BIOINFORMÁTICA Y BIOLOGÍA COMPUTACIONAL

Curso de la Escuela Complutense de Verano (Julio 2007)

Docking de Proteínas

Juan Fernández Recio

Barcelona Supercomputing Center (BSC)

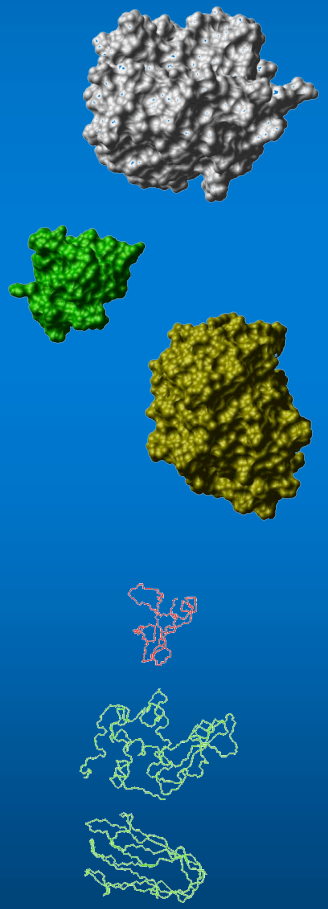
- **Introduction**
- Computational protein-protein docking
- Geometric docking algorithms
- Docking by global energy optimization
- Comparison of docking methods
- Present and future challenges in protein-protein docking

Protein-Protein Interactions: Importance

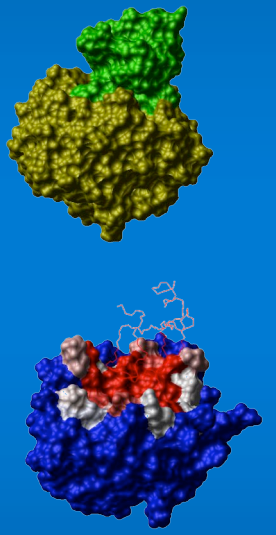
gene
(genomics)

AGFHICVQVYENK
ASFHICVQVYENK
MGFHICVQVYQNK
LGFHICVQVY.NK

protein
(proteomics)



function
(interactomics)

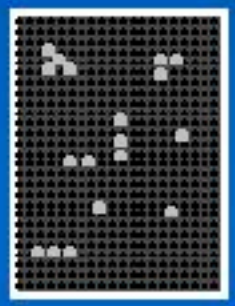


Protein-Protein Interactions: Analysis

Interface Identification

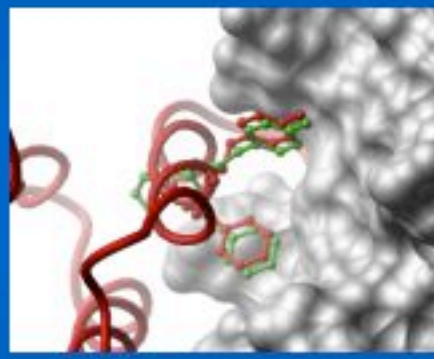
- surface analysis
- gene analysis
- NMR
- mutants

Protein Interaction Detection



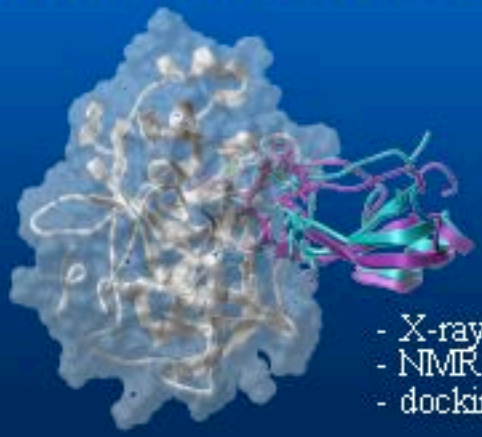
- two-hybrid test
- affinity column, gel assays...
- BIAcore
- mass-spectrometry
- electron microscopy
- cross-linking
- co-immunoprecipitation
- immunofluorescence
- knock-out
- phylogenetic profiles, gene fusion events...
- ...

Applications



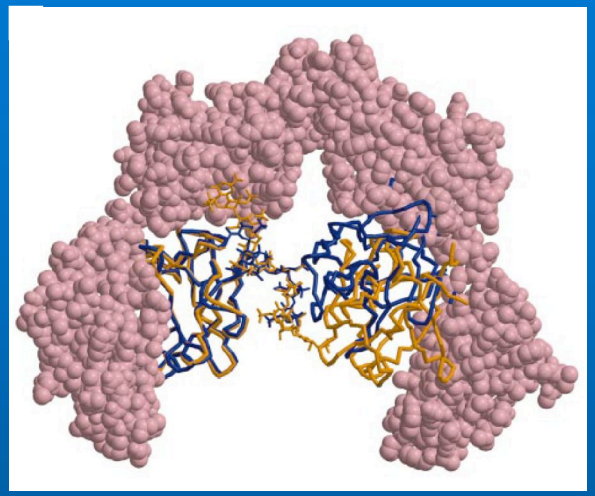
- protein design
- drug discovery
- mechanism of associat

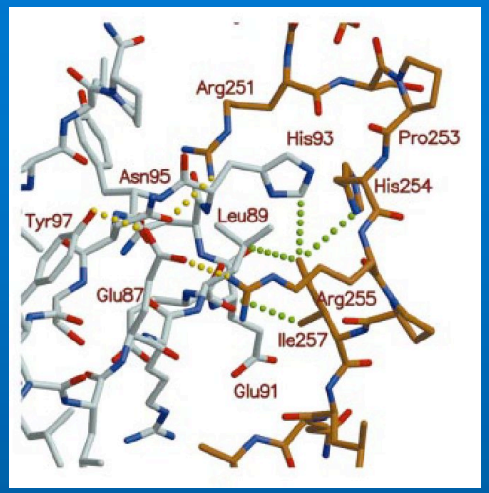
Structural Characterization

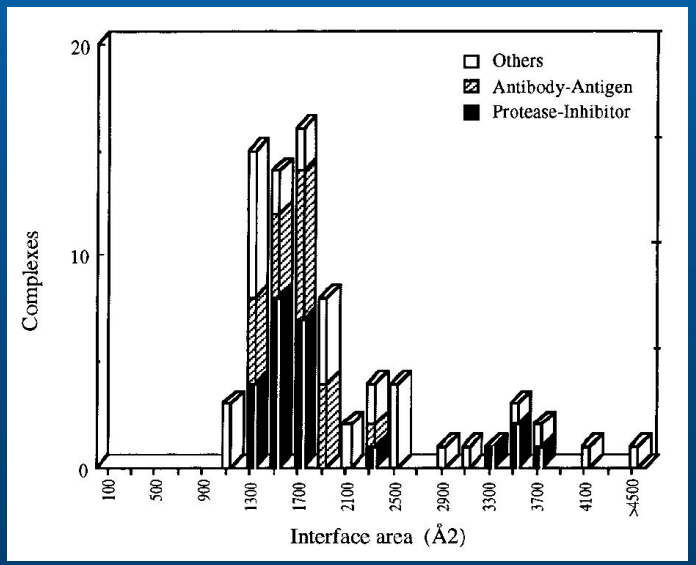


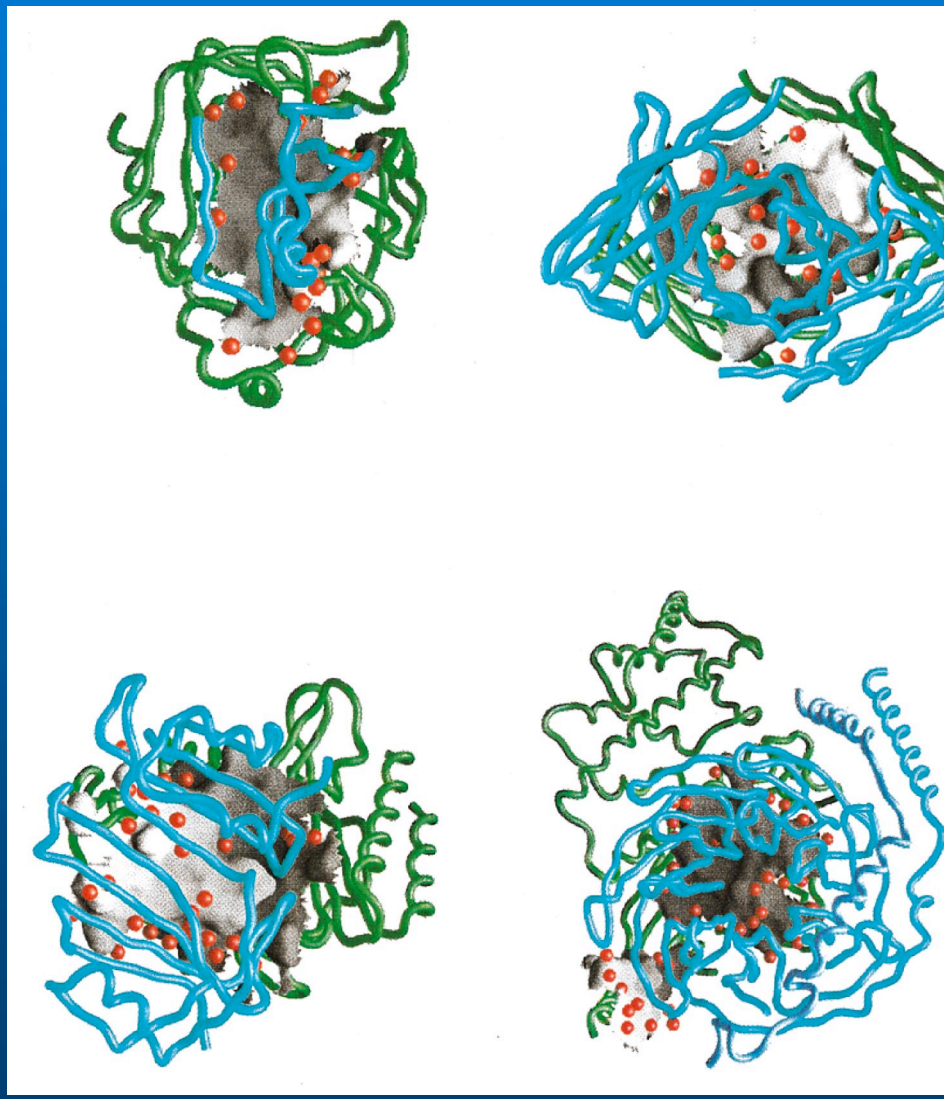
- X-ray
- NMR
- docking simulations

Structural Analysis at Atomic Resolution: NMR and X-ray









Structural Analysis at Atomic Resolution: NMR and X-ray

TABLE II. Amino Acid Composition of Protein-Protein Interfaces

	Number (a)			Area (b)			Propensities (c)			
Residue	Interface	Core	Rim	Interface	Core	Rim	Core	Rim	Lo Conte et al. (d)	Jones and Thornton (e)
All	100.0	100.0	99.9	99.9	100.0	100.0				
Ala	3.9	4.0	3.8	2.8	2.7	3.1	−0.40	−0.26	−0.43	−0.17
Arg	6.4	5.9	7.0	10.1	10.1	9.9	0.13	0.11	0.13	0.27
Asn	5.9	5.4	6.4	5.7	5.4	6.4	−0.14	0.03	−0.12	0.12
Asp	6.6	5.4	8.0	5.1	4.5	6.6	−0.46	−0.07	−0.31	−0.38
Cys	3.5	4.7	2.1	1.7	1.9	1.3	1.00	0.62	0.76	0.43
Gln	3.7	3.7	3.8	4.3	4.3	4.2	−0.34	−0.36	−0.36	−0.11
Glu	6.5	4.6	8.6	6.0	4.4	10.0	−0.80	0.02	−0.47	−0.13
Gly	8.1	7.5	8.7	4.8	4.2	6.4	−0.08	0.35	0.02	−0.07
His	3.4	4.4	2.3	3.8	4.4	2.4	0.84	0.23	0.64	0.41
Ile	3.6	4.1	3.1	4.6	4.9	3.5	0.71	0.38	0.56	0.44
Leu	5.0	5.5	4.5	5.7	5.8	5.3	0.34	0.25	0.29	0.40
Lys	5.7	3.7	8.0	6.5	5.2	9.7	−0.82	−0.20	−0.57	−0.36
Met	2.0	2.6	1.4	3.2	3.7	2.0	1.13	0.51	0.98	0.66
Phe	3.5	5.1	1.7	4.1	5.5	1.1	1.01	−0.60	0.79	0.82
Pro	3.8	3.4	4.2	3.6	3.5	4.1	−0.38	−0.22	−0.25	−0.25
Ser	7.9	7.8	8.1	5.4	4.8	7.3	−0.56	−0.14	−0.42	−0.33
Thr	6.2	5.7	6.8	5.0	4.7	5.9	−0.44	−0.21	−0.35	−0.18
Trp	2.8	4.1	1.3	4.2	5.3	1.6	1.41	0.21	1.25	0.83
Tyr	6.8	8.1	5.4	9.4	10.9	5.3	1.22	0.50	1.04	0.66
Val	4.5	4.3	4.7	3.8	3.8	3.9	0.08	0.11	0.09	0.27

Database of Protein-Protein Complexes

http://pqs.ebi.ac.uk

PQS Form - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://pqs.ebi.ac.uk/ Go

EMBL-EBI

Macromolecular Structure Database

contact msd

home > searches > pqs-search

PQS Protein Quaternary Structure Query Form at the EBI

about help

Click Titles for Help

- Sample Output
- Output Files
- Complex Searches
- References related to Quaternary Structure
- FTP download index files
- FTP download directory for all files
- NMR representative model search

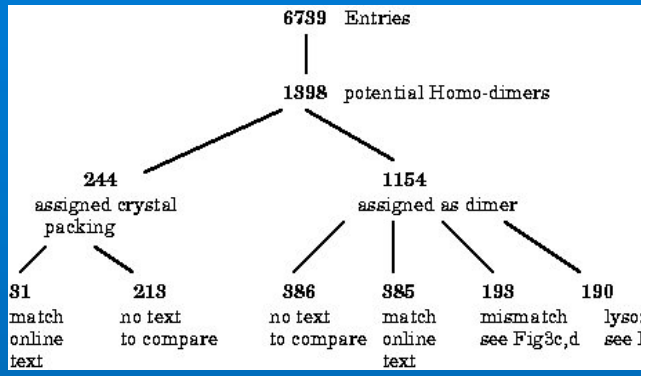
PDBidcode	KeyWords	Authors
Quaternary Type	**Homo OR Hetero**	**Num. Independent Molecules**
No condition	Both Homo and Hetero	No condition
mean Delta ASA per chain	**Delta Solvation Energy**	**Find SaltBridges**
No condition	No condition	No condition
Find DiSulphides	**Interface Buried SideChains**	**Total Num Residues**
No condition	No condition	No condition
SpaceGroup	**ChainFormula**	**Percent ASA**
		No Condition
OutPut ordered by	**ReturnStatistics**	RESET FORM
delta_asa	No Stats for simple query	SEARCH

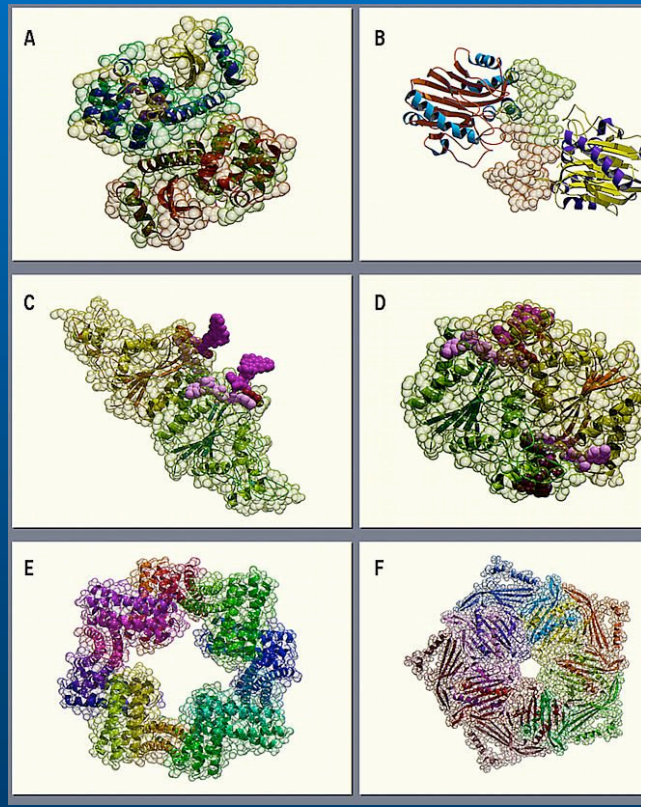
A table is available **from Sue Jones at UCL** to enable you to compare the oligomers generated here with known oligomers analysed by the UCL Protein-Protein Interaction Server.

primary developer: Kim Henrick

last modified: 28/02/03

Done Internet



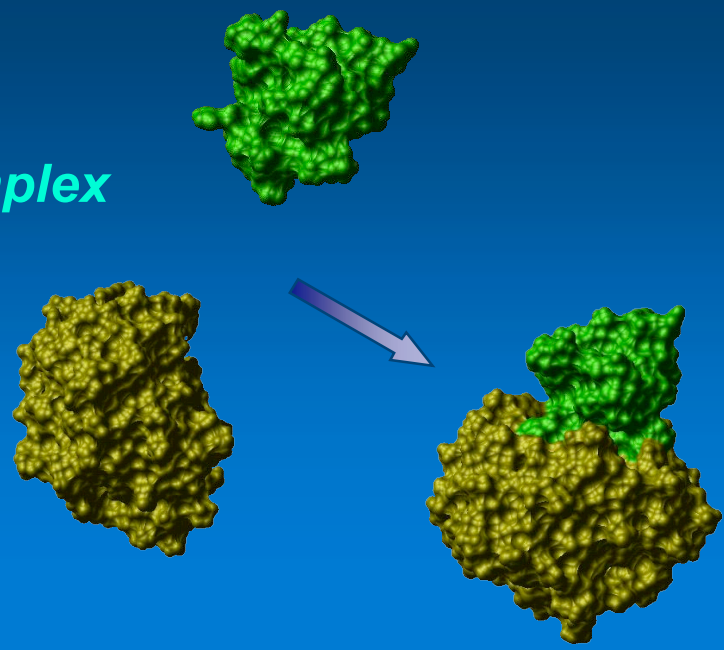


- Introduction
- **Computational protein-protein docking**
- Geometric docking algorithms
- Docking by global energy optimization
- Comparison of docking methods
- Present and future challenges in protein-protein docking

Protein-Protein Docking

Generation of the structure of a protein-protein complex

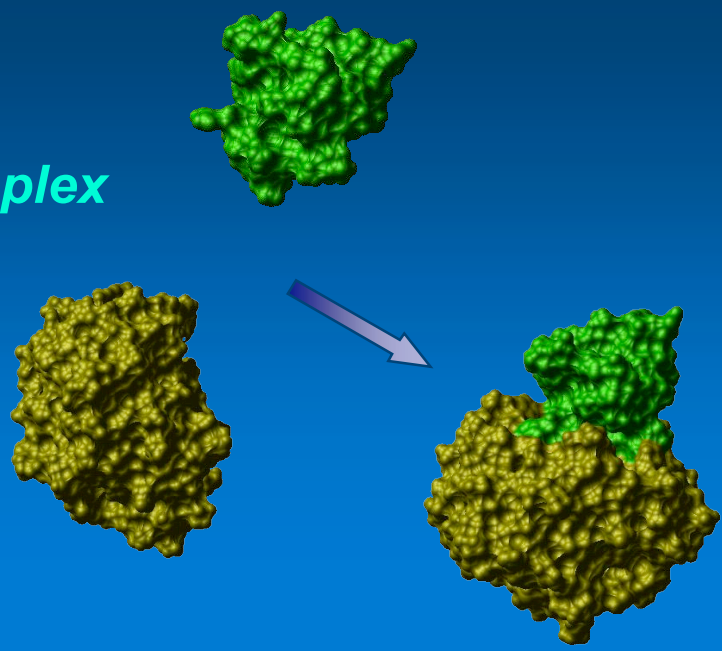
from the individual protein structures



Motivation …

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

Protein-Protein Docking



Generation of the structure of a protein-protein complex

from the individual protein structures

Motivation …

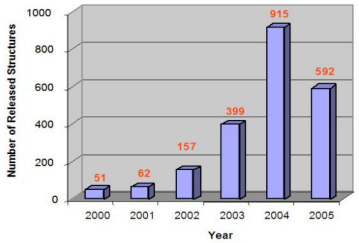
- X-ray, NMR: Determination of complex structures remains difficult

Structural Genomics (2000-2005)

2566 structures

(65% non-redundant)

~10% of 2000-2005 PDB



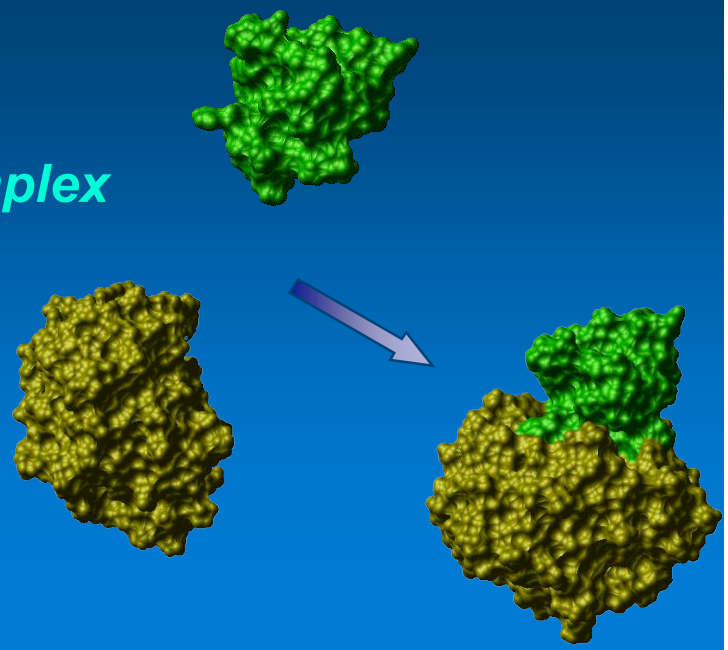
Structural Genomics (2000-2003)

1% protein complexes (!?)

