M1 ISDD (Feb 2018)

PROTEIN DOCKING

Lesson 5

Benchmarking and experimental cases

Juan Fernández-Recio

CSIC, BSC

juanf@bsc.es

- **Protein-protein docking benchmarks**
- **CAPRI**
- **Application to experimental cases**

Docking software (I)

FRODOCK	Fast Rotational DOCKing. Generates very efficiently many potential predictions of how two proteins could interact	online
DOCK/PIE(RR)	Web Server for Structure Prediction of Protein-Protein Complexes	online
ZDOCK	ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers	online
dDFIRE/DFIRE2	Energy calculation	online
PRISM	PRotein Interactions by Structural Matching	online
PREDDIMER	PREDDIMER: a web server for prediction of transmembrane helical dimers	online
SPRING	SPRING is a template-base algorithm for protein-protein structure prediction	.
COTH	COTH (CO-THreader) is a multiple-chain protein threading algorithm which is designed to identify and recombine protein complex structures from both tertiary and complex structure libraries	.
TACOS	TACOS (Template-based Assembly of Complex Structures) is designed to model the structure of protein-protein complexes based on a hierarchical approach of template identification and structural refinement	.
HOMCOS	HOmology Modeling of protein COmplex Structure	.
Udock	The Interactive Docking Entertainment System	standalone (windows)
DockTrina	Docking triangular protein trimers	docking, standalone
DockRank	Ranking docked conformations using partner-specific sequence homology based protein interface prediction	online scoring
MEGADOCK	An All-to-all Protein-protein Interaction Prediction System Using Tertiary Structure Data	PPI docking, standalone

http://www.vls3d.com/index.php/links/bioinformatics/protein-protein-interaction/protein-protein-docking

Docking software (II)

ATTRACT	Docking Program (Fortran-Version, full source code and manual)	standalone
SwarmDock	A server for flexible protein-protein docking	online docking
PRUNE and PROBE	Two modular web services for protein-protein docking	docking online
3D-GARDEN	Global and Restrained Docking Exploration Nexus 3D-GARDEN is a state-of-the-art comprehensive software suite and server for protein-protein docking with full high-performance computing functionality	online docking
DOCKGROUND	The resource implements a comprehensive database of co-crystallized (bound-bound) protein-protein complexes	Database of experimental complexes
FireDock	The server addresses the refinement problem of protein-protein docking solutions	Macromolecular docking online
Grammx	Tools for protein-protein docking. GrammX: web interface of Gramm	Macromolecular docking online, see also standalone
LIGIN	Molecular docking using surface complementarity. The LIGIN program is also available as part of the WHATIF software package	Macromolecular docking standalone
PatchDock	Protein docking tools (PatchDock) and related. PatchDock, webserver for macromolecules and small molecules docking based on shape complementarity criteria. There are many other tools here including tools for peptides, flexibility, comparing binding pockets...	online docking
AquaSAXS	A web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models	online
DOCK/PIE(RR)	Protein Docking Server	docking online
PBSword	A web server designed for efficient and accurate comparisons and searches of geometrically similar protein-protein binding sites from a large-scale database	online

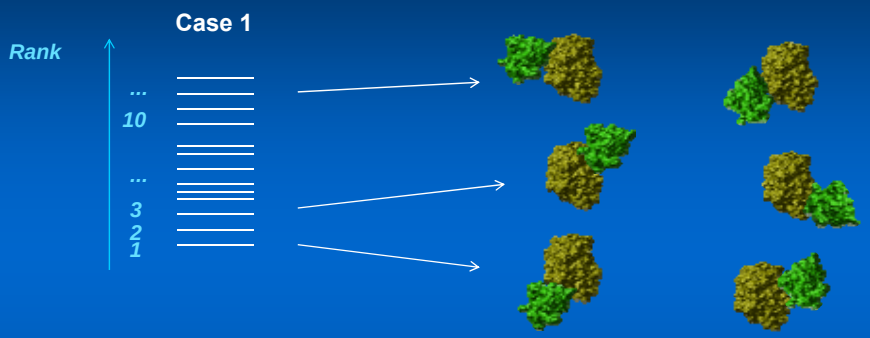
http://www.vls3d.com/index.php/links/bioinformatics/protein-protein-interaction/protein-protein-docking

Docking software (III)

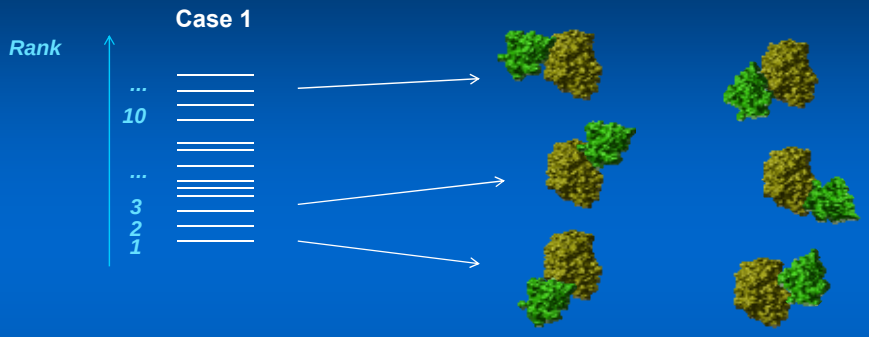
pyDockWEB	A web server for rigid-body protein-protein docking using electrostatics and desolvation scoring	online
PyDock	Tool for protein-protein docking. The module ODA can help to predict potential protein-protein interaction regions. pyDock is a fast protocol which uses electrostatics and desolvation energy to score docking poses generated with FFT-based algorithms	standalone
HINT	HINT (High-quality INTeractomes) is a database of high-quality protein-protein interactions in different organisms	database
HADDOCK	Docking driven by interface restraints	docking, also online
SmoothDock	Protein docking	docking online
Bipdock	Bielefeld Protein Docking Software	docking standalone
ZDOCK	docking based on FFT search	docking, standalone
ClusPro	Protein-protein docking webserver using 3 docking programs - DOT ZDOCK GRAMM	docking online
PIC	Protein Interactions Calculator	Protein-protein energy computation, online, structural analysis
DOT	Protein-protein docking software	docking standalone
ROSIE	ROSIE, including rosetta Protein-protein docking	docking online
CombDock	Combinatorial assembly of multiprotein complexes by multiple docking (see also Firedock)	docking standalone
RosettaDock	The RosettaDock server	docking
BDOCK	Protein-protein docking software integrating the degree of burial of surface residues into protein-protein docking	docking standalone
MolFit	Protein-protein docking software estimating the extent of geometric and chemical surface complementarity	docking
Hex	Protein-protein docking and molecular superposition program	docking online
ESCHER-NG	Protein-protein and DNA-protein docking software	Macromolecular docking
FTDock	Fourier Transform Dock	docking standalone

http://www.vls3d.com/index.php/links/bioinformatics/protein-protein-interaction/protein-protein-docking

Assessment of docking performance

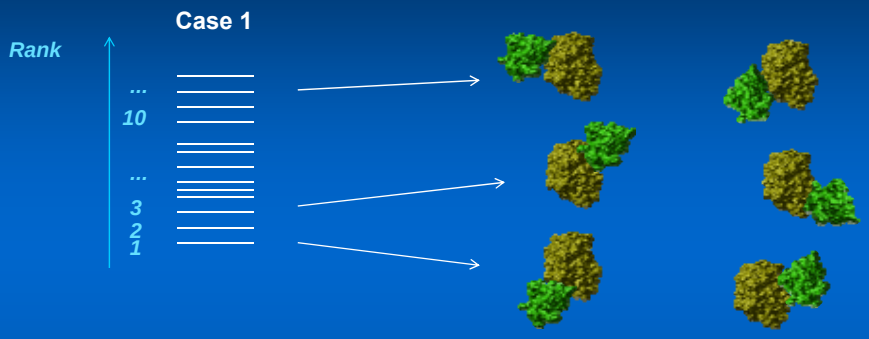


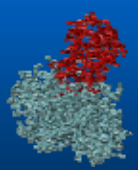
Assessment of docking performance



good solution ??

Assessment of docking performance





reference complex

good solution ??

Assessment of docking performance

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

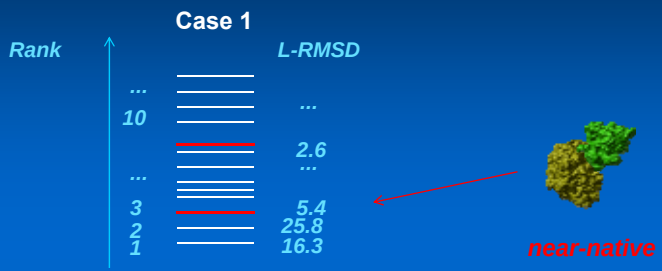
$$F_{nat} = \frac{\text{Correctly predicted contacts}}{\text{Total number of contacts in the target}}$$

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

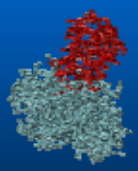
near-native: L-RMSD < 10Å

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

Assessment of docking performance

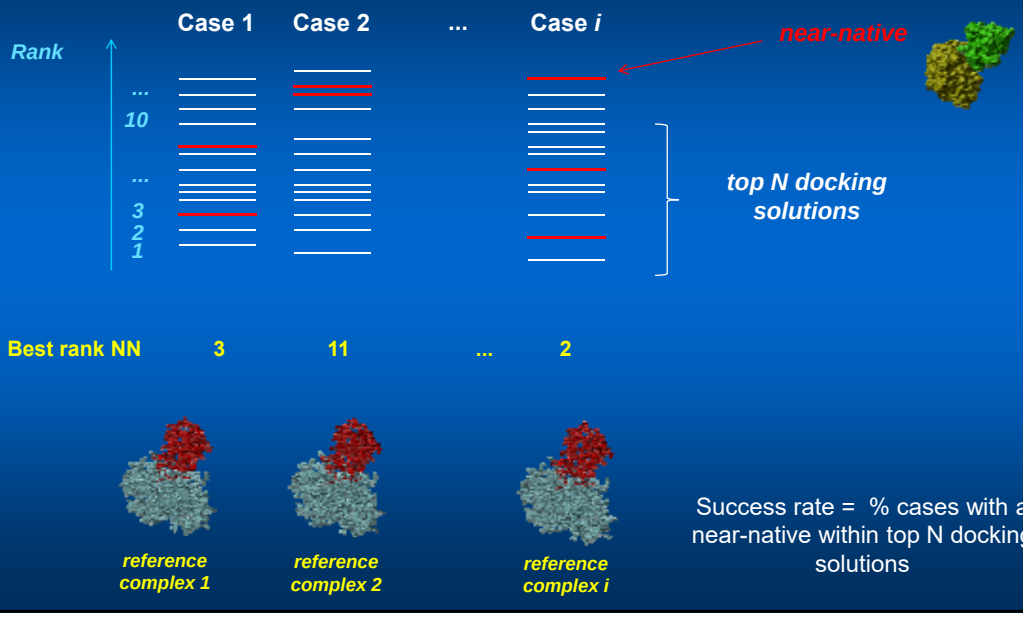


Best rank NN **3**



reference complex

Assessment of docking performance



Success rate = % cases with a near-native within top N docking solutions

Protein-protein benchmarking - early days

before 2002 < 11 unb-unb cases

FTDock (1997)

