log2(N^3)

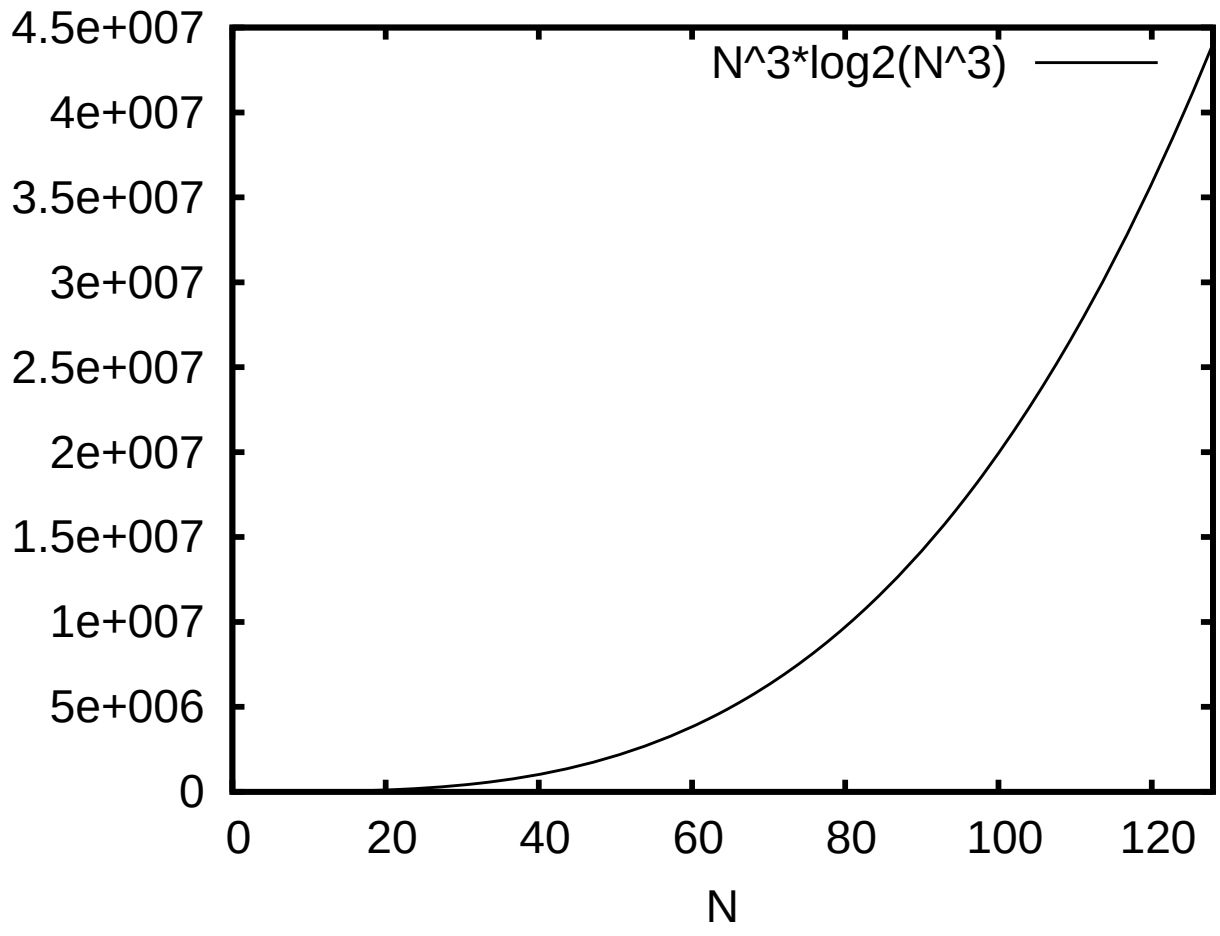
FFT 3D


4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
0
0.5
1
1.5
2
2.5
3
3.5
4
N
N^6
N^3*log2(N^3)

FFT 3D


1e+014
1e+012
1e+010
1e+008
1e+006
10000
100
1
0
20
40
60
80
100
120
N
N^6
N^3*log2(N^3)

FFT 3D


4.5e+007
4e+007
3.5e+007
3e+007
2.5e+007
2e+007
1.5e+007
1e+007
5e+006
0
0
20
40
60
80
100
120
N
N^3*log2(N^3)

Grid size in ZDOCK

- Grid spacing: 1.2 Å
- Grid points N = 128 for the largest protein (about 150 Å cube side length), otherwise N = 100
- 128^3 = 2 million grid points => 2 million different translation vectors (α, β, γ)
- Without FFT => $128^6 = 4.4 \cdot 10^{12}$ = 4400 billion elementary operations (addition or multiplication)
- With FFT => $128^3 \cdot log_2(128^3) = 2.1 \cdot 10^6 \cdot 21$ = 44 million elementary operations

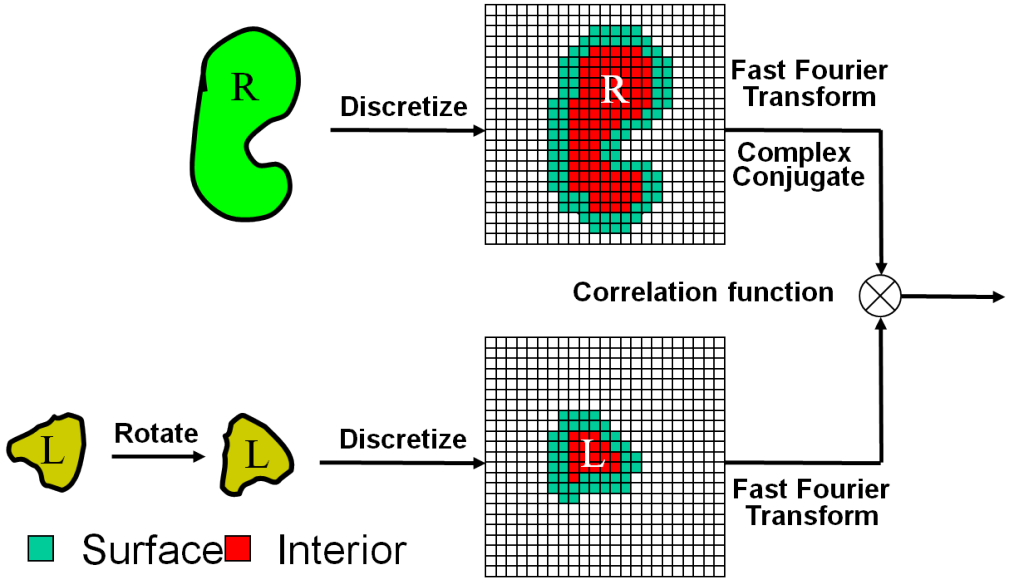
=> 10^5 times faster with FFT !

[Chen and Weng, 2002]

Ligand rotations

ZDOCK 2.3-3.x => two rotational sampling options (non-redundant rotations, uniform sampling of the sphere):

1. $\Delta = 15^\circ$ => $M_{rot} = 3600$
   => $M_{rot} \cdot N^3$ = 7.5 billion docking poses
2. $\Delta = 6^\circ$ => $M_{rot} = 54000$
   => $M_{rot} \cdot N^3$ = 113 billion docking poses



Total number of operations

$$M_{trans+corr} = N^3 \cdot log_2(N^3) \tag{15}$$

$$M_{total} = M_{rot} \cdot M_{trans+corr} = M_{rot} \cdot N^3 \cdot log_2(N^3) \tag{16}$$

ZDOCK 2.3-3.x =>
M_{total} = 160 billion operations with M_{rot} = 3600 => average runtime (2.3: 1h, 3.0: 3h)
M_{total} = 2300 billion operations with M_{rot} = 54000 => average runtime (2.3: 15h, 3.0: 45h)

[Pierce et al., 2011]

Run-time improvement with Conv3D

Table 1. Average running time, running time fold improvement, and memory usage of optimized ZDOCK versions.

Name	Optimization[1]	Running Time (min)	Fold Improvement[2]	Memory (MB)
ZDOCK 3.0	-	167.1	-	700
ZDOCK 3.0.1	Conv3D	26.5	6.4	303
ZDOCK 3.0.2f	Conv3D+Cent	23.2	7.2	282
ZDOCK 3.0.2	**Conv3D+Cent+Rot+Switch**	**18.9**	**8.6**	**256**
ZDOCK 2.3	-	53.2	-	296
ZDOCK 2.3.1	Conv3D	13.1	4.0	215
ZDOCK 2.3.2f	Conv3D+Cent	11.2	4.7	203
ZDOCK 2.3.2	**Conv3D+Cent+Rot+Switch**	**9.3**	**5.5**	**191**

All values are averages from running ZDOCK on 176 unbound docking test cases, each run using a single 2.8 GHz 64-bit Opteron processor with 8 GB available RAM.

[Pierce et al., 2011], [Nukada et al., 2007]

Introduction

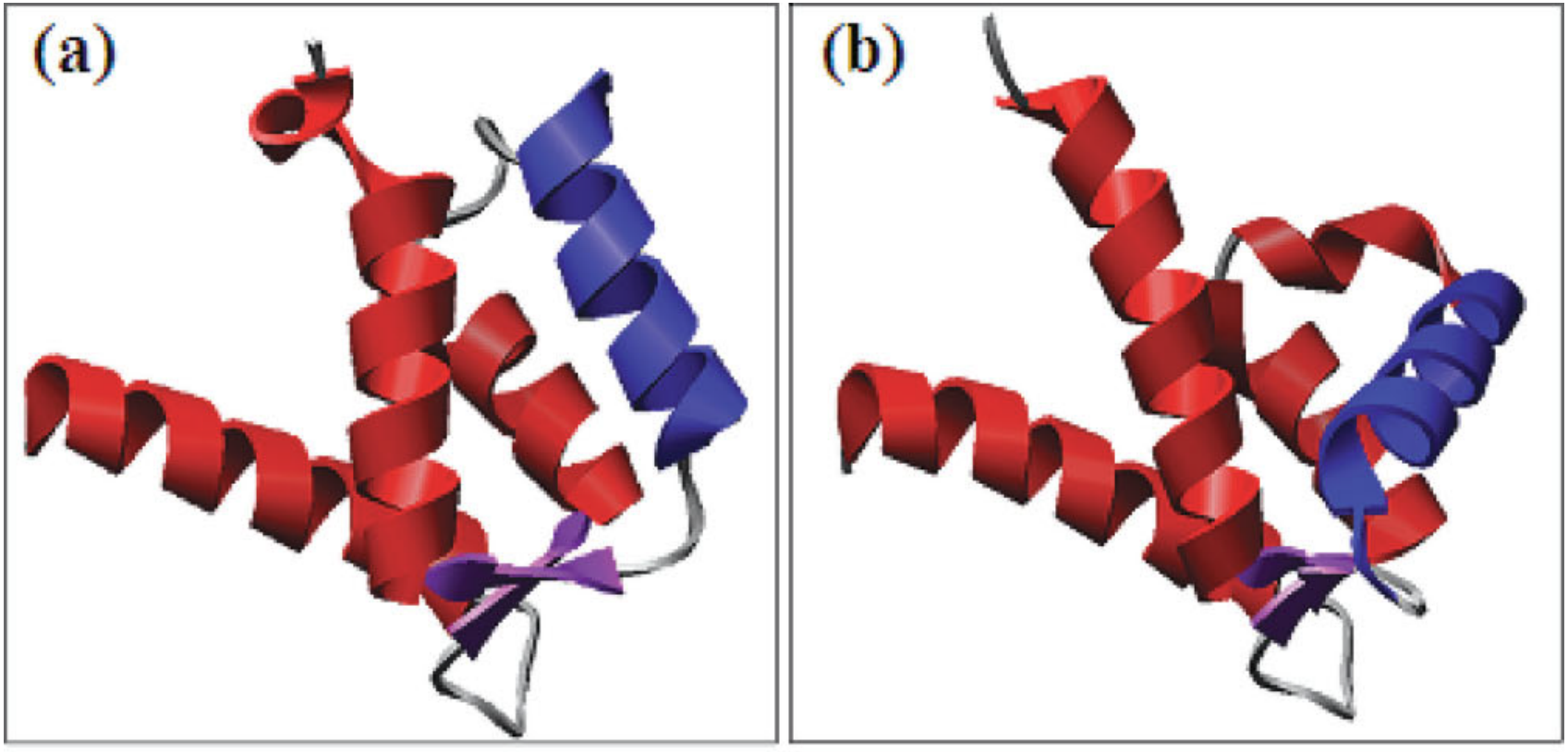
- Flexibility makes the docking problem harder
  - Increased number of degrees of freedom
  - Scoring more difficult
- Difficult to predict a-priori conformational changes
- Current docking methodology can mainly deal with small conformational changes

Reviews

- Bonvin, A. M. J. J. (2006). Flexible protein-protein docking. *Curr. Opin. Struct. Biol.*, 16(2):194–200. PMID: 16488145
- Andrusier, N., Mashiach, E., Nussinov, R., and Wolfson, H. J. (2008). Principles of flexible protein-protein docking. *Proteins*, 73(2):271–289. PMID: 18655061
- Zacharias, M. (2010). Accounting for conformational changes during protein-protein docking. *Curr. Opin. Struct. Biol.*, 20(2):180–186. PMID: 20194014
- Tuffery, P. and Derreumaux, P. (2012). Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches. *J R Soc Interface*, 9(66):20–33. PMID: 21993006

Large-scale domain motions

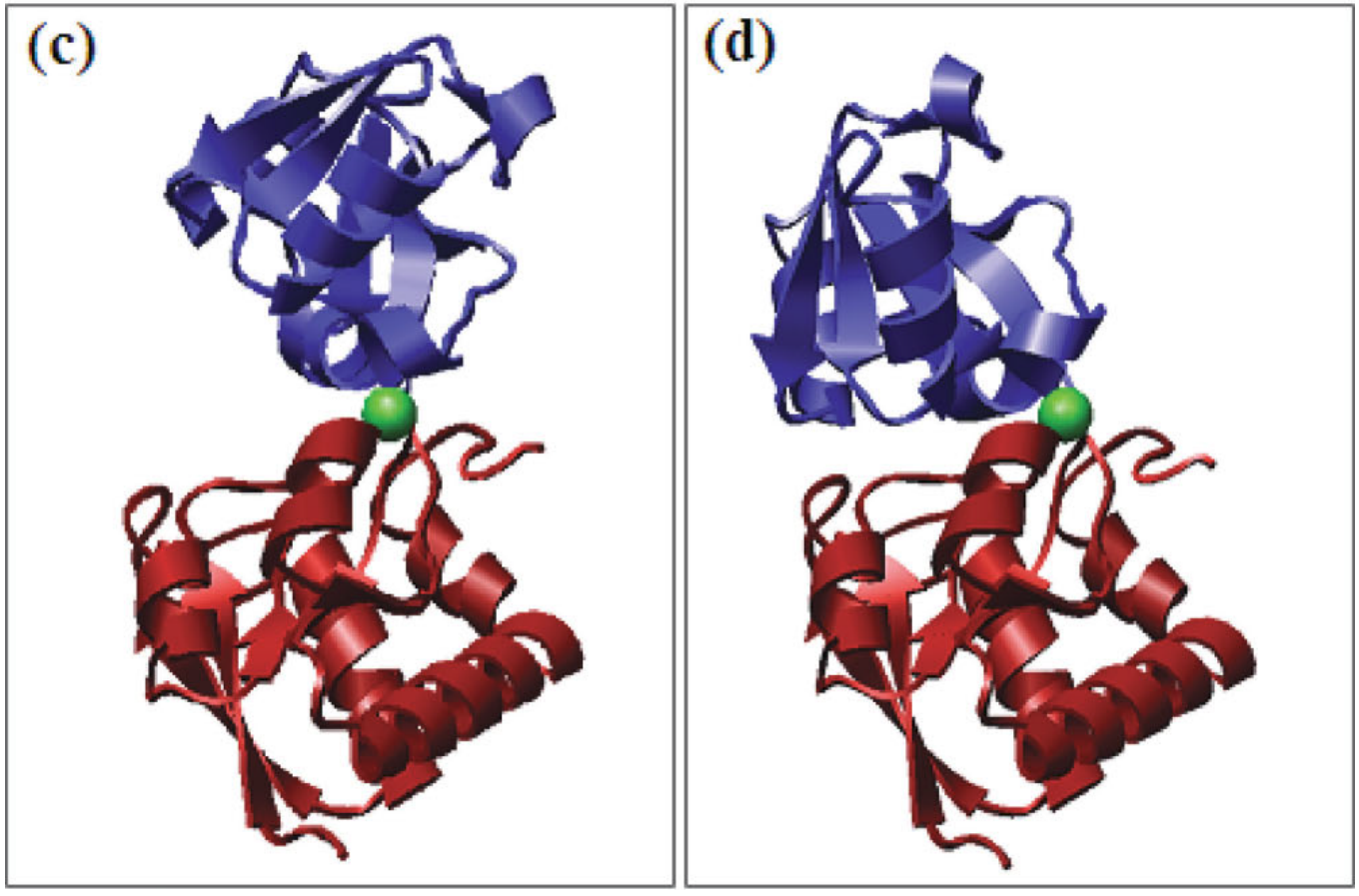
Shear motion



[Andrusier et al., 2008]

Large-scale domain motions

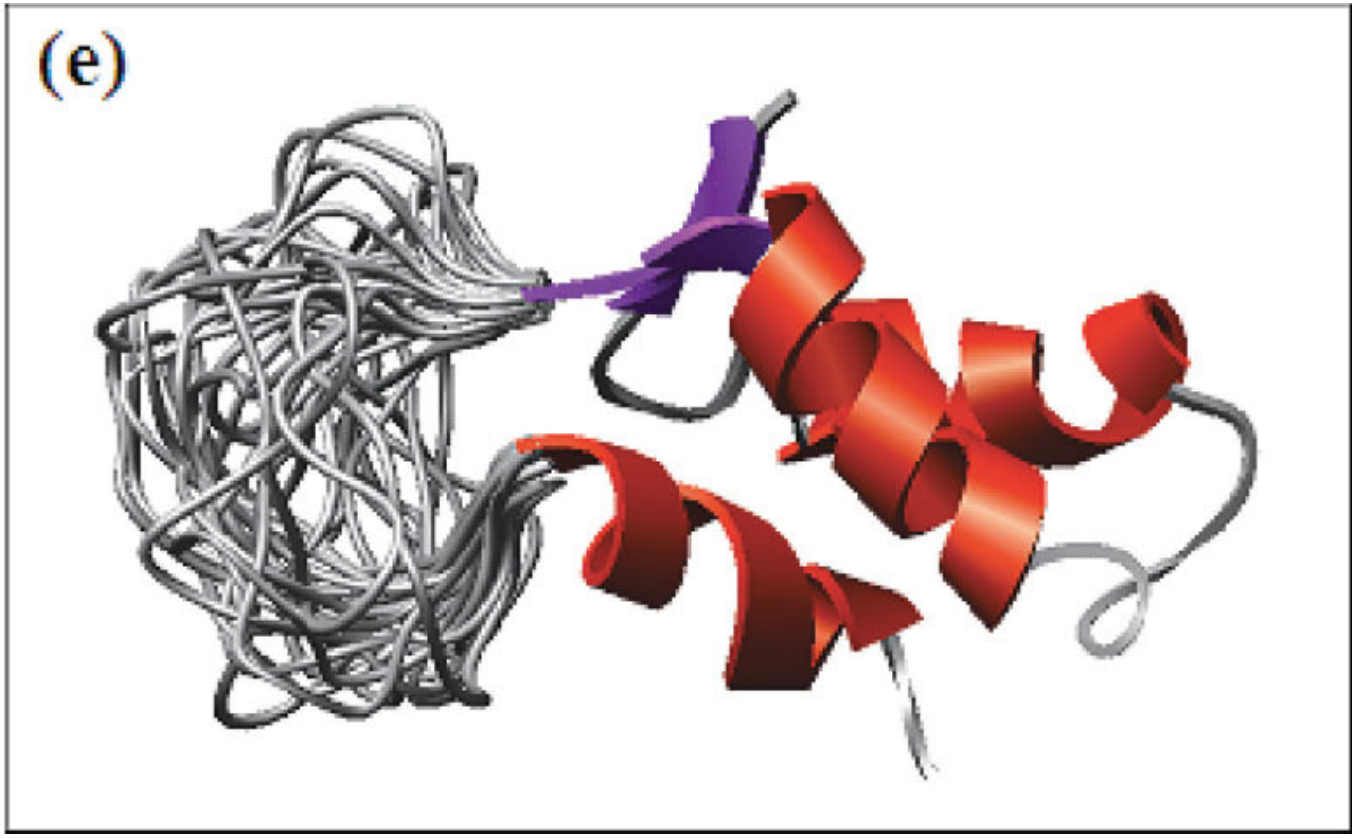
Hinge motion



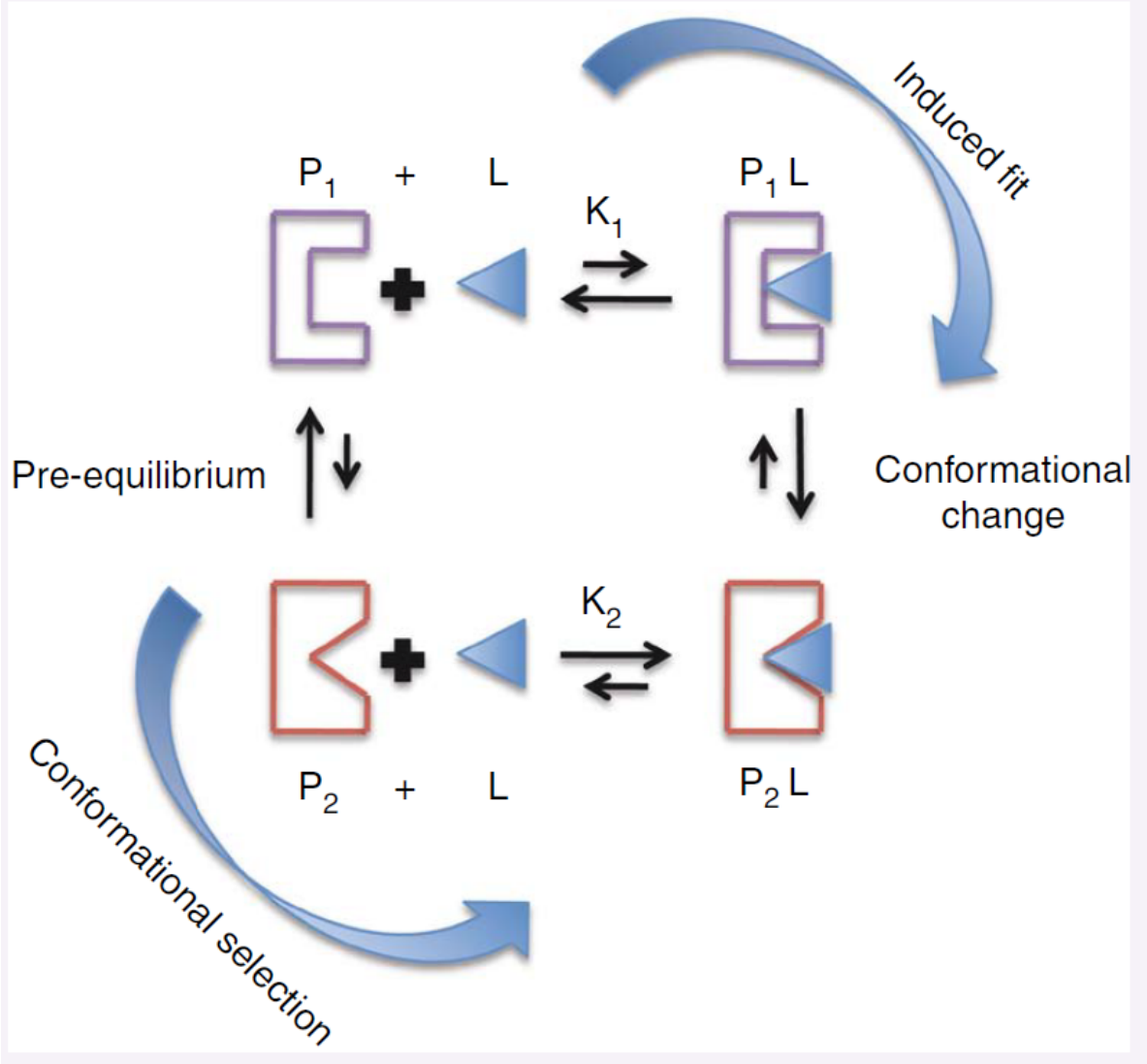
[Andrusier et al., 2008]

