

Predicting Specificity in Protein-Protein Interactions

Marc F. Lensink
Interdisciplinary Research Institute
CNRS USR3078 – Villeneuve d'Ascq

Genome

collection of hereditary information

Genomics

Transcriptome

collection of genes expressed under environmental conditions of the cell

Transcriptomics
(Expression Profiling)

Proteome

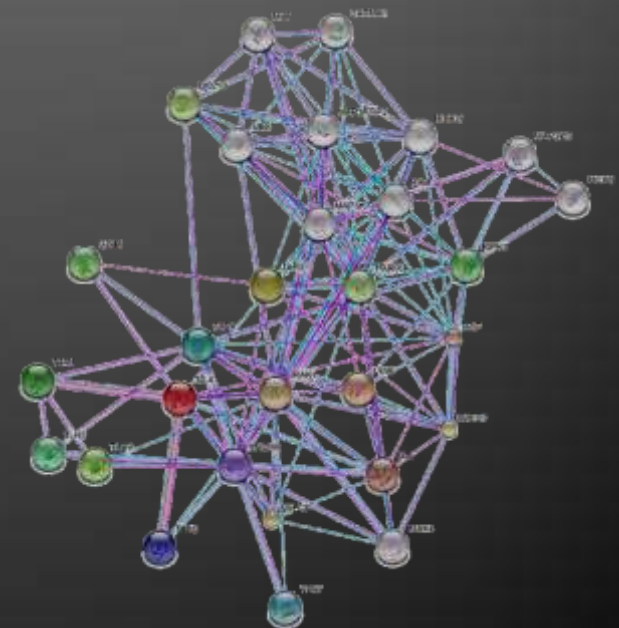
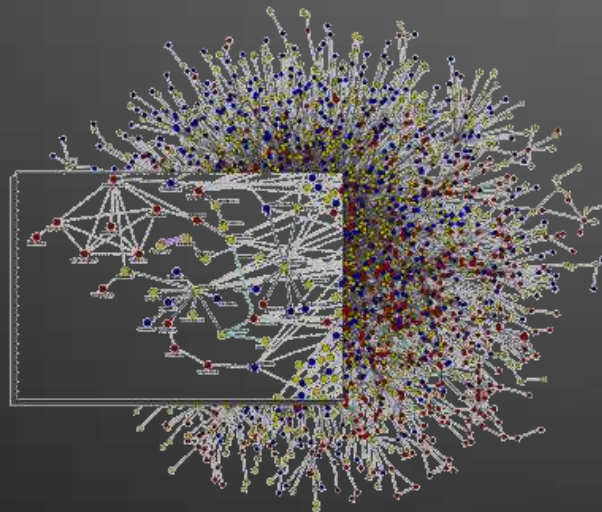
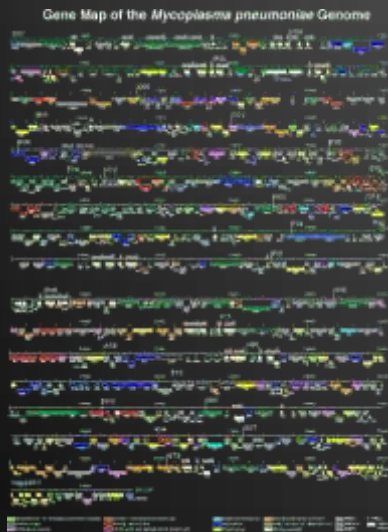
collection of expressed proteins

Proteomics

Interactome

whole set of molecular interactions

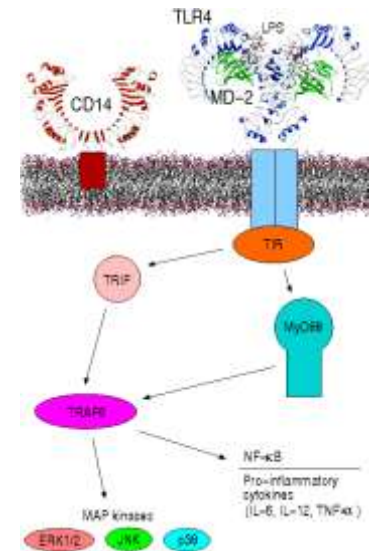
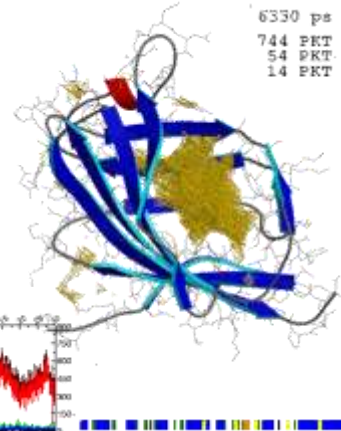
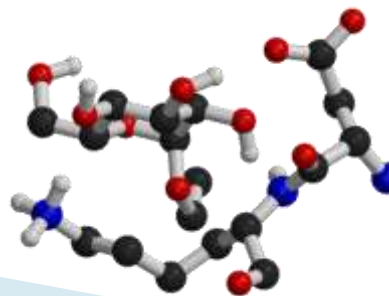
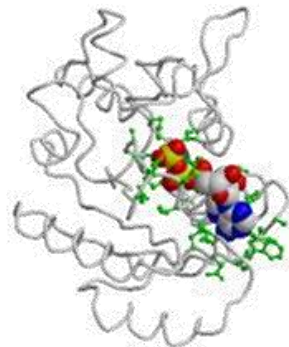
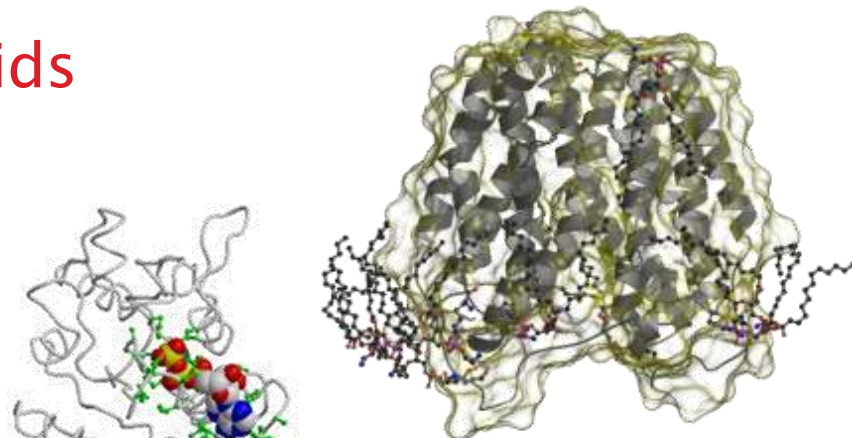
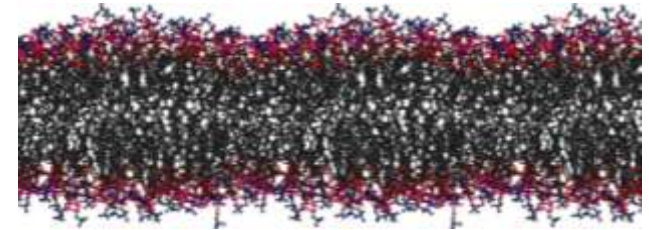
Interactomics
(Systems Biology)



Molecular interactions

► Molecuome:

- Proteins
- Nucleic acids
- Lipids
- Ions
- Metals
- Sugars

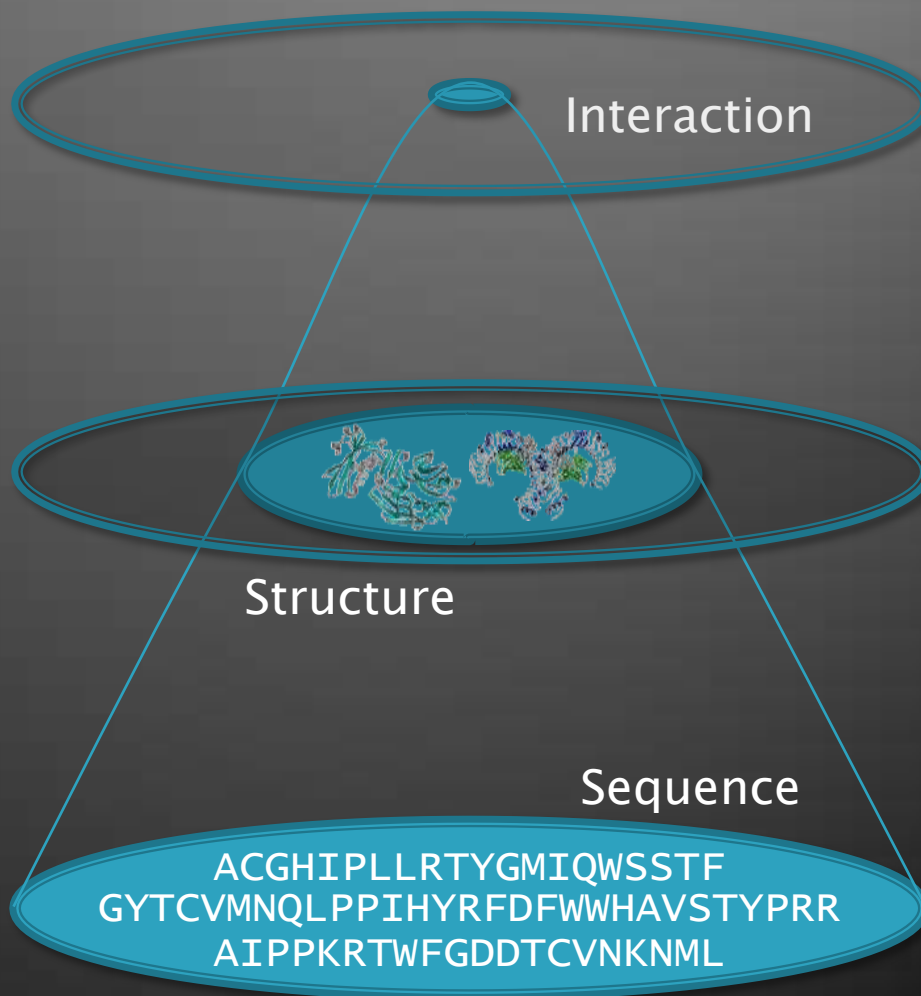


Use for computational docking

Space of individual structures is being progressively filled

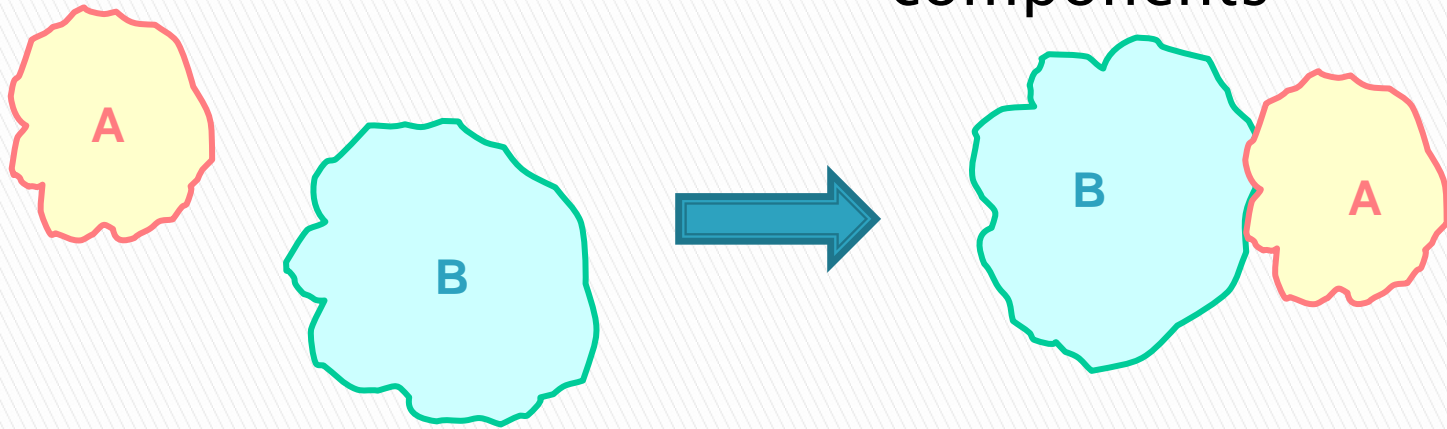
Sampling of the interactome remains sparse

