Centre d'Orsay

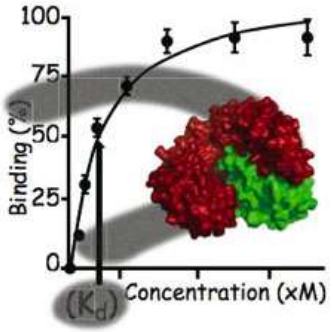
MACROMOLECULAR ASSEMBLY & INTERACTION NETWORKS
LAPTH, Annecy-le-Vieux, 15-16 Sept. 2011

IBBMC

Protein-protein interactions: modelling structure and affinity

Joël Janin

IBBMC UMR8619 CNRS
Université Paris-Sud
Orsay, France



Celebrating fifty years of allostery

Allostery is all about protein-protein interaction

1961

Changeux (1961) *Cold Spring Harbor Symp. Qu. Biol.* 26:313

Monod & Jacob (1961) *Cold Spring Harbor Symp. Qu. Biol.* 26:389

Monod, Changeux & Jacob (1963) *J. Mol. Biol.* 6:306

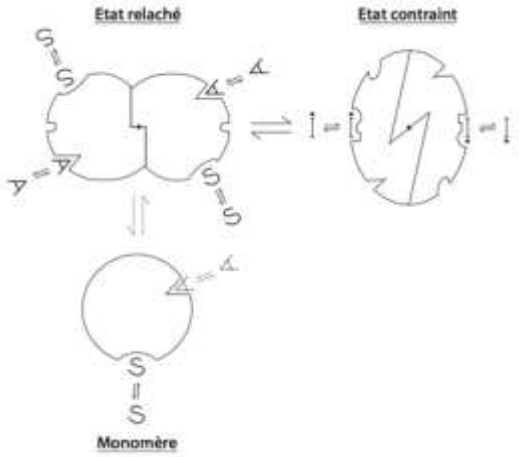
Monod, Wyman & Changeux (1965) *J. Mol. Biol.* 12:88

….

Changeux & Edelstein (2005) *Science* 308:1424

2011

Changeux (2011) *Protein Sci.* 20:1119



1.Protein-protein interaction (PPI)

in the Protein Data Bank (PDB)

2. Analyzing protein-protein interfaces

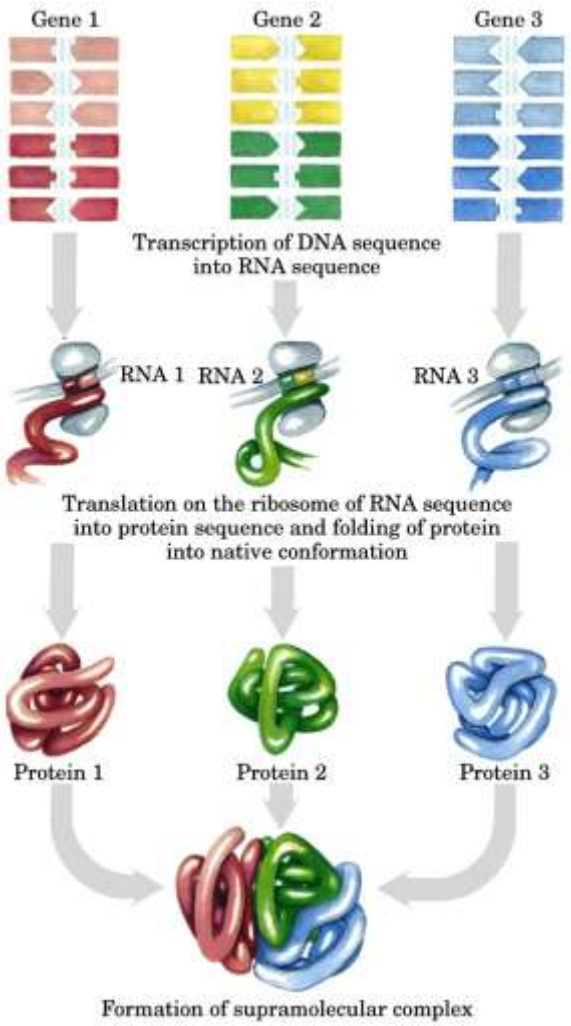
3. Modeling PPI: the CAPRI experiment

4. Modeling affinity: a structure/affinity benchmark

Protein-protein interaction (PPI), an essential component of biological structure and function

DNA codes for polypeptide chains, not proteins !

Most proteins are multi-chain assemblies, held together by non-covalent PPI



Quaternary structure:

- In bacteria, >50% proteins are ***oligomeric***; they contain 2 to 12 polypeptide chains.
- In yeast, >50% proteins are part of ***stable assemblies*** that contain 2-15 chains

Transient assemblies:

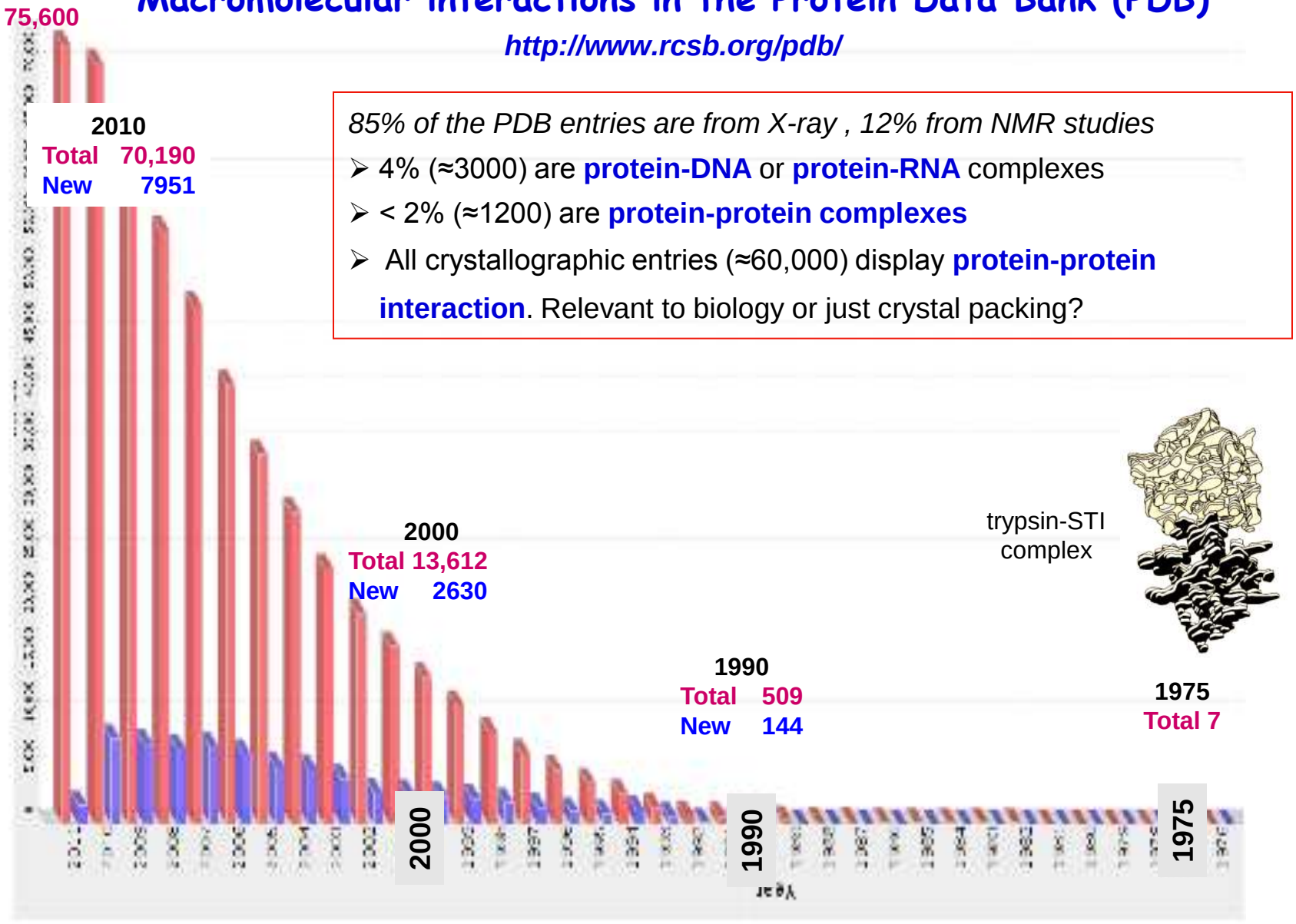
- proteins A and B fold and assemble separately
- A and B "recognize" each other and form a non-covalent complex C when they meet

Examples

- Antigen-antibody recognition
- Enzyme-inhibitor interaction

Macromolecular interactions in the Protein Data Bank (PDB)

http://www.rcsb.org/pdb/



85% of the PDB entries are from X-ray , 12% from NMR studies

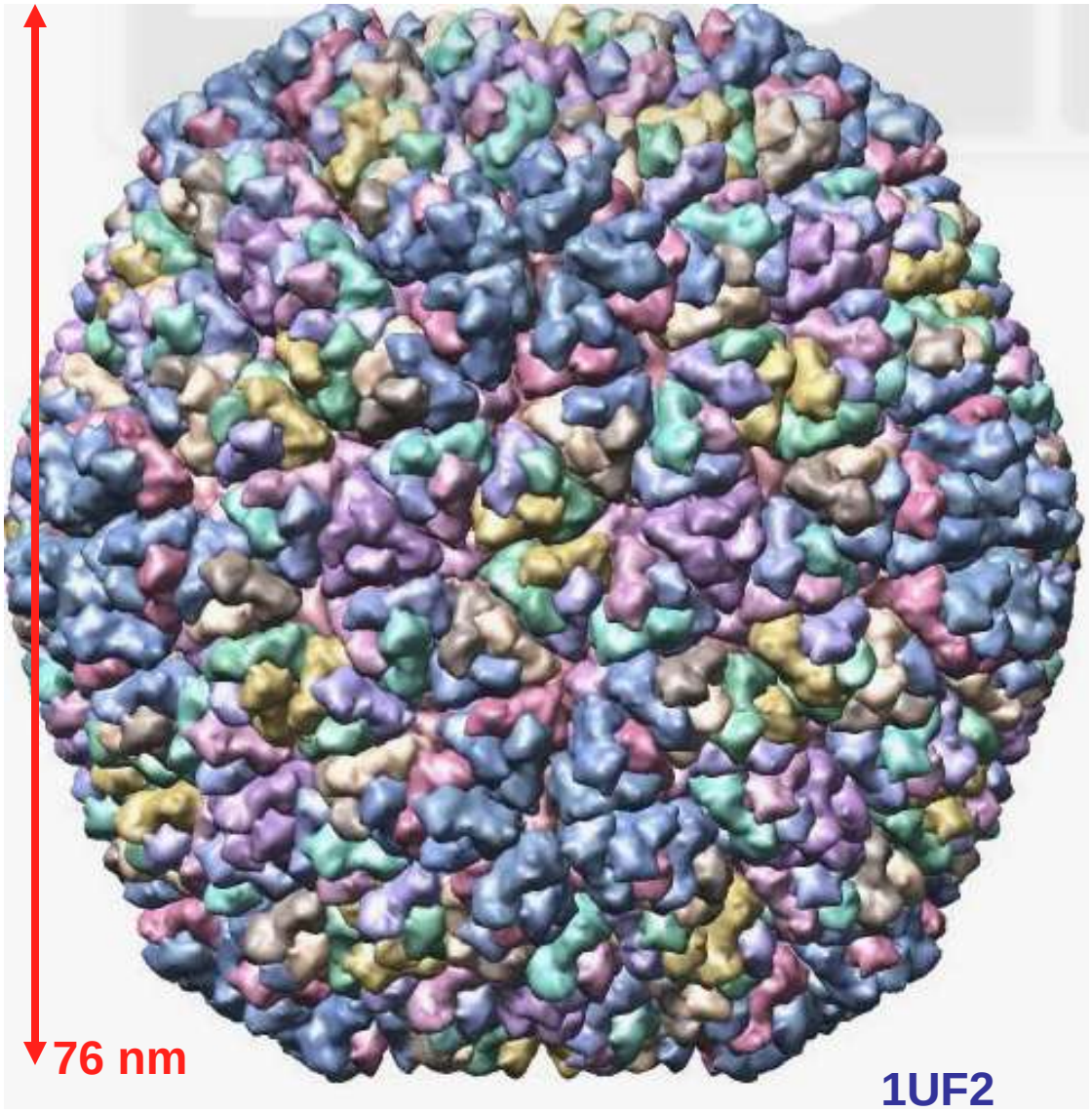
- 4% (≈3000) are **protein-DNA** or **protein-RNA** complexes
- < 2% (≈1200) are **protein-protein complexes**
- All crystallographic entries (≈60,000) display **protein-protein interaction**. Relevant to biology or just crystal packing?

