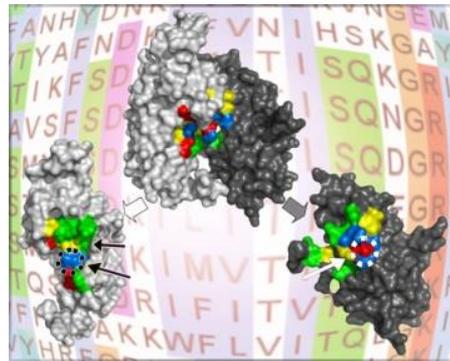
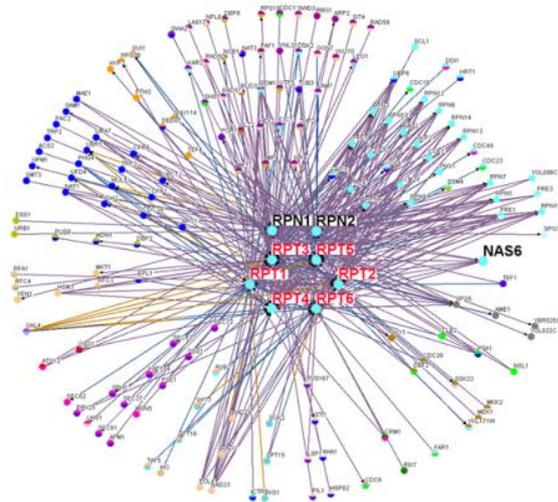


EXPLORING STRUCTURAL INTERACTOMES IN THE LIGHT OF EVOLUTION

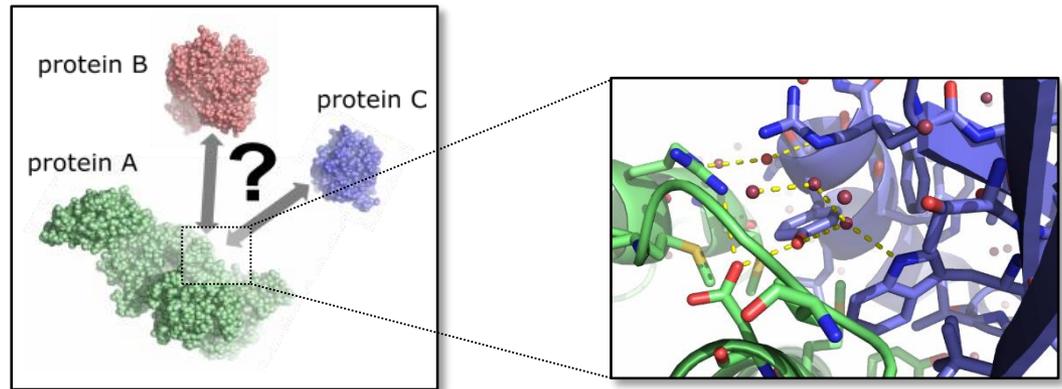


Jessica ANDREANI
Raphaël GUEROIS

Molecular Assemblies and Genome Integrity team
STRUCTURAL BIOLOGY AND RADIOBIOLOGY LAB

I2BC - CEA Saclay

PROTEIN INTERACTIONS AT THE HEART OF MOST CELLULAR PROCESSES

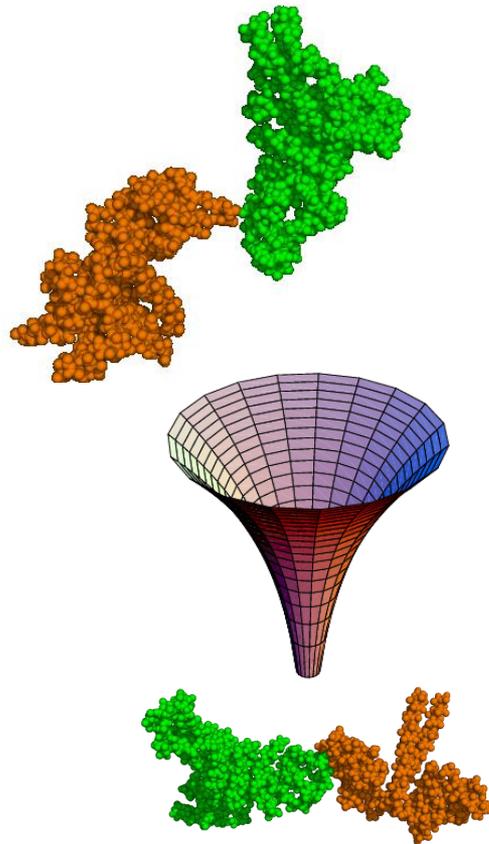


- The activation/inhibition of cellular pathways rely on **synergies** and **competitions** at the binding surfaces of macromolecules
- An atomic/residue scale analysis
 - ➔ Steric hindrance or co-association between partners
 - ➔ Specific design of disruptive and compensatory mutants
 - ➔ Design of PPI modulators