Needs to clarify how proteins interact



Docking procedures

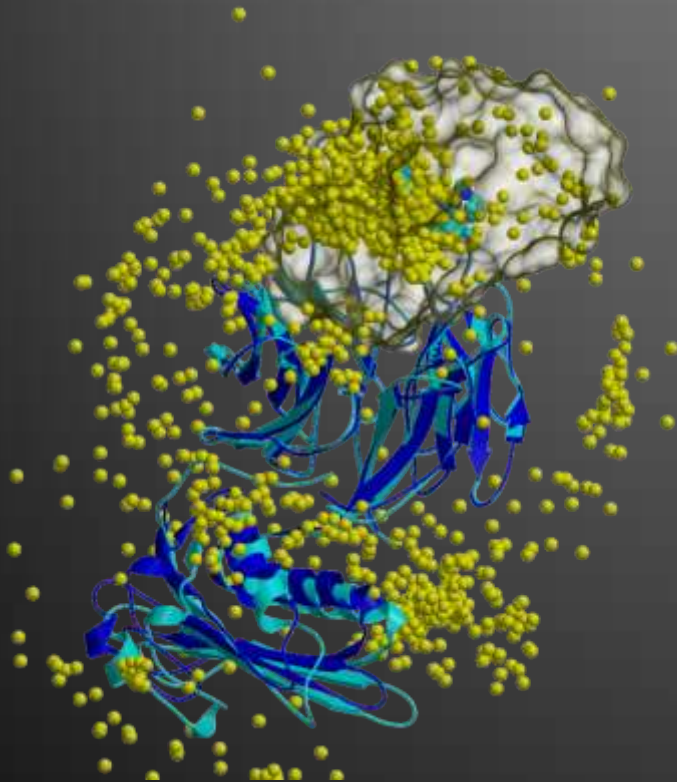
- ▶ Assemblies for which the individual components exist in the free form
- ▶ Produces an atomic model of the complex from the 3D coordinates of these components



Protein complexes

Docking

Critical Assessment of Predicted Interactions

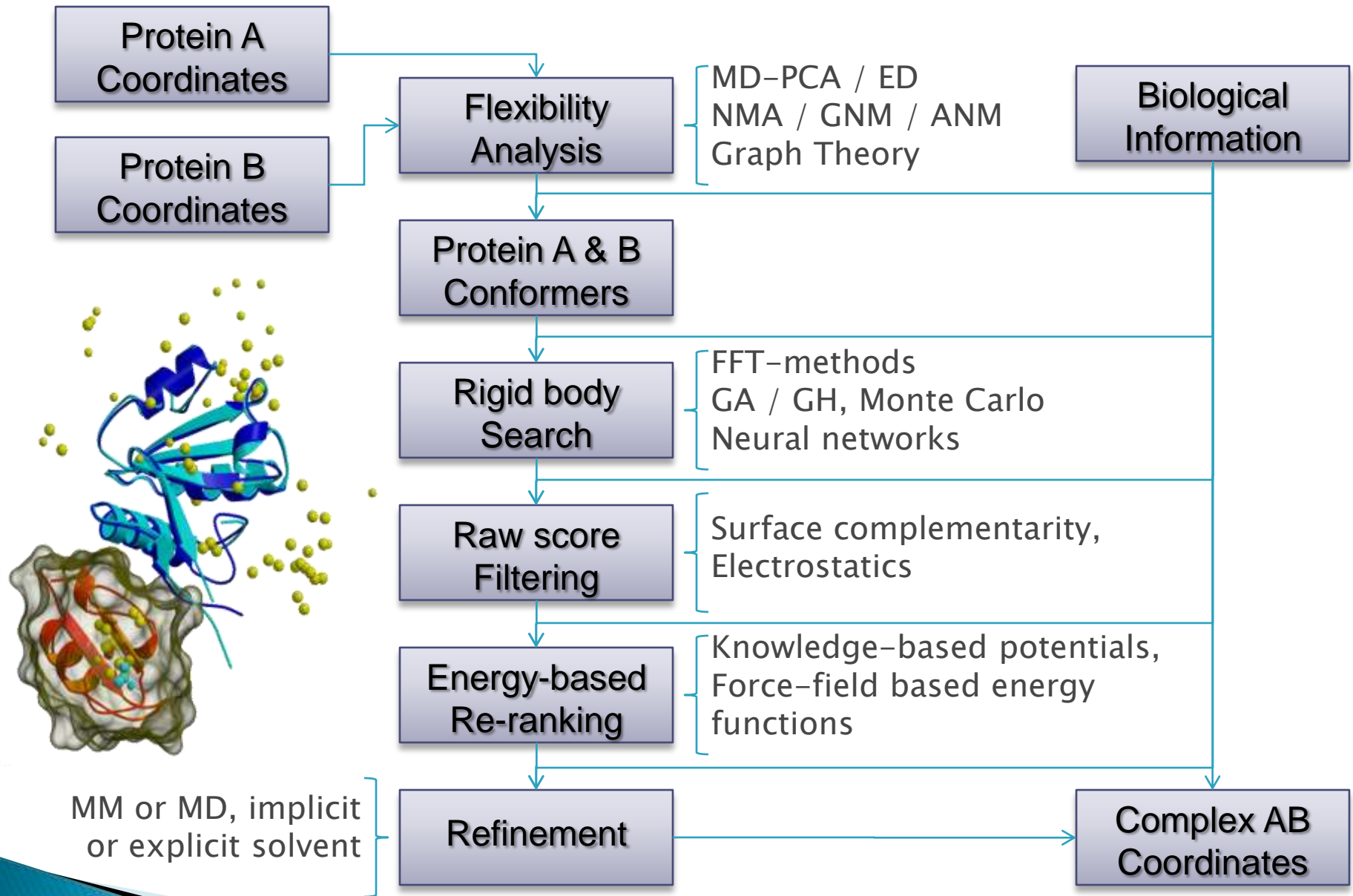


- ▶ Improvement of protein–protein interaction models
 - computational methods
 - unbound components
 - or homology models
- ▶ Project started in 2001
 - 49 targets
 - 90+ groups
 - 12+ servers
 - Scoring experiment

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

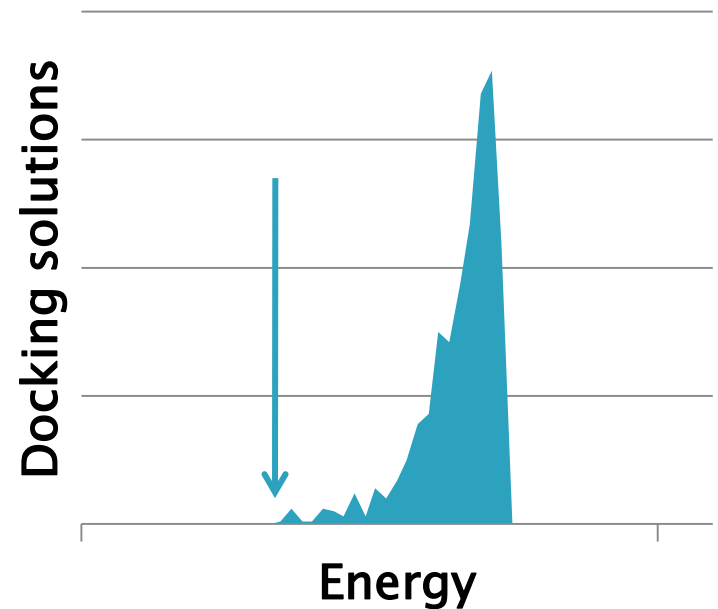
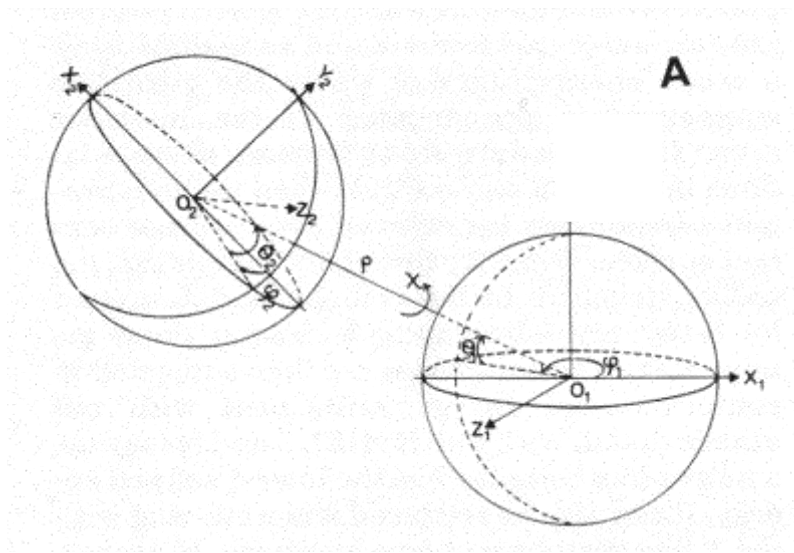
Participants (2007) Servers (2009)