JMB 348, 1235-60 (2005)

Protein-Protein Docking

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



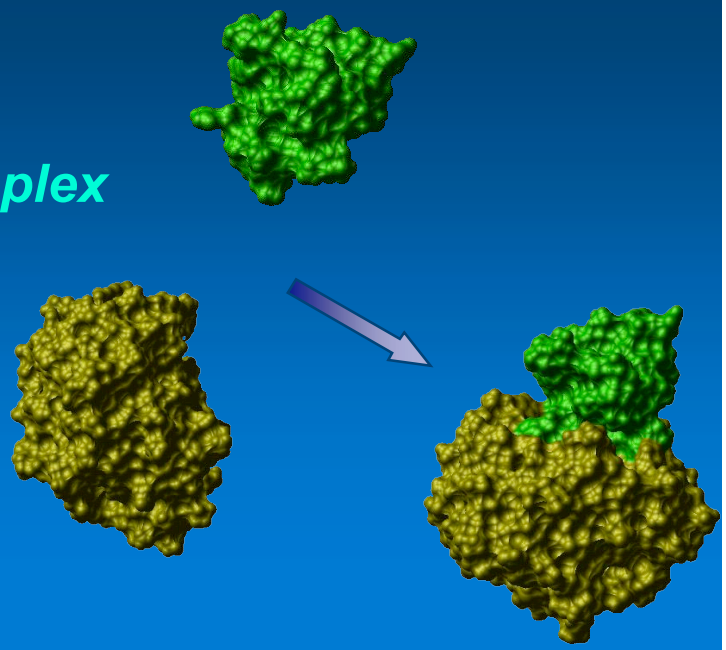
Motivation …

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

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

Protein-Protein Docking

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



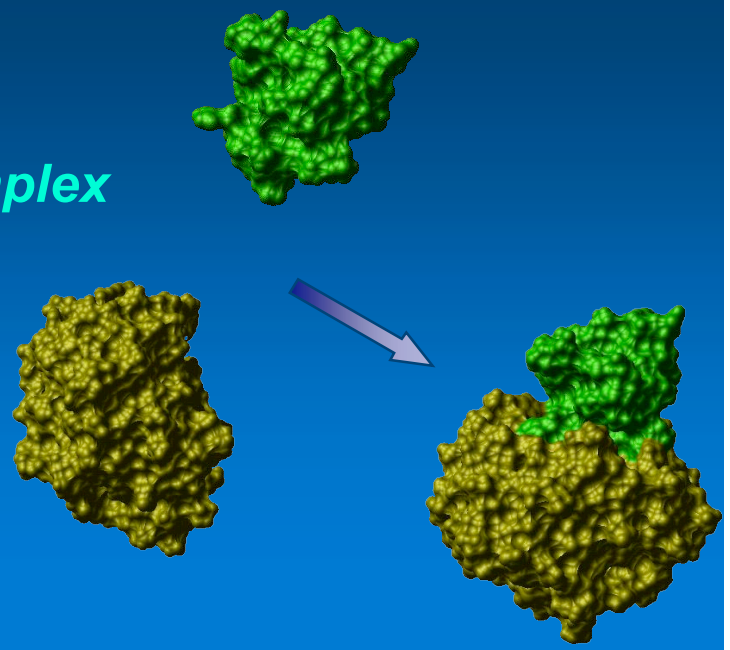
Motivation …

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

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

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

Protein-Protein Docking



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

Motivation …

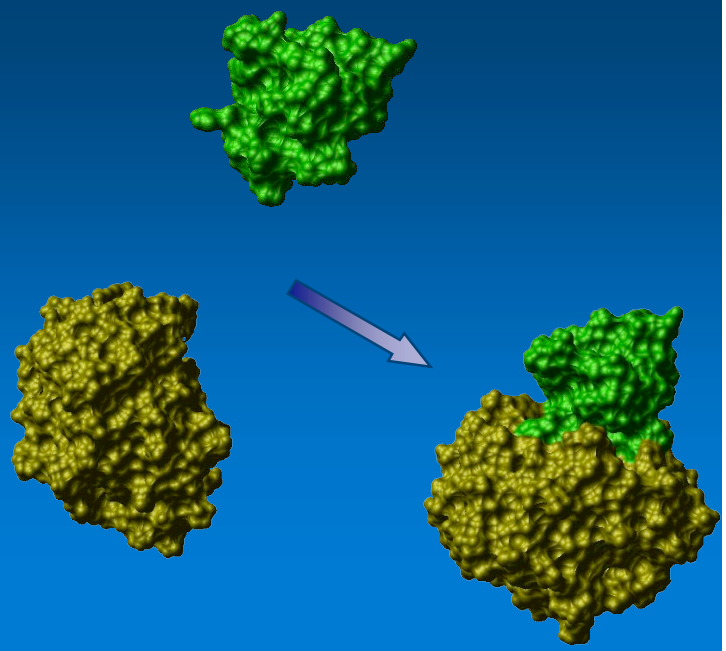
- *X-ray, NMR: Determination of complex structures remains difficult*
- *Low-resolution data on PPI available (cryo-EM, MS…)*
- *Understand energetics and mechanism of protein-protein association*
- *Protein design (diagnostic, environment) and drug discovery*

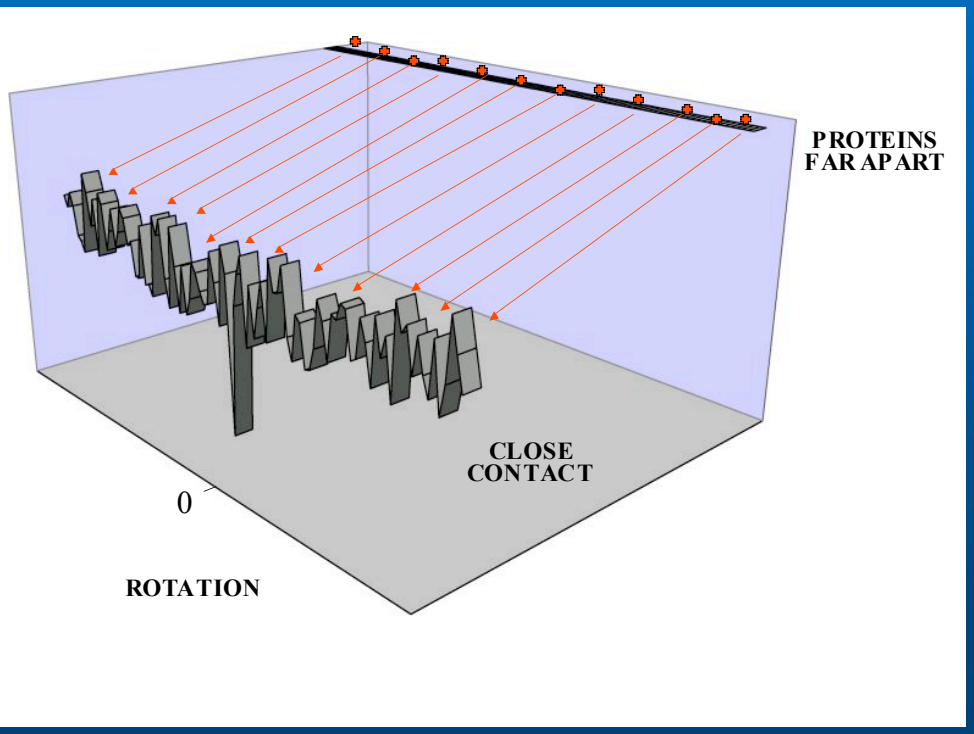
Why Is Docking A Hard Problem?

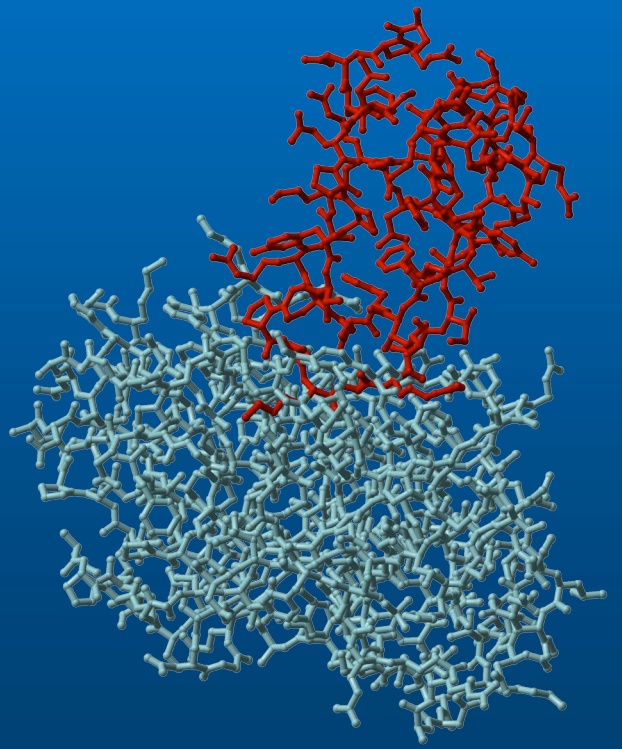
- Large Search Space $\longrightarrow$ "Hard"
- "The Curse of Dimensionality"

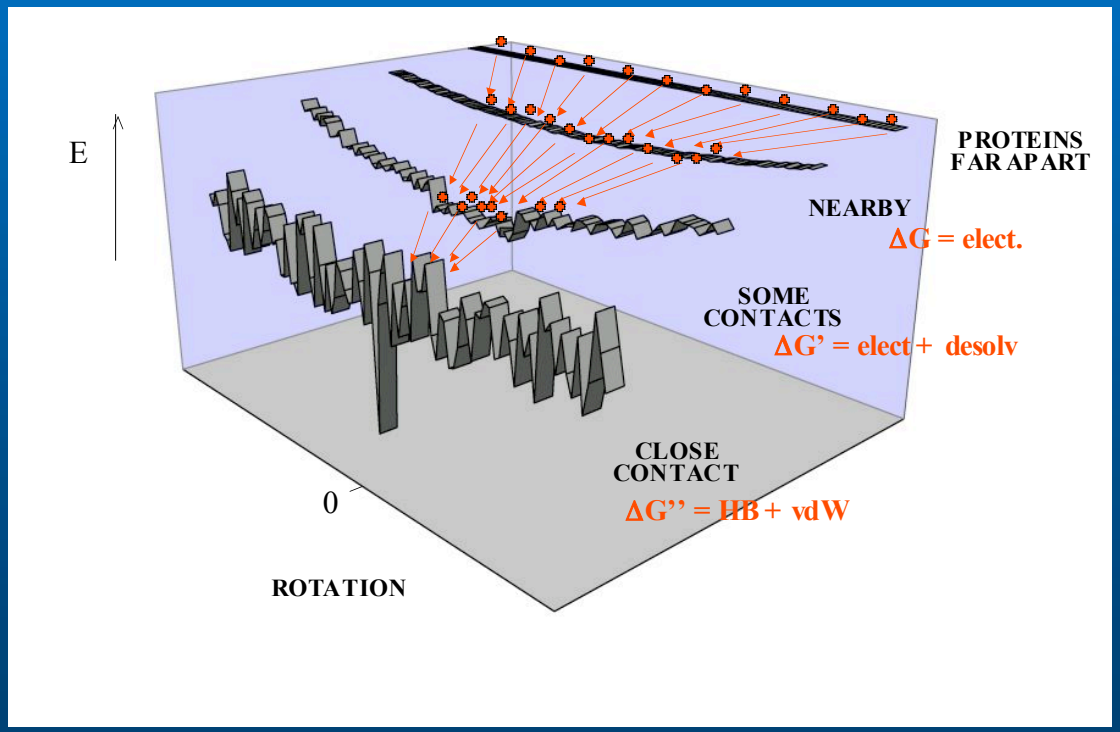
- Sequence Alignment: 1D - Easy!
- Structure Alignment: 3-6D - Hard
- Rigid Body Docking: 6D - Hard!
- Flexible Docking: 3N - "Impossible!"

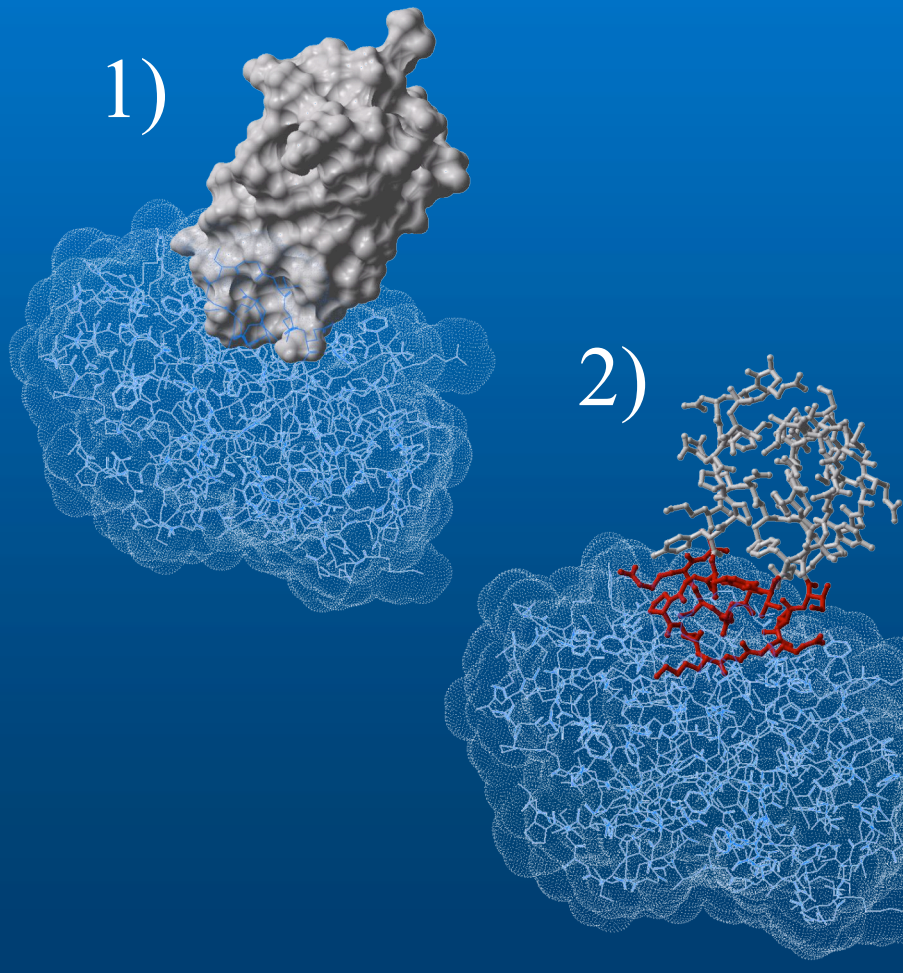
- Sequence Space: Discrete
- Structure Space: Continuous




E
PROTEINS
FAR APART
CLOSE
CONTACT
0
ROTATION

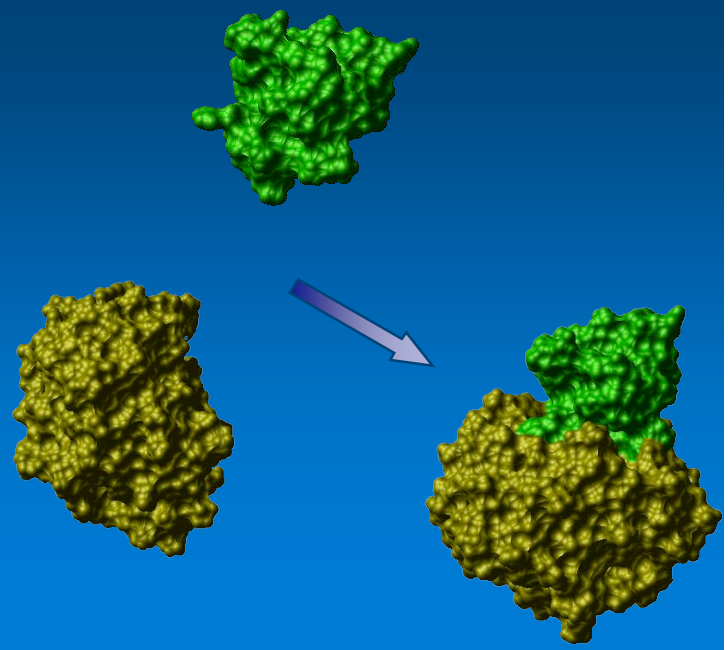



E
PROTEINS
FAR APART
NEARBY
ΔG = elect.
SOME
CONTACTS
ΔG' = elect + desolv
CLOSE
CONTACT
ΔG'' = HB + vdW
0
ROTATION


1)
2)

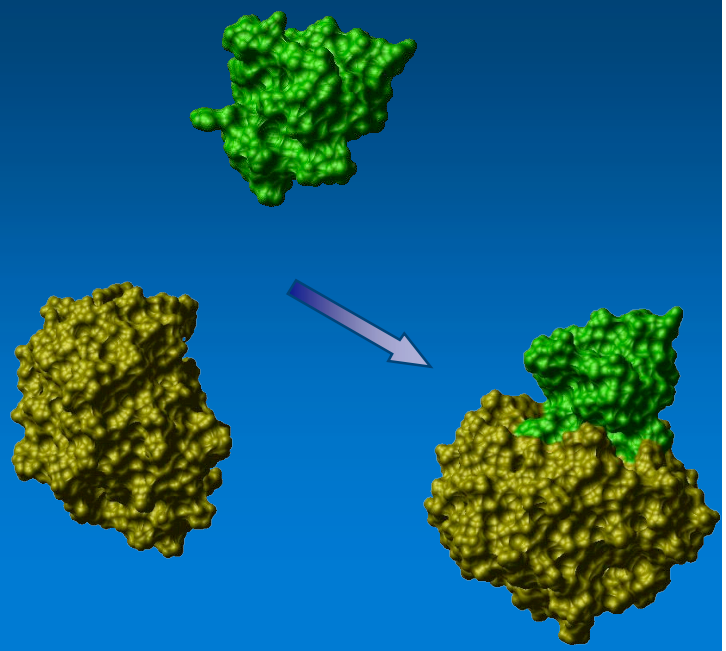
Docking Search Strategies

- **Pseudo Random**
  - Simulated Annealing / Monte Carlo
  - Genetic Algorithms
- **Directed Search**
  - Geometric Hashing
  - Spherical Harmonic Surface Triangles
- **Brute-Force Search**
  - Explicit Grid Correlations
  - Fast Fourier Transform (FFT) Correlations
  - Spherical Polar Fourier Correlations
- **Refinement Phase**
  - Classical or Soft Potentials (+/- Electrostatics)
  - Desolvation, Solvent Dipoles...
  - Visual Inspection!!



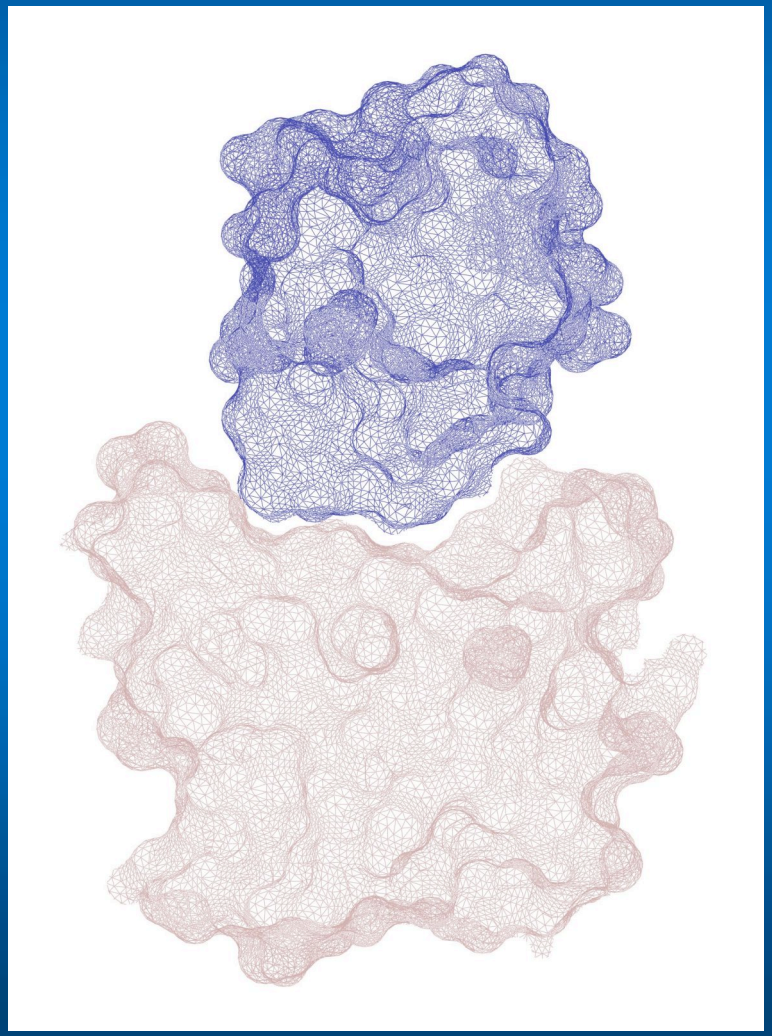
Criteria for Good Docking Orientations

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

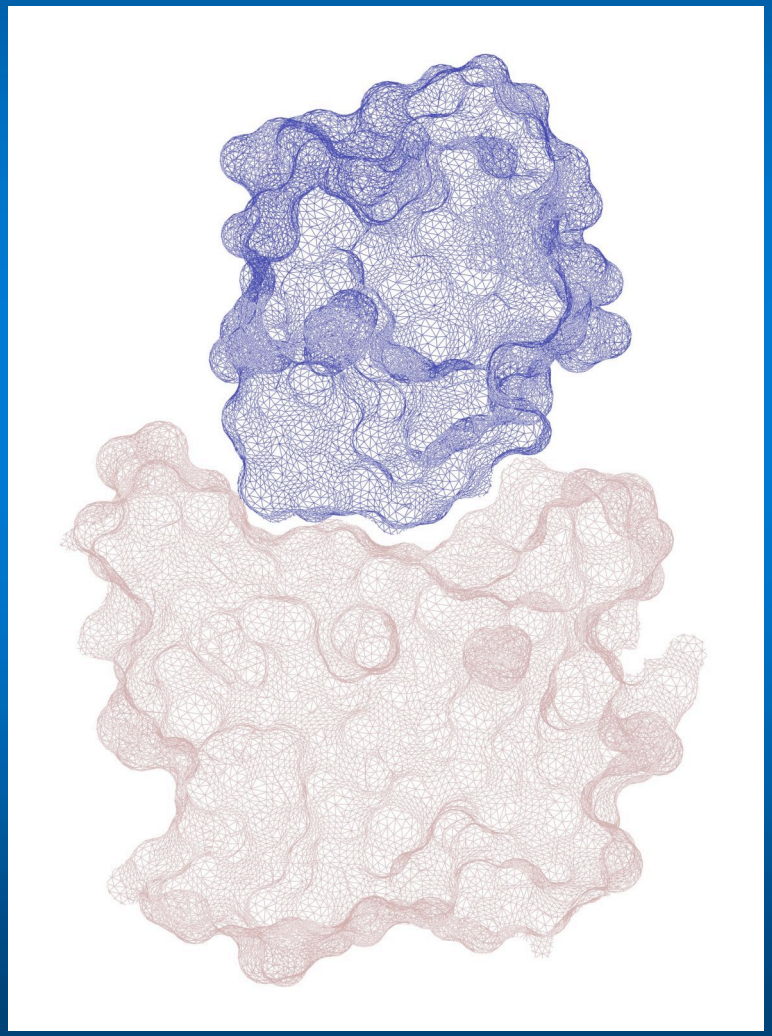


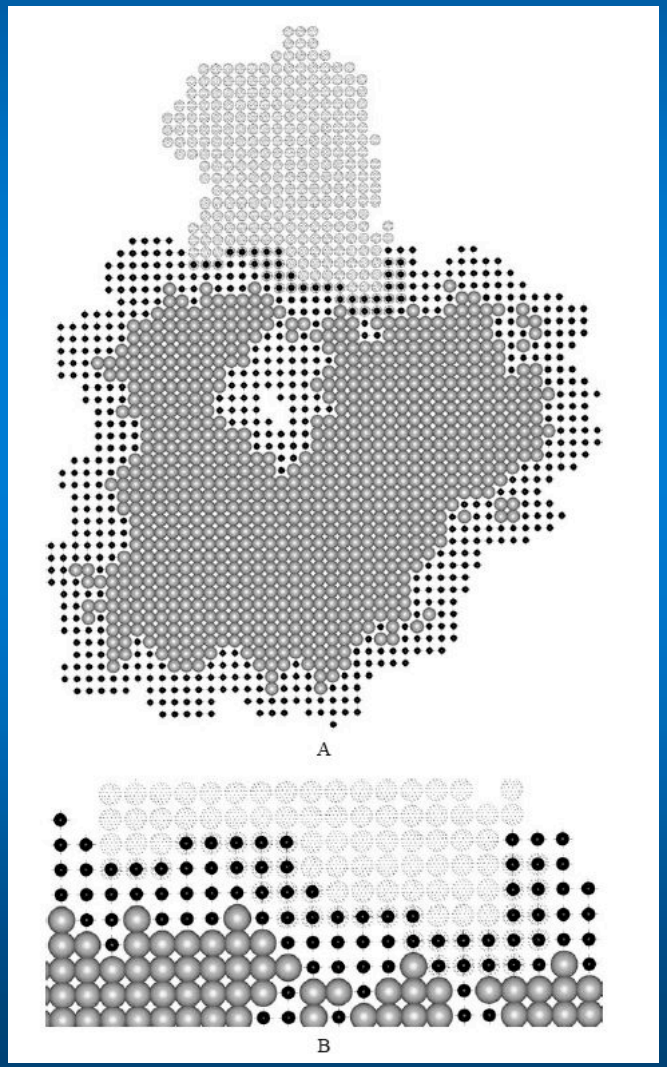
- Introduction
- Computational protein-protein docking
- **Geometric docking algorithms**
- Docking by global energy optimization
- Comparison of docking methods
- Present and future challenges in protein-protein docking