FINDING A CORRECT INTERFACE IN THE HAYSTACK OF POSSIBLE ASSEMBLIES

Protein A $\leftarrow ? \rightarrow$ Protein B

PRINCIPLES OF MOLECULAR DOCKING



$> 10^4$ decoys

STEP 1: Low resolution step.
Coarse-grained rigid body docking

Filters

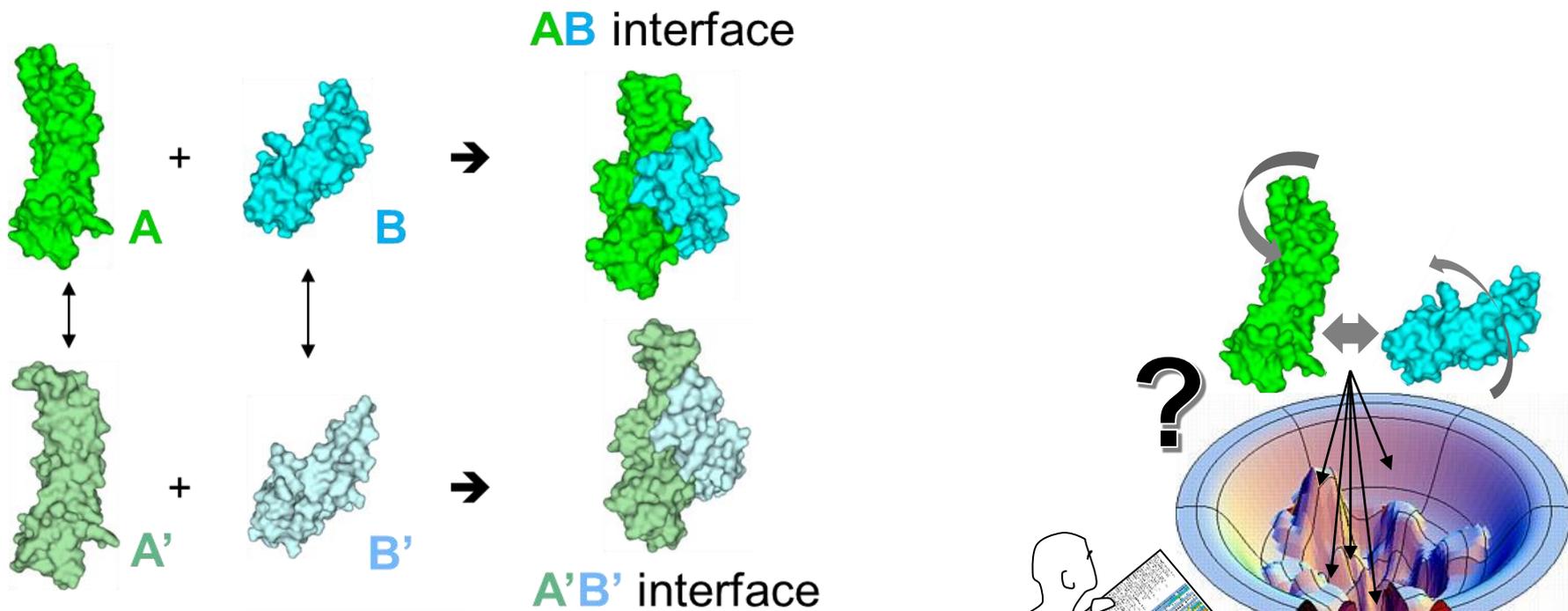
EVOLUTIONARY CONSTRAINTS ??

~ 10 clusters

STEP 2: refinement step.
Flexibility, atomic details

1 most likely model

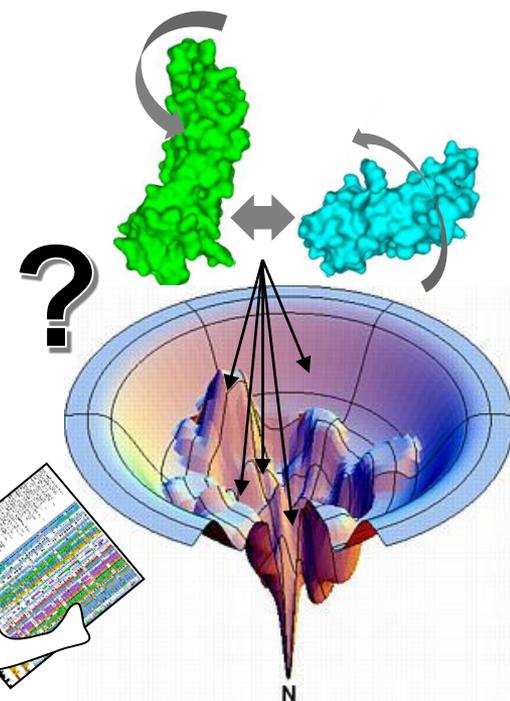
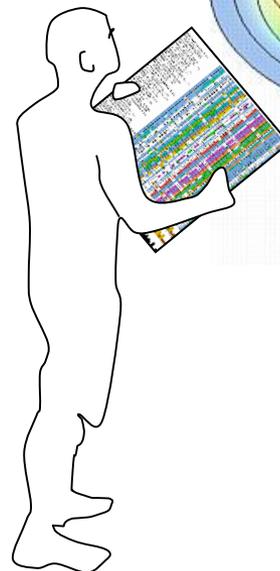
CONSERVATION OF BINDING MODES DURING THE COURSE OF EVOLUTION



M	F	Y	H	I	S	L	E	H	E	I	L	L	H
M	F	Y	H	I	S	L	E	Q	E	I	L	L	H
M	F	F	H	I	V	L	E	R	N	M	Q	L	H
M	F	F	H	I	V	L	E	R	N	M	Q	L	H
M	F	Y	H	I	A	L	E	H	E	I	L	L	H
F	Q	I	S	L	D	H	E	I	L	L	H	P	R
M	F	Y	H	I	S	L	E	H	E	I	L	L	H
M	F	Y	H	I	S	L	E	H	E	I	L	L	H
M	F	F	H	V	T	L	E	H	E	I	T	L	H
M	F	F	H	L	T	L	E	K	D	L	H	M	H
M	F	F	I	K	E	L	T	H	T	I	L	L	H
M	P	F	F	L	K	E	L	S	L	T	I	S	L
M	F	F	L	K	E	E	T	K	V	I	S	L	H
M	F	F	L	K	E	E	T	K	V	V	T	L	H
M	F	F	L	K	D	L	S	L	N	L	T	L	H
M	D	G	R	W	R	L	Q	R	S	L	A	Q	R

