What can the PDB tell us about how proteins recognize and bind to each other?

Charles Robert[1]

Joël Janin[2]

[1] Institut de Biologie Physico-Chimique, Paris

[2] Université Paris XI, Orsay

Preface: ways to approach existing knowledge

- Observations as a function of parameters
  - coded as alphabets, integers, real numbers, …
  - anywhere from 1 to N dimensions
  - noise

What

- Pattern recognition
  - Machine learning
  - Intuition (supervision)
- Models
  - quantitative, qualitative, heuristic
  - More data, better models
  - More *types* of data, better models

Why

Overall idea

Today I'll talk about three types of experience with biological macromolecules

- Biochemical/biophysical
- Structural
- Mechanical / Dynamic

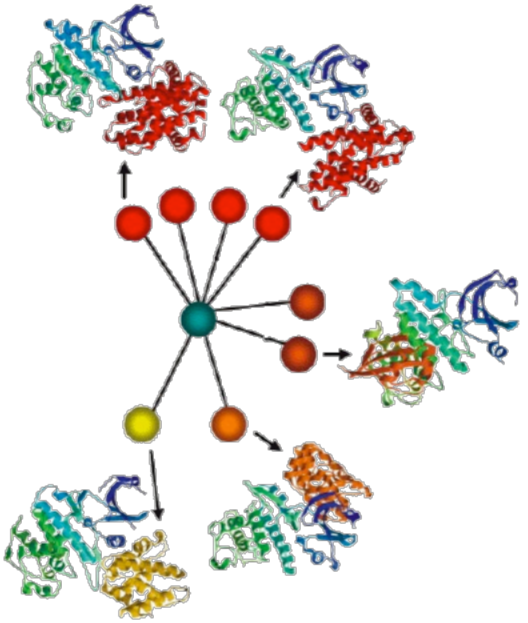
What

Why

WHAT

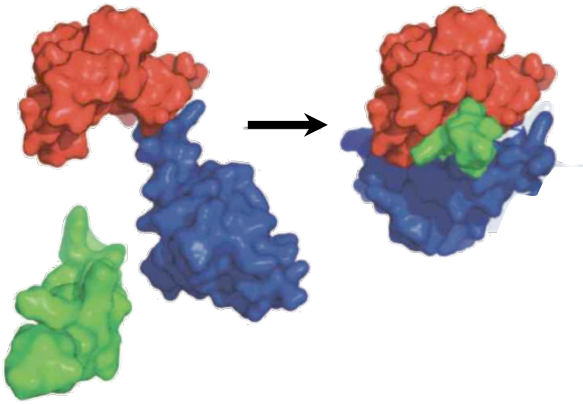
Macromolecular interactions in Biology

- **Macromolecular interactions**
  - Define the morphology of the organism, its function, its pathologies
  - Interactions generally manifested by a complex
  - Repetition ofs interfaces → assemblies (filaments, envelopes, …)
  - Therapeutic targets



- **100 000's of protein-protein interactions (PPIs)**
  - identified by biochemical approaches (TAP/Tag)
  - results in form of graphs
    edges = interactions

- **Current challenges**
  - structural information missing
    complexes < 5 % of the PDB
  - large assemblies
  - pairwise info – interface overlap
  - flexible association
  - dynamic information missing