Disordered regions

Flexible loop



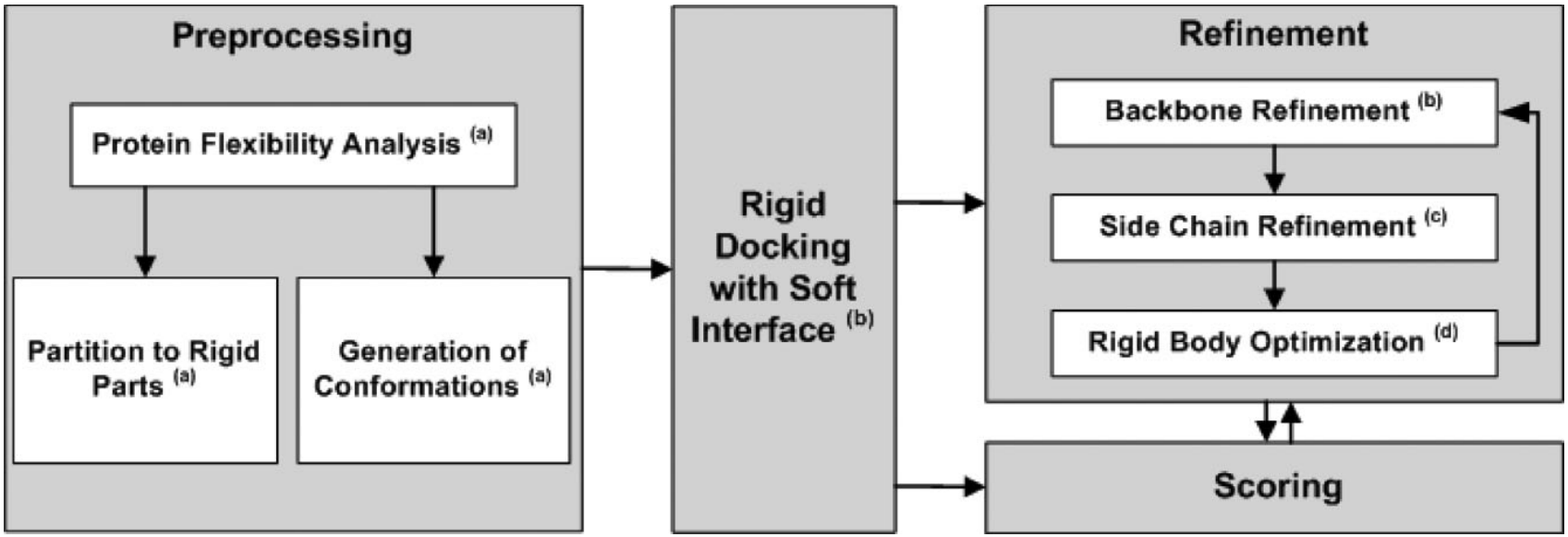
[Andrusier et al., 2008]



[Boehr et al., 2009]

Four major stages

1. Preprocessing => conformational ensemble / selection
2. Rigid body "soft"-docking
3. Refinement => induced fit
4. Scoring



[Andrusier et al., 2008]

Step 1: Flexibility analysis

Methods can be grouped in three major categories:

1. Generate an ensemble of discrete conformations
   - Conformational ensemble analysis of solved structures
   - Snapshots of Molecular Dynamics (MD) simulations
2. Continuous protein conformational space
   - Normal Modes Analysis (NMA)
   - Essential Dynamics
3. Identification of rigid and flexible regions
   - Rigidity theory
   - Hinge detection algorithms

[Andrusier et al., 2008]

Step 1: Flexibility analysis

Conformational ensemble analysis

- Instead of a single unbound structure use an ensemble of slightly different unbound structures
- Use experimentally solved 3D-structures of different conformations of the same protein or homologs
  - Morphing techniques: linear interpolation, with limited biological relevance
  - Detect rigid domains and hinge locations (DynDom, HingeFind, FlexProt)

[Andrusier et al., 2008]

Step 1: Flexibility analysis

Molecular Dynamics (MD)

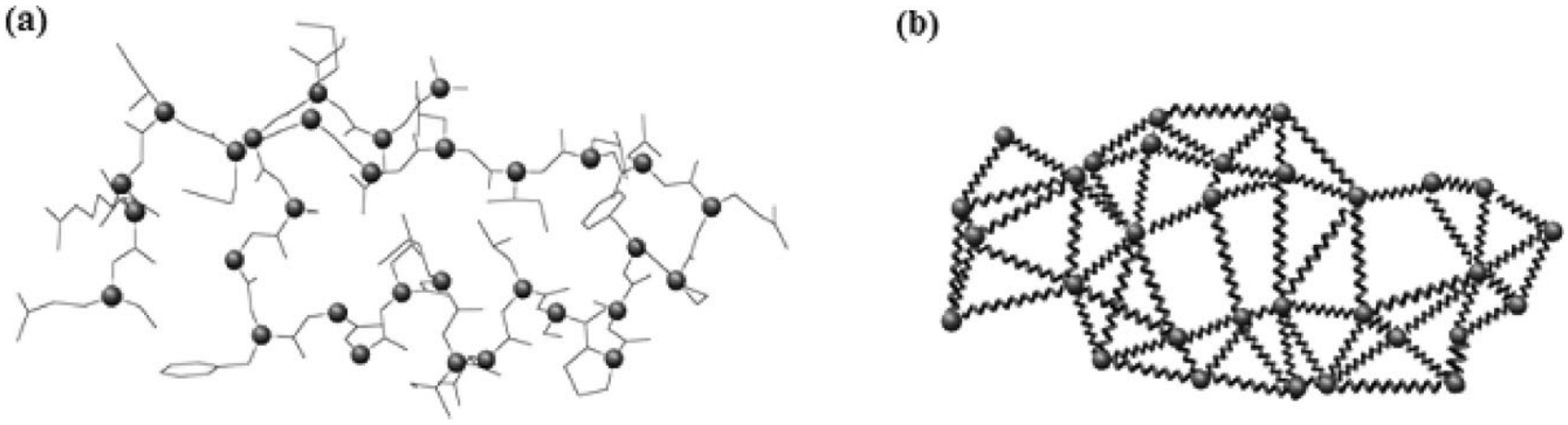
Problems and solutions in using MD:

- MD simulates only small-scale movements (ns timescale)
- Protein conformational changes take up to 1 ms
  - *Solutions:* restricting degrees of freedom (ex: torsional space)
- Energy barriers may trap the MD simulation in certain conformations
  - *Solutions:* Simulated annealing (ex: HADDOCK), scaling methods, biased methods, flooding technique (used in GROMACS), puddle-jumping

[Andrusier et al., 2008]

Step 1: Flexibility analysis

Normal Modes Analysis (NMA)



- (a) Polypeptide chain with C_α atoms as spheres
- (b) Simplified spring model

[Andrusier et al., 2008]

Normal Modes Analysis (NMA) - Models

- Goal: Study equilibrium fluctuations
- Common setup:
  - Simplified spring model which relies primarily on the geometry and mass distribution of a protein
  - every two atoms (or residues) within a distance below threshold are connected by a spring
  - all springs usually have a single force constant

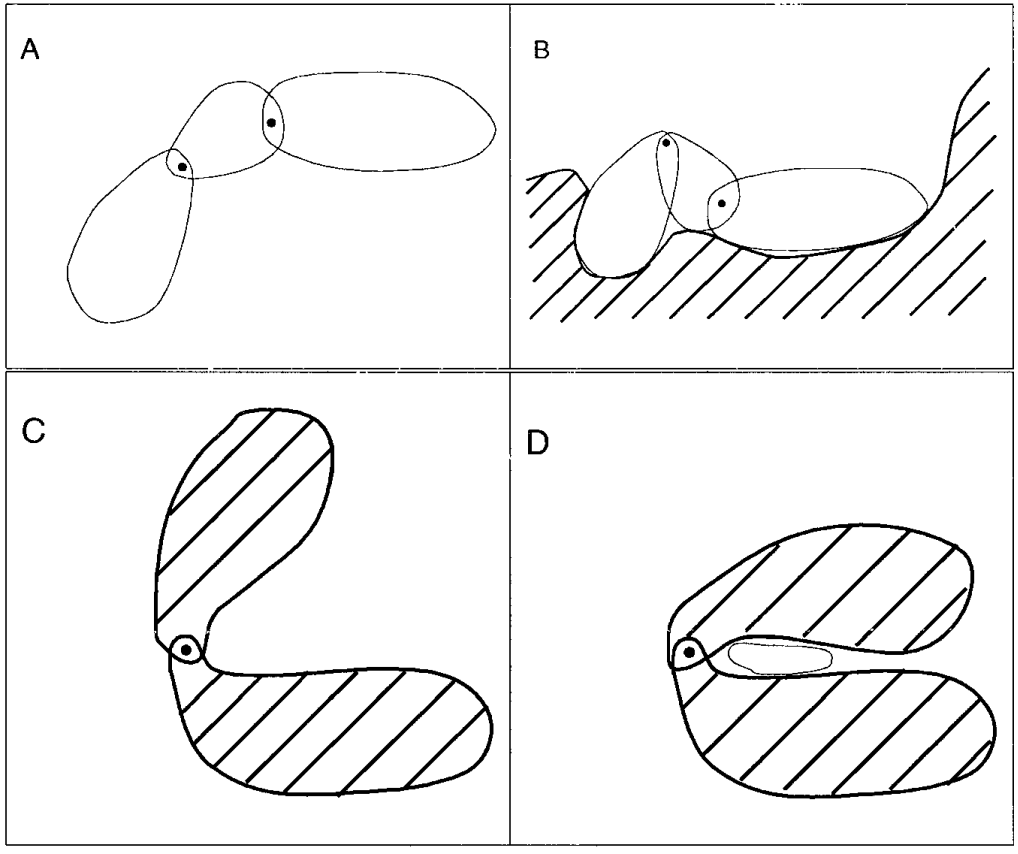
[Andrusier et al., 2008]

Normal Modes Analysis (NMA) - Models

- Gaussian Network Model (GNM)
  - Gaussian-distributed fluctuations about mean positions
  - Isotropic fluctuations
  - Coupling with harmonic potentials
  - Yields an analytical solution
  - Yields mean-square displacements and cross-correlations between fluctuations
  - Motion is projected to a mode space of N dimensions
- Anisotropic Network Model (ANM)
  - Extension of the GNM
  - Account for anisotropic fluctuations
  - Yields directional preferences
  - Motion is projected to a mode space of 3N-6 dimensions
  - More time-consuming than GNM

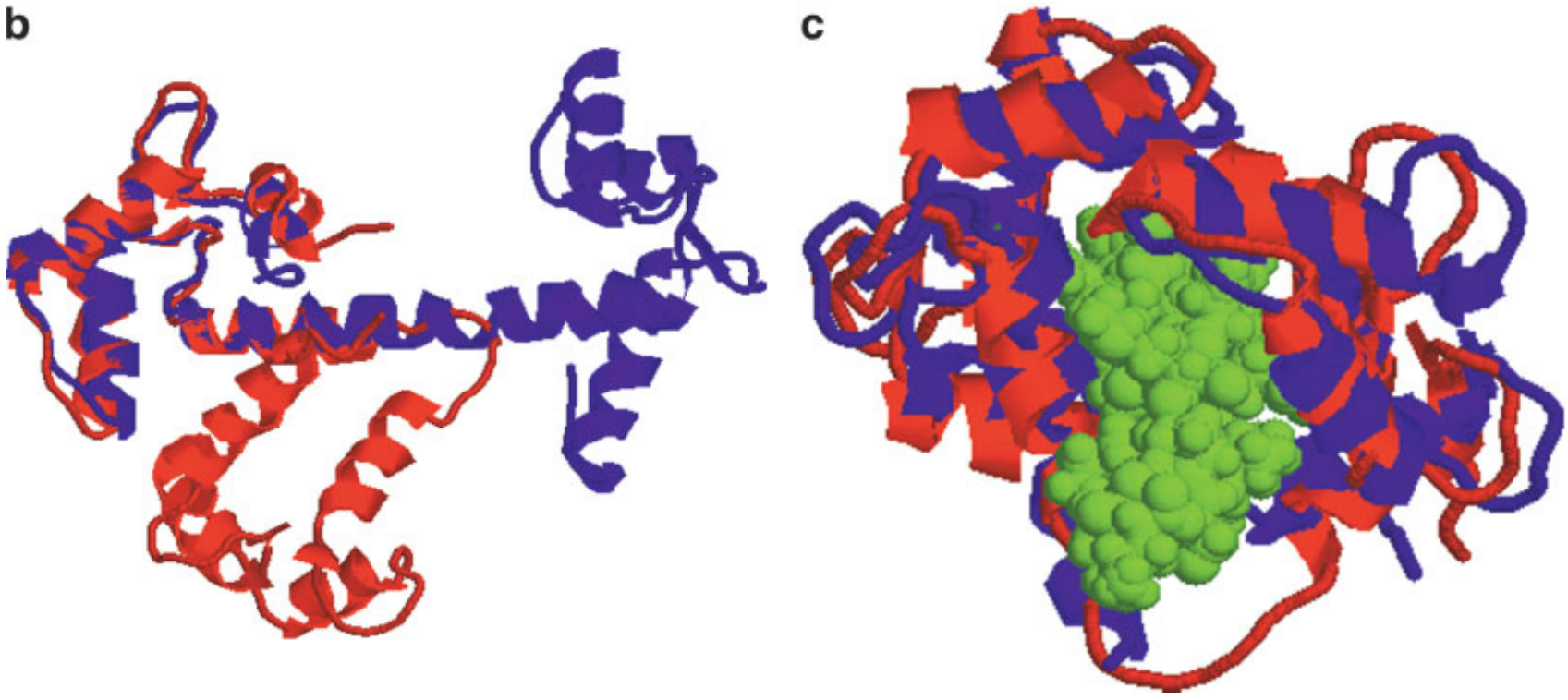
[Atilgan et al., 2001]

Hinges



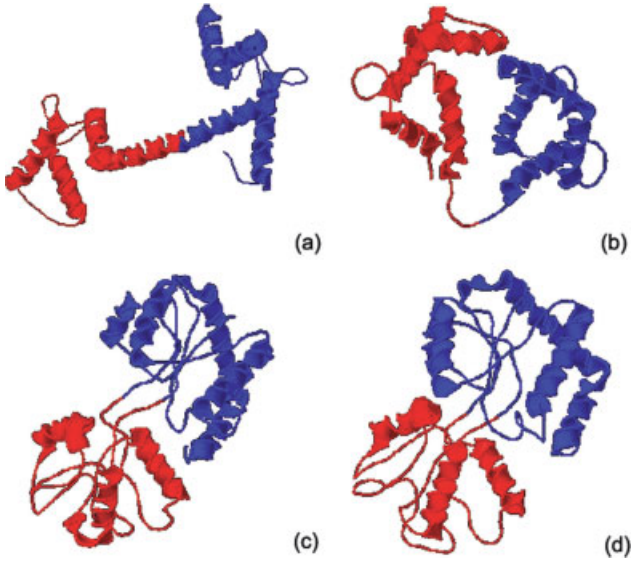
[Sandak et al., 1998]

Example: Calmodulin <=> myosin kinase peptide



[Schneidman-Duhovny et al., 2005]

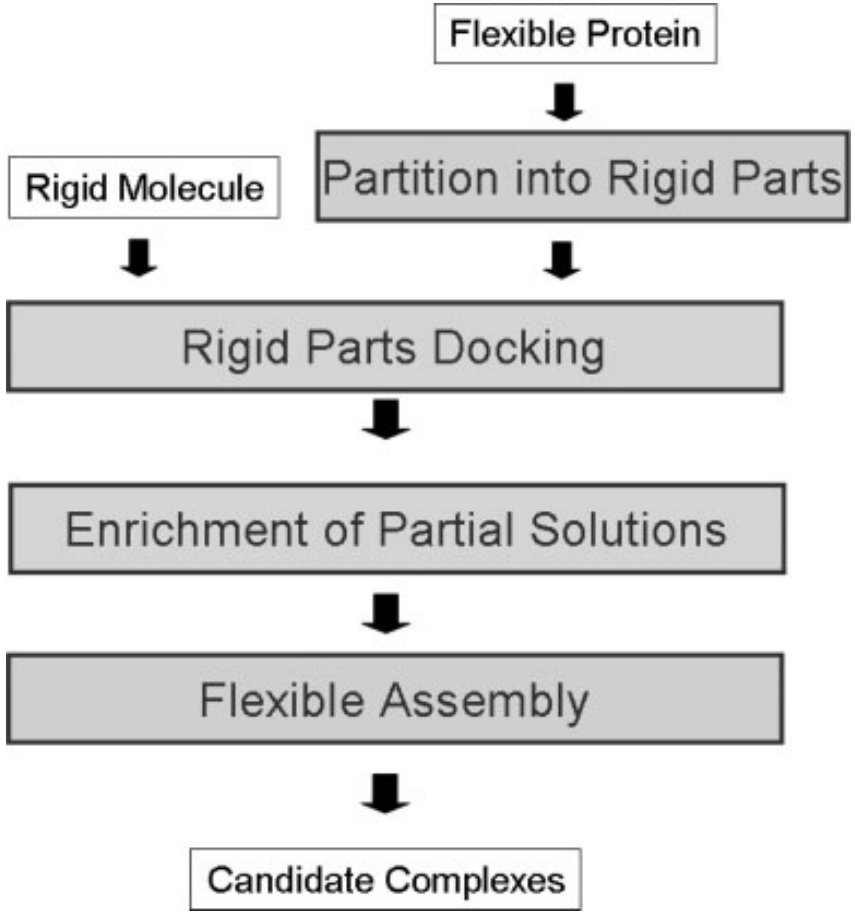
HingeProt



- Predicts locations of hinges and rigid parts
- HingeProt employs the Elastic (Gaussian) Network Model, based on normal mode analysis (NMA)
- Fully automated analysis of NMA results
- Using the two slowest modes, it calculates to correlation between the fluctuations of each pair of residues, that is their tendency to move in the same direction
- A change in the sign of the correlation value between two consecutive regions in the protein suggests a flexible joint that connects rigid units

