

Protein-protein docking

Arrimage protéine-protéine

Dirk Stratmann

https://stratmann.fr/cours/homology_modeling
dirk.stratmann@upmc.fr

Master M2 PSF, Sorbonne Université

octobre 2024

Outline

- 1 Introduction
 - Motivation
 - Steps of protein-protein docking
- 2 Protein-protein interaction
 - Interactions protéine-protéine (PPI)
 - Models
 - Types of complexes
- 3 Scoring
 - Scoring Functions
 - Shape complementarity
- 4 Rigid-body docking
 - Geometric docking
 - Fast Fourier Transform (FFT) docking
- 5 Evaluation
 - Performance of docking programs
 - CAPRI
- 6 Inclusion of experimental data
 - NMR - chemical shifts
 - CS-HADDOCK
- 7 Bibliography

1

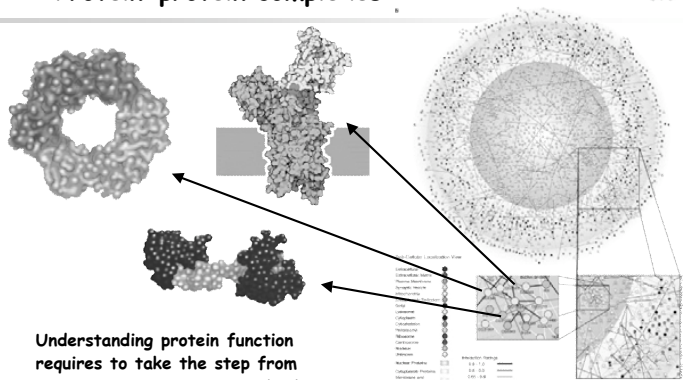
Introduction

- Motivation
- Steps of protein-protein docking

Motivation

Protein function

Protein-protein complexes



AB/10-07

- ✓ PNAS 100, 12123 (2003)
- ✓ Science 302, 1727 (2003)

Free proteins - Structural genomics

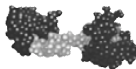
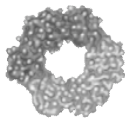
- 3D structure of a large number of unbound/free proteins solved => PDB
- Only about 1000 types of folds, almost all known.
- => Comparative modeling / Homology modeling

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

Protein-docking problem

M L Connolly (July 1986). In: *Biopolymers* 25.7

- Connolly 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).

Steps of protein-protein docking

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:

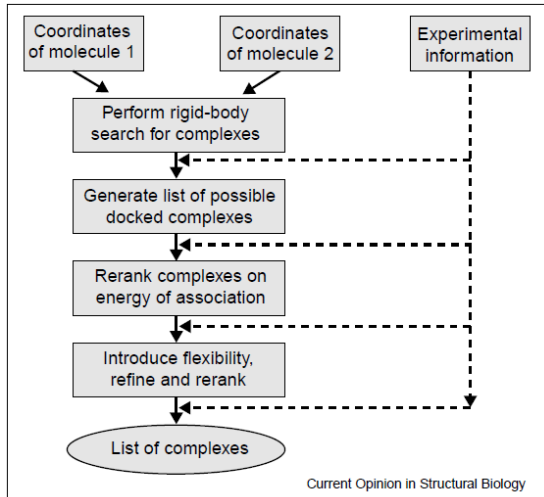
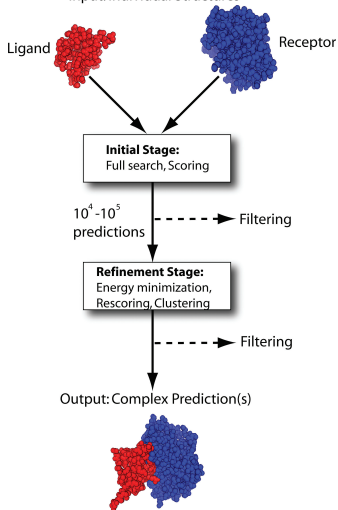
Graham R Smith and Michael J E Sternberg (Feb. 2002). In: *Curr. Opin. Struct. Biol.*

12.1

Inbal Halperin et al. (June 2002). In: *Proteins* 47.4

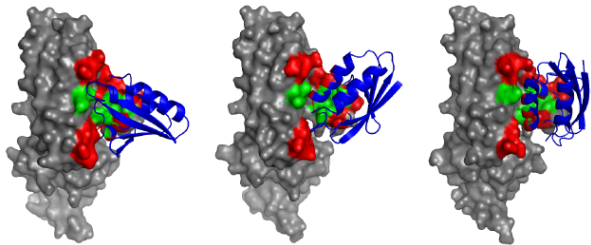
Protein Docking: General Methodology

Input: Individual Structures



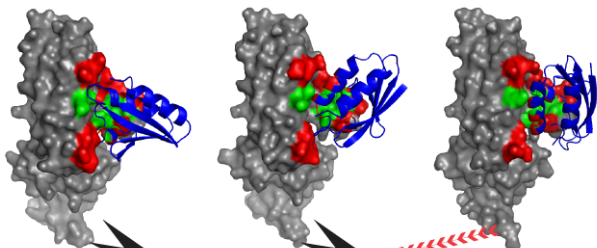
Sampling and Scoring

Sampling

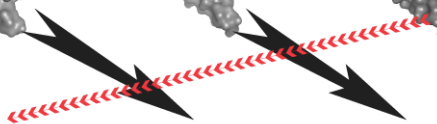


Sampling and Scoring

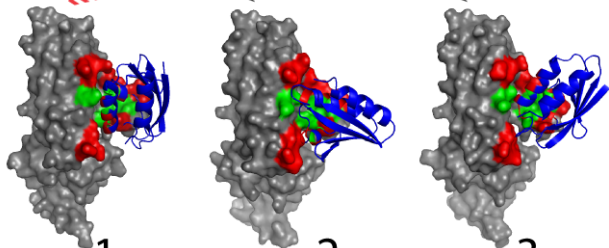
Sampling



Score Function



Scoring



Rank

1

2

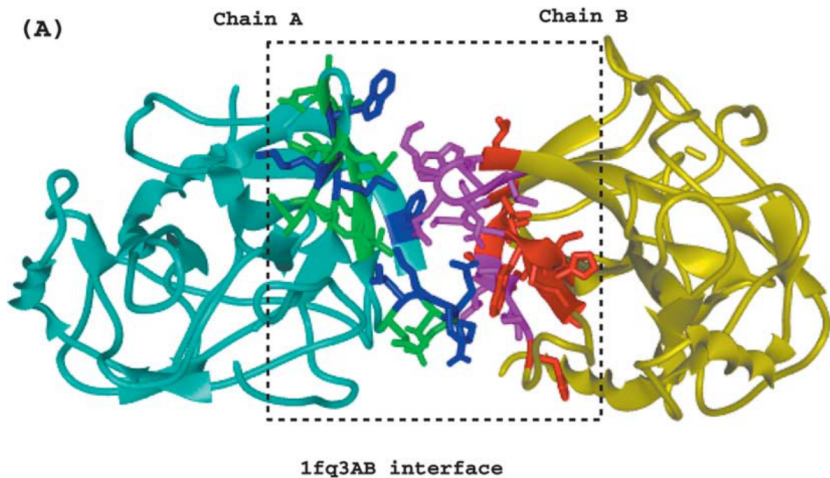
3

2 Protein-protein interaction

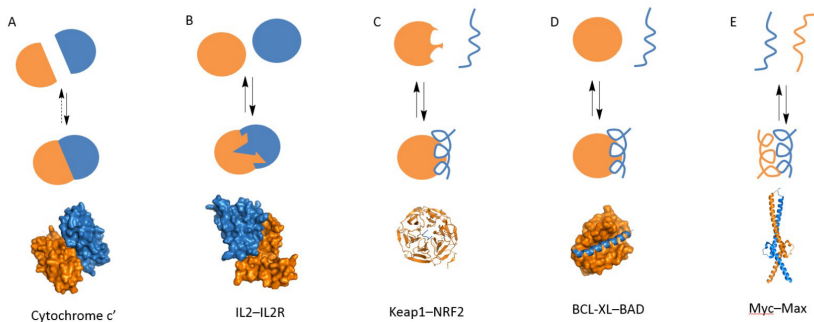
- Interactions protéine-protéine (PPI)
- Models
- Types of complexes

Interactions protéine-protéine (PPI)

Interface

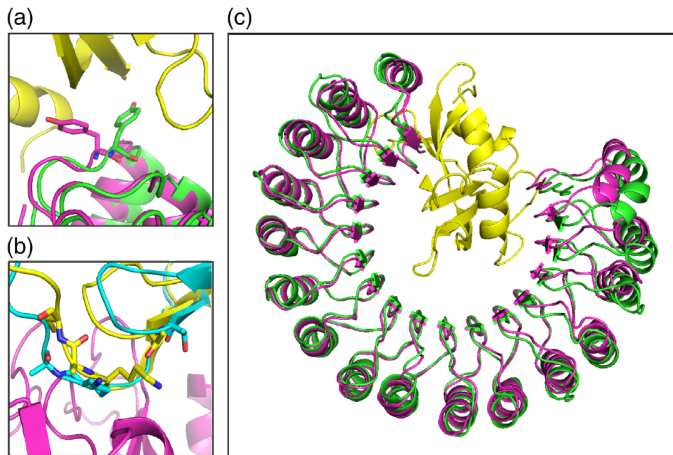


Changement conformationnel lors de l'interaction protéine-protéine



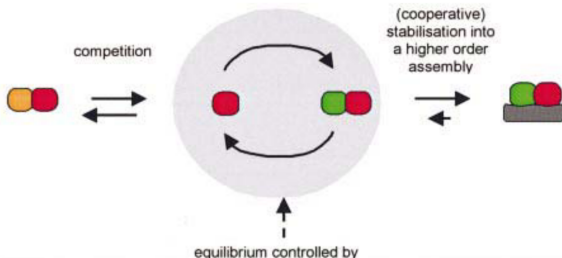
Scott et al. 2016

Changement conformationnel lors de l'interaction protéine-protéine



Facteurs qui contrôlent l'association des protéines

A



localisation and protein concentration(s)

and/or

binding energy ΔG

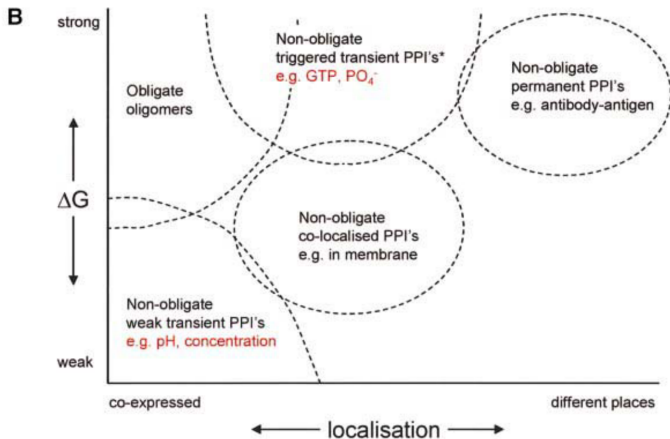
co-expression, subcellular localisation or compartmentalisation

level of gene expression/secretion, degradation, temporary storage, local molecular environment, diffusion or viscosity

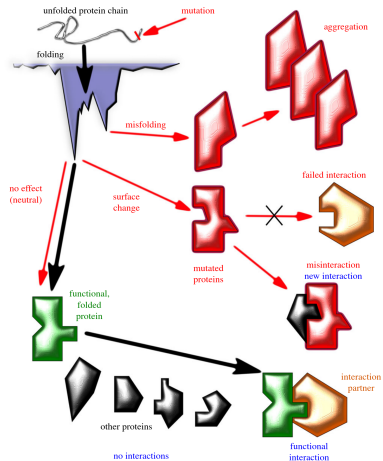
molecular (cooperative/allosteric) binding i.e. concentration of metabolite, protein or ions (e.g. ATP, Ca^{2+}) or covalent modification through enzymatic activity (e.g. PO_4)

pH, temperature, ionic strength

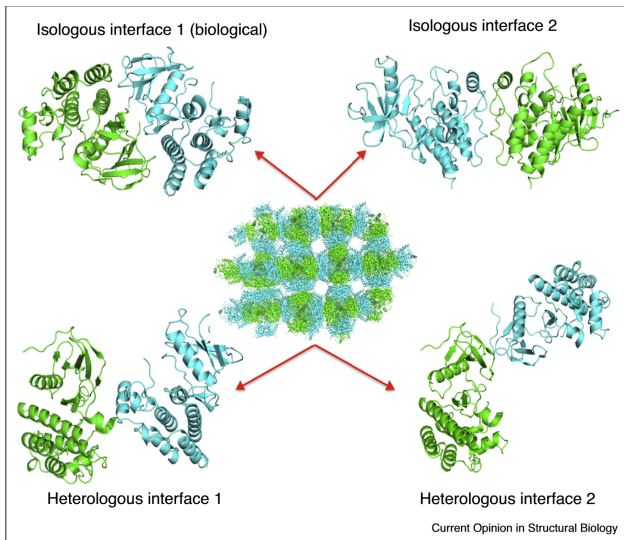
Relation entre les types de PPI, la localisation et l'affinité de liaison