trypsin-STI complex

Large protein assemblies : Rice dwarf virus

- Rice dwarf virus is **the largest biological assembly known at the atomic level**
- It has 780 polypeptide chains, plus the RNA, and a total momecular weight
  ≈ **70 millions Da**,
- The first virus X-ray structure was Tomato Bushy Stunt Virus (Harrison et al., 1978)

Trypsin /STI
on the same scale



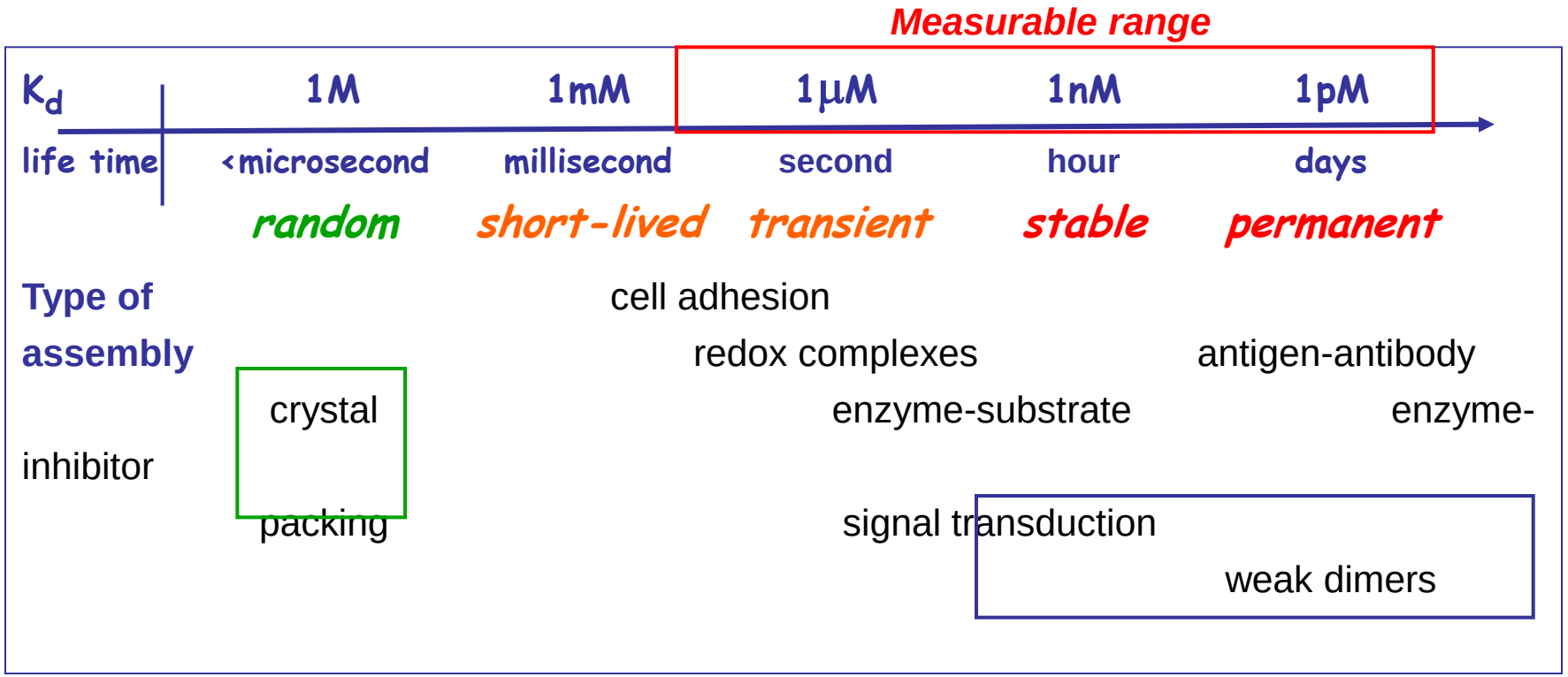
Nakagawa et al. & Tsukihara (2003)

PPI in real life: the scale of time and affinity

Affinity is defined by the dissociation equilibrium constant K_d, or the free energy $\Delta G_d = -RT \ln K_d / c^\circ$

It determines:

- which assemblies **actually exist** in a cell or an organism
- whether an assembly is **permanent or transient** (and can exchange components)



Oligomeric proteins (Monod, Wyman & Changeux, 1965)

In the cell, most proteins are oligomeric

- **Homo-oligomers:** the polypeptide chains have the same sequence; they come from the same gene
- **Hetero-oligomers:** the polypeptide chains have different sequences; they come from different genes

In the PDB, 62% of the proteins are oligomeric

#chains			Homo	Hetero
1	Monomers	38%		
2	**Dimers**	**35%**	**30%**	5%
3	Trimers	6%		
4	Tetramers	12%	9%	3%
5	Pentamers	0.5%		
6	Hexamers	4%		
7	Heptamers	0.1%		
8 to 900		4%		

3D-Complex database: *http://supfam.mrc-lmb.cam.ac.uk/elevy/3dcomplex/*

Levy , Pereira-Leal, Chothia & Teichmann (2006)

Analyzing PPI in the PDB

The information in the PDB is **chemical** and **geometric**.

What does it tell us about the **physics** (thermodynamic stability) and the **biology** (functional relevance and evolution) of PPI ?

Validated sets of PDB entries allow us make the best use of this information.

Validated set	set size	ref.
Transient hetero complexes	75 to 144	Lo Conte et al. (1999); Janin et al. (2008) Kastritis et al. (2011)
Homodimers	315	Janin et al. (1988); Dey et al. (2010)
" **weak**	42	Dey et al. (2010)
Large crystal contacts (monomers)	188	Bahadur et al. (2004)

1. Protein-protein interaction (PPI)

   in the Protein Data Bank (PDB)

2. Analyzing protein-protein interfaces

3. Modeling PPI: the CAPRI experiment

4. Modeling affinity: a structure/affinity benchmark

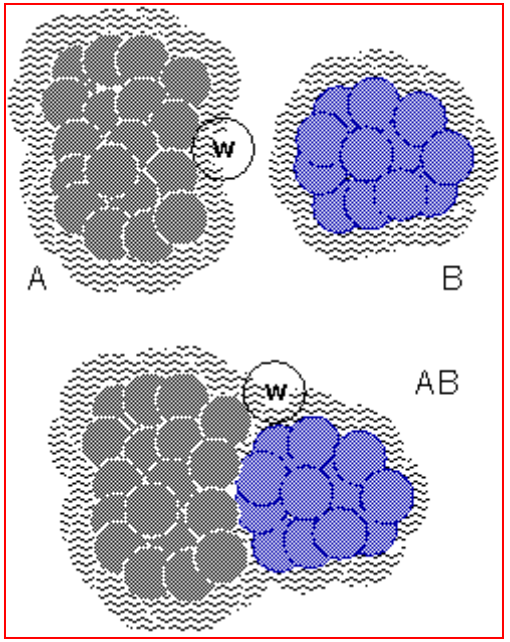
ASA as a tool to analyze PPI in structures

ASA **solvent accessible surface area**

measures protein-solvent contacts

Lee & Richards (1971)

$$BSA = ASA_A + ASA_B - ASA_{AB}$$



BSA **buried surface area (=interface area)**

measures protein-protein contacts

Chothia & Janin (1975)

The hydrophobic effect

The free energy of desolvating non-polar (aliphatic or aromatic) groups scales linearly with their ASA

$$\Delta G_{np} = \gamma\, ASA$$

Chothia (1974) $\gamma = 24$ cal/mol.Å^2

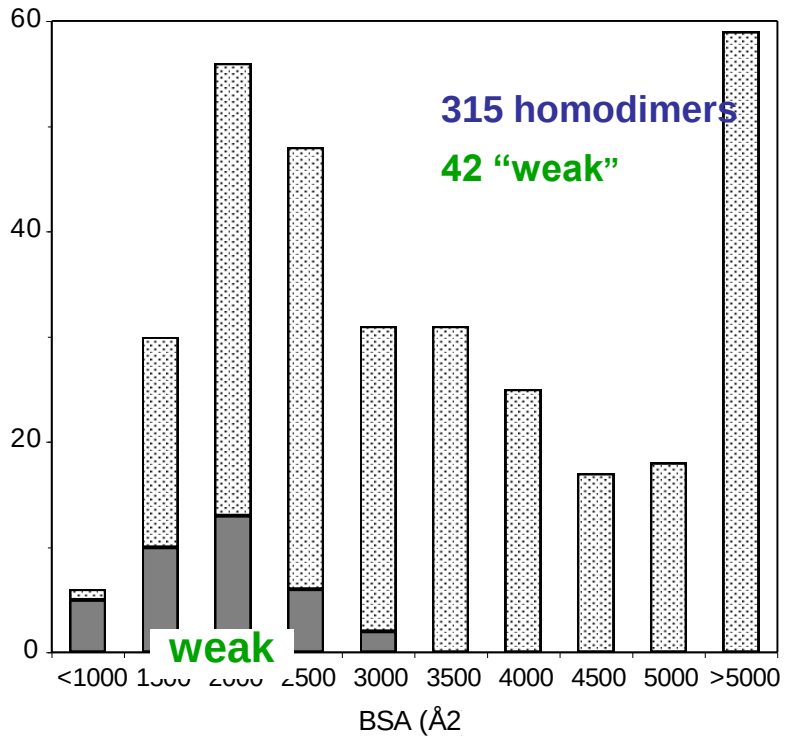
Other accepted values $\gamma =$ 30 to 50 cal/mol.Å^2

BSA and ΔG_{np} : do they explain the stability of PPI ?

A transient protein-protein complex typically buries 900 Å^2 of non-polar protein surface.

Thus: ΔG_{np} **> 20 kcal/mol** in favor of association !!!

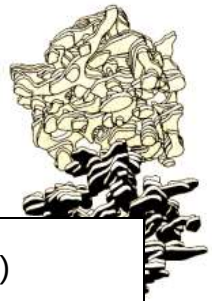
Interface size and stability in homodimers

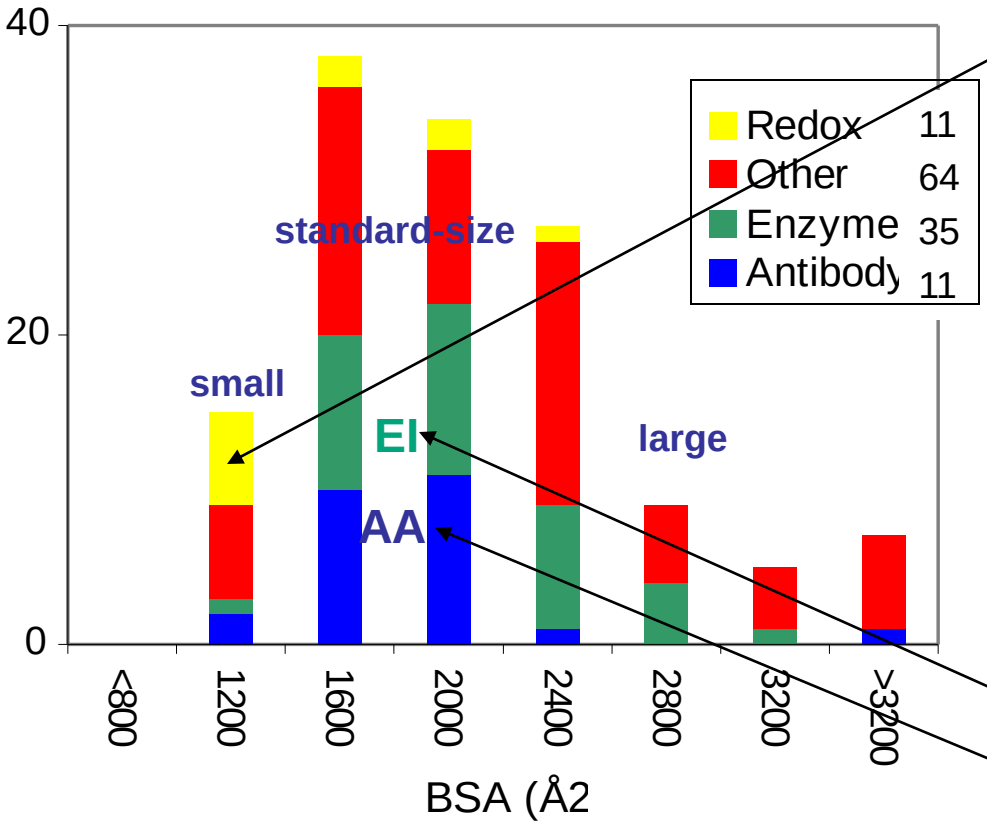


- Most homodimeric proteins assemble while their subunits fold; their assembly is **permanent** and stable in solution
- Most also have **large interfaces**: 72% have a *BSA*>2000 Å^2 , up to 10,000 Å^2
- In homodimers (and in transient complexes), the **smallest interfaces** have a *BSA* ≈900 Å^2
- Nearly all the homodimers with *BSA* <1000 Å^2 are **weak:** in solution, they are in equilibrium with the monomers, with K_d in the range 10^{-8} to 10^{-5} M.