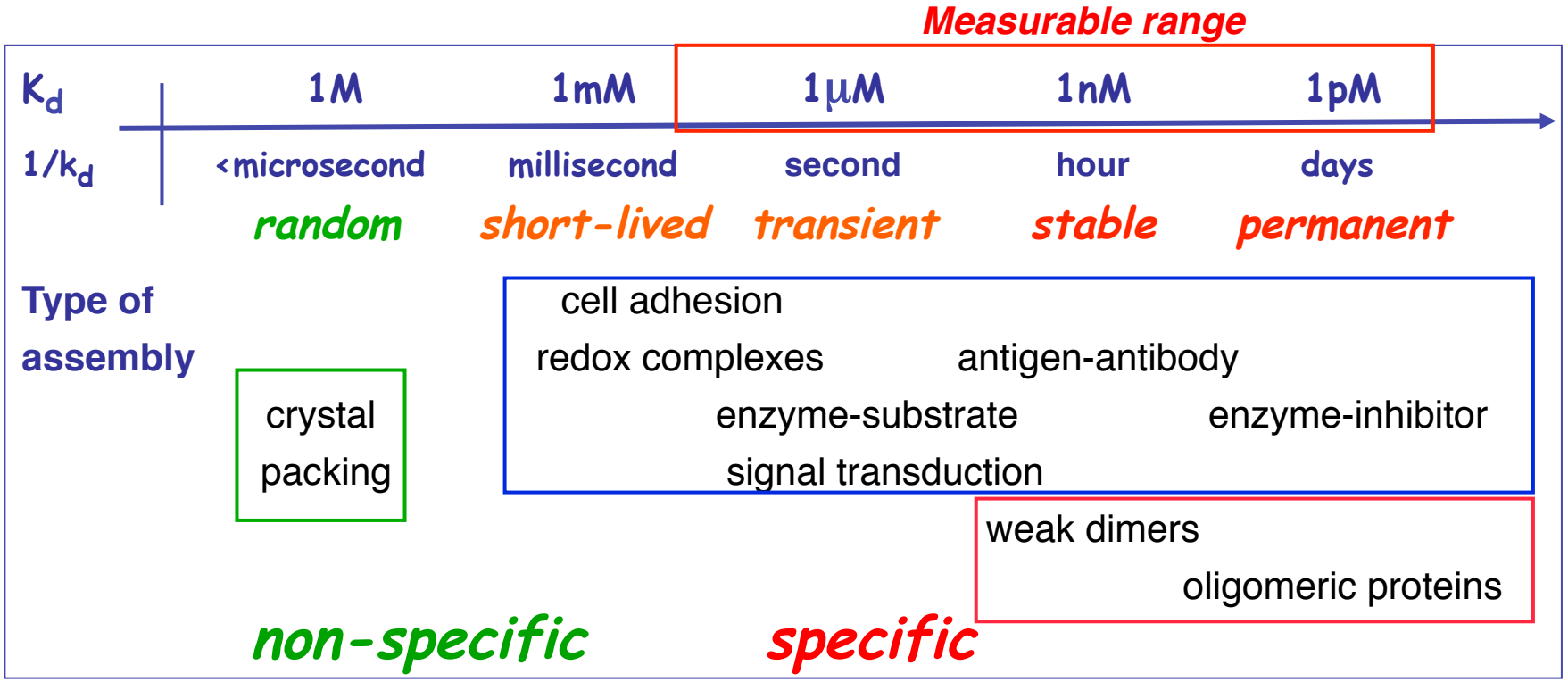
Protein-protein recognition: affinity and time scales

Affinity :

- equilibrium constant K_d; the dissociation reaction has a free energy $\Delta G_d = -RT \ln K_d/c°$ (c°=1M in standard state)
- at a given free component concentration, K_d determines the fraction bound

Time scale:

- fixed by rate constants k_a (bimolecular) and k_d (monomolecular); $K_d = k_d / k_a$
- $k_d$determines whether an assembly is **permanent** or **transient** (life time **1/k_d**)



Interface size and stability

ASA

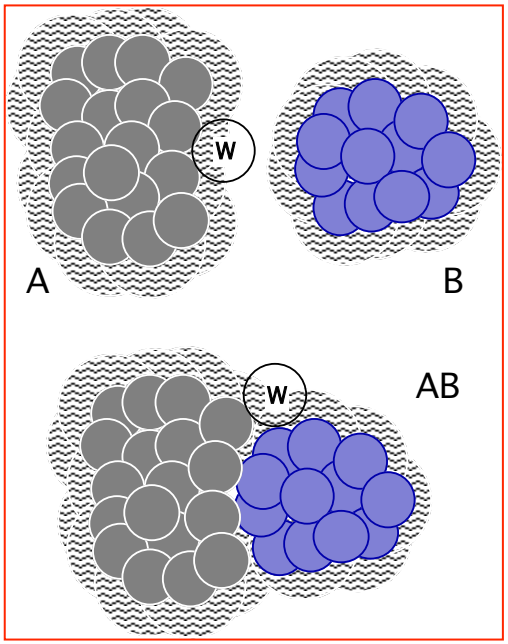
The **solvent accessible surface area** assesses molecule-solvent contacts. (*Lee & Richards 1971*)