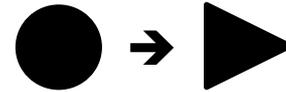
G	P	N	L	N	T	V	K	Q	K	L	F	T	E	
G	P	Q	L	L	E	T	V	K	Q	K	L	Y	S	E
G	R	N	L	K	E	N	L	V	S	K	L	M	K	D
G	P	H	L	R	D	K	L	V	S	K	L	I	K	D
G	P	Q	L	M	E	T	V	K	Q	K	L	F	T	E
Q	L	E	E	T	V	K	Q	K	L	A	E	V	E	
G	P	N	L	L	D	T	V	K	Q	K	L	F	T	E
G	P	Q	L	I	E	T	V	K	Q	K	L	Y	T	E
G	P	Q	L	I	E	T	V	K	Q	K	L	Y	T	E
G	P	N	L	L	D	T	V	R	R	K	L	F	S	E
G	P	N	L	N	F	I	A	T	Q	Q	L	Y	S	E
G	P	R	M	L	N	F	L	E	S	K	L	Y	S	E
F	G	P	R	M	D	Y	L	K	A	K	L	L	A	
G	P	N	M	R	E	Y	L	I	N	R	L	N	E	E
G	P	N	V	K	E	Y	L	V	N	R	L	N	E	E
G	P	Q	M	D	Q	Y	L	R	D	K	L	L	S	D
S	T	E	L	T	Q	K	D	L	T	L	S	L	T	L

Threshold for interface conservation
~ 30% seq. id
(Aloy et al JMB 2003)



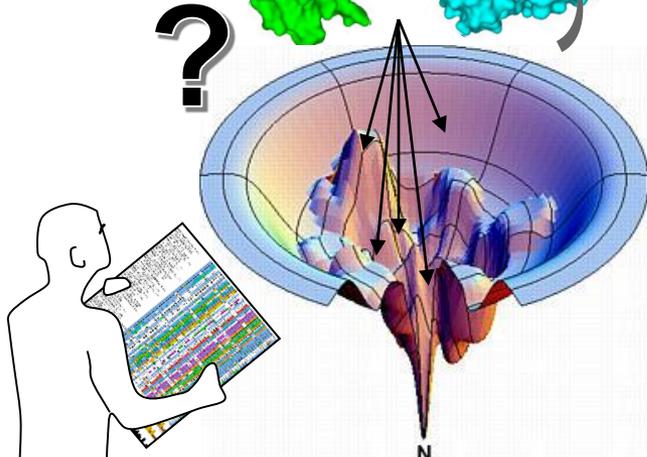
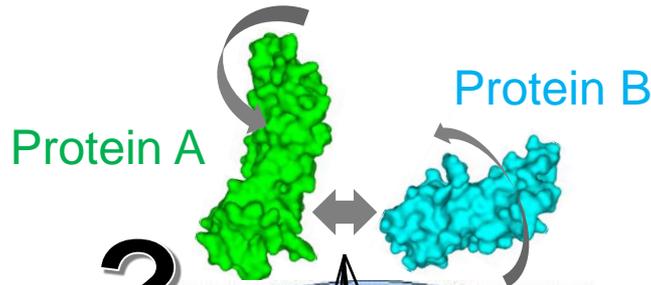
IS THE EVOLUTIONARY TRACE OF INTERACTING PARTNERS USEFUL TO PREDICT HOW THEY BIND ?

Substitutions in the alignment



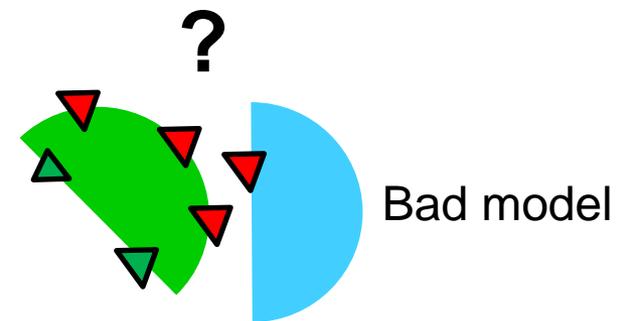
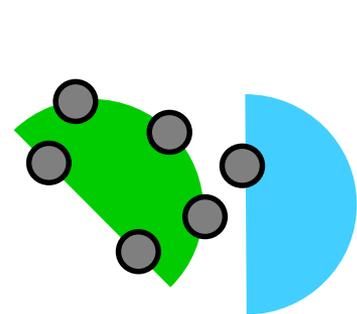
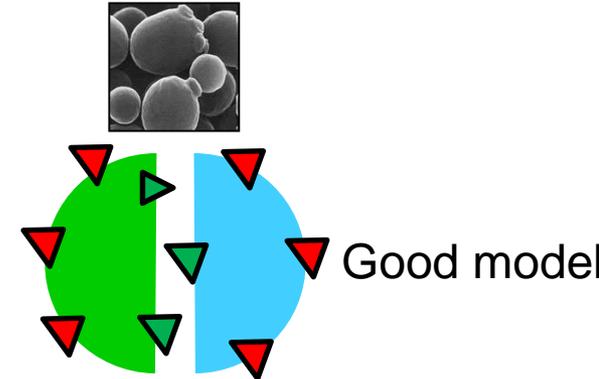
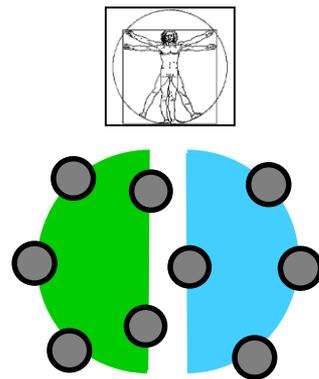
● → ▼ Deleterious mutation