Dey, Chakrabarti & Janin (2010) *JMB* 398:146

Interface size in transient complexes





124 transient protein-protein complexes

Janin, Bahadur & Chakrabarti (2008)

Quat. Rev. Biophysics 2:133-180.

WEAK: Redox (electron transfer) complexes make short-lived interactions; most have a **small interface:**

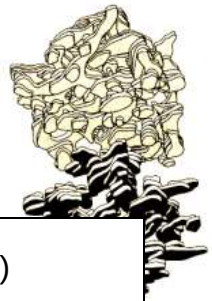
BSA = 900-1200 Å^2 0-3 H-bonds

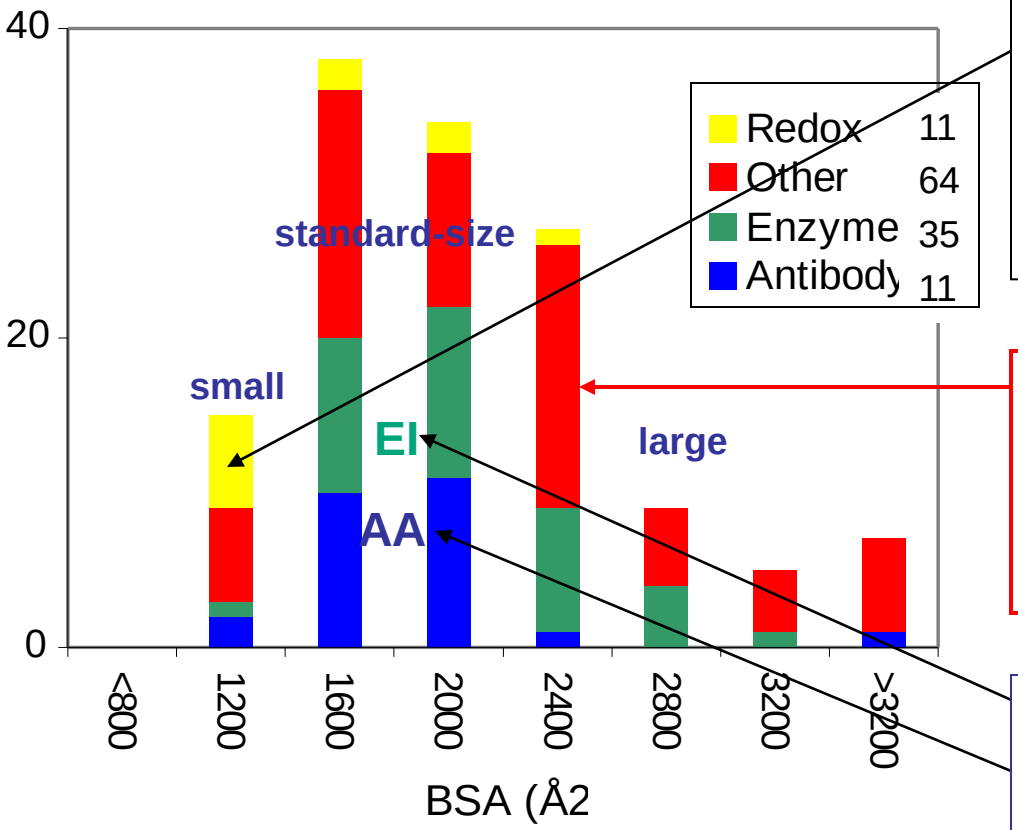
Crawley & Carrondo (2004))

STRONG: Enzyme/Inhibitor and **Antigen/Antibody** complexes are long-lived and highly specific; most have a **standard-size** interface

BSA = 1200-2000 Å^2 6-15 H-bonds

Interface size in transient complexes





124 transient protein-protein complexes
Janin, Bahadur & Chakrabarti (2008)
Quat. Rev. Biophysics 2:133-180.

WEAK: Redox (electron transfer) complexes make short-lived interactions; most have a **small interface:**

BSA = 900-1200 Å^2 0-3 H-bonds

Crawley & Carrondo (2004))

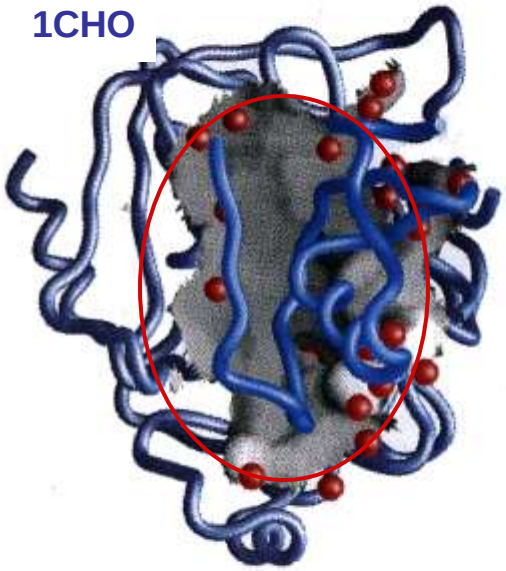
Signal transduction complexes are often short lived. Yet, many have a **large interface**:

BSA >2000 Å^2 .

STRONG: Enzyme/Inhibitor and **Antigen/Antibody** complexes are long-lived and highly specific; most have a **standard-size** interface

BSA = 1200-2000 Å^2 6-15 H-bonds

Rigid-body *vs.* flexible recognition



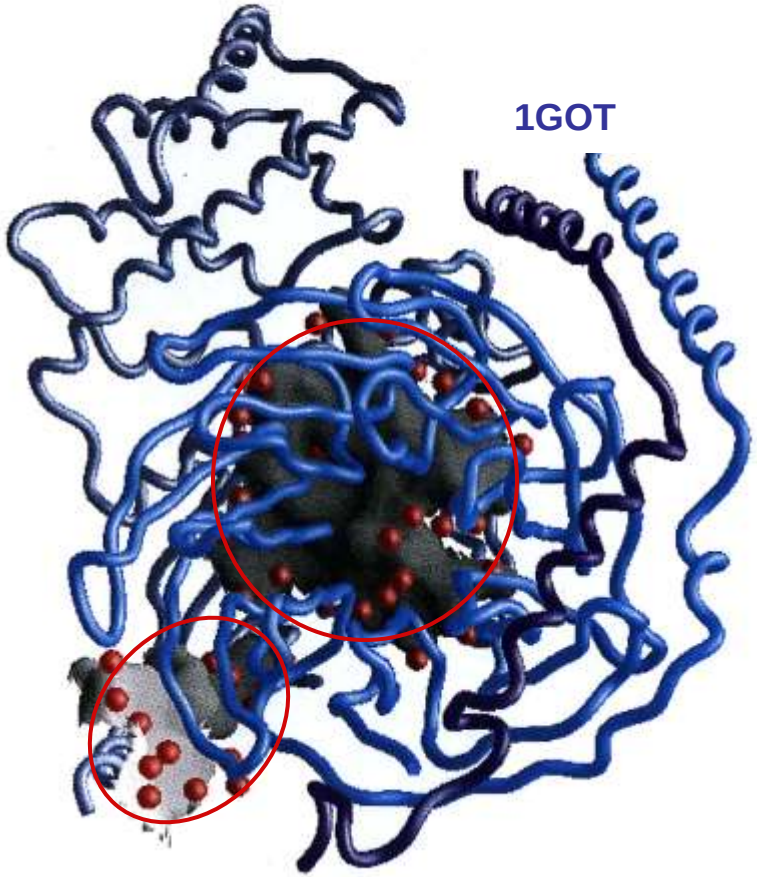
Rigid-body recognition:

chymotrypsin-inhibitor complex

A standard-size, single patch interface:

BSA = 1470 Å^2

Very little change in conformation between the free and bound proteins: 0.6 Å Cα RMS

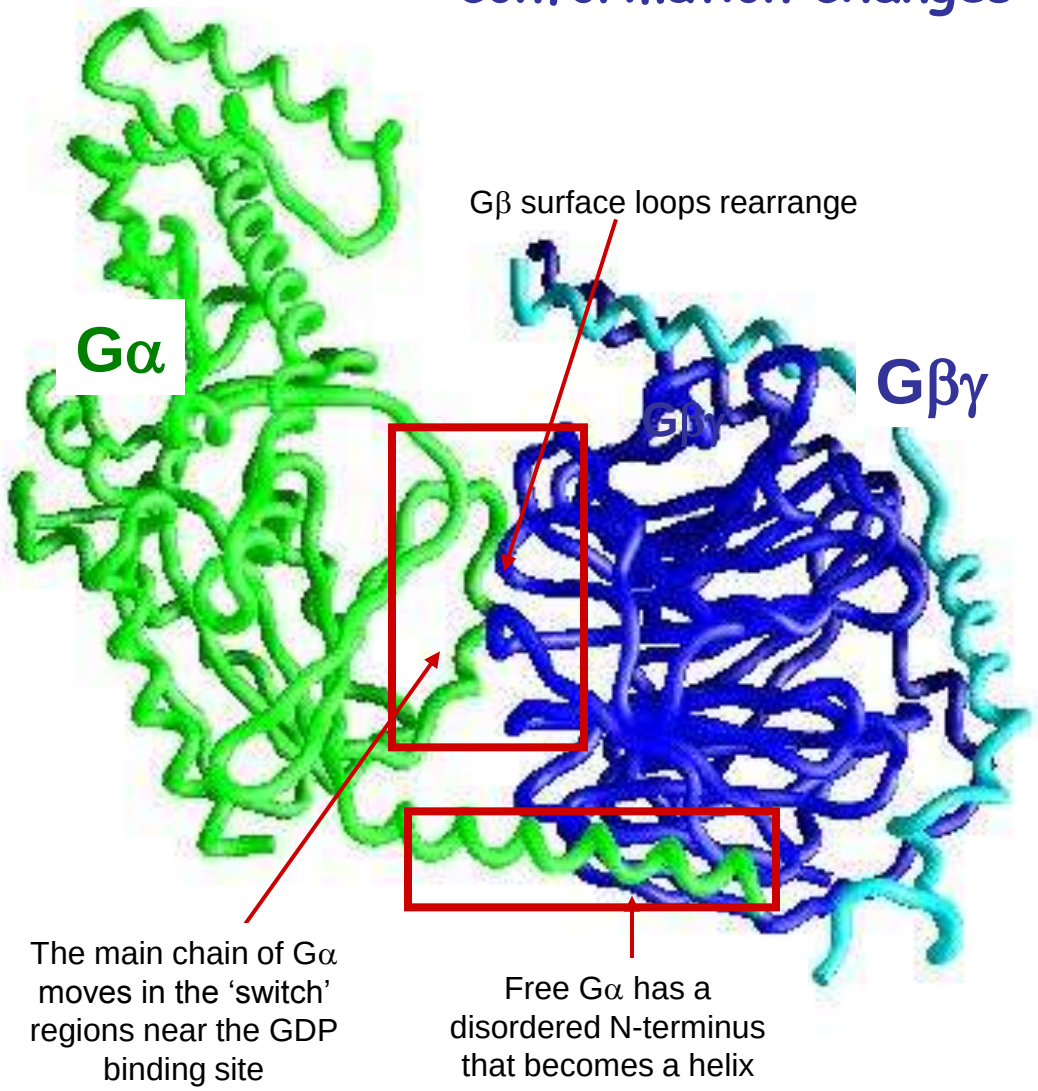


Flexible recognition: Transducin Gα-Gβγ

A large interface (*BSA* =2500 Å2) in two patches.