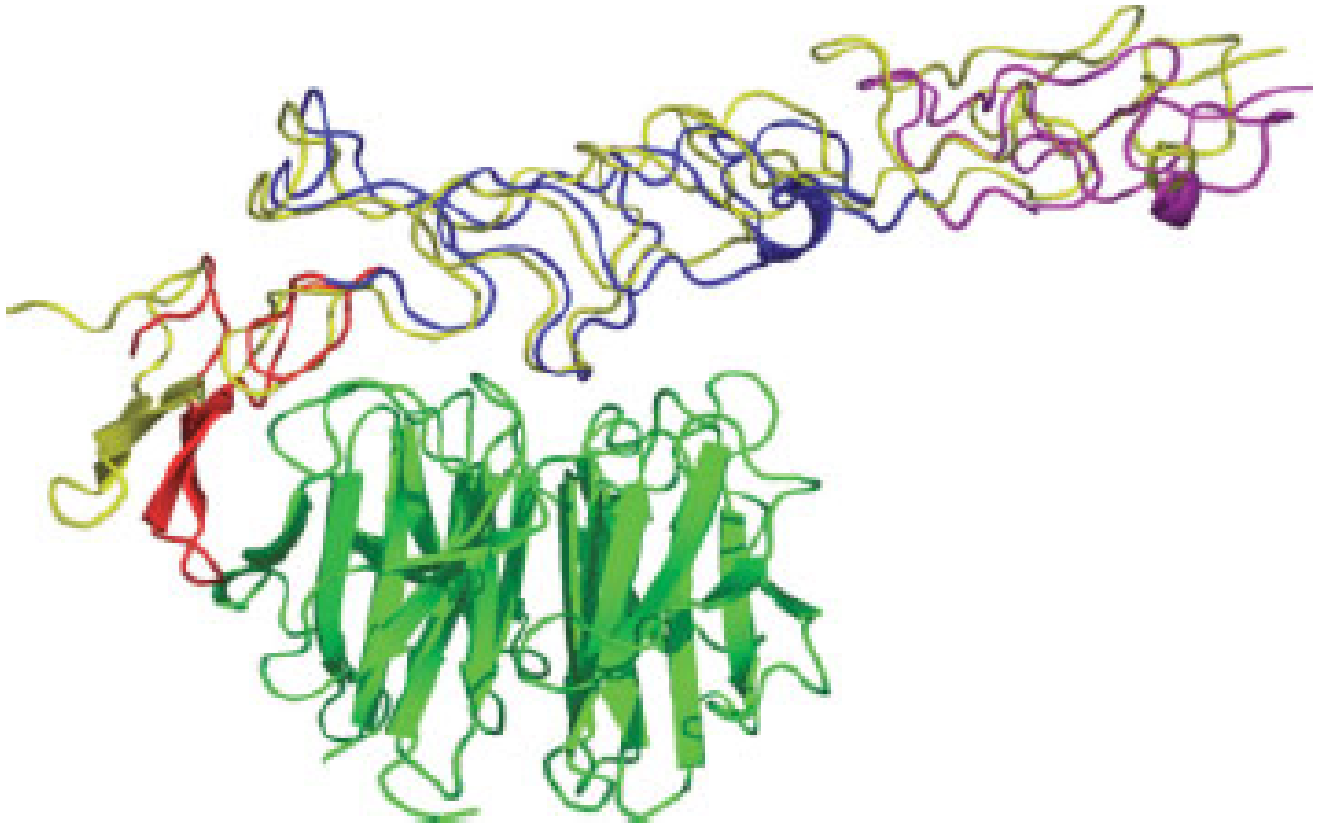
[Emekli et al., 2008], [Andrusier et al., 2008]

FlexDock



[Schneidman-Duhovny et al., 2007]

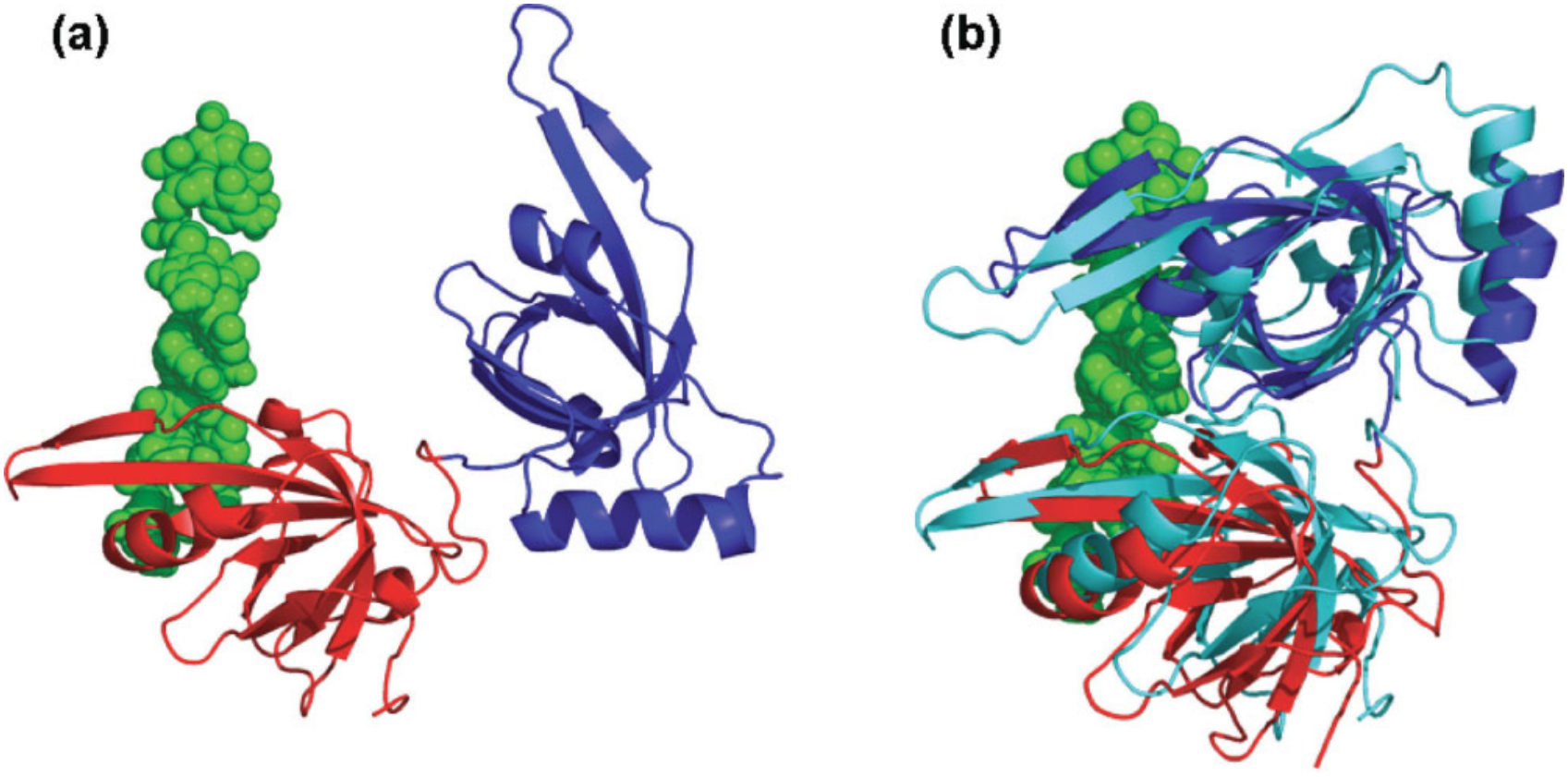
FlexDock - CAPRI target 8



FlexDock uses HingeProt to identify the hinges.

[Schneidman-Duhovny et al., 2007]

FlexDock - Replication Protein A (1FGU) + DNA



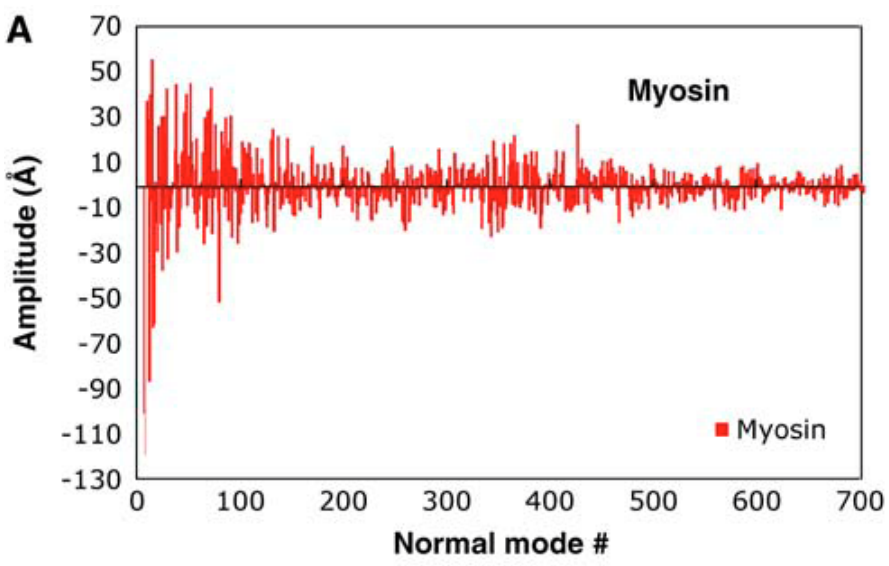
FlexDock uses HingeProt to identify the hinges.

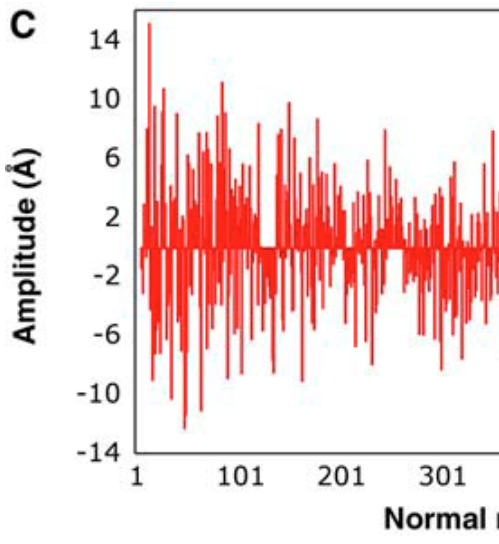
[Schneidman-Duhovny et al., 2007]

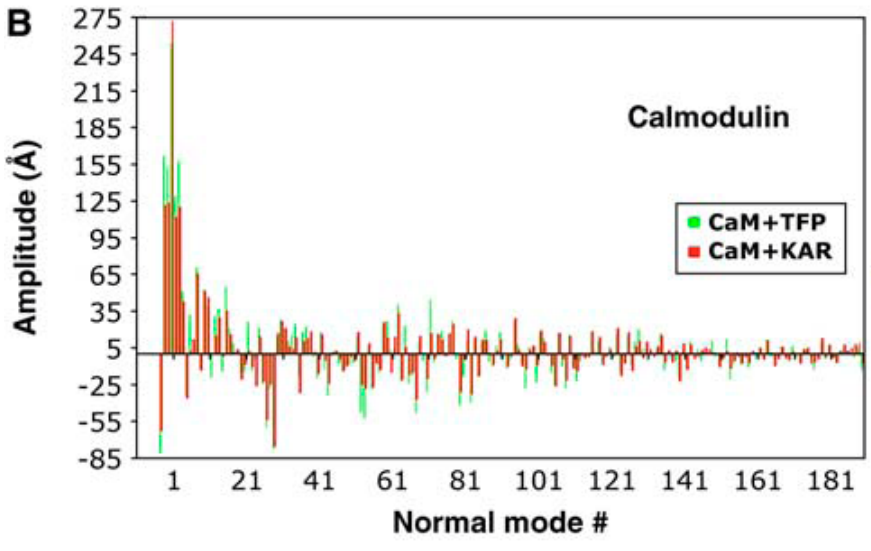
Normal Modes Analysis (NMA) - cautions

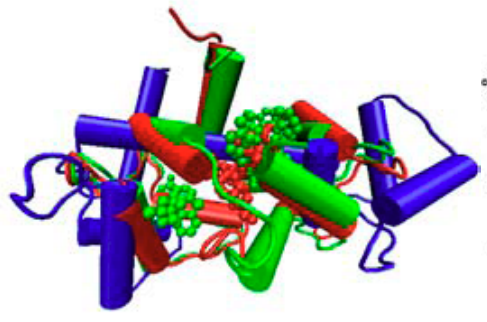
- When bound to a structure, a ligand can :
  - stabilize a conformation that is generally unpopulated in the ligand-free state
  - stretch the structure along the direction of certain normal modes that were irrelevant in the unbound state
- => Difficulty to predict which modes are relevant
- => Use as many modes as possible
- Fortunately, in the majority of cases ligand binding perturbs a system along its lower-frequency normal modes

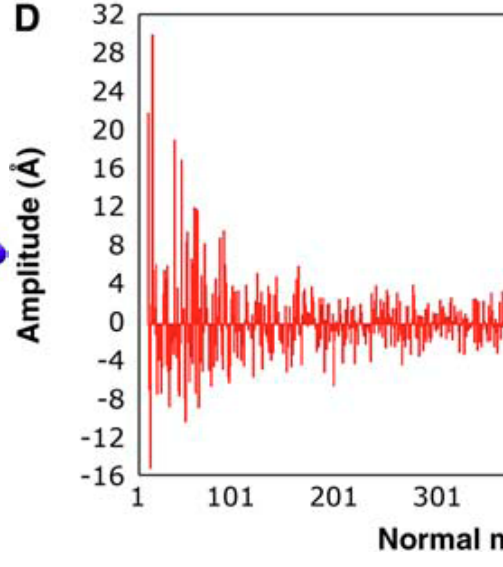
[Petrone and Pande, 2006]











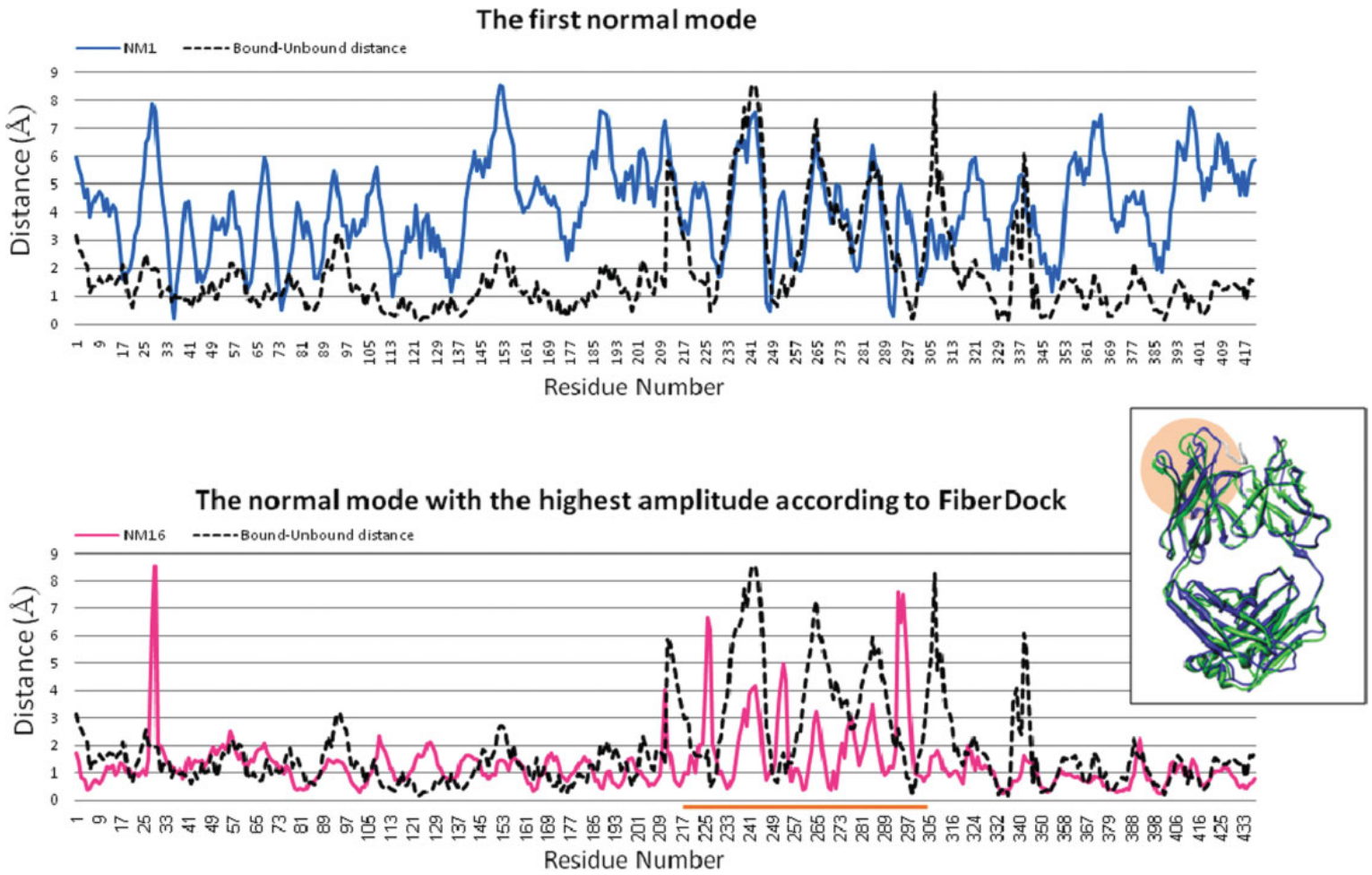
[Petrone and Pande, 2006]

Normal Modes Analysis (NMA) - loops

- Binding site of proteins often contains loops which undergo relatively small conformational changes triggered by an interaction (ex: protein kinase binding pockets)
- Loop movements can only be characterized by high-frequency normal modes
- Cavasotto et al. developed a method for measuring the relevance of a mode to a certain loop
- Goal: Flexible ligand (here small molecule) - flexible receptor docking

[Cavasotto et al., 2005]

Normal Modes Analysis (NMA) - FiberDock



[Mashiach et al., 2010a], [Mashiach et al., 2010b]

Step 1: Flexibility analysis

Methods can be grouped in three major categories:

1. Generate an ensemble of discrete conformations
   - Conformational ensemble analysis of solved structures
   - Snapshots of Molecular Dynamics (MD) simulations
2. Continuous protein conformational space
   - Normal Modes Analysis (NMA)
   - **Essential Dynamics**
3. Identification of rigid and flexible regions
   - Rigidity theory
   - Hinge detection algorithms

[Andrusier et al., 2008]

Step 1: Flexibility analysis

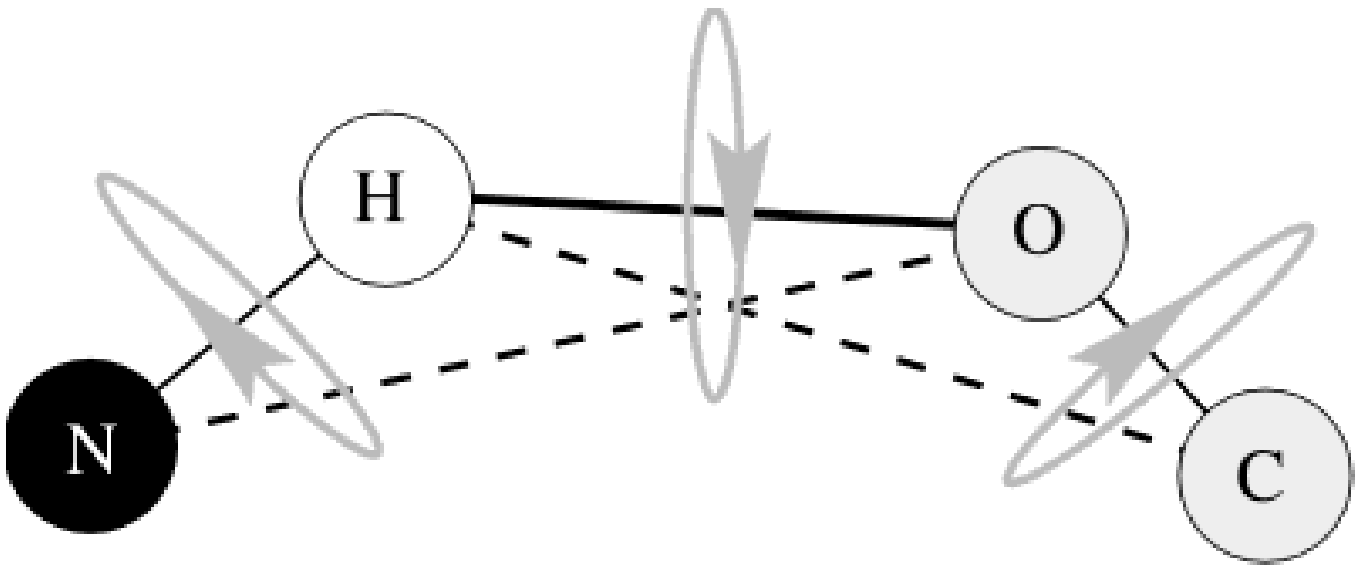
Essential dynamics

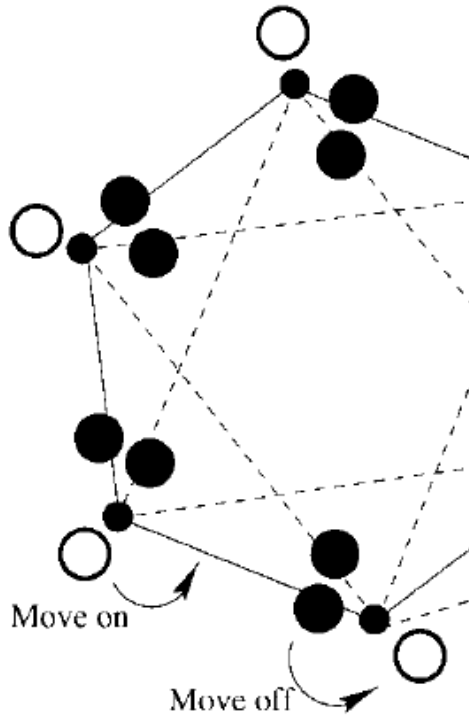
- Capture the main flexible degrees of freedom of a protein, given a set of its feasible conformations
- Degrees of freedom are described by vectors, called *essential modes* or principal components (PC)
- Set of conformations => (3N x 3N) covariance matrix (N = number of atoms) of the deviation of each atom from its average position
- Matrix is diagonalized => eigenvectors (= PC of flexibility), eigenvalues (= amplitude)
- Applied by Ritchie et al. (LORIA Nancy)

[Andrusier et al., 2008], [Mustard and Ritchie, 2005]

Step 1: Flexibility analysis

Rigidity theory





Pebble (=galet) game on graph

[Jacobs et al., 2001]

http://gepard.bioinformatik.uni-saarland.de/old_html/html/ProSeminarWS0607/JanChristoph/ProteinFlexibilityPredictions_JanChristoph.pdf

Methods for flexibility analysis

Table I
Some Methods for Flexibility Analysis

Method	Flexibility type	Description
DynDom[17]	Hinge bending	Given two conformations, clusters rotation vectors of short backbone segments and detects the rigid domains.
HingeFind[18]	Hinge bending	Compares given conformational states using sequence alignment and detects hinge locations.
FlexProt[20,21]	Hinge bending	Compares given conformational states, preforms structural alignment and detects hinge locations.
HingeProt[48]	Hinge bending	Detects hinge locations using GNM.
CONCOORD[61]	General flexibility	Generates conformations that fulfill distance constraints.
Dynamite[63]	General flexibility	Generates conformations using the essential dynamics approach.
FIRST[65]	General flexibility	Identifies rigid and flexible substructures using Rigidity Theory.