Protein-Protein Docking: Geometry Approach

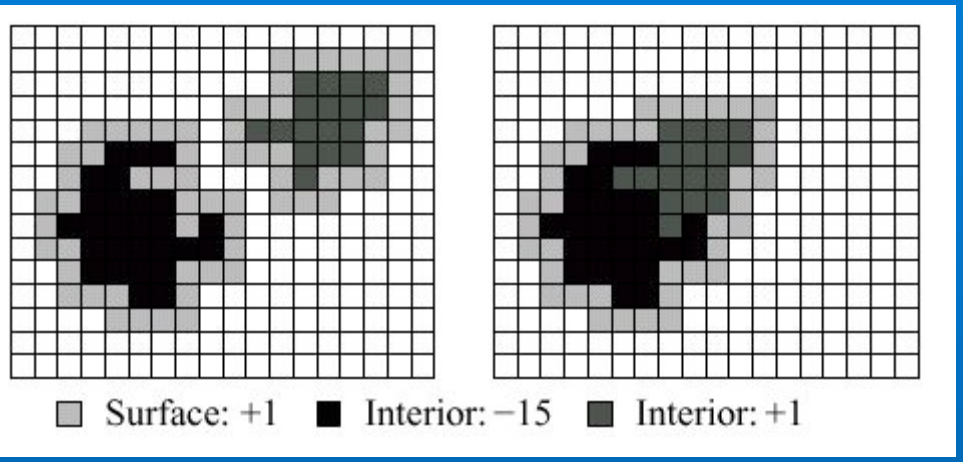


Protein-Protein Docking: Geometry Approach





MolFit



$$\bar{a}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

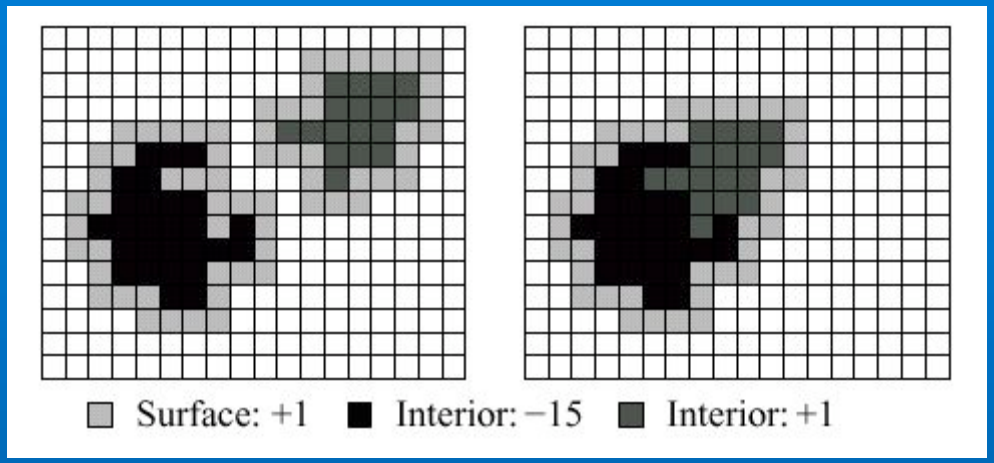
$$\bar{b}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

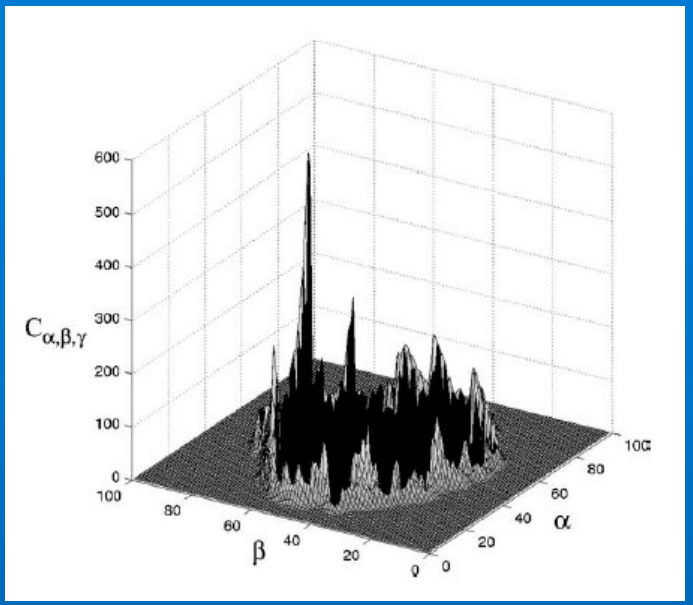
$$l, m, n = (1, \ldots, N)$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

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

MolFit





$$\bar{a}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$\bar{b}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$l, m, n = (1, \ldots, N)$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

a

b

Corr(a,b)

c

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

Fourier Transform

It re-express a function in terms of sinusoidal basis functions

Fourier transform equations:

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2\pi i k x}\, dk$$

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i k x}\, dx.$$

$F(k)$ = FT($f(x)$)

$f(x)$ = IFT($F(k)$)

Correlation function:

$$c(t) = \mathrm{Corr}(g,h) \equiv \int_{-\infty}^{\infty} g(\tau + t) h(\tau)\, d\tau$$

"Correlation Theorem":

$C(f) = G(f)\, H(-f)$

If h(t) is real then $H(-f) = [H(f)]^*$

$C(f) = G(f)\, [H(f)]^*$

$c(t)$ = IFT(FT($g(\tau)$) [FT($h(\tau)$)]*)

FFT (Fast Fourier Transform)

Algorithms for efficient calculation of FT and IFT

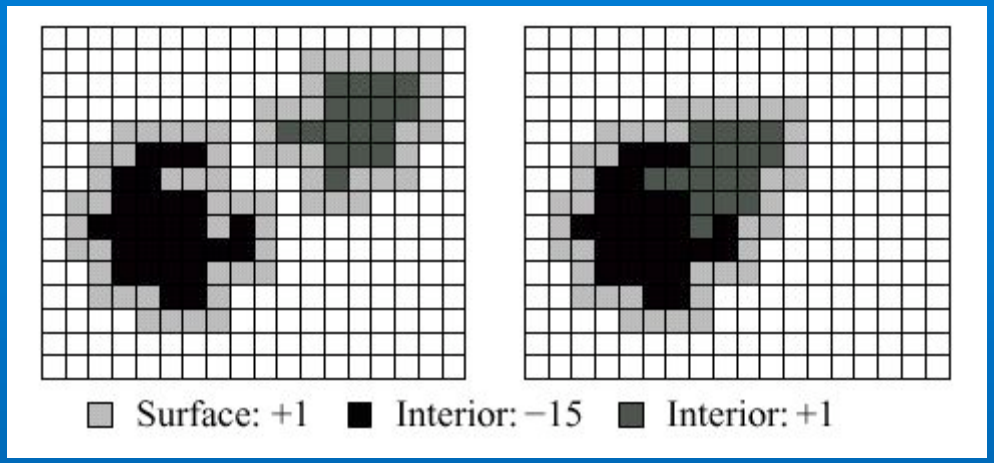
- Most common: Cooley-Tukey FFT (divide and conquer)
- Other: Prime-factor, Bruun's, Rader's, Bluestein's

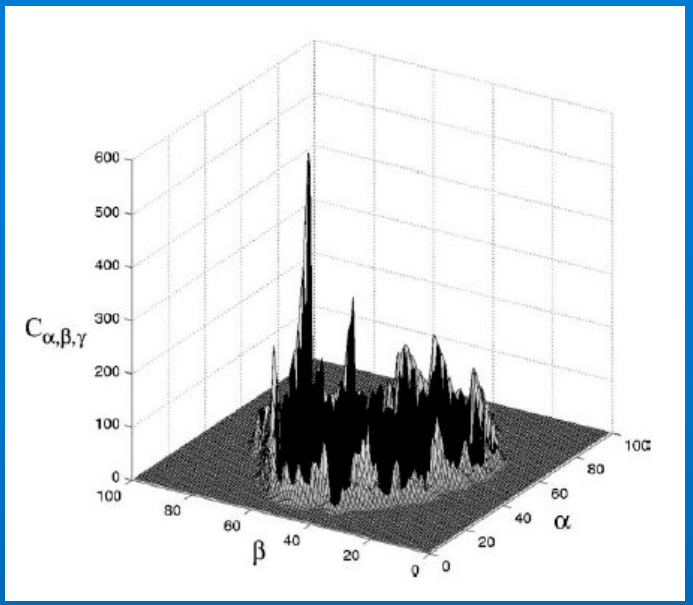
Fourier transform timing: N^2 **(if** $N = 10^6$ **1MHz CPU time ~2 weeks)**

Fast Fourier transform: $N\log_2 N$ **(if** $N = 10^6$ **1MHz CPU time ~ 30 sec)**

http://www.fftw.org/

MolFit





$$\bar{a}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$\bar{b}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$l, m, n = (1, \ldots, N)$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

a

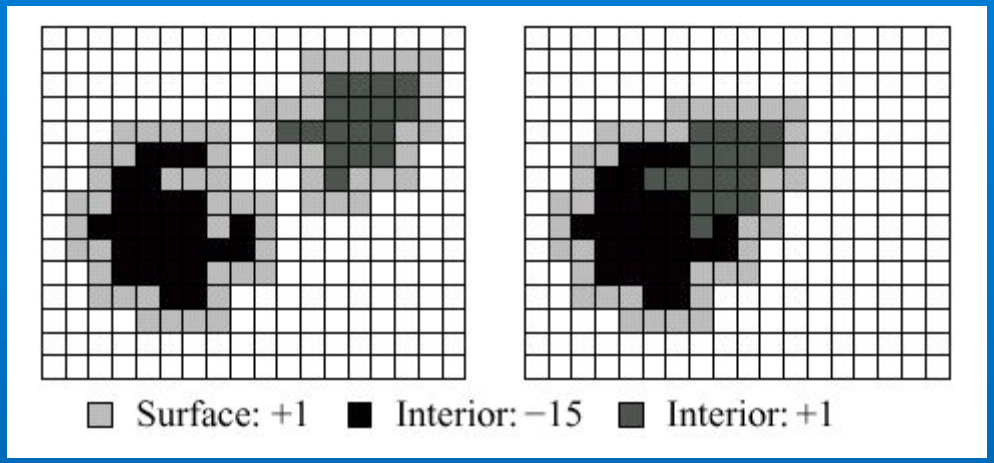
b

Corr(a,b)

c

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

MolFit



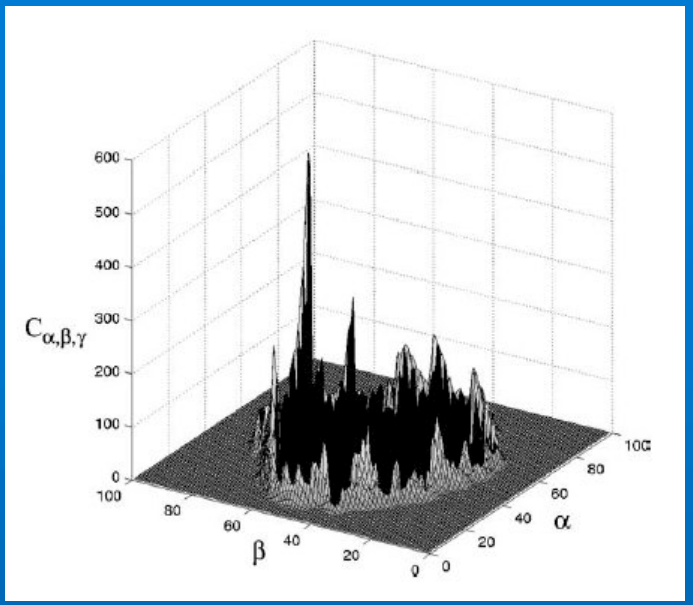
$$\bar{a}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

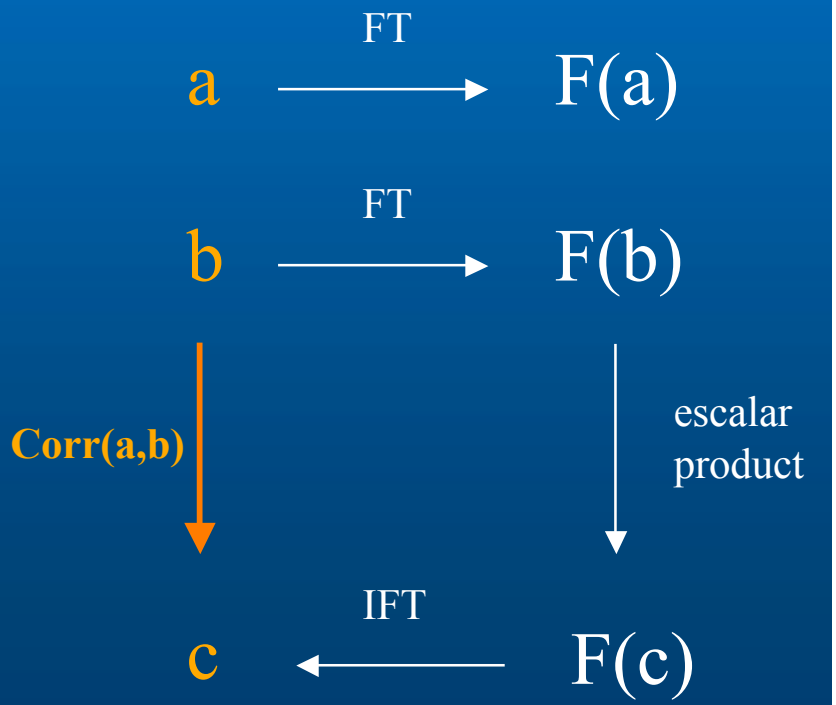
$$\bar{b}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$l, m, n = (1, \ldots, N)$$

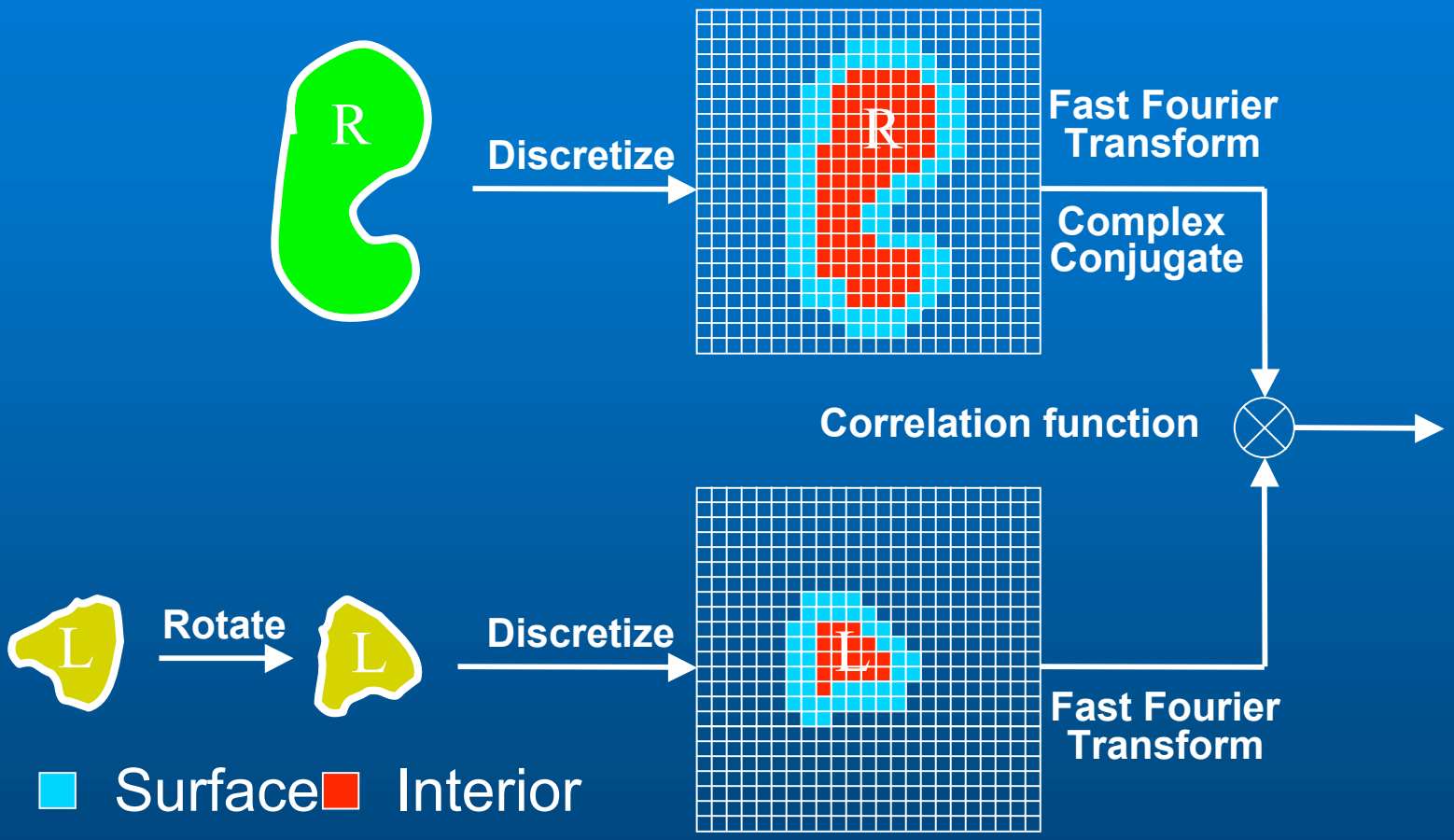
$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