Table 3. Docking unbound subunits: global search and local refinement

	Filtered global search				Local refinement (1.5 Å surface)			
	N	Rank	*N* ⩽ 2.5 Å	Cα RMS	*N*	Rank	*N* ⩽ 2.5 Å	Cα RMS
CGI	172	5	2	2.1/1.9	94	3	1	1.8/2.0
CHO	234	12	7	1.4/1.3	86	11	5	1.2/1.1
KAI	377	64	19	2.2/1.8	364	130	18	1.5/1.2
PTC	242	39	8	1.9/2.0	229	16	8	1.5/1.7
SNI	29	19	2	1.8/1.6	26	8	2	1.8/1.6
SIC	–	–	–	–/–	–	–	–	–/–
FDL	717	22	2	2.1/1.8	707	176	2	2.1/1.8
MLC	623	13	5	2.2/1.9	590	41	4	1.2/1.6
HFL	544	460	2	2.3/1.8	519	228	2	1.8/1.5
HFM	837	27	7	1.7/1.5	762	65	6	1.1/1.1

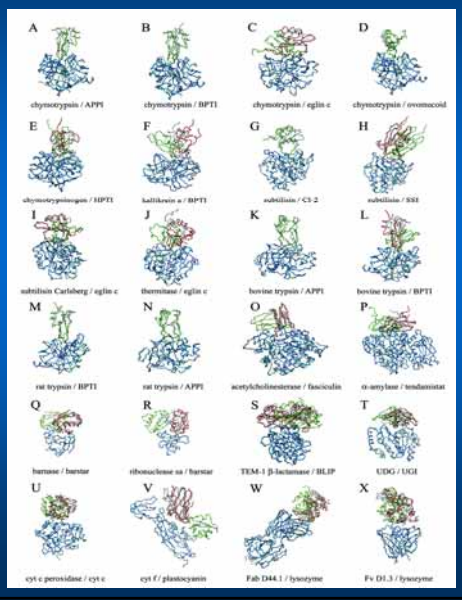
Nussinov (1999)

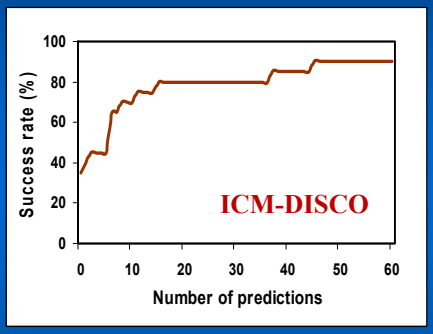
BiGGER (2000)

Dock name	Scoring Rank	RMS
BOUND		
2SICXX	2	3.76
1SBNXX	31	2.23
1TECXX*	77	3.57
1ACBXX*	18	0.61
1SDHXX*	1	3.17
1OCPXX*	1	3.32
PSEUDO-UNBOUND		
1SDHXF	12	2.33
1UNGXF*	2	0.58
4EBXXF	200	3.54
2MIPXF	82	2.21
1HFLXF*	43	3.70
1HPMXF	13	3.89
1CTAXF*	1	1.36
UNBOUND		
2PTCFF*	52	2.73
2SICFF*	15	3.33
1BRIFF*	35	1.89
2PCCFF	50	3.87
2PCBFF*	18	2.36
2SNIFF*	16	1.32
1FSSFF*	11	3.20
1CHOFF	6	2.93
1MLCFF	—	—
1FDLFF	—	—
2KAIFF	—	—
1CGIFF*	9	3.72

Protein-protein benchmarking - early days

2002 24 cases

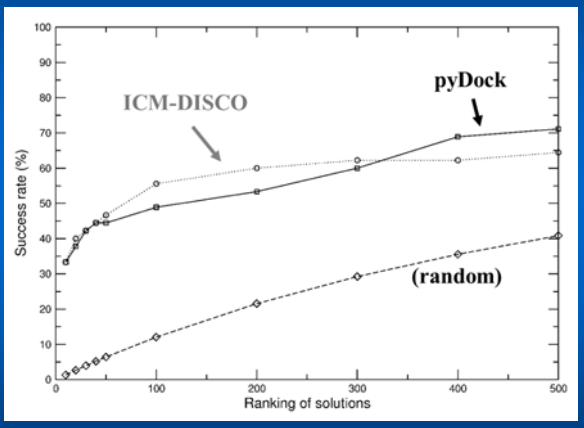




Protein-protein benchmarking - early days

2004 45 cases

Complex[a]
Parameterization set
1ca0
1cbw
1acb
1cho
1cgi
2kai
2sni
2sic
1cse
2tec
1taw
2ptc
3tgi
1bre
1fss
1bvn
1bgs
1ay7
TEM1[e]
1ugh
2pcb
2jel
1mlc
1vfb
Test set
1ewy
1eer
1kkl
VP6FAB[f]
1ken
1kxv
1kxt
1kxq
1l0x
1avw
1dfj
1bgs
1ahw
1dqj
1wej
1avz
1wq1
2mta
1bth
1fin
1fq1



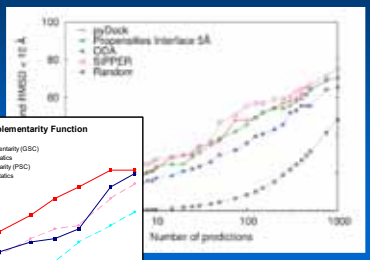
ICM-DISCO's docking benchmark (45 unbound cases)

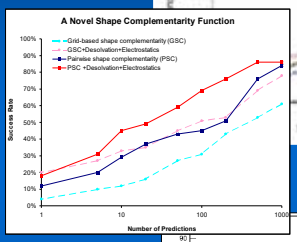
Fernandez-Recio et al. (2004) JMB 335, 843-865

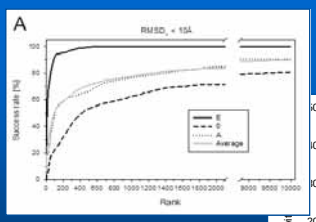
Protein-protein benchmarking

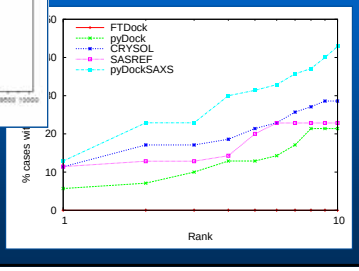
2002-2018 **Weng's benchmark series**

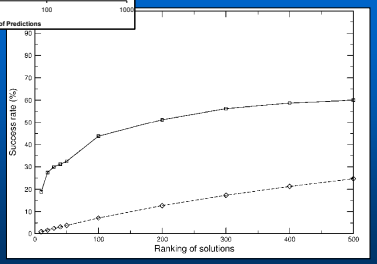
0.0: 54 cases
1.0: 59 cases
2.0: 84 cases
3.0: 124 cases
4.0: 176 cases
5.0: 230 cases





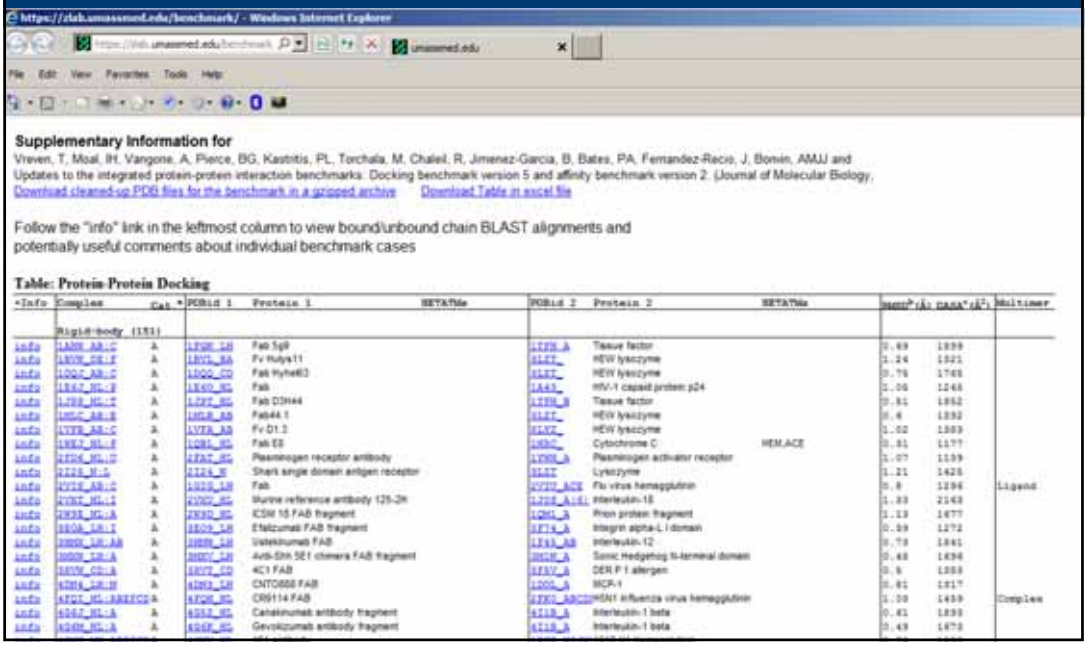






http://zlab.umassmed.edu/benchmark/

https://zlab.umassmed.edu/benchmark/



Supplementary Information for

Vreven, T, Moal, IH, Vangone, A, Pierce, BG, Kastritis, PL, Torchala, M, Chaleil, R, Jimenez-Garcia, B, Bates, PA, Fernandez-Recio, J, Bonvin, AMJJ and Updates to the integrated protein-protein interaction benchmarks. Docking benchmark version 5 and affinity benchmark version 2. (Journal of Molecular Biology,

Download cleaned-up PDB files for the benchmark in a gzipped archive Download Table in excel file

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

Table: Protein-Protein Docking

*Info	Complex	Cat *	PDBid 1	Protein 1	RETAINs	PDBid 2	Protein 2	RETAINs	Multimer
	Rigid-body (153)								
info	1ADW_AB:C	A	1FQK_LH	Fab 5g9		1TFH_A	Tissue factor		
info	1BVN_DE:F	A	1BVI_BA	Fv Hulya11		6LZT_	HEW lysozyme		
info	1DQJ_AB:C	A	1DQQ_CD	Fab Hyhel63		6LZT_	HEW lysozyme		
info	1E6J_HL:P	A	1E6O_HL	Fab		1A43_	HIV-1 capsid protein p24		
info	1JPS_HL:T	A	1JPT_HL	Fab D3H44		1TFH_B	Tissue factor		
info	1MLC_AB:E	A	1MLB_AB	Fab44.1		6LZT_	HEW lysozyme		
info	1VFB_AB:C	A	1VFA_AB	Fv D1.3		6LZT_	HEW lysozyme		
info	1WEJ_HL:F	A	1QBI_HL	Fab E8		1HRC_	Cytochrome C	HEM,ACE	
info	2FD6_HL:U	A	2FAT_HL	Plasminogen receptor antibody		1PWH_A	Plasminogen activator receptor		
info	2I25_N:P	A	2I24_H	Shark single domain antigen receptor		6LZT	Lysozyme		
info	2VIR_AB:C	A	1RUZ_LH	Fab		2VIU_ACE	Flu virus hemagglutinin		Ligand
info	2VXT_HL:I	A	2VXU_HL	Murine reference antibody 125-2H		1JPS_A:161	Interleukin-18		
info	2W9E_HL:A	A	2W9D_HL	ICSW 15 FAB fragment		1QM1_A	Prion protein fragment		
info	3EOA_LH:I	A	3EO9_LH	Efalizumab FAB fragment		1FT4_A	Integrin alpha-L I domain		
info	3HEE_LH:AB	A	3HEM_LH	Ustekinumab FAB		1F45_AB	Interleukin-12		
info	3M0N_IH:A	A	3M0Y_LH	Anti-Shh 5E1 chimera FAB fragment		3HCM_A	Sonic Hedgehog N-terminal domain		
info	3RTM_CD:A	A	3RYT_CD	4C1 FAB		1FSV_A	DER P 1 allergen		
info	4DN4_IH:W	A	4DN3_LH	CNTO888 FAB		1JOL_A	MCP-1		
info	4FQI_HL:ABEFCD	A	4FQH_HL	CR9114 FAB		4FNK_ABCDE	H5N1 influenza virus hemagglutinin		Complex
info	4G6J_HL:A	A	4G6J_HL	Canakinumab antibody fragment		4I1B_A	Interleukin-1 beta		
info	4G6M_HL:A	A	4G6F_HL	Gevokizumab antibody fragment		4I1B_A	Interleukin-1 beta		