major conformation changes (1.8 Å Cα RMS)

Signal transduction, large interfaces and conformation changes

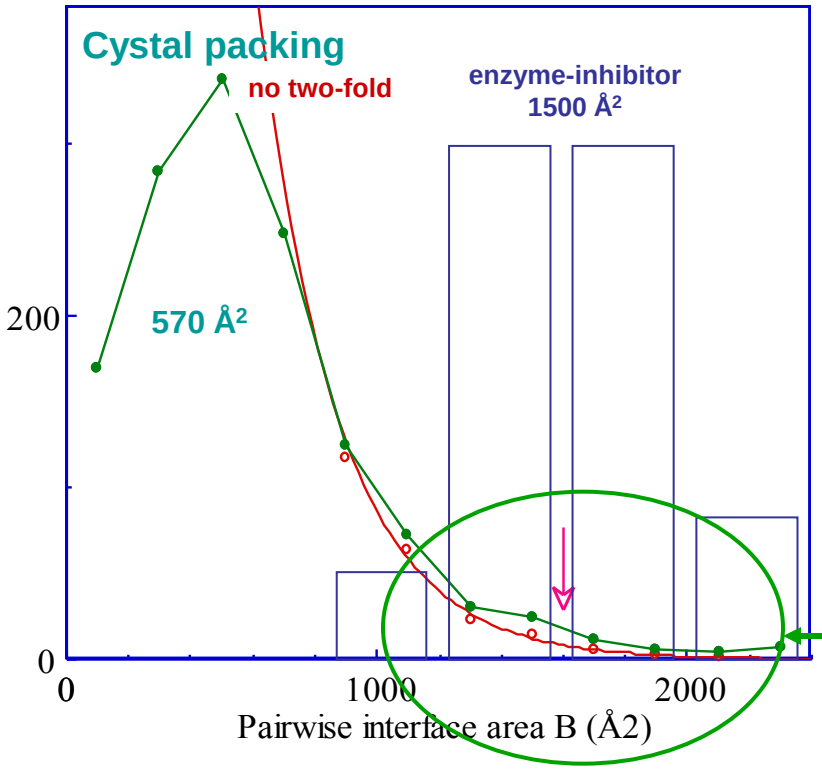


The retina, a paradigm of signal transduction

The visual signal is initiated in **rhodopsin** as retinal absorbs a photon.
Rhodopsin interacts with the **transducin Gα** subunit.
Gα binds and hydrolyzes GTP, then associates with Gβγ.
Gβγ can also interact with the cyclase that produces the chemical signal cyclic-GMP.

X-ray structures by the group of the late P. Sigler (Yale)

Non-specific PPI at crystal contacts



Most crystal packing interfaces are much smaller than biological interfaces: their average BSA is only 570 Å^2

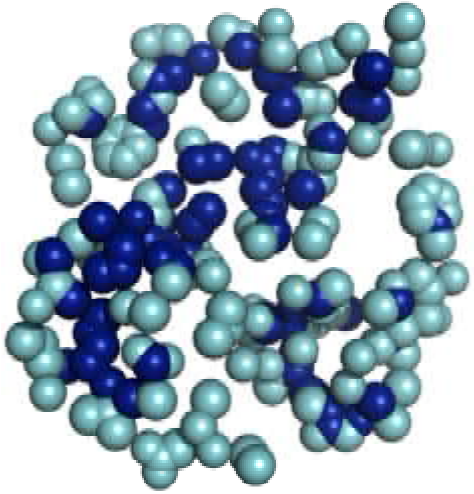
But there are many exceptions: large crystal packing interfaces, which cannot be distinguished from real interface on the basis of their size only. They illustrate **non-specific PPI**

In crystals with no two-fold symmetry, packing interfaces with *BSA*>800 Å^2 follow an extreme value distribution. They are more frequent in crystals with two-fold symmetry, where they form **crystal dimers**, often mistaken as biological.

Janin (1997) *Nature Struct Biol* 4:973

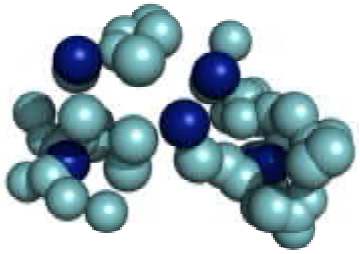
Bahadur et al. (2004) *JMB* 336:943

Atomic packing at homodimer and crystal packing interfaces



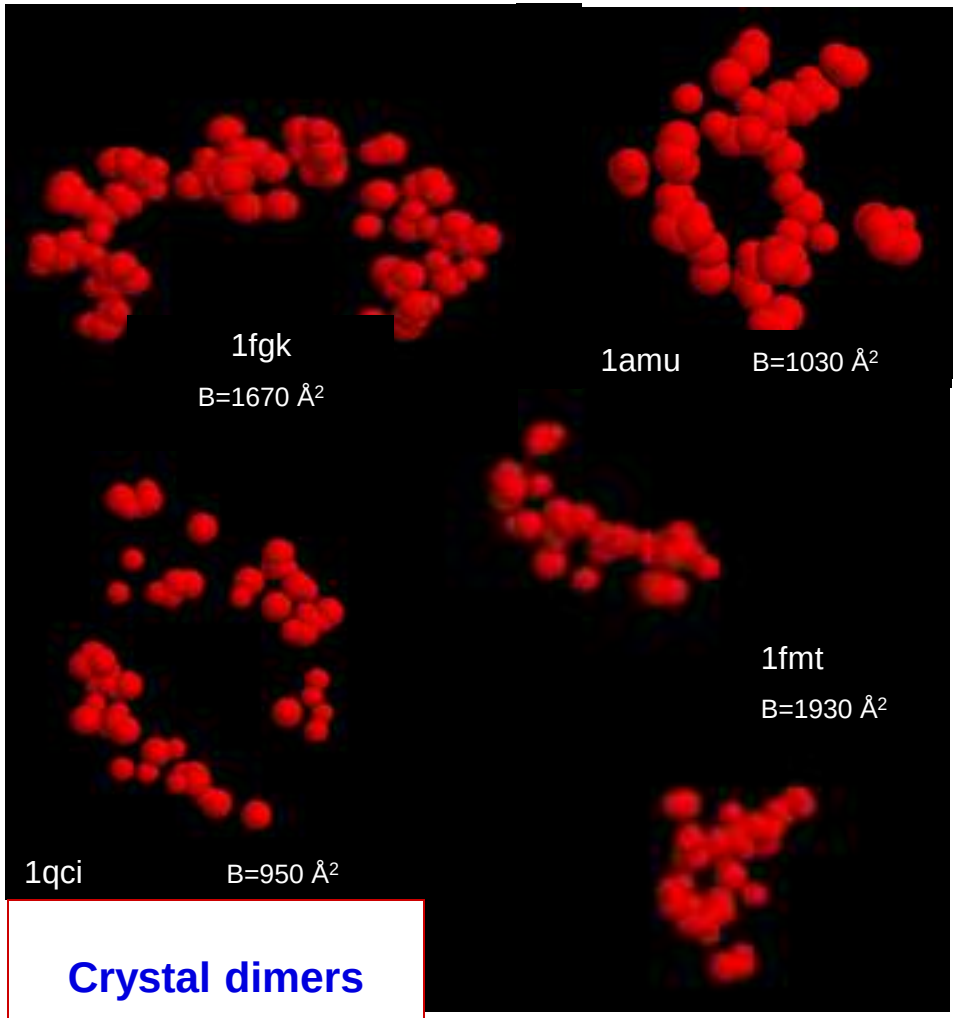
A strong homodimer

1ctt cytidine deaminase



A weak homodimer

1b8e β-lactoglobulin

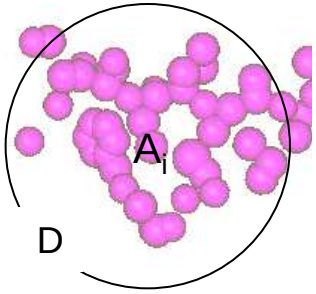


Crystal dimers

Weak protein-protein interfaces are loosely packed

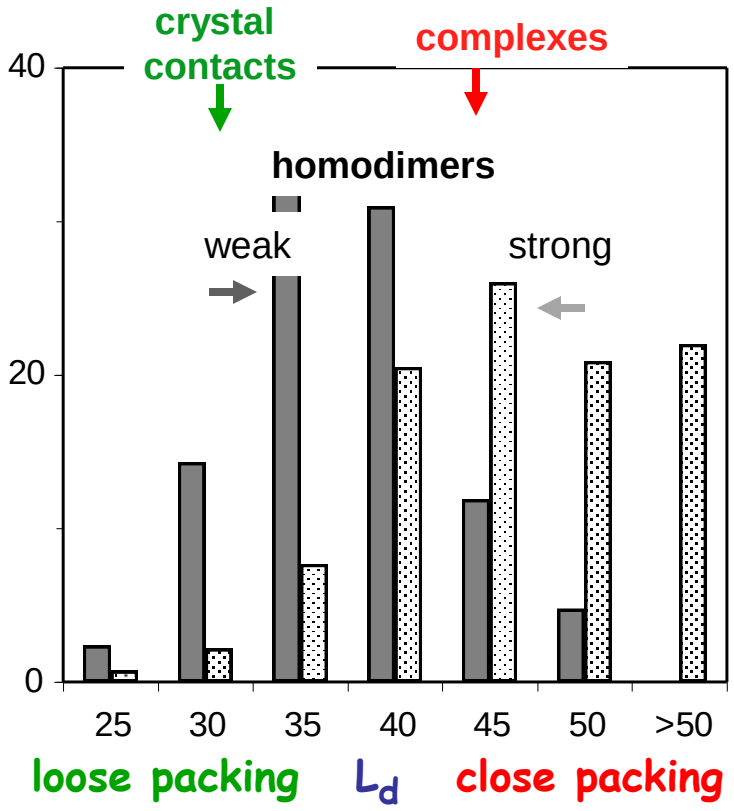
L_d local density index (Bahadur et al., 2004)

- count the number of interface atoms within D = 12 Å of a given interface atom A_i
- **L_d** is the average over all interface atoms A_i.



Interface	L_D
homodimers	
strong	43±8
weak	35±5
transient complexes	42
crystal contacts	32

Dey et al. 2010



Conclusion (part 2)

There is a relation between *stability* and *interface size*

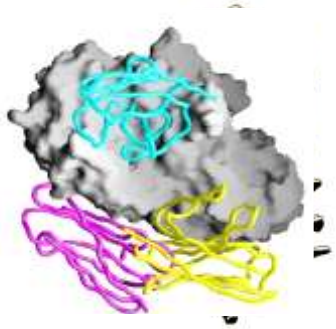
- biologically relevant interfaces have a minimum size with a *BSA* ≈ 900 Å^2
- small interfaces (*BSA* ≈1000 Å^2) form weak homodimers and short-lived complexes
- standard-size interfaces (*BSA* =1200-2000 Å^2) yield stable, specific assemblies

…but it is often masked by *conformation changes* and/or the *atomic packing*

- large conformation changes always accompany the formation of large interfaces
- stable assemblies (transient complexes and strong homodimers) have close-packed interfaces
- weak homodimers and crystal packing contacts have loosely packed interfaces.