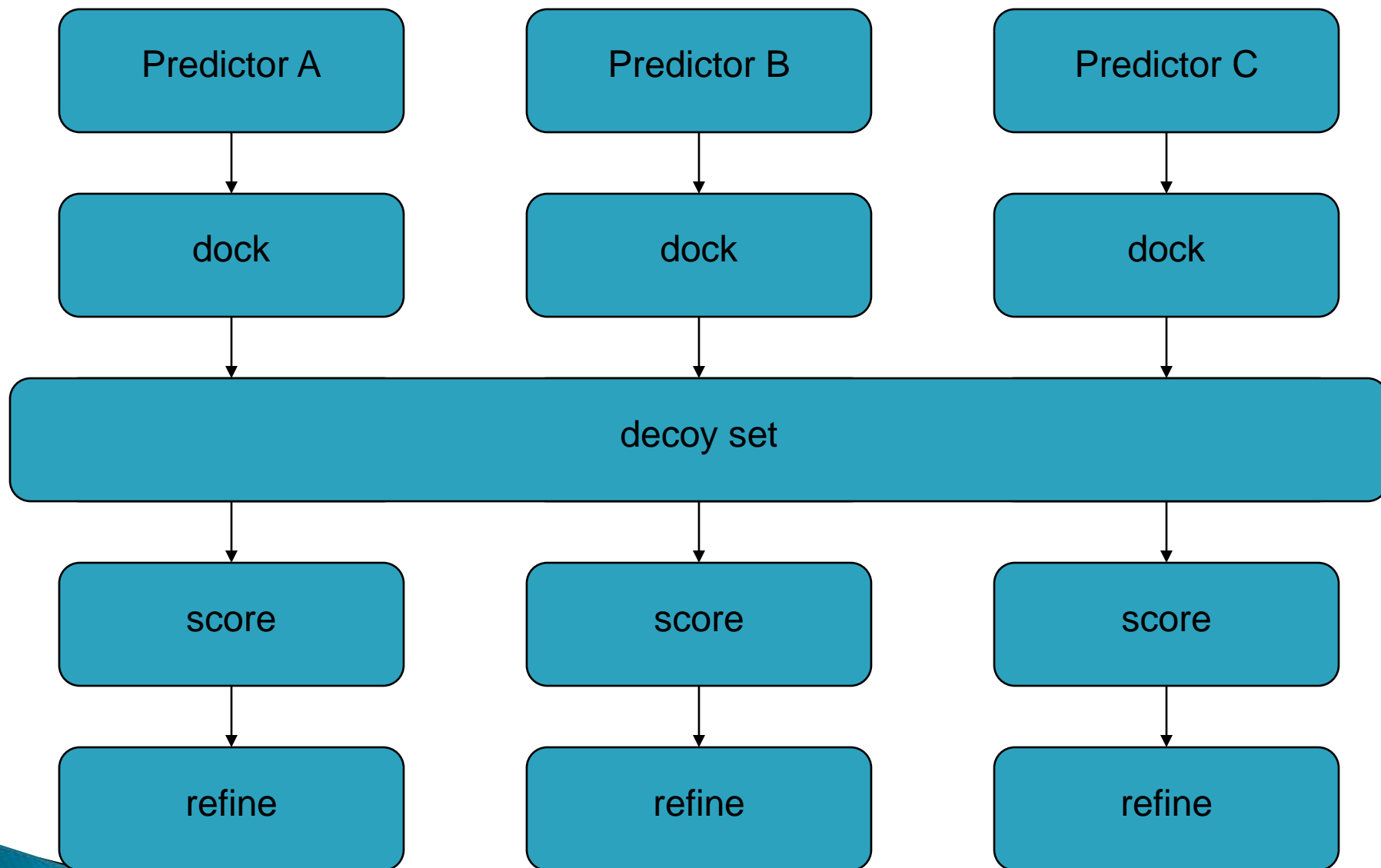
3D-Garden	ClusPro	FiberDock	FireDock
GRAMM-X	HADDOCK	HEX	PatchDock
SAMSON+HEX	SKE-DOCK	SmoothDock	TOP DOWN

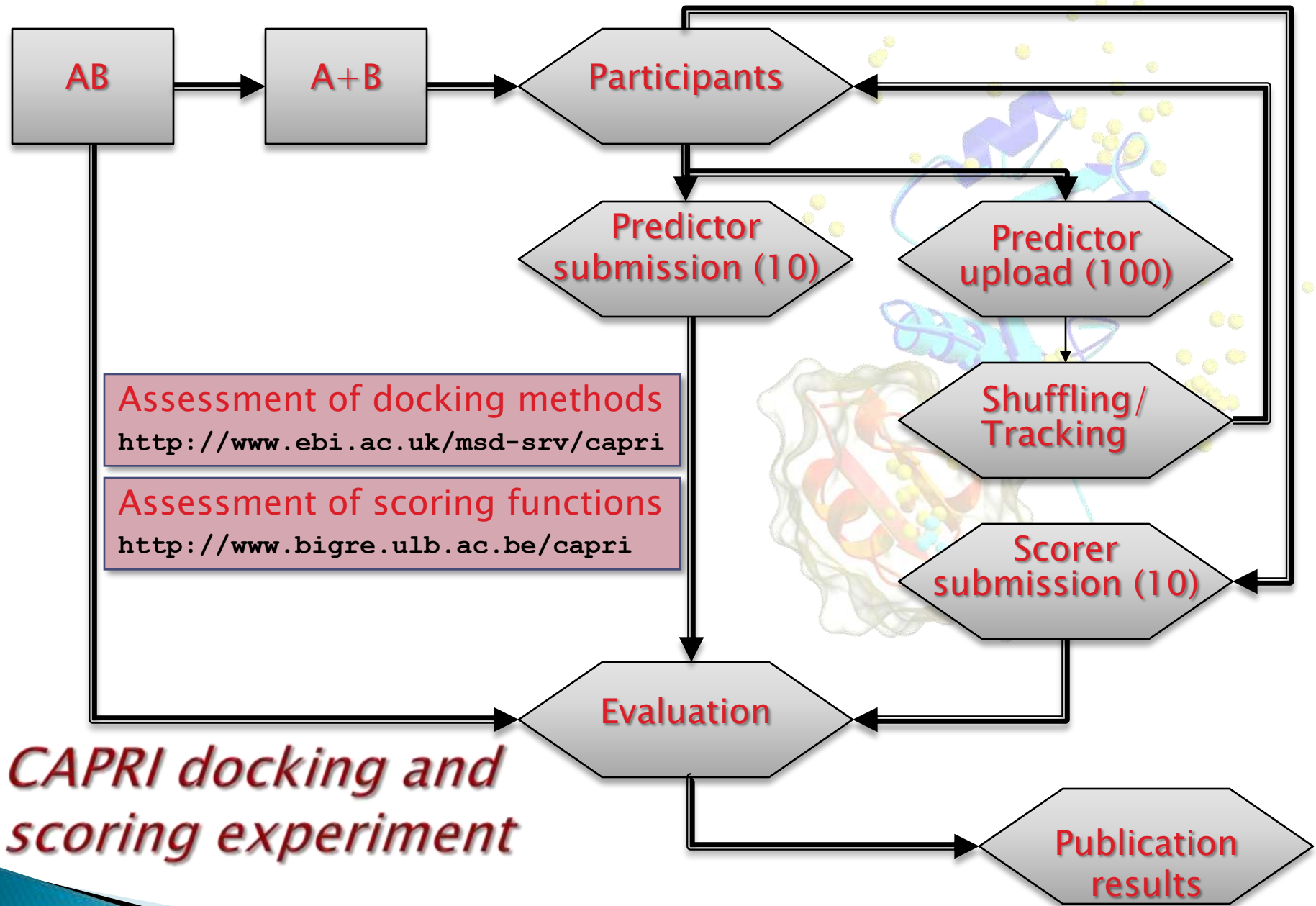


Docking procedures

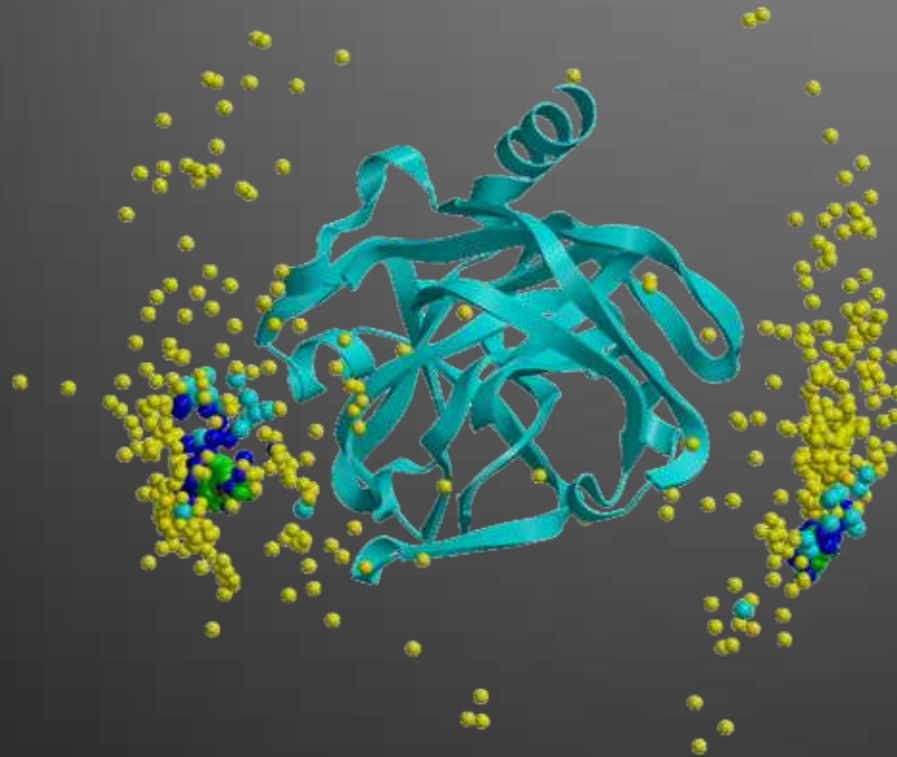
- ▶ Protein-protein docking involves two main steps:
 1. Producing and exploring a **very** large number of association modes
 2. Criteria to identify the correct association mode from a number of incorrect alternatives