Benchmark 5.0 - Weng

230 cases:

151 rigid-body
45 medium difficulty
34 difficult

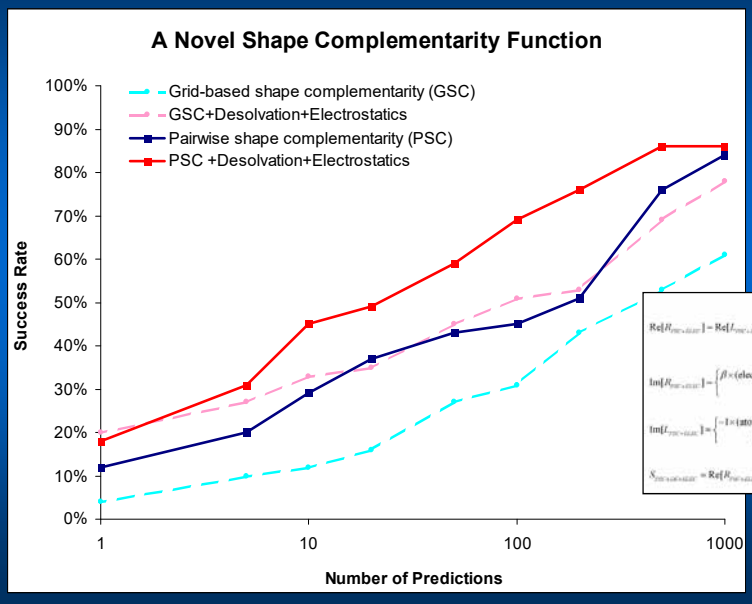
Rigid body:
I-RMSD $\leq$ 1.5 Å and $f_{\text{non-nat}} \leq 0.4$

Medium difficulty:
[1.5 Å < I-RMSD $\leq$ 2.2 Å] or [I-RMSD $\leq$ 1.5 Å and $f_{\text{non-nat}} > 0.4$]

Difficult:
I-RMSD > 2.2 Å

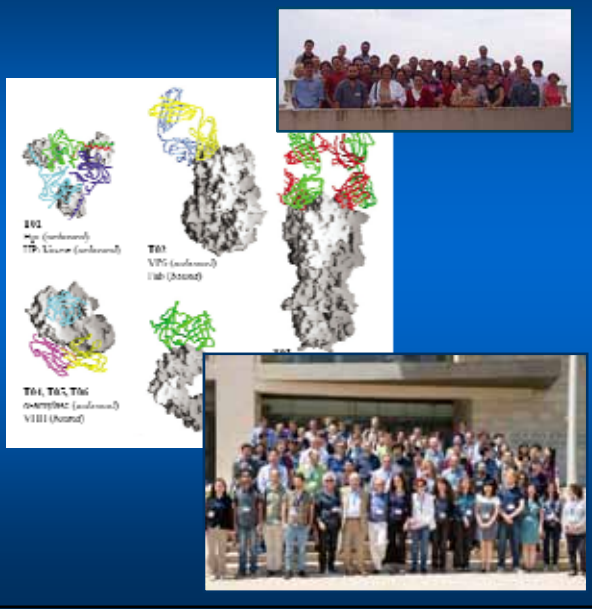
http://zlab.umassmed.edu/benchmark/

An example of benchmark use: ZDOCK performance



- Protein-protein docking benchmarks
- CAPRI
- Application to experimental cases

CAPRI: A Critical Assessment of PRedicted Interactions



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

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

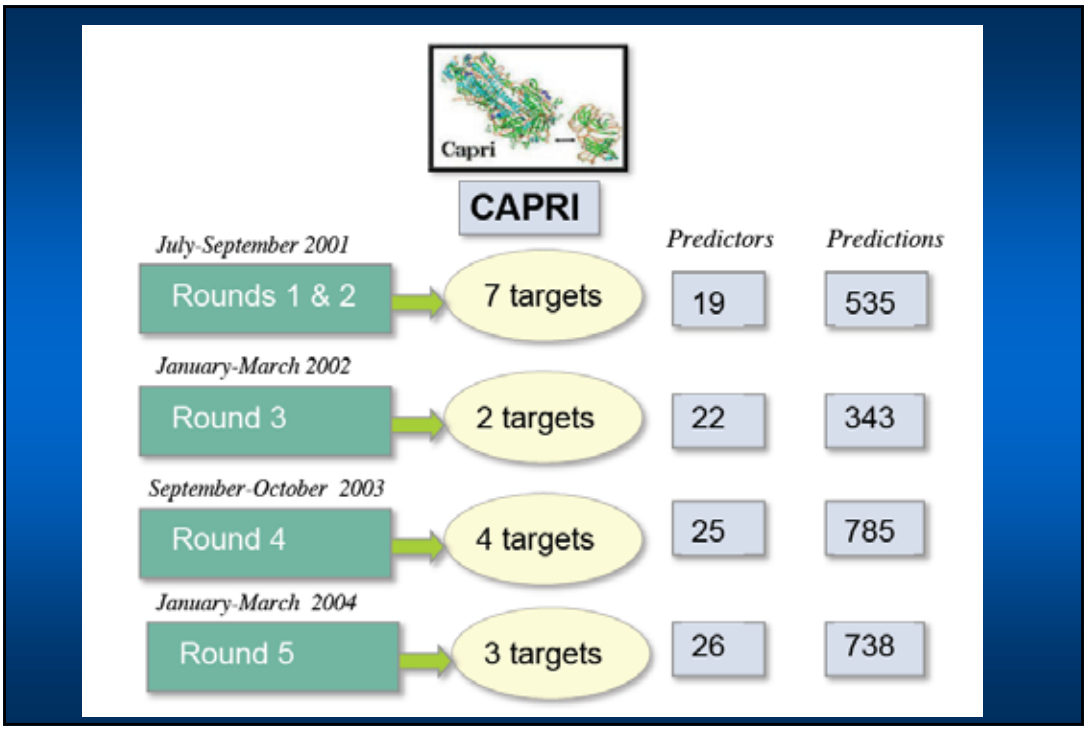
2nd CAPRI – Dec04 Gaeta (Italy)
PROTEINS: Structure, Function, and Bioinformatics 60 (July 2005)

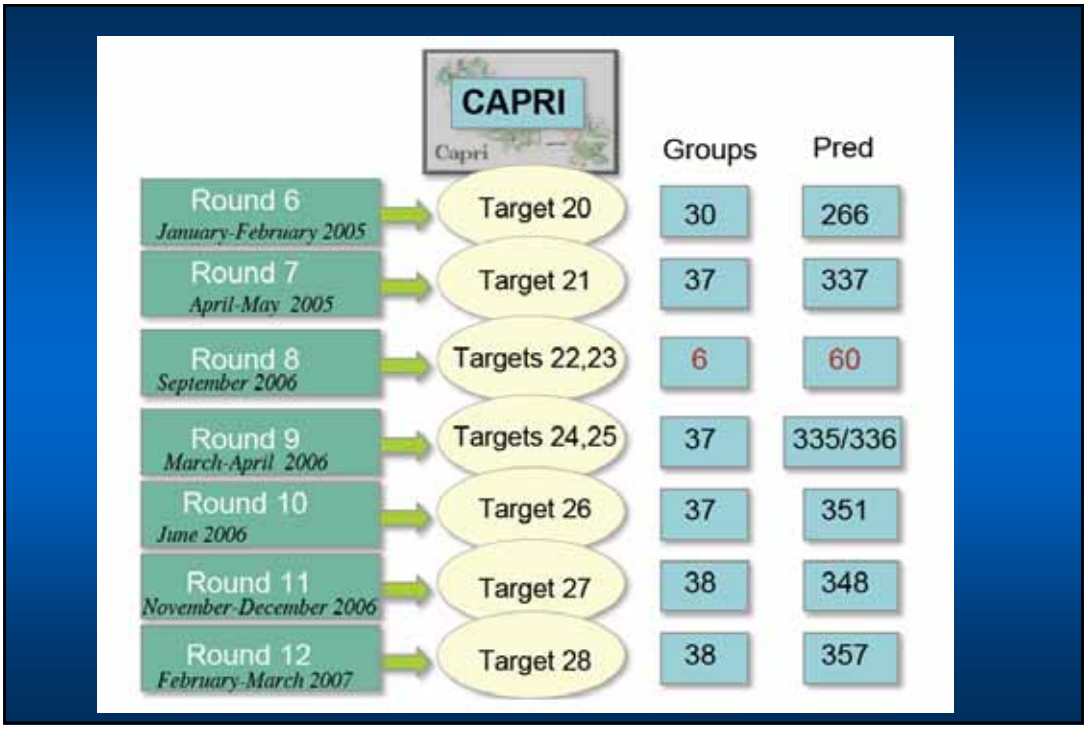
3rd CAPRI – Apr07 Toronto (Canada)
PROTEINS: Structure, Function, and Bioinformatics 69 (December 2007)

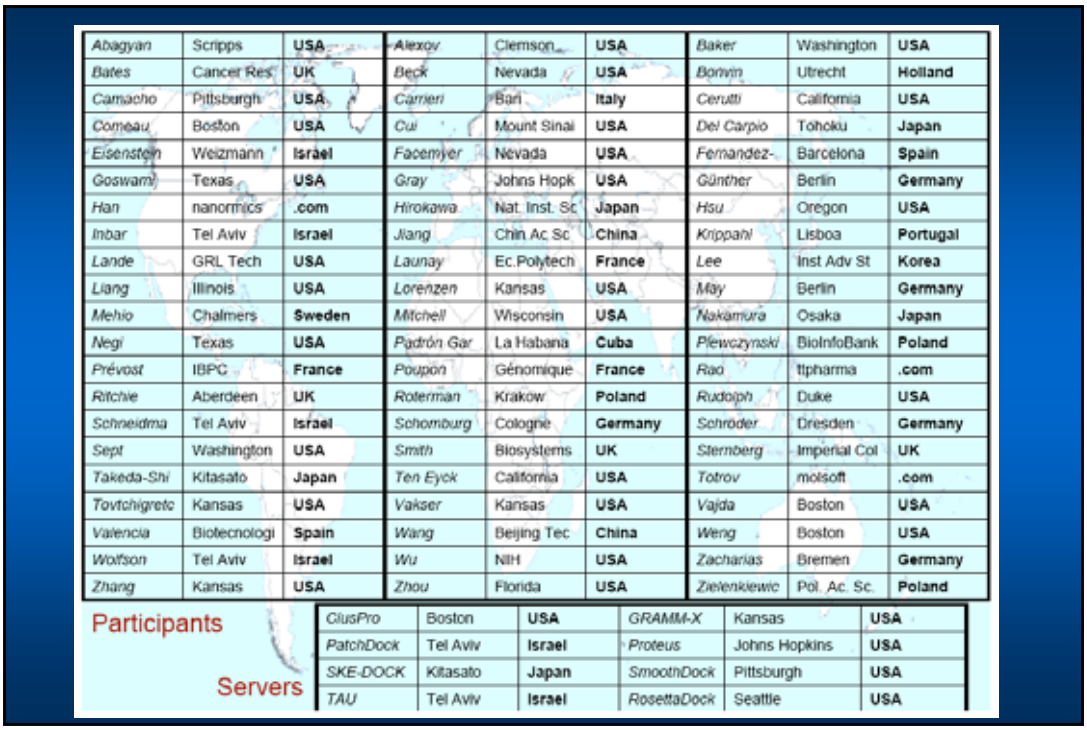
4th CAPRI – Dec10 Barcelona (Spain)
PROTEINS: Structure, Function, and Bioinformatics 78 (November 2010)