● → ▼ Tolerable/Favorable mutation



Protein A Protein B

10	20	30
MFYHISLEHEILLH	GNLNTVKQKLFTE	
MFYHISLEQEILLH	GPQLLETVKQKLYSE	
MFFHIVLERNMQLH	GRNLKENLVS KLMKD	
MFFHIVLERNMQLH	GPQLLETVKQKLYTE	
MFYHIALEHEILLH	GPQLLETVKQKLYTE	
FQISLDHEILLHPR	QLEETVKQKLYAEVE	
MFYHISLEHEILLH	GNLLETVKQKLFTE	
MFYHISLEHEILLH	GPQLLETVKQKLYTE	
MFYHISLEHEILLH	GPQLLETVKQKLYTE	
MFFHVTLEHEITLH	GNLLDTVRRKLFSE	
MFFHLLTLEKDLHMH	GNLFTIATQQLYSE	
MFFIKELTHTILLH	GPRLNPLESKLYSD	
MPFFLKELSLSLISL	FGRRMQDY LKAKLLA	
MFFLKEETKVISLH	GNMRYLINKLNNE	
MFFLKEETKVMTLH	GNVKEYLVNRLNEE	
MFFLKLDLSLNLTLH	GPQMDQYLRDKLLSD	
MDGRWRLLQRS LAQR	STELTKDQLTSLTL	



PRINCIPLES IN INTERFACE COEVOLUTION : STATISTICAL ANALYSES → QUANTIFICATION OF INTERFACE PLASTICITY



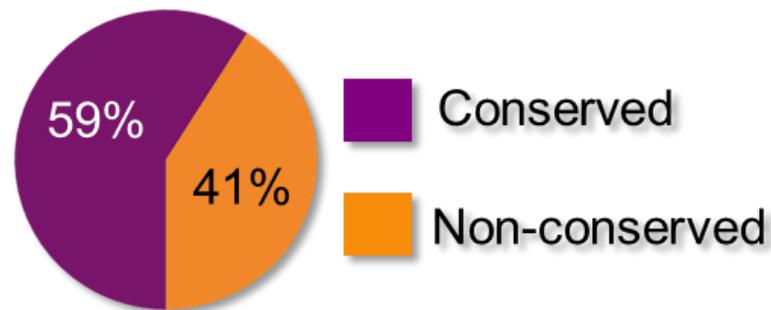
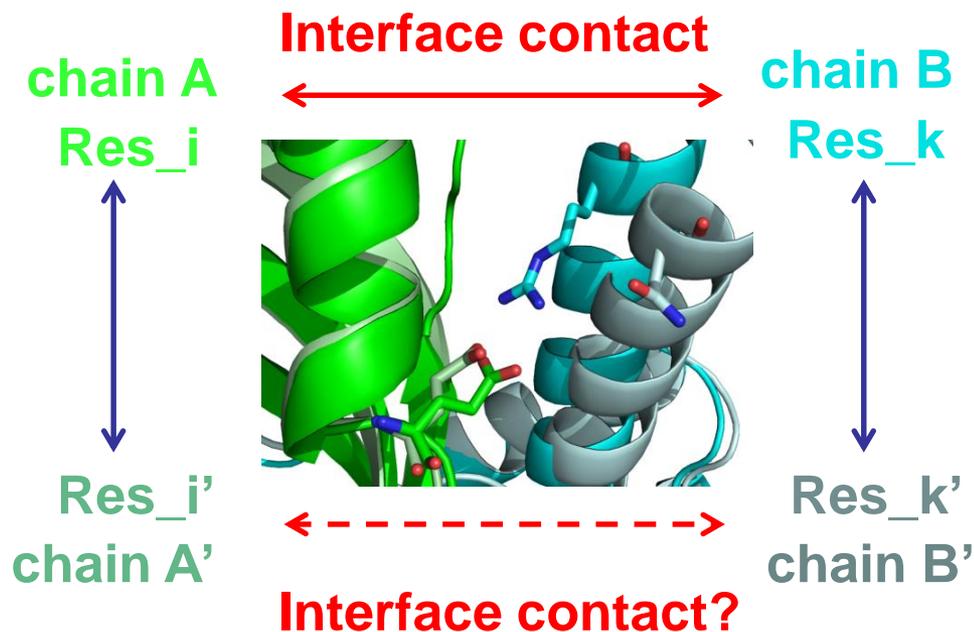
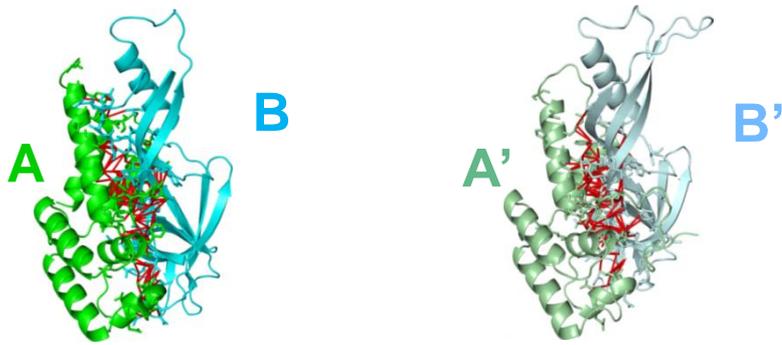
Database of
interfaces with structural &
evolutionary information