[Andrusier et al., 2008]

Backbone flexibility

Four groups of methods:

1. Soft interface
2. Ensemble docking
3. Hinge bending motions
4. Heuristic search for energetically favored conformations in a wide conformational space

[Andrusier et al., 2008]

Backbone flexibility

Soft interface

- Rigid-body docking which allows a certain amount of steric clashes
- Accounts only for side chain flexibility and small scale backbone movements
- Assumes that the proteins are capable of performing the required conformational changes which avoid the steric clashes
- The actual changes are not modeled explicitly
- Results of soft docking usually contain steric clashes => need further refinement

Three major groups:

1. Brute force techniques speeded up by FFT
2. Randomized methods
3. Shape complementary methods

[Andrusier et al., 2008]

Backbone flexibility

Ensemble docking

- Prior to docking: generate an ensemble of conformations for the binding partners
- Docking of the whole ensemble:
  1. Cross-docking: dock one-by-one
  2. Dock all together: mean-field approach

[Andrusier et al., 2008]

Ensemble docking

Mean-field MC2 (Multi-Copy/Monte-Carlo) method

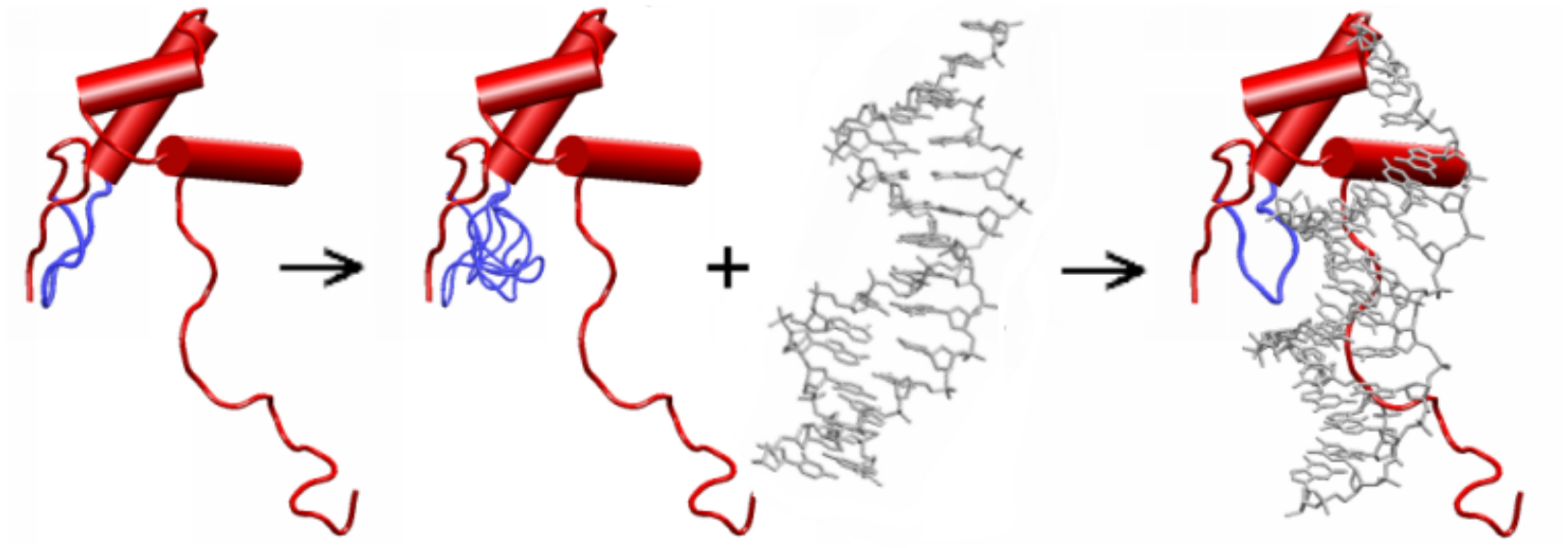
- Predict conformation of flexible loop in the interface
- Multiple copy representation of the loop
- Side-chains conformations are samples by Monte Carlo Simulated Annealing process
- Multiple copy representation and Monte Carlo simulation are coupled via copy weights
- Initially equal, these weights are recalculated at the end of each Monte Carlo cycle
- A unique loop copy is selected at the end
- Introduced into ATTRACT docking program

[Bastard et al., 2003], [Bastard et al., 2006]

csb.stanford.edu/karine/thesis-k-bastard.pdf

Ensemble docking

Mean-field MC2 (Multi-Copy/Monte-Carlo) method



http://www.ibpc.fr/chantal/www/MC2/mc2.html

Backbone flexibility

Heuristic search

- Energy minimization + normal modes (ATTRACT)
- Flexibility tree: hierarchical data structure which represents conformational sub-spaces of proteins and full flexibility of small ligands
- Monte Carlo methods:
  - Monte Carlo minimization (MCM) used in RosettaDock

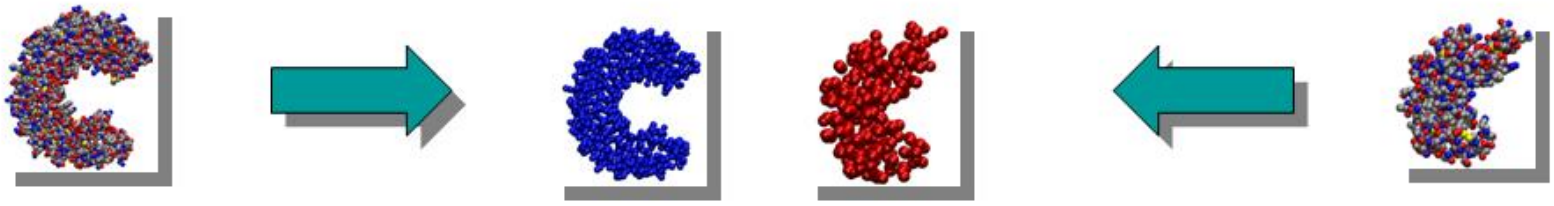
[Andrusier et al., 2008]

Example: ATTRACT

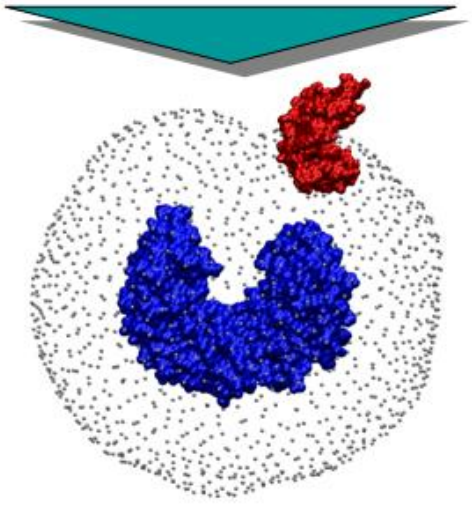
- Coarse grained: Three pseudo atoms per amino acid residue
- Side-chain flexibility: multicopy approach

[Zacharias, 2003]

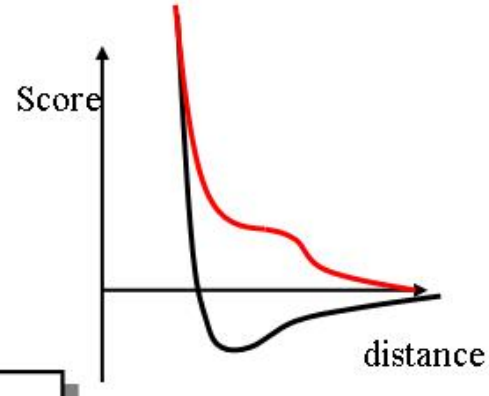
The ATTRACT approach



- 31 LJ-atom types
- Real charges



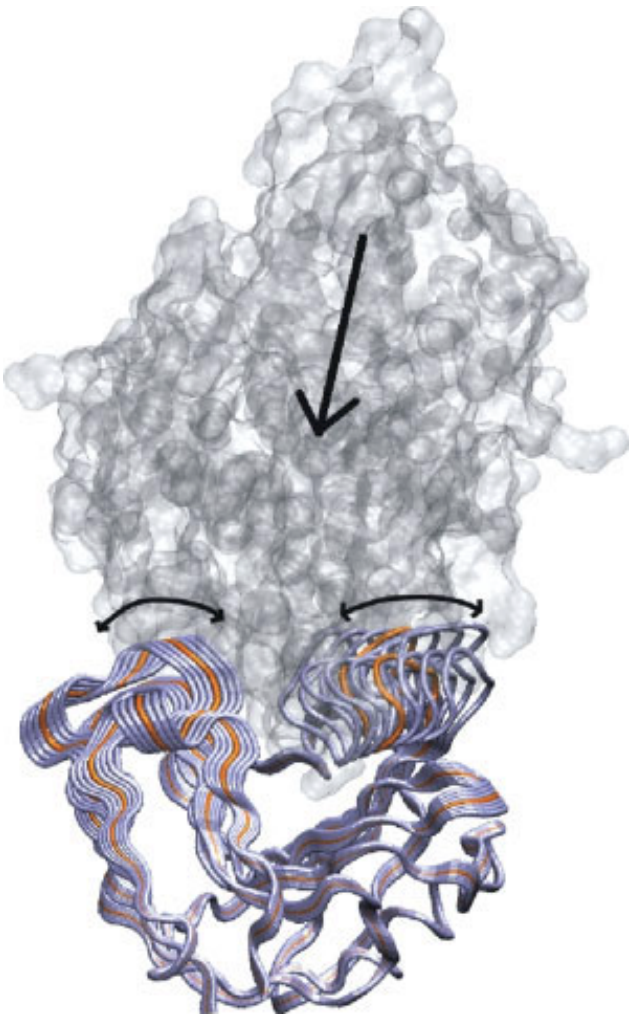
$$V = \sum \left(\frac{R_{AB}}{r_{ij}} \right)^8 - \left(\frac{R_{AB}}{r_{ij}} \right)^6 + \frac{q_i q_j}{\varepsilon r_{ij}}$$



Multi-start systematic search by Energy Minimization

Zacharias, Protein Science. 2003, 12, 1271.

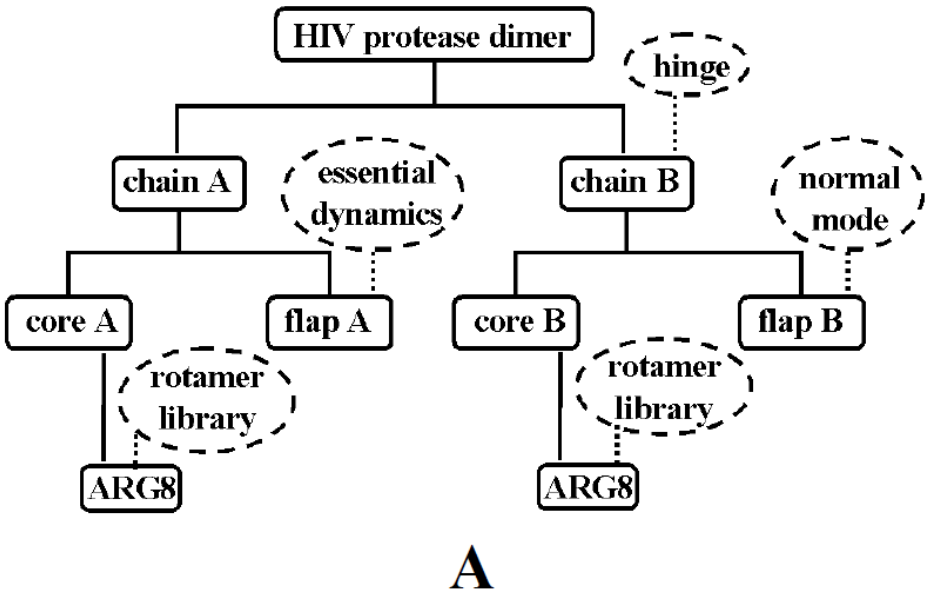
Energy minimization in low-frequency normal modes

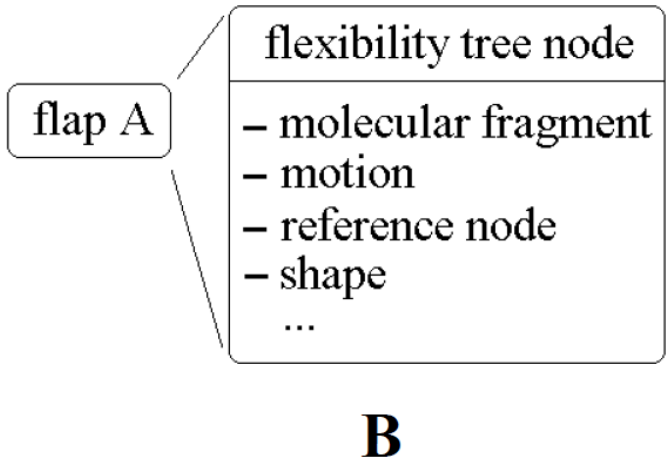


- Docking in 6 + n-dimensional space (n is the number of modes (up to five) + 6 rotational and translational degrees of freedom)
- About 300000 starting structures

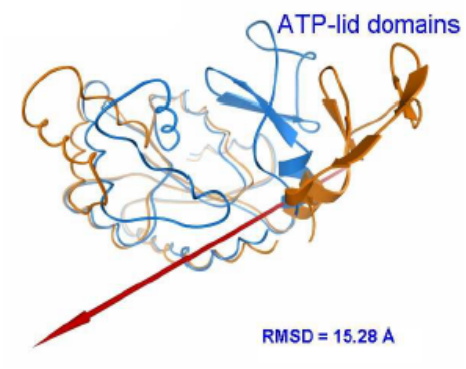
[May and Zacharias, 2008]

Flexibility Tree (FT)

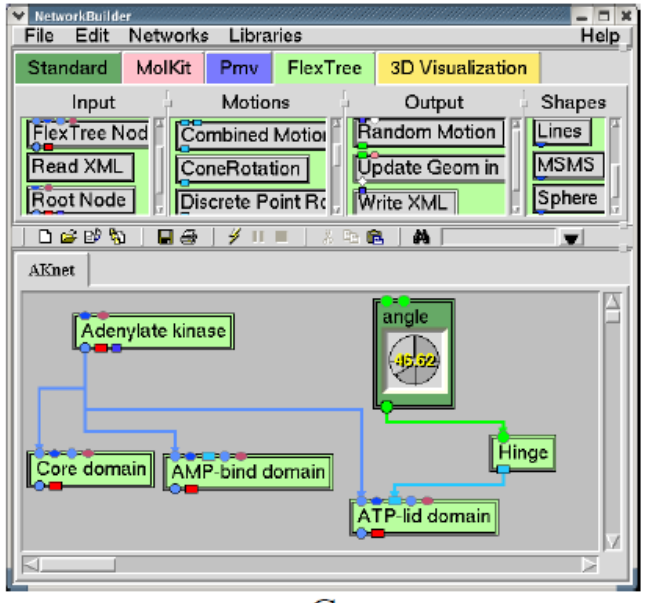




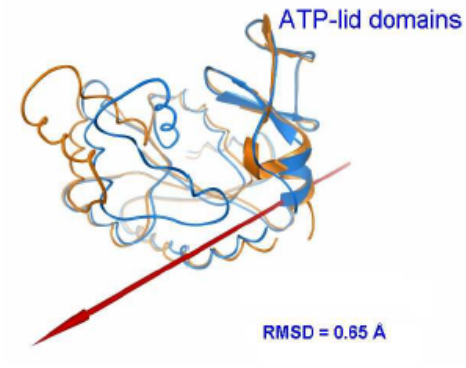
Used in FLIPDock [Zhao et al., 2006], [Zhao and Sanner, 2007]


ATP-lid domains
RMSD = 15.28 Å

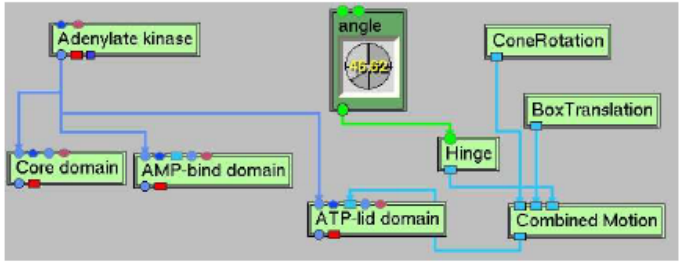
A


NetworkBuilder
File Edit Networks Libraries Help
Standard MolKit Pmv FlexTree 3D Visualization
Input Motions Output Shapes
FlexTree Nod
Read XML
Root Node
Combined Motio
ConeRotation
Discrete Point Rc
Random Motion
Update Geom in
Write XML
Lines
MSMS
Sphere
AKnet
Adenylate kinase
angle
45.62
Hinge
Core domain
AMP-bind domain
ATP-lid domain

C


ATP-lid domains
RMSD = 0.65 Å

B


Adenylate kinase
angle
45.62
ConeRotation
BoxTranslation
Hinge
Core domain
AMP-bind domain
ATP-lid domain
Combined Motion

D

Monte Carlo minimization (MCM) used in RosettaDock

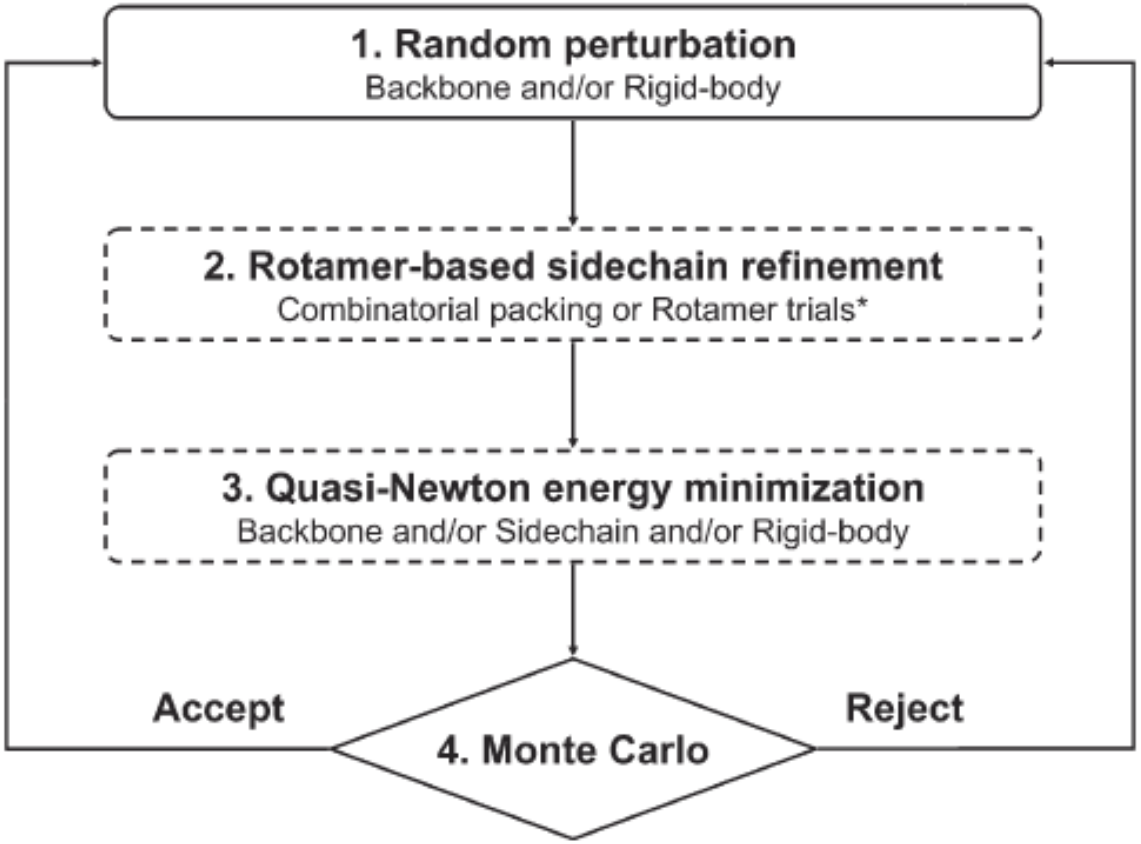


Table 1. Fold-tree-based sampling strategies fo

Modeling task	Fold tree
Protein folding	1a
Domain assembly	1b
Loop modeling	1c
Fixed-backbone docking	1d
Composite type	
Docking with backbone relaxation	1e
Folding and docking	1e
Docking with small hinge motion	1f
Docking with large hinge motion	1f
Docking with loop refinement	1g
Docking with loop rebuilding	1g

The flexible regions in the fold trees in Fig. 1 can b
combinations of fold trees with sampling strategies a

[Wang et al., 2007]

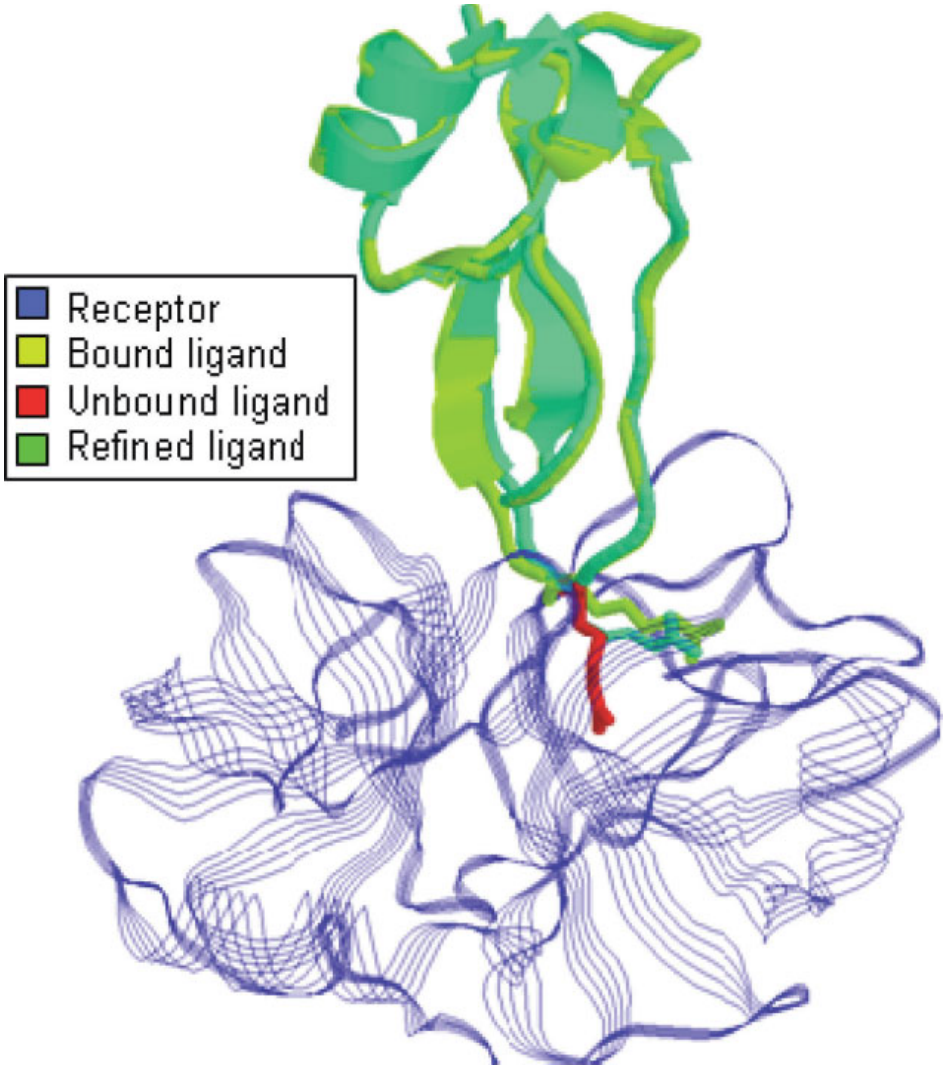
Methods for docking with backbone flexibility

Table II
Some Methods for Docking with Backbone Flexibility

Method	Flexibility type	Description
MC2[81]	Flexible loops	Chooses the best loop conformations from an ensemble using the Mean-Field approach.
ATTRACT[51,83]	Flexible loops	Chooses the best loop conformations from an ensemble using the Mean-Field approach.
	General flexibility	Energy minimization on degrees of freedom derived from the lowest frequency normal modes.
FlexDock[86]	Hinge bending	Allows hinge bending in the docking. The rigid subdomains are docked separately and consistent results are assembled.
FLIPDock[92]	General flexibility	Searches favored conformations by a genetic algorithm and a divide and conquer approach. Uses FT data structure.
HADDOCK[32,33]	General flexibility	Handles backbone flexibility in the refinement stage, by simulated annealing MD.
RosettaDock[10,93,118]	General flexibility	Handles backbone flexibility in the refinement stage, by Monte Carlo minimization.