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

6th CAPRI – Apr16 Tel Aviv (Israel)


Capri
CAPRI
Predictors
Predictions
July-September 2001
Rounds 1 & 2
7 targets
19
535
January-March 2002
Round 3
2 targets
22
343
September-October 2003
Round 4
4 targets
25
785
January-March 2004
Round 5
3 targets
26
738


CAPRI
Capri
Groups
Pred
Round 6
January-February 2005
Target 20
30
266
Round 7
April-May 2005
Target 21
37
337
Round 8
September 2006
Targets 22,23
6
60
Round 9
March-April 2006
Targets 24,25
37
335/336
Round 10
June 2006
Target 26
37
351
Round 11
November-December 2006
Target 27
38
348
Round 12
February-March 2007
Target 28
38
357



Abagyan	Scripps	USA	Alexov	Clemson	USA	Baker	Washington	USA
Bates	Cancer Res	UK	Beck	Nevada	USA	Bonvin	Utrecht	Holland
Camacho	Pittsburgh	USA	Carneiri	Bari	Italy	Cerutti	California	USA
Comeau	Boston	USA	Cui	Mount Sinai	USA	Del Carpio	Tohoku	Japan
Eisenstein	Weizmann	Israel	Facemyer	Nevada	USA	Fernandez-	Barcelona	Spain
Goswami	Texas	USA	Gray	Johns Hopk	USA	Günther	Berlin	Germany
Han	nanormics	.com	Hirokawa	Nat. Inst. Sc	Japan	Hsu	Oregon	USA
Inbar	Tel Aviv	Israel	Jiang	Chin Ac Sc	China	Krippahl	Lisboa	Portugal
Lande	GRL Tech	USA	Launay	Ec.Polytech	France	Lee	Inst Adv St	Korea
Liang	Illinois	USA	Lorenzen	Kansas	USA	May	Berlin	Germany
Mehio	Chalmers	Sweden	Mitchell	Wisconsin	USA	Nakamura	Osaka	Japan
Negi	Texas	USA	Padrón Gar	La Habana	Cuba	Pilewczynski	BioInfoBank	Poland
Prévost	IBPC	France	Poupon	Génomique	France	Rao	ttpharma	.com
Ritchie	Aberdeen	UK	Roterman	Krakow	Poland	Rudolph	Duke	USA
Schneidma	Tel Aviv	Israel	Schomburg	Cologne	Germany	Schröder	Dresden	Germany
Sept	Washington	USA	Smith	Biosystems	UK	Sternberg	Imperial Col	UK
Takeda-Shi	Kitasato	Japan	Ten Eyck	California	USA	Totrov	molsoft	.com
Tovtchigrete	Kansas	USA	Vakser	Kansas	USA	Vajda	Boston	USA
Valencia	Biotecnologi	Spain	Wang	Beijing Tec	China	Weng	Boston	USA
Wolfson	Tel Aviv	Israel	Wu	NIH	USA	Zacharias	Bremen	Germany
Zhang	Kansas	USA	Zhou	Florida	USA	Zielenkiewic	Pol. Ac. Sc.	Poland

ClusPro	Boston	USA	GRAMM-X	Kansas	USA
PatchDock	Tel Aviv	Israel	Proteus	Johns Hopkins	USA
SKE-DOCK	Kitasato	Japan	SmoothDock	Pittsburgh	USA
TAU	Tel Aviv	Israel	RosettaDock	Seattle	USA

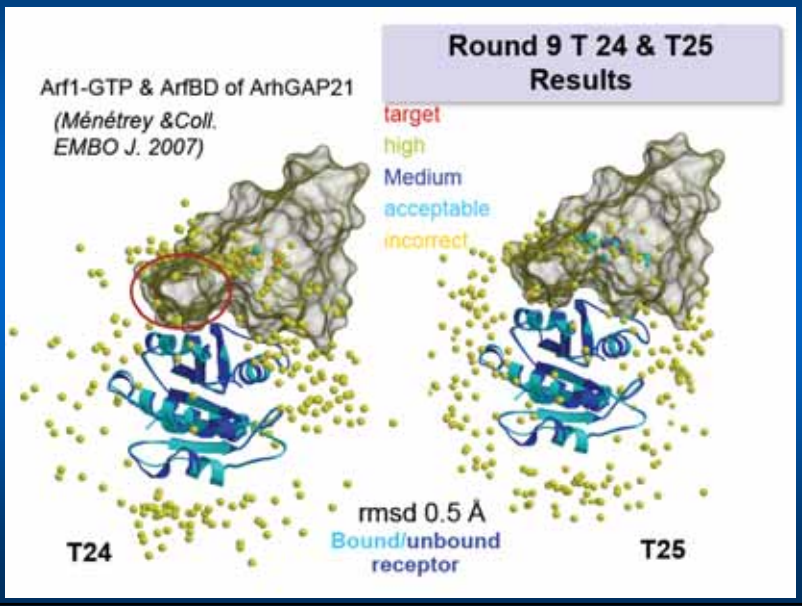
DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

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

DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

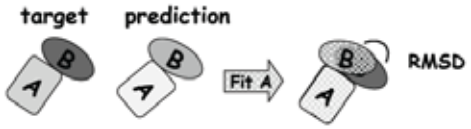


DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

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

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

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



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

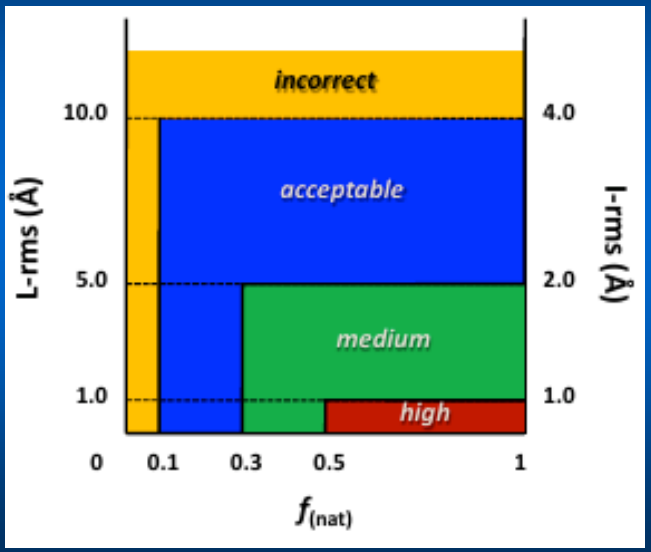
DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

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

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

DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

- $f_{(nat)}$ residue-residue 5 Å
- L-rms ligand residues
- l-rms interface residues 10 Å
- n_{clash} atom-atom 3 Å
- d_L
- ϑ_L



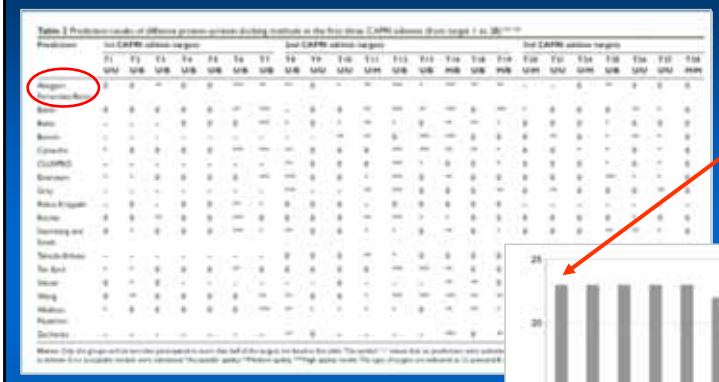
DOCKING VALIDATION
CAPRI: A Critical Assessment of PRedicted Interactions

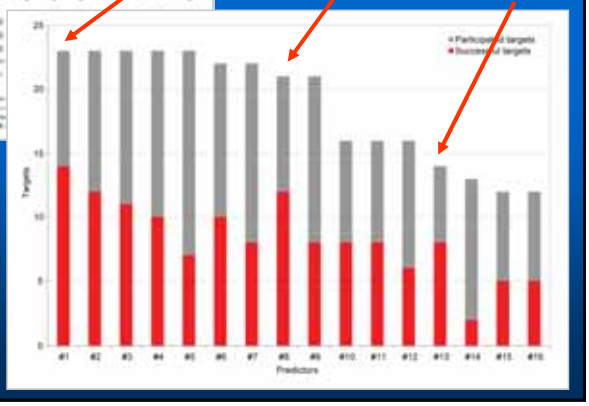
ROUND 9 T25 Results

Predictor groups	37
Evaluated predictions	336
High accuracy (***)	1

Model	Predictor	f_{nat}	$f_{n\text{-}nat}$	$f_{IR\text{-}R}$	$f_{IR\text{-}L}$	#Cl	$L\text{-}_{rmsd}$	$l\text{-}_{rmsd}$	Θ	D_L
10 H	Eisenstein	0.827	0.246	0.913	0.840	4	2.203	0.904	2.70	1.482
01 M	Schomburg	0.808	0.276	0.913	0.920	5	1.829	1.062	3.69	0.407
10 M	PATCHDOCK	0.692	0.357	0.913	0.960	19	2.334	1.166	6.28	1.139
01	GRAMM-X	0.827	0.295	0.913	0.960	10	2.824	1.246	8.94	1.294
03	Vajda	0.635	0.400	0.913	0.960	2	2.832	1.297	9.10	1.293
03	Fernandez-Recio	0.788	0.281	0.913	0.880	35	2.844	1.234	9.36	1.563
02	Totrov	0.692	0.122	0.783	0.960	2	3.018	1.246	10.70	1.643
01	Facemyer	0.788	0.211	0.957	0.920	21	3.153	1.415	14.03	1.151
07	SKE-DOCK	0.788	0.281	0.913	0.920	9	3.307	1.330	11.07	2.029
07	Takeda-Shitaka	0.808	0.323	0.913	1.000	8	3.324	1.335	11.08	2.059
07	Weng	0.808	0.364	0.913	0.960	6	3.804	1.506	13.01	2.450
02	Smith	0.673	0.426	0.913	0.960	23	3.929	1.510	9.27	3.085
01	SMOOTHDOCK	0.385	0.592	0.739	0.760	15	4.849	2.192	23.98	1.966
05	Negi	0.673	0.470	0.913	0.920	55	5.745	1.862	17.36	4.029
01 A	Bonvin	0.365	0.708	0.913	0.960	10	6.644	3.522	21.05	4.809
04 A	Camacho	0.308	0.765	0.826	0.840	7	7.533	4.088	34.68	4.202
06	CLUSPRO	0.654	0.477	0.739	0.840	12	7.564	2.845	34.45	4.634
05	Bates	0.442	0.681	0.957	0.880	17	9.825	3.178	24.42	8.412
	Unbound	0.923	0.094	0.957	0.960	0	0.282	1.134	0.65	0.109

CAPRI 1-3: summary of results

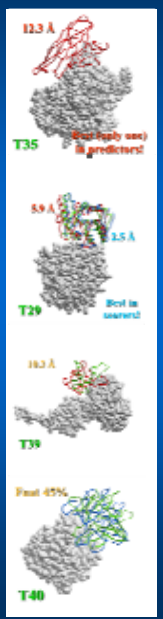




4th CAPRI

Predictor	Predictor Summary
Vajda	6/4***/2**
Zacharias	6/4***/1**
Zou	6/3***/2**
Eisenstein, Wolfson	6/3***/1**
Weng, Zhou	6/2***/2**
Bonvin	6/1***/4**
CLUSPRO	5/1***/3**
Fernandez-Recio	5/2**
Gray	4/2***/1**
Bates, Camacho, HADDOCK	4/1***/1**
Nakamura, Baker	3/2***/1**
Wang	3/1***/1**
Ritchie	3/3**
GRAMM-X, Takeda-Shitaka, Xiao	2/2***
Tovchigrechko, Ten Eyck, Vakser	2/1***/1**
SKE-DOCK	2/1***
Mitchell	2/2**
F_Jiang	2
Comeau, PATCHDOCK, FIBERDOCK, FIREDOCK	1/1***
Elofsson, TOPDOWN, Maigret, Leclerc, Alexov, Bajaj	1/1**
Smith, Elber, Kihara, Gunther, Kinoshita, Del Carpio	1
21 other groups	**0**