~18,000 non-redundant interfaces
among which

~1,000 pairs of interologs

Faure, Andreani & Guerois
Nucleic Acids Res. (2012)

<http://biodev.cea.fr/interevol>



Andreani, Faure & Guerois
PLoS Comp Biol (2012)

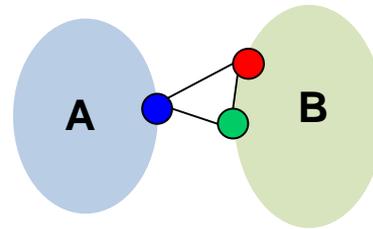


Database of
interfaces with structural &
evolutionary information

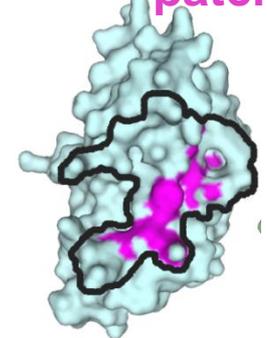


InterEvScore : Discriminate co-evolved interfaces

Multi-body
statistical potential



Apolar
patches



Evolutionary
information

≥ 10 sequences

Andreani, Faure & Guerois
PLoS Comp Biol (2012)
Bioinformatics (2013)

Sequence logos for regions A and C:

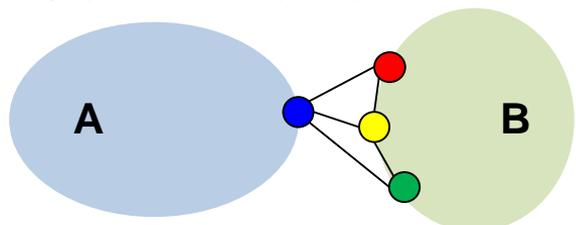
```

Region A:
T A E T L L N S E V H M L L E R H Q Q N E
S A E S L L N S E V Y M L L E R H T Q N E
N A E T L L I S E V H M L L E R H N Q N E
N A E T L L I S E V H M L L E R H N Q N E
N A E T L L I S E V M L L L E R H N Q N E
Q A N T L T S E V Y L L L E H N Q Q S E
T C H A L L T A V Y L L L E H N Q S S E
N A N F L L N S E V A I L L E H N G E S E
N A G C L L I S E V Y L L E N L E H T A
N E D M L T V S E A I L I E T V L A Q T A
G V E T L S V S E A E L V L E A L E R H R
N V P T L S I S E A A L V I N E V L D L R
R V E T L I S E A A L V I N E A L A H R
R V E T L I S E A S L V I N A L A H R
R V E T L I S E A S L V I N A L N T V R
G V E T L I S E A S L V I D A L M L R
A N C L M N C E V A I I E R R Y E Q I Q

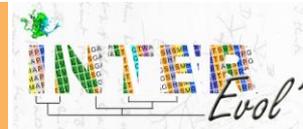
Region C:
L N T V Q A L F T E V E G T C T G Y
L I Q T V Q A L F T E V E G T C S G Y
L I E T V Q A L Y T E V E G T C T G Y
L I E V T A L Y T E V E G T C T G Y
L L I T V Q A L Y F E V E G T C T G Y
L L E T V Q A L F N E V E G T C T G Y
L N E T I K M L F N E V E G T C T G Y
L E T I A T Q L Y S E V E G T C T G Y
M L S Y L E T L Y N D V E G T C S G Q F
M Q Y L A K A L L A I V E G T C S G Q Y
R H E I V V Q L L A I T E G T C O G F
V H E Y G T C L L D D V E G I C T G B Y
R H E L V T G L L L D V E G T C A G S Y
R T E L V T T L V K D V E G T C T G Y
R H E L V T T L V K D V E G T C A G Y
R H E L V T T L V K D V E G T C A G Y
L E D L V S K L I K D V E G T C S G Y
    
```

Development of InterEvScore, a docking score taking evolution into account

Inter-molecular interface contacts
(every pair or triplet)



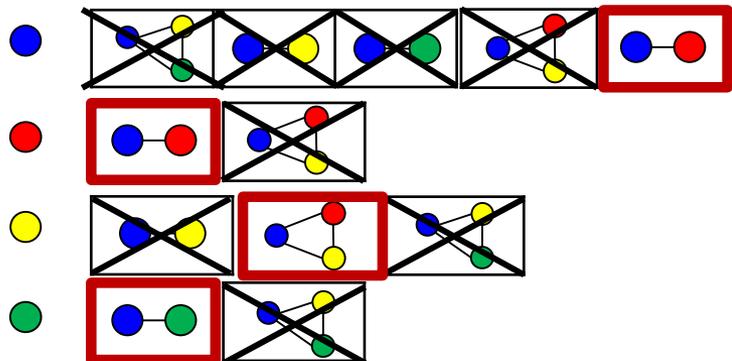
Contact propensities
derived from InterEvol
(statistics on 1,289 interfaces)



	0.53		0.04
	-0.05		-0.26
	-0.11		

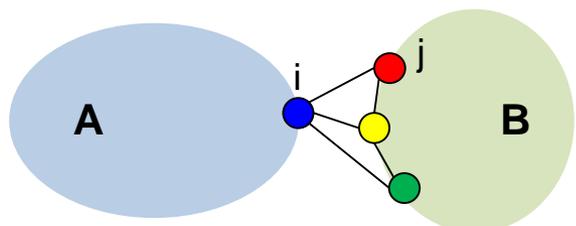
Find the most favorable
environment for each residue...