[Andrusier et al., 2008]

Refinement of side-chains: FireDock



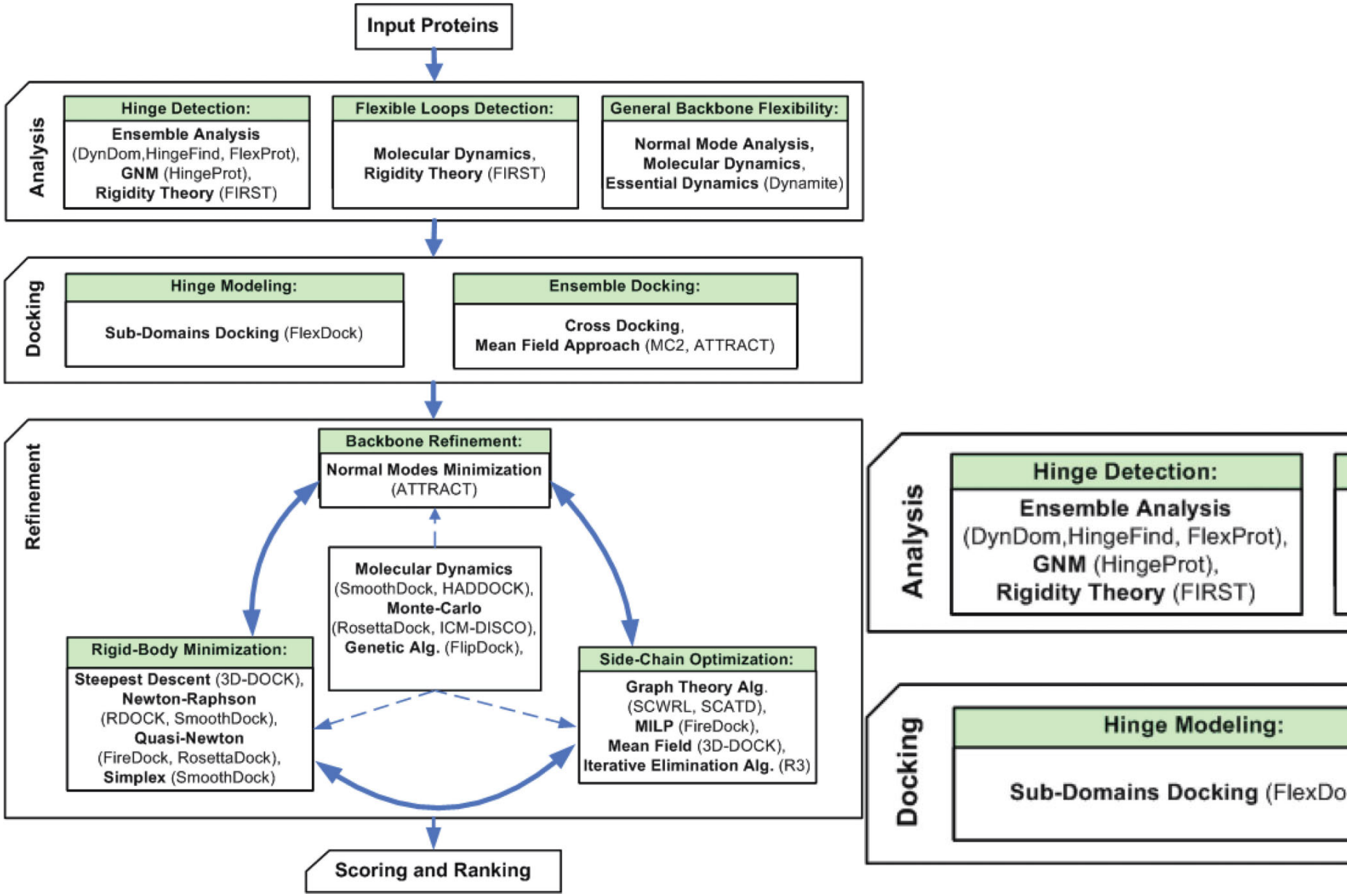
[Mashiach et al., 2008], [Andrusier et al., 2008], [Andrusier et al., 2007]

Docking and refinement methods with side-chain and rigid-body optimization

Table III
Some Docking and Refinement Methods with Side-Chain and Rigid-Body Optimization

Method	Side-chain flexibility	Rigid-body optimization	Scoring function terms
RosettaDock[10,93,118]	MC on rotamers and minimization of rotamer torsion angles	MC with DFP quasi-Newton minimization[147,148]	Linear repVdW, attrVdW, EEF1 (SASA), rotamer probability, hydrogen bonds, residue pair potentials, and electrostatics.
ICM-DISCO[128]	Biased probability MC on internal coordinates	Biased probability MC on internal coordinates	Truncated VdW, electrostatics, solvation, hydrogen bonds, and hydrophobicity.
3D-DOCK[121]	SCMF	Steepest-descent minimization[139]	VdW, electrostatics, and Langevin dipole salvation.
SmoothDock[11,129]	Pre-docking MD and ABNR minimization in the refinement	Simplex[142] and ABNR minimization	VdW, electrostatics, and ACE.
HADDOCK[32,33]	Simulated annealing MD	Steepest-descent minimization[139]	VdW, electrostatics, binding site restriction, and buried surface area.
RDOCK[149]	ABNR minimization	ABNR minimization	Electrostatics and ACE.
FireDock[113]	MILP	MC with BFGS quasi-Newton minimization[150,151]	Linear repVdW, attrVdW, ACE, electrostatics, π-stacking and aliphatic interactions, hydrogen and disulfide bonds, and insideness measure.

[Andrusier et al., 2008]



[Andrusier et al., 2008]

Assessing structural predictions in community-wide experiments:

CAPRI and CASP

- **CASP (Critical Assessment of methods of Structure Prediction)**:
  - predict the mode of **folding** of a protein based on the amino acid sequence
  - compare to an unpublished X-ray or NMR structure.
  - J. Moult (CARB, Rockville MD) launched CASP in 1994
  - round of predictions once every two years (CASP8 in 2008) with 50-100 targets
- **CAPRI (Critical Assessment of PRedicted Interactions)**:
  - predict the **mode of recognition** of two proteins by docking their 3D structures
  - compare to unpublished X-ray structures of **protein-protein complexes**.
  - CAPRI started in 2001
  - a round of prediction begins **any time a target is made available**

http://capri.ebi.ac.uk/

Running CAPRI

The Management Committee

Web site K. Henrick, S. Velenkar (EBI, Hinxton, UK)

Targets J. Janin (Orsay, France)

Assessors S. Wodak (Toronto), M. Lensink (Brussels)

M. Sternberg (Imperial College London)

S. Vajda (Boston University)

I. Vakser (Kansas University)

L. Ten Eyck (UC San Diego)

Special Issues of *Proteins: structure, fonction and bioinformatics*

1: Vol. 52-1, July 1, 2003

2: Vol. 60-2, Aug.1, 2005

3: Vol. 69-4, Dec. 2007

4: **Vol. 78, Nov. 15, 2010**

Evaluation meetings

La Londe des Maures, France Sept. 19-21, 2002

Gaeta, Italy, Dec. 8-10, 2004

Toronto, Canada, April 20-21, 2007

Barcelona, Spain, Dec. 9-11, 2009

CAPRI star evaluation

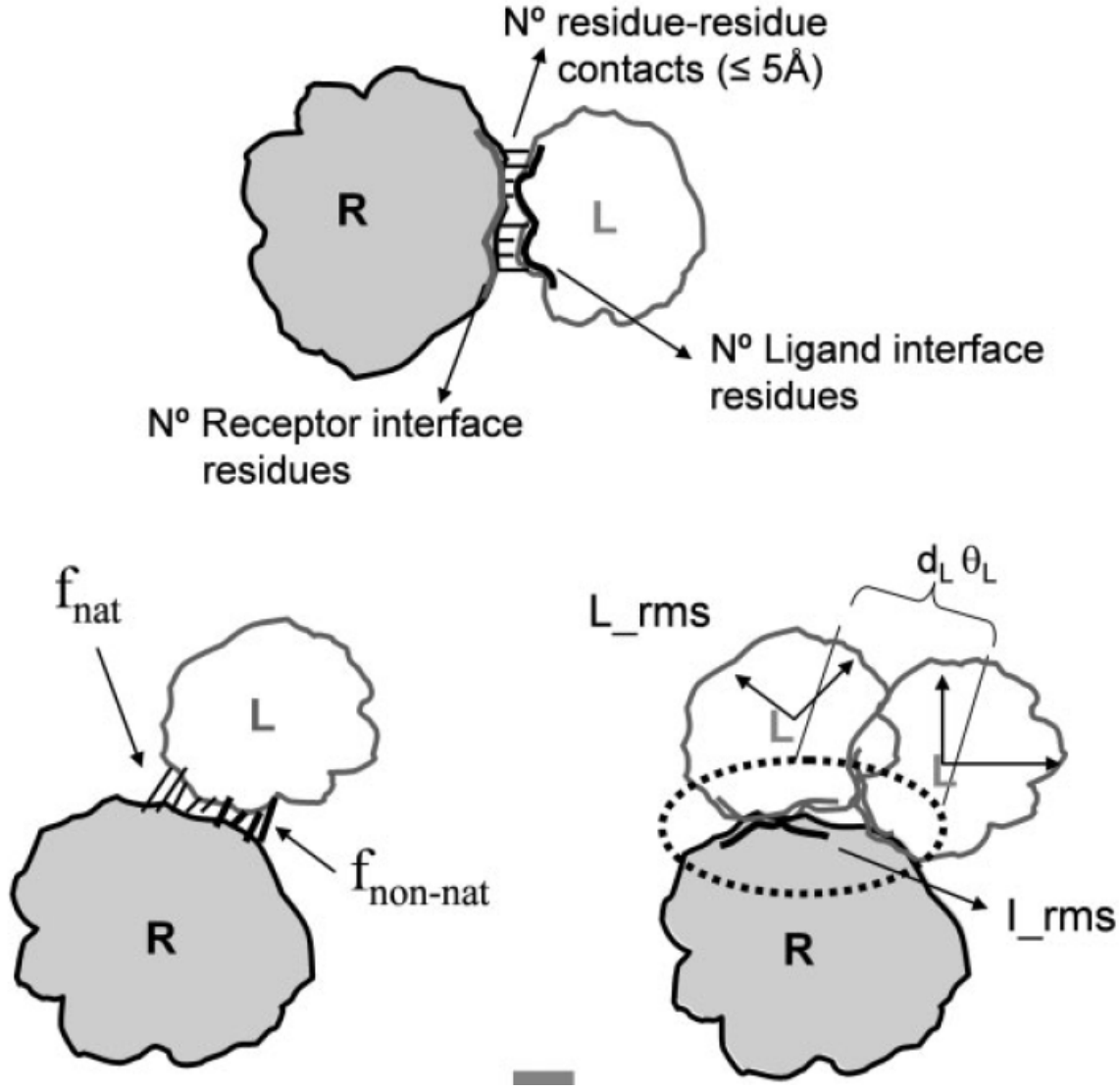
The CAPRI star system

Mendez, Leplae, Wodak 2003

Lensink et al. 2005, 2007, 2010

Model quality		% native contacts (correctly predicted residue pairs) f_{nc}	main chain RMSD (Å) Ligand L_{rms}		Interface I_{rms}
High	**(three-star)**	**> 50%**	< 1 Å	or	**< 1Å**
Good	**(two-star)**	**>** 30%	< 5	or	< 2
Acceptable	(one-star)	> 10%	< 10	or	< 4
Incorrect		< 10%	>10	and	> 4

Source: Janin, LIX 2010



[Méndez et al., 2005]

CAPRI rules

1. Each group gets the input structures (bound, unbound or sequence only).
2. Some weeks later they have to submit 10 models for the complex.
3. Exception: web-servers have to submit within 24h to prevent "human scoring".
4. The best model out of the 10 models is used to evaluate the performance of one group or web-server.
5. Group $\neq$ Program: each group can use the programs they like, but usually they are using their own programs.

Table III

Summary of Target Prediction Performance in CAPRI Rounds 13–19

			***			**			*		
	L-rms (Å)	R-rms (Å)	P	U	S	P	U	S	P	U	S
T29	1.7	**B**	0	2	1	9	78	13	8	87	13
T30	1.7	2.3	0	0	0	0	0	0	2	2	0
T32	0.3	2.1	15	0	0	13	3	0	6	12	2
T33	2.0	2.6	0	0	0	0	0	0	0	0	0
T34	2.0	**B**	0	0	0	25	13	4	40	165	26
T35	2.9	2.9	0	0	0	0	0	0	1	2	1
T36	2.9	**B**	0	0	0	0	0	0	1	0	0
T37	0.6	0.4	1	8	5	7	34	13	13	34	11
T38	3.2	1.9	0	0	0	0	0	0	0	0	0
T39	3.2	**B**	1	0	0	2	3	0	0	1	0
T40	**B**	0.4	79	176	39	54	163	40	31	149	13
T41	2.0	1.5	24	2	2	58	99	16	67	198	51
T42	1.5	1.5	9			5			6		

[Lensink and Wodak, 2010]

Web-server

Table V
Prediction Performance of Web-Servers

Target	29	30	32	33	34	35	36	37	38	39	40	41	42
ClusPro	0	0	0	0	1*	0	0	0	0	1**	2/1**	1**	1***
FiberDock												10/1***	0
FireDock			0	0	0	0	0	0	0	0	2/1***		
GRAMM-X	0	0	0	0	0	0	0	0	0	0	2***	1***	0
HADDOCK			0	0	7*	0	0	0	0	0	1***	4/1**	1*
SKE-DOCK	0	0	0	0	0	0	0	2*	0	0	2/1***	0	0
Top down								0	0		2/1**	0	0

[Lensink and Wodak, 2010]

Conclusion

Is the protein-protein docking problem solved ?
Not really:

- Final goal: best structure at first rank
- CAPRI results:
  - Best structure at top 10 => still up to 90% (worst case) false positives
  - No program works for all complexes
  - Bad performance of non-human scores, i.e. web-servers
  - Scores are only a first help for "human scorers"

Conclusion

Is the protein-protein docking problem solved ?
Challenges:

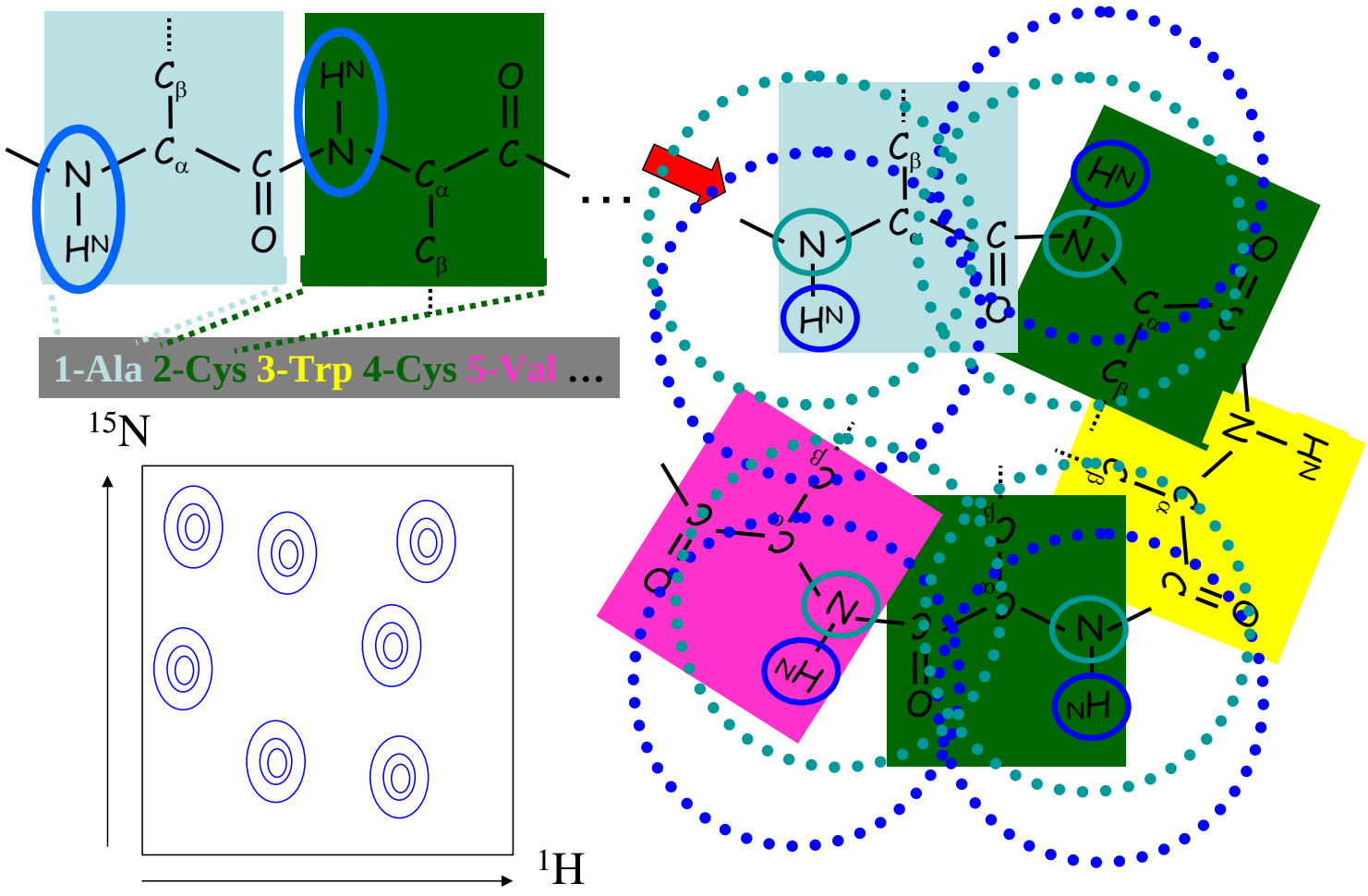
- Better sampling and scoring
- Conformational changes upon binding
- Predicting domain motions
- Folding upon binding
- Large scale docking => Interactome, Large molecular assemblies
- Predicting which proteins interact => Predicting binding affinities

Conclusion

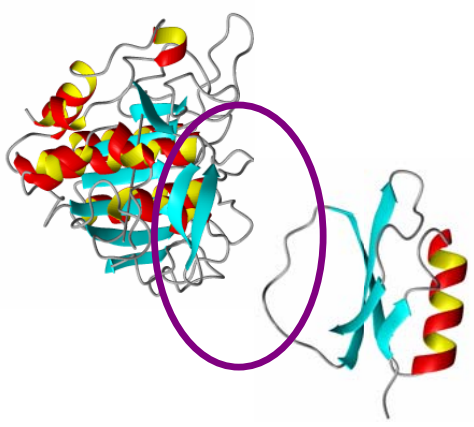
Is the protein-protein docking problem solved ?
Not really and a there are still a lot of challenges.
One possible solution:

- Combine docking with experimental data (NMR, mutagenesis, cryo-EM, SAXS, ...)