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



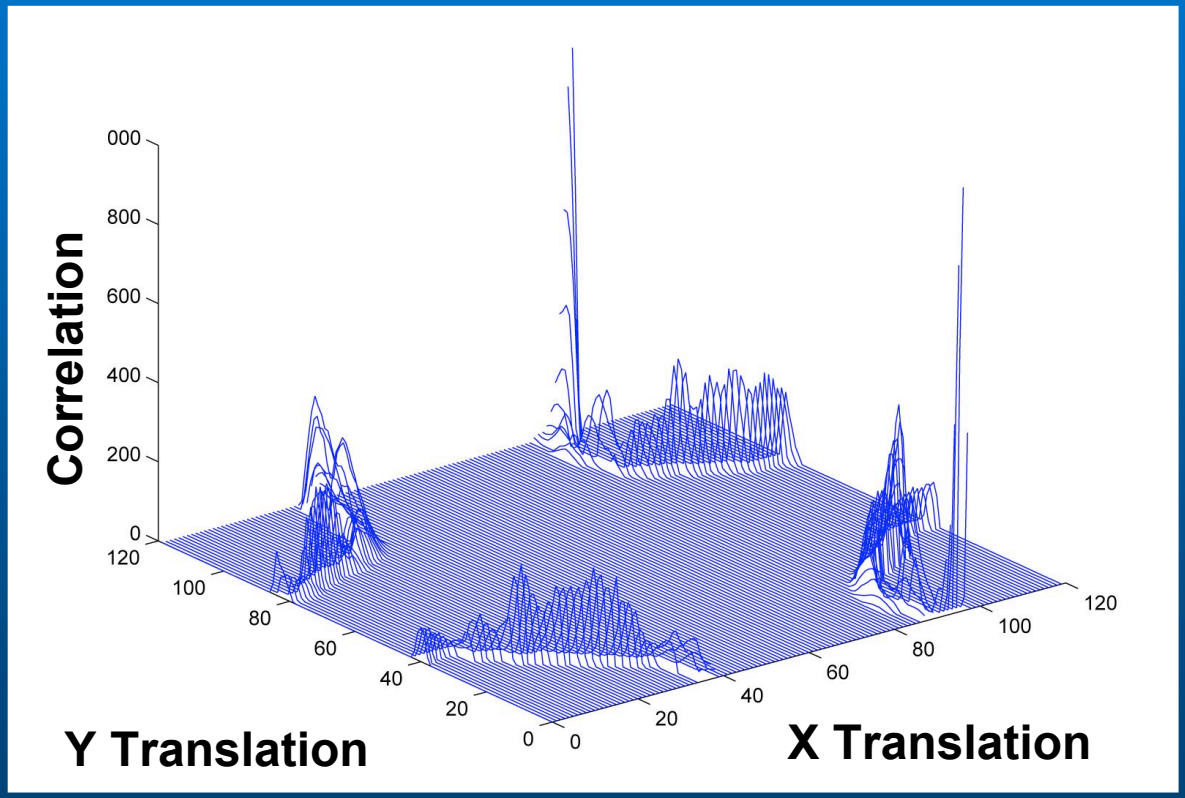


Protein Docking Using FFT

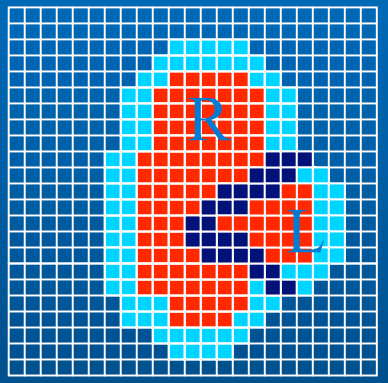


Protein Docking Using FFT

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



IFFT



FTDOCK

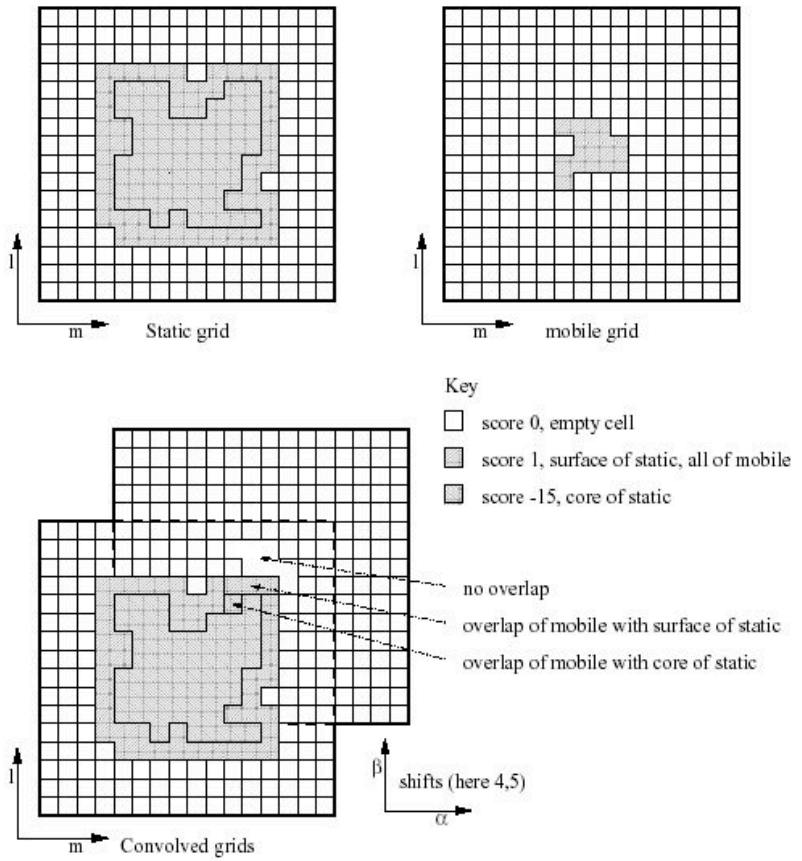
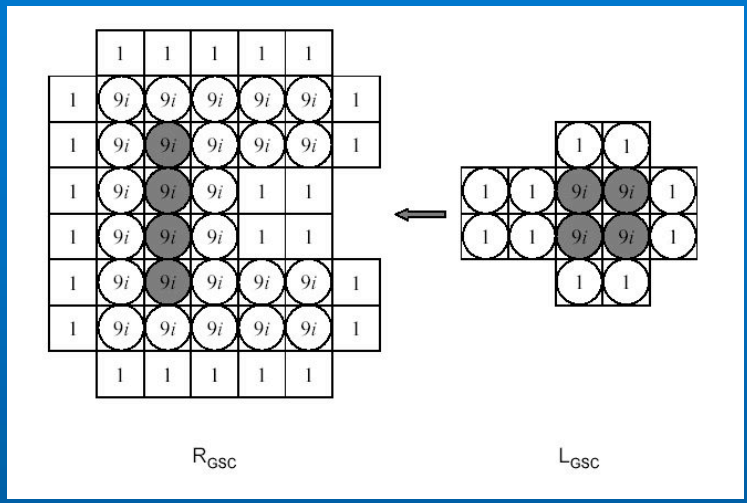
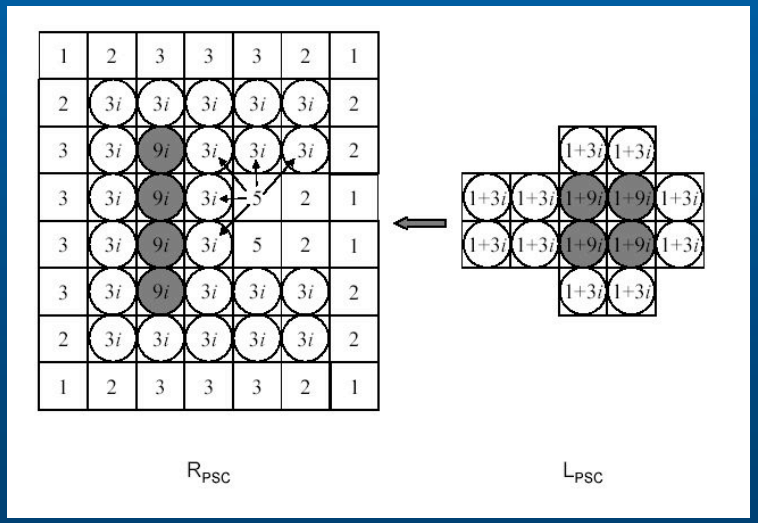


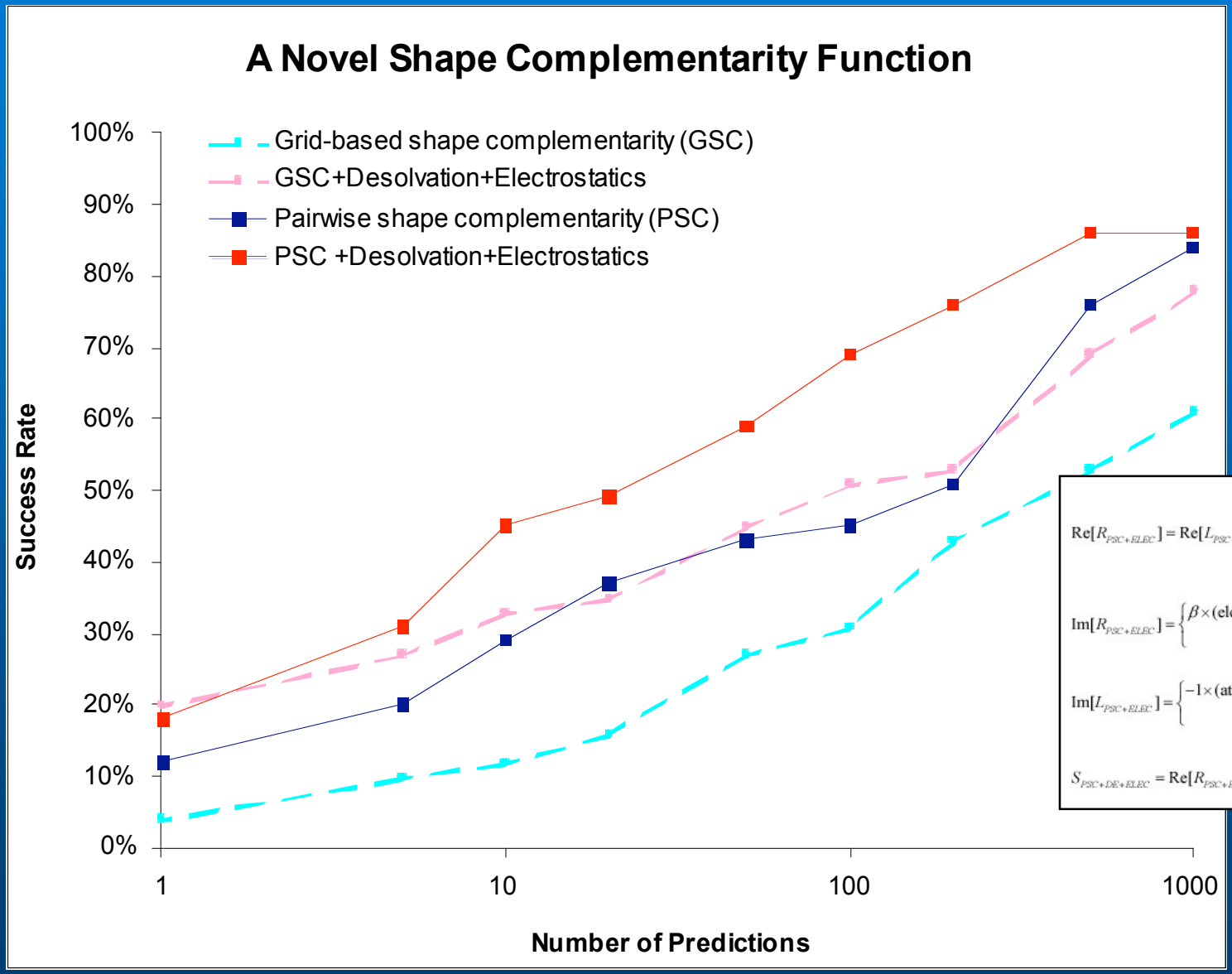
Figure 2: Grid discretisation of molecules and calculation of surface complementarity

ZDOCK





ZDOCK performance



$$\mathrm{Re}[R_{PSC+ELEC}] = \mathrm{Re}[L_{PSC+ELEC}] = \begin{cases} 3.5 & \text{solvent excluding surface layer of the protein} \\ 3.5^2 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

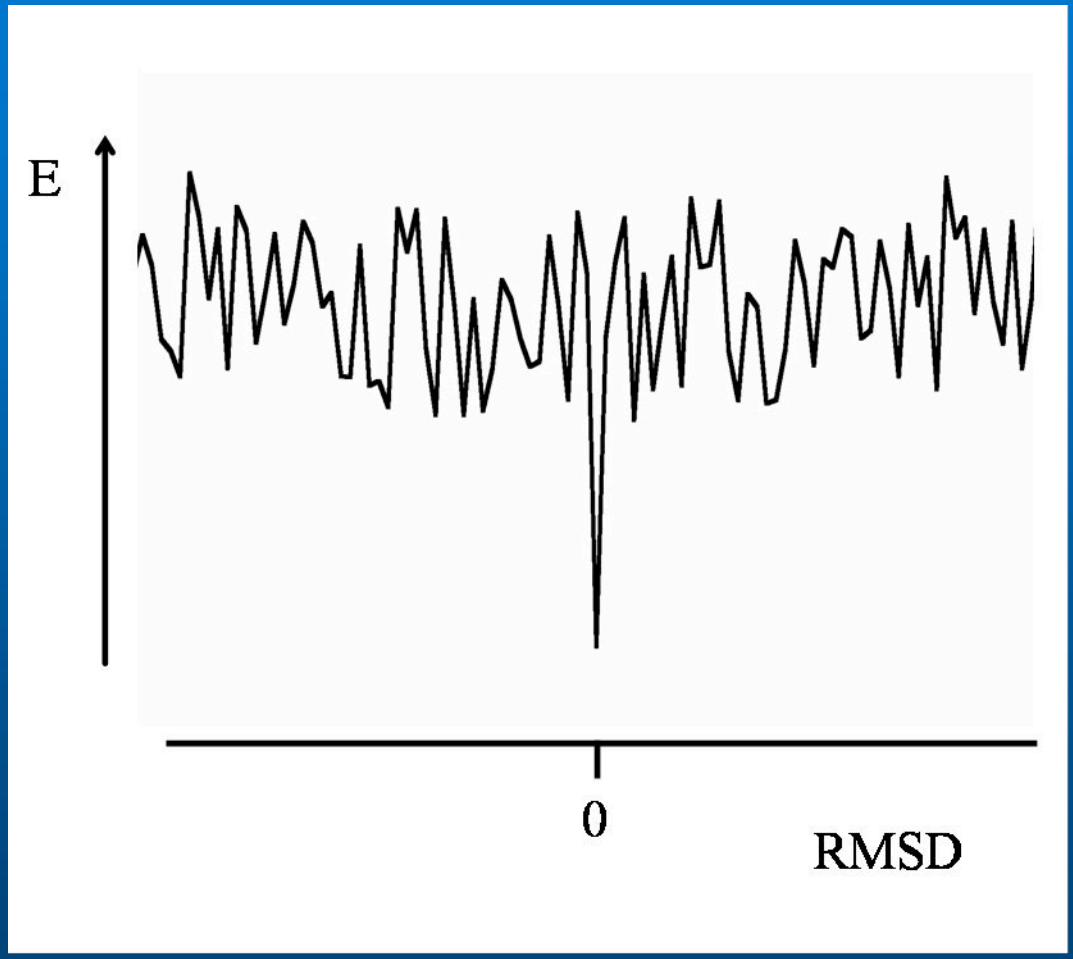
$$\mathrm{Im}[R_{PSC+ELEC}] = \begin{cases} \beta \times (\text{electric potential of all receptor atoms}) & \text{open space} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{Im}[L_{PSC+ELEC}] = \begin{cases} -1 \times (\text{atom charge}) & \text{if this grid point is the nearest grid point of a ligand a} \\ 0 & \text{otherwise} \end{cases}$$

$$S_{PSC+DE+ELEC} = \mathrm{Re}[R_{PSC+ELEC} \bullet L_{PSC+ELEC}] + \frac{1}{2} \times \mathrm{Im}[R_{DE} \bullet L_{DE}]$$

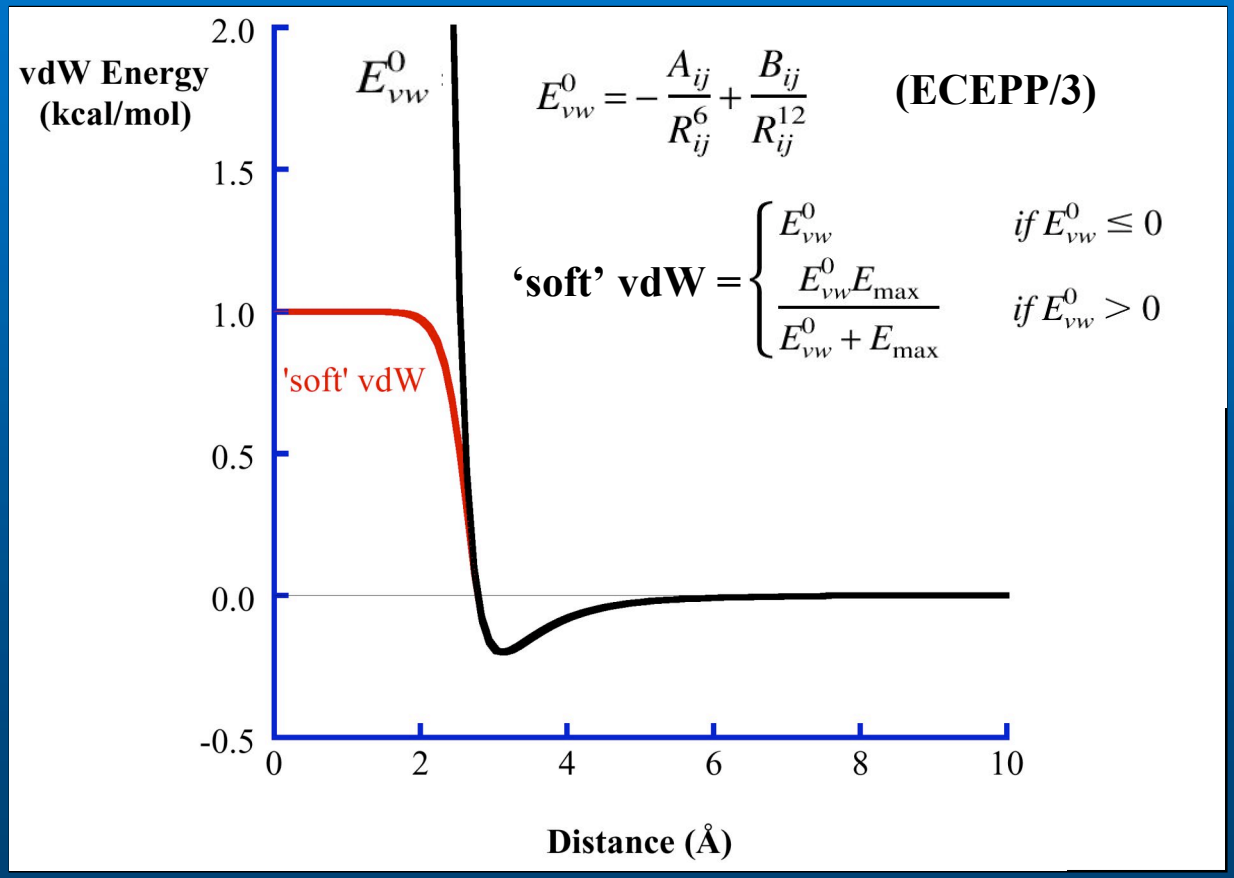
- Introduction
- Computational protein-protein docking
- Geometric docking algorithms
- **Docking by global energy optimization**
- Comparison of docking methods
- Present and future challenges in protein-protein docking

Protein-Protein Docking Energy



Protein-Protein Docking Energy

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



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

Max E_{el} = 20 kcal/mole

Min E_{el} = -20 kcal/mole

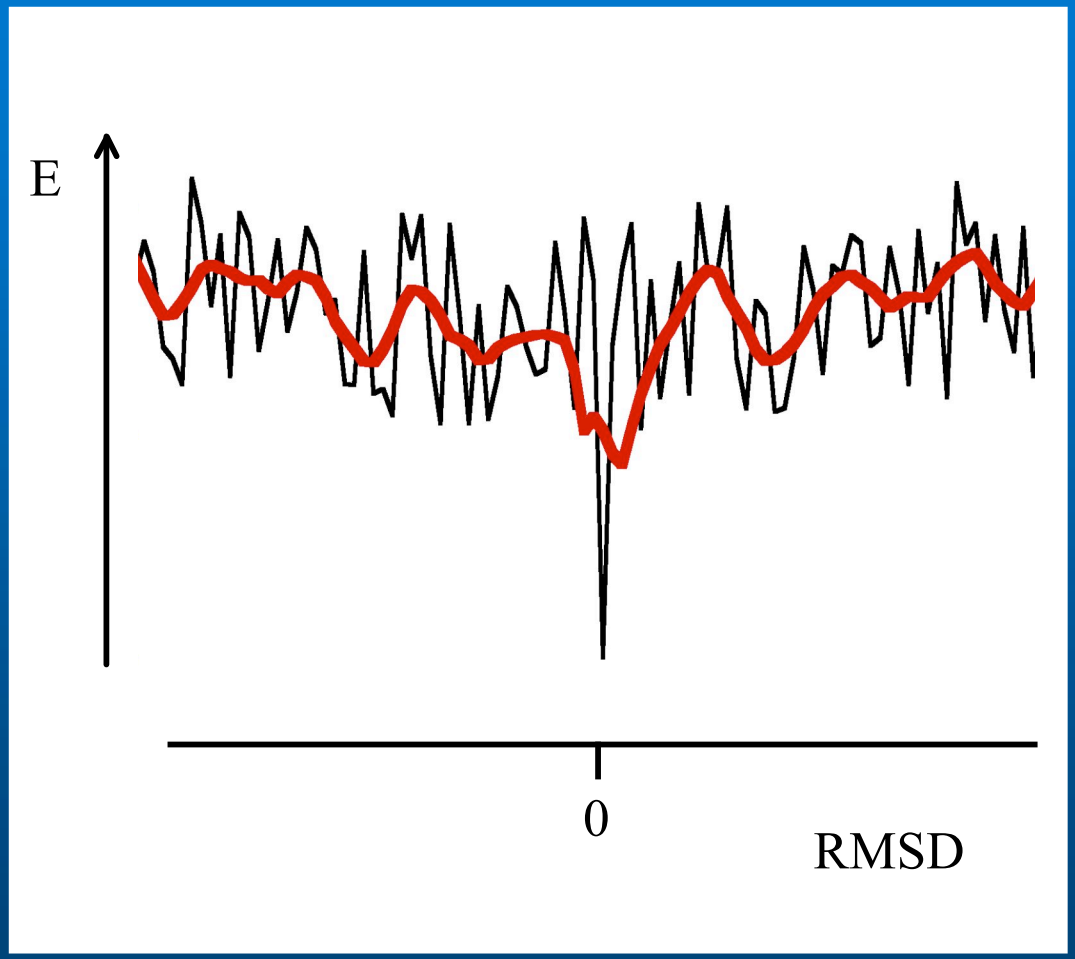
$$E_{hb} = E^0_{hb} e^{-[(r - r_{ep})^2 / d^2_{hb}]}$$