Other features relevant to stability

- The interface is enriched in **hydrophobic** (aromatic/aliphatic) groups relative to the free protein surface In homodimers, but not in transient complexes; it is depleted of electric charges.
- The interface core has a specific **amino acid composition**; the rim is like the protein surface
- Residues of the interface core are **conserved in evolution**; the rim is not conserved



1.Protein-protein interaction (PPI)

in the Protein Data Bank (PDB)

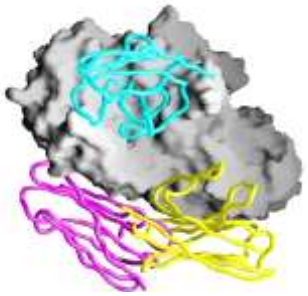
2. Analyzing protein-protein interfaces

3. Modeling PPI: the CAPRI experiment

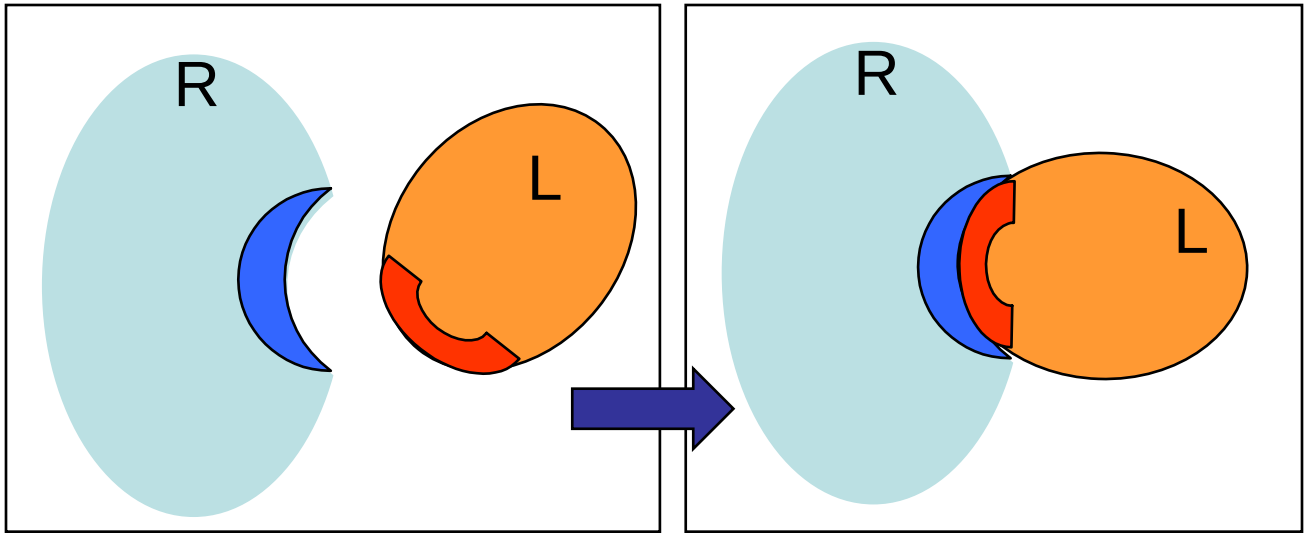
4. Modeling affinity: a structure/affinity benchmark

Modeling protein-protein interaction by docking

Wodak & Janin (1978)

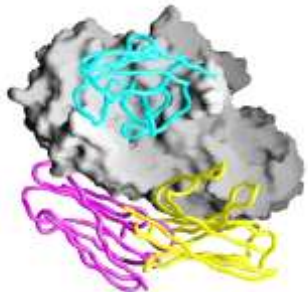


Given the 3D structure of proteins R and L (*Receptor* and *Ligand*),
build the R/L complex



- If R and L are in ***bound*** conformation (=taken from the complex), docking has only six degrees of freedom
- but in real life, proteins *change conformation* when they associate
- To be realistic, the prediction must start from ***unbound*** (=free) protein structures

Macromolecular docking: the procedure



Molecular description

(surface/volume, interactions)

- make molecules « soft »
- **generate conformers**

6D rigid-body search

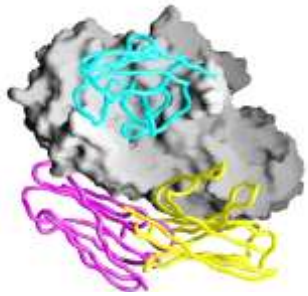
(rotations/translations)

Scoring

rank docking models and drive false positives out

- empirical potentials
- energy refinement (MM, MD)
- biochemical data, sequence conservation etc…

What can docking do for you?



A reliable* and efficient* macromolecular docking procedure complements biochemical and structural studies

- allow building complex assemblies from the known structure of their components
- tell meaningful interactions apart from crystal contacts
- **help in the interpretation of low resolution (EM, SAXS) data**

and possibly:

- suggest possible interactions that have not been detected experimentally
- Inform on the nature and strength of the interaction

CAPRI is a community-wide experiment to test docking procedures in blind predictions

Assessing structural predictions in community-wide experiments:

CAPRI and CASP

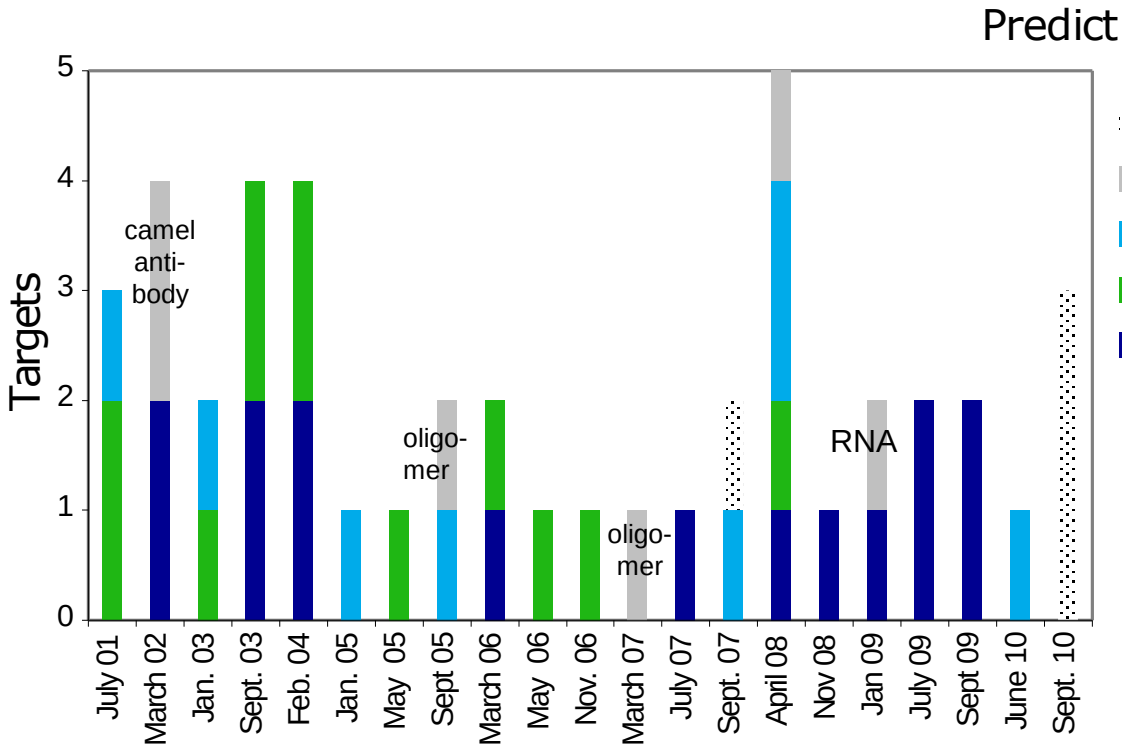
- **CASP (Critical Assessment of methods of Structure Prediction)**:
  - predict the mode of **folding** of a protein based on the amino acid sequence
  - compare to an unpublished X-ray or NMR structure.
  - J. Moult (CARB, Rockville MD) launched CASP in 1994
  - round of predictions once every two years (CASP9 in 2010) with >100 targets
- **CAPRI (Critical Assessment of PRedicted Interactions)**:
  - predict the mode of **recognition** of two proteins by docking their 3D structures
  - compare to unpublished X-ray structures of **protein-protein complexes**.
  - CAPRI started in 2001
  - Targets are few: a round of prediction begins any time one is made available

http://capri.ebi.ac.uk/

The CAPRI star system

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

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



70% of the targets have had good predictions. **15%** failed:

- oligomeric proteins
- unbound RNA
- two antigen/antibody

T37: Signalling & membrane traffic

Arf6 / LZ2 of JNK-interacting protein 2

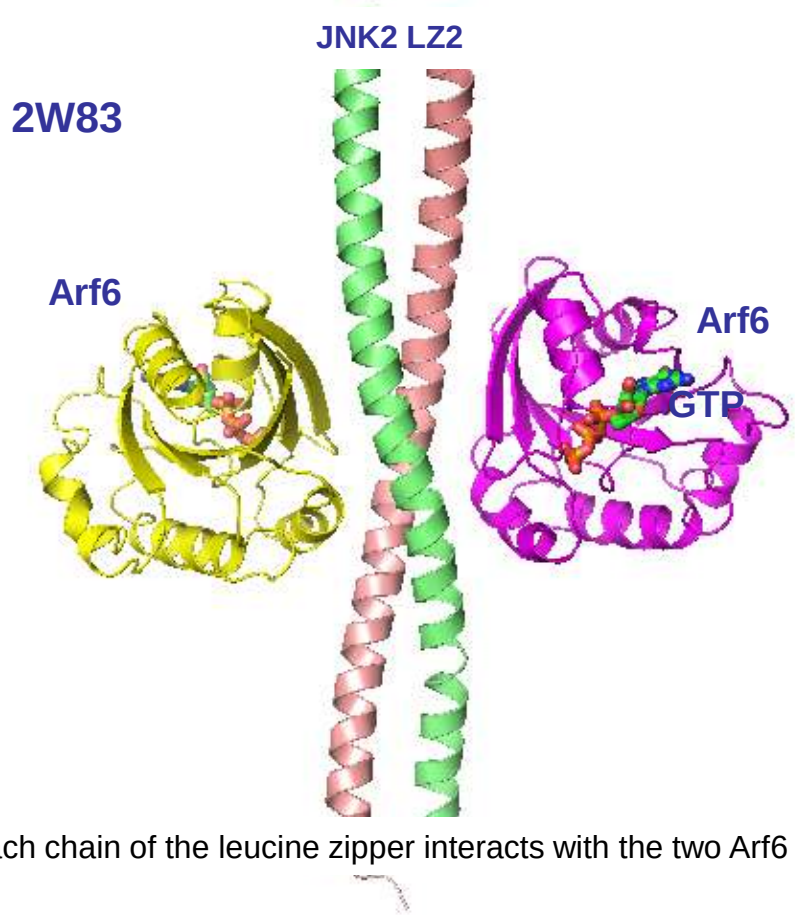
Isabet et al. 2009 *EMBO J.* 28:2835

T37 is a gift of J. Ménetrey (Institut Curie, Paris)

Round 16

Nov. 2008 47 predictors, 18 scorers

Model quality	High	Med.	Accept
Predictors	1	7	13
Scorers	**5**	13	11

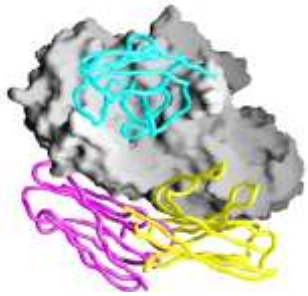


Each chain of the leucine zipper interacts with the two Arf6

New in CAPRI !

LZ2 had to be model-built from the non-homologous GCN4 leucine zipper.

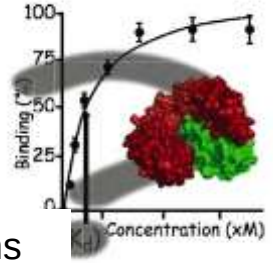
A great success for predictors and scorers!



The current state of docking methods

- Make docking algorithms more efficient
- Improve the scoring functions for docking, identifying correct models, and refinement
- Use non-structural information more efficiently (sequence, mutation data etc…)
- Predict and simulate conformation changes
- **Model and predict affinity**

Now, predict affinity ?

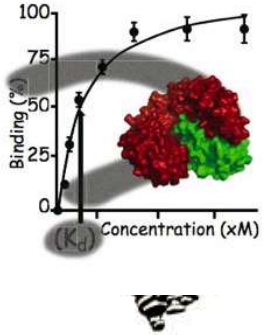


David Baker and Sarel Fleishman (University of Washington, Seattle) engineer proteins that form novel protein-protein interactions. They make gene constructs, express on the surface of yeast cells, and test if they interact with target proteins (Fleishman et al. 2011, *Science* 332:816)

David and Sarel gave us large sets of designed models, and challenged CAPRI participants to **predict which ones make a stable interaction** (very few actually do).

- CAPRI Round 20 (Feb. 2010): 42 designs, one that binds
- CAPRI Round 21 (April-June 2010): 87 designs, one that binds

38 predictors CAPRI groups participated
the results were essentially random…
Scoring functions designed to rank docking models of proteins known to bind fail to predict whether two proteins bind or not.
To do that, we must predict affinity !