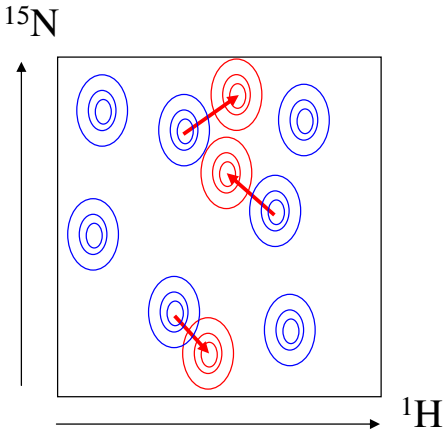
Chemical shift

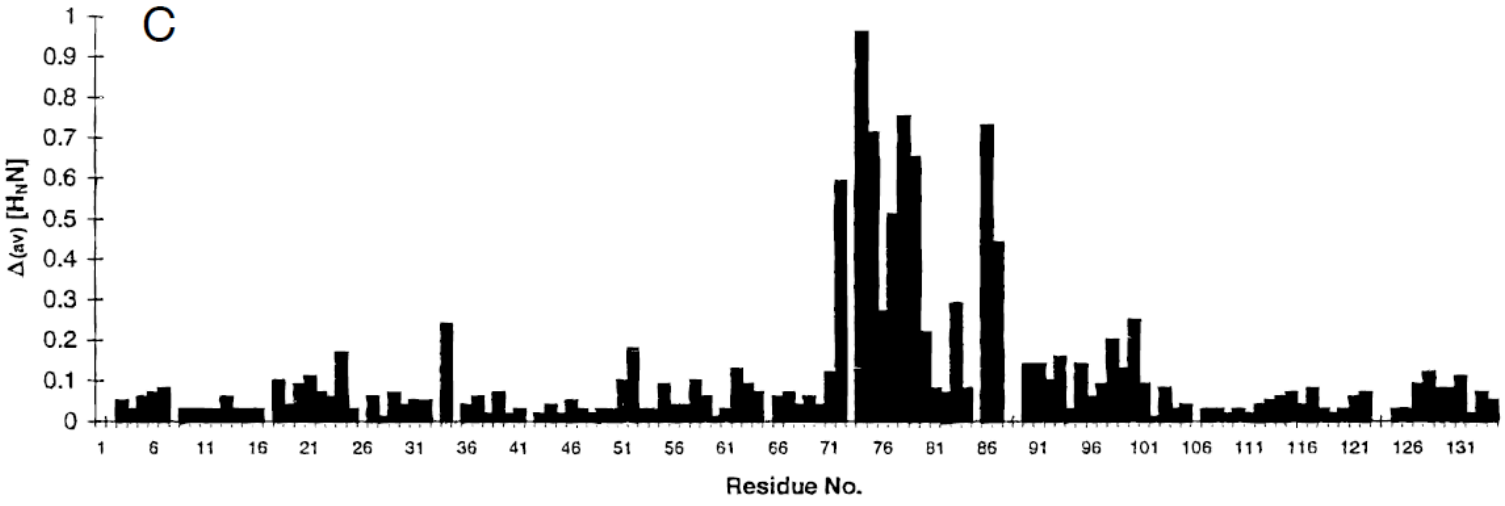


Chemical Shift Perturbation (CSP)

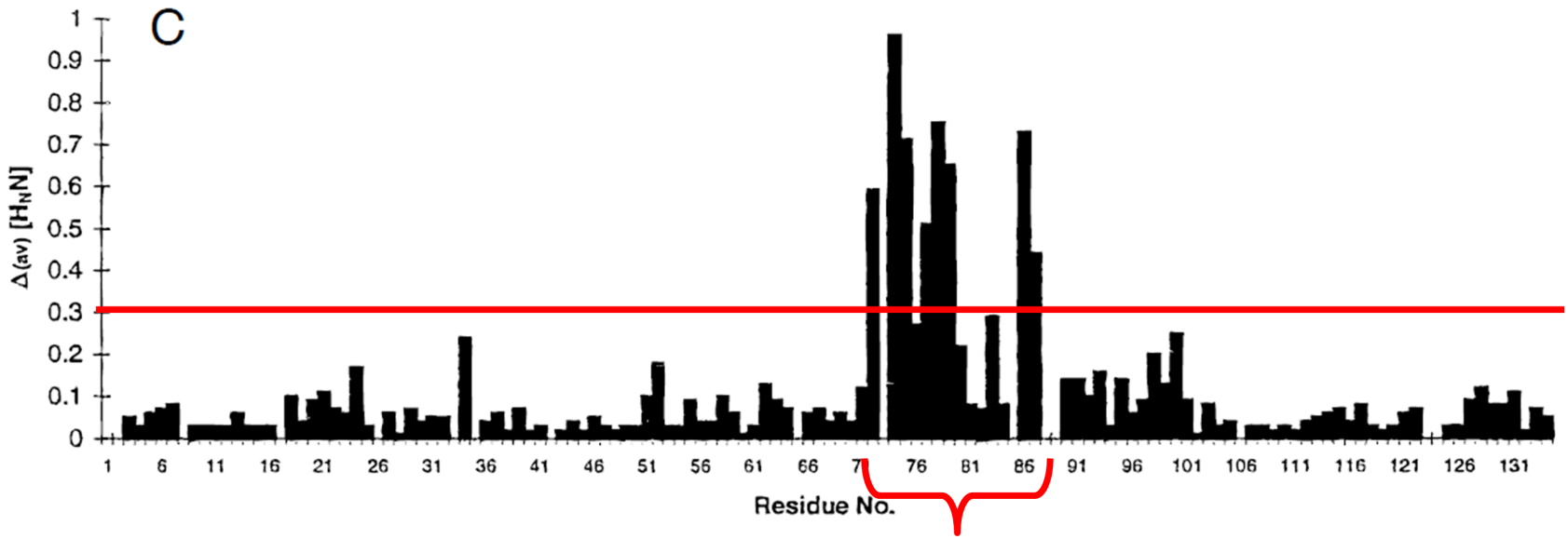


1-Ala 2-Cys 3-Trp 4-Cys 5-Val …

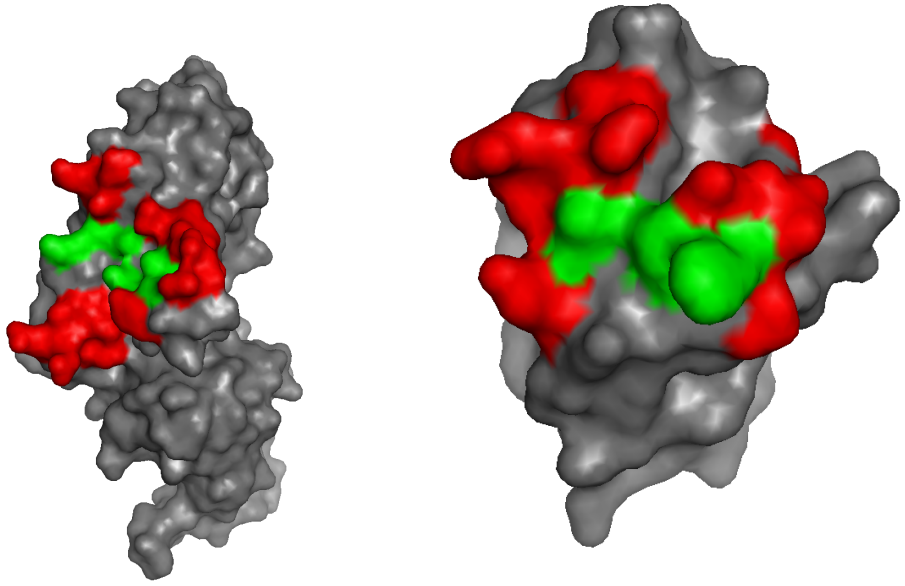




Chemical Shift Perturbation (CSP)



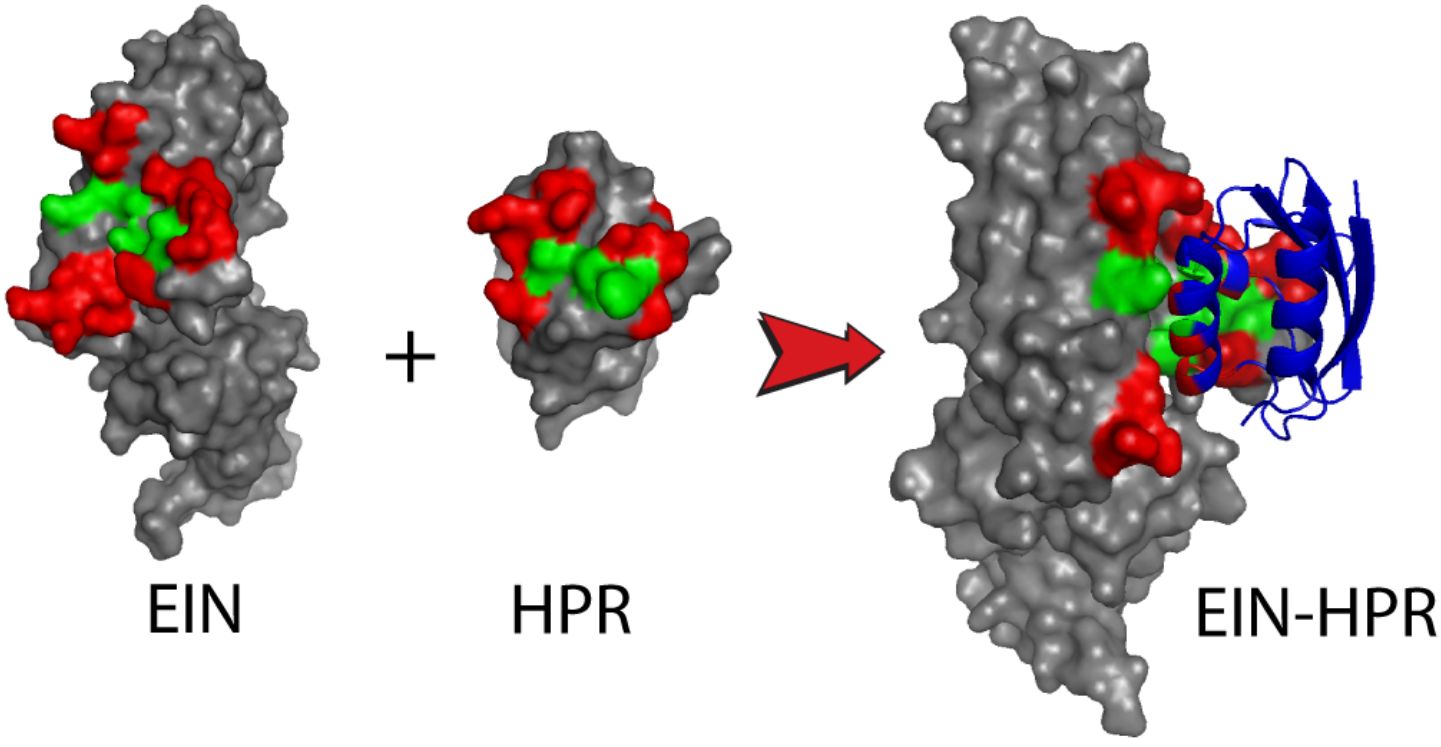
Interface localization on 3D structures



EIN HPR

red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues

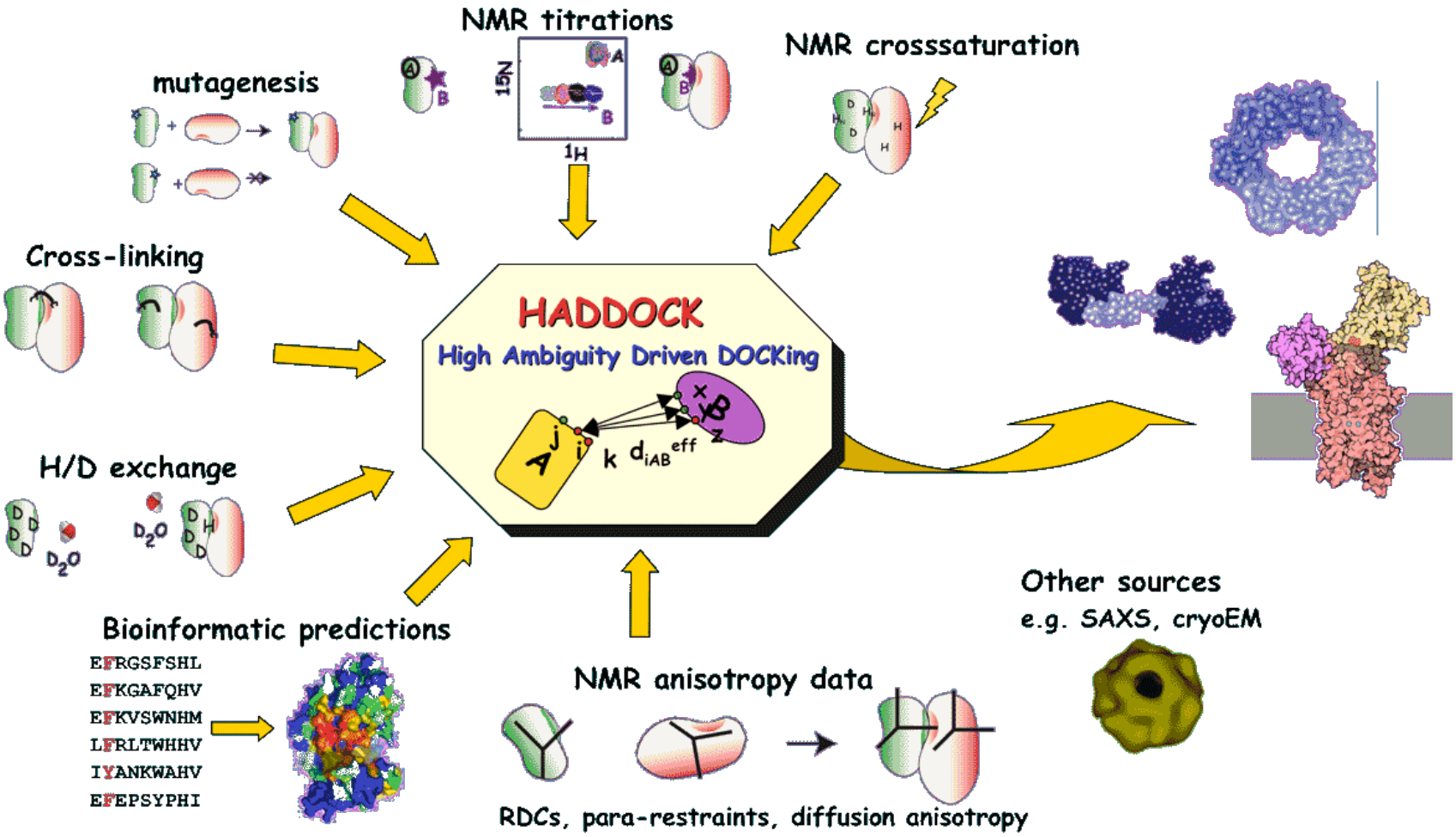
Docking



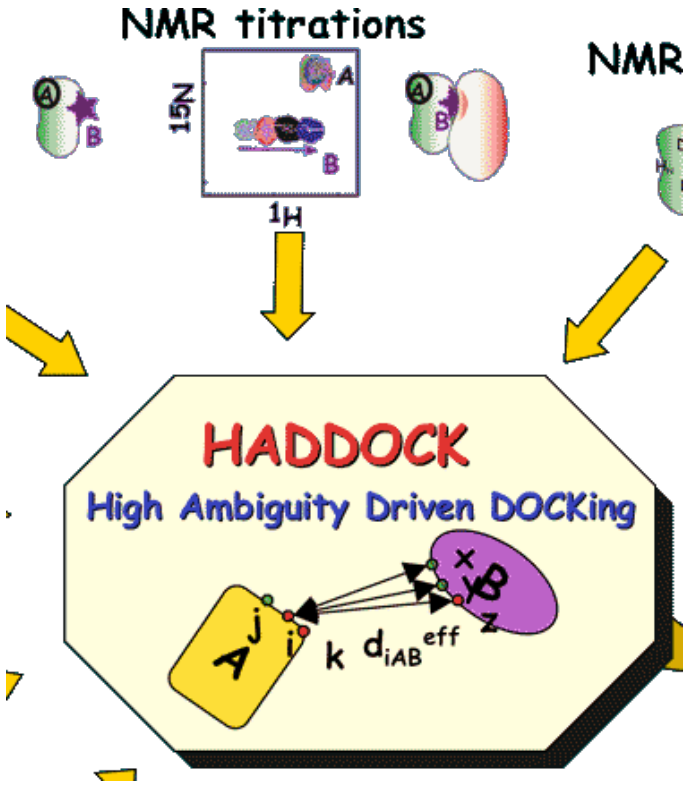
red = active residues derived from CSP data and surface accessibility

green = passive residues, i.e. the surface neighbors of the active residues

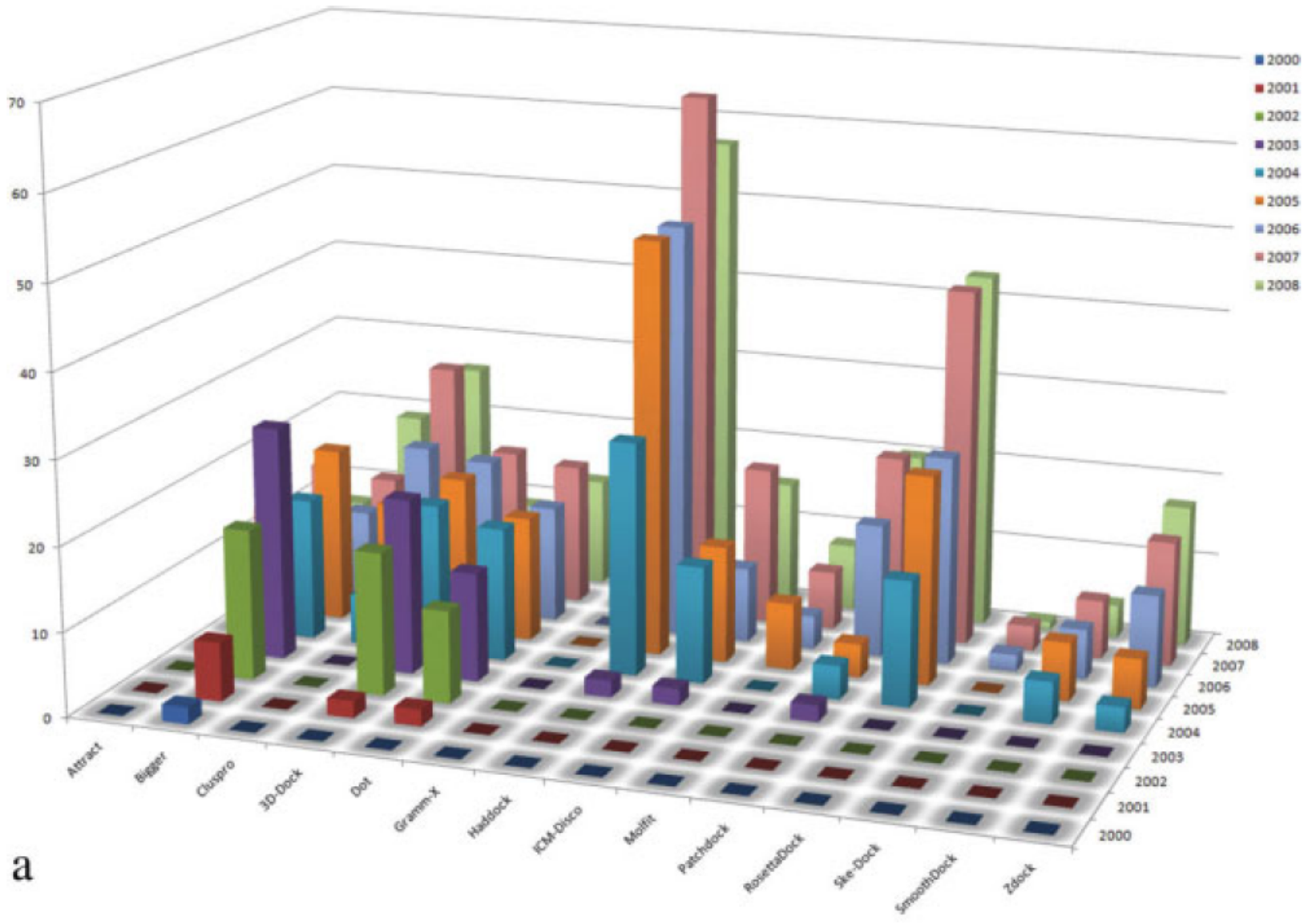
Haddock - http://haddock.chem.uu.nl



Haddock - http://haddock.chem.uu.nl



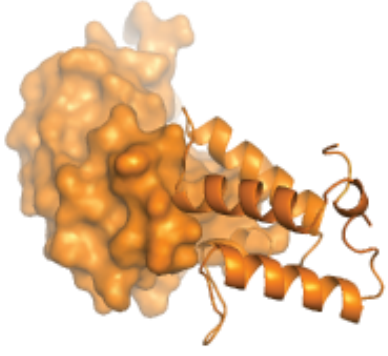
$$E_{Haddock} = E_{vdW} + E_{elec} + E_{AIR} + E_{desolv}$$