4th CAPRI – *pyDock results*



Domain assembly

T35

Homology

T36

T42

Dimer

Bound

T29
T39
T40

T37
T38

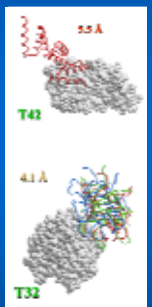
Unbound
T30
T32
T41

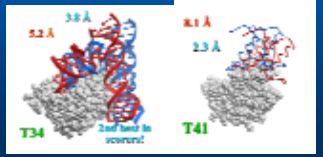
T34

T33

Protein-RNA

Success
(Almost) Success





5th CAPRI

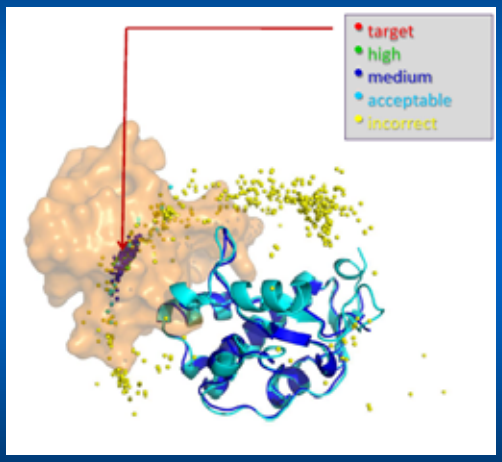
HADDOCK
SwarmDock
GRAMM/homol
PIPER
pyDock, SDU
ITScore+entropy
ATTRACT
CLUSPRO
MolFit
HADDOCK server, ZDOCK, RosettaDock
MLSBF
SwarmDock server
PatchDock/FireDock
...

Predictor Group – Per Target Data	Predictor Summary
Bonvin	9/1***/3**
Bates	8/2**
Vakser	7/1***
Vajda	6/2***/3**
Fernandez-Recio, Shen	6/1***/3**
Zou	6/1***/2**
Zacharias	6/1***
CLUSPRO	6/4**
Eisenstein, Grudinin	5/1***/2**
HADDOCK, Weng, Seok, Gray	4/1***/1**
Nakamura	4/1***
SWARMDOCK	4/1**
Wolfson	3/2**
DOCK_PIE, Zhou	3/1**
Cui	2/1***/1**
HEXSERVER, Luethy	2/1***
Mitchell, Elber	2/1**
FIBERDOCK, Ritchie, Wang	1***
Ten Eyck, Camacho, LZERD	1**
Pal, Zhang, SURFIT, Poupon, Kihara, Dokholyan	1*
26 other groups	0

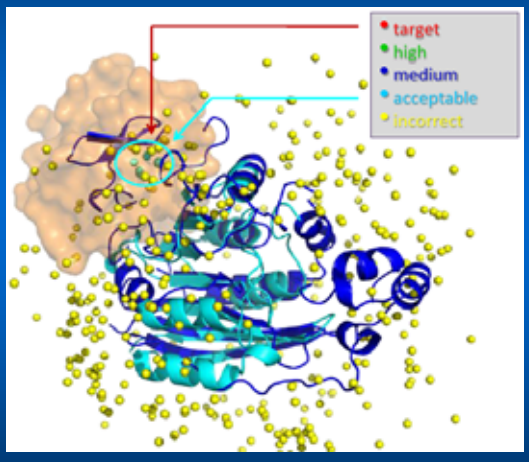
Total 62 groups

5th CAPRI

Easy target (T47)



A difficult target (T46)



5th CAPRI

Target	Area	RMS	Seq Identity	Summary	
T47	1700 Å^2	0.5 Å	75%	25/20***/5**	trivial
T53	1750 Å^2	0.5 Å	96%	20/1***/8**	easy
T50	1540 Å^2	0.8 Å	84%	14/1**	easy
T48/T49	1830 Å^2	0.9 Å	100%	14/6**	easy / average
T54	1440 Å^2	0.5 Å	100%	4	difficult
T46	1980 Å^2	3.2 Å	12%	2	difficult
T51	840 Å^2 / 1440 Å^2	3.4 Å	29%	2	difficult
T57	1020 Å^2	0.8 Å	100%	14/4**	easy
T58	2360 Å^2	0.8 Å	100%	11/5**	easy

5th CAPRI – *pyDock*

organizers' classification	Target	Type	Predictors	
			Quality	Submission Rank
Difficult	T46	H-H	-	-
Trivial	T47	H-U	***	1
Easy	T48	U-U	*	3
Easy	T49	U-U	*	4
Easy	T50	U-H	**	1
Difficult	T51	D-H-D	-	-
Easy	T53	U-H	**	3(1)
Difficult	T54	U-H	-	-
Easy	T58	U-U	**	5

67% success rate

6th CAPRI

PREDICTORS

Courtesy of Marc Lensink

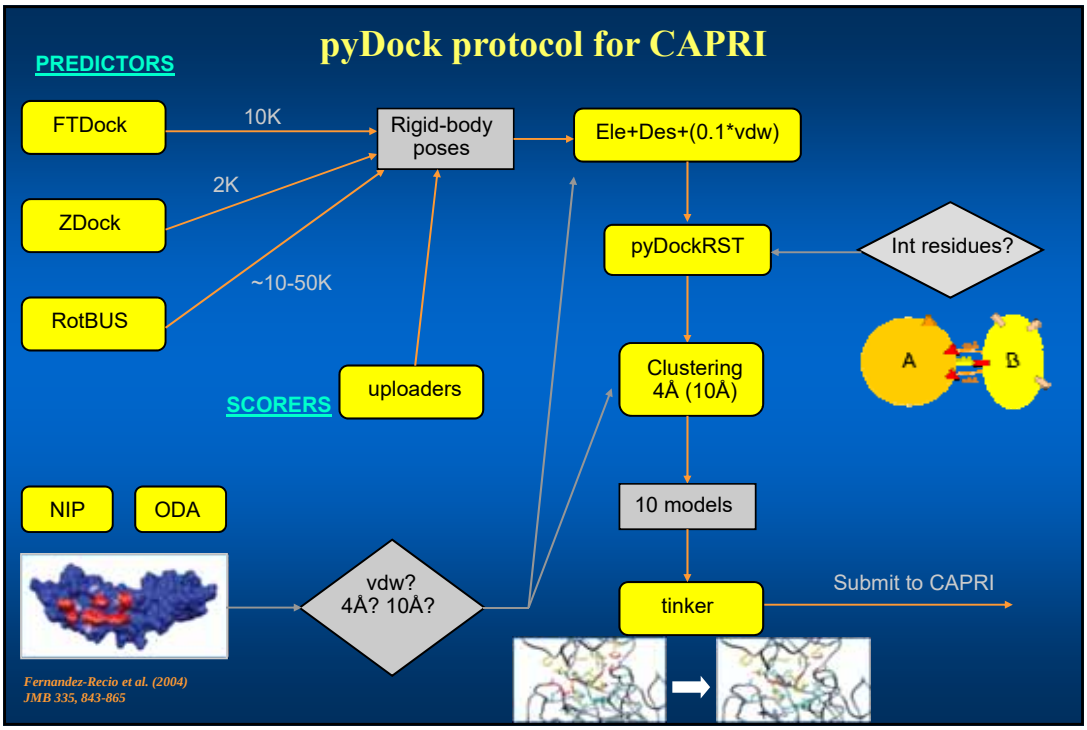
Predictor	Participation	Rank
Guerois	**18**	**10/1***/8****
Zacharias	16	**10/3***/2****
Vajda/Kozakov, Seok	17, **18**	**8/3***/2****
Weng	17	**6/1***/4****
Fernandez-Recio	**18**	**7/1***/3****
Vakser	**18**	**6/2***/2****
Eisenstein	9	**4/2***/2****
Zou	**18**	**7/1***/2****
Bates	**18**	**6/3****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

6th CAPRI

SERVERS

	Participation	Rank
ClusPro	17	**9/3****
PyDockWeb	13	**6/2****
LZerD	**18**	**4/1***/1****
HADDOCK	**18**	**4/2****
SwarmDock	**18**	**3/2****
GalaxyPPDock	10	**1****
PatchDock-FiberDock, DOCK/PIERR, MegaDock, GRAMM-X	2, 2, 4, 7	**1**
SurFit	1	**0**