E^0_{hb} = 2.5 kcal/mole

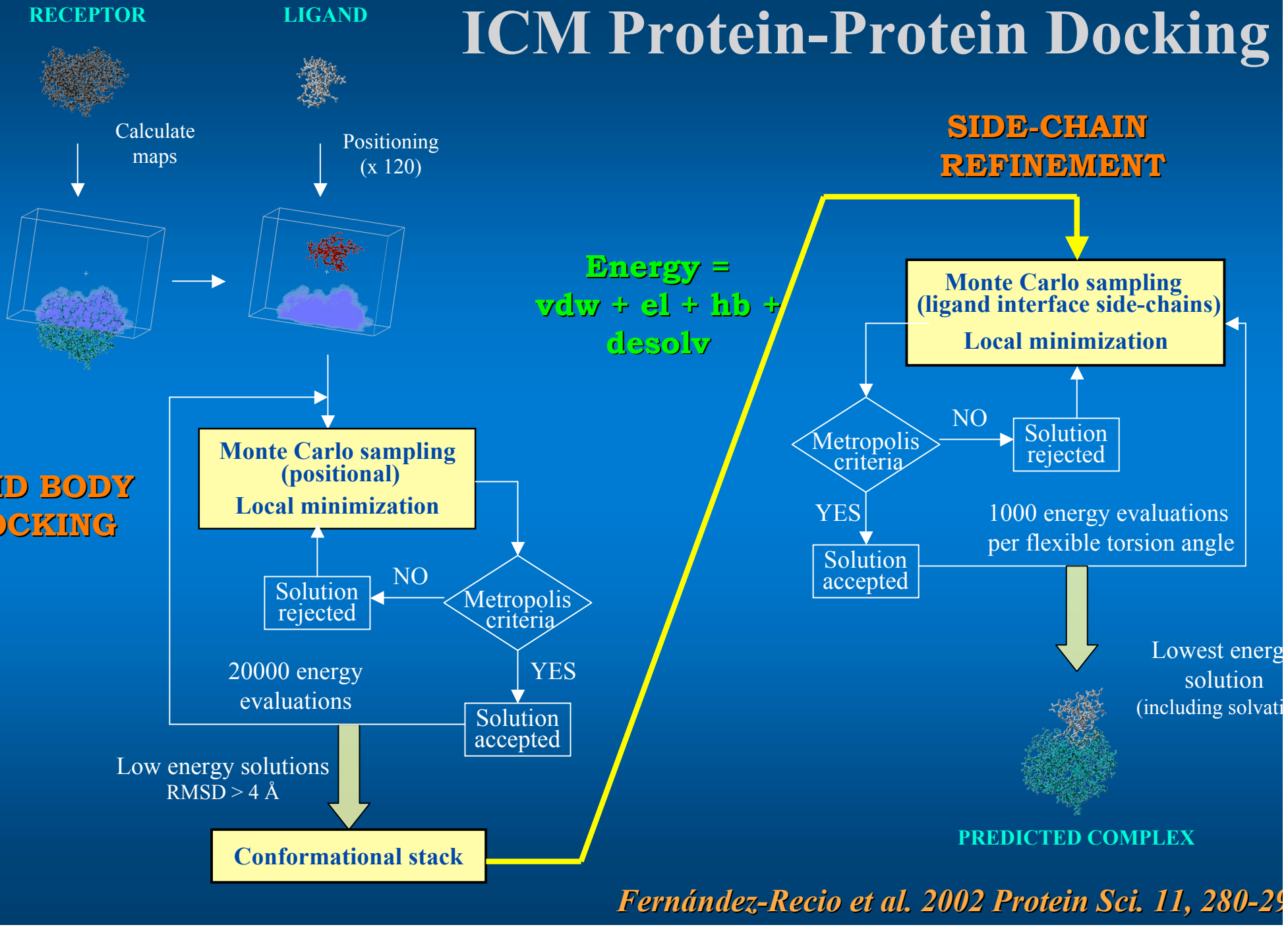
d_{hb} = 1.4 Å

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

Protein-Protein Docking Energy



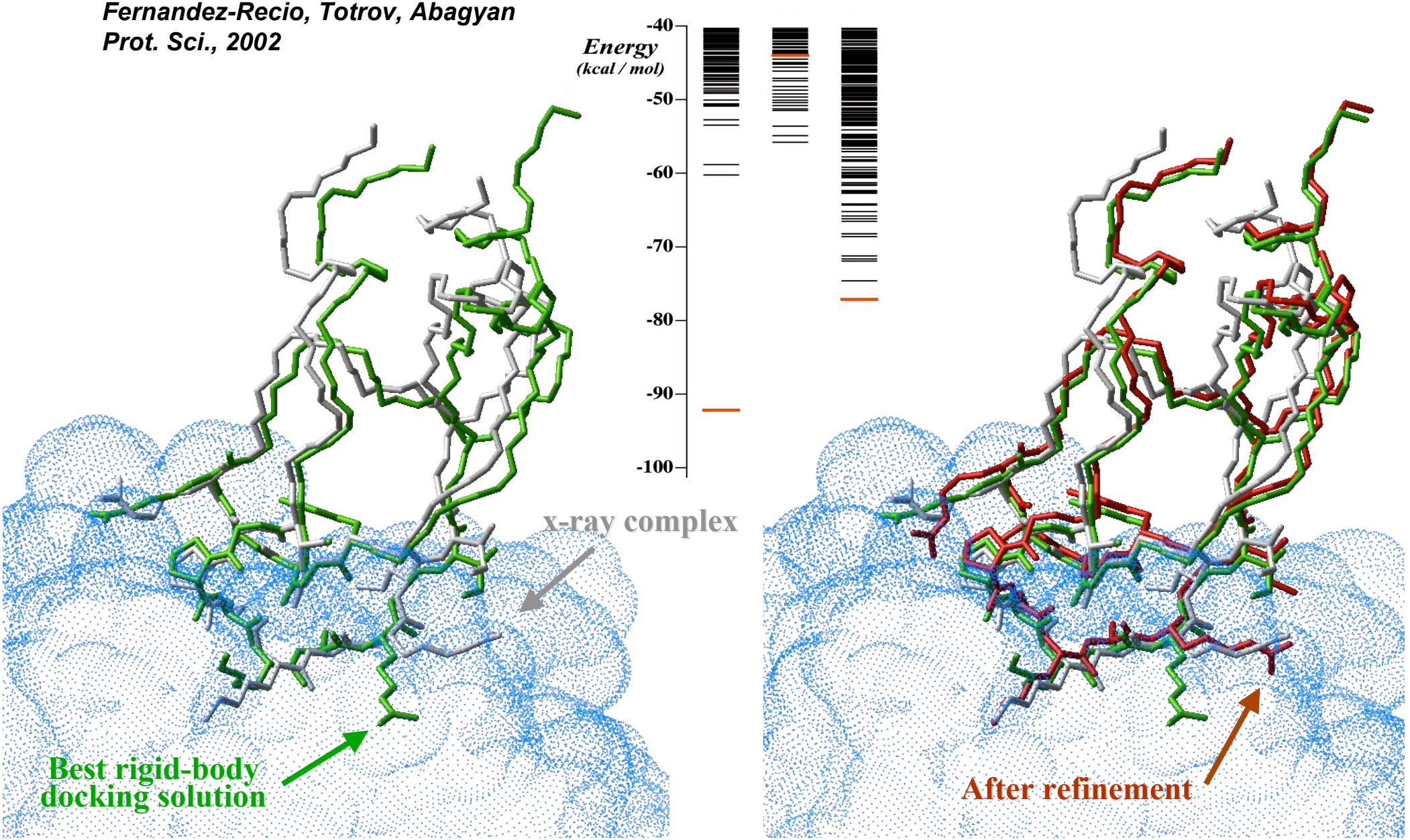
ICM Protein-Protein Docking



Fernández-Recio et al. 2002 Protein Sci. 11, 280-29

Interface Refinement - *unbound trypsin/BPTI*

Fernandez-Recio, Totrov, Abagyan
Prot. Sci., 2002



pyDock: scoring of rigid-body docking orientations by electrostatics + desolvation

$$E = E_{el} + E_{solv}$$

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

Max E_{el} = +1 kcal/mole

Min E_{el} = -1 kcal/mole

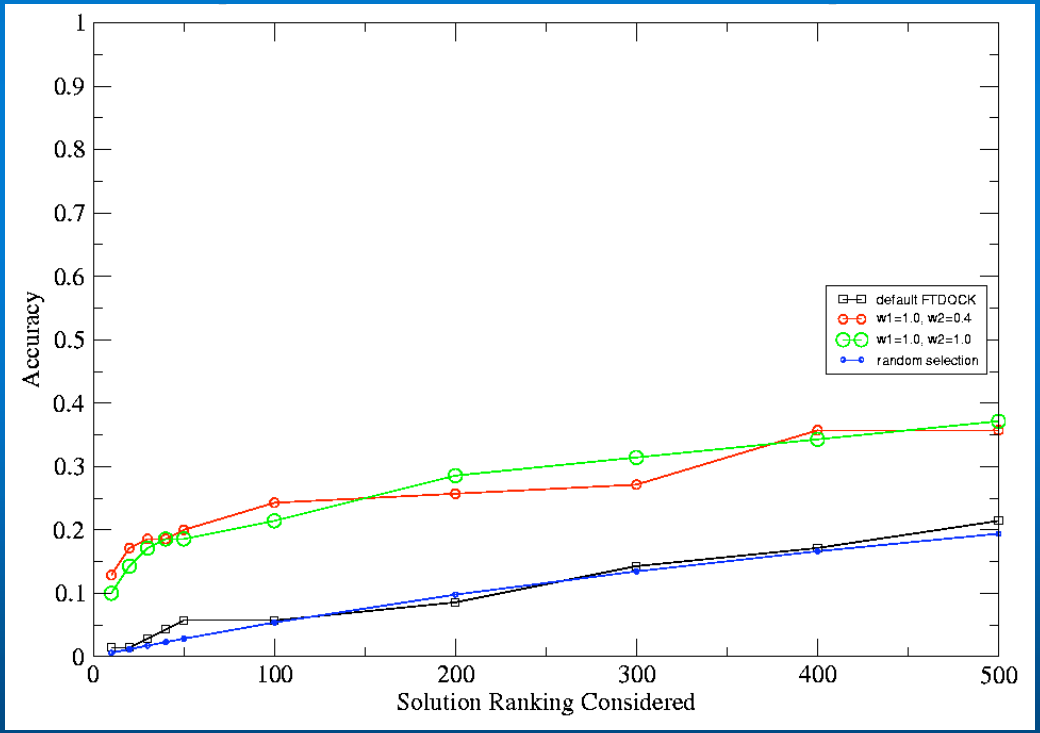
$$\Delta G_{sol} = \sum_{k=1}^{N} \sigma_k A_k$$

water/interface

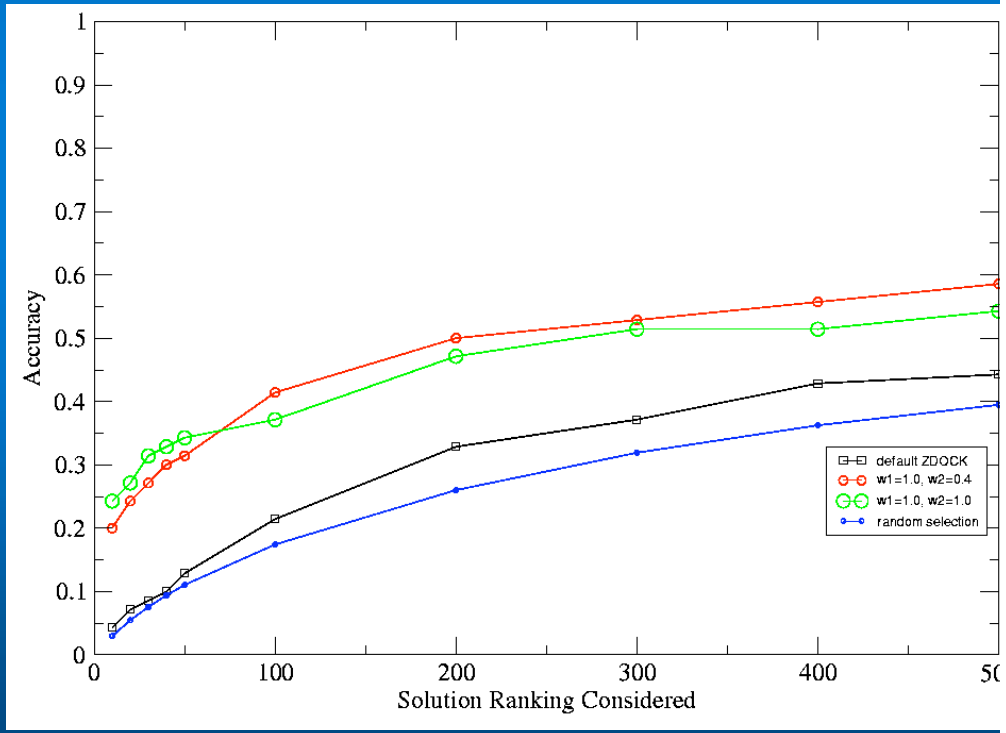
(C_{ali}, C_{aro}, N, N^+_{lys},N^+_{arg}, O_{-oh}, O_{-co}, O^-, S_{-sh}, S_{-s-})

pyDock: scoring of rigid-body docking orientations by electrostatics + desolvation

FTDOCK's docking sets:
(70 unbound cases)



ZDOCK's docking sets:
(70 unbound cases)



- Introduction
- Computational protein-protein docking
- Geometric docking algorithms
- Docking by global energy optimization
- **Comparison of docking methods**
- Present and future challenges in protein-protein docking

Docking software (I)

program	algorithm	laboratory	details
3D-Dock (alternative site)	global:fft ; rescore:residue potentials ; flexible refinement:mean-field sidechain multicopy	Imperial College (Sternberg)	free to academic; mostly source (C) distribution ; refinement sgi/linux executables
HEX	global: fourier correlation of spherical harmonics	Aberdeen University (Richie)	free to academic; sgi/sun/linux executables
DOT	global: fft for shape complementarity and poission-boltzmann electrostatics	UCSD (Ten Eyck)	free to academic; parallelized under MPI; source (C/fortran) and sun/sgi/dec/ibm executables; site also has molec shape programs fade/padre
ICM and input scripts for docking	global: rigid-body pseudo-Brownian MC with grid-based energy funtion ; refinement: Biased MC Minimization using ICM (internal coordinate mechanics)	Scripps (Abagyan)	academic licence for ICM with ICM-Dock and chemistry modules about $2000; sgi, compaq(dec alpha), sun, linux/intel, linux/ppc, winNT executables
GRAMM	global:fft ; clustering and rescoring decoys also available	SUNY/MUSC (Vakser)	free to academic ; sgi/sun/ibm/dec/linux/win32 executables
PPD	global:geometric hashing rescoring:multiple	Columbia (Honig)	free to academic; sgi executables
BIGGER (Chemera) (Download from biotecnol.com)	global:bit mapping; rescoring: multiple filters	Universidade Nova de Lisboa (Palma/Moura)	free to academic; win32 executables
Vajda/Camacho refinement	refinement protocol only: constrained minimization; desolvation calculations ; modified CharMM forcefield	University of Boston(MERL)	free; charmm input scripts, source code (f77)

Docking software (II)

program	algorithm	laboratory	details
*DOCK 4.0	grid based energy function (LJ+ES) flexible docking - random search plus incremental construction	UCSF (Kuntz group)	free to academic; sgi executables
*Autodock	grid based empirical potential flexible docking via MC search and incremental construction	Scripps inst (Olson).	free to academic; source code and executables for sgi/dec/sun
*FlexX	fragment assembly energy function:(Boehm potential)	GMD-SCAI (Lengauer), BioSolveIT GmbH.	licence required; commercial licence from Tripos.com; academic licence from BioSolveIT GmbH. Linux/SGI/Sun/HP Risc/HP Itanium executables.
DARWIN	GA	University of Pennsylvania	contact Roger Burnett, burnett@wistar.upenn.edu
ZDOCK/RDOCK	fft for global search (ZDOCK), and reranking/refinement (RDOCK). Decoys also available.	University of Boston (Weng)	free to academic; IBM/SGI/Linux executables; source code available for atomic contact energy (ACE) part
RosettaDock (see also Rosetta and Robetta)	fragment assembly; sidechain and rigid-body refinement with statistical h-bonding potential; computational ala scanning with Robetta	Washington Univ (Baker)/Johns Hopkins Univ (Gray)	use with academic licence agreement
Haddock	Multiple-stage real-space docking with NMR/mutagenesis restraint information; later stages fully flexible	University of Utrecht (Bonvin)	free to academic; python/CNS and other scripts

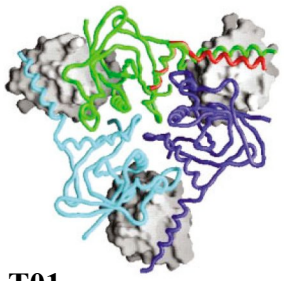
see also servers from Camacho, Wolfson and Zhou groups

Docking Servers

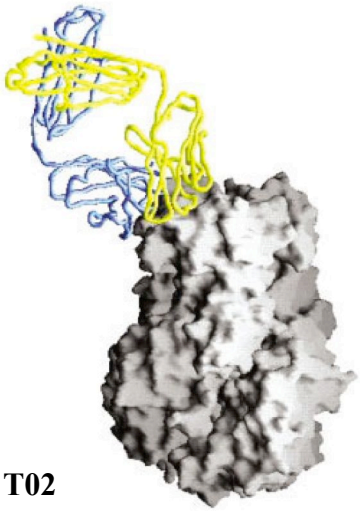
Servers (True Docking Servers)

ClusPro Server (beta-test)	Boston University, C. Camacho/S. Vajda et al.	Carries out protein-protein docking from molecules that the user may upload, using either ZDOCK or DOT. The user may then re-rank using using filtering, energy scoring and spatial and angular clustering.	For licensing contact C. Camacho	Bioinfor-matics 20 (1): 45-50 Jan 2004
PatchDock Server (beta-test)	Tel Aviv University, H. Wolfson et al	Carries out protein-protein docking from molecules that the user may upload, using geometric hashing methods and clustering.	no	Proteins. 2003 Jul 1; 52(1): 107-12 & notes from web site
Dcomplex Server	Buffalo University, Y. Zhou, C. Zhang et al	Scores complexes from pdbs that the user may upload, using statistically derived DFIRE potential	no	Protein Science (2004), 13:400-411 etc

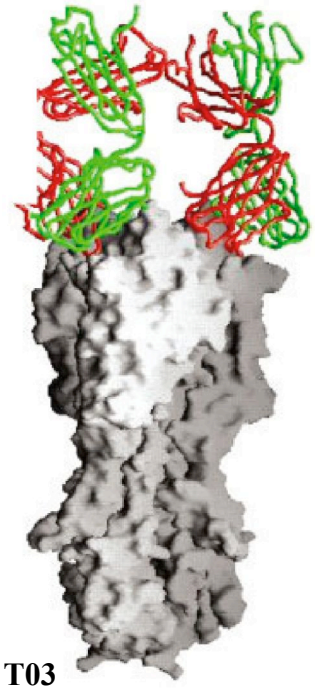
DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions



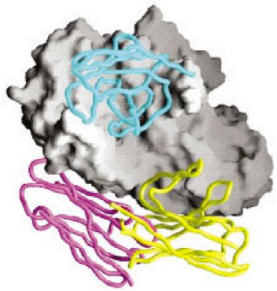
T01
Hpr (*unbound*)
HPr kinase (*unbound*)



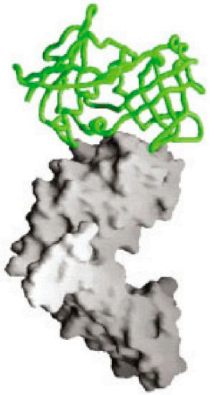
T02
VP6 (*unbound*)
Fab (*bound*)



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



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



T07
TCRβ (*unbound*)
speA (*unbound*)

http://capri.ebi.ac.uk

1st CAPRI – Sep02 La Londe (Franc
Special issue, in:
PROTEINS: Structure, Function, and
Genetics 52 (July 2003)

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

3rd CAPRI – Apr07 Toronto (Canad
Special issue, in progress

DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

CAPRI Participants & Algorithms

Predictor	Affiliation	Software	Algorithm
Abagyan	Scripps	ICM	Force Field
Camacho/Vajda	Boston	CHARMM	Force Field Refinement
Gardiner	Sheffield	GAPDOCK	Shape + Area GA
Sternberg/Smith	Imperial	FTDOCK	FFT
Bates/Fitzjohn	ICRF	Guided Docking	Force Field
Ten Eyck/Mitchell	SDSC	DOT	FFT
Vakser/Tovchigrechko	SUNY/MUSC	GRAMM	FFT
Olson	Scripps	Harmony	Spherical Harmonics ?
Weng/Chen	Boston	ZDOCK	FFT
Eisenstein	Weizmann	MolFit	FFT
Wolfson/Nussinov	?	BUDDA/PPD	Geometric Hashing
Iwadate	Kitasato	TSCF	Force Field + Solvent
Ritchie/Mustard	Aberdeen	Hex	Spherical Polar Fourier
Palma	Lisbon	BIGGER	Geometric + Electrostatic
Gray/Baker	Washington	?	Monte Carlo + Flexibility

Several other participants (not shown) attempted a small no. of targets

1st CAPRI - Predictions

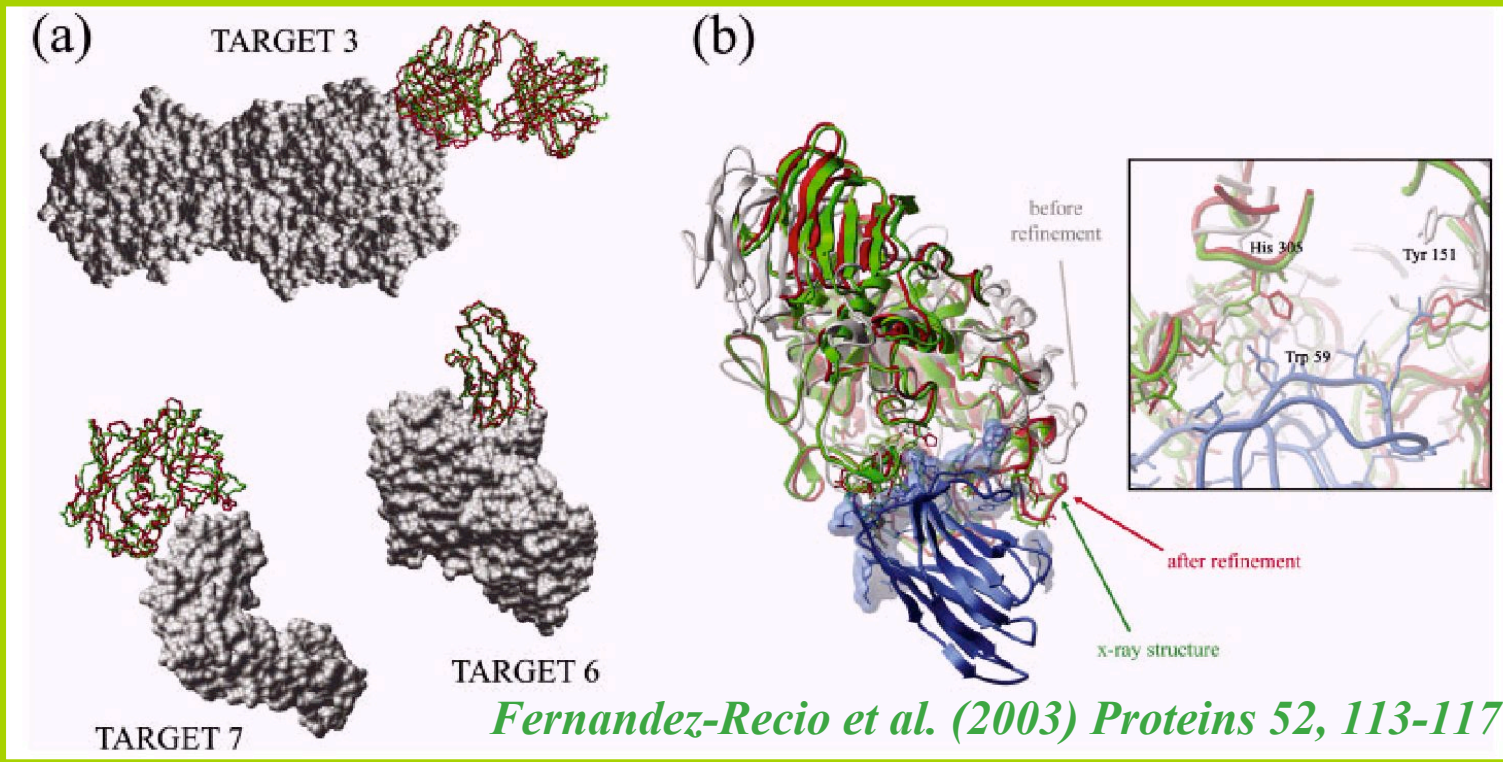
Predictor group	T1	T02	T03	T04	T05	T06	T07	Predictor summary
Scripps US (*Abagyan*)	0	0	**	0	0	***	**	3/2**/1***
Boston U. US (*Camacho*)	*	0	0	0	0	***	***	3/2***
Weizmann Inst. IL (*Eisenstein*)	*	*	0	0	0	0	***	3/1***
Imperial Coll. UK (*Sternberg*)	0	*	0	0	0	***	*	3/1***
UCSD, US (*Ten-Eyck*)	*	*	0	0	0	**	0	3/1**