Residue

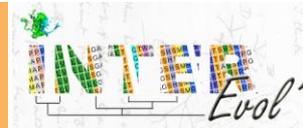


Development of InterEvScore, a docking score taking evolution into account

Inter-molecular interface contacts



Contact propensities
derived from InterEvol
(statistics on 1,289 interfaces)

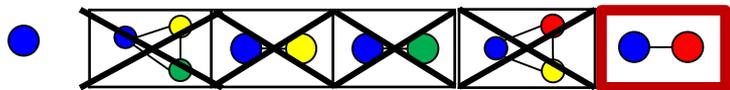


	0.53		0.04
	-0.05		-0.26
	-0.11		

Find the most favorable
environment for each residue...

...and take evolution into account

Residue



≥ 10 sequences

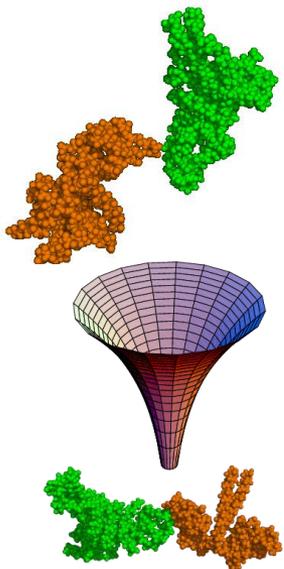
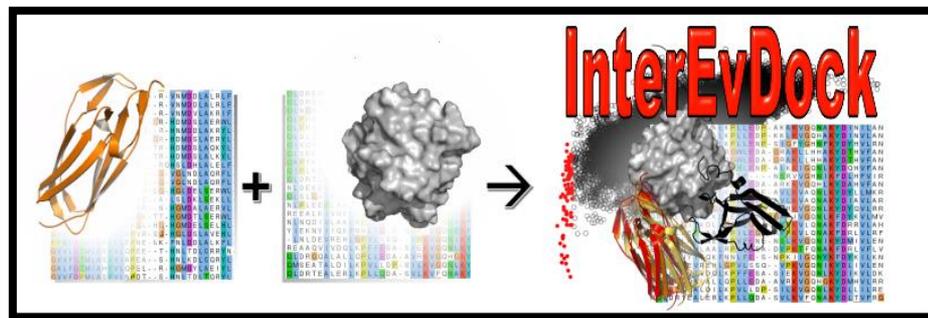
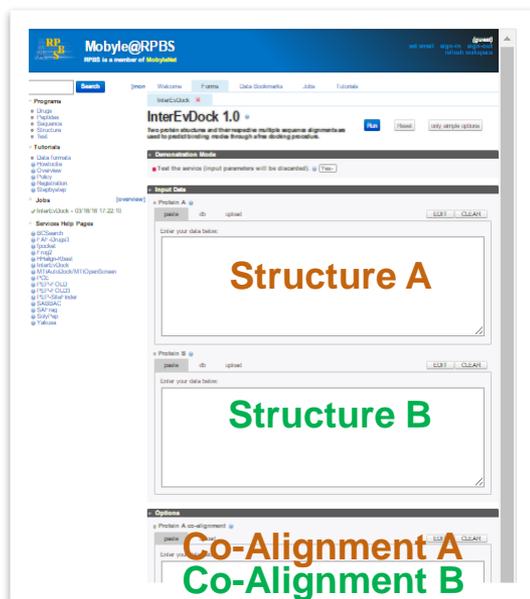
Alignment for A

Alignment for B

	Alignment for A					Alignment for B						
			i					j				
<i>H. sapiens</i>	○	○	●	○	○	○	○	○	●	○		0.53
<i>M. musculus</i>	○	○	●	○	○	○	○	○	●	○	+	0.53
<i>D. rerio</i>	○	○	●	○	○	○	○	○	●	○	+	0.53
⋮	○	○	●	○	○	○	○	○	○	○	+	0.03
	○	○	○	○	○	○	○	○	○	○	+	-0.20
<i>S. cerevisiae</i>	○	○	●	○	○	○	○	○	●	○	+	0.53

1.95

INTEREVDOCK SERVER TO ACCOUNT FOR CO-EVOLUTION INFORMATION IN DOCKING



Returns 10 models (~1h) selected from a consensus among the top solutions of :

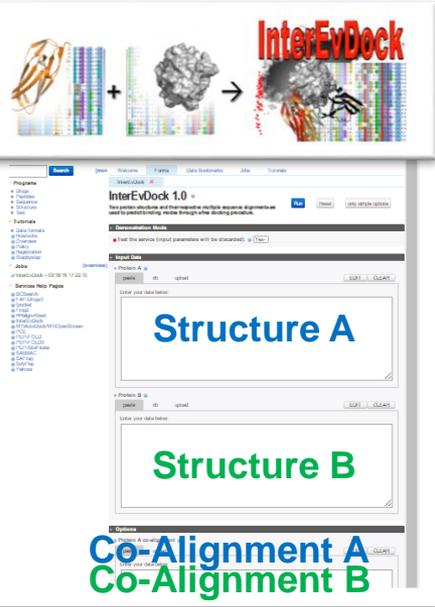
- **InterEvScore** (residue based / coevolution),
- **SOAP-PP** (statistical atomic based) (*A. Sali's lab*),
- **FRODOCK** (rigid-body + physics based) (*P. Chacon's lab*)