Effets possible d'une mutation



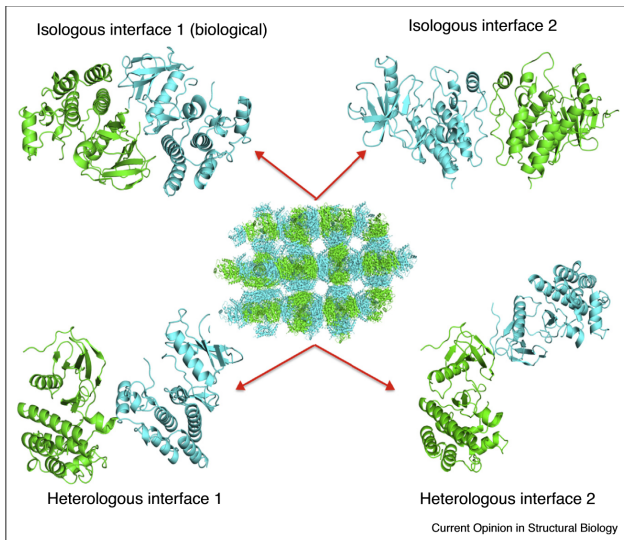
Isologues / Hétérologues



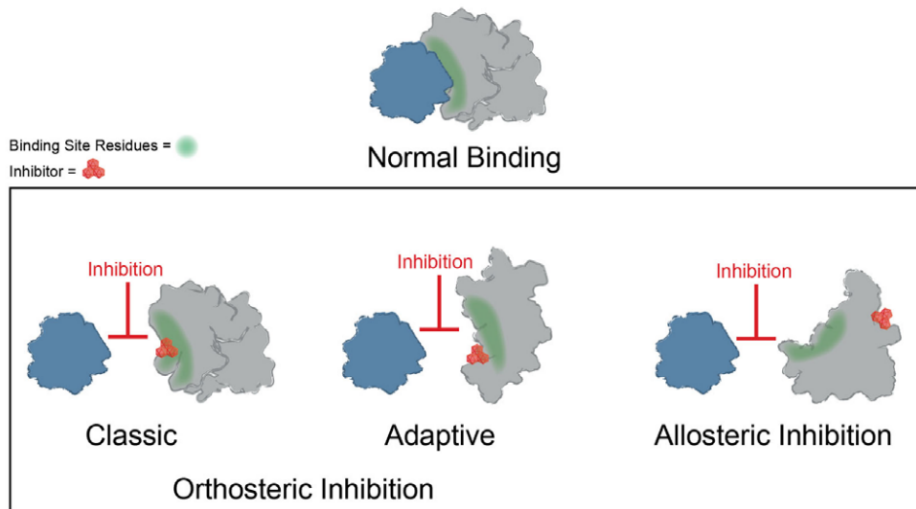
Isologues / Hétérologues

- *Associations isologues*: Le domaine de liaison est composé de deux unités identiques
- *Associations isologues*: Leurs interfaces ont une symétrie C2 (rotation 180 degrés permet de générer un monomère à partir de l'autre)
- *Associations hétérologues*: Le domaine de liaison est composé jusqu'à deux unités différentes
- *Associations hétérologues*: Leurs interfaces sont asymétriques

Isologues / Hétérologues



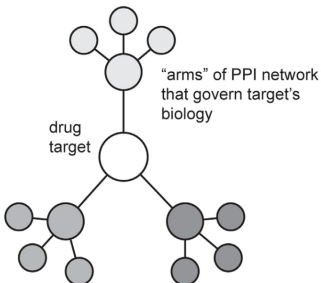
Inhibition des PPI avec des petites molécules



Inhibition des PPI avec des petites molécules

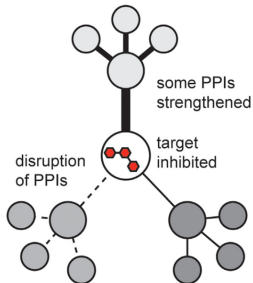
A

Unperturbed PPI Network



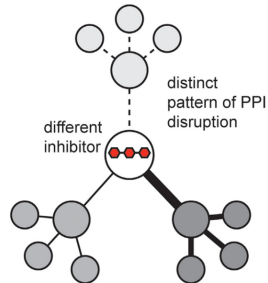
B

Propagation of PPI network disruption after treatment with inhibitor A

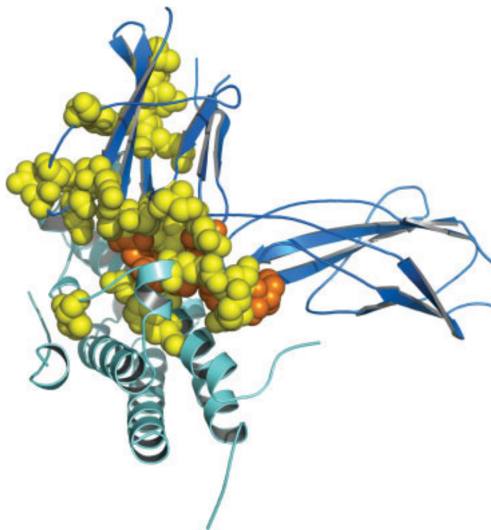


C

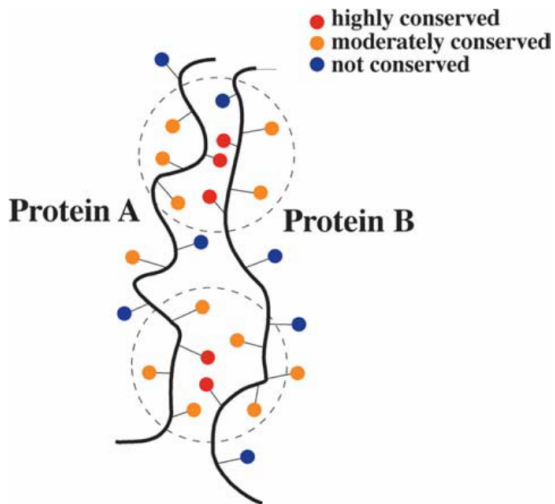
Propagation of PPI network disruption after treatment with inhibitor B



Hot spots

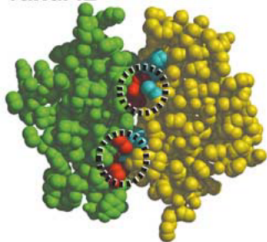


Hot regions

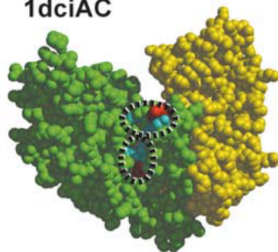


Hot regions

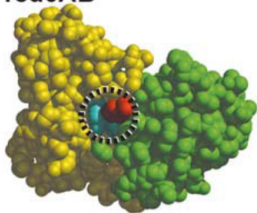
1axdAB



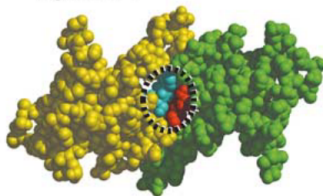
1dciAC



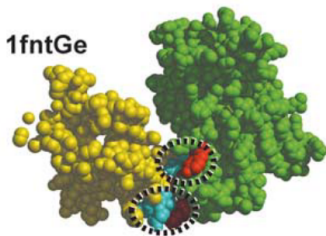
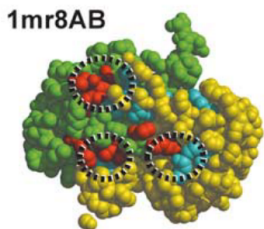
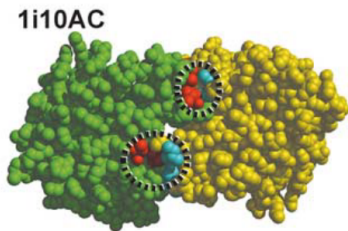
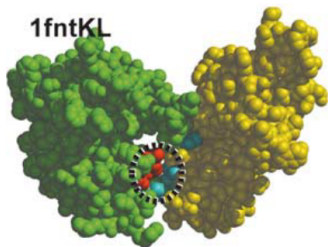
1cd0AB



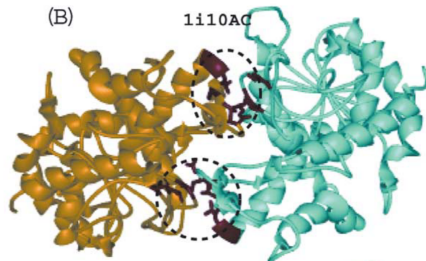
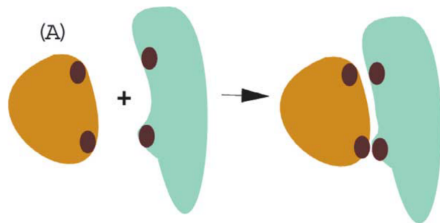
1qmoAB



Hot regions



Hot regions



Models

Lock and Key

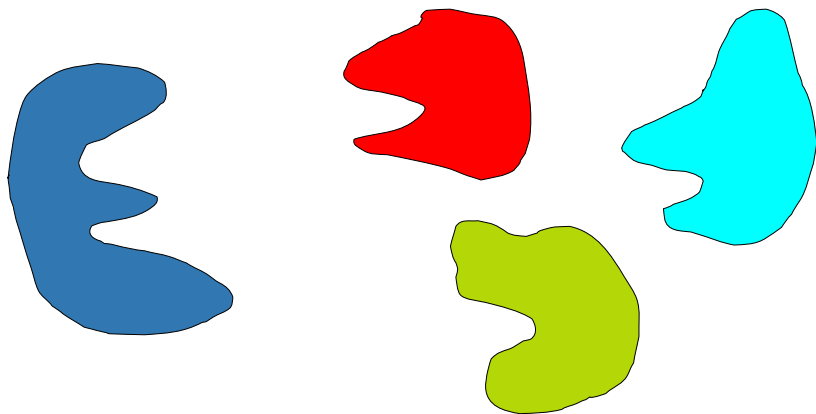


Emil Fischer 1894

“ To use an image, I would say that enzyme and glycoside have to fit into each other like a lock and a key, in order to exert a chemical effect on each other.”