…

6 groups:
2 acceptable models

3 groups:
1 acceptable models

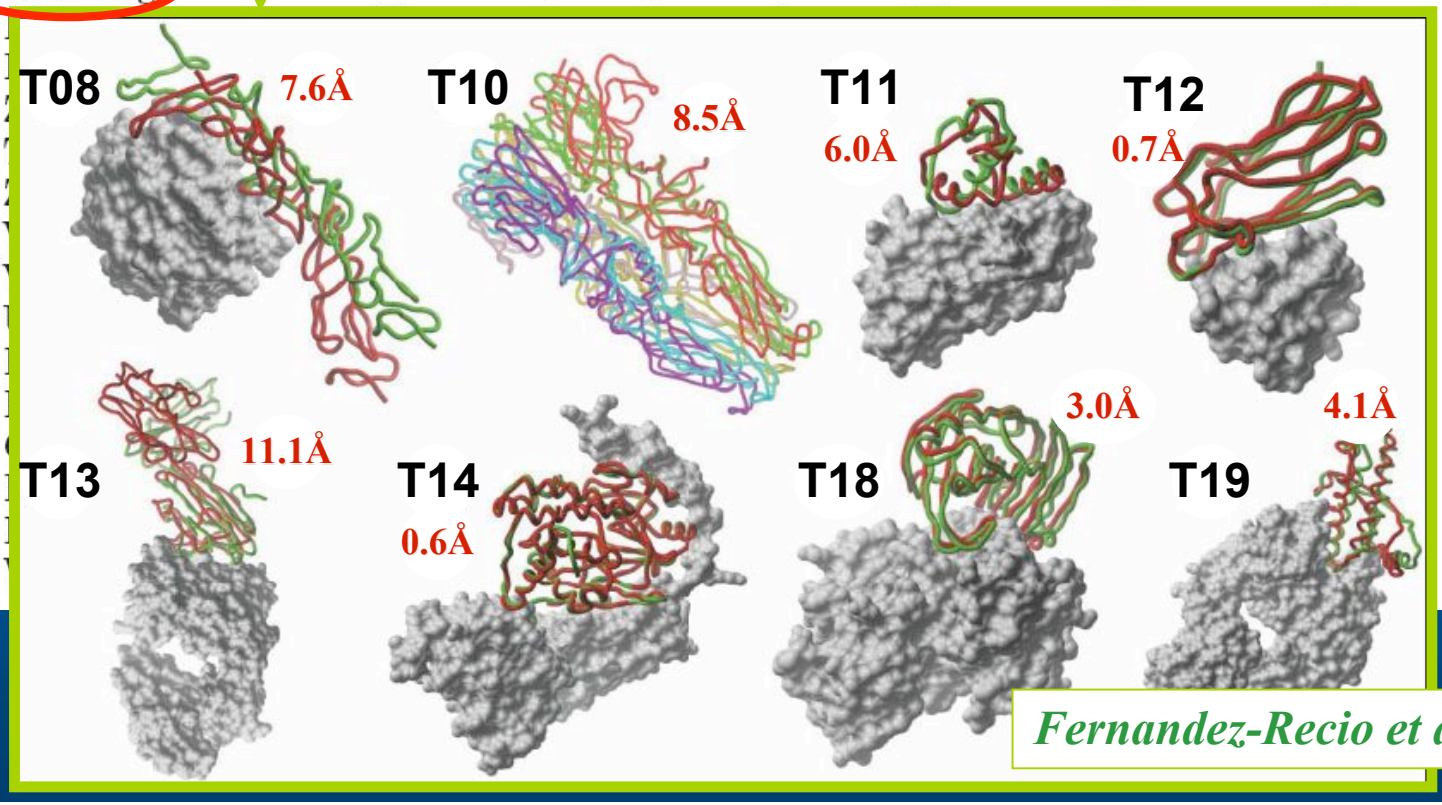
5 groups:
no acceptable models



Fernandez-Recio et al. (2003) Proteins 52, 113-117

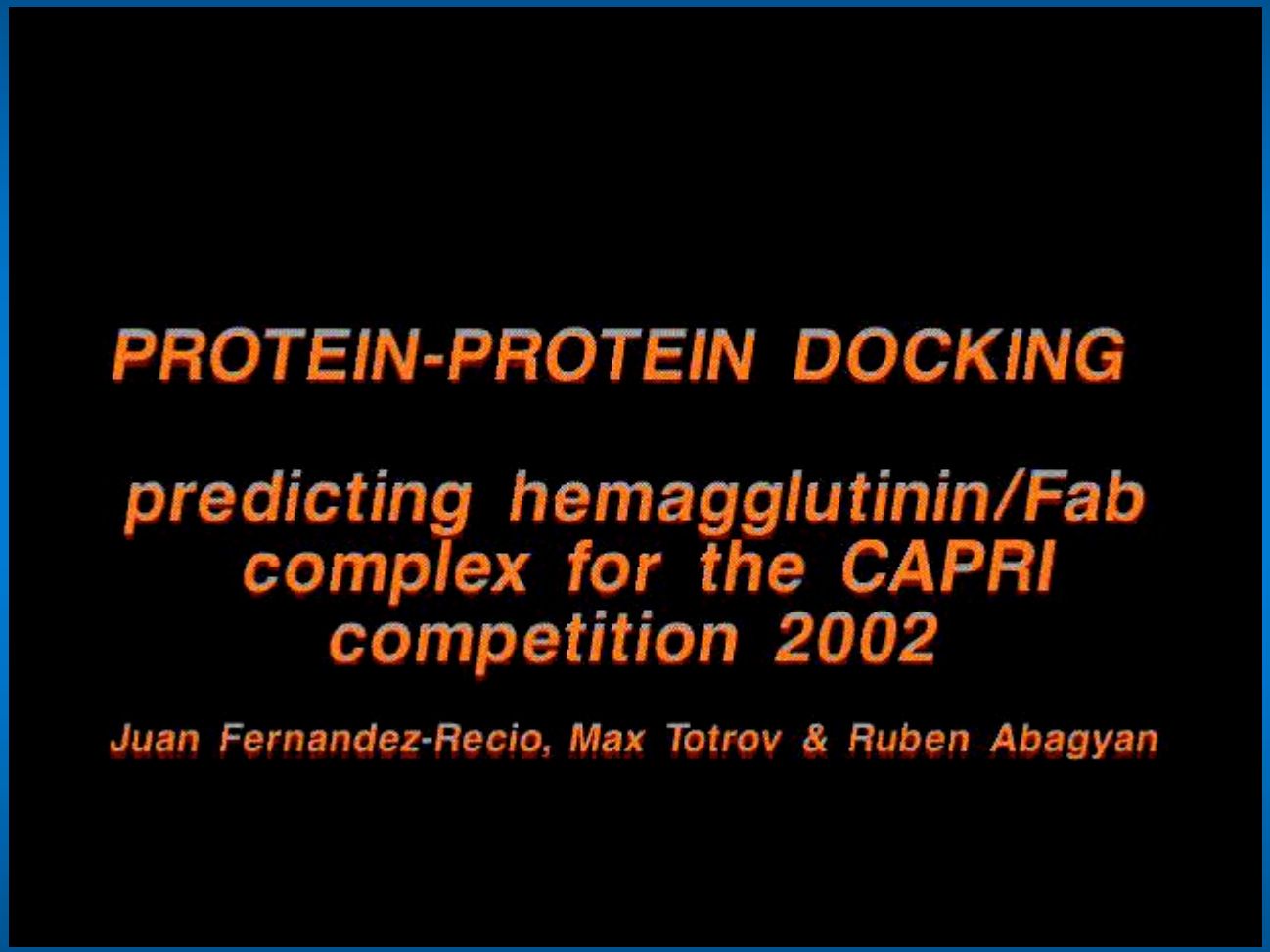
2nd CAPRI - Predictions

Predictor group	T08	T09	T10	T11	T12	T13	T14	T18	T19	Predictor summ
Abagyan	**	0	*	**	***	*	***	**	**	8/4**/2***
Wolfson	**	*	*	*	*	0	**	**	*	8/3**
Weng	**	0	0	*	***	***	***	**	**	7/3**/3***
Bates	*	0	*	**	*	0	**	**	*	7/3**
Baker	—	0	0	**	***	**	***	0	***	6/2**/4***
Camacho	**	0	0	0	***	***	**	**	*	6/3**/2***
Gray	***	—	—	**	***	0	0	0	**	5/2**/3***
Bonvin	—	—	**	**	0	***	***	0	0	5/3**/2***
CamPro	**	0	0	0	***	*	0	0	*	5/2**/1***
Sternberg	**	0	0	*	*	0	**	0	*	5/2**
								0	0	4/1**/2***
								0	0	4/1**/1***
								*	0	4/1**/1***
								0	0	3/1**/2***
								0	**	3/2**/1***
								0	—	3
								**	0	2/2**
								0	0	2/1**
								0	0	1/1***
								0	0	1
								—	—	1
								0	0	1
								0	0	1
								0	0	1



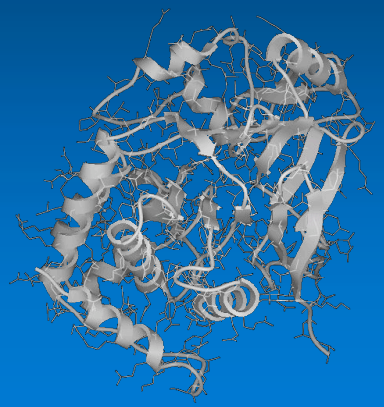
Fernandez-Recio et al. (2005) Proteins 60, 308-31

1st CAPRI – *Target 3*

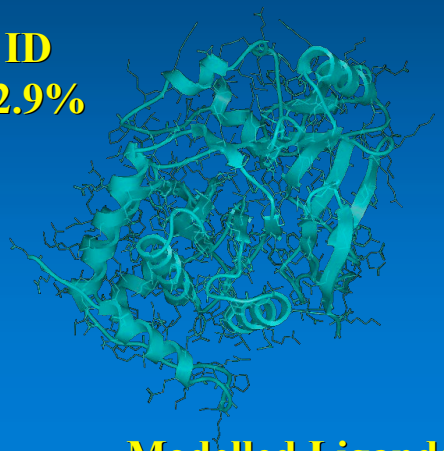


Fernandez-Recio et al. (2003) Proteins 52, 113-117

2nd CAPRI – *Target 14*

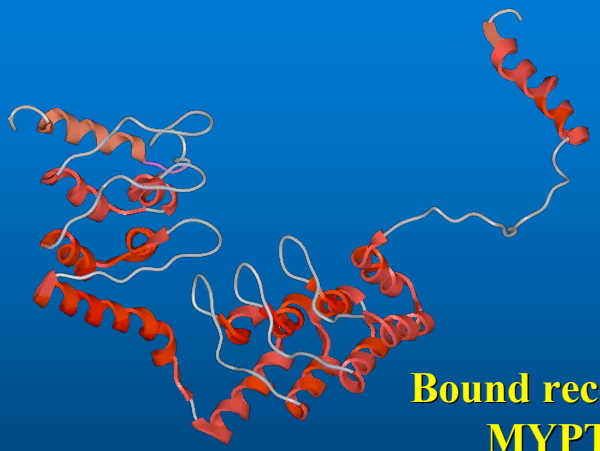


ID
92.9%



Template:
protein phosphatase 1α

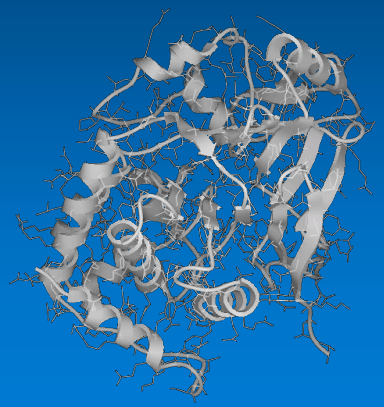
Modelled Ligand:
protein phosphatase 1β



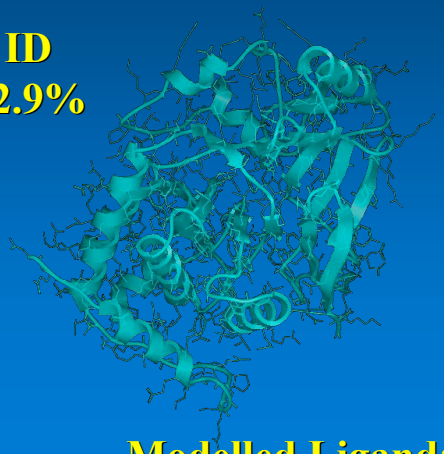
Bound receptor:
MYPT1

Fernandez-Recio et al. (2005) Proteins 60, 308-313

2nd CAPRI – *Target 14*



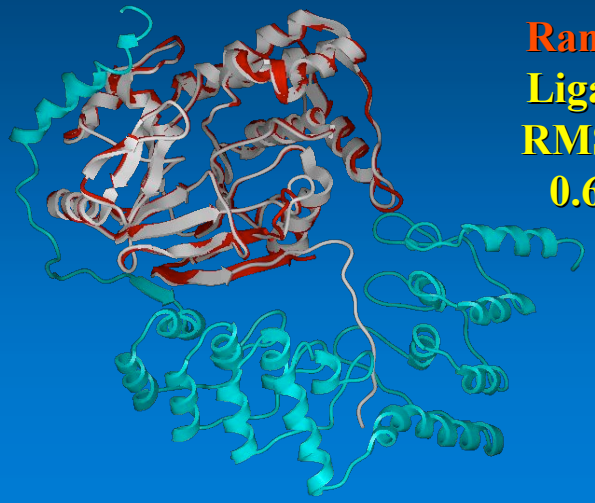
ID
92.9%



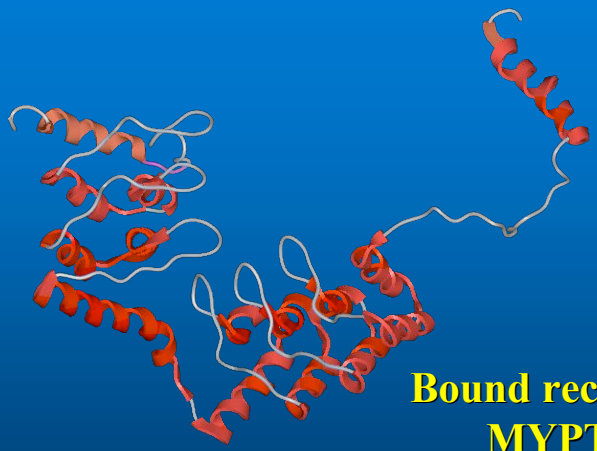
Template:
protein phosphatase 1α

Modelled Ligand:
protein phosphatase 1β

Docking



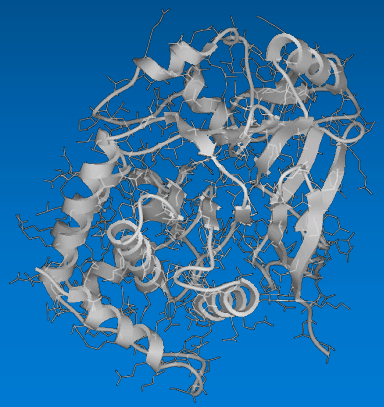
Rank 1
Ligand
RMSD:
0.6Å



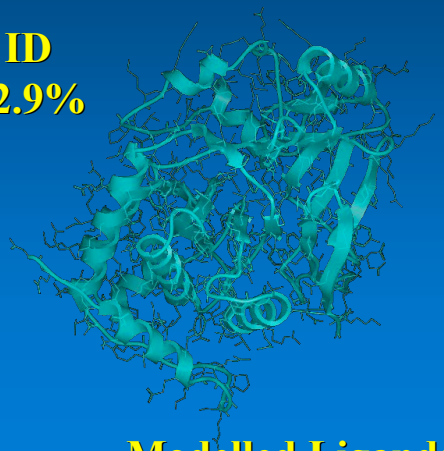
Bound receptor:
MYPT1

Fernandez-Recio et al. (2005) Proteins 60, 308-313

2nd CAPRI – *Target 14*



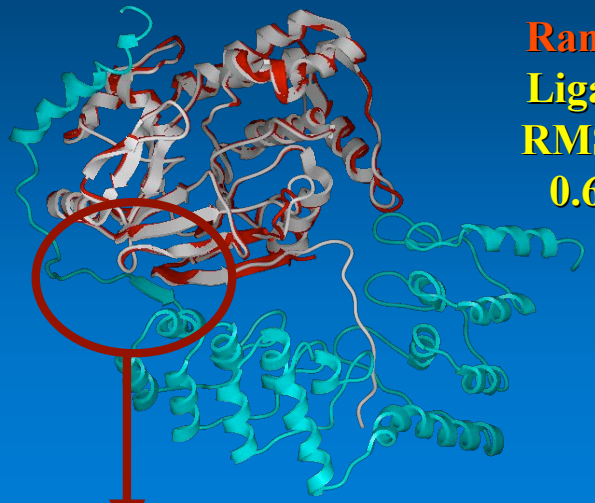
ID
92.9%



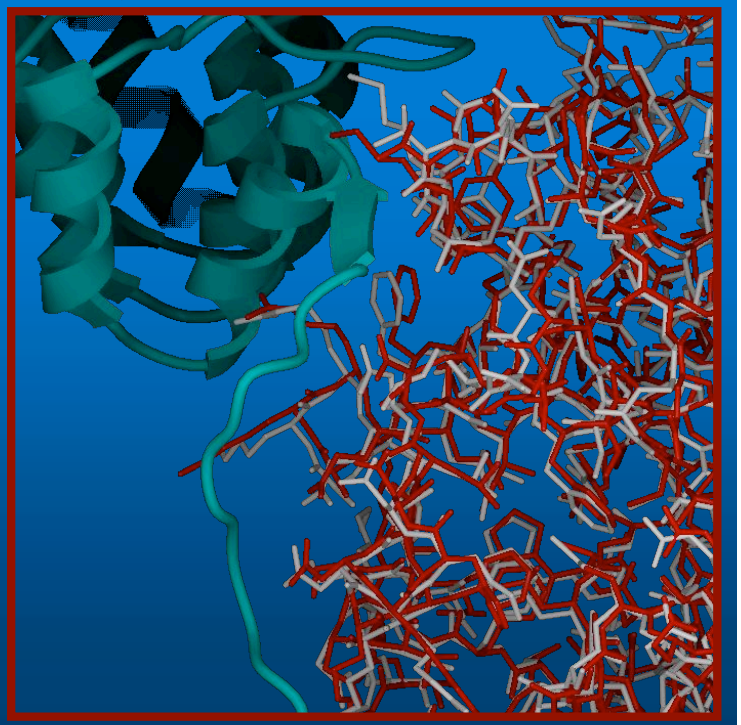
Template:
protein phosphatase 1α

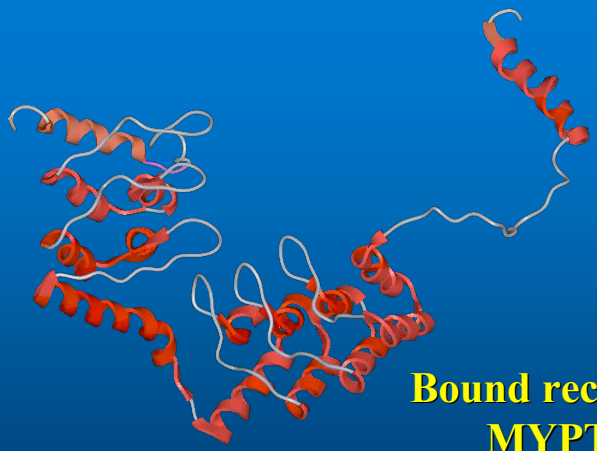
Modelled Ligand:
protein phosphatase 1β

Docking



Rank 1
Ligand
RMSD:
0.6Å





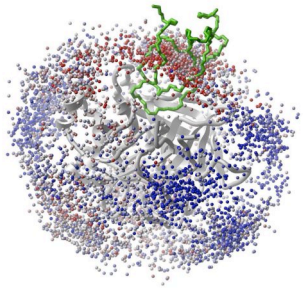
Bound receptor:
MYPT1

Fernandez-Recio et al. (2005) Proteins 60, 308-313

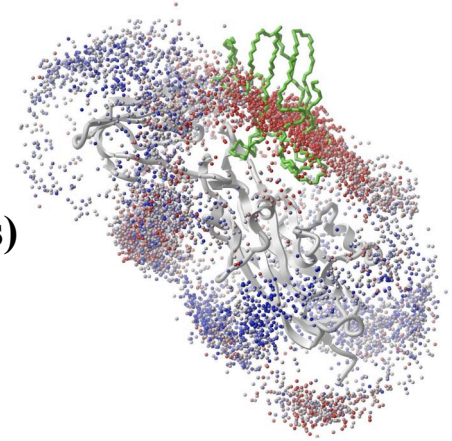
- Introduction
- Computational protein-protein docking
- Geometric docking algorithms
- Docking by global energy optimization
- Comparison of docking methods
- **Present and future challenges in protein-protein docking**

Interface propensity maps from docking landscape

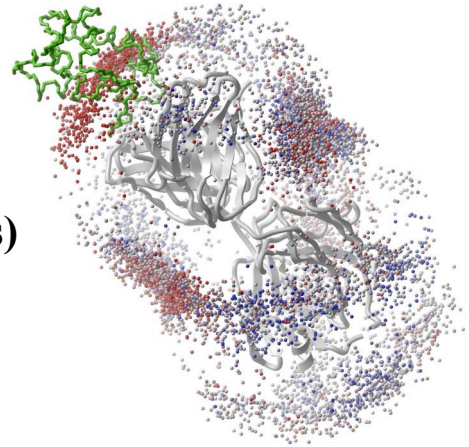
1ca0
(all docking poses)

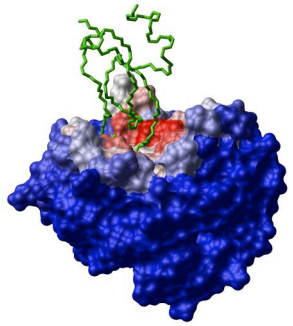


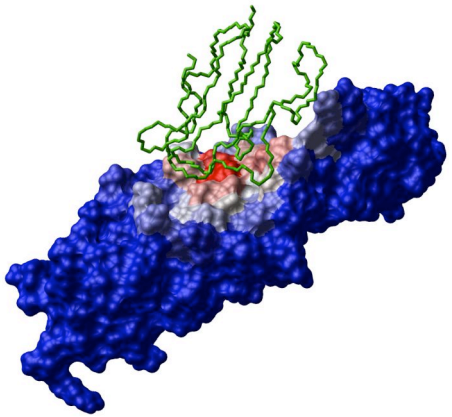
2pcf
(all docking poses)