BSA

The **buried surface area (=interface area)** assesses molecule-molecule contacts (*Chothia & Janin, 1975)*

Each interface atom contributes an average **≈ 10 Å²**

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



The hydrophobic effect

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

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

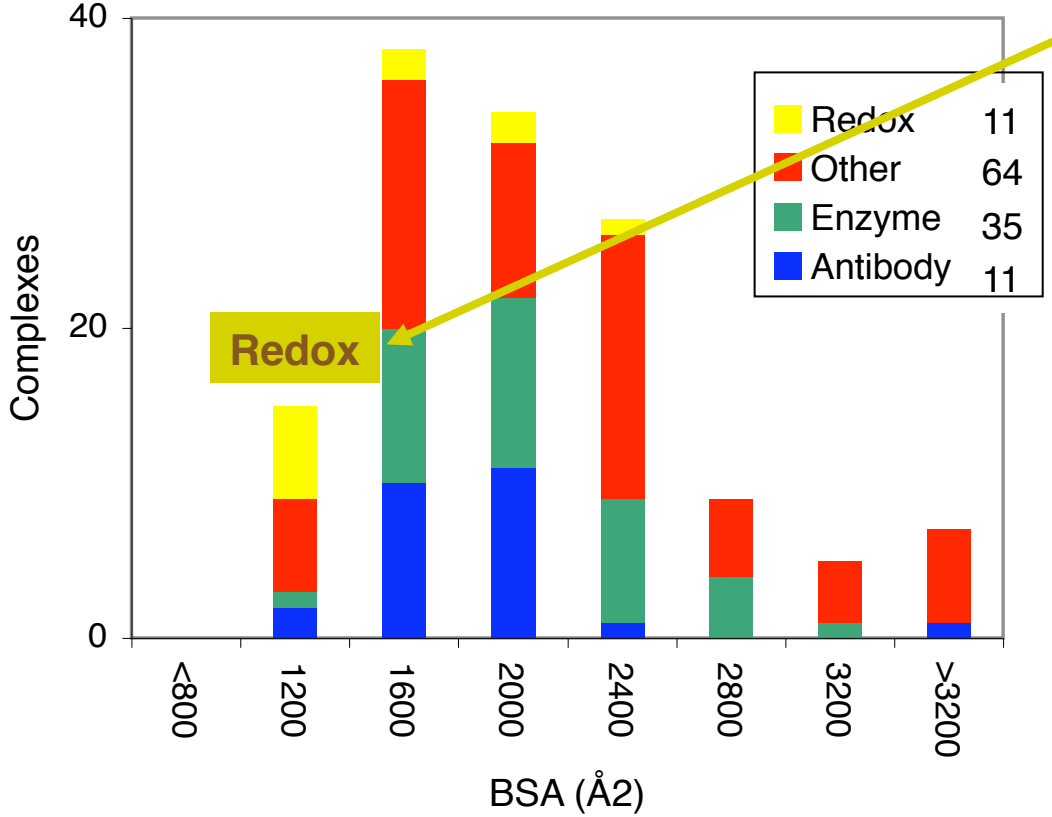
accepted values γ = 24 (*Chothia, 1974*) to 50 cal/mol.Å^2

Polar and non-polar interactions

- the number of Van der Waals interactions at the interface scales linearly with the *BSA*
- Interfaces have about 1 H-bond per 200 Å^2 *BSA*, but the correlation is mediocre.

Can we relate BSA and stability?