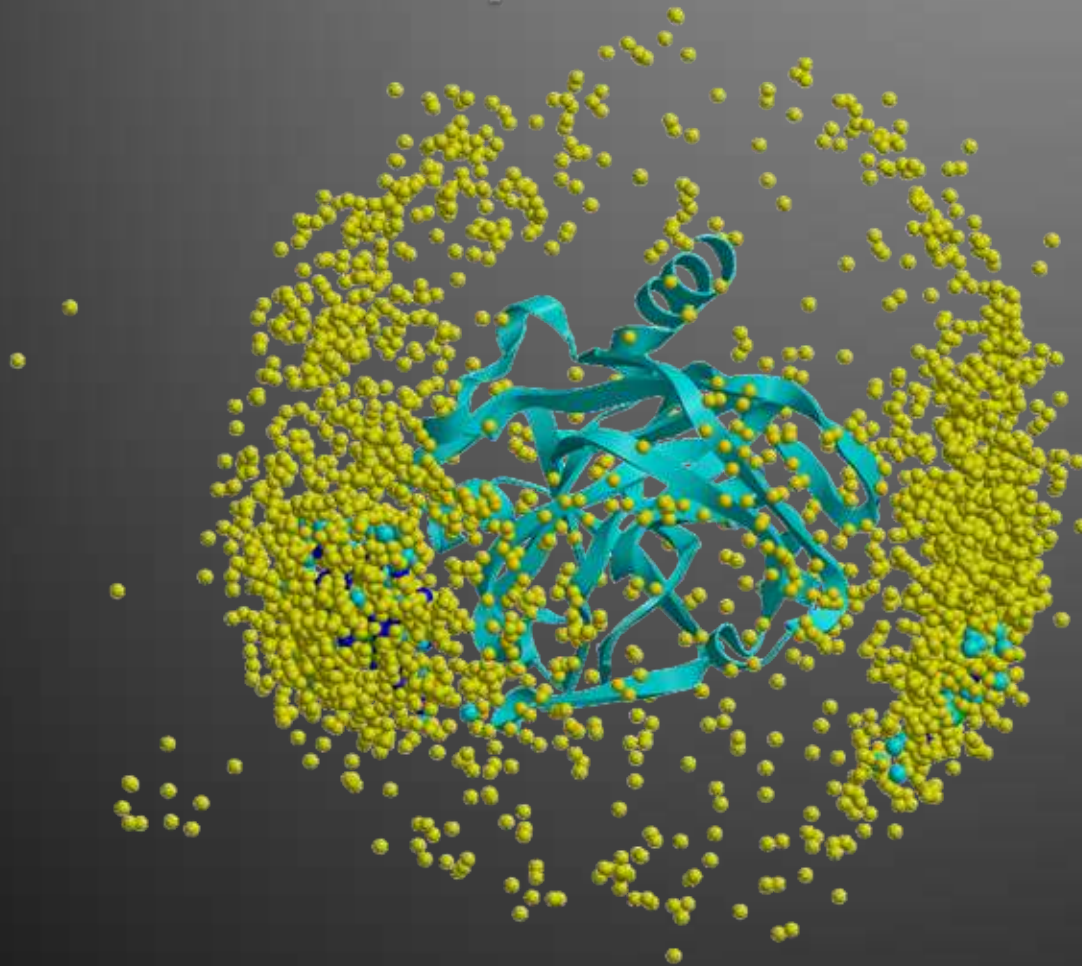




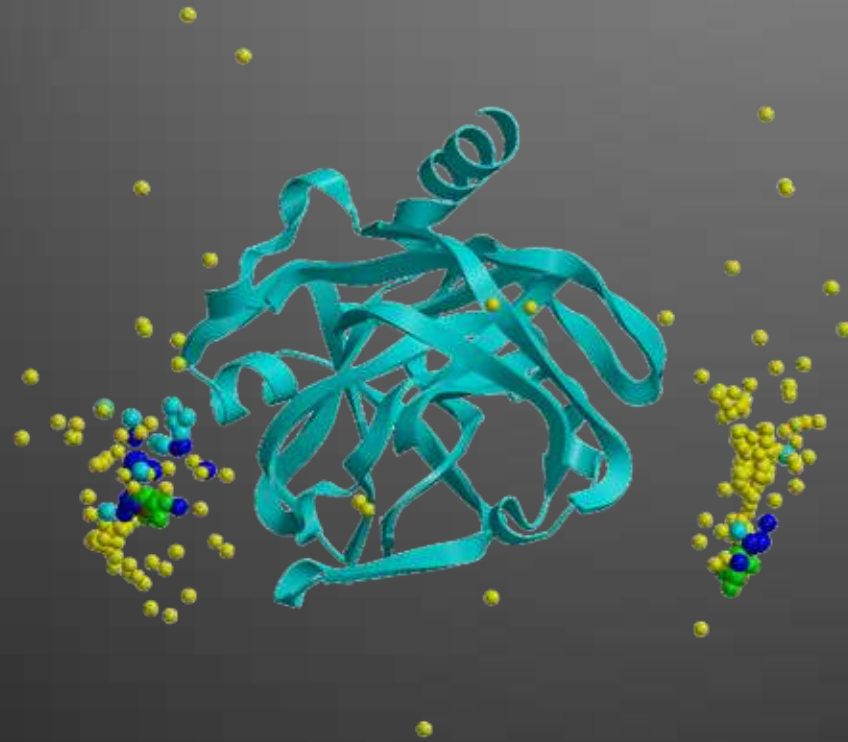
Predictors



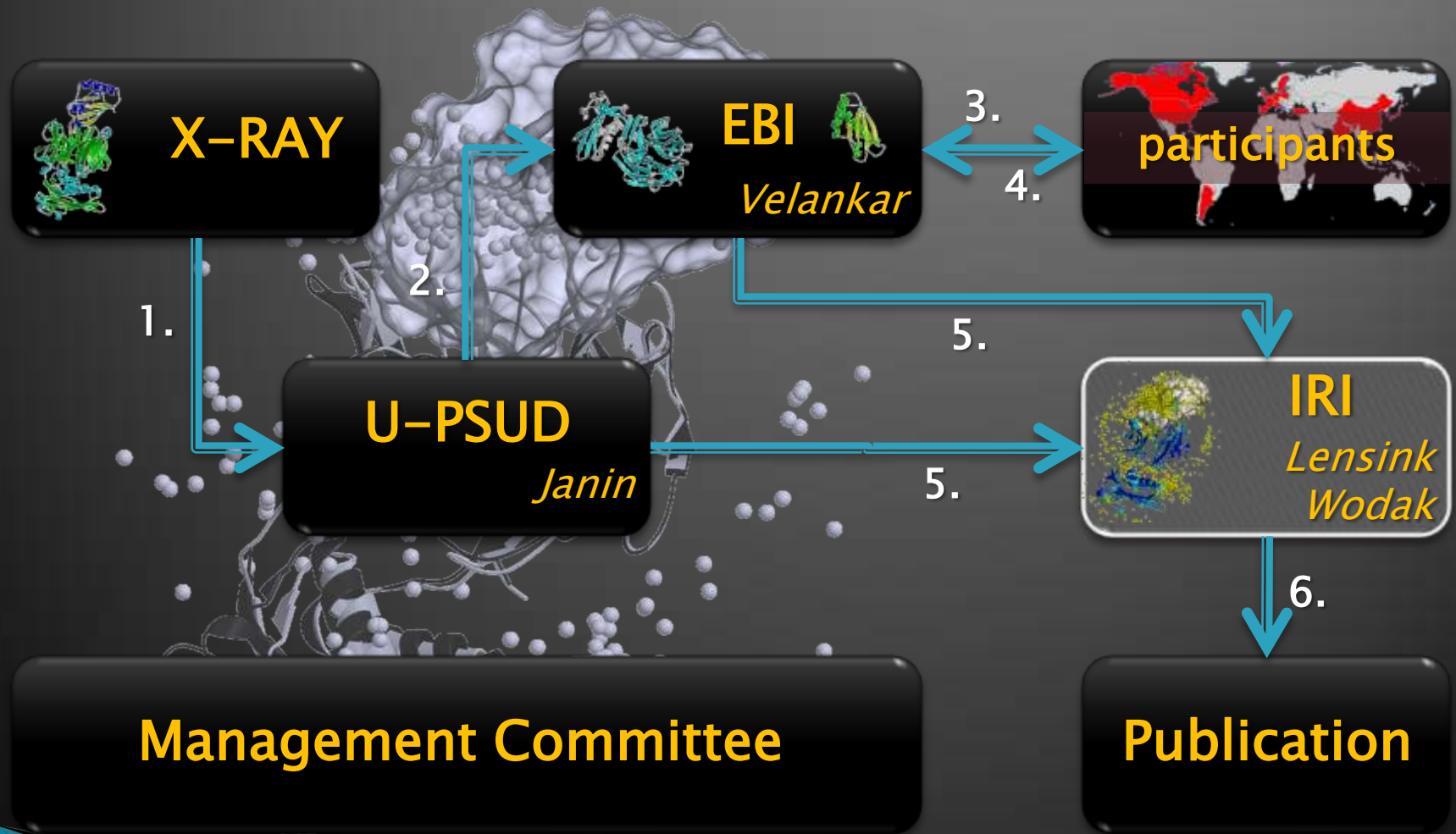
Predictors → Uploaders



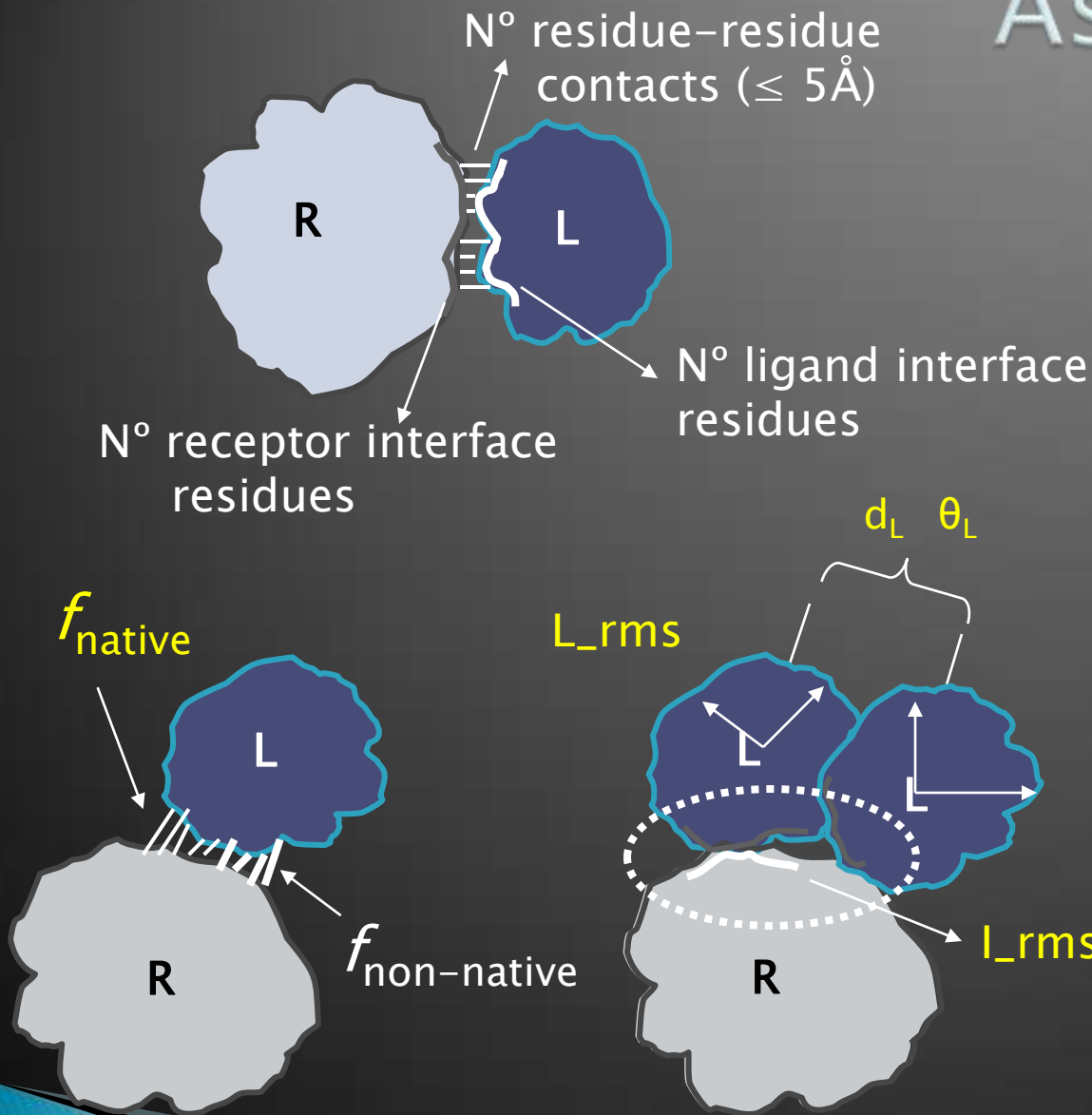
Predictors → Uploaders → Scorers



CAPRI organization



Assessment

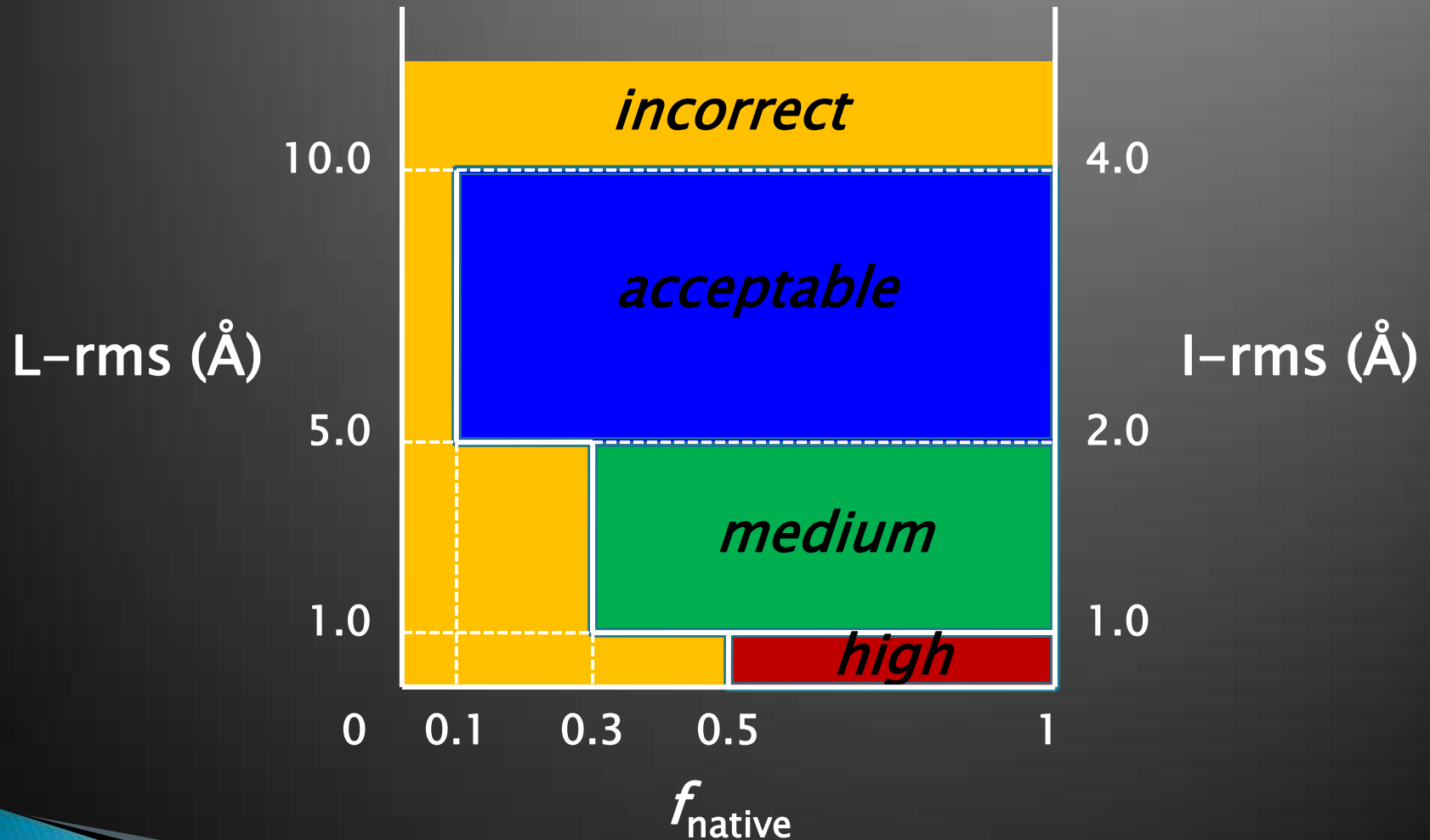


- ▶ f_{native}
residue-residue
5 Å
- ▶ L_{rms}
ligand residues
- ▶ l_{rms}
interface residues
10 Å
- ▶ n_{clashes}
atom-atom
3 Å
- ▶ d_L
- ▶ θ_L

CAPRI classification

incorrect	$f_{\text{nat}} < 0.1$	OR	L-RMS > 10.0	AND	I-RMS > 4.0
acceptable	$f_{\text{nat}} \geq 0.3$	AND	L-RMS > 5.0	AND	I-RMS > 2.0
OR	$(0.1 \leq f_{\text{nat}} < 0.3)$	AND	$(\text{L-RMS} < 10.0$	OR	$\text{I-RMS} < 4.0)$
medium	$f_{\text{nat}} \geq 0.5$	AND	L-RMS > 1.0	AND	I-RMS > 1.0
OR	$(0.3 \leq f_{\text{nat}} < 0.5)$	AND	$(\text{L-RMS} < 5.0$	OR	$\text{I-RMS} < 2.0)$
high	$f_{\text{nat}} \geq 0.5$	AND	L-RMS < 1.0	AND	I-RMS < 1.0

Classification criteria



CAPRI development

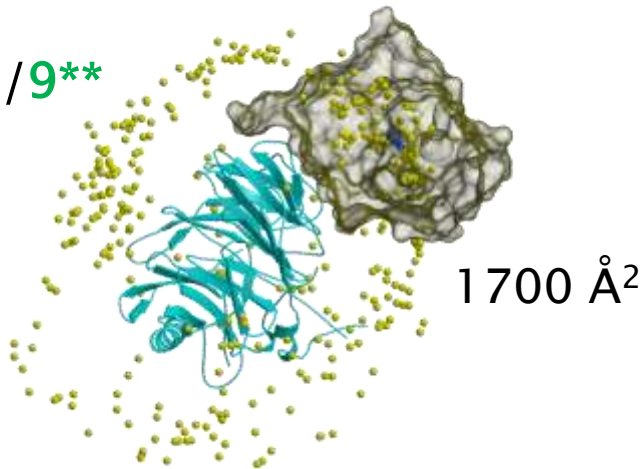
- ▶ Rounds 1,2 (2003)