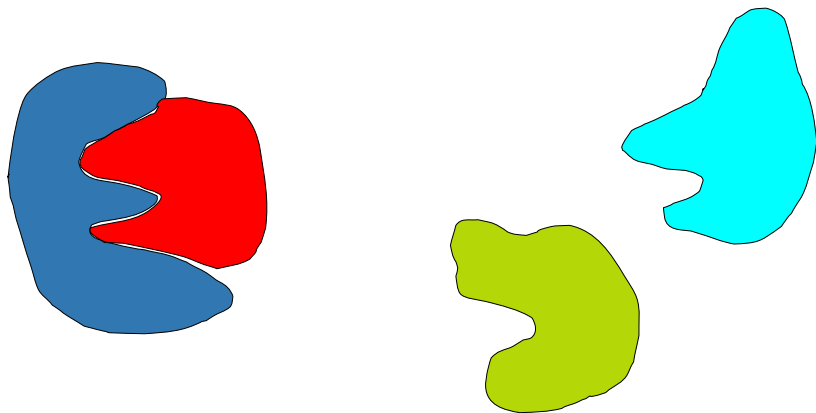
Source: Kohlbacher and Lenhof

Lock and Key



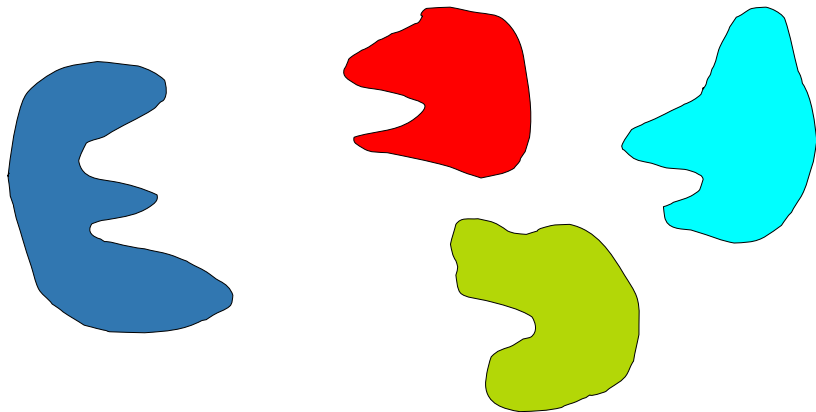
Source: Kohlbacher and Lenhof

Lock and Key



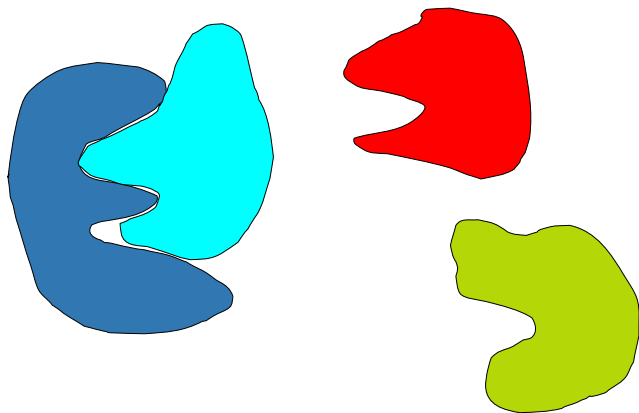
Source: Kohlbacher and Lenhof

Lock and Key



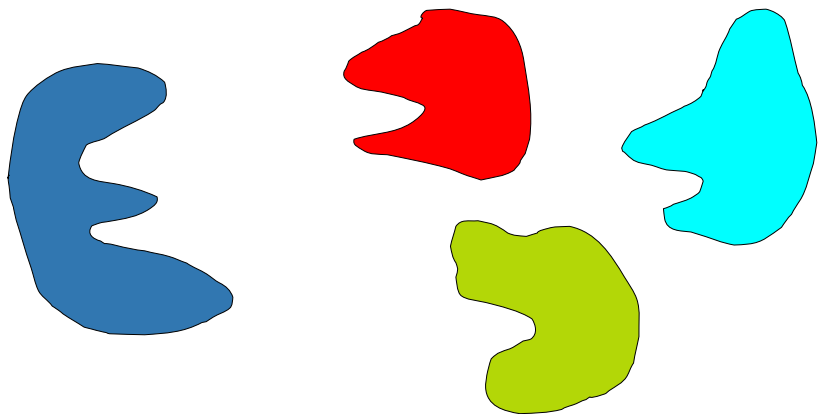
Source: Kohlbacher and Lenhof

Lock and Key



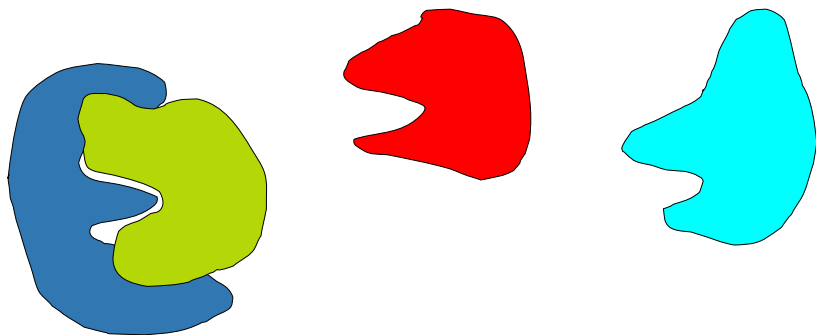
Source: Kohlbacher and Lenhof

Lock and Key



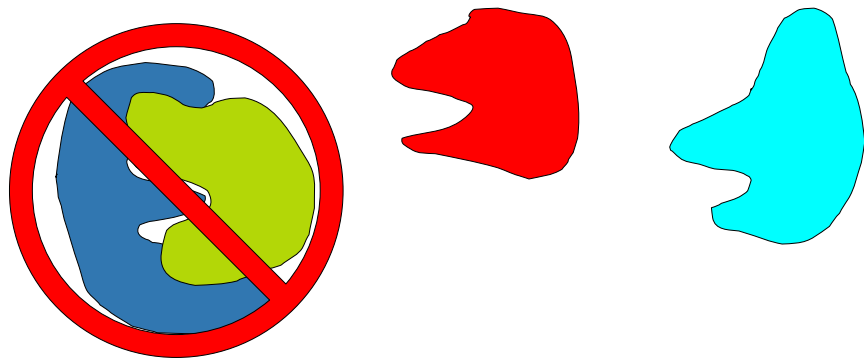
Source: Kohlbacher and Lenhof

Lock and Key



Source: Kohlbacher and Lenhof

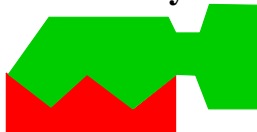
Lock and Key



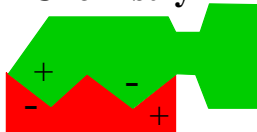
Source: Kohlbacher and Lenhof

Lock and Key

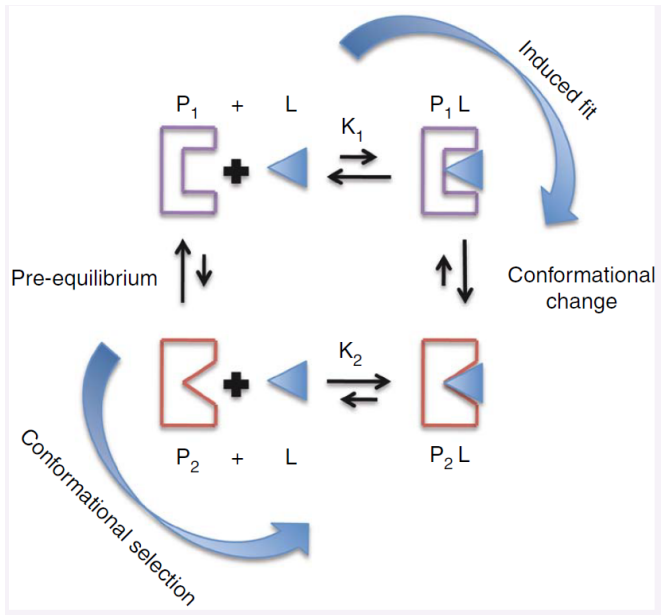
Geometry



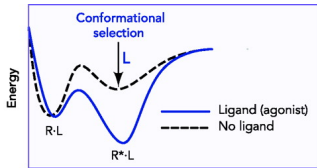
Chemistry



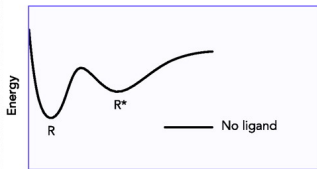
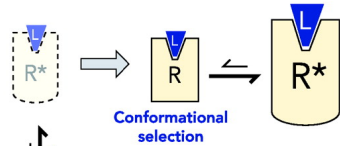
Source: Kohlbacher and Lenhof



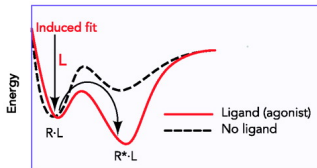
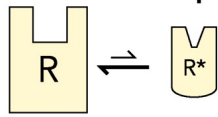
David D Boehr, Ruth Nussinov, and Peter E Wright (Nov. 2009). In: *Nat. Chem. Biol.*



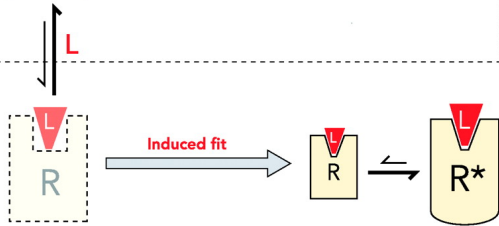
Reaction coordinate (activation)



Reaction coordinate (activation)

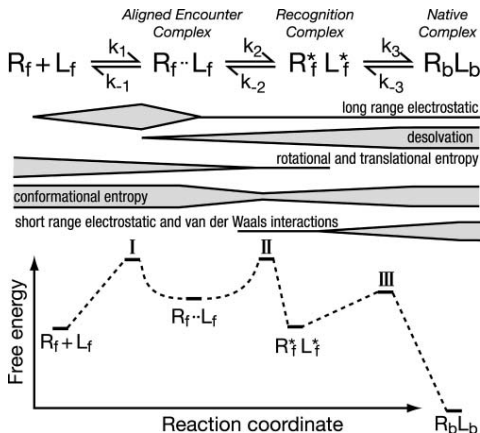


Reaction coordinate (activation)



Flexible Protein Recognition

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



Raik Grünberg, Johan Leckner, and Michael Nilges (Dec. 2004). In: *Structure* 12.12

Types of complexes

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}	$f_{non-nat}$	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]

Protein-Protein Docking Benchmark 5.0 and Affinity Benchmark 2.0

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

Table 2. Composition of the updated docking and affinity benchmarks (in parentheses are values for the previous versions of the benchmarks, docking version 4 and affinity version 1).

	Docking		Affinity	
	<i>N</i>	%	<i>N</i>	%
All	230 (175)		179 (144)	
Enzyme containing	88 (71)	38% (41%)	69 (61)	39% (42%)
Antibody-antigen	40 (24)	17% (14%)	33 (19)	18% (13%)
Others	102 (80)	45% (45%)	77 (64)	43% (45%)
Rigid body ^a	151 (119)	65% (68%)		
Medium ^a	45 (29)	20% (17%)		
Difficult ^a	34 (27)	15% (15%)		
Rigid (I-RMSD < 1.0 Å) ^a			93 (75)	52% (52%)
Flexible (I-RMSD > 1.0 Å) ^a			86 (69)	48% (48%)

^a See the materials and methods section for definition.

Thom Vreven et al. (Sept. 2015). en. In: *Journal of Molecular Biology* 427.19

- 3 Scoring
 - Scoring Functions
 - Shape complementarity

Scoring Functions

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.

Brian Pierce and Zhiping Weng (Jan. 2007). en. In: *Computational Methods for Protein Structure Prediction and Modeling*. Biological and Medical Physics, Biomedical Engineering

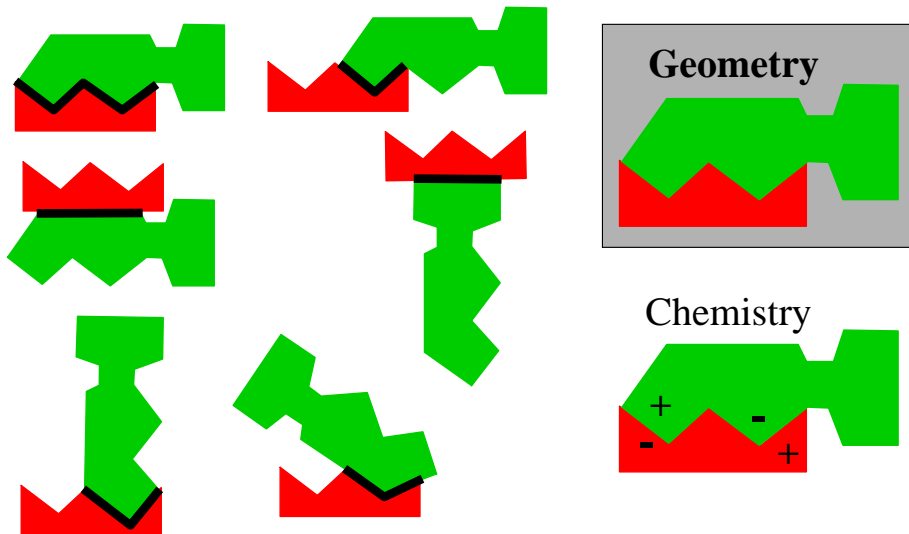
Alternative: Scoring Functions

- 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

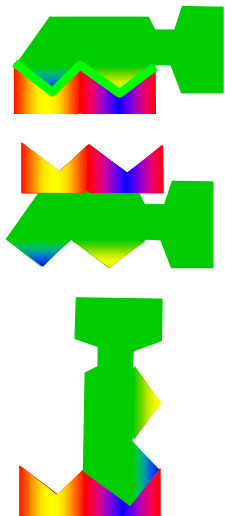
Scoring functions must be accurate and fast at the same time to evaluate several billions of docking poses.

Scoring functions based only on geometry or only on chemistry are not successful in general.

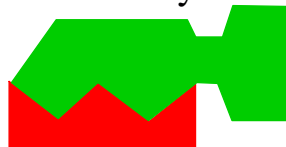
Geometry and Chemistry



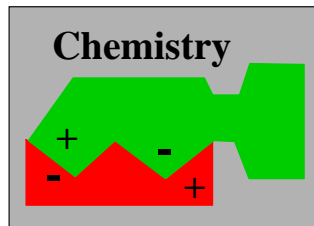
Geometry and Chemistry



Geometry



Chemistry



Geometry

- 1 Steric complementarity of shapes
- 2 Buried surface area (BSA) = $SAS_A + SAS_B - SAS_{AB}$, typical values for complexes: 1200-2200 Å²

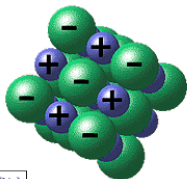
Chemistry

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

Pont salin (salt bridge, ion-pairing)

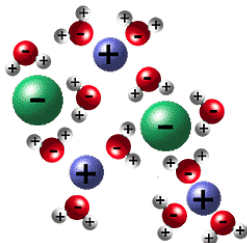
- interaction électrostatique entre deux (ou plus) résidus chargés
- peut inclure des liaisons hydrogènes, ce n'est pas un terme orthogonal à "liaison hydrogène"
- plus qu'une simple liaison hydrogène, forte interaction monopole
- ne dépend pas de la directionnalité
- atténué par l'eau qui fait écran
- pas souvent enfoui dans la protéine, malgré l'absence d'écran, car énergie libre favorable si en contact avec l'eau.