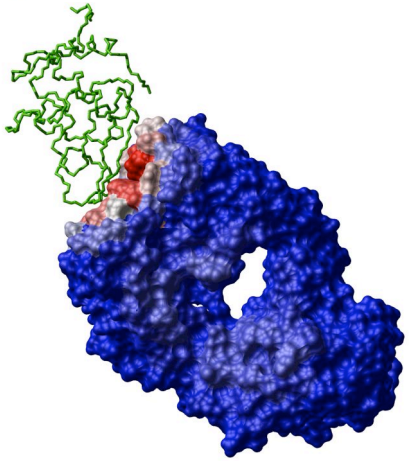


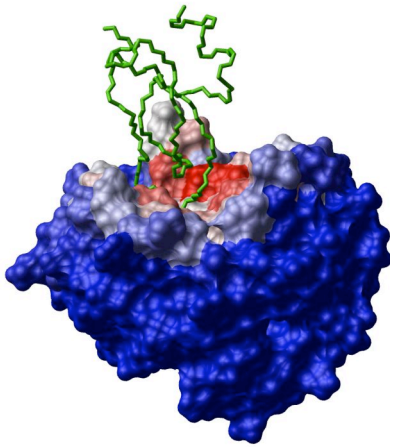
1mlc
(all docking poses)



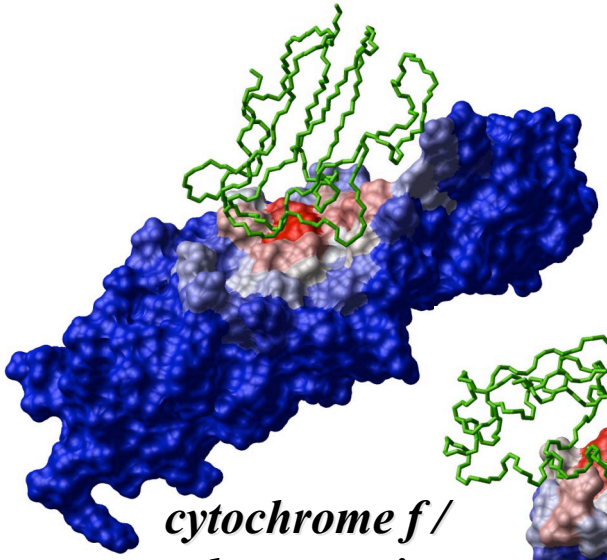




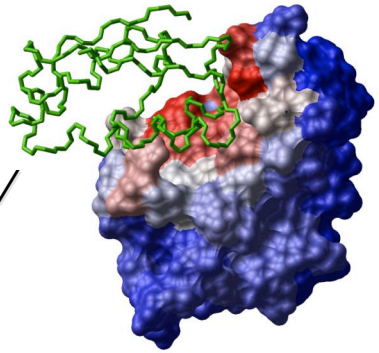




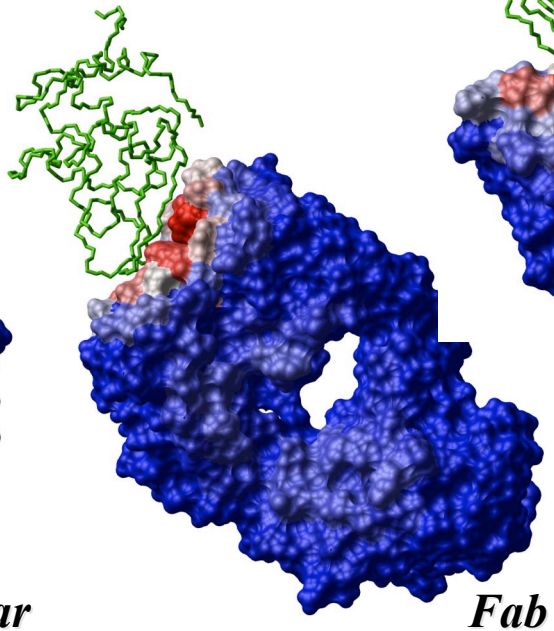
chymotrypsin / APPI



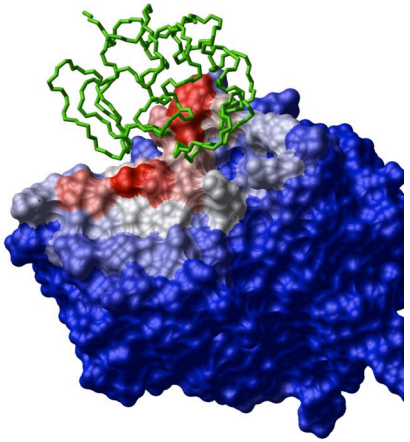
cytochrome f / plastocyanin



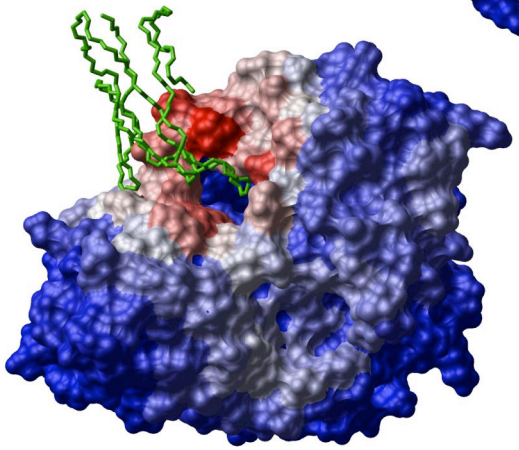
barnase / barstar



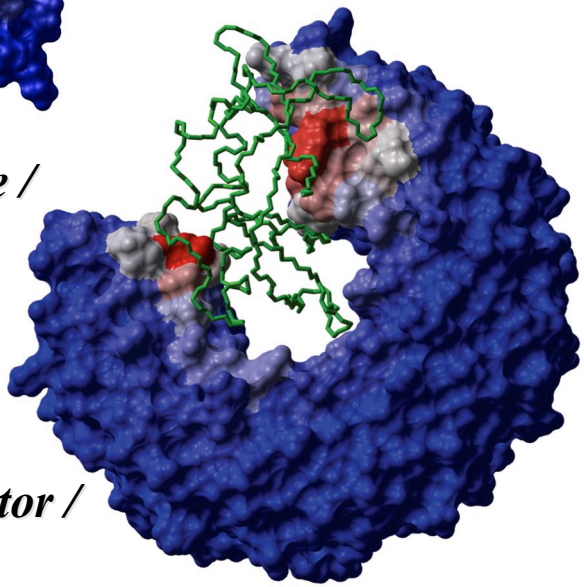
Fab / lysozyme



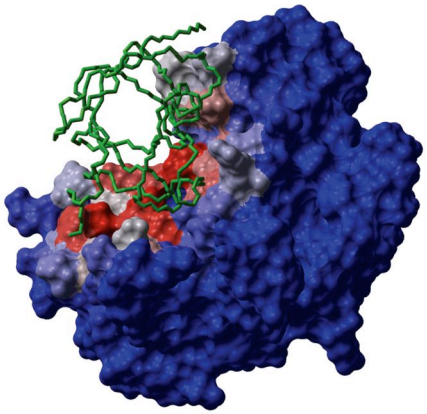
ccp / cytochrome c



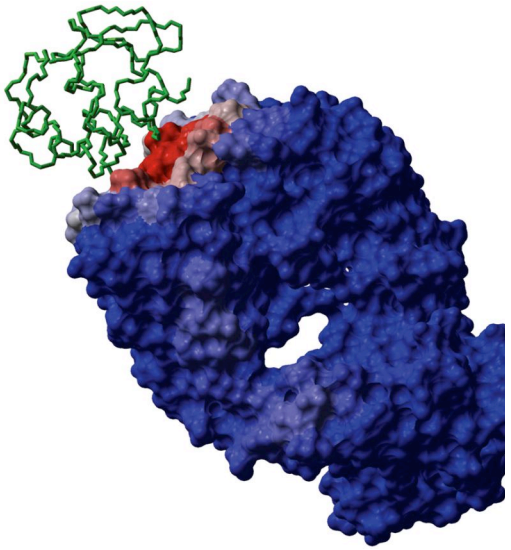
acetylcholinesterase / fasciculin II



ribonuclease inhibitor / ribonuclease A

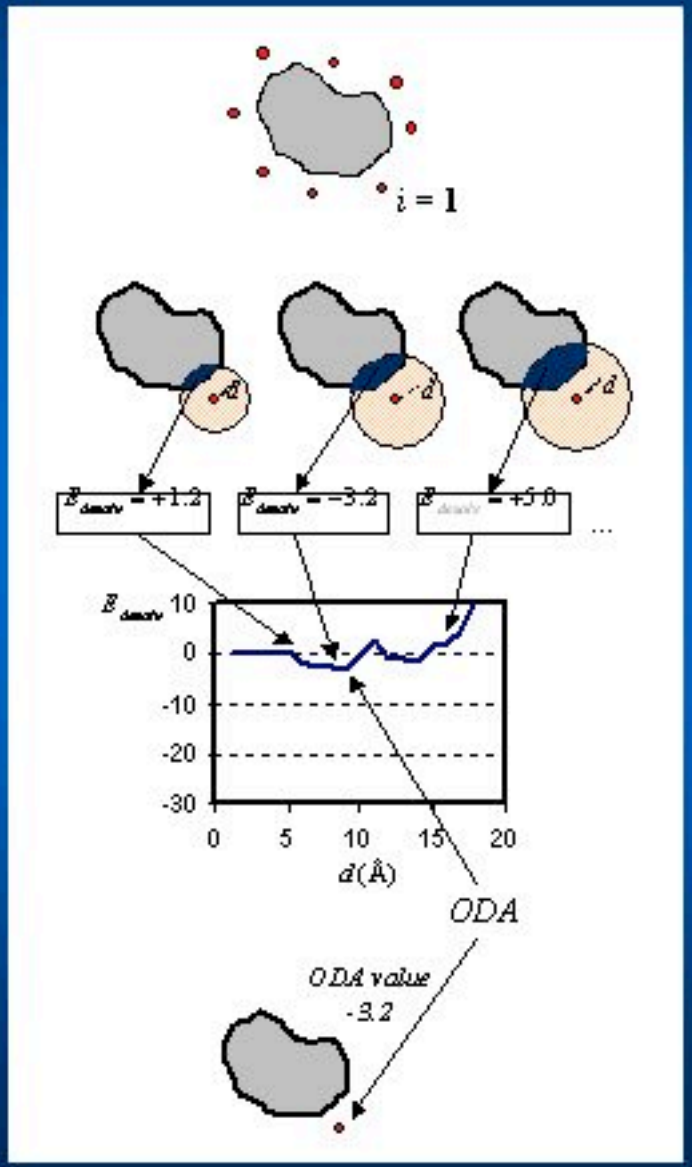


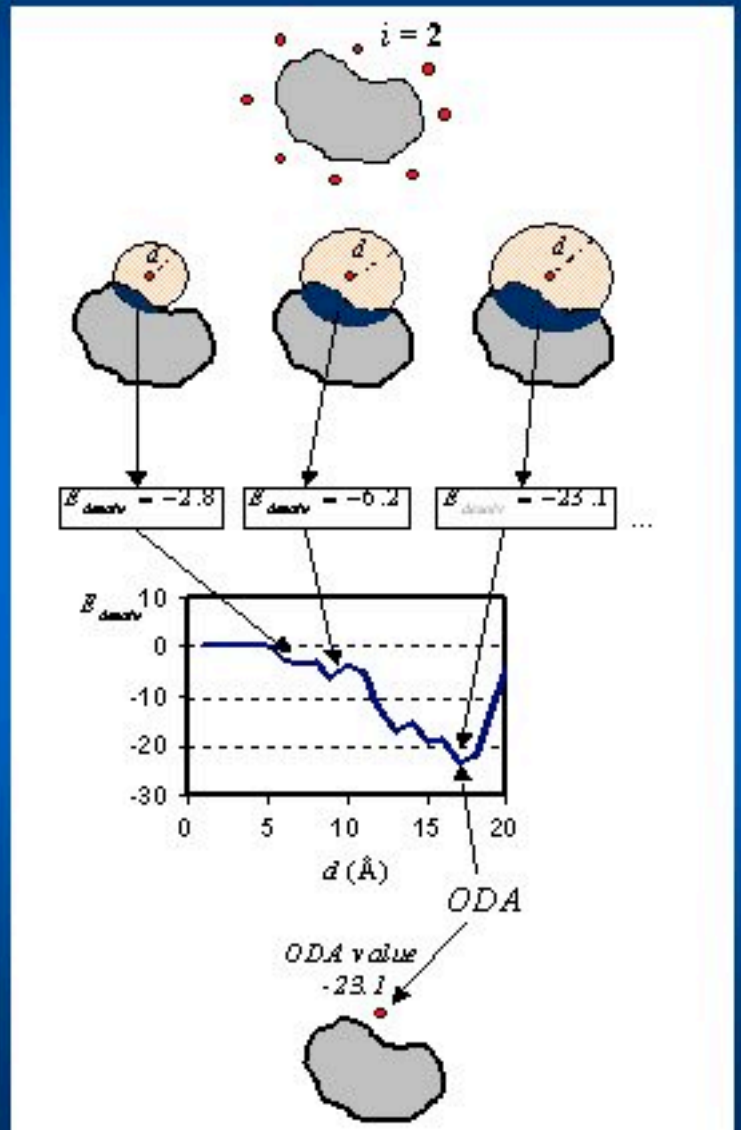
FNR / ferredoxin

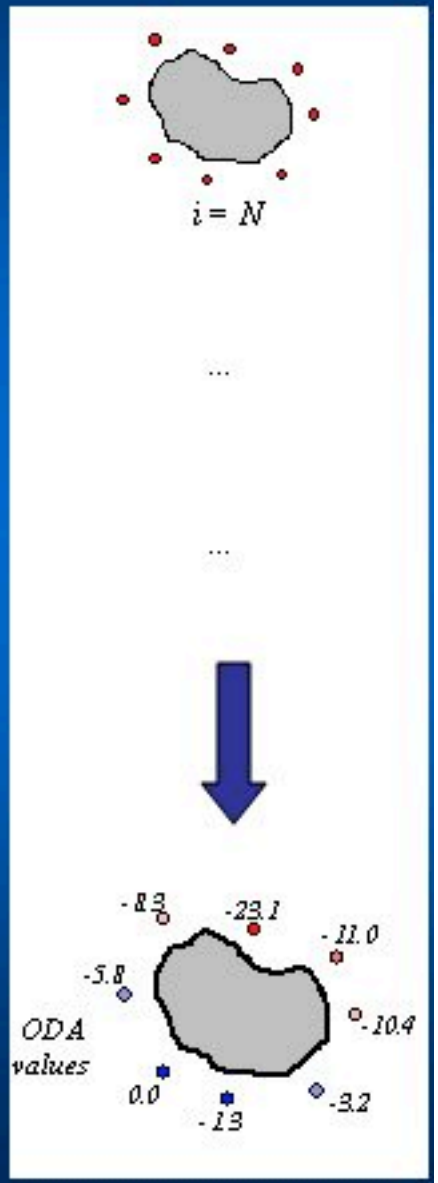


Fab / cytochrome c

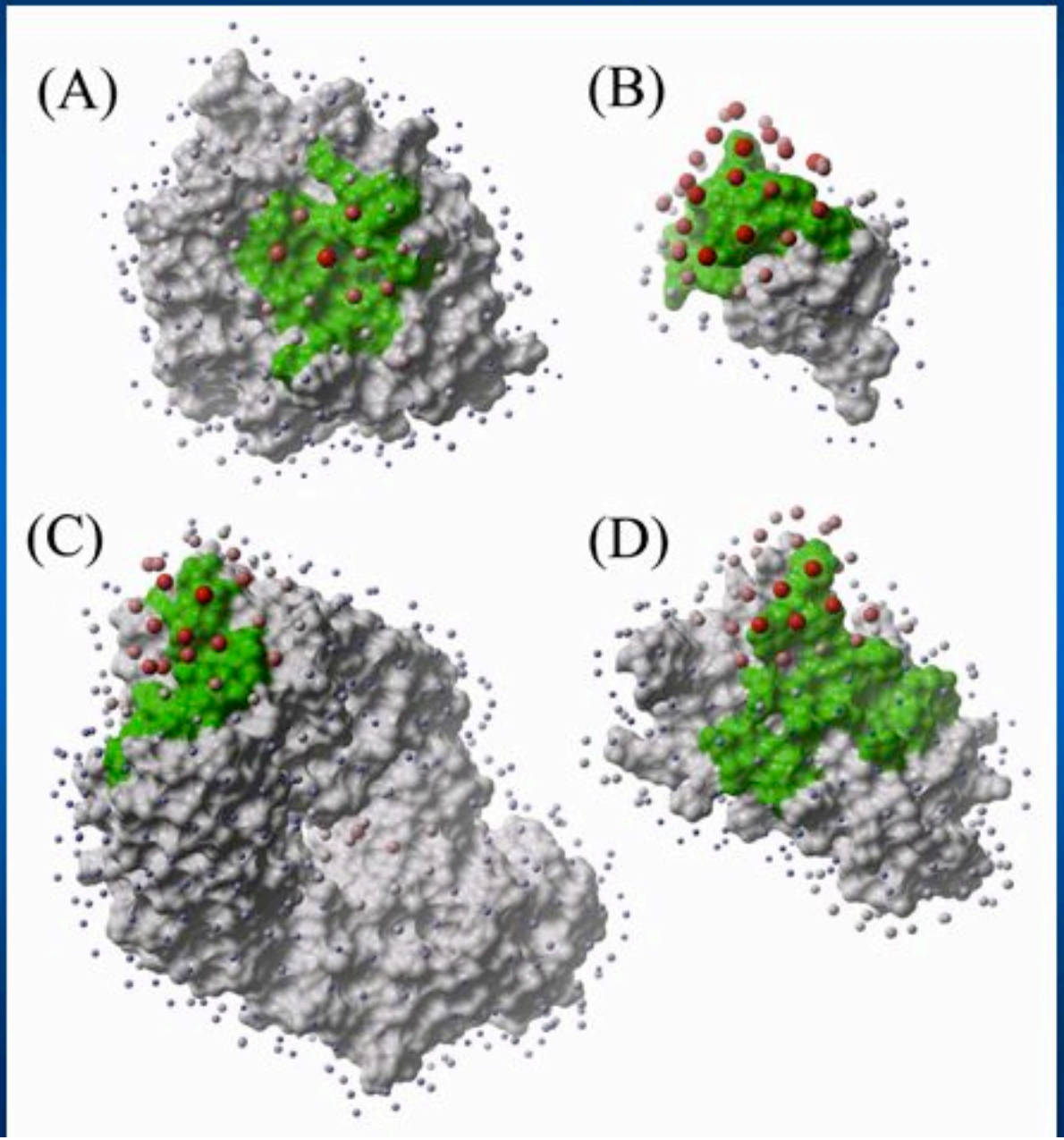
Optimal Docking Areas







Optimal Docking Areas



Benchmark: 66 unbound cases

- ODAs in 50% proteins

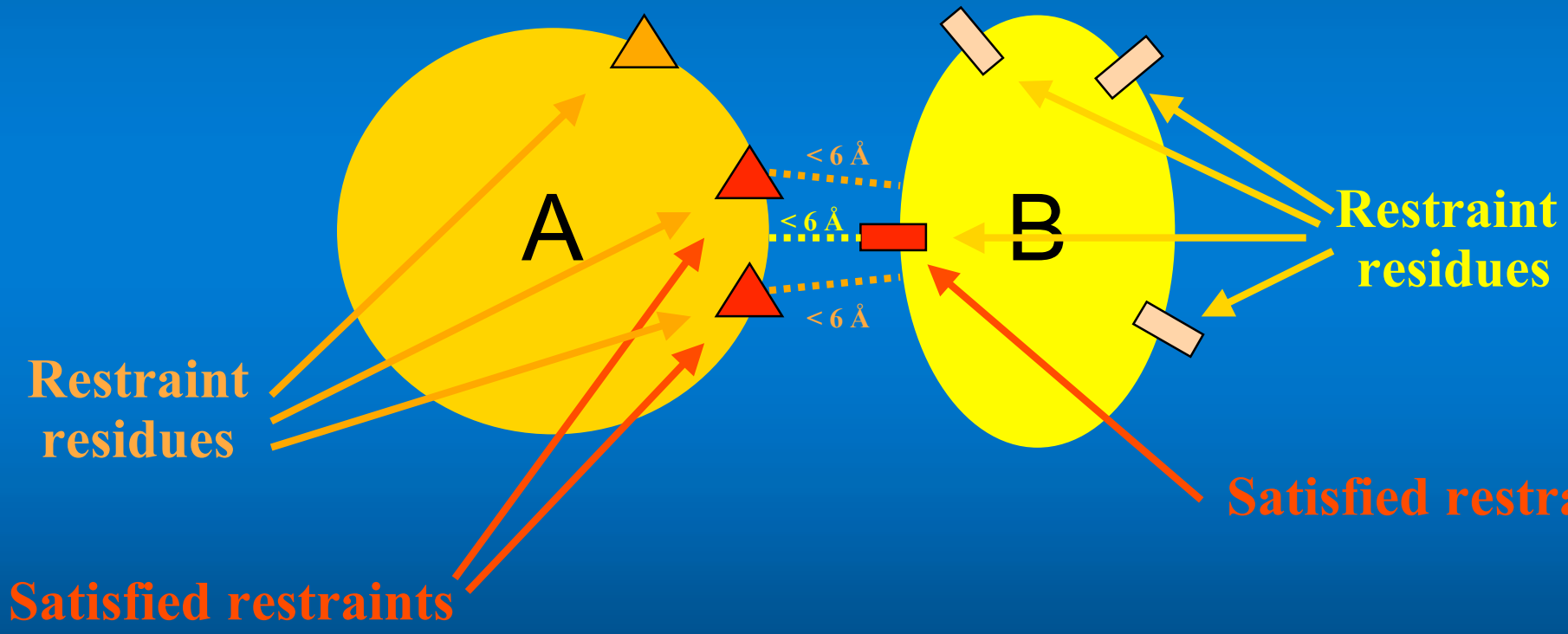
- 80% correctly located in binding sites