 - Development of docking
- ▶ Rounds 3–5 (2005)
 - Analysis of docking results shows progress
- ▶ Rounds 6–12 (2007)
 - Introduction of scoring
 - Protein docking of very diverse targets
- ▶ Rounds 13–19 (2010)
 - Protein docking becomes a routine matter

Target Evolution

Early Targets	Enzyme/Inhibitor	Unbound/Bound	Crucial for development
or	Antigen/Antibody	Rigid-body	
T20/21/22	Gene expression	Unbound/Homology	Most targets now contain a homology-modeling step
T24/25/30	G-protein signaling		
T37/38/39	G-protein signaling		
T26	Membrane maintenance		
T27	Protein degradation	Multiple binding modes	
T23/28	Dimerization	Soluble monomer	
T29/33/34	RNA maturization	Protein/RNA complex	
T32/40	Enzyme/Inhibitor	Standard targets	
T35/36/41	Enzymes		
T42	Designed peptide repeat		

CAPRI targets : Gene expression

17/9**



T29 : tRNA methyltransferase
(RNA maturation)

T21 : Yeast Orc1 and Sir1 domains
(Eukaryotic gene expression)

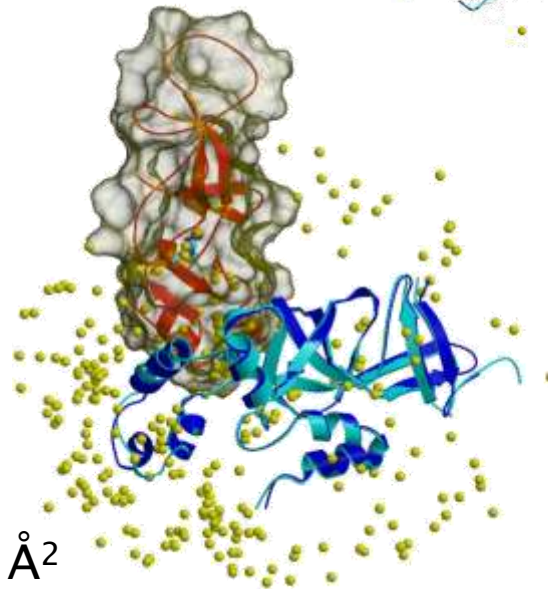
11/4**

1300 Å²

T34 : Protein – RNA
(RNA maturation)