NaCl crystal structure

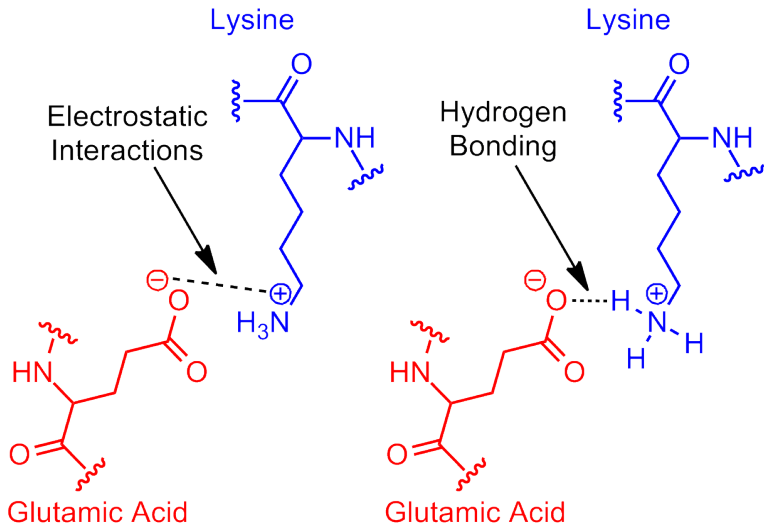


sodium (Na)
chlorine (Cl)

NaCl in water



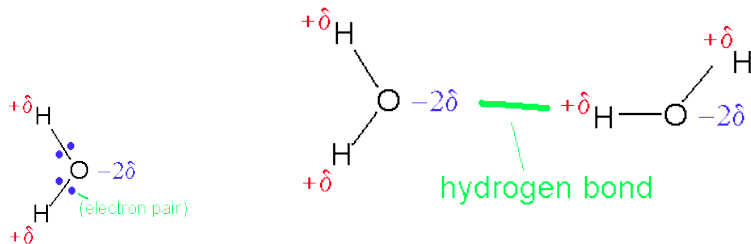
Pont salin (salt bridge or ion-pair)

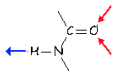


Liaisons hydrogènes (hydrogen bonds)

Liaisons hydrogènes:

- Interaction électrostatique entre des groupements sans charge formelle
- Polarisation: Électrons sont attirés par les atomes électronégatifs (O, N et un peu aussi S), mais restent partagés à travers la liaison covalente
- Ceci crée un moment dipolaire
- => Formation d'une liaison hydrogène
- Distance entre les deux O de H₂O: 2,8 Å





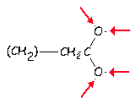
peptide



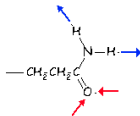
serine



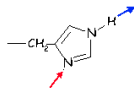
threonine



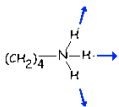
glutamic/aspartic acid



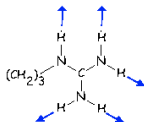
glutamine/asparagine



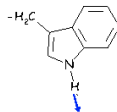
histidine



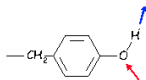
lysine



arginine



tryptophan



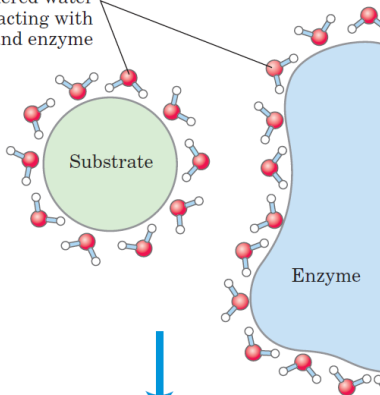
tyrosine

hydrogen bond donor →
hydrogen bond acceptor ←

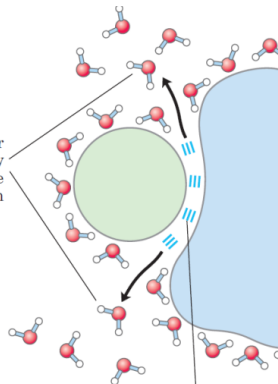
Désolvatation

Désolvatation:

Ordered water interacting with substrate and enzyme



Disordered water displaced by enzyme-substrate interaction



Enzyme-substrate interaction stabilized by hydrogen-bonding, ionic, and hydrophobic interactions

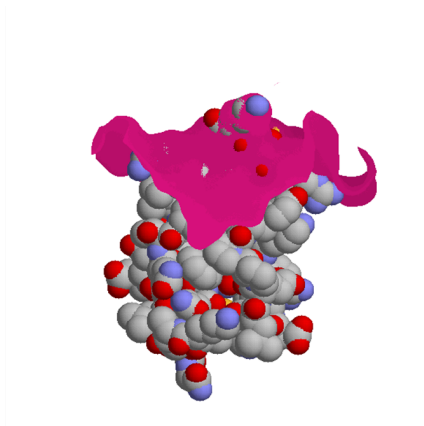
Categories of scoring functions

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

Irina S Moreira, Pedro A Fernandes, and Maria J Ramos (Jan. 2010). In: *J Comput Chem* 31.2

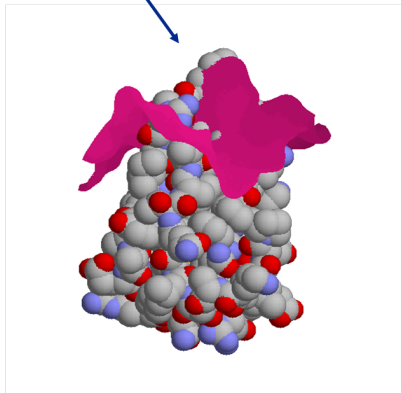
Shape complementarity

Bound VS unbound



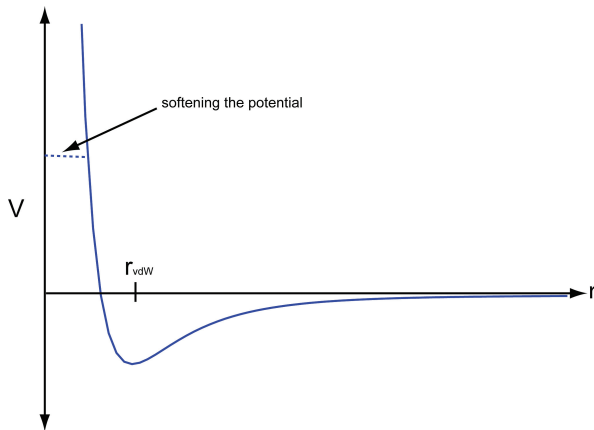
Kallikrein A/trypsin inhibitor complex (PDB codes 2KAI,6PTI)

10 highly penetrating residues

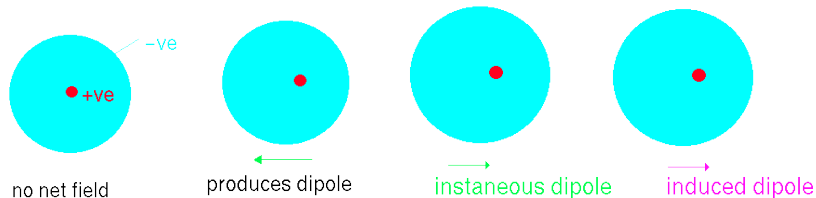


Soft van der Waals

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



Force de dispersion ou van der Waals



- Fluctuations de la densité des électrons autour du noyau d'un atome, créent un moment dipolaire temporaire.
- Ce moment dipolaire induit à un autre atome très proche aussi un moment dipolaire transitoire
- Résultat: force attractive $F_{disp} = -B_{ij}/r_{ij}^6$, faible mais additive

van der Waals - potentiel de Lennard-Jones

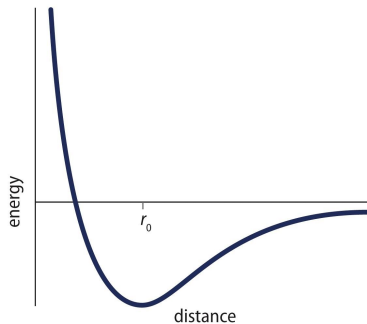


Figure 1.20 How Proteins Work (©2012 Garland Science)

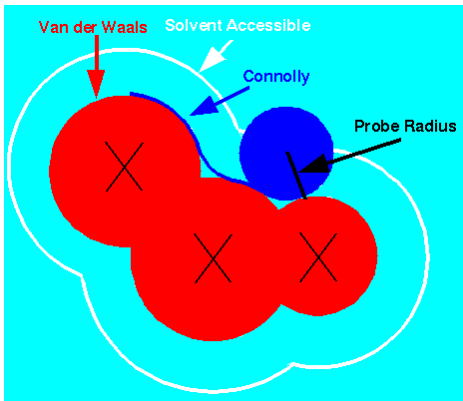
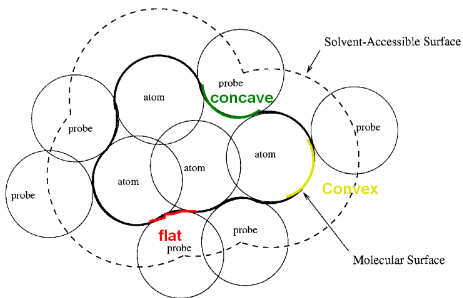
- Résultat: force attractive $F_{disp} = -B_{ij}/r_{ij}^6$, faible mais additive
- Répulsion (principe de Pauli): $F_{rep} = A_{ij}/r_{ij}^{12}$
- Total: $F_{total} = A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6$
- Rayons de van der Waals (minimum du potentiel r_0):
C (1,85Å), O (1,60Å), H (1,0Å), N (1,75Å)

- 4 Rigid-body docking
 - Geometric docking
 - Fast Fourier Transform (FFT) docking

Geometric docking

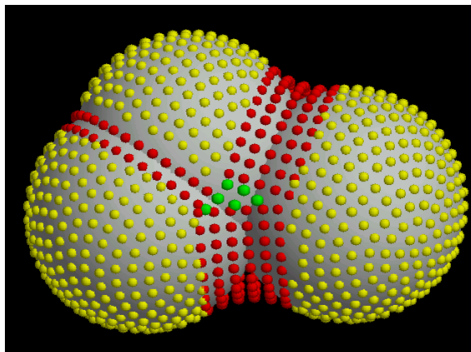
Solvent accessible surface - SAS

Connolly's MS (molecular surface) algorithm

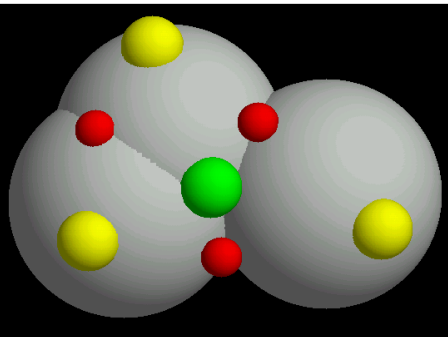


Cai 1998 / <http://www.simbiosys.ca/sprout/eccc/cangaroo.html>

Dot surface VS critical points

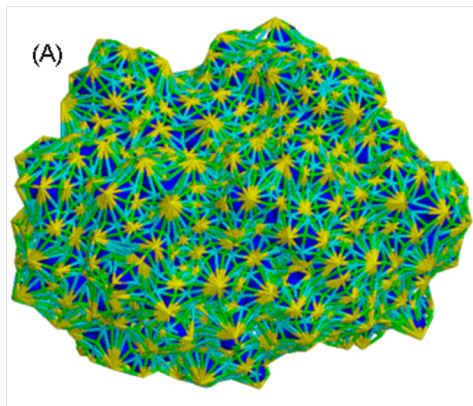
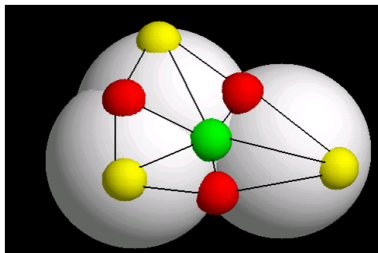


(a) dense, Connolly



(b) sparse, Lin et al. 1994

green = concave, yellow = convex, red = flat

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/>

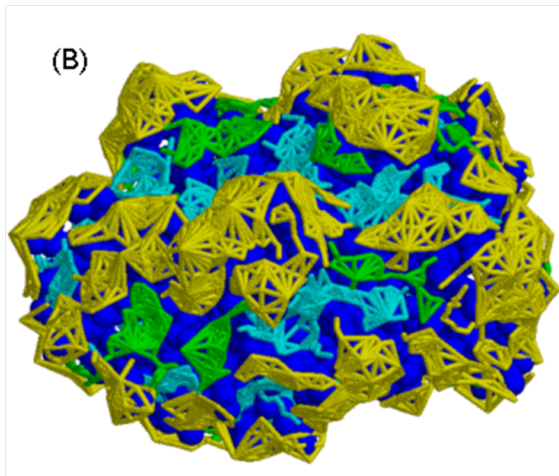
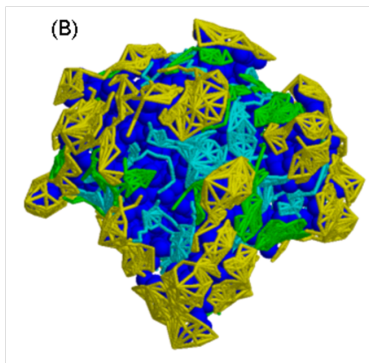
Group critical points as patches