Yu et al, *Nucleic Acids Research* (2016)

Running at RPBS server (coll. P. Tufféry)

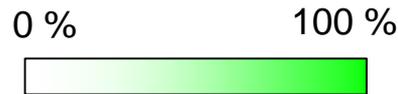
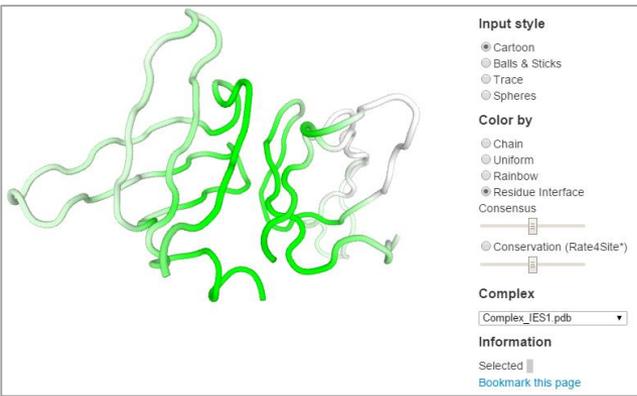
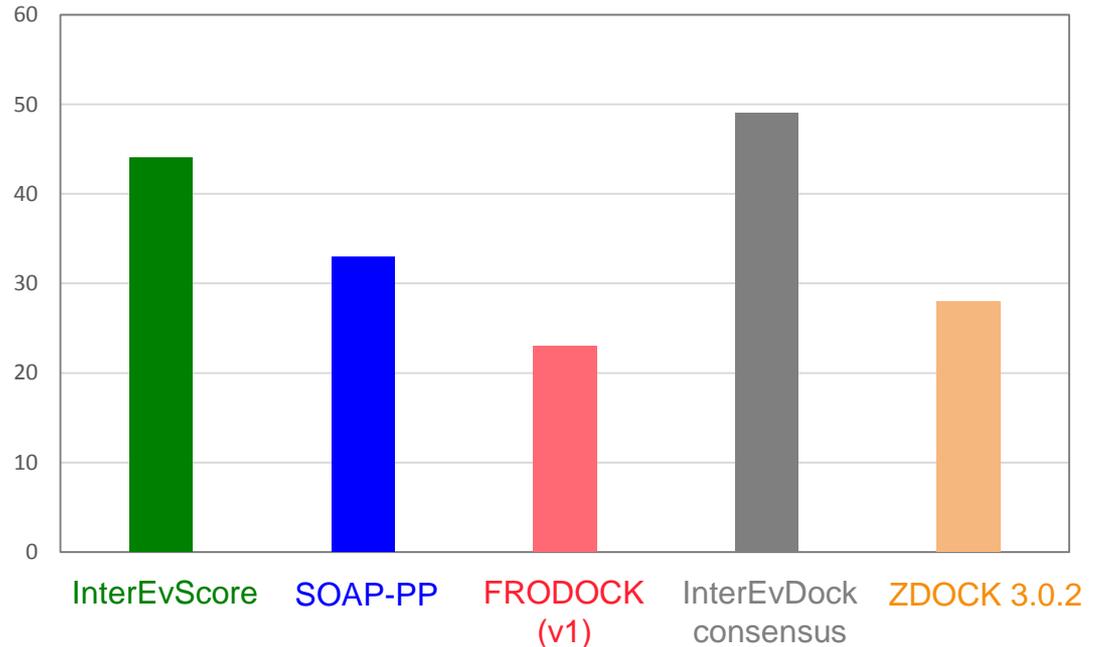
<http://bioserv.rpbs.univ-paris-diderot.fr/services/InterEvDock/>

INTEREVDOCK : A SERVER FOR PROTEIN RIGID-BODY DOCKING AND SCORING USING EVOLUTION



Top10 success rate (%)

Performance for 43 rigid-body docking cases from the Weng database for which evolutionary information can be gathered