Interface size and stability : short-lived complexes



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

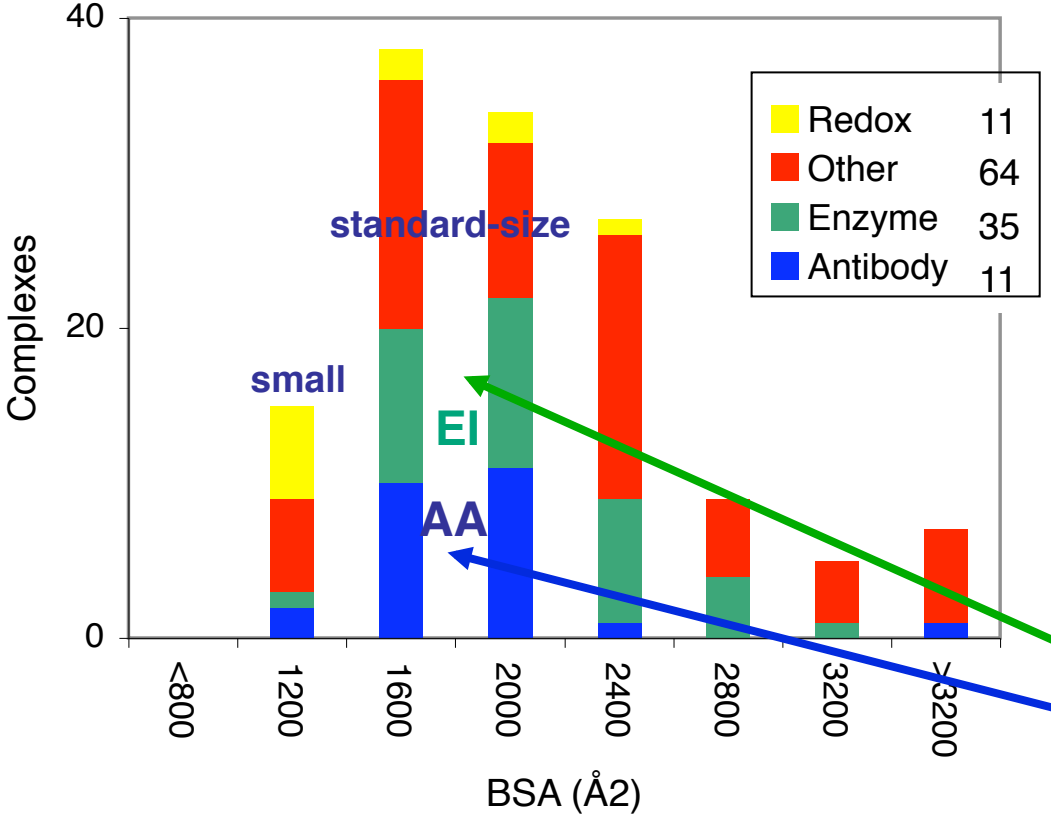
Redox (electron transfer) complexes make short-lived interactions;

Most have a **small interface**

BSA = 900-1200 Å^2 0-3 H-bonds

Crawley & Carrondo (2004)