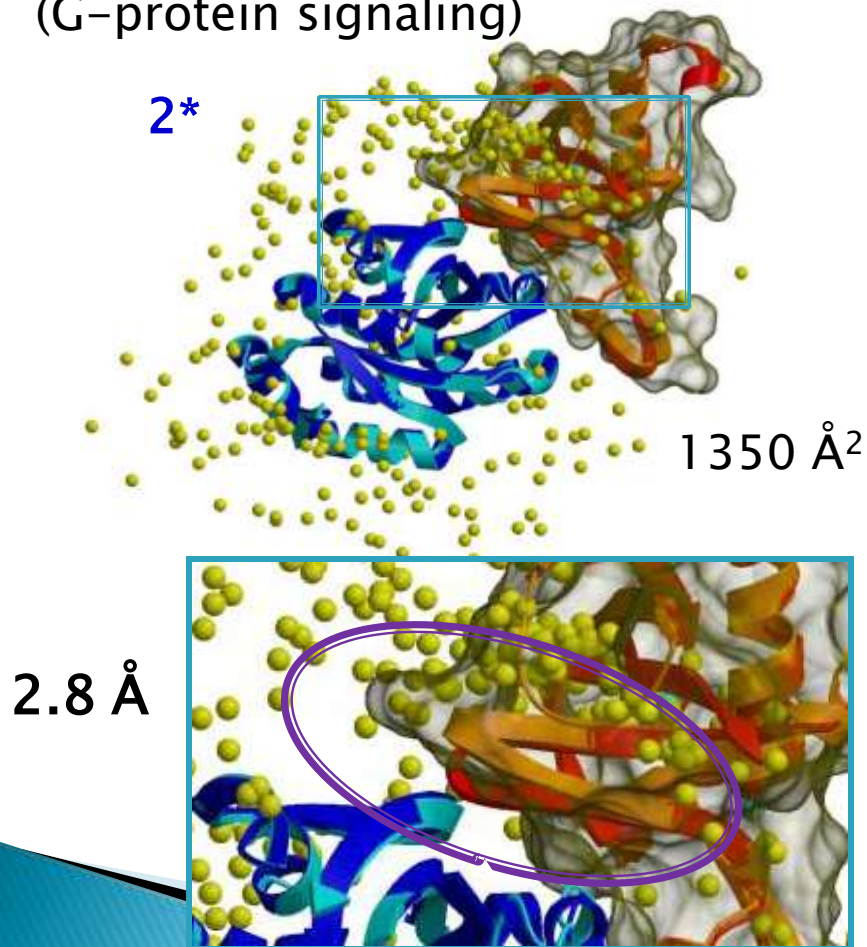
3700 Å²

16/9**

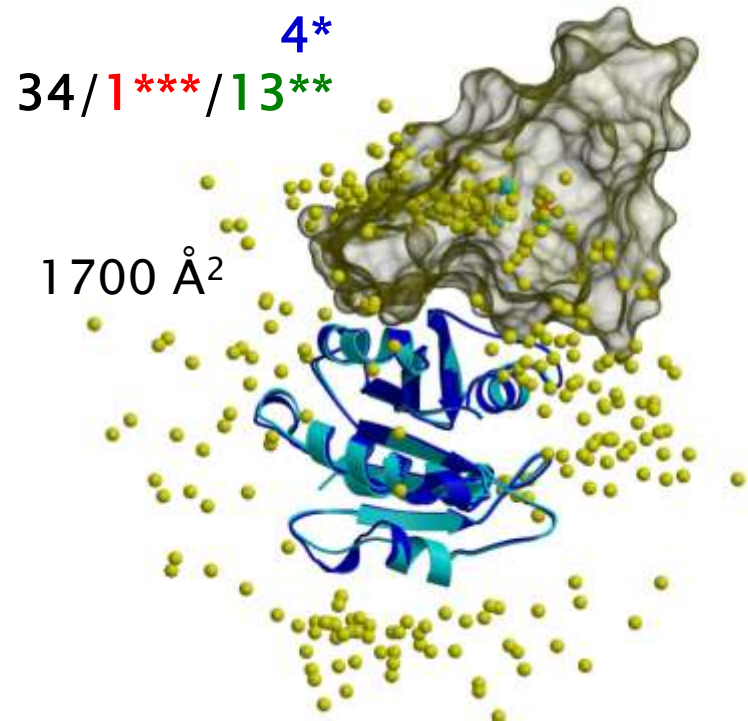


CAPRI targets : G-protein signaling

T30 : Rnd1 – RhoBD
(G-protein signaling)



T24/25 : Arf1 & ArfBD
(G-protein signaling)



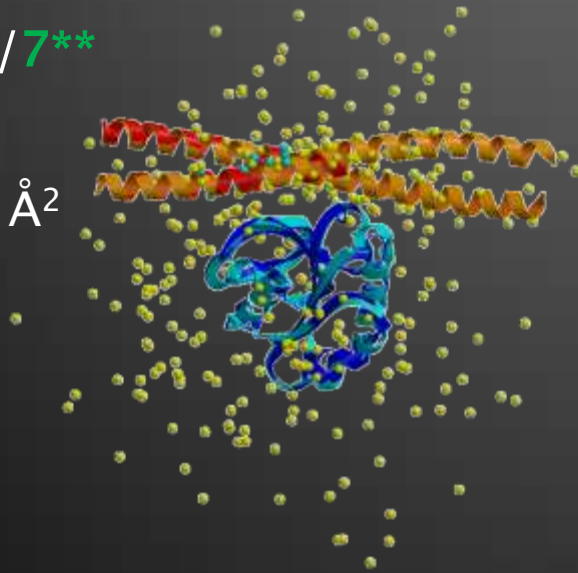
CAPRI targets : G-protein signaling

T38 : Centaurin- α – forkhead-associated domain
(G-protein signaling)

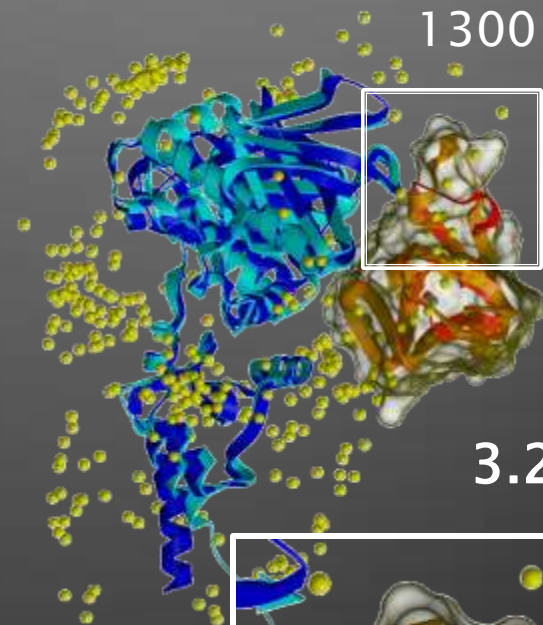
T37 : Arf6 – JIP4-LZ2
(G-protein signaling)

21/1***/7**

1500 Å²

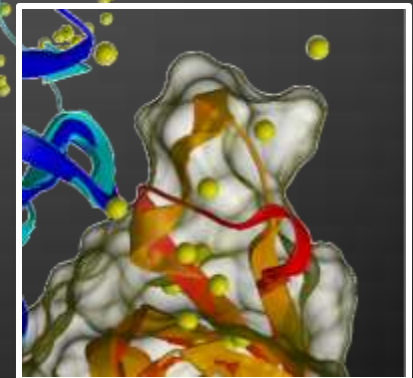


1300 Å²

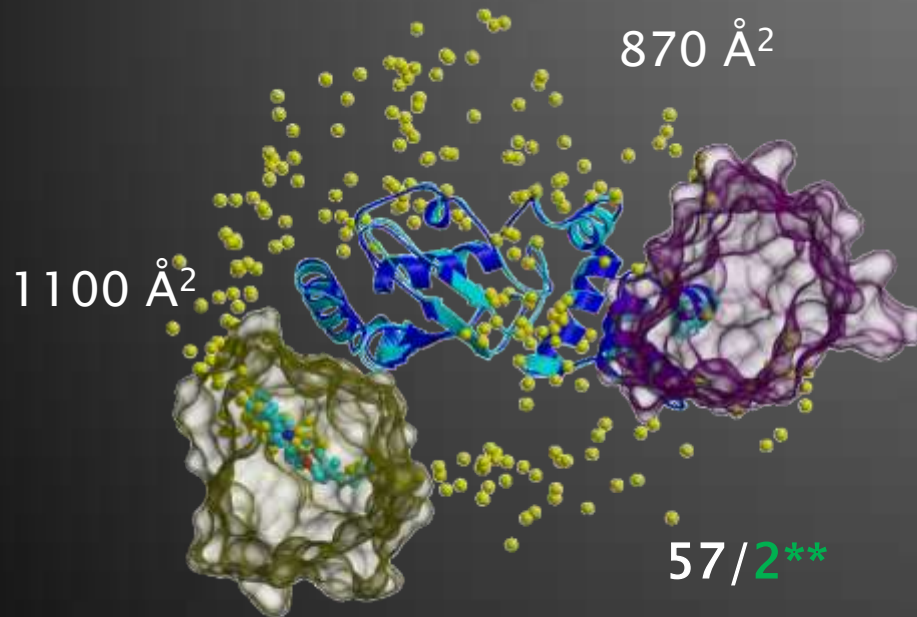


3.2 Å

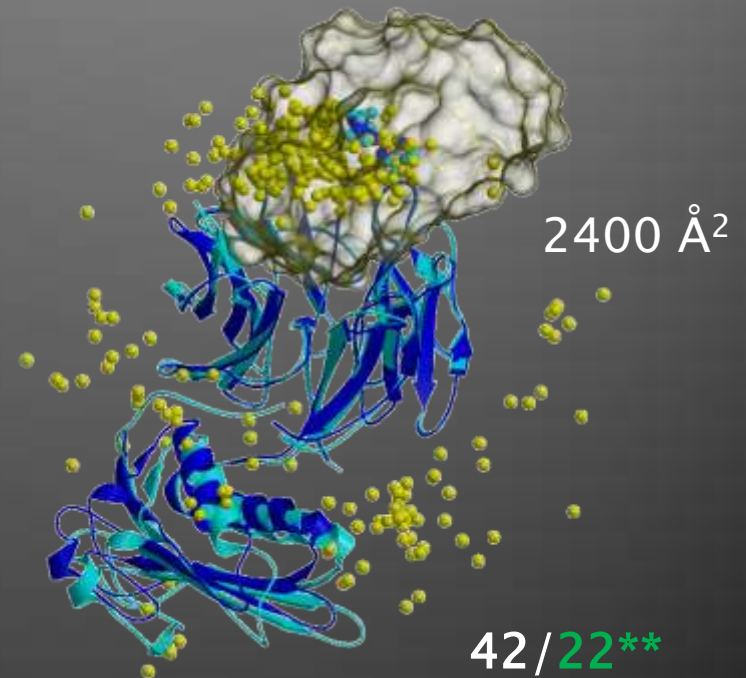
3/1***/2**



CAPRI targets : other



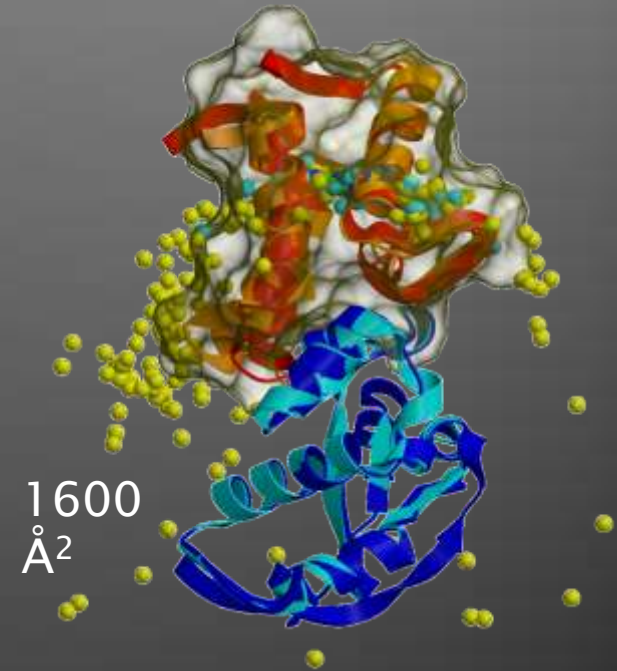
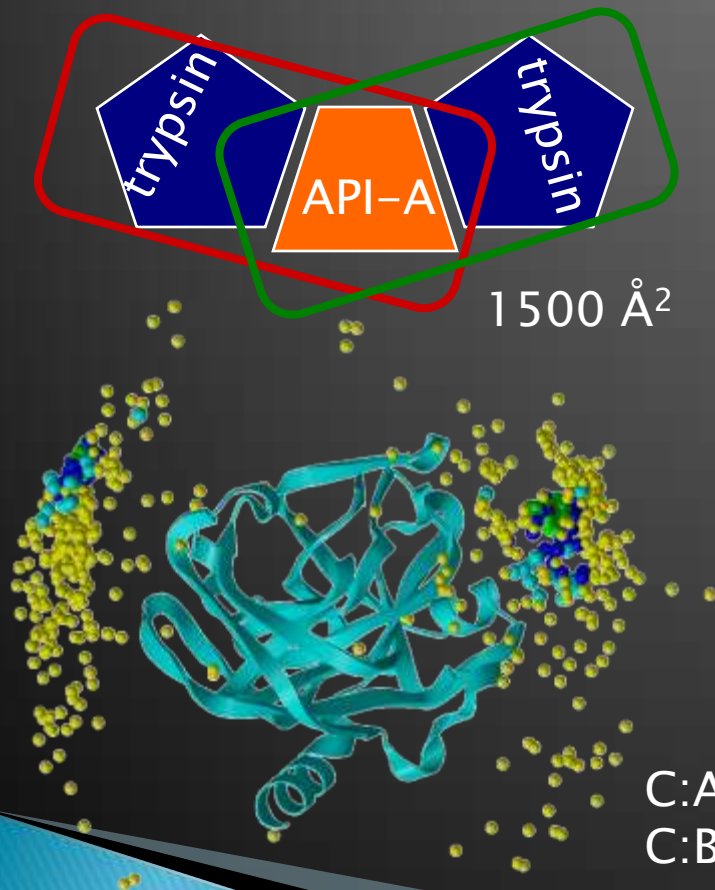
T27 : Ubiquitin and SUMO conjugating enzymes
(Protein degradation)