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 \leftrightarrow hole patches and flat patches \leftrightarrow any patch

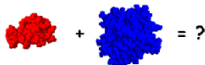
- 1 *Single Patch Matching*: One patch of the receptor with one patch of the ligand, for small ligands
- 2 *Patch-Pair Matching*: Two patches of the receptor with two patches of the ligand, for protein-protein complexes

Match critical points within patches by computer vision techniques:

- Geometric Hashing
- Pose Clustering

Dina Duhovny, Ruth Nussinov, and Haim J. Wolfson (2002). In: *In WABI '02: Proceedings of the Second International Workshop on Algorithms in Bioinformatics*

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:

Complex Type:

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

Advanced Options:

[\[Show\]](#) [\[Hide\]](#)

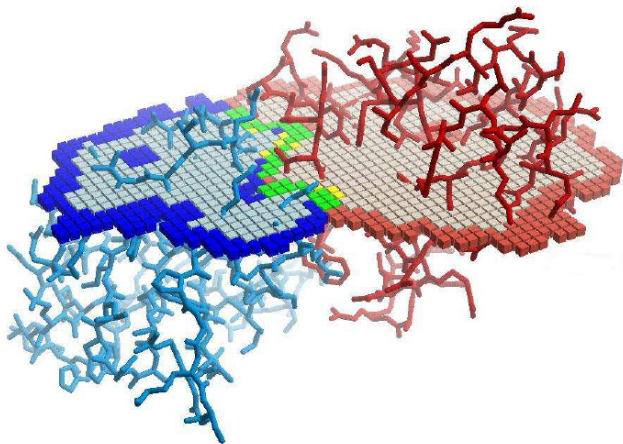
FireDock - Fast Interaction Refinement in Molecular Docking

SymmDock - An Algorithm for Prediction of Complexes with C_n Symmetry

Beta 1.3 Version, Contact: duhovka@gmail.com

Fast Fourier Transform (FFT) docking

3D grid

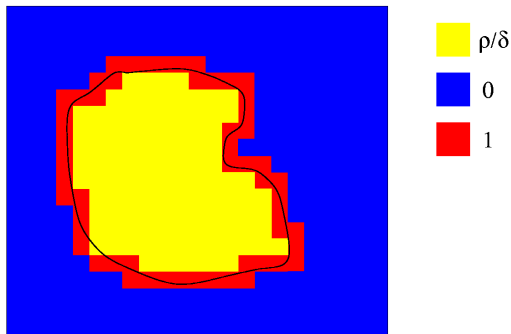


P N Palma et al. (June 2000). In: *Proteins* 39.4
P Nuno Palma (July 2003). In: *Proteins* 52.1

Ludwig Krippahl, José J Moura, and

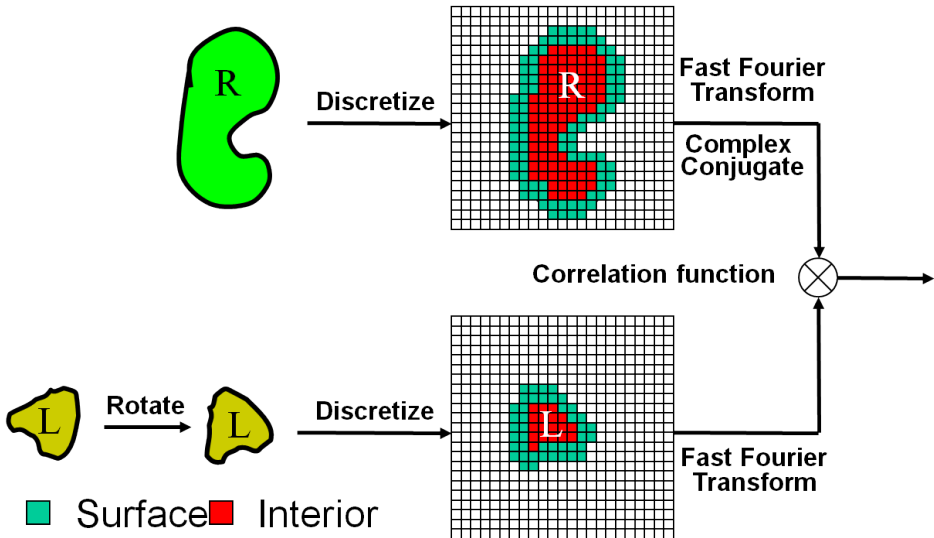
Katchalski-Katzir et al., PNAS 1992

- Protein on grid
- Assign values
 - $a_{i,j,k} =$
 - 1 at the surface of A
 - $\rho \ll 0$ inside A
 - 0 outside
 - $b_{i,j,k} =$
 - 1 at the surface of B
 - $\delta > 0$ inside B
 - 0 outside B

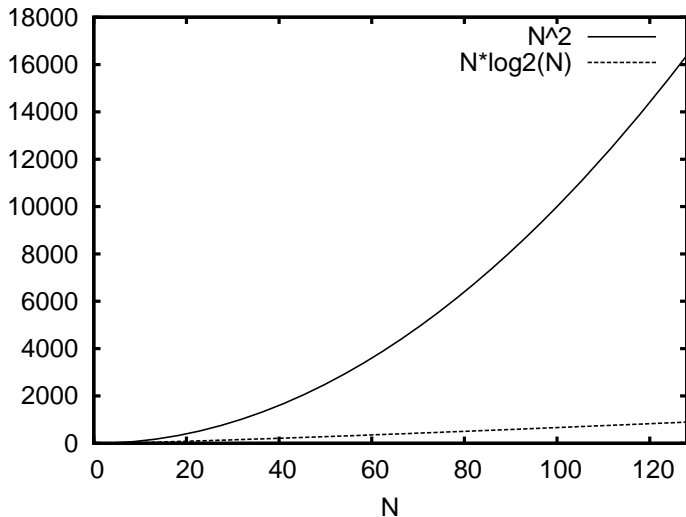


A \ B	inside	surface	outside
inside	$\rho^* \delta < 0$	$\rho < 0$	0
surface	$\delta > 0$	1	0
outside	0	0	0

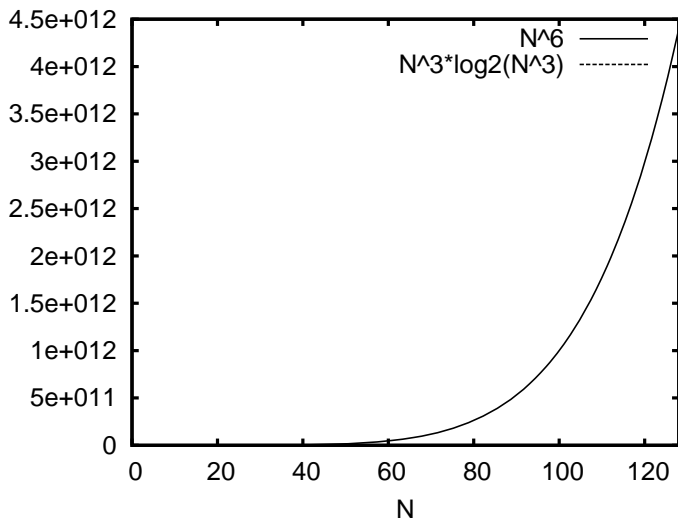
Discrete Fast Fourier Transform



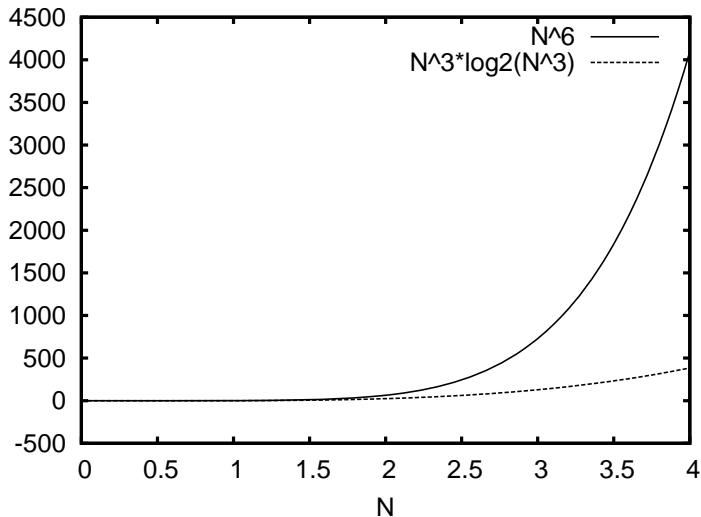
FFT speedup - 1D



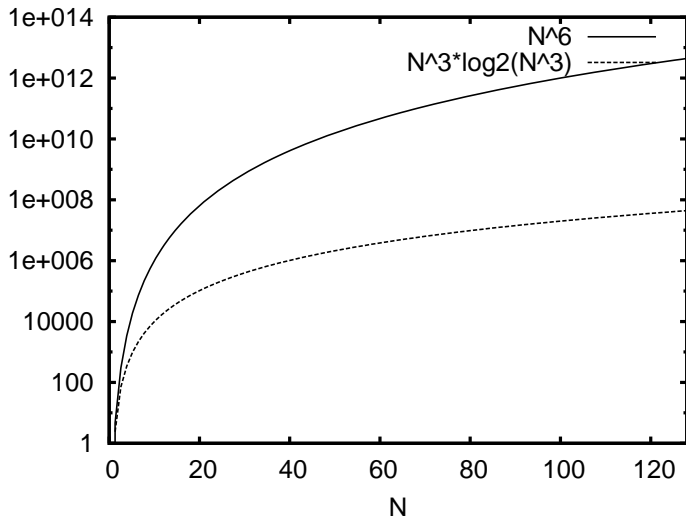
FFT speedup - 3D



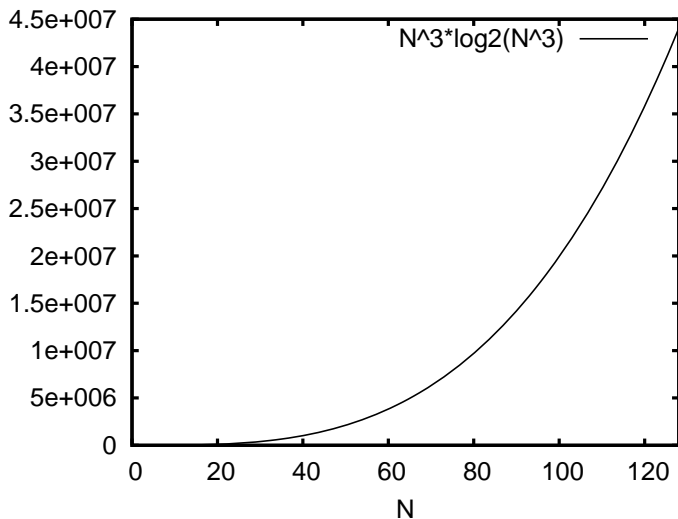
FFT speedup - 3D



FFT speedup - 3D



FFT speedup - 3D



ZDOCK: a FFT docking program

- 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 \Rightarrow 2 million different translation vectors (α, β, γ)
- Without FFT $\Rightarrow 128^6 = 4.4 \cdot 10^{12} = 4400$ billion elementary operations (addition or multiplication)
- With FFT $\Rightarrow 128^3 \cdot \log_2(128^3) = 2.1 \cdot 10^6 \cdot 21 = 44$ million elementary operations

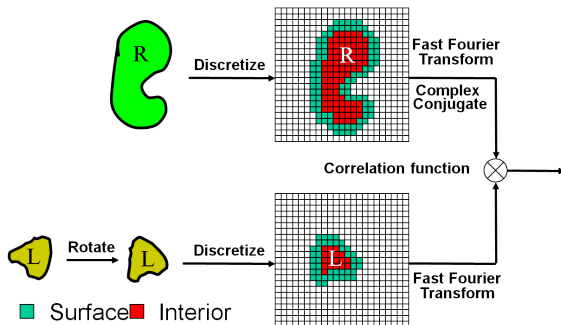
$\Rightarrow 10^5$ times faster with FFT !