Interface size and stability : long-lived complexes

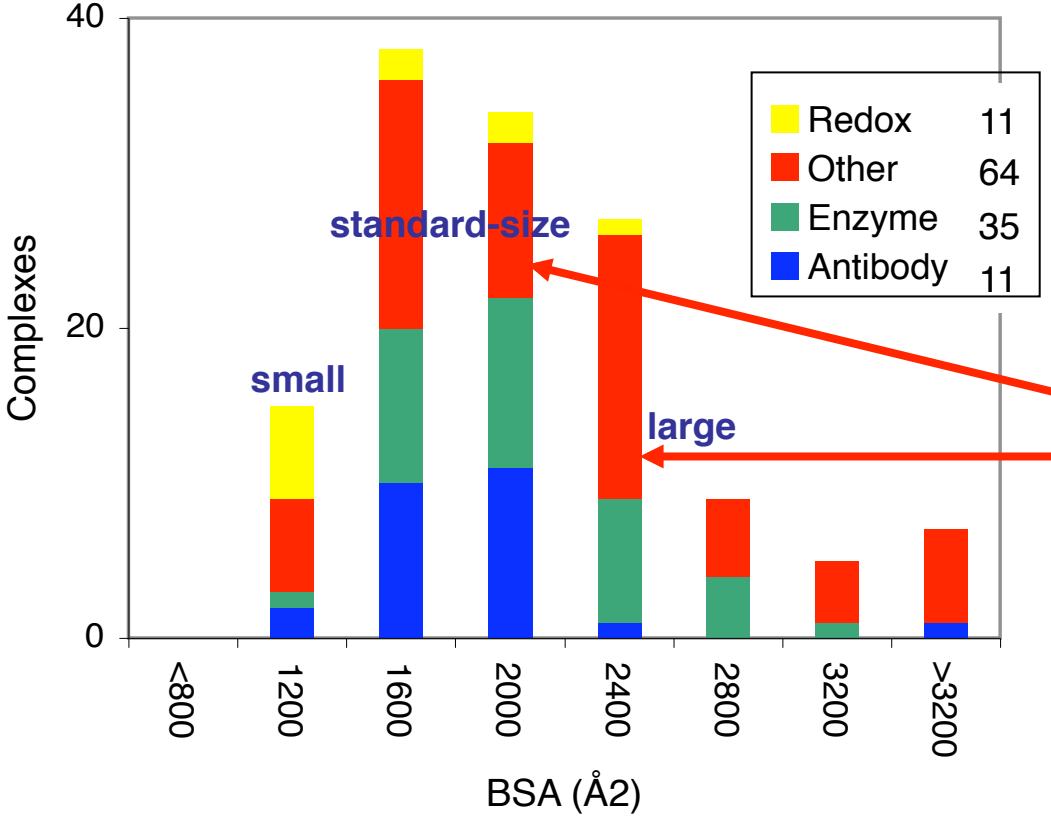


124 protein-protein complexes
Janin, Bahadur & Chakrabarti (2008)
Quat. Rev. Biophysics 2:133-180.

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))

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 and stability : transient complexes



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

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. They have standard-size or **large interfaces**:

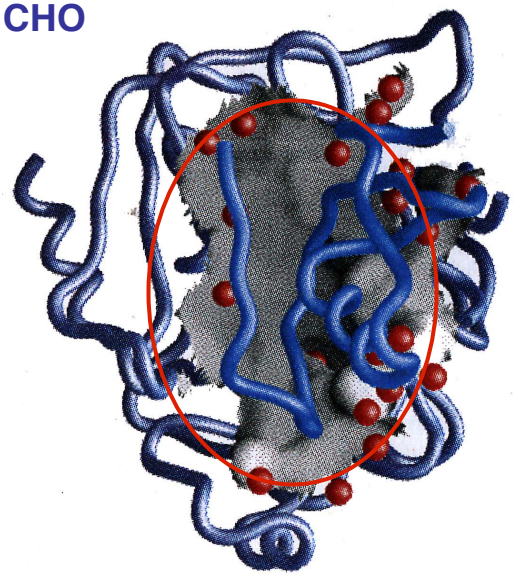
BSA >2000 Å^2 .