Average % of interface implication among Top10 models

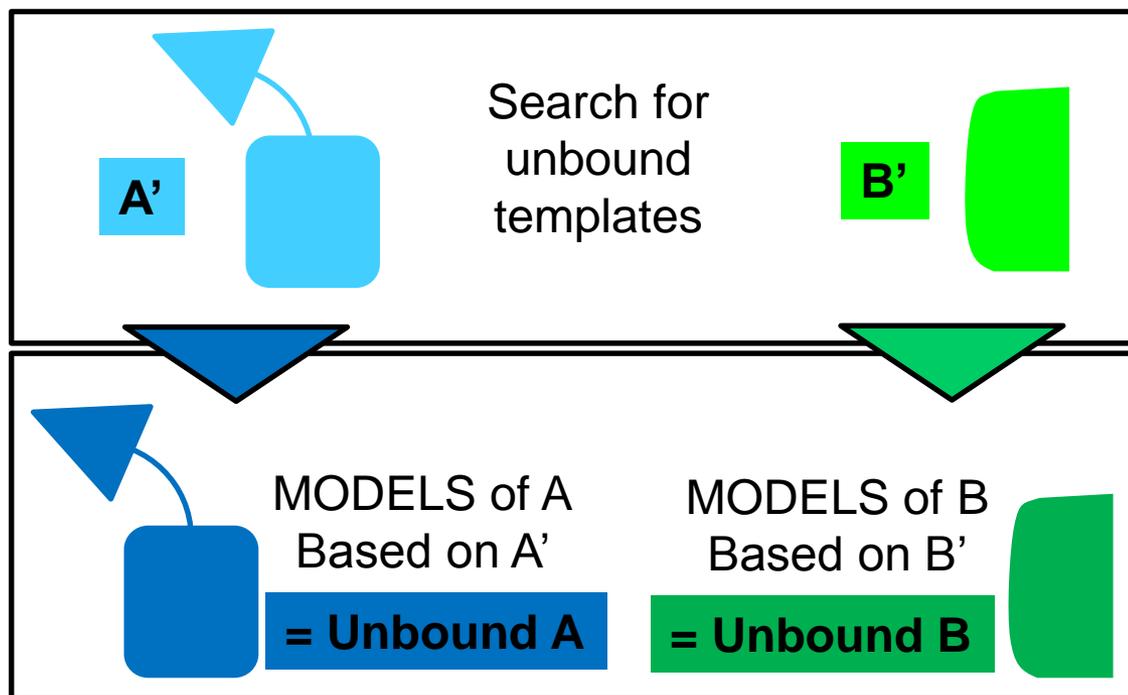
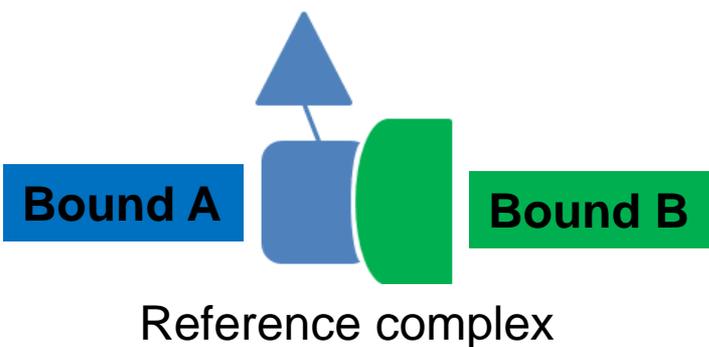
PREDICTOR GROUP PERFORMANCE BASED ON THE THREE MOST RECENT CAPRI EVALUATION MEETINGS.

CAPRI evaluation meeting, year and targets						
	2009, 12 targets		2013, 14 targets		2016, 20 targets	
Rank	Group	Success	Group	Success	Group	Success
1	Vajda/Kozakov	6/4 ^{***} /2 ^{**}	Bonvin	9/1 ^{***} /3 ^{**}	Guerois	10/1^{***}/8^{**}
2	Zacharias	6/4 ^{***} /1 ^{**}	Bates	8/2 ^{**}	Zacharias	10/3 ^{***} /2 ^{**}
3	Zou	6/3 ^{***} /2 ^{**}	Vakser	7/1 ^{***}	ClusPro	9/3 ^{**}
4	Eisenstein	6/3 ^{***} /1 ^{**}	Kozakov/ Vajda	6/2 ^{***} /3 ^{**}	Kozakov/ Vajda	8/3 ^{***} /2 ^{**}
5	Wolfson	6/3 ^{***} /1 ^{**}	Shen	6/1 ^{***} /3 ^{**}	Seok	8/3 ^{***} /2 ^{**}
6	Weng	6/2 ^{***} /2 ^{**}	Fernandez-Recio	6/1 ^{***} /3 ^{**}	Fernandez-Recio	7/1 ^{***} /3 ^{**}
7	Zhou	6/2 ^{***} /2 ^{**}	ClusPro	6/4 ^{**}	Zou	7/1 ^{***} /2 ^{**}
8	Bonvin	6/1 ^{***} /4 ^{**}	Zou	6/1 ^{***} /2 ^{**}	Weng	6/1 ^{***} /4 ^{**}
9	ClusPro	5/1 ^{***} /3 ^{**}	Zacharias	6/1 ^{***}	Vakser	6/2 ^{***} /2 ^{**}
10	Fernandez-Recio	5/2 ^{**}	Eisenstein	5/1 ^{***} /2 ^{**}	Bates	6/3 ^{**}

*Kozakov D et al (2017)
Nat. Protocols*

HOW TO GET A HIGHER NUMBER OF TARGETS ?

PPI4DOCK : A BENCHMARK FOR LARGE SCALE DOCKING ASSESSMENT BASED ON COMPARATIVE MODELS



PPI4DOCK contains 1417 docking targets

<http://biodev.cea.fr/interevol/ppi4dock/>

- Ensure MODELS represent unbiased unbound state.
- 822 docking targets can be used with coevolution information
- Success rate by Rigid-Body docking methods with these MODELS ?

Acknowledgements



Seydou Traoré
Arun Nadaradjane
Chloé Quignot

Raphaël Guerois

Françoise Ochsenbein

Gwenaëlle Moal

Laure Plancon-Arnould

Jinchao Yu

May Bakail

Danni Liu

Fouad Ouasti

Nadège Guinot

I2BC
Arnaud Martel

U. Paris 7
Julien Rey
Pierre Tufféry

**Structural Biology and
Radiobiology Lab**

I. Jacques Monod
Alessandro Berto
Valérie Doye