Rong Chen and Zhiping Weng (May 2002). In: *Proteins* 47.3

Ligand rotations

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

- 1 $\Delta = 15 \text{ degrees} \Rightarrow M_{rot} = 3600$
 $\Rightarrow M_{rot} \cdot N^3 = 7.5 \text{ billion docking poses}$
- 2 $\Delta = 6 \text{ degrees} \Rightarrow M_{rot} = 54000$
 $\Rightarrow M_{rot} \cdot N^3 = 113 \text{ billion docking poses}$



Total number of operations

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

$$M_{total} = M_{rot} \cdot M_{trans+corr} = M_{rot} \cdot N^3 \cdot \log_2(N^3) \quad (4)$$

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)

Brian G Pierce, Yuichiro Hourai, and Zhiping Weng (2011). In: *PLoS ONE* 6.9

- 5 Evaluation
 - Performance of docking programs
 - CAPRI

Performance of docking programs

CAPRI

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/>

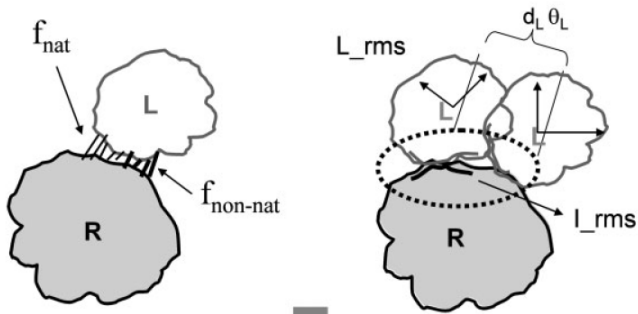
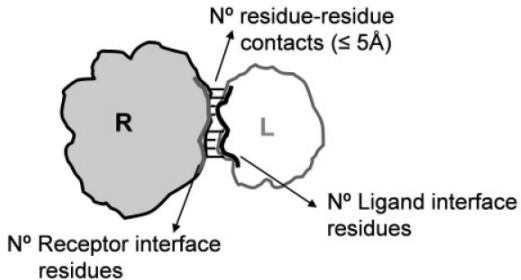
CAPRI star evaluation

The CAPRI star system

Mendez, Lepplae,
Wodak 2003
Lensink et al.
2005, 2007, 2010

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

Source: Janin, LIX 2010



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		

Web-server (en 2010)

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

Marc F Lensink and Shoshana J Wodak (Nov. 2010). In: *Proteins* 78.15

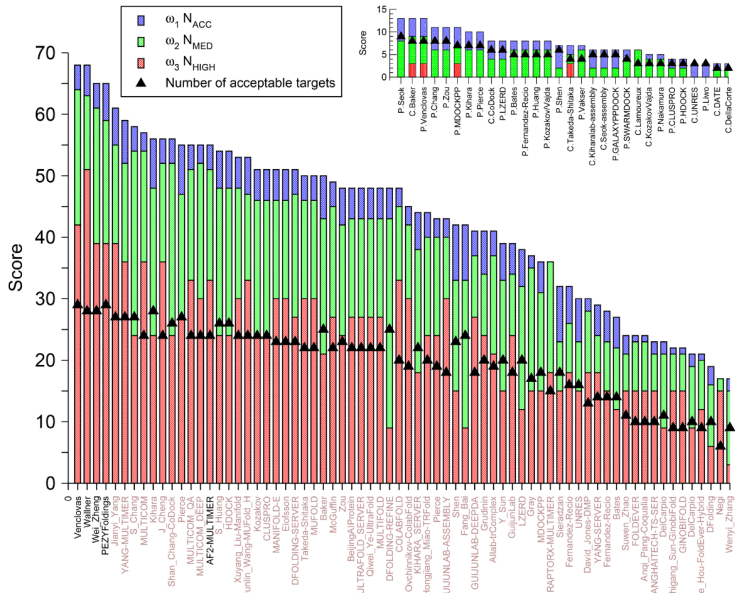
Conclusion (before AlphaFold)

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"

CASP 15 (2022): joint CASP-CAPRI experiment



CASP 15 (2022): joint CASP-CAPRI experiment

- Huge improvement from previous CASP-CAPRI (2020) experiment, due to AlphaFold 2 and AlphaFold-Multimer
- for 40% of the targets high quality models have been obtained (previously: only 8%)
- Webservers are on par with human groups
- Antibodies and Nanobodies remain challenging
- Targets with important conformational flexibility remain challenging

Marc F. Lensink et al. (2023). en. In: *Proteins: Structure, Function, and Bioinformatics*

91.12

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 there are still a lot of challenges.

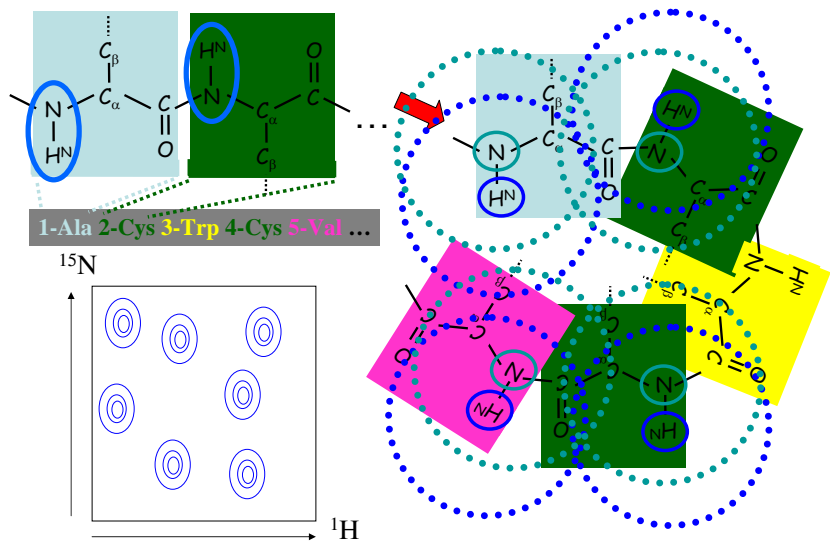
One possible solution:

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

- 6 Inclusion of experimental data
 - NMR - chemical shifts
 - CS-HADDOCK

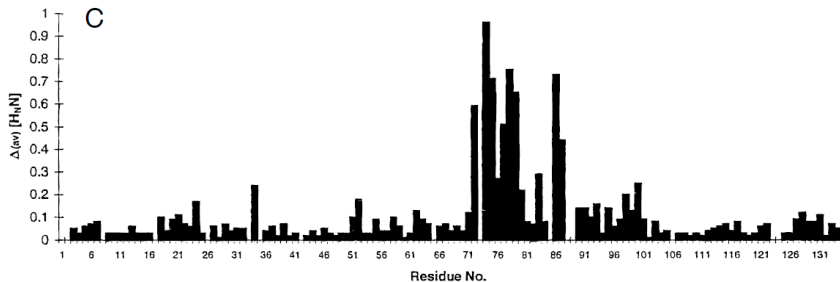
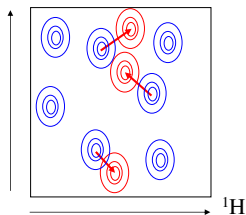
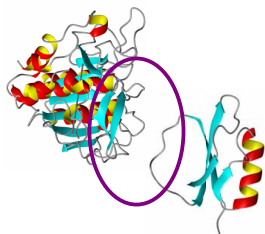
NMR - chemical shifts

Chemical shift

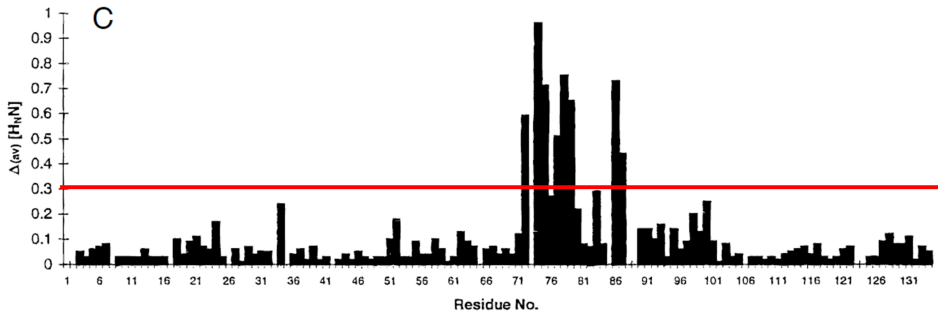


Chemical Shift Perturbation (CSP)

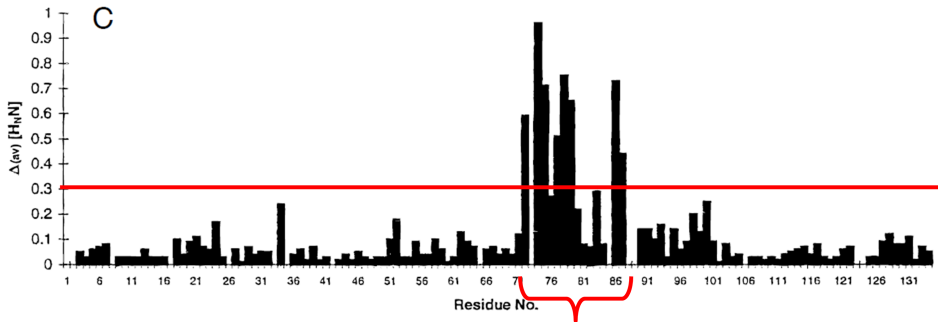
1-Ala 2-Cys 3-Trp 4-Cys 5-Val ...

 ^{15}N 

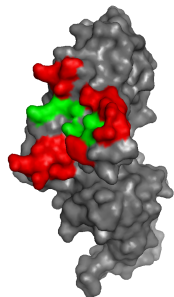
Chemical Shift Perturbation (CSP)



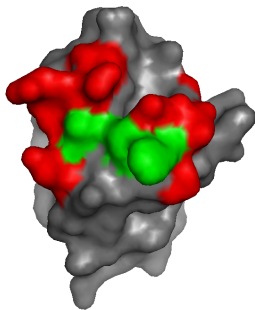
Chemical Shift Perturbation (CSP)



Interface localization on 3D structures



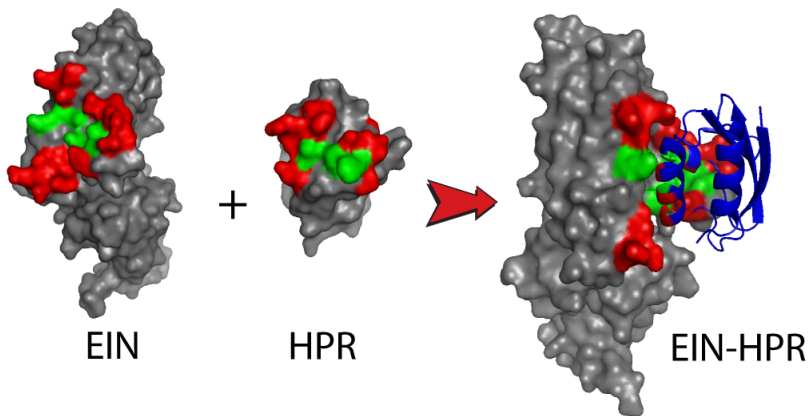
EIN



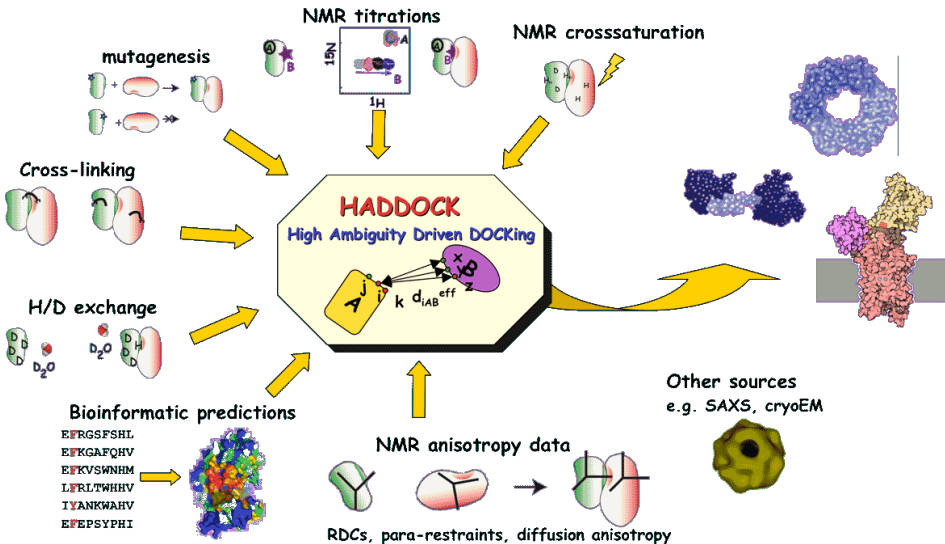
HPR

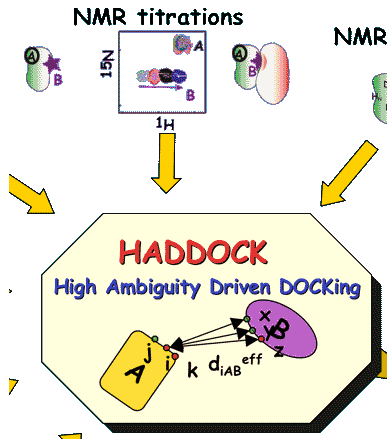
red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues