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

1CHO



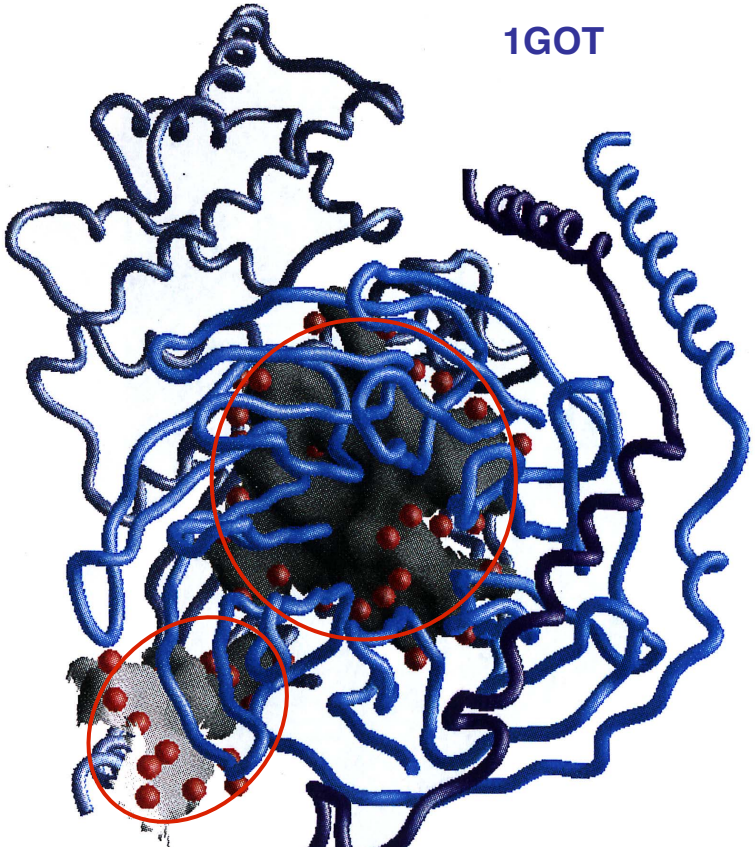
Rigid-body recognition: chymotrypsin-inhibitor complex

High affinity: $K_d \approx 0.1$ nM

A standard-size (*BSA* = 1470 Å^2), single patch interface:

No change in conformation between the free and bound components: 0.6 Å Cα RMS

1GOT



Flexible recognition: Transducin Gα-Gβγ

Low affinity: $K_d \approx 1$ μM

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

Major conformation changes (1.8 Å Cα RMS)

Conclusion (1)

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

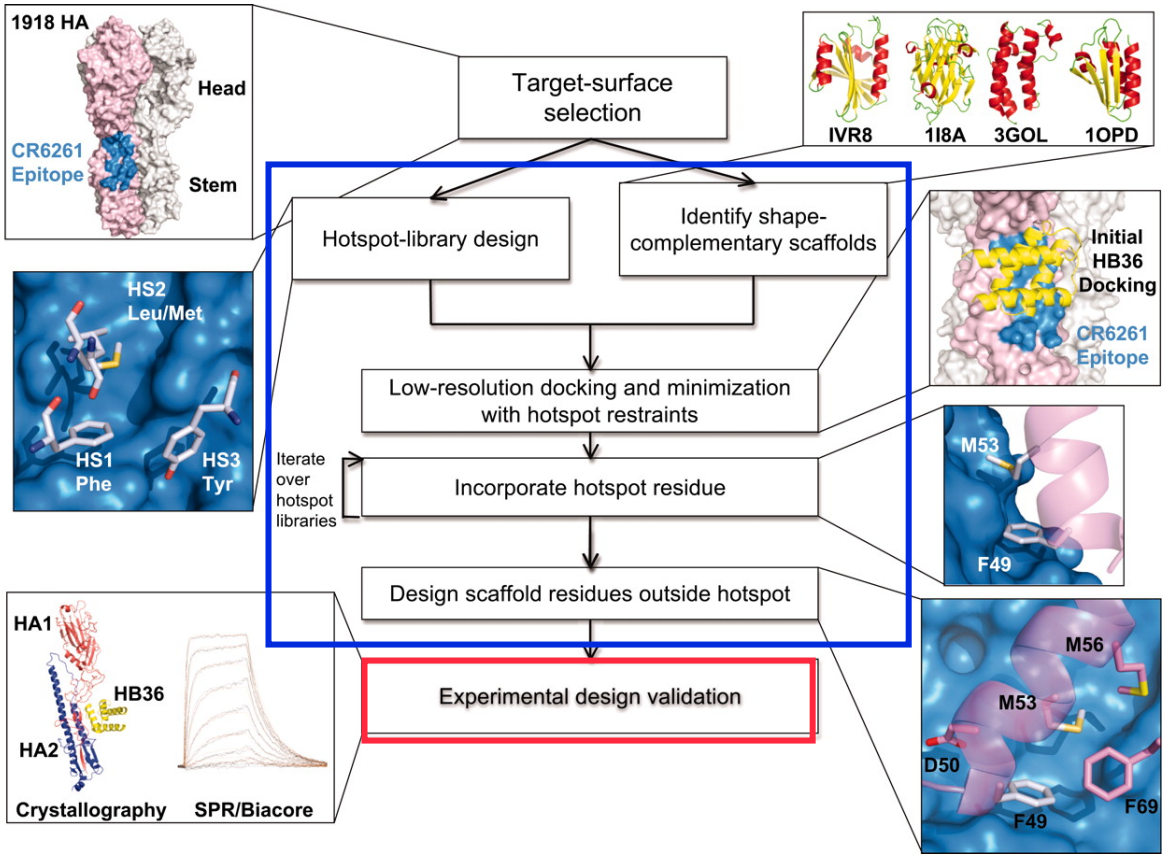
…but it may be masked by the conformation changes that accompany the formation of large interfaces (flexible recognition)