T26 : TolB Pal
(Membrane maintenance)

CAPRI targets : Enzyme/Inhibitor

T40 : Trypsin – protease inhibitor API-A
(Enzyme/Inhibitor)

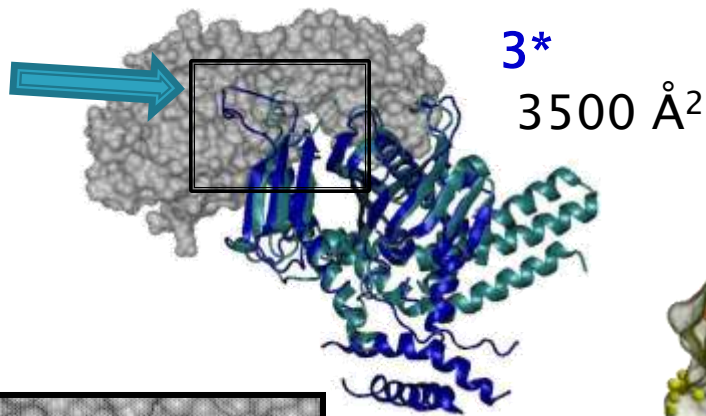


T41 : Colicin E9 – Im2 immunity
protein (Enzyme/Inhibitor)

26/8***/13**

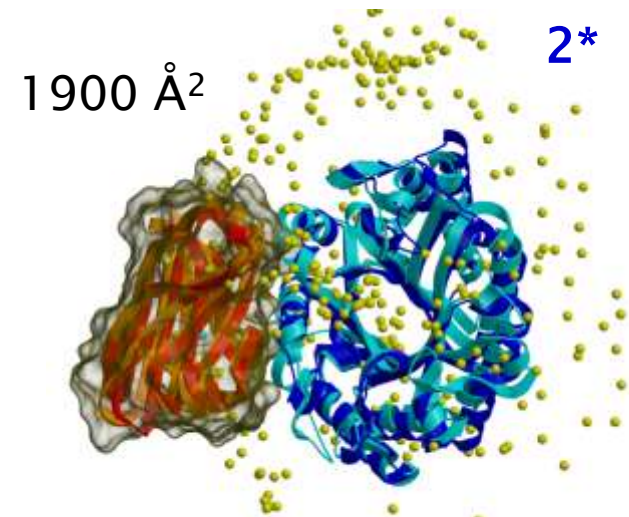
CAPRI targets : Enzyme/Inhibitor

T20 : *E. coli* release factor 1 – protein methyltransferase HEMK complex
(Enzyme/substrate complex involved in Gene expression)



1750 Å²

T32 : Savinase/BASI
(Enzyme/Inhibitor)



T35 : Xylanase carbohydrate binding module CBM22 and catalytic module GH10, covalently linked
(Enzyme, probably cleaved during crystallization)

34/15***/13**

Recent CAPRI results



Target	Predictors	Scorers
T29	7/4**	7/1***/5**
T30	2	0
T32	11/5***/3**	1
T33	0	0
T34	16/9**	4/3**
T35	1	1
T36	1	0
T37	16/1***/5**	13/4***/6**
T38	0	0
T39	3/1***/2**	0

Target	Predictors	Scorers
T20	3	
T21	11/4**	–
T24	4	0
T25	34/1***/13**	19/6**
T26	42/22**	14/7**
T27	57/2**	49/2**
T28	0	0
T40CA	24/18***/5**	11/6***/5**
T40CB	26/20***/3**	9/7***/2**
T41	26/8***/13**	11/1***/5**
T42AB	9/6***/2**	–
T42BC	4/2***/1**	–



Who are the consistent...

- ▶ Vajda, Zacharias, Zou Eisenstein, Wolfson, Weng Zhou, Bonvin, Gray, Bates Fernandez-Recio Baker, Camacho, Ten Eyck Wang, Vakser, Ritchie Smith, Totrov
- ▶ Bonvin, Bates, Zou Weng, Wang, Wolfson Fernandez-Recio
- ▶ Servers:
ClusPro, HADDOCK
GRAMM-X, SKE-DOCK
PatchDock, FiberDock

... dockers...

... and scorers...

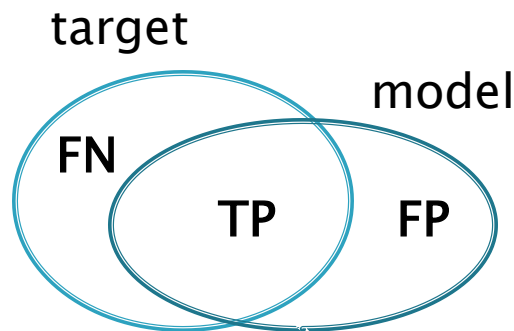
Prediction of interaction interface

- ▶ Can protein–protein docking help in predicting the interface of protein–protein interactions?
 - Insight into protein function
 - Inform mutagenesis studies
 - Drug design applications



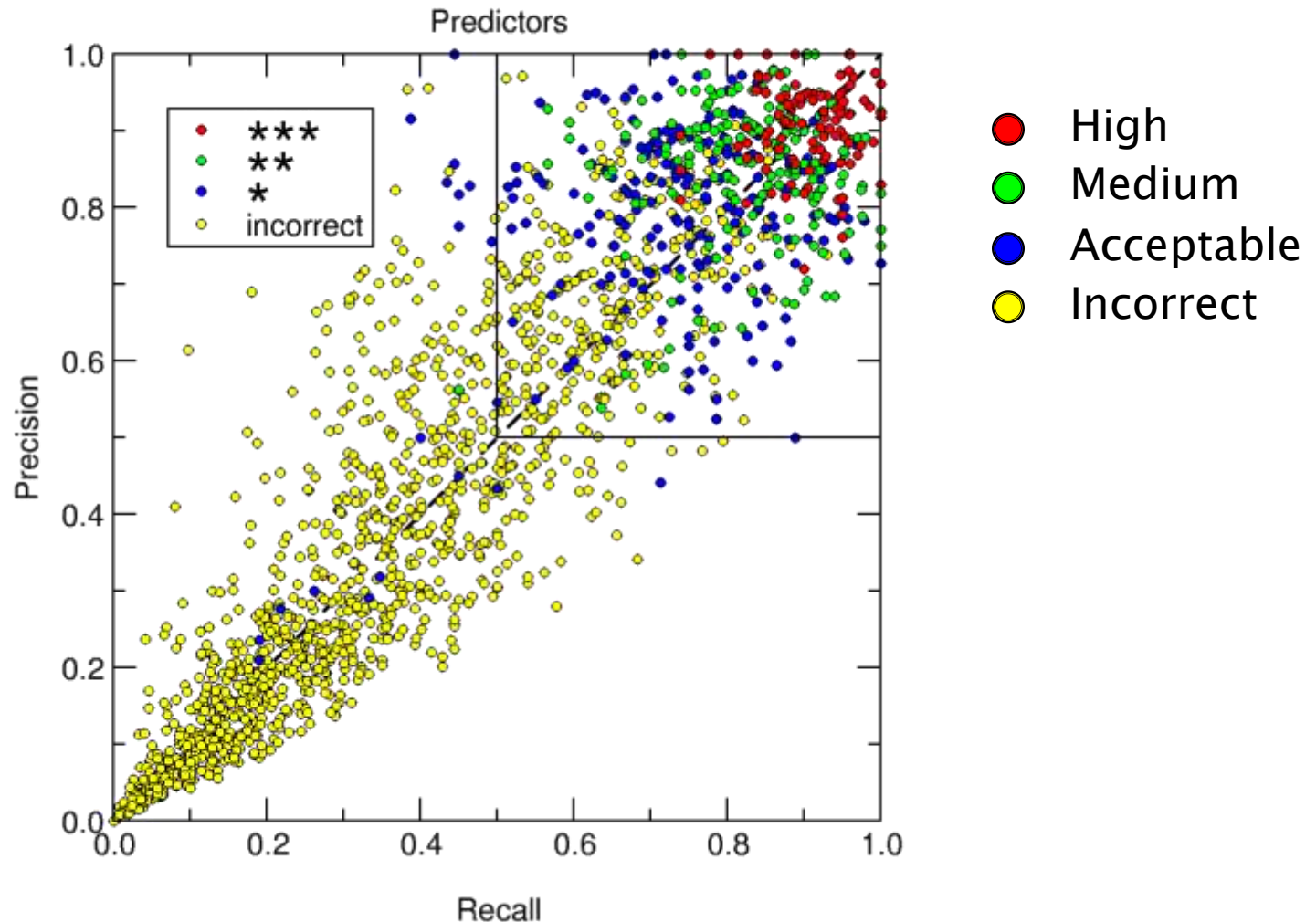
Interface prediction

- ▶ Measure overlap of protein docking interfaces with true interfaces
- ▶ $Recall = TP / (TP + FN)$
 - Fraction of true interface residues reproduced by prediction
- ▶ $Precision = TP / (TP + FP)$
 - Fraction of the model interface correctly predicted