Docking

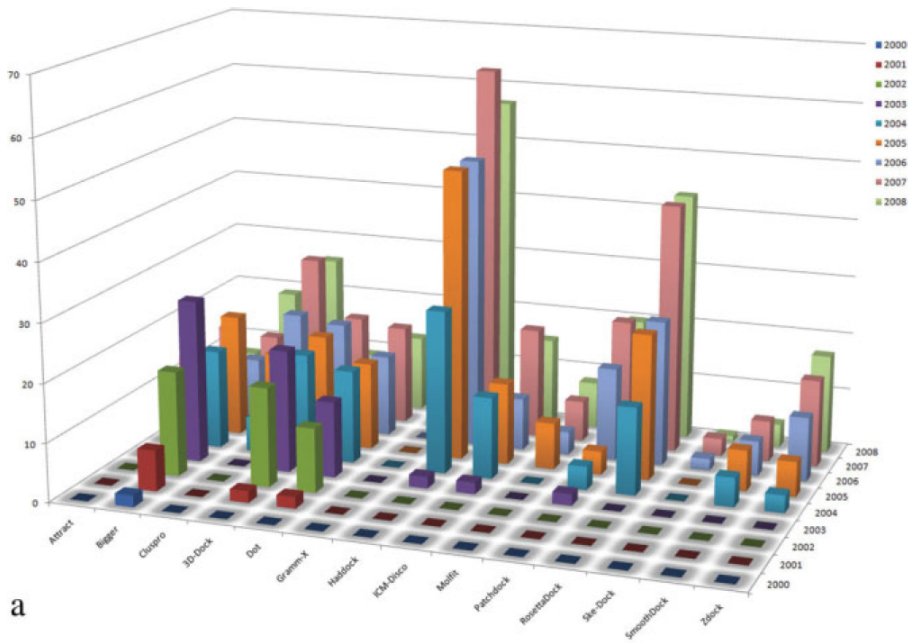


red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues



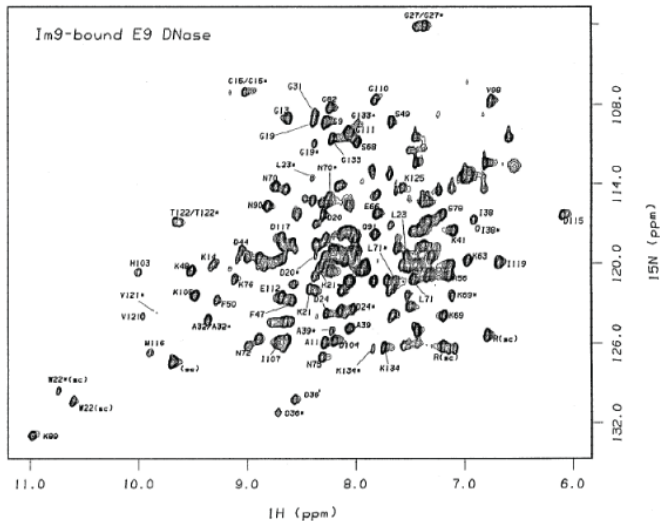


$$E_{\text{Haddock}} = E_{\text{vdW}} + E_{\text{elec}} + E_{\text{AIR}} + E_{\text{desolv}}$$

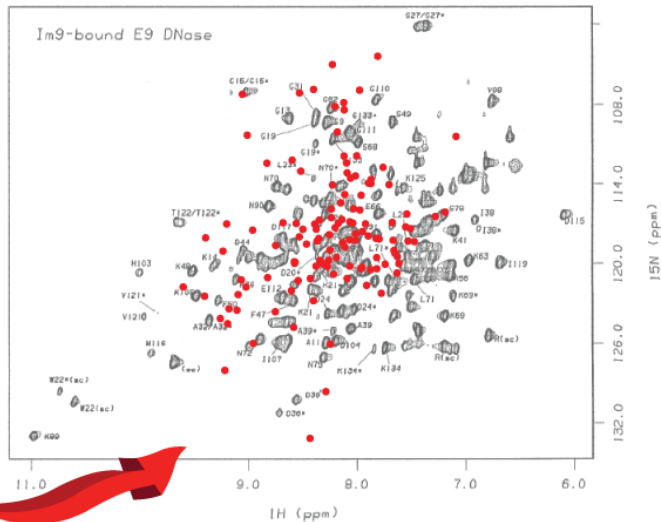
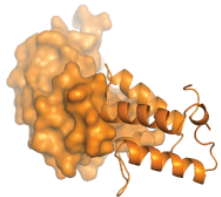


CS-HADDOCK

3D to CS



3D to CS

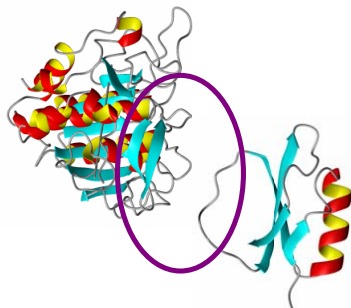


3D to CS with ShiftX

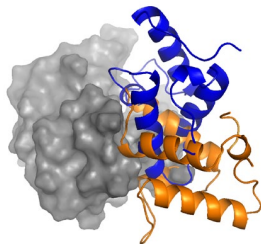
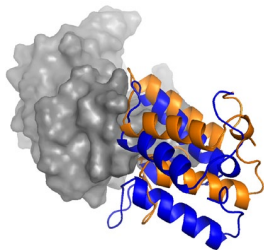
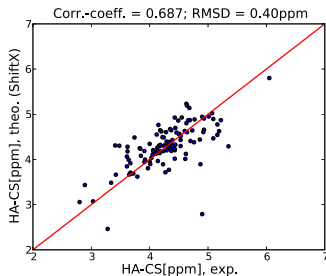
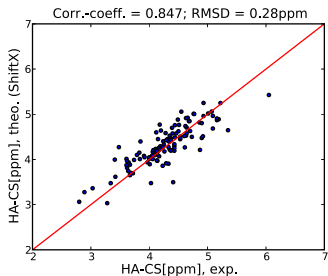
Contributions to calculated CS δ_{calc} :

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

- 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 $^1H^\alpha$ -CS

Protocole d'arrimage CS-HADDOCK

Structures 3D protéines libres

Données RMN (CSP)

Arrimage avec HADDOCK 2.1

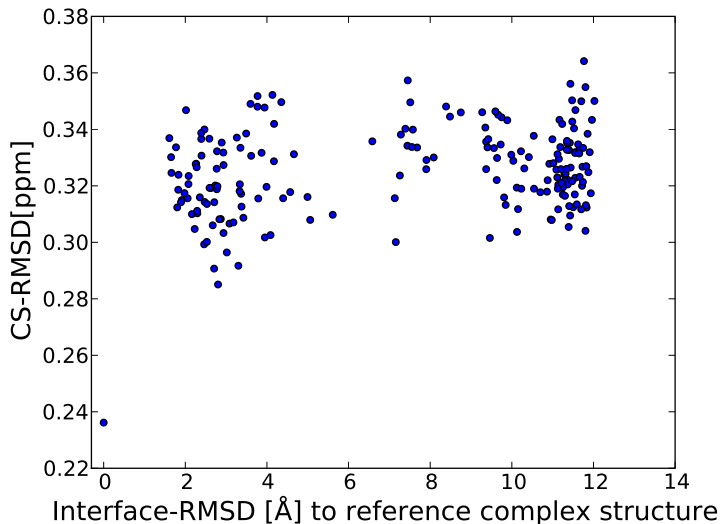
Calculer H α CS avec ShiftXH α CS exp. du complexe

Calculer CS-RMSDs

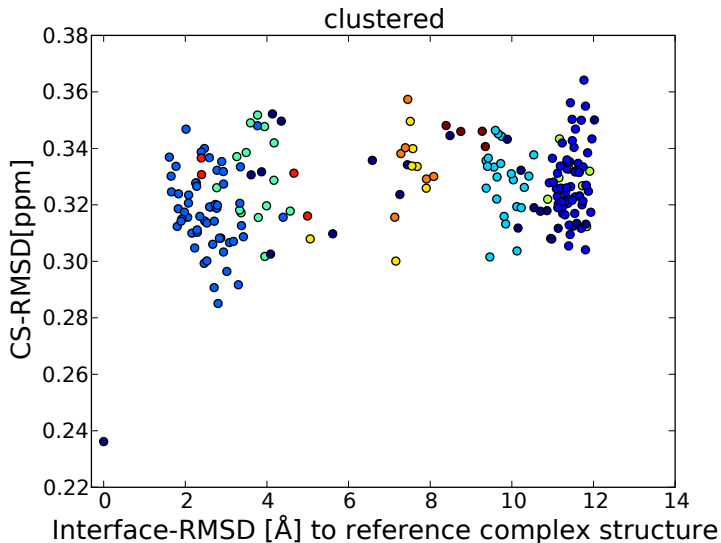
CS-RMSD =

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

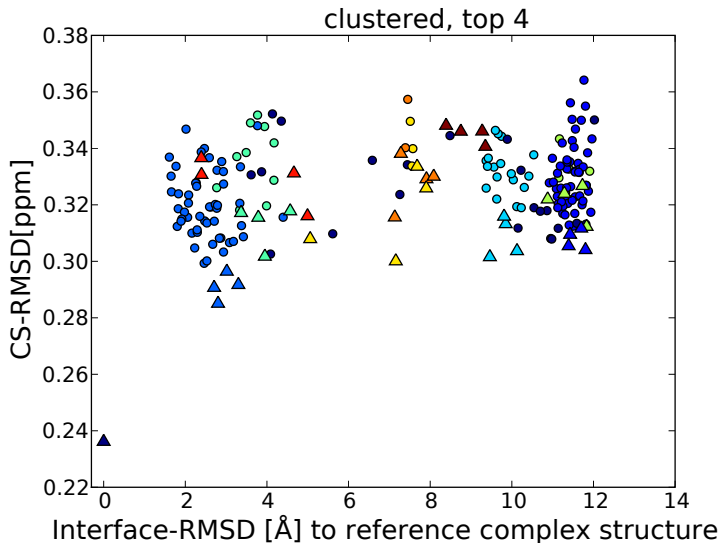
CS-RMSD scoring on all generated structures



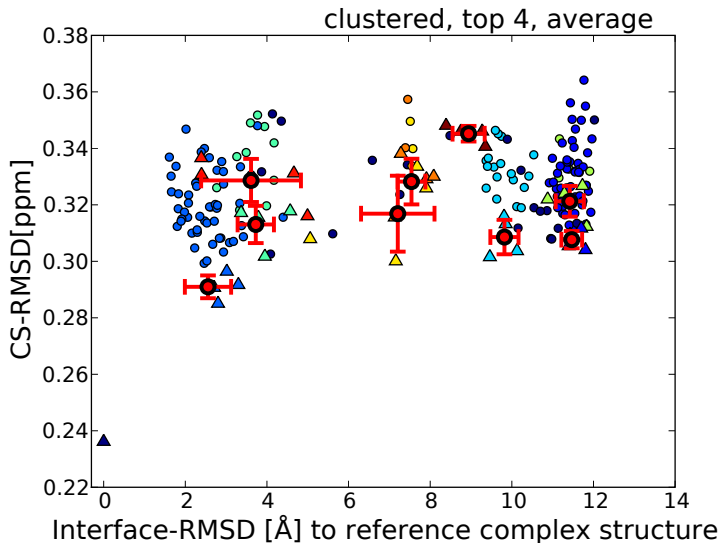
CS-RMSD scoring on all generated structures



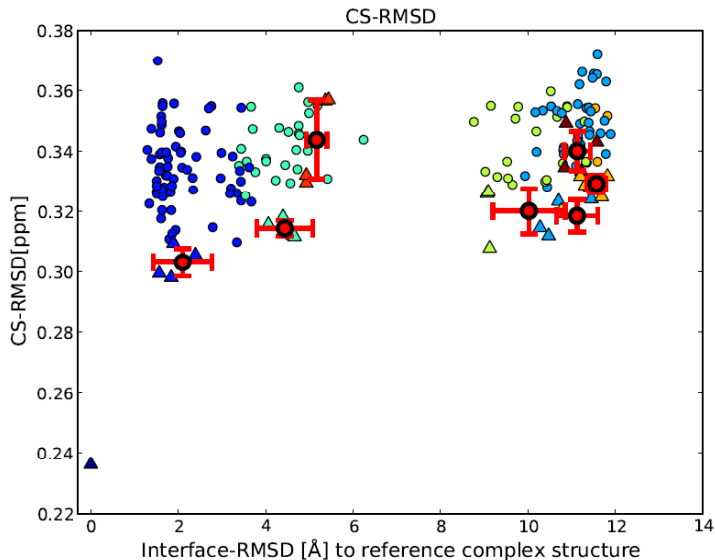
CS-RMSD scoring on all generated structures