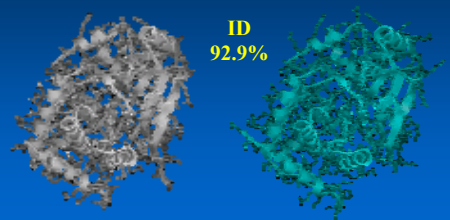
Courtesy of Marc Lensink



CAPRI

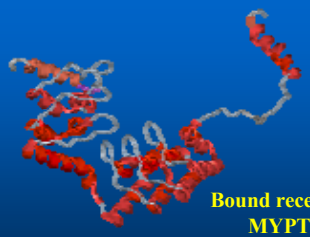
Some examples from ICM/pyDock

2nd CAPRI – *Target 14*



Template:
protein phosphatase 1α

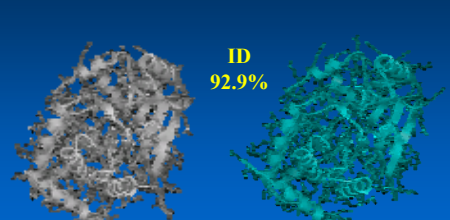
Modelled Ligand:
protein phosphatase 1β



Bound receptor:
MYPT1

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

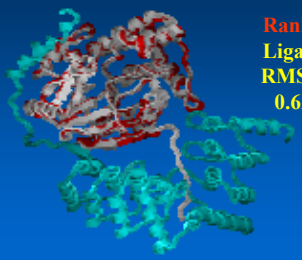
2nd CAPRI – *Target 14*



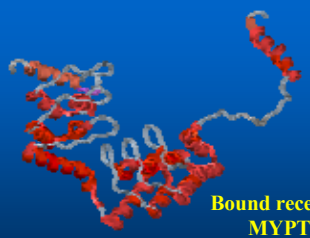
Template:
protein phosphatase 1α

Modelled Ligand:
protein phosphatase 1β

Docking



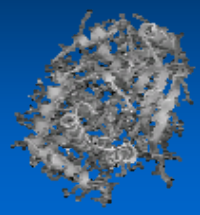
Rank 1
Ligand
RMSD:
0.6Å



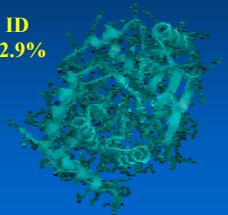
Bound receptor:
MYPT1

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

2nd CAPRI – *Target 14*



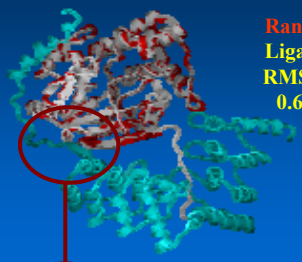
ID
92.9%



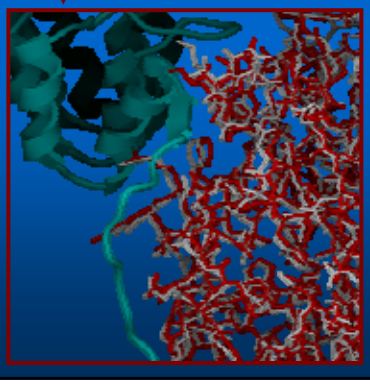
Template:
protein phosphatase 1α

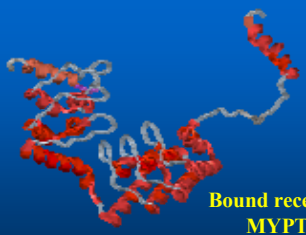
Modelled Ligand:
protein phosphatase 1β

Docking



Rank 1
Ligand
RMSD:
0.6Å

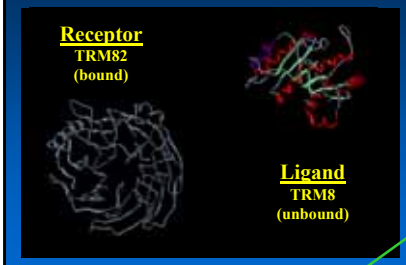




Bound receptor:
MYPT1

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

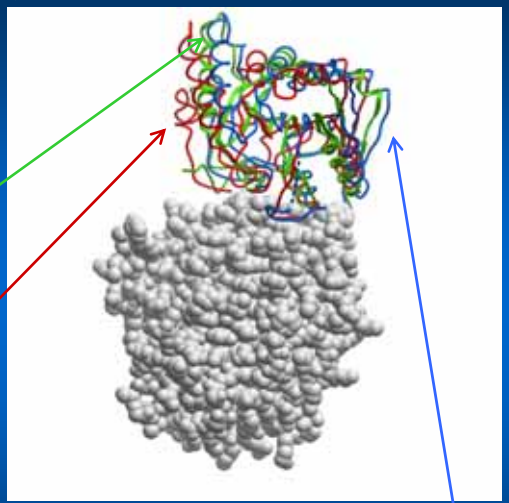
Target 29 (bound / unbound)



Receptor
TRM82
(bound)

Ligand
TRM8
(unbound)

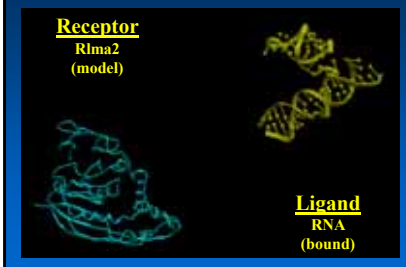
PDB complex 2VDU

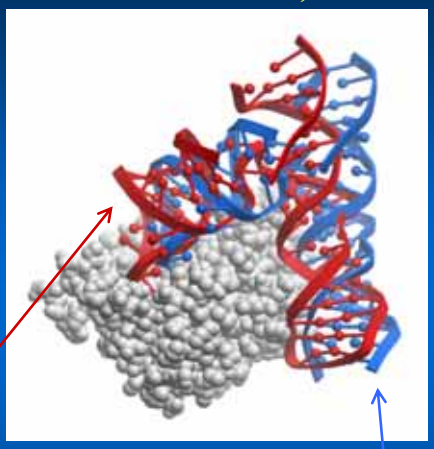


Predictors (total 41 groups):
0.1*vdw
PyDock Model 1 (** medium)
40% native contacts
Ligand RMSD: 5.9Å; Interface RMSD 1.9Å
[4th best of any group !!]
From zdock

Scorers (total 18 groups):
0.1*vdw
PyDock Model 4 (*** high) [best of any group !!]
71% native contacts
Ligand RMSD: 2.5Å; Interface RMSD 1.0Å

Targets 34 (protein model / RNA bound)

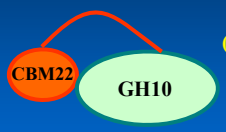




Predictors (total 44 groups):
FTDock & RotBUS
Filter: RNA-loop G745,G748,A749,A750,A752,G753 >83%
dist F167-G748<8Å
Sorted by RNA Restraints A705,A727,A730,G731,A734
[Douthwaite et al JMB (2008); Lebars et al JMB (2007)]
PyDock Model 3 (* acceptable)
22% native contacts
Ligand RMSD: 5.2Å; Interface RMSD 2.9Å
From RotBUS

Scorers (total 20 groups):
0.1*vdw
Energy+Restraints: RNA-loop G..GAA.AG
Filter: dist F167-G748<8Å
PyDock Model 2 (** medium) [2nd best of any group !!]
47% native contacts
Ligand RMSD: 3.8Å; Interface RMSD 1.9Å
(+ 6 * acceptable models, including model 1)

Target 35 (model / model)



Xylanase
(domain assembly)

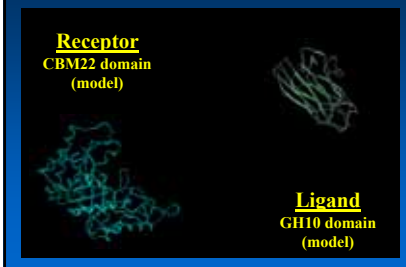
Flexible
linker

CBM22
domain

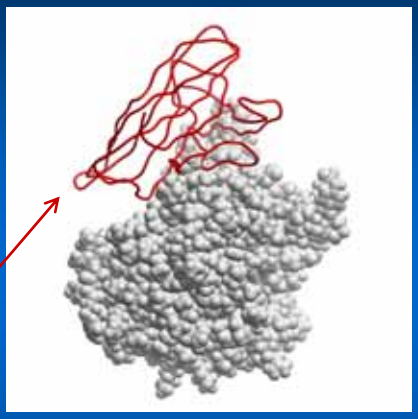
MGSSHHHHHHSSGLVPRGSHMDYEVVHDTFEVNFDGWCNLGVDTYL
TAVENEGNNGTRGMMVINRSSASDGAYSEKGFYLDGGVEYKYSVFV
KHNGTGTETFKLSVSYLDSETEEENKEVIATKDVVAGEWTEISAKY
KAPKTAVNITLSITTDSTVDFIFDDVTITRKGMAEANTVYLKDMYA
NYFRVGSVLNSGTVNNSSIKALILREFNSITCENEMKPDATLVQSG
STNTNIRVSLNRAASILNFCAQNNIAVRGHTLVWHSQTPQWFFKDN
FQDNGNWVSQSVMDQRLESYIKNMFAEIQRQYPSLNLYAYDVVNAA
VSDDANRTRYYGGAREPGYGNGRSPWVQIYGDNKFIEKAFTYARKY
APANCKLYYNDYNEYWDHKRDCIASICANLYNKGLLDGVGMQSHIN
ADMNGFSGIQNYKAALQKYINIGCDVQITELDISTENGKFSLQQQA
DKYKAVFQAAVDINRTSSKGKVTAVCVWGPNDANTWLGSQNAPLLF
NANNQPKPAYNAVASIIPQSEWGDGNNPA

GH10
domain

Target 35 (model / model)



PDB complex 2W5F
(not yet available)

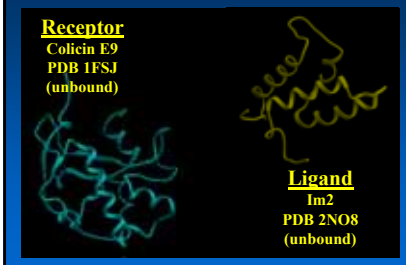


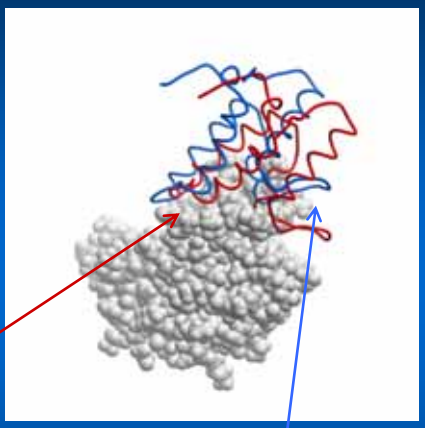
Predictors (total 44 groups):
CBM22 modeled on 1DYO_A; GH10 modeled on 1N82_A
0.1*VDW (needed to remove impossible linker distances)
pyDockTET
PyDock Model 1 (* acceptable) [best (and the only one) of any group]
14% native contacts
Ligand RMSD: 12.3Å; Interface RMSD 3.8Å
From ZDock

Scorers (total 20 groups):
0.1*VDW
pyDockTET
PyDock Model 1 (incorrect)
Ligand RMSD: 47.5Å
[only 1 acceptable from all groups]
Why did not we identify any correct model?

T36: bound/mod …why worse results?
Predictors: - [only 1 acceptable from all groups]
Scorers: N/A

Target 41 (unbound / unbound)

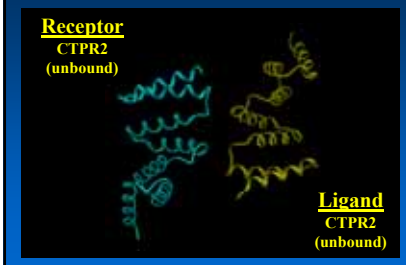


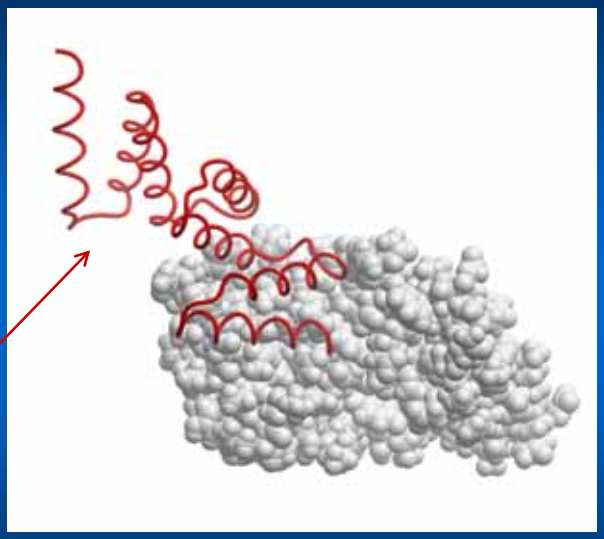


Predictors (total 33 groups):
Energy+Restraints (Im2 E30,V37,E41,S50,D51,Y54,Y55,P56; colicin E9 R54,N75,F86)
0.1*VDW (satisfied better restraints)
PyDock Model 4 (* acceptable)
61% native contacts
Ligand RMSD: 8.1Å; Interface RMSD 2.2Å
From ZDock
[similar results without restraints]

Scorers (total 13 groups):
Energy+Restraints (Im2 E30,V37,E41,S50,D51,Y54,Y55,P56; colicin E9 R54,N75,F86)
0.1*VDW (satisfied better restraints)
PyDock Model 9 (** medium)
68% native contacts
Ligand RMSD: 2.3Å; Interface RMSD 1.4Å

Target 42 (unbound dimer)

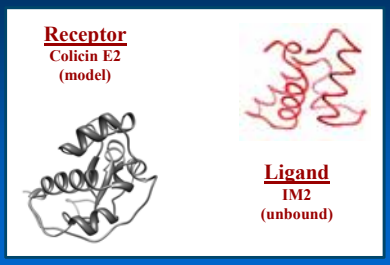


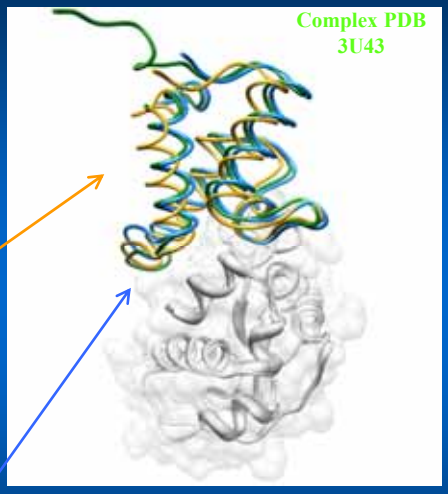


Predictors (total 28 groups):
Model based on PDB 1na0
Filter symmetry 180±10° (rise <5Å)
0.1*VDW (more symmetric solutions)
PyDock Model 1 (** medium)
39% native contacts
Ligand RMSD: 5.5Å; Interface RMSD 1.7Å
From ZDock
[similar results without symmetry filtering]