1. Protein-protein interaction (PPI)

   in the Protein Data Bank (PDB)

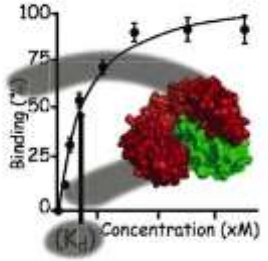
2. Analyzing protein-protein interfaces

3. Modeling PPI: the CAPRI experiment

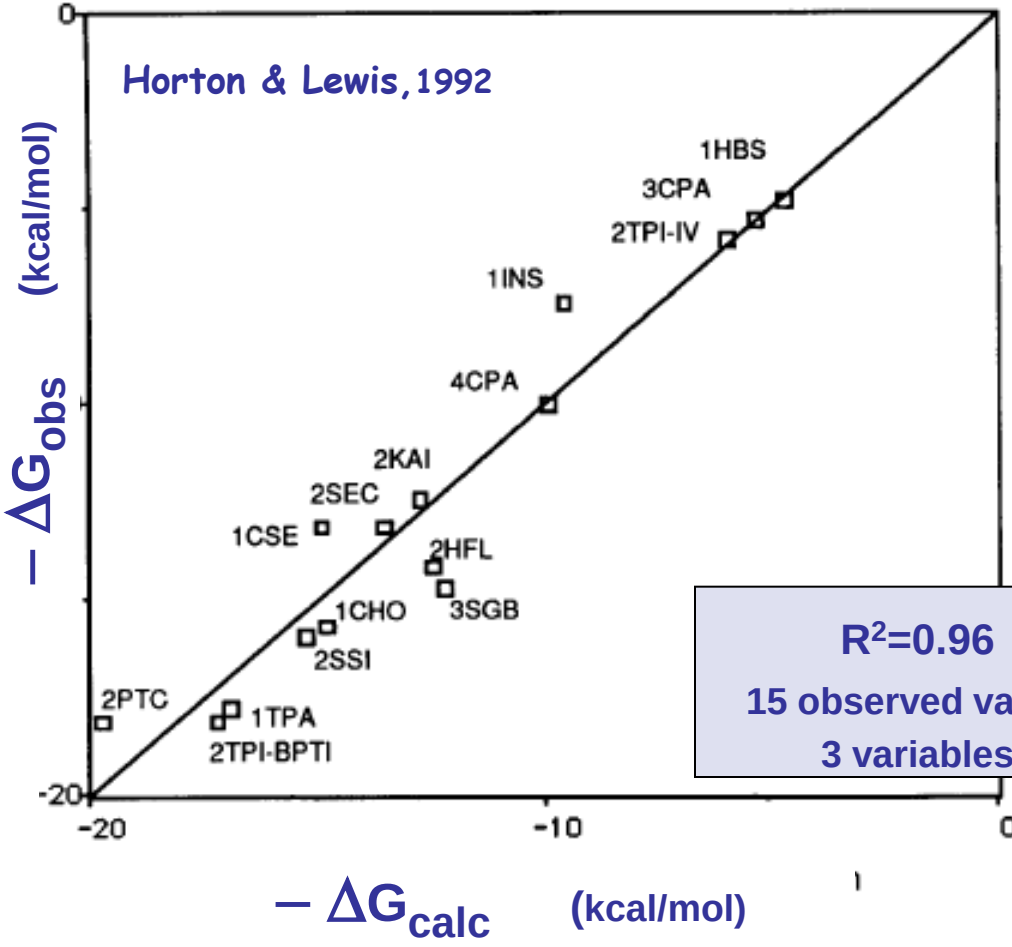
4. Modeling affinity: a structure/affinity benchmark

Modeling ΔG from structure

Horton & Lewis (1992)



$$\Delta G_{obs} = -RT \ln (K_d/c^\circ) \qquad \Delta G_{calc} = \alpha \Delta G_{np} + \beta \Delta G_{pol} + \Delta G_{rt}$$



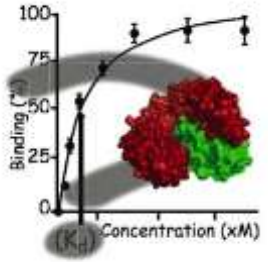
R^2=0.96

15 observed values

3 variables

- $\Delta G_{np} \approx 25$ cal/mol.Å^2
  similar to Chothia (1974)
- ΔG_{pol} based on the atomic desolvation coefficients of Eisenberg & Mclachlan (1986)
  - $\beta = -1.2$ **wrong sign !**
- $\Delta G_{rt} = -6$ kcal/mol
  (external degrees of freedom)

Later attempts to fit ΔG didn't do so well…

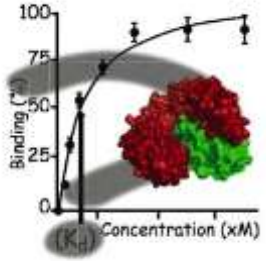


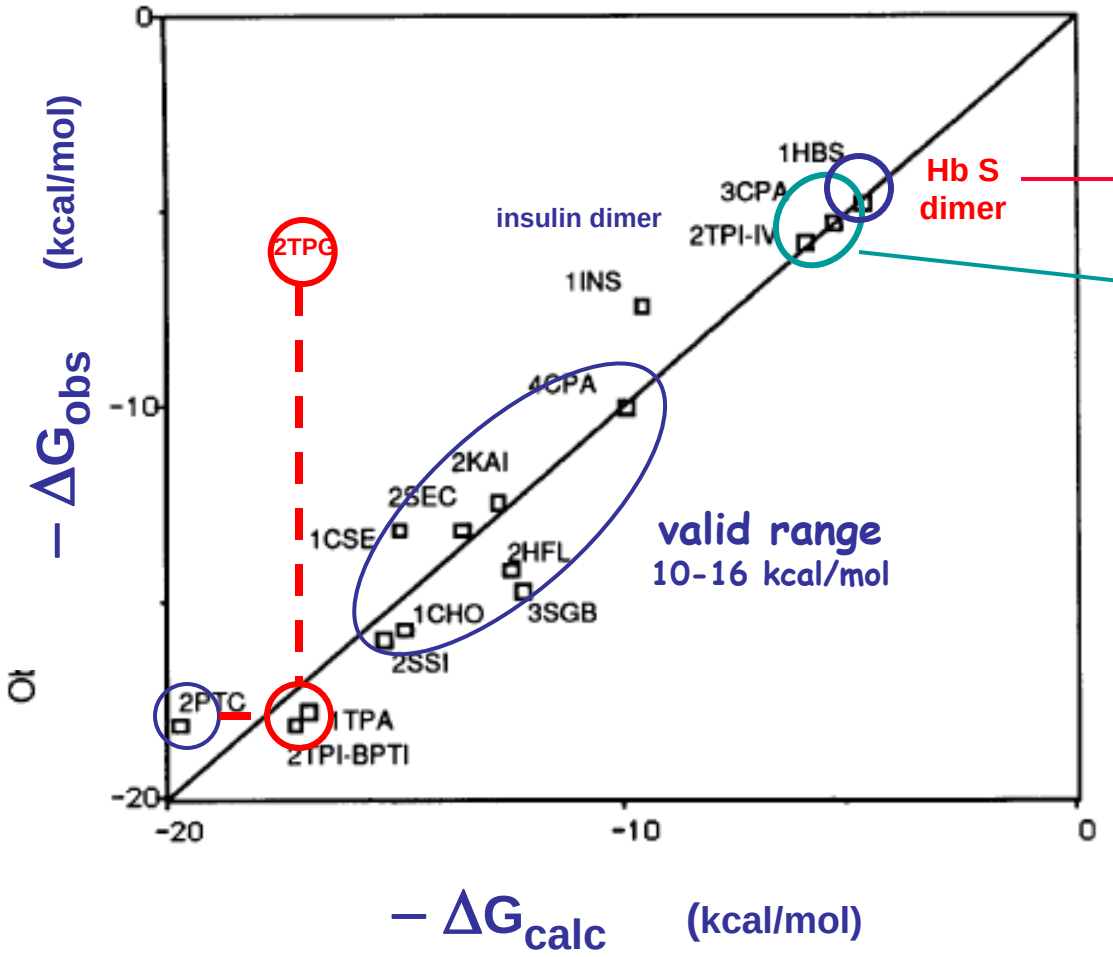
		Sample size	R^2 correl. coeff	$\langle \Delta G_{calc} - \Delta G_{obs} \rangle$ (kcal/mol)
Horton & Lewis (1992)		15	0.96	0.8
Audie & Scarlata (2007)				
	training set	24	0.97	0.6
	test set	35	0.53	2.4
Zhang et al. (2005)		82	0.53	2.2
Su et al. (2009)	test set 5	82	0.53	2.2
	test set 6	86	0.58	2.2

- **Wrong models** based on the reaction product, but not the reactants
- **Wrong data** - and errors that propagate from one paper to the next

Problems with the experimental data

in Horton & Lewis (1992)





Hb S dimer

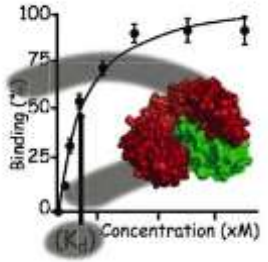
dipeptides

1HBS Hemoglobin S

- the dimer does not exist, except in crystal
- there is no K_d in literature only a critical concentration

BPTI / trypsin or trypsinogen		ΔG (kcal/mol)
2PTC	trypsin	18
2TPI	trypsinogen/IleVal	no data
2TPG	trypsinogen	7

Building a structure-affinity benchmark



We* start from the **Docking Benchmark** version 4.0 (Hwang et al. 2010), which includes the structures of 176 complexes and their **unbound** components, and **collect K_d values** from the literature.

We do our best NOT to

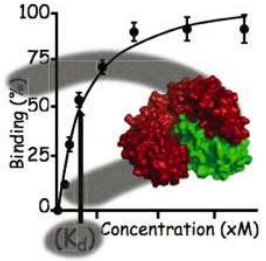
- associate a K_d with the **wrong proteins** or the wrong complex
- use **second hand data** that can't be traced to an actual measurement
- or data obtained *in vivo*, or under poorly defined conditions (IC_{50})
- copy typos (including typos in original papers)

while keeping track of

- method artefacts in K_d measurement (immobilization,reporter groups etc.)
- the **conditions of the measurement** : pH, ionic strength etc.
- differences between the proteins in crystal and solution studies (genetic constructs, mutations, covalent modifications)
- **allosteric** ligand effects

We*: Panos Kastritis (Utrecht), Iain Moal (London)
All the data were triple-checked by the rest of the team…

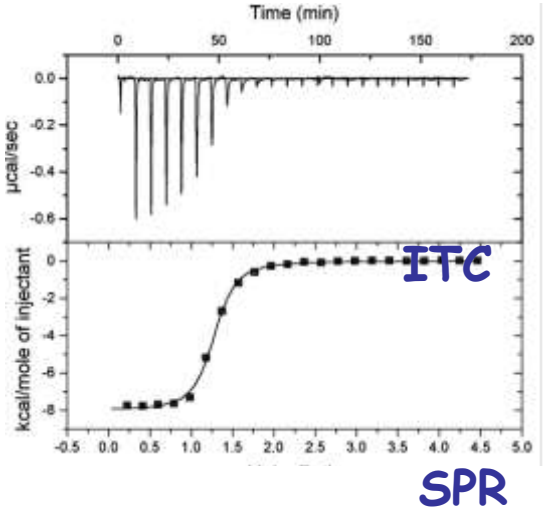
Benchmark composition: Measuring K_d



144 experimental values:

40% Titration **(Langmuir isotherm)**

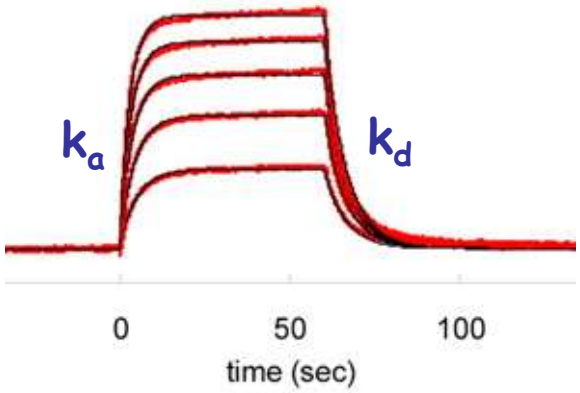
- **Spectroscopy: fluorescence, UV absorbance, NMR etc…**
- **Calorimetry (ITC)**



40% Kinetics **($K_d = k_d/k_a$)**

- **Surface plasmon resonance (SPR)**
- **Fast kinetics (stopped-flow)**

SPR

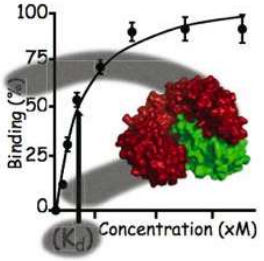


15% Enzyme inhibition

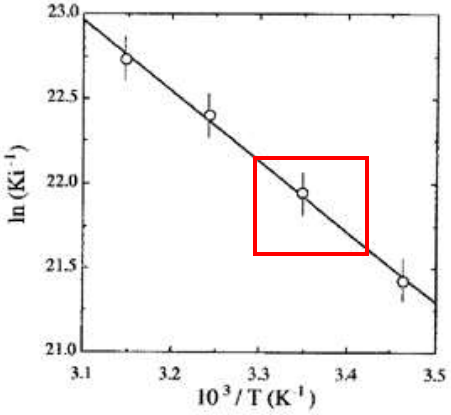
- **K_i corrected for competition with substrate and slow binding;**

5% Other methods (AUC etc.)