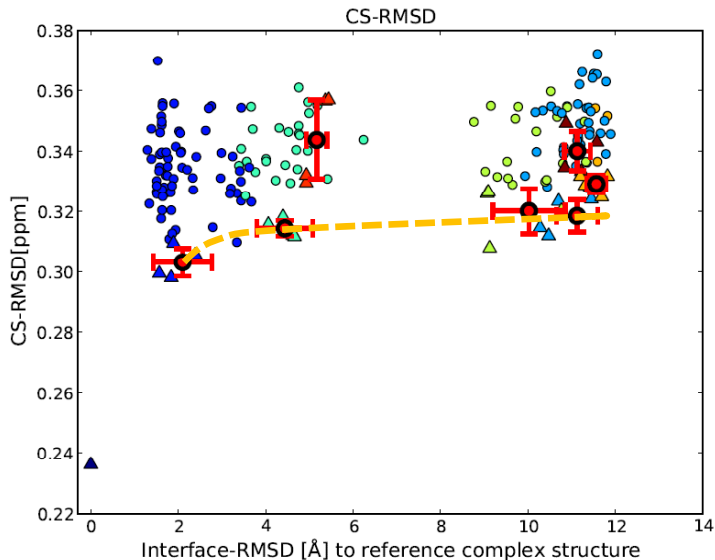
CS-RMSD scoring on all generated structures



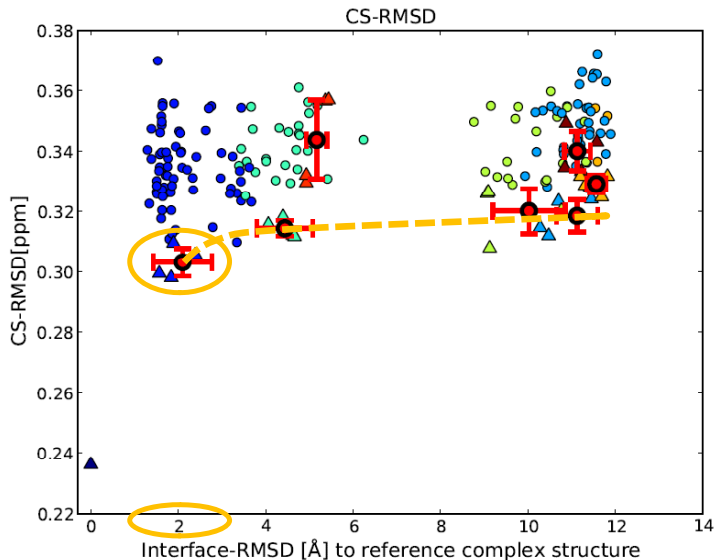
Classement des clusters de structures par CS-RMSD



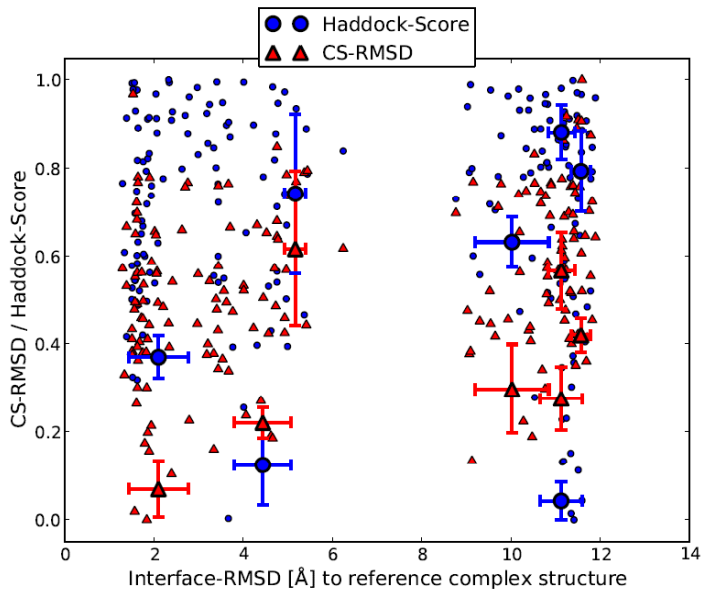
Classement des clusters de structures par CS-RMSD



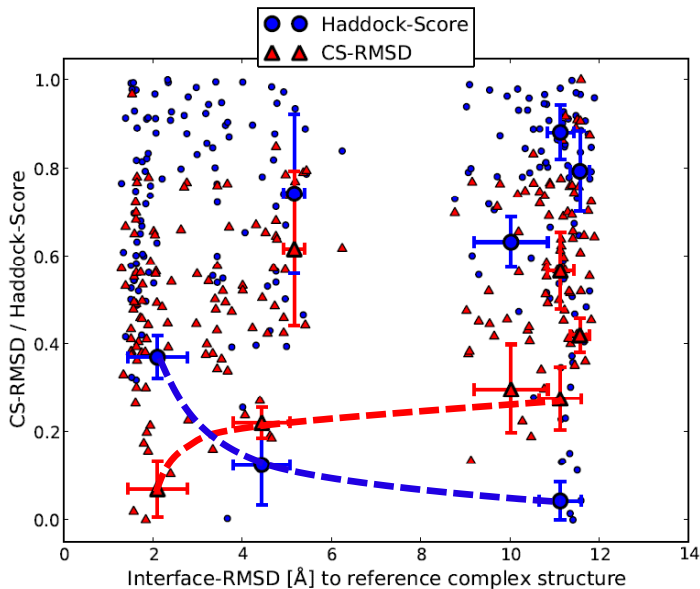
Classement des clusters de structures par CS-RMSD



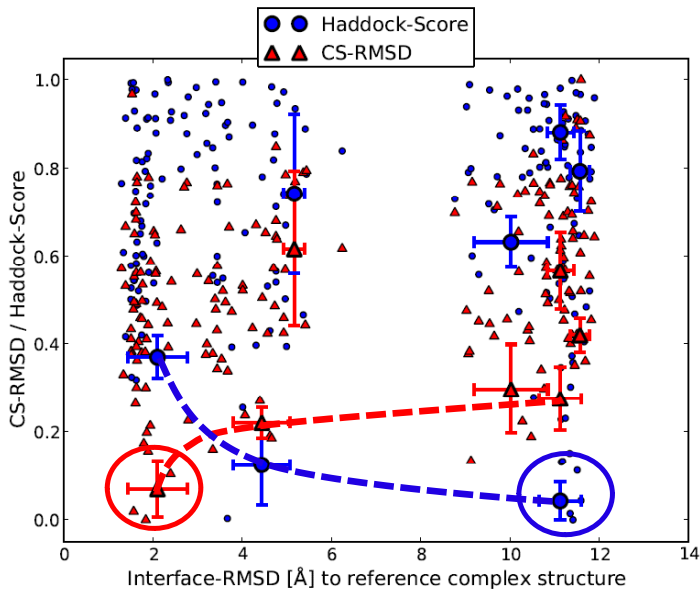
CS-HADDOCK vs HADDOCK



CS-HADDOCK vs HADDOCK

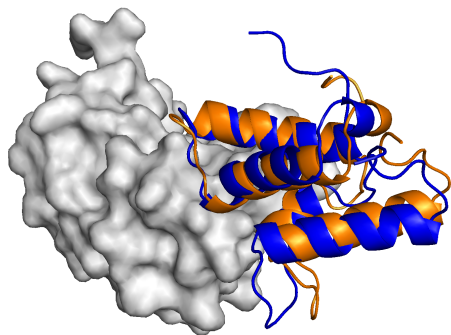


CS-HADDOCK vs HADDOCK

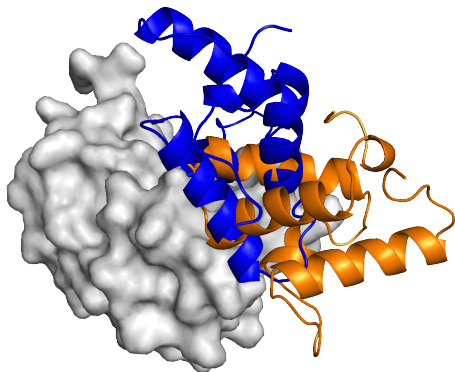


CS-HADDOCK vs HADDOCK

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









(c) CS-RMSD score









(d) HADDOCK score

7 Bibliography








Bibliography I

- 
- Boehr, David D, Ruth Nussinov, and Peter E Wright (Nov. 2009). “The role of dynamic conformational ensembles in biomolecular recognition”. In: *Nat. Chem. Biol.* 5.11. PMID: 19841628, pp. 789–796.
- 
- Chen, Rong and Zhiping Weng (May 2002). “Docking unbound proteins using shape complementarity, desolvation, and electrostatics”. In: *Proteins* 47.3. PMID: 11948782, pp. 281–294.
- 
- Connolly, M L (July 1986). “Shape complementarity at the hemoglobin alpha 1 beta 1 subunit interface”. In: *Biopolymers* 25.7. PMID: 3741993, pp. 1229–1247.
- 
- Deupi, Xavier and Brian K. Kobilka (Jan. 2010). “Energy Landscapes as a Tool to Integrate GPCR Structure, Dynamics, and Function”. en. In: *Physiology* 25.5, pp. 293–303.
- 
- Duhovny, Dina, Ruth Nussinov, and Haim J. Wolfson (2002). “Efficient unbound docking of rigid molecules”. In: *In WABI '02: Proceedings of the Second International Workshop on Algorithms in Bioinformatics*. Springer Verlag, pp. 185–200.
- 
- Grünberg, Raik, Johan Leckner, and Michael Nilges (Dec. 2004). “Complementarity of structure ensembles in protein-protein binding”. In: *Structure* 12.12. PMID: 15576027, pp. 2125–2136.





Bibliography II

-  Halperin, Inbal et al. (June 2002). “Principles of docking: An overview of search algorithms and a guide to scoring functions”. In: *Proteins* 47.4. PMID: 12001221, pp. 409–443.
-  Krippahl, Ludwig, José J Moura, and P Nuno Palma (July 2003). “Modeling protein complexes with BiGGER”. In: *Proteins* 52.1. PMID: 12784362, pp. 19–23.
-  Lensink, Marc F and Shoshana J Wodak (Nov. 2010). “Docking and scoring protein interactions: CAPRI 2009”. In: *Proteins* 78.15. PMID: 20806235, pp. 3073–3084.
-  Lensink, Marc F. et al. (2023). “Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment”. en. In: *Proteins: Structure, Function, and Bioinformatics* 91.12. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26609>, pp. 1658–1683.
-  Méndez, Raúl et al. (Aug. 2005). “Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures”. In: *Proteins* 60.2. PMID: 15981261, pp. 150–169.
-  Moreira, Irina S, Pedro A Fernandes, and Maria J Ramos (Jan. 2010). “Protein-protein docking dealing with the unknown”. In: *J Comput Chem* 31.2. PMID: 19462412, pp. 317–342.

Bibliography III

-  Nooren, I. M.A. (July 2003). “NEW EMBO MEMBER’S REVIEW: Diversity of protein-protein interactions”. en. In: *The EMBO Journal* 22.14, pp. 3486–3492.
-  Palma, P N et al. (June 2000). “BiGGER: a new (soft) docking algorithm for predicting protein interactions”. In: *Proteins* 39.4. PMID: 10813819, pp. 372–384.
-  Pierce, Brian and Zhiping Weng (Jan. 2007). “Structure Prediction of Protein Complexes”. en. In: *Computational Methods for Protein Structure Prediction and Modeling*. Ed. by Ying Xu, Dong Xu, and Jie Liang. Biological and Medical Physics, Biomedical Engineering. Springer New York, pp. 109–134.
-  Pierce, Brian G, Yuichiro Hourai, and Zhiping Weng (2011). “Accelerating protein docking in ZDOCK using an advanced 3D convolution library”. In: *PLoS ONE* 6.9. PMID: 21949741, e24657.
-  Scott, Duncan E. et al. (Aug. 2016). “Small molecules, big targets: drug discovery faces the protein–protein interaction challenge”. en. In: *Nat Rev Drug Discov* 15.8. Number: 8 Publisher: Nature Publishing Group, pp. 533–550.
-  Siebenmorgen, Till and Martin Zacharias (May 2020). “Computational prediction of protein–protein binding affinities”. en. In: *WIREs Comput Mol Sci* 10.3, e1448.
-  Sikosek, Tobias and Hue Sun Chan (Nov. 2014). “Biophysics of protein evolution and evolutionary protein biophysics”. en. In: *J. R. Soc. Interface*. 11.100, p. 20140419.

Bibliography IV

-  Smith, Graham R and Michael J E Sternberg (Feb. 2002). “Prediction of protein-protein interactions by docking methods”. In: *Curr. Opin. Struct. Biol.* 12.1. PMID: 11839486, pp. 28–35.
-  Stratmann, Dirk, Rolf Boelens, and Alexandre M J J Bonvin (Sept. 2011). “Quantitative use of chemical shifts for the modeling of protein complexes”. In: *Proteins* 79.9. PMID: 21744392, pp. 2662–2670.
-  Vreven, Thom et al. (Sept. 2015). “Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2”. en. In: *Journal of Molecular Biology* 427.19, pp. 3031–3041.
-  Xu, Qifang and Roland L Dunbrack (Apr. 2019). “Principles and characteristics of biological assemblies in experimentally determined protein structures”. In: *Current Opinion in Structural Biology. Macromolecular Assemblies * Theory and simulation: Demystifying GPCRs* 55, pp. 34–49.