Target 47 (model / pseudo-unbound)

Regular prot-prot assessment



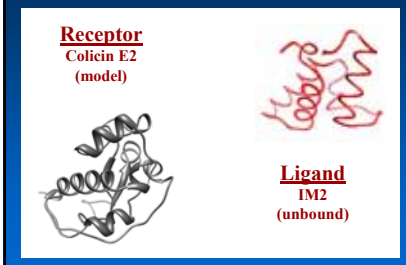


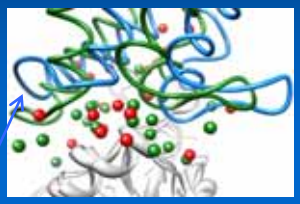
Predictors:
E2 modeled on 1EMV_B 85% SI (0.8Å CA-RMSD from bound)
IM2 2WPT_A in complex with homol E9
FTDock+Zdock+RotBUS / pyDockSER
pyDockRST: REC F85 – LIG Y54
pyDock Model 1 (*** high)
75% native contacts
Ligand RMSD: 2.5Å; Interface RMSD 0.8Å
[same results without restraints]
From Zdock

Scorers:
pyDockSER / pyDockRST: REC F85 – LIG Y54
pyDock Model 2 (*** high)
(Model 1: **) (Model 5: ***) [3rd best of all groups]
77% native contacts
Ligand RMSD: 0.9Å; Interface RMSD 0.4Å
From uploader Weng

Target 47 (model / pseudo-unbound)

Water prediction





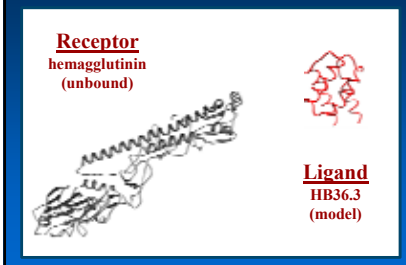
Complex PDB
3U43

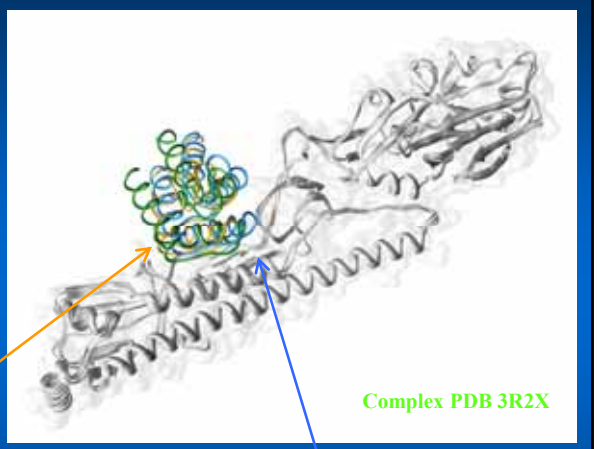
Predictors:
E2 modeled on 1EMV_B 85% SI (0.8Å CA-RMSD from bound)
IM2 2WPT_A in complex with homol E9
FTDock+Zdock+RotBUS / pyDockSER
pyDockRST: REC F85 – LIG Y54

DOWSER
pyDock Model 1 (+ fair)
20% water native contacts
From Zdock

Scorers:
pyDockSER / pyDockRST: REC F85 – LIG Y54
pyDock Model 8 (++ good)
43% water native contacts
From uploader Bates

Target 50 (unbound / model)



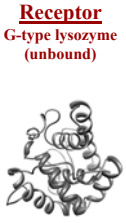


Predictors:
HB36.3 modeled on 1U84 83% SI (0.7Å CA-RMSD from bound)
FTDock (>92K)+Zdock / pyDockSER
pyDock Model 1 (** medium)
[model 4 **, best ligRMSD of all groups]
47% native contacts
Ligand RMSD: 6.1Å; Interface RMSD 1.8Å
From FTDock

Scorers:
pyDockSER
pyDock Model 4 (** medium)
45% native contacts
Ligand RMSD: 4.7Å; Interface RMSD 1.9Å
From uploader Bonvin

Target 58 (unbound / unbound)

Receptor
G-type lysozyme
(unbound)



Ligand
Lysozyme inhibitor
(unbound)

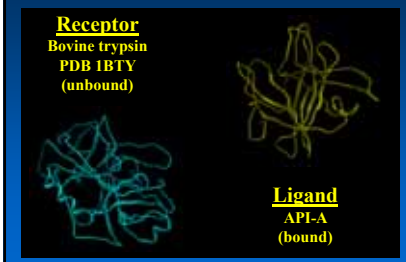
Complex PDB
4G9S

Predictors:
FTDock+Zdock+SwarmDock / pyDockSER
pyDockSAXS
pyDockRST: REC E73,D86,D97; LIG R119,Y47
pyDock Model 5 (** medium) [4th best model of all groups]
43% native contacts
Ligand RMSD: 4.9Å; Interface RMSD 1.8Å
[without SAXS: Model 10 (** medium)]
From SwarmDock

SAXS slightly helped

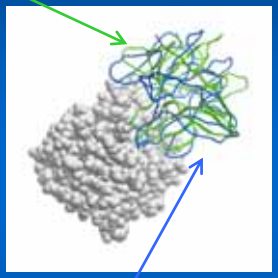
Target 40 (unbound / bound)

Receptor
Bovine trypsin
PDB 1BTY
(unbound)



Ligand
API-A
(bound)

PDB complex 3E8L



Predictors (total 38 groups):
0.1*vdw
Energy+Restraints (inhibitor L87 AND K145)
PyDock Model 2 (incorrect)
Ligand RMSD: 12.5Å

Incorrect restraints !!!

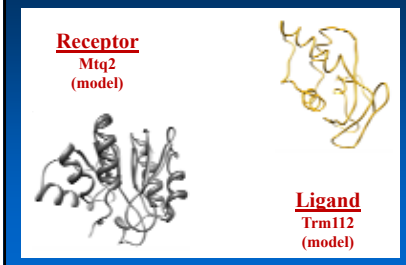
Energy+Restraints: inhibitor L87 OR K145
PyDock Model 8 (* acceptable?)

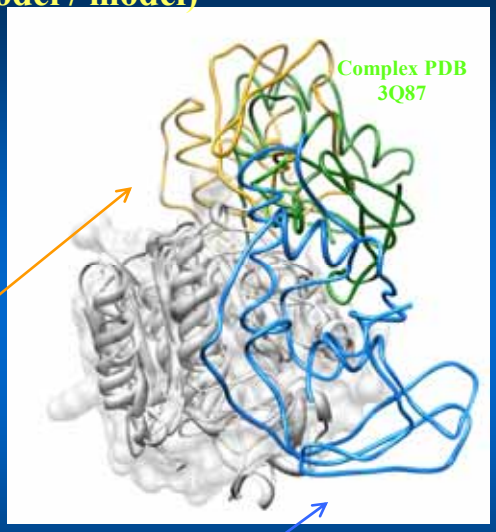
Scorers (total 14 groups):
0.1*vdw
Energy+Restraints (inhibitor L87 AND K145)
PyDock Model 6 (incorrect)
Ligand RMSD: 12.5Å

Incorrect restraints !!!

Energy+Restraints: inhibitor L87 OR K145
PyDock Model 1 (* acceptable)

Target 46 (model / model)





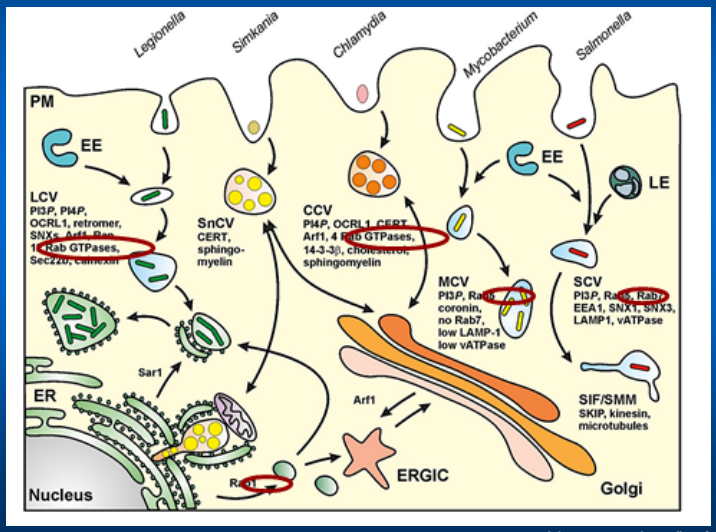
Predictors:
Mtq2 modeled on 1T43 28% SI (6.8Å CA-RMSD from bound)
Trm112 modeled on 2J6A 36% SI (7.3Å CA-RMSD from bound)
FTDock+Zdock+RotBUS / pyDockSER
pyDock Model 6 (incorrect)
Ligand RMSD: 31.2 Å

Scorers:
pyDockSER
pyDock Model 8 (incorrect)
Ligand RMSD: 28.0 Å

- Protein-protein docking benchmarks
- CAPRI
- Application to experimental cases

A Model of the SidD / Rab1 Interaction

Host/pathogen interactions: strategies of pathogens to invade host cells

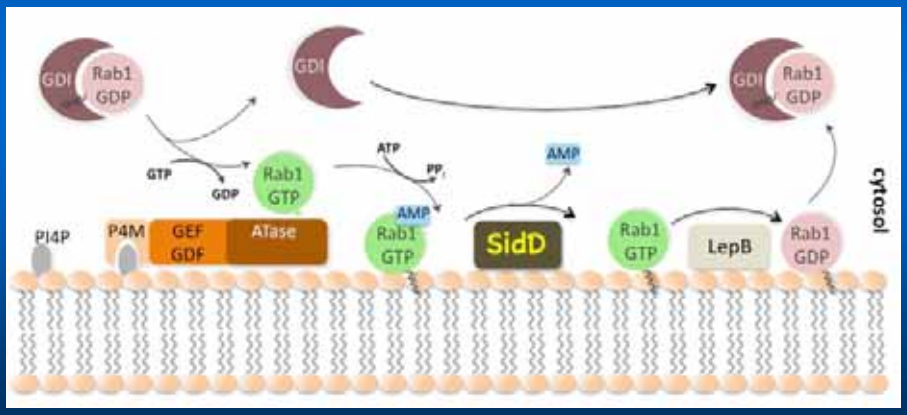


Herweg et al. *Cell. Infect. Microbiol.* (2015)

A Model of the SidD / Rab1 Interaction

AMPylation is a mechanism of microbial pathogens to manipulate host cell proteins

Legionella SidD catalyzes de-AMPylation of Rab1, a regulatory host cell protein, key for endocytosis

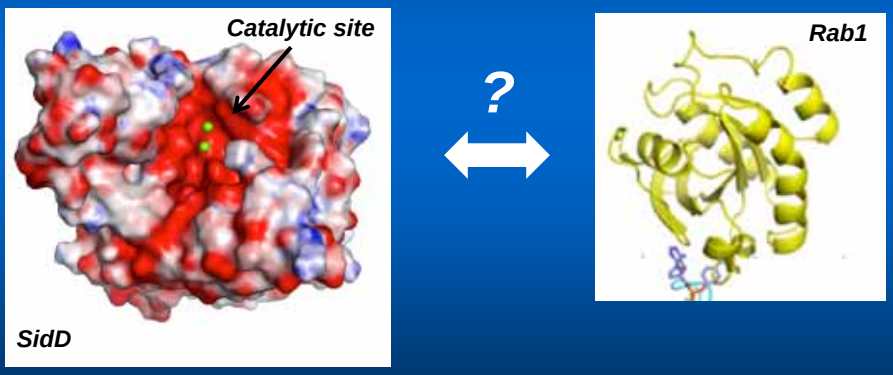


Henzler-Wildman K et al. Nature (2007)