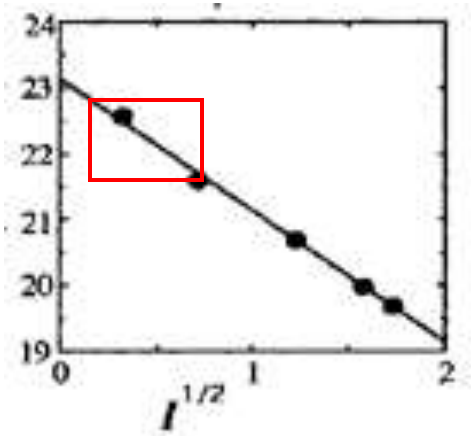
How experimental conditions affect K_d



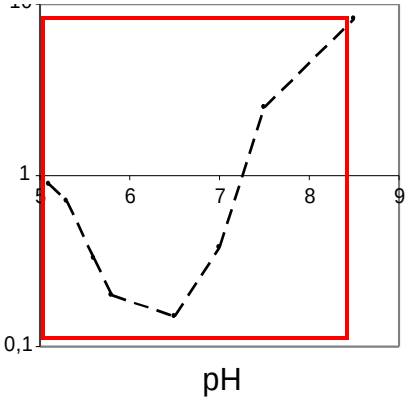
Temperature



Ionic strength



pH

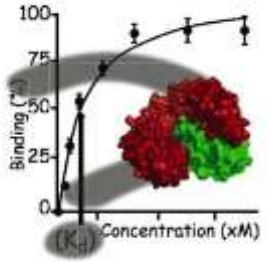


	range	K_d range (nM)	ratio
Temperature	20 - 35 ° C	0.2 - 0.4	2
Ionic strength	0.1 - 0.5 M	0.12 - 0.38	3
pH	5 - 8.5	0.15 - 8	53

Data on *Streptomyces* inhibitor / thermolysin

Kunugi et al. 1999 *FEBS Lett* 259:815

Error bars in K_d data



Source of discrepancy	$\sigma(K_d) / K_d$	$\sigma(\Delta G)$ kcal/mol
Experimental error (*as reported*)	20-50%	0.1-0.25
Discrepancy between methods	2-10	0.4-1.4
Protein sequence, modifications etc…	1-10	<1.4
Dependence on		
temperature (20-35° C)	2	0.4
ionic strength (0.1-0.5 M)	2-10	0.4-1.4
pH (6-8.5)	10-50	1.4-2.3

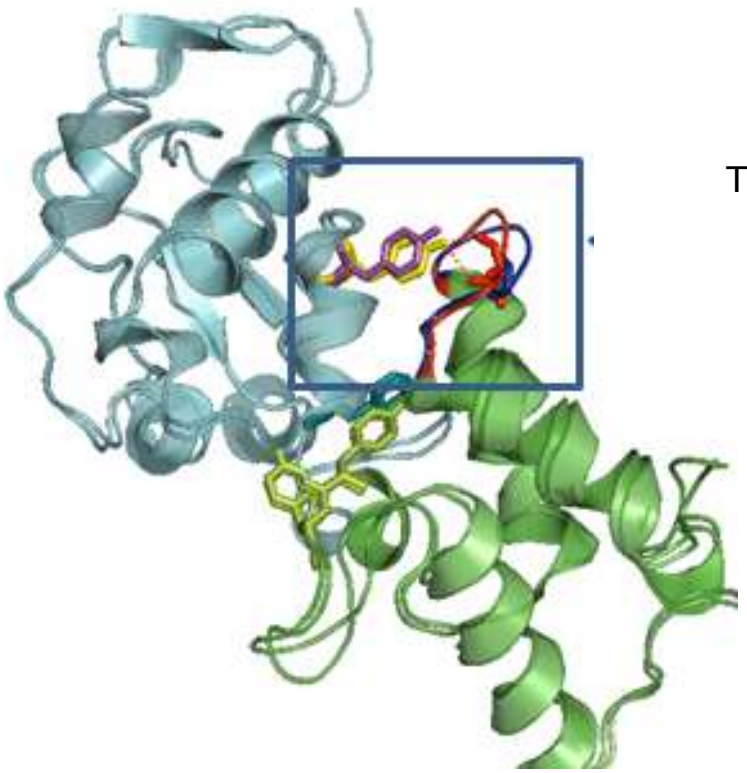
Conclusion:

- **Most K_d values in our set are defined to within one order of magnitude**
- **It makes no sense to model or predict ΔG to within better than 1.4 kcal/mol unless one can also model its pH dependence**

Same structure, different K_d: colicin DNase/immunity protein

Kleanthous et al. (1998) *Mol. Microbiol.* 28:227; Meenan et al. (2010) *PNAS* 107:10080

Colicins are protein weapons excreted by *E. coli* strains to kill other bacteria; they carry DNase (or other) enzymic activities. To protect itself against its own colicin, each strain also produces an **immunity protein** that inhibits the cognate colicin very efficiently (K_i < 1 pM), and other (non-cognate) colicins poorly (K_i > 1 nM). Cell survival requires K_i < 0.1 nM.



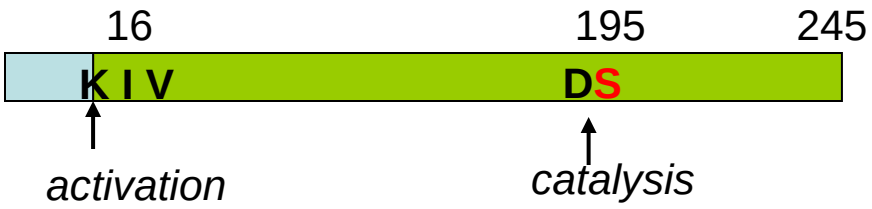
The DNase domain of colicin **E9** has been crystallized in complex with the cognate **Im9** and the non-cognate **Im2** (68% seq id).

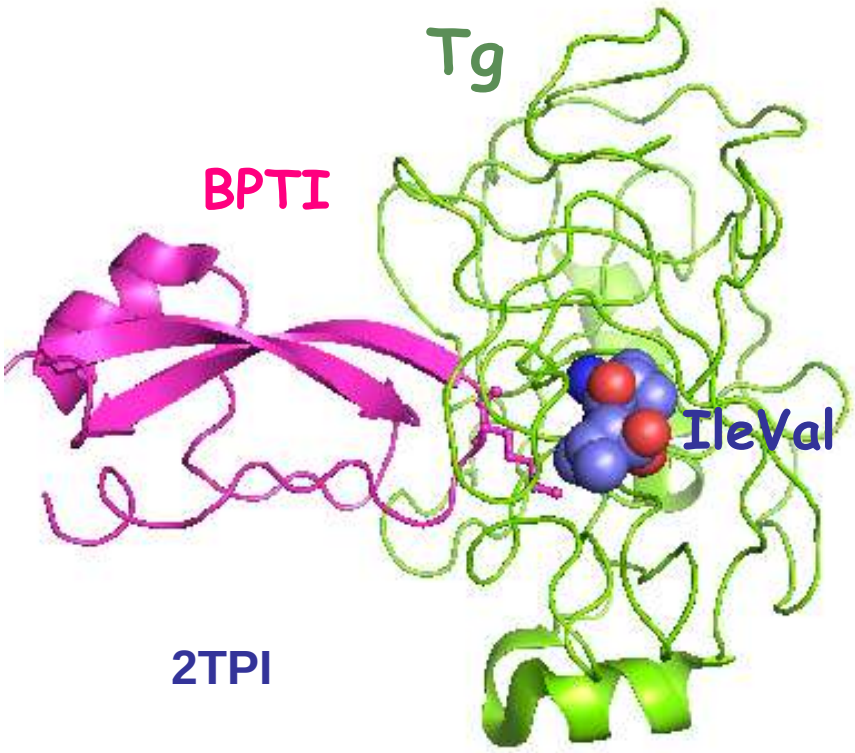
The two complexes have a **very similar structure** (rmsd = 0.4 Å), but extremely **different affinities**

PDB	complex	K_d
1EMV	**E9 / Im9**	$2.4\ 10^{-14}$ M
2WPT	**E9 / Im2**	10^{-7} M
K_d ratio **= 4.10^6**		**$\Delta\Delta G$ = 9.2** kcal/mol

Trypsinogen as an allosteric protein

Bode (1979) *JMB* 127:357





How trypsinogen becomes trypsin:

Proteolytic cleavage of the Lys-Ile16 peptide bond releases a -NH_3^+ that can interact with Asp194 at the active site, triggering a major **conformation change**. The protein becomes fully ordered, a substrate binding site forms, and the enzyme becomes active

- **BPTI binding induces the same change**
- **addition of the *IleVal* dipeptide also !**

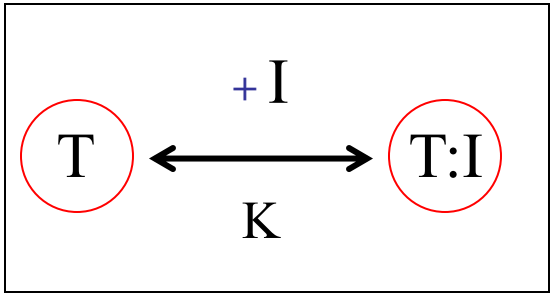
Allosteric interaction:

BPTI binding raises the affinity of trypsinogen for *IleVal* by > 5 orders of magnitude.