- pyDockODA (centered in residues); web server in INB

Fernández-Recio et al. (2005) Proteins 58, 134-143

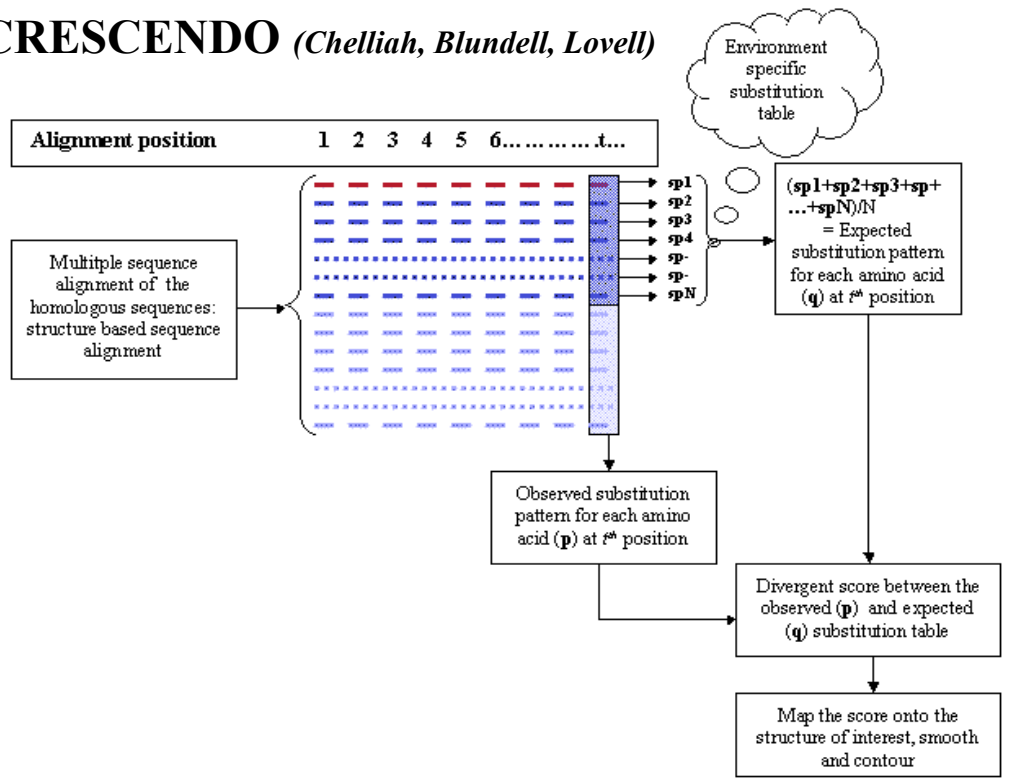
pyDockRST: use of restraints to filter docking solutions

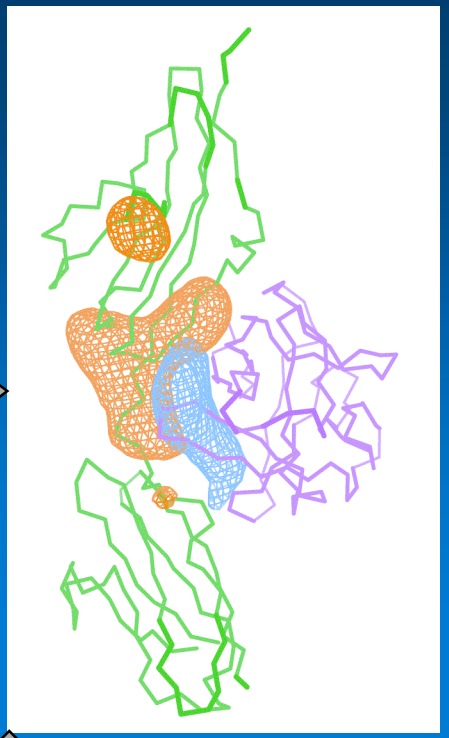
Docking solution *i*



Pseudo-Energy = -100*(satisfied restraints / total restraint residues)

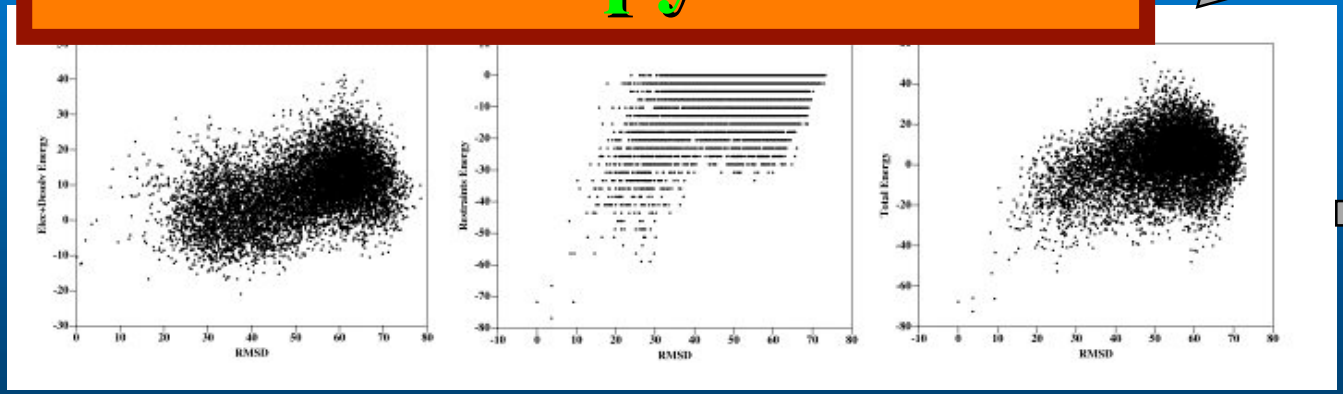
CRESCENDO *(Chelliah, Blundell, Lovell)*

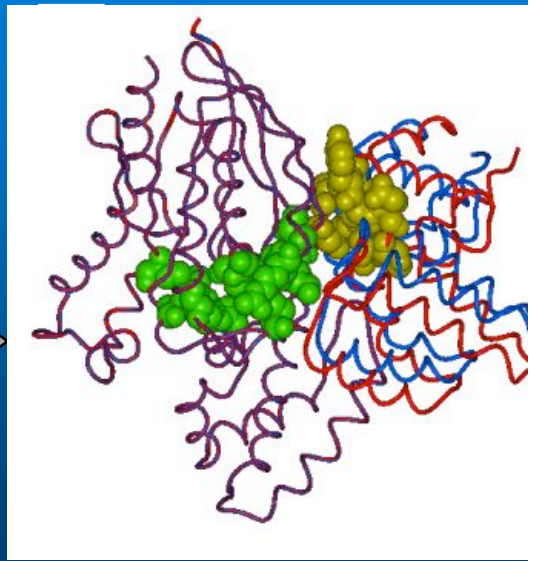




Crescendo + pyDockRST

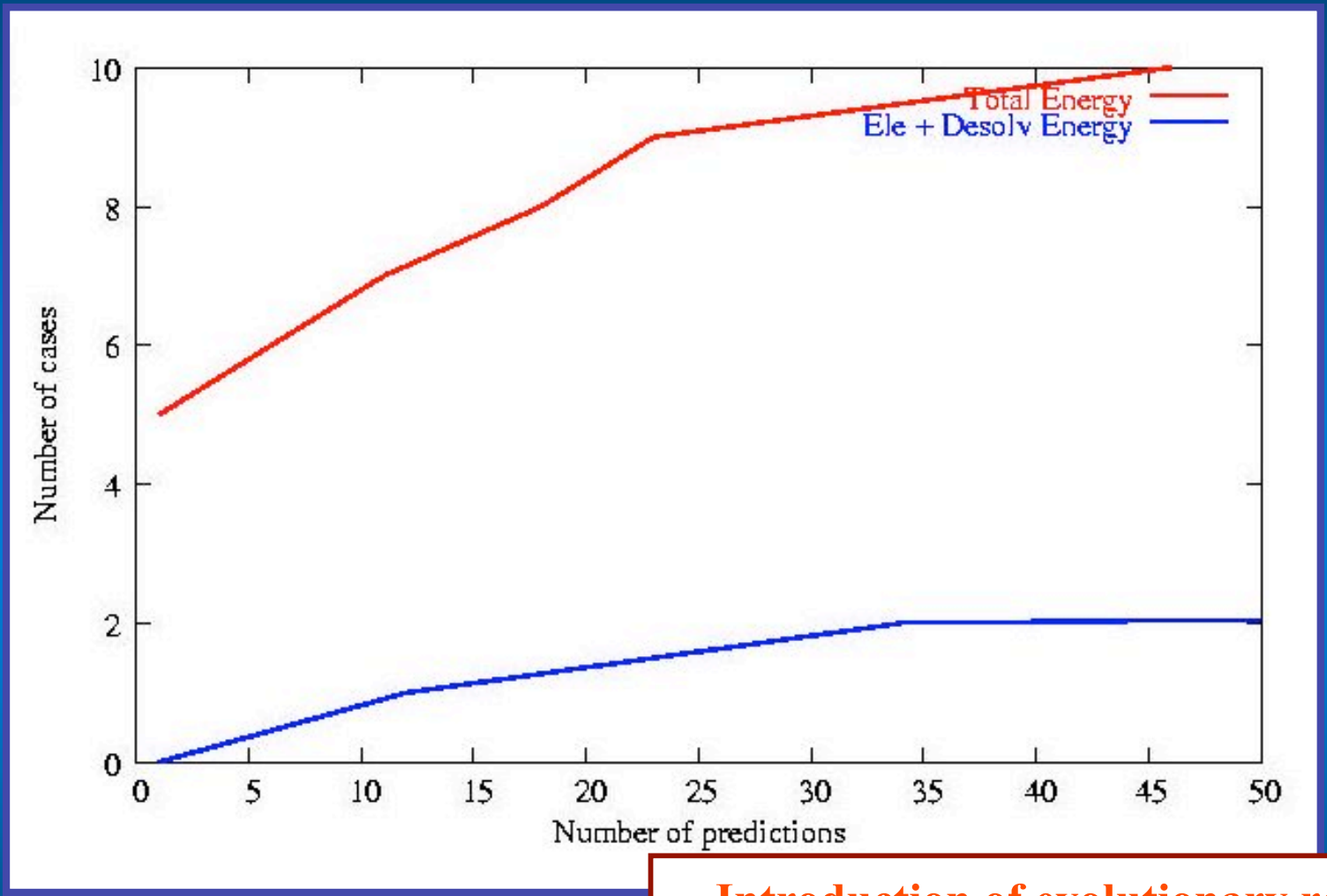
pyDockRST





Chelliah et al. (2006) JMB 357, 1669-1682

Crescendo + pyDockRST



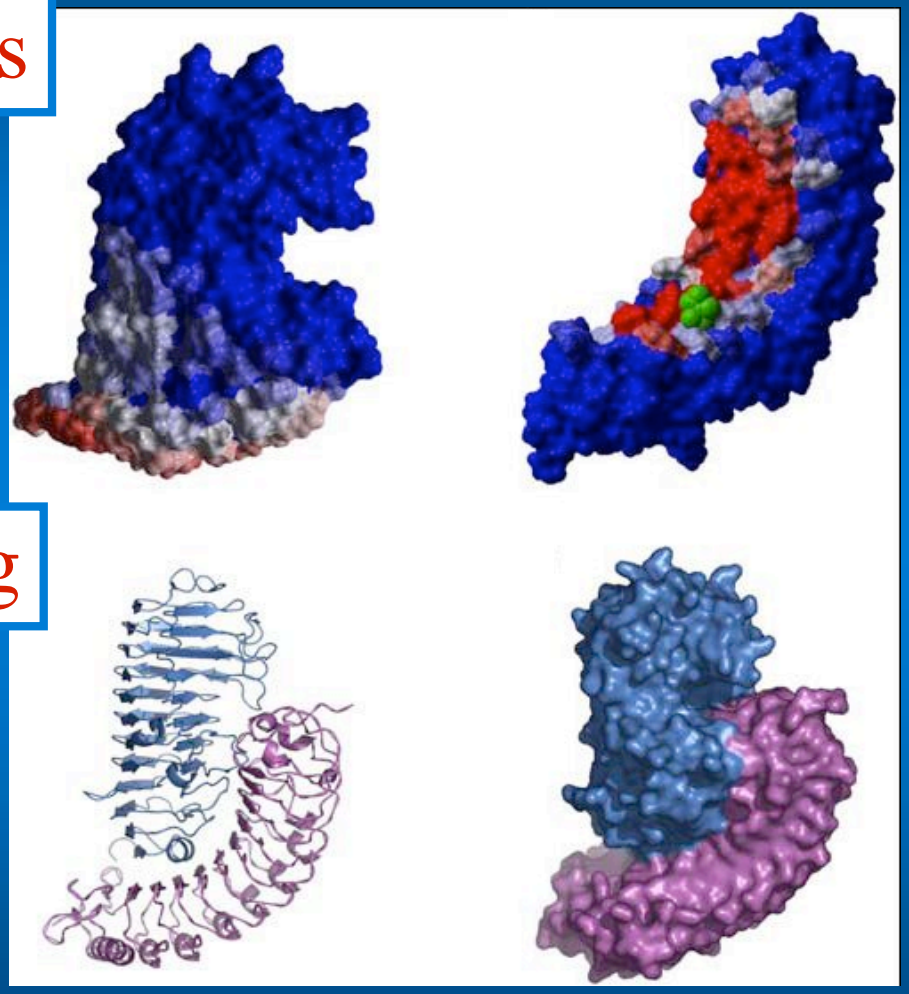
Introduction of evolutionary restraints dramatically improves the docking results

Chelliah et al. (2006) JMB 357, 1669-1682

One example of modelling by docking: PG / PGI

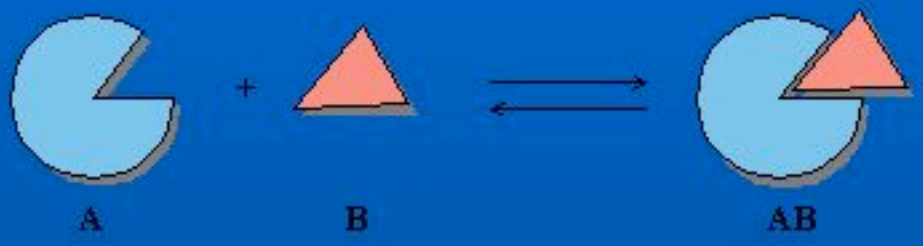
NIP surfaces

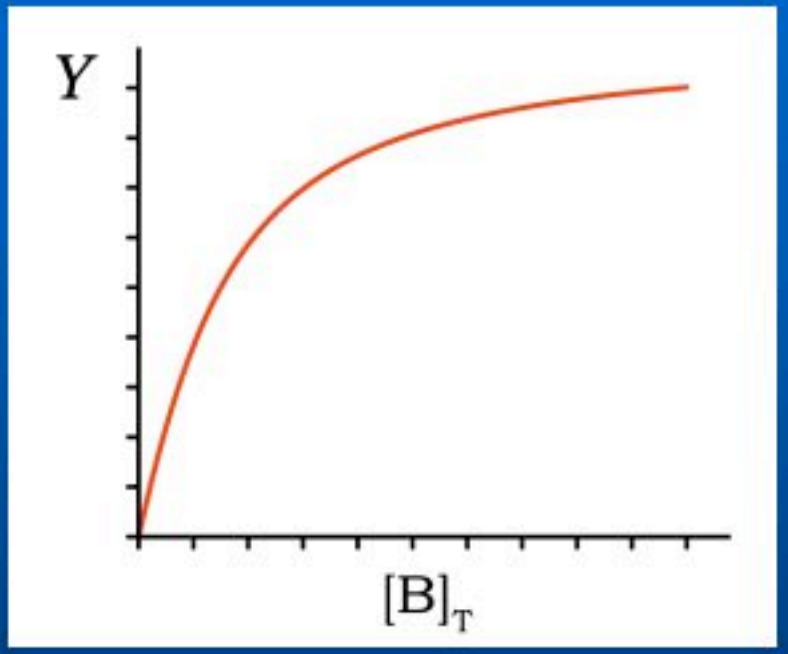
docking



Sicilia et al. (2005) Plant Physiol. 139, 1380-1388
Federici et al. (2006) Trends Plant Sci. 11, 65-70

Protein-Protein Interaction: Thermodynamics



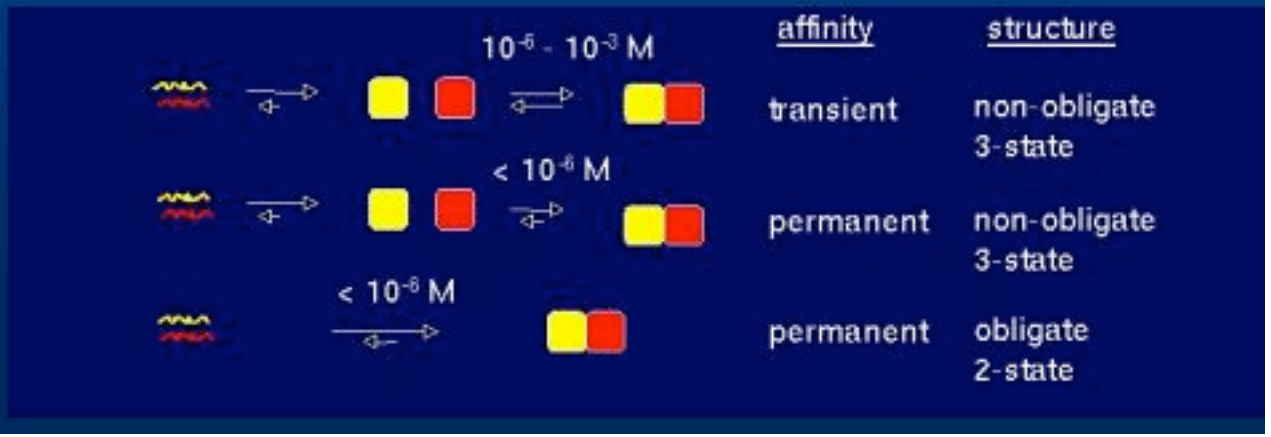


$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} AB$$

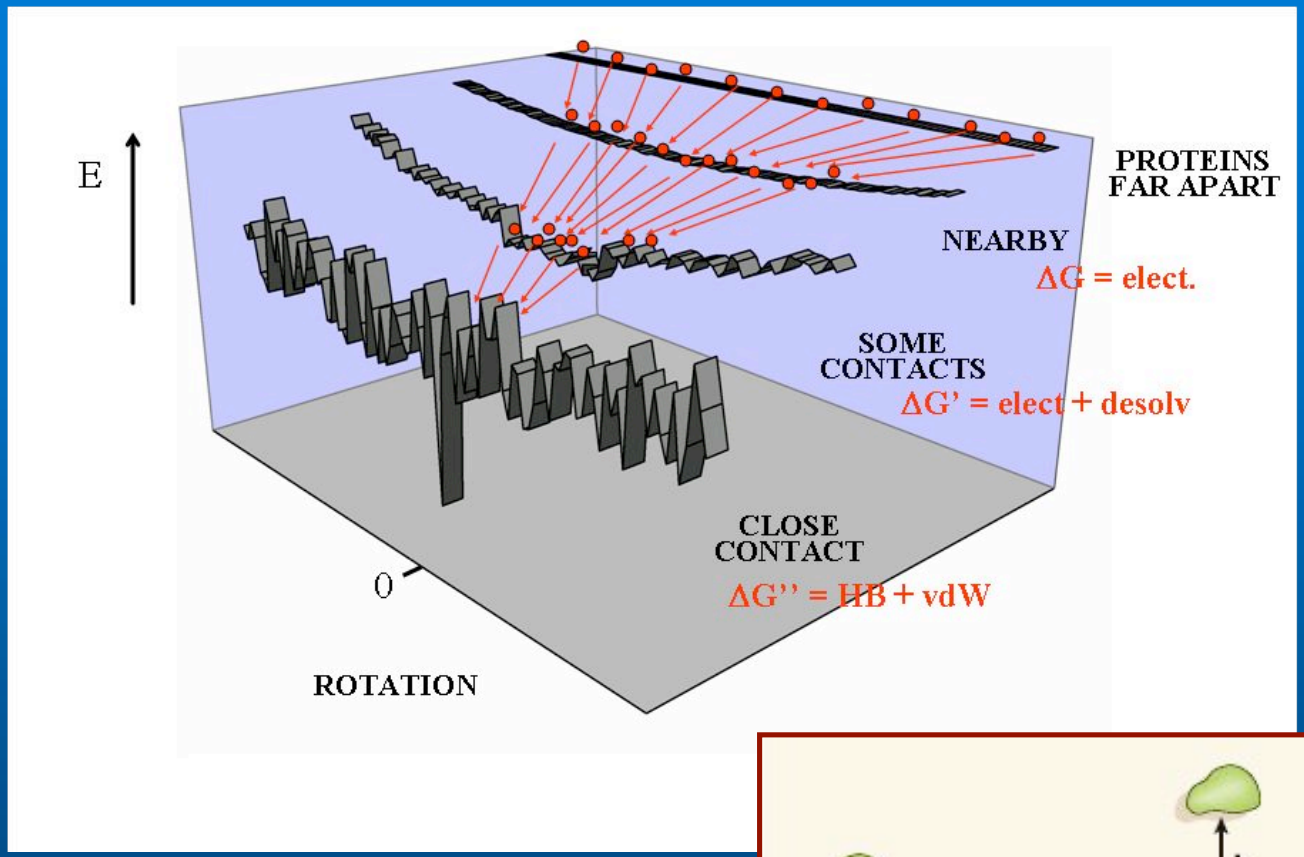
$$K_a = \frac{1}{K_d} = \frac{k_a}{k_d}$$

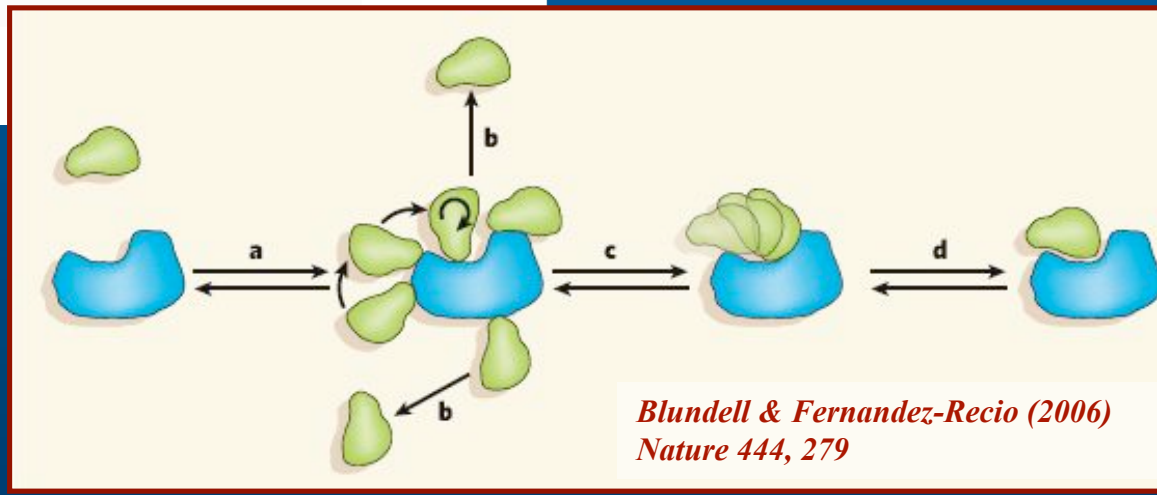
$$K_a = \frac{1}{K_d} = \frac{[AB]}{[A][B]}$$

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



Protein-Protein Docking Mechanism





Blundell & Fernandez-Recio (2006)
Nature 444, 279

References

Protein-Protein Interaction - General

- Protein-Protein Recognition, C. Kleanthous ed., Oxford University Press
- Conte *et al.* (1999) J. Mol. Biol 285, 2177-2198
- Estructura de Proteínas, C. Gómez-Moreno & J. Sancho coord., Editorial Ariel

Docking Simulations

- Katchalski-Katzir *et al.* (1992) PNAS 89, 2195-2199
- Halperin *et al.* (2002) *Proteins* 47, 409-443
- Smith & Sternberg (2002) COSB 12, 28-35

CAPRI

- Proteins Special Issues (July 2003, July 2005)