A Model of the SidD / Rab1 Interaction

AMPylation is a mechanism of microbial pathogens to manipulate host proteins

SidD catalyzes de-AMPylation of Rab1, a regulatory protein

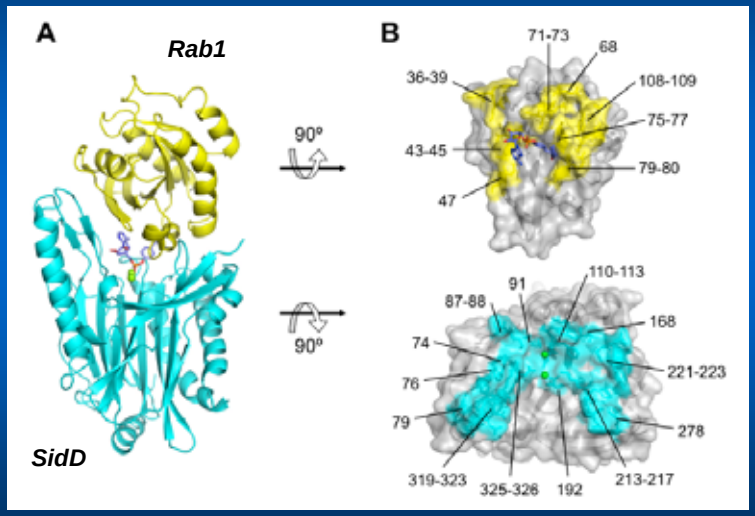


New X-ray structure

Chen et al. (2013) PLOS Pathogens 9, e1003382

A Model of the SidD / Rab1 Interaction

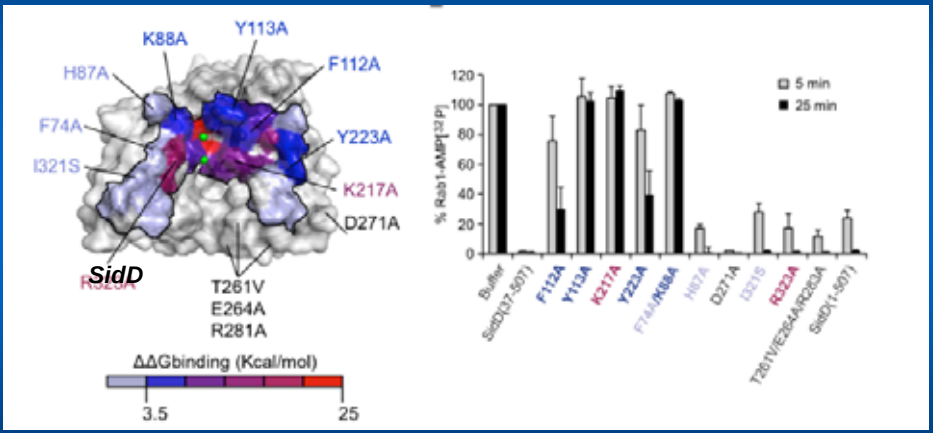
Docking model suggests interface residues



Chen et al. (2013) PLOS Pathogens 9, e1003382

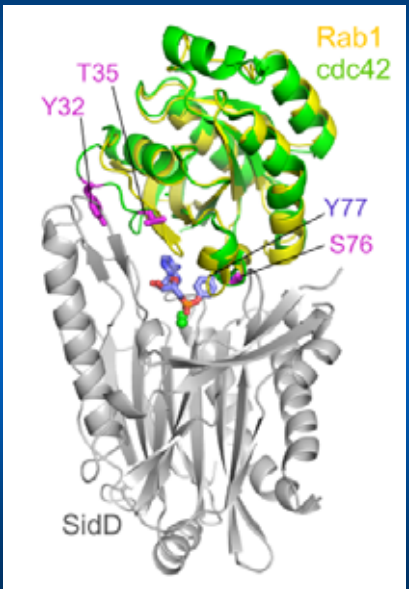
A Model of the SidD / Rab1 Interaction

Mutation experiments confirms docking model



Chen et al. (2013) PLOS Pathogens 9, e1003382

A Model of the SidD / Rab1 Interaction



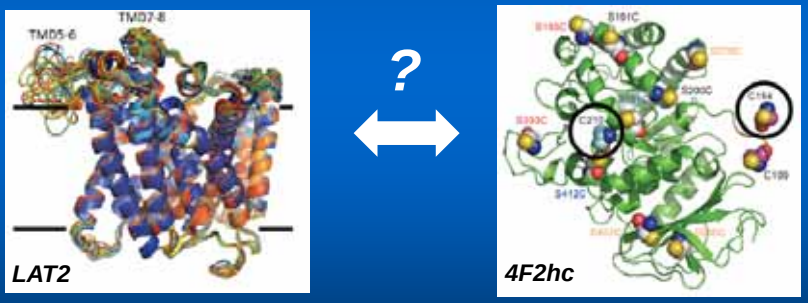
First structural model for de-AMPylation by SidD

Chen et al. (2013) PLOS Pathogens 9, e1003382

A Model of the 4F2hc / LAT2 Interaction

Heteromeric Aminoacid Transporters (HATs) are involved in amino acidurias, tumor growth and invasion, viral infection...

HAT proteins are composed of two domains (LAT2 and 4F2hc), but the structural details of their interaction were not known

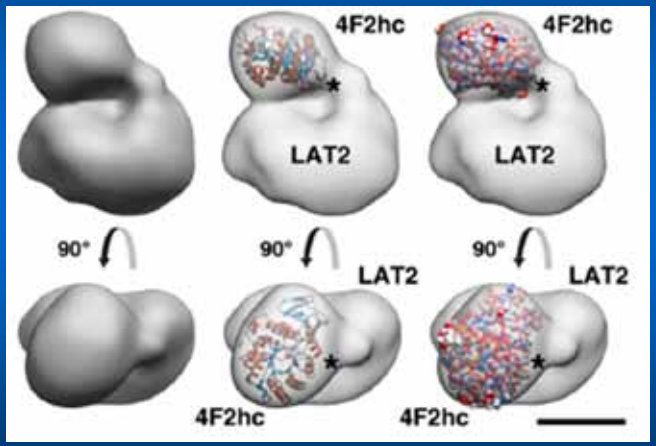


Homology-based models *X-ray structure*

Rosell et al. (2014) PNAS

A Model of the 4F2hc / LAT2 Interaction

Electron Microscopy indicates that 4F2hc is positioned over LAT2

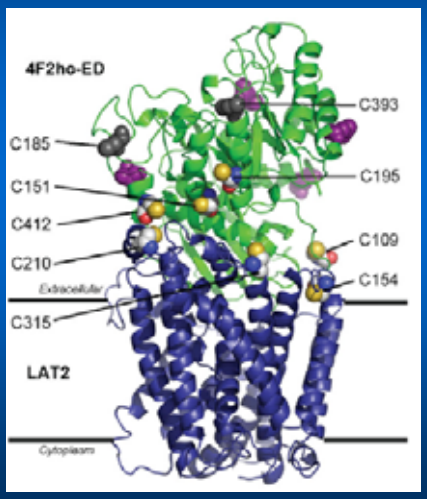


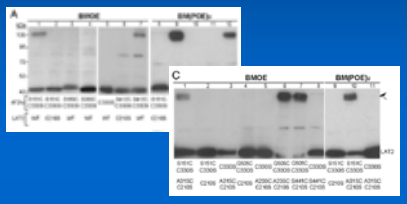
Rosell et al. (2014) PNAS

A Model of the 4F2hc / LAT2 Interaction

20 LAT2 models x 4F2hc: 200,000 docking models

Lowest-energy docking model is confirmed by cross-linking experiments

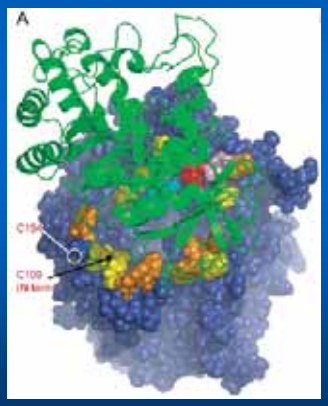




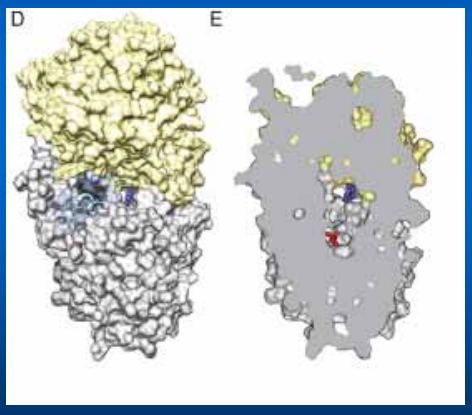
Rosell et al. (2014) PNAS

A Model of the 4F2hc / LAT2 Interaction

NIP predictions show a hydrophobic binding ring highly conserved in other transporters



The model explains how substrate can enter to the channel



Rosell et al. (2014) PNAS

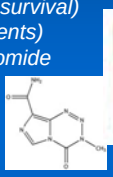
Modeling HAT family for drug discovery

Discovery of new inhibitors of the glutamate transporter xCT for the treatment of human gliomas

Fundació La Marató de TV3

Human Brain Gliomas

- *Poor prognosis (for high-grade gliomas, 1 year survival)*
- *Glioma-associated epilepsy seizures (80% patients)*
- *Current treatments: surgery, radiation, temozolomide*



Glioma-related seizures and tumor growth: glutamate is the key

Reported inhibitor: Sulfasalazine
(successfully reduced epilepsy seizures **in mice**)

Sulfasalazine as drug for human gliomas:

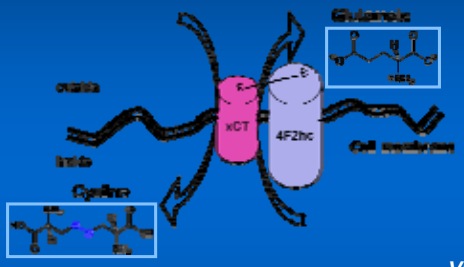
- no reports of any antiepileptic activity
- not specific
- no benefit (phase I study)
- short biological half-life
- CNS toxicity

Modeling HAT family for drug discovery

Discovery of new inhibitors of the glutamate transporter xCT for the treatment of human gliomas

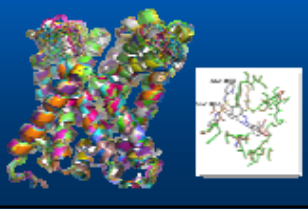
Fundació La Marató de TV3

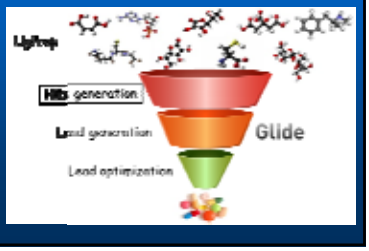
Glutamate/Cystine exchanger in glioma cells



Virtual screening

Structural models of xCT





Conclusions

- best docking programs: HADDOCK, SwarmDock, GRAMM/homology, PIPER, pyDock, SDU...
- how to introduce biological information is important
- difficult cases are: flexible ones, with modelled subunits...
- docking models are increasingly important to guide and interpret experimental results