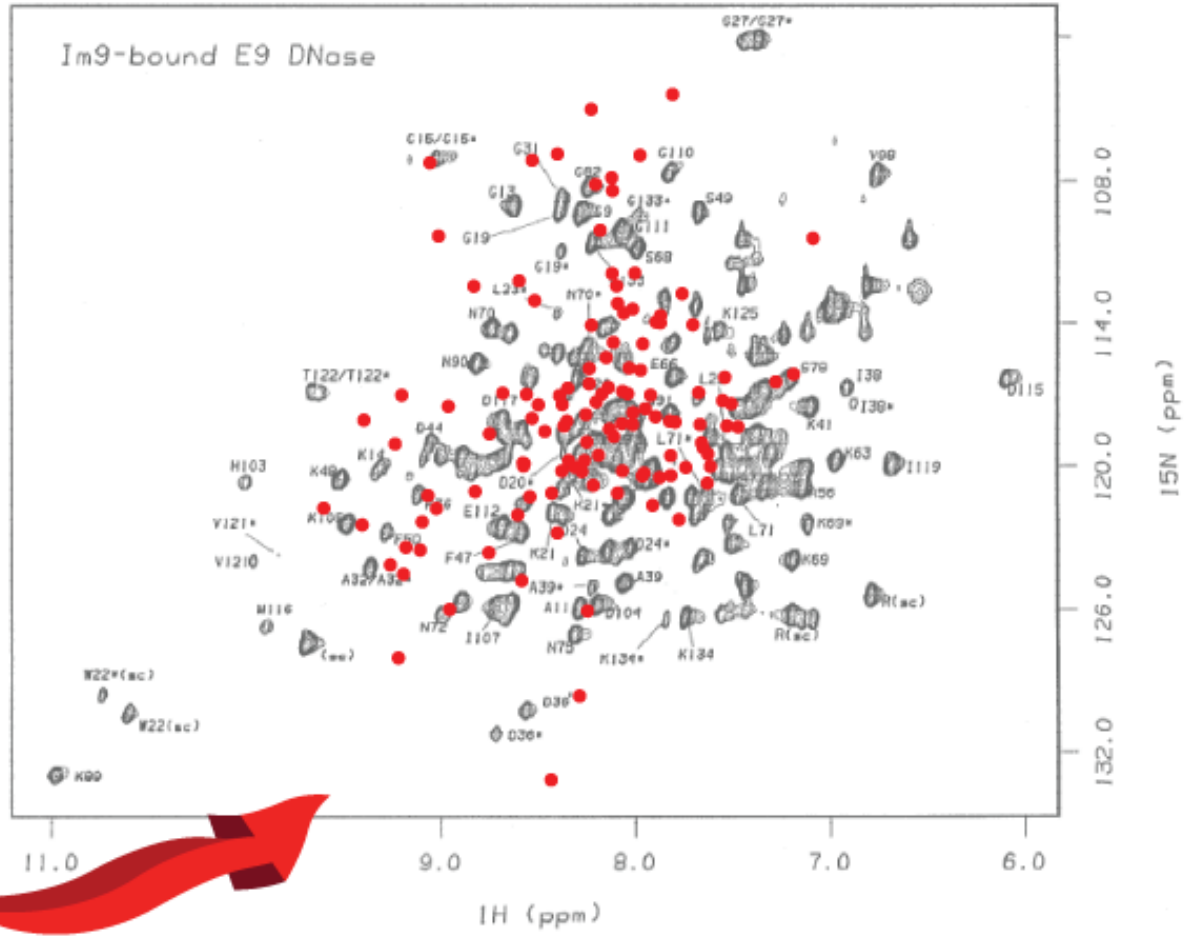

70
60
50
40
30
20
10
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
Attract
Bigger
Cluspro
3D-Dock
Dot
Gramm-X
Haddock
ICM-Disco
Molfit
Patchdock
RosettaDock
Ske-Dock
SmoothDock
Zdock

a

CS-HADDOCK

3D to CS





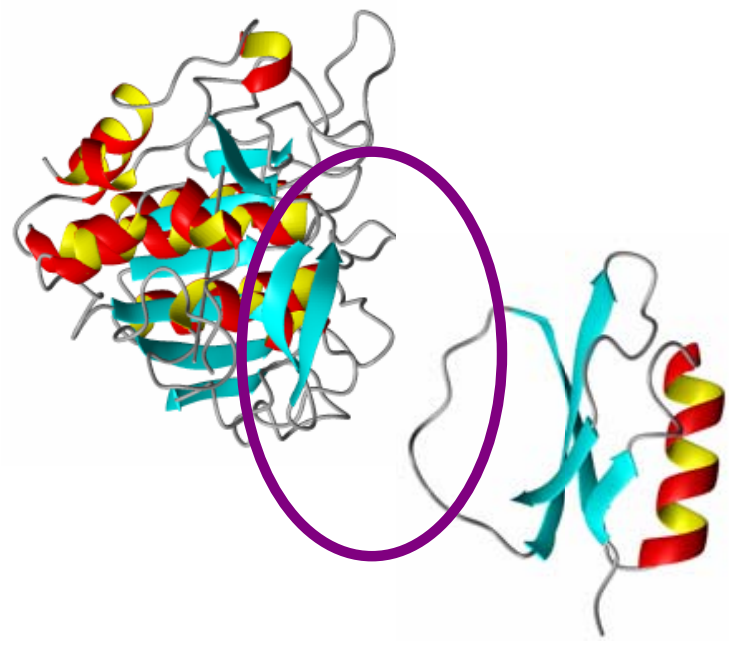
CS-predictor

3D to CS with ShiftX

Contributions to calculated CS δ_{calc}:

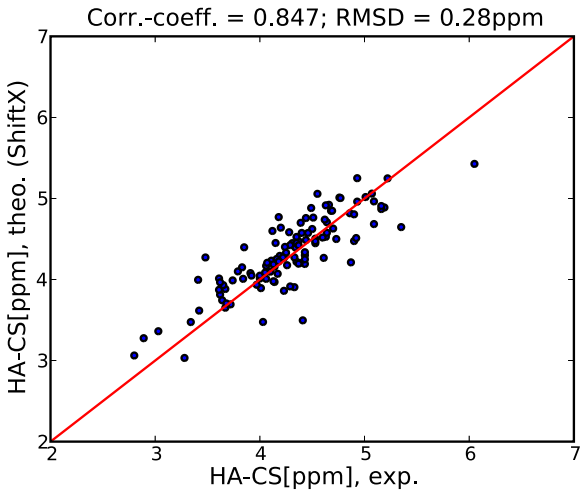
$$\delta_{calc} = \delta_{coil} + \delta_{RC} + \delta_{EF} + \delta_{HB} + \delta_{HS}$$

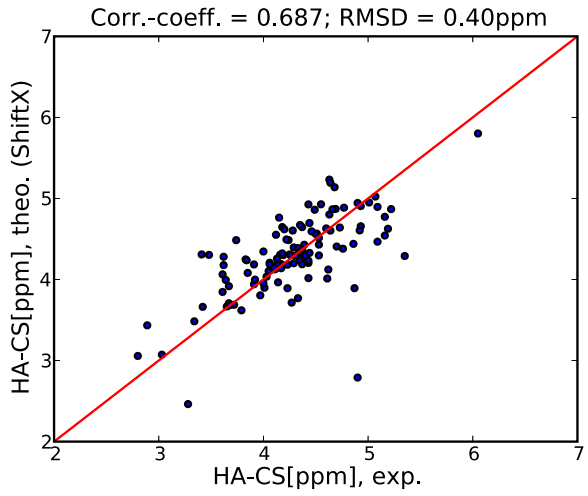
- δ_{coil} - random coil (amino acid type)
- δ_{RC} - ring current
- δ_{EF} - electric field
- δ_{HB} - hydrogen bonding
- δ_{HS} - empirical hypersurfaces (backbone dihedral angles)

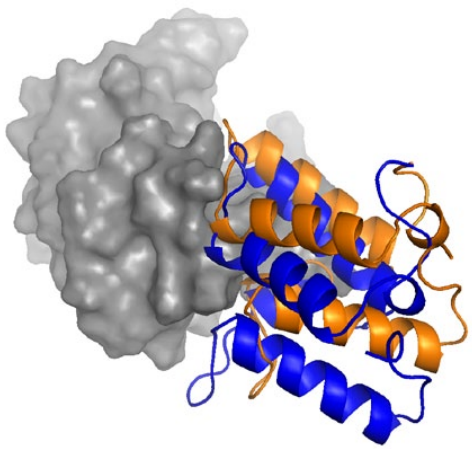


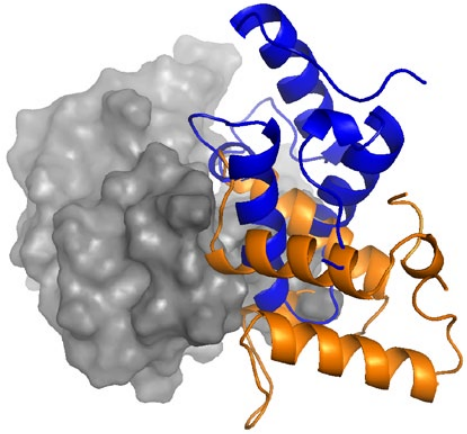
Neal et al., *J. Biomol. NMR* 26: 215-240, 2003

RMSD between δ_{calc} and δ_{exp} for $^{1}H^{\alpha}$-CS

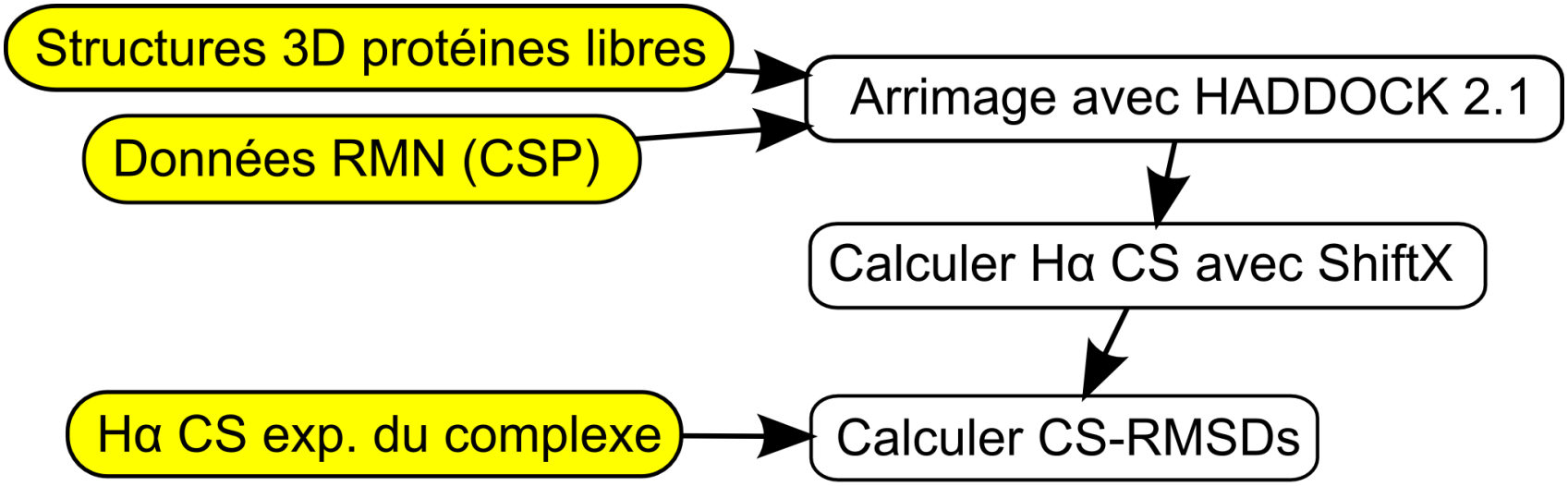








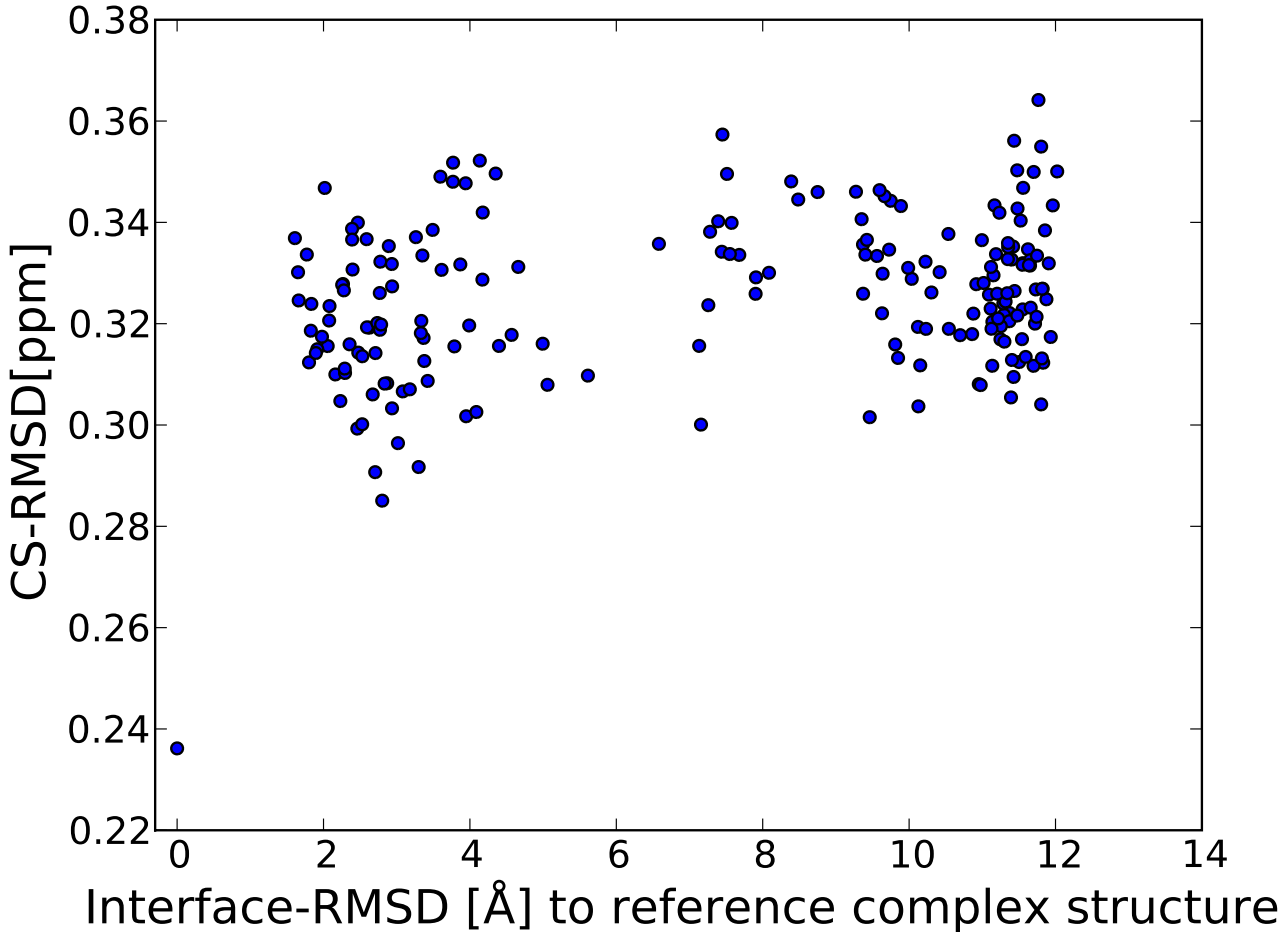
Protocole d'arrimage CS-HADDOCK



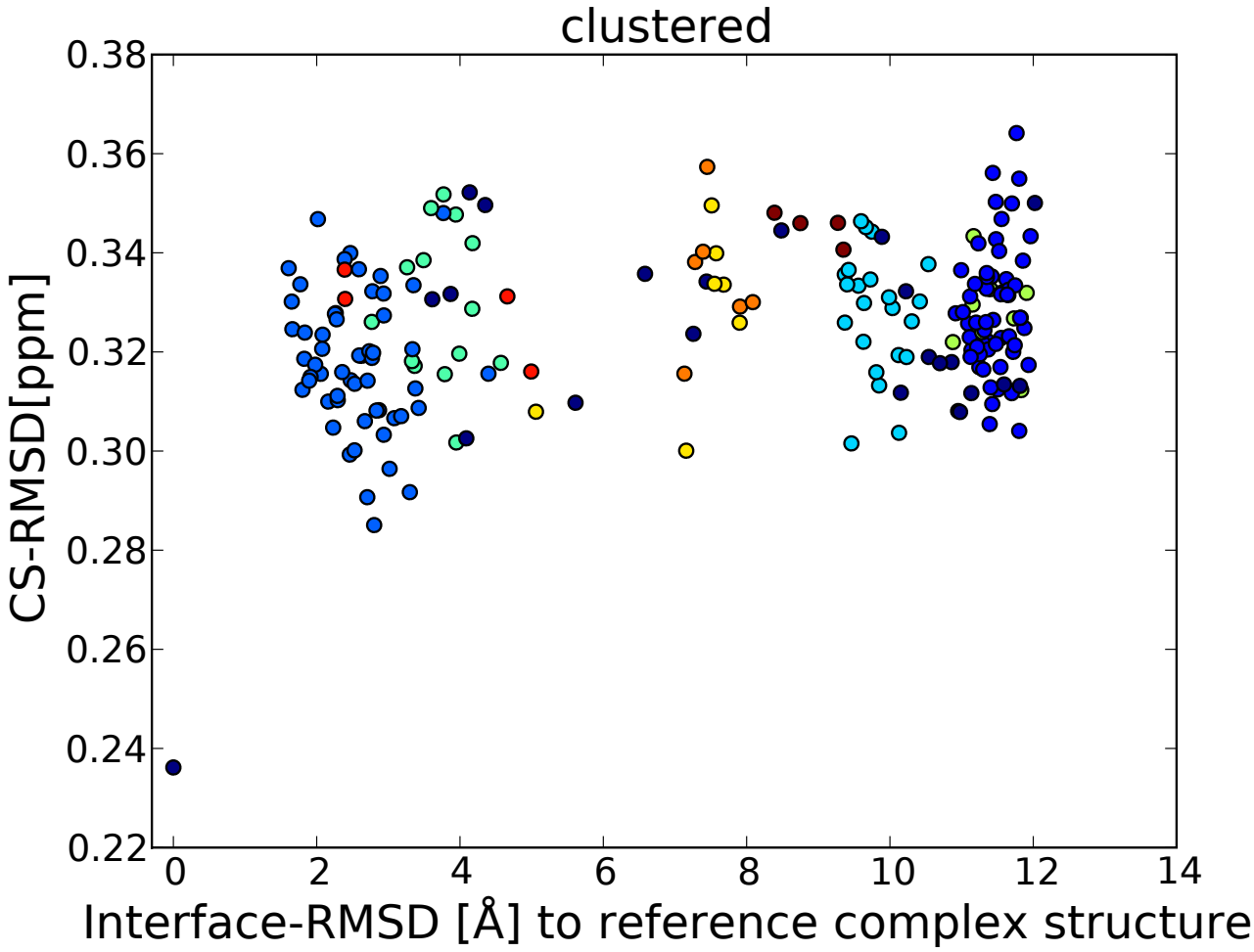
$$\text{CS-RMSD} = \frac{\sqrt{\frac{\sum_{i=1}^{n_A}(\delta_i^{exp}-\delta_i^{theo})^2}{n_A}} + \sqrt{\frac{\sum_{i=1}^{n_B}(\delta_i^{exp}-\delta_i^{theo})^2}{n_B}}}{2}$$

[Stratmann et al., 2011]

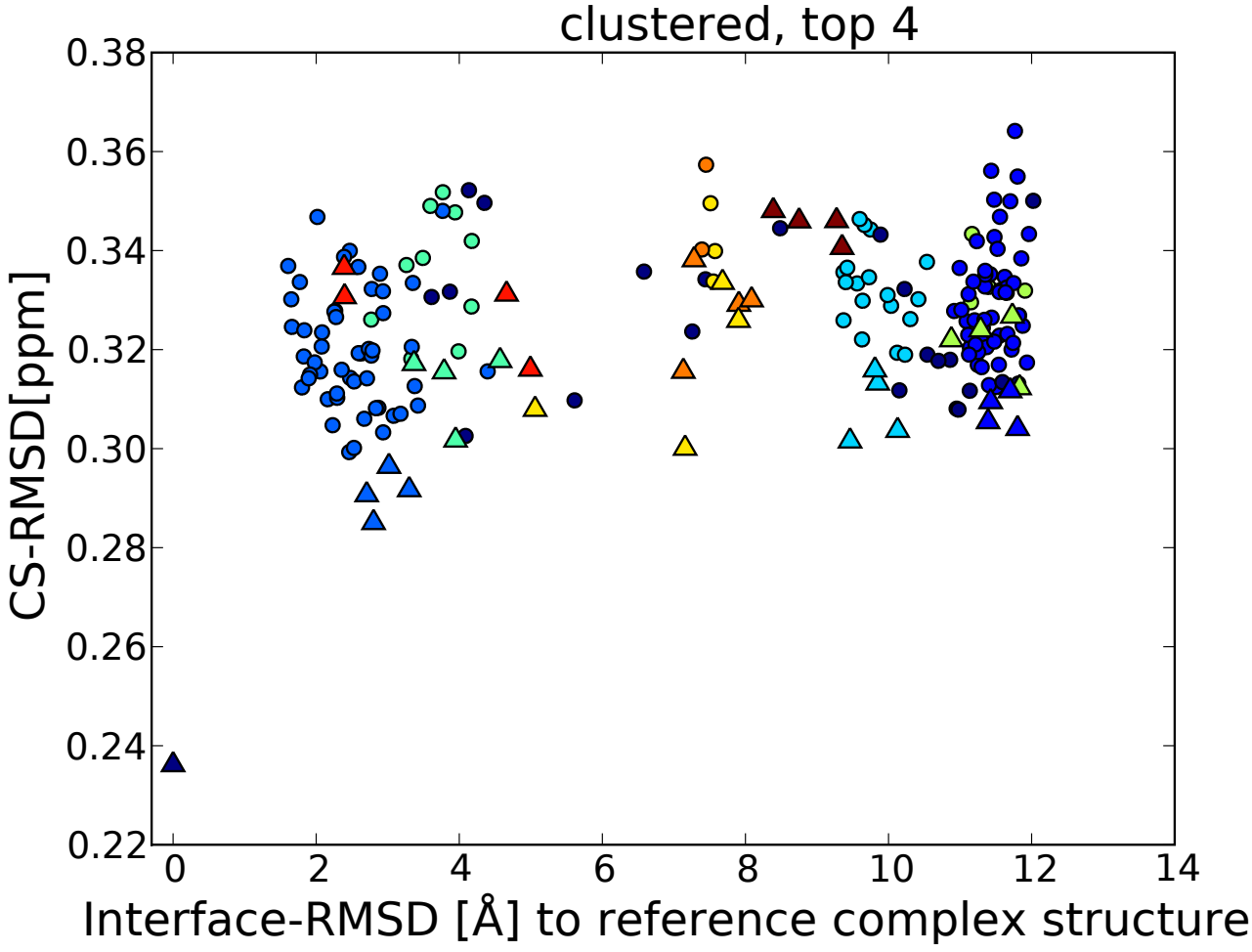
CS-RMSD scoring on all generated structures