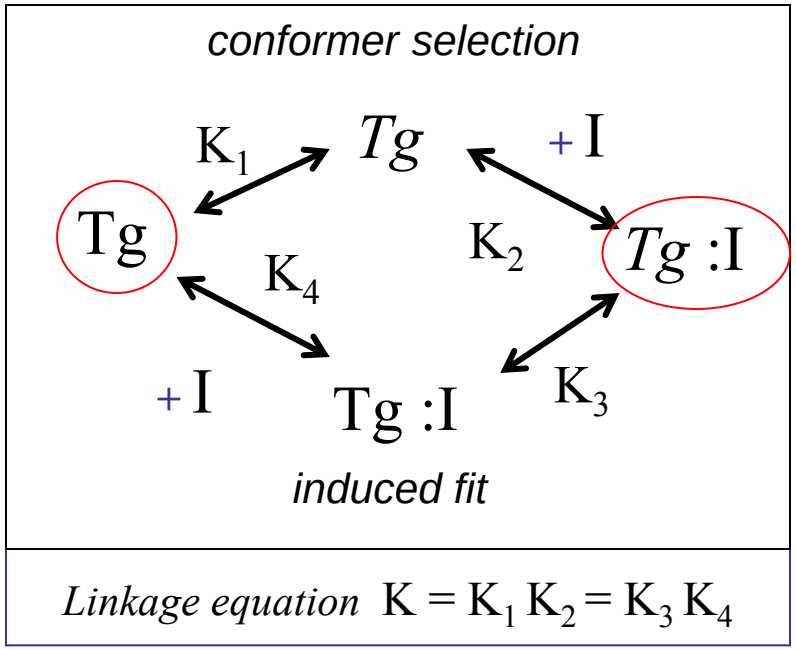
Same structure, different K_d: trypsin and trypsinogen / BPTI

Felhammer, Bode, Huber (1977) *JMB* 111:415

trypsin



trypsinogen



X-ray structures (W. Bode)

T, T:I and *Tg*:I are all very similar, but

Tg is different (no substrate binding site, disordered loops)

Affinity for BPTI

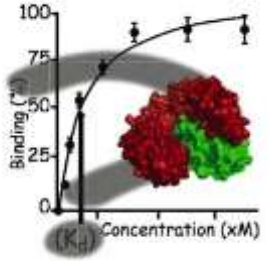
Experimental

trypsin T	$K = 6\ 10^{-14}$ M
trypsinogen Tg	$K = 2.3\ 10^{-6}$ M
Assuming Tg ≈ T	$K_2 \approx 6\ 10^{-14}$ M

From the linkage equation

isomerization Tg -> *Tg*	$K_1 \approx 3\ 10^{-8}$

Allosteric systems



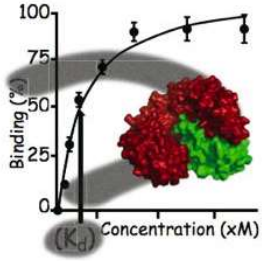
- There are many allosteric proteins our affinity benchmark
- They play essential roles in regulation
- Their affinity for a protein ligand can be highly dependent on the presence (and concentration) of another ligand :

  H^+ (pH effect), Ca^{++}, small molecule, DNA etc…

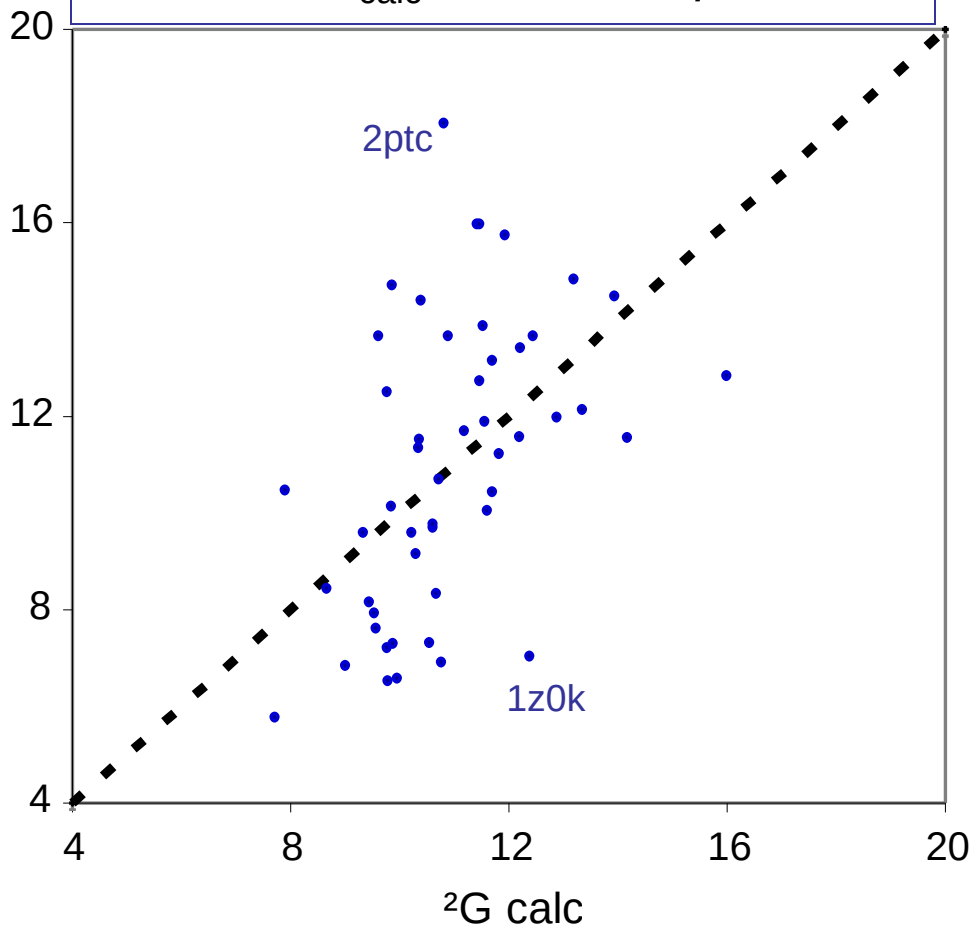
Example: **G-proteins**

- G-proteins bind GTP and hydrolyze it to GDP in a highly regulated way
- they change conformation when GDP or GTP binds ('switch' regions)
- they interact with many partner proteins (GAP, GEF, kinases etc…)
- K_d for a partner protein may change by 3 to 5 orders of magnitude between the 'empty' form, the GDP complex and the GTP complex
- **Signaling** depends on that allosteric effect

Fitting ΔG in the rigid-body case



$$\Delta G_{calc} = \alpha\, BSA + \beta$$



48 of the 145 complexes (33%) display small changes at the interface

($\Sigma\, \delta x^2$ <35 Å^2, I_rmsd below ≈1 Å)

On those, ΔG_d correlates with the interface size: the *BSA* accounts for ≈1/3 of the variance

N	$\langle \Delta G_{calc} - \Delta G_{obs} \rangle$ (kcal/mol)	R^2	Outliers
48	2.4	0.32	2

The outliers

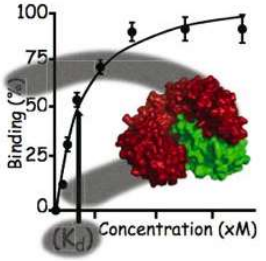
2ptc (trypsin/BPTI)

electrostatics?

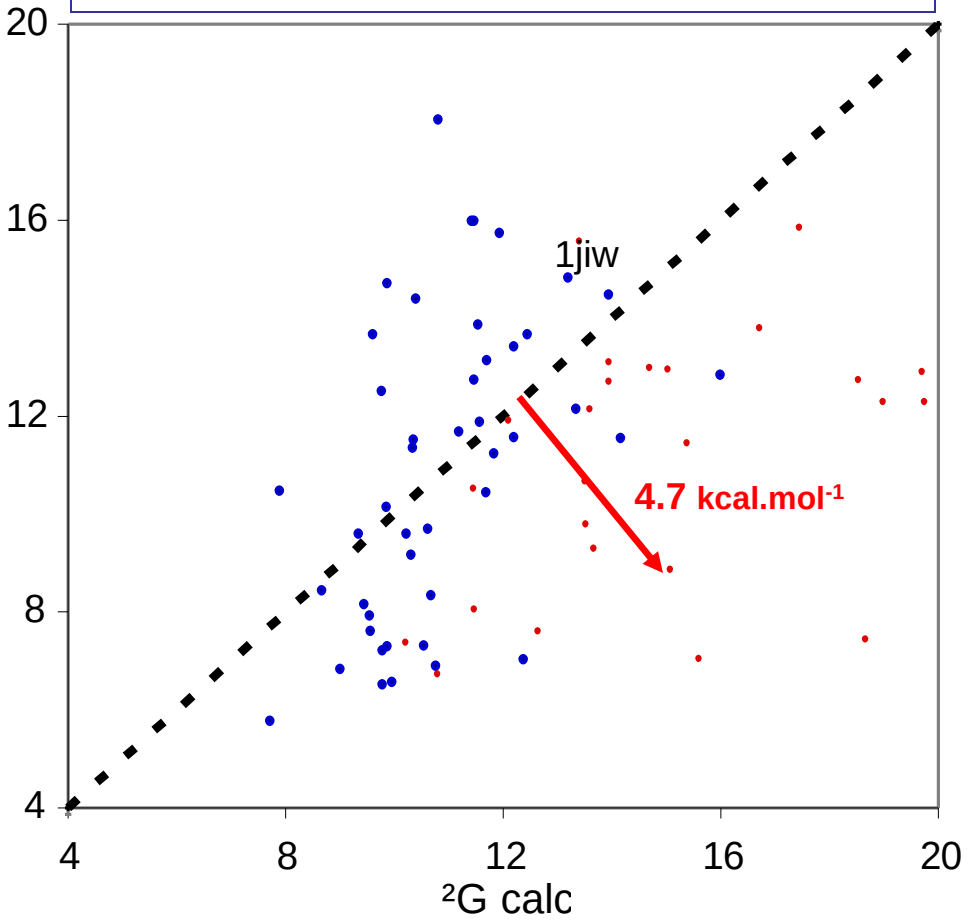
1z0k (Rab4/rabenosyn-5)

poor packing?

Fitting ΔG : the cost of conformation changes



$$\Delta G_{calc} = \alpha\, BSA + \beta$$



27 of the 145 complexes (20%) display very large movements and/or disorder-to-order transitions

($\Sigma\ \delta x^2$ >165 Å², I_rmsd = 1.5 to 9 Å)

All have $\Delta G_{obs} < \Delta G_{calc}$

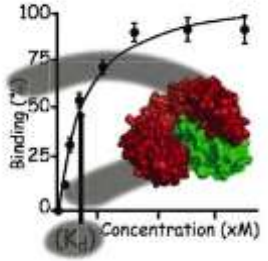
1jiw (UEV/ubiquitin) has a Zn metal bond.

$\Delta G_{obs} - \Delta G_{calc}$ is an estimate of the average **free energy cost of the conformation changes** :

$\langle \Delta G_{conf} \rangle$ = **4.7 kcal.mol^{-1}**

(the maximum difference is 34 kcal.mol^{-1})

What is new in the structure/affinity benchmark ?



- We found a reliable K_d value for ≈ 80% of the complexes in the Docking Benchmark - but only **after a lot of cross-checking !**
- We now have the first version of a database
- Along with the complexes, the **unbound structures** are available
- Nine entries represent **cognate/non-cognate pairs** of complexes,
- Many proteins are **allosteric** and have ligand binding sites that interact (trypsinogen, G-proteins, receptors etc…)
- Many displays **large conformation changes** … and that costs free energy !

http://bmm.cancerresearchuk.org/~bmmadmin/Affinity

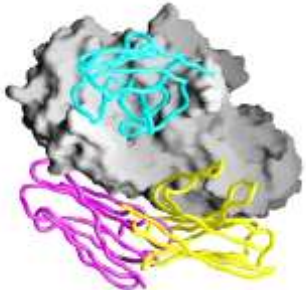
Kastritis et al. (2011) *Protein Sci.* 20:482

I'm grateful to...

Protein-protein interfaces in the PDB

P. Chakrabarti, R. Bahadur, S. Dey (Bose Institute, Calcutta)

CEFIPRA and ARCUS-Ile de France for financial support



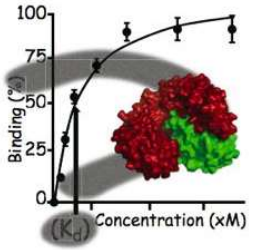
Managing CAPRI

S. Wodak (Toronto), M. Lensink (Brussels-Lille)

K. Henrick, S. Velenkar (EBI, Hinxton)

I. Vakser (Lawrence, Kansas), S. Vajda (Boston, Mass)

Structure/affinity benchmark

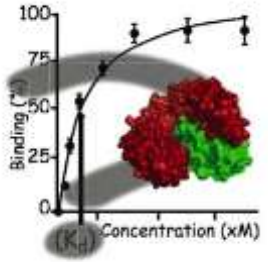


A. Bonvin, P. Kastritis (Utrecht)

I. Moal, P. Bates (Cancer Research UK, London)

H. Hwang, Z. Weng (U Mass, Worcester, Mass)

Benchmark composition: functional classes



Class		Number	non-cognate	ΔG (kcal/mol)*	s.d.
A	Antigen/antibody	19	2	12.2	1.3
EI	Enzyme/inhibitor	40	4	13.8	2.2
		41%			
ES	Enzyme/substrate	11	1	9.1	1.8
OG	G-proteins	19	-	9.0	2.
OR	Receptors	14	-	11.4	1.6
OX	Miscellaneous	41	2	9.1	2.3
	All	144	9	10.7	2.9

* cognate only

Within a functional class, the mean value of ΔG is a predictor

- at least as good as any published physical model: s.d. < 2.4 kcal/mol
- nearly optimal for A (antibodies) and OR (receptors): s.d. ≈ 1.4 kcal/mol)