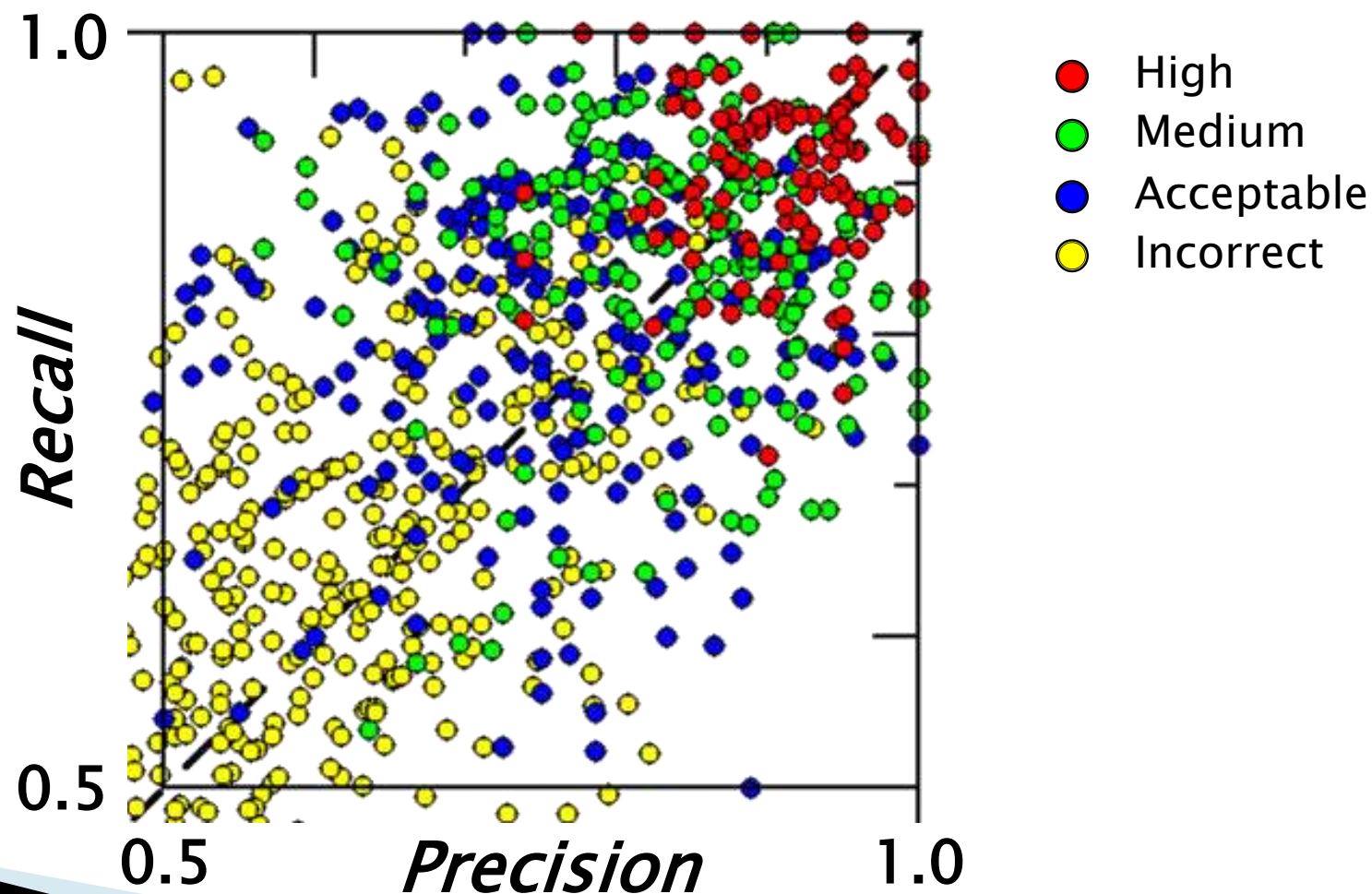


An interface residue is defined as any residue that loses ASA upon binding

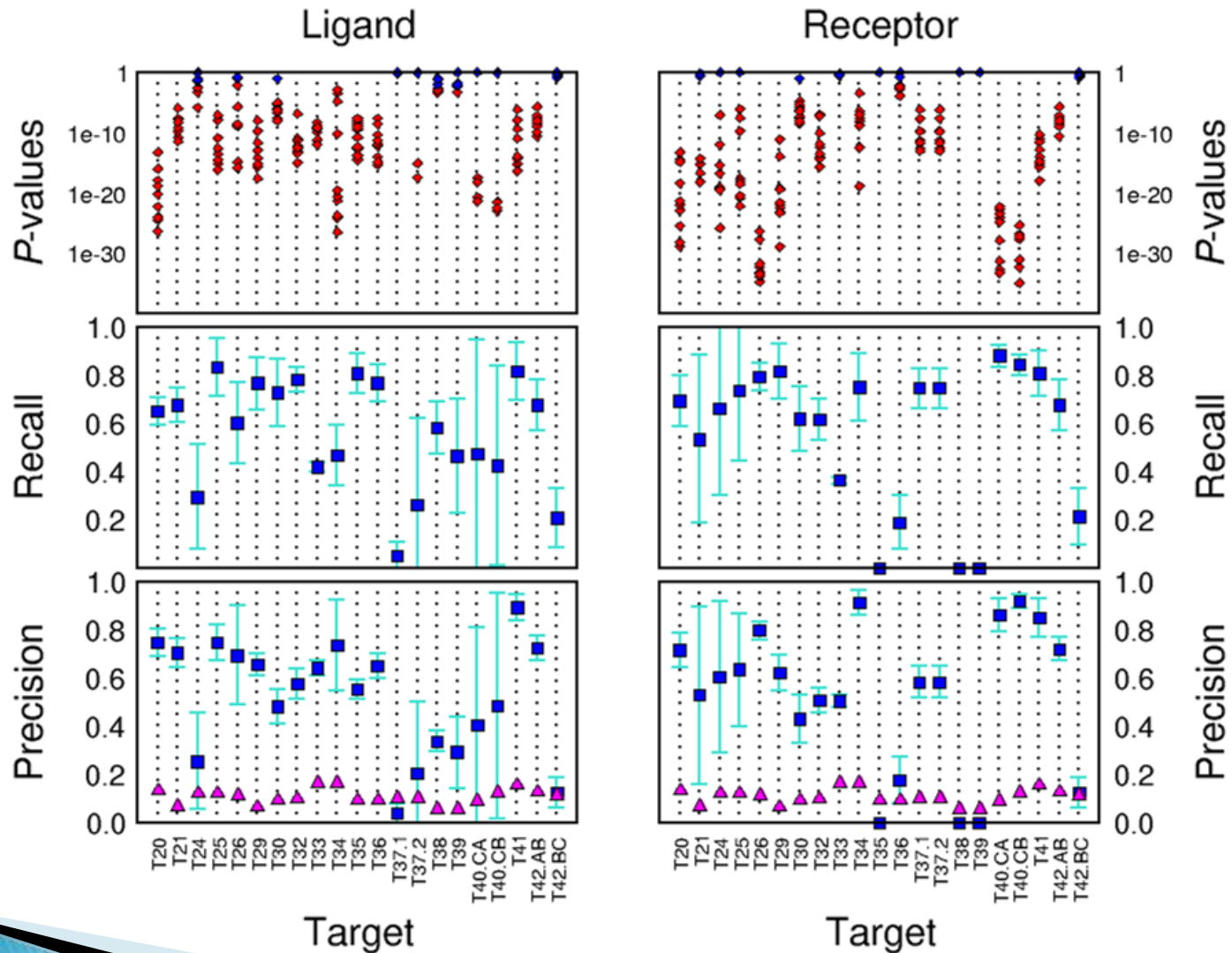
Global prediction landscape



Many incorrect docking models may show a correct interface prediction!



Quality of predicted interfaces



Conclusions

- ▶ Interfaces predicted by CAPRI participants show *recall* and *precision* levels at 60%
- ▶ Versus 30% *precision* and 50% *recall* for most recent interface prediction methods
- ▶ 24% of interface in incorrect models show *recall* and *precision* $\geq 50\%$
- ▶ Interfaces predicted in CAPRI are truly blind predictions
- ▶ Non-obligate complexes
- ▶ Docking calculations may be used to predict protein-protein interaction interfaces with higher reliability
 - Inclusion of interaction data
 - Improvement expected from meta-prediction methods

Conclusions

- ▶ Enzyme/Inhibitor docking is straightforward and reliable
- ▶ Docking of proteins to nucleic acids poses no problem
- ▶ Homology-modeling is now routinely used
- ▶ Docking of standard-sized interfaces (1000 – 1800 Å²) produces acceptable, or even high-quality results
 - Provided there is no conformational change at the interface
 - Moderate conformational changes (< 2 Å) can be handled
- ▶ Multiple protein interfaces are recognized
 - Protein interface preferred over crystal packing
- ▶ Even if unsuccessful in producing a high-quality three-dimensional model of the assembly, docking can help answer many biological questions.
 - Especially when combined with biochemical interaction data

CAPRI sets the Standard in Protein Docking

- ▶ Significantly increased reliability
 - *Docking servers give reliable results*
- ▶ Catalyst for Development
 - **ZDOCK, HEX, ClusPro, HADDOCK**
- ▶ Deposition in Protein Data Bank
- ▶ Difficulties
 - **Conformational change**
 - including displacement of loops
 - **Singling out (scoring) the best models**
 - **Required information**
 - biochemical interaction data to “guide” the docking
- ▶ Keywords

Flexibility

Interface

For more information

- ▶ Wodak & Janin, *Adv Protein Chem* 2002;61:9
“The structural basis of macromolecular recognition”
- ▶ Janin *et al*, *Proteins* 2003;52:2
“Critical Assessment of PRedicted Interactions”
- ▶ Wodak & Méndez, *Curr Opin Struct Biol* 2004;14:242
“Predictions of protein–protein interactions”
- ▶ Lensink *et al*, *Proteins* 2007;69:704
“Docking and scoring protein complexes CAPRI 3rd edition”
- ▶ Lensink & Méndez, *Curr Pharm Biotechnol* 2008;9:77
“Recognition–induced conformational changes in protein–protein docking”
- ▶ Lensink & Wodak, *Proteins* 2010;78:3073 & 3085
 - a) *“Docking and scoring protein interactions: CAPRI 2009”*
 - b) *“Blind predictions of protein interfaces by docking calculations in CAPRI”*

Thanks to

- ▶ M Graille, N Leulliot, H van Tilbeurgh; Z Hou, J Keck; T Nielsen, R Ficner; A Ghosh, C Herrmann; J Ménétrey, T Isabet, A Housusse; D Bonsor, A Keeble, N Meenan, A Krachler, C Kleanthous; J Walker, S Dhe-Paganon, A Bochkarev; Y Tong, P Hota, M Buck; K Venkateswarlu, A Chishti; P Micheelsen, M Skjød; R Bao, Y Chen; S Najmudin, C Fontes
- ▶ (T20 – T42 only)
- ▶ M Zacharias, X Zou, M Eisenstein, H Wolfson, Z Weng, H Zhou, A Bonvin, J Fernandez–Recio, J Gray, P Bates, C Camacho, J Bernauer, H Nakamura, D Baker, C Wang, D Ritchie, Y Xiao, M Takeda–Shitaka, A Poupon, J Mitchell, F Jiang, S Comeau, E Alexov, A Tovchigrechko, C del Carpio, S Günther, M Totrov, R Abagyan
- ▶ ... and many others
- ▶ J Janin, K Henrick, J Moult, S Wodak, L Ten Eyck, M Sternberg, S Vajda, I Vakser

Crystallographers

Participants