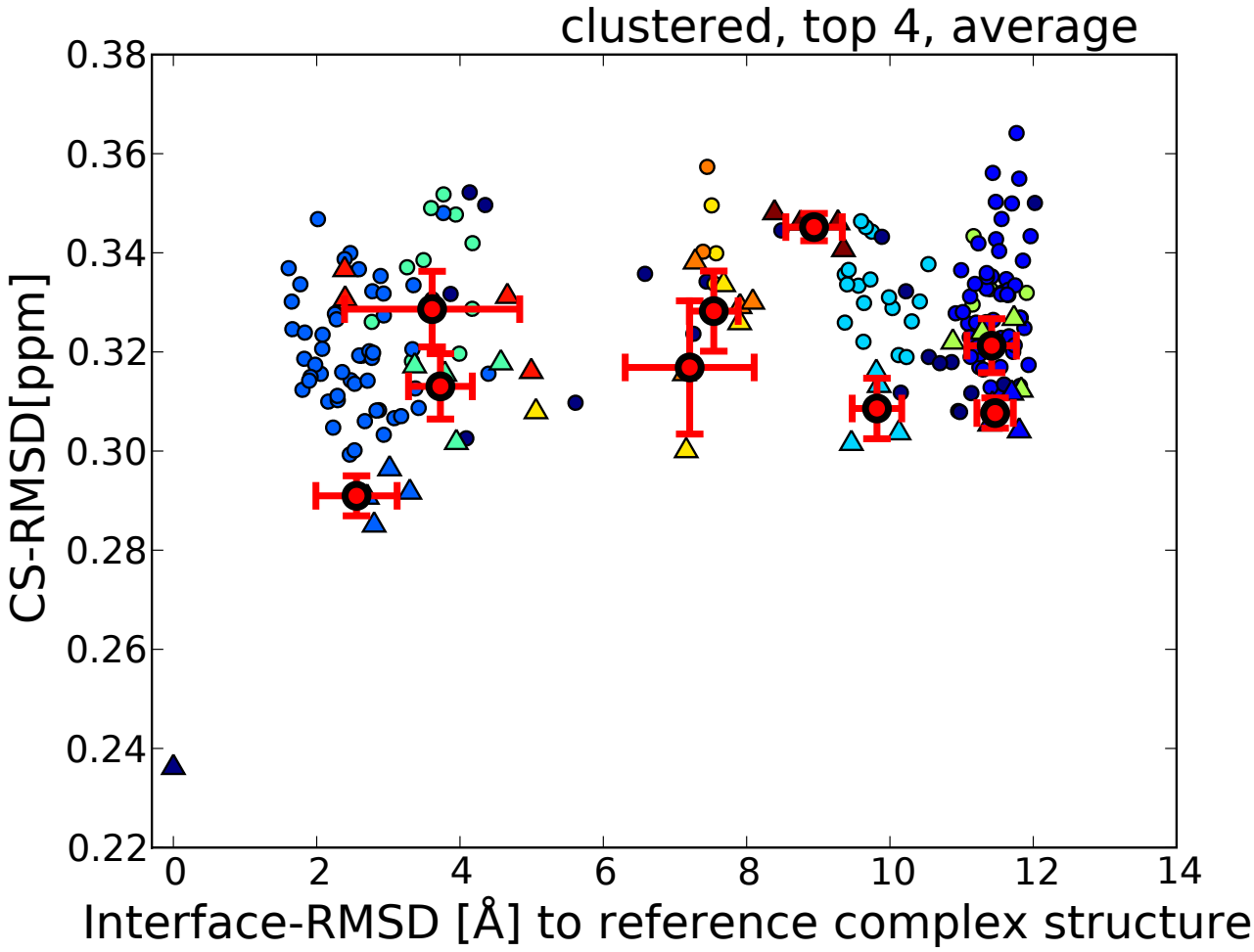
CS-RMSD scoring on all generated structures



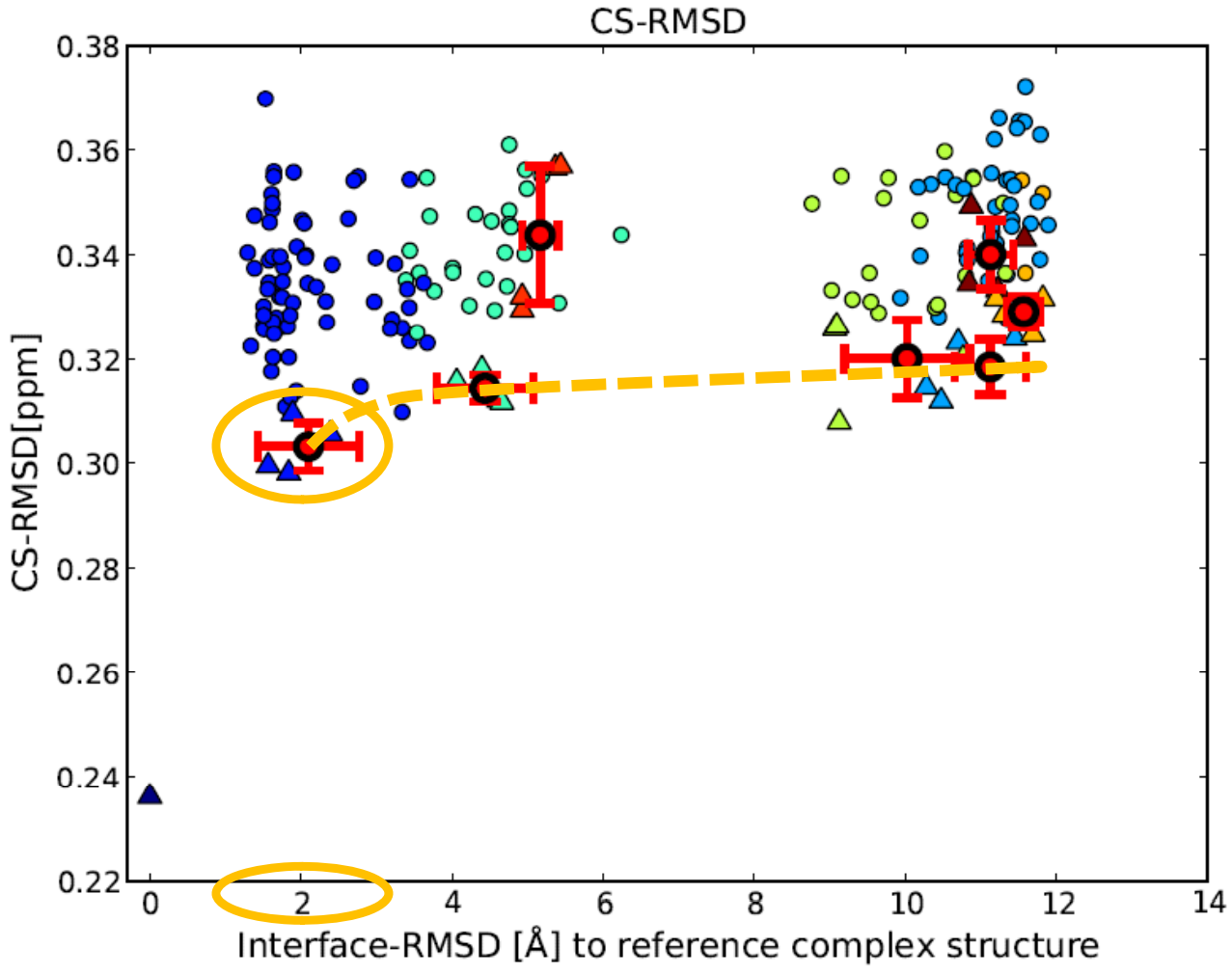
CS-RMSD scoring on all generated structures



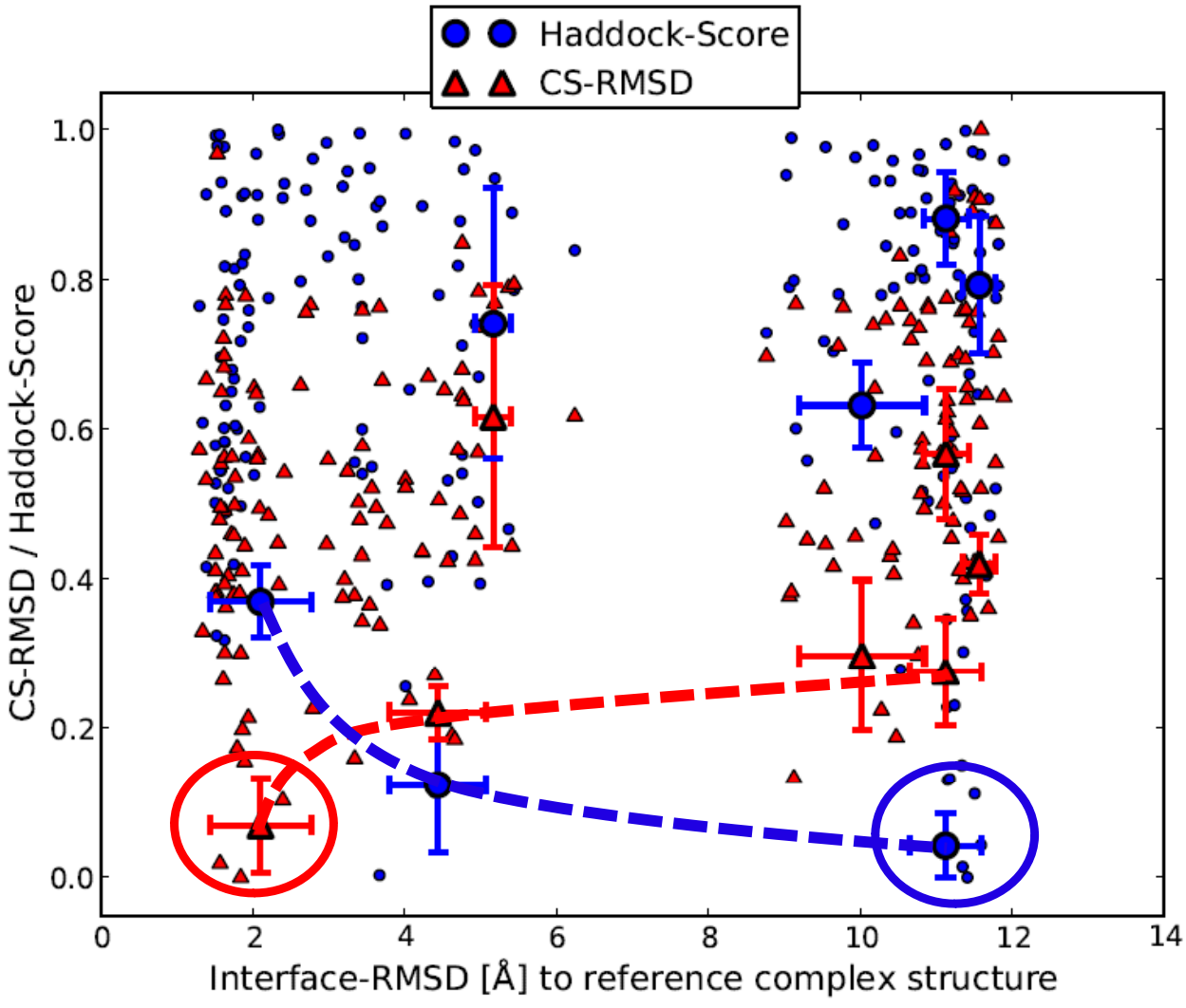
CS-RMSD scoring on all generated structures



Classement des clusters de structures par CS-RMSD

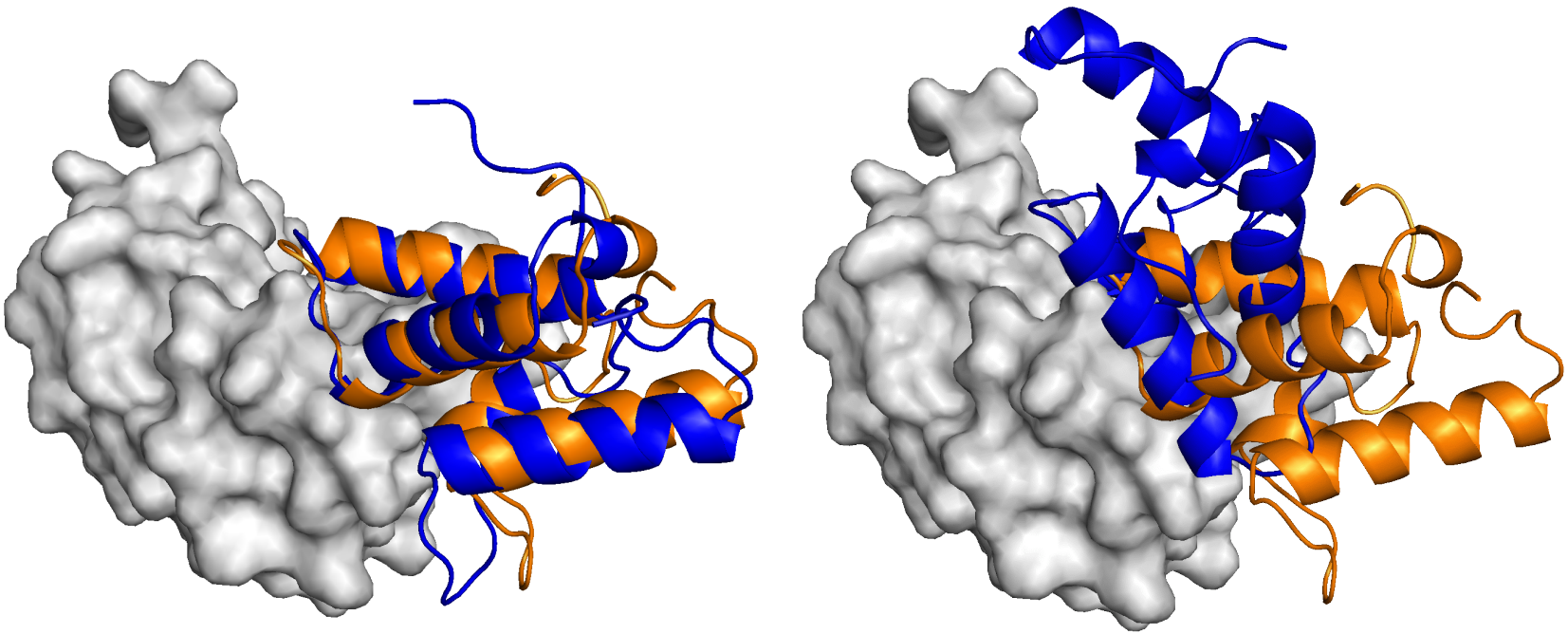


CS-HADDOCK vs HADDOCK



CS-HADDOCK vs HADDOCK

Meilleure structure (en bleu) par rapport à la référence (en orange):



(c) CS-RMSD score

(d) HADDOCK score

Andrusier, N., Mashiach, E., Nussinov, R., and Wolfson, H. J. (2008). Principles of flexible protein-protein docking. *Proteins*, 73(2):271–289.
PMID: 18655061.

Andrusier, N., Nussinov, R., and Wolfson, H. J. (2007). FireDock: fast interaction refinement in molecular docking. *Proteins*, 69(1):139–159.
PMID: 17598144.

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J*, 80(1):505–515.
PMID: 11159421 PMCID: PMC1301252.

Bastard, K., Prévost, C., and Zacharias, M. (2006). Accounting for loop flexibility during protein-protein docking. *Proteins*, 62(4):956–969.
PMID: 16372349.

Bastard, K., Thureau, A., Lavery, R., and Prévost, C. (2003). Docking macromolecules with flexible segments. *J Comput Chem*, 24(15):1910–1920.
PMID: 14515373.

Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, 5(11):789–796.
PMID: 19841628.

Bonvin, A. M. J. J. (2006). Flexible protein-protein docking. *Curr. Opin. Struct. Biol.*, 16(2):194–200.
PMID: 16488145.

Cavasotto, C. N., Kovacs, J. A., and Abagyan, R. A. (2005). Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.*, 127(26):9632–9640.
PMID: 15984891.

Chen, R. and Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins*, 47(3):281–294.
PMID: 11948782.

Connolly, M. L. (1986). Shape complementarity at the hemoglobin alpha 1 beta 1 subunit interface. *Biopolymers*, 25(7):1229–1247.
PMID: 3741993.

Deupi, X. and Kobilka, B. K. (2010). Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology*, 25(5):293–303.

Duhovny, D., Nussinov, R., and Wolfson, H. J. (2002). Efficient unbound docking of rigid molecules. In *In WABI '02: Proceedings of the Second International Workshop on Algorithms in Bioinformatics*, page 185–200. Springer Verlag.

Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R., and Haliloglu, T. (2008). HingeProt: automated prediction of hinges in protein structures. *Proteins*, 70(4):1219–1227.
PMID: 17847101.

Grünberg, R., Leckner, J., and Nilges, M. (2004). Complementarity of structure ensembles in protein-protein binding. *Structure*, 12(12):2125–2136.
PMID: 15576027.

Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–443.
PMID: 12001221.

Jacobs, D. J., Rader, A., Kuhn, L. A., and Thorpe, M. (2001). Protein flexibility predictions using graph theory. *Proteins*, 44(2):150–165.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., and Vakser, I. A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.*, 89(6):2195–2199.
PMID: 1549581.

Krippahl, L., Moura, J. J., and Palma, P. N. (2003). Modeling protein complexes with BiGGER. *Proteins*, 52(1):19–23.
PMID: 12784362.

Lensink, M. F. and Wodak, S. J. (2010). Docking and scoring protein interactions: CAPRI 2009. *Proteins*, 78(15):3073–3084.
PMID: 20806235.

Mashiach, E., Nussinov, R., and Wolfson, H. J. (2010a). FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking. *Nucleic Acids Res.*, 38(Web Server issue):W457–461.
PMID: 20460459.

Mashiach, E., Nussinov, R., and Wolfson, H. J. (2010b). FiberDock: flexible induced-fit backbone refinement in molecular docking. *Proteins*, 78(6):1503–1519.
PMID: 20077569.

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H. J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.*, 36(Web Server issue):W229–232.
PMID: 18424796.

May, A. and Zacharias, M. (2008). Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins*, 70(3):794–809.
PMID: 17729269.

Miyazawa, S. and Jernigan, R. L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, 256(3):623–644.
PMID: 8604144.

Méndez, R., Leplae, R., Lensink, M. F., and Wodak, S. J. (2005). Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, 60(2):150–169.
PMID: 15981261.

Moont, G., Gabb, H. A., and Sternberg, M. J. (1999). Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins*, 35(3):364–373.
PMID: 10328272.

Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2010). Protein-protein docking dealing with the unknown. *J Comput Chem*, 31(2):317–342.
PMID: 19462412.

Mustard, D. and Ritchie, D. W. (2005). Docking essential dynamics eigenstructures. *Proteins*, 60(2):269–274.
PMID: 15981272.

Norel, R., Lin, S. L., Wolfson, H. J., and Nussinov, R. (1994). Shape complementarity at protein-protein interfaces. *Biopolymers*, 34(7):933–940.
PMID: 8054472.

Nukada, A., Hourai, Y., Nishida, A., and Akiyama, Y. (2007). High performance 3D convolution for protein docking on IBM blue gene. In Stojmenovic, I., Thulasiram, R. K., Yang, L. T., Jia, W., Guo, M., and Mello, R. F. d., editors, *Parallel and Distributed Processing and Applications*, number 4742 in Lecture Notes in Computer Science, pages 958–969. Springer Berlin Heidelberg.

Palma, P. N., Krippahl, L., Wampler, J. E., and Moura, J. J. (2000). BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins*, 39(4):372–384.
PMID: 10813819.

Petrone, P. and Pande, V. S. (2006). Can conformational change be described by only a few normal modes? *Biophys J*, 90(5):1583–1593.
PMID: 16361336 PMCID: PMC1367309.

Pierce, B. and Weng, Z. (2007). Structure prediction of protein complexes. In Xu, Y., Xu, D., and Liang, J., editors, *Computational Methods for Protein Structure Prediction and Modeling*, Biological and Medical Physics, Biomedical Engineering, pages 109–134. Springer New York.


Pierce, B. G., Hourai, Y., and Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE*, 6(9):e24657.
PMID: 21949741.


Sandak, B., Wolfson, H. J., and Nussinov, R. (1998). Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. *Proteins*, 32(2):159–174.
PMID: 9714156.


Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). Geometry-based flexible and symmetric protein docking. *Proteins*, 60(2):224–231.
PMID: 15981269.


Schneidman-Duhovny, D., Nussinov, R., and Wolfson, H. J. (2007). Automatic prediction of protein interactions with large scale motion. *Proteins*, 69(4):764–773.
PMID: 17886339.


Smith, G. R. and Sternberg, M. J. E. (2002). Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.*, 12(1):28–35.
PMID: 11839486.


Stratmann, D., Boelens, R., and Bonvin, A. M. J. J. (2011). Quantitative use of chemical shifts for the modeling of protein complexes. *Proteins*, 79(9):2662–2670.
PMID: 21744392.


Tuffery, P. and Derreumaux, P. (2012). Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches. *J R Soc Interface*, 9(66):20–33.
PMID: 21993006.


Wang, C., Bradley, P., and Baker, D. (2007). Protein-protein docking with backbone flexibility. *J. Mol. Biol.*, 373(2):503–519.
PMID: 17825317.

Zacharias, M. (2003). Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.*, 12(6):1271–1282.
PMID: 12761398.

Zacharias, M. (2010). Accounting for conformational changes during protein-protein docking. *Curr. Opin. Struct. Biol.*, 20(2):180–186.
PMID: 20194014.

Zhang, C., Vasmatzis, G., Cornette, J. L., and DeLisi, C. (1997). Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.*, 267(3):707–726.
PMID: 9126848.

Zhao, Y. and Sanner, M. F. (2007). FLIPDock: docking flexible ligands into flexible receptors. *Proteins*, 68(3):726–737.
PMID: 17523154.

Zhao, Y., Stoffler, D., and Sanner, M. (2006). Hierarchical and multi-resolution representation of protein flexibility. *Bioinformatics*, 22(22):2768–2774.
PMID: 16984893.