Other determinants of affinity and specificity

- stable assemblies (transient complexes and strong homodimers) have close-packed interfaces
- weak homodimers and crystal packing interfaces are loosely packed
- 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

Engineering novel interactions

Baker lab: Design proteins to bind the epitope of **Spanish flu virus hemagglutinin (**HA) recognized by a broadly neutralizing antibody (Fleishman et al. 2011, *Science* 332:816)



865 candidate scaffolds taken from the PDB (shape complementarity)

260,000 docking models mutated *in silico* and refined with Rosetta

88 candidate binders tested by yeast display. Derive from 79 scaffolds and carry an average of 11 surface mutations.

2 successful designs

Making high affinity Spanish flu HA binders

Fleishman et al. 2011, *Science* 332:816

Create HA binders

- make synthetic gene Aga2-design-cMyc
- express on the yeast surface
- incubate cells with biotinylated HA
- label with fluorescent biotin & cMyc antibody
- run flow cytometer to select binders

88 candidates

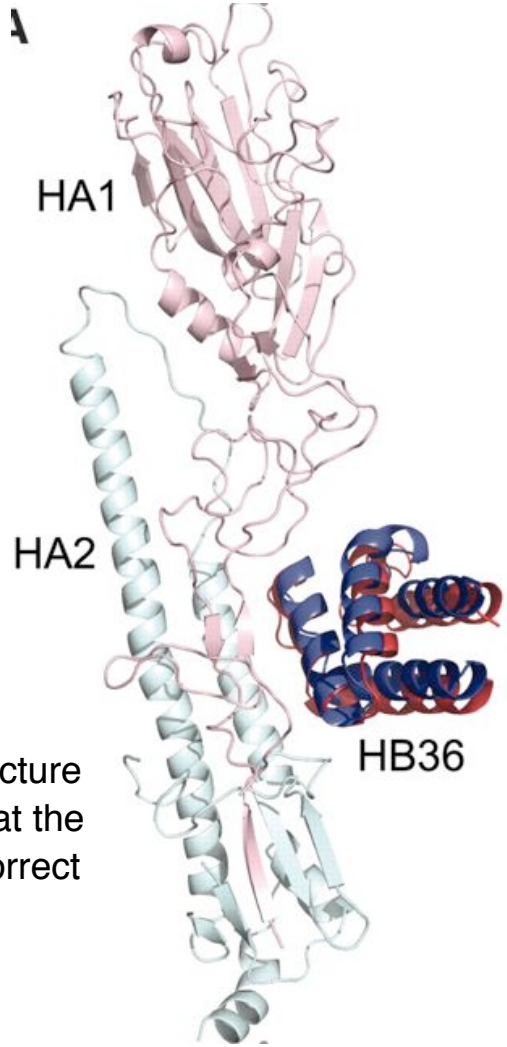
73 express

2 designs bind

Affinity maturation

mutate key residues and perform
two rounds of selection by yeast display

Design	Kd (nM) by SPR
HB36	200
HB36.3 (2 mutations)	4
HB36.4 (3 mutations)	4
HB80	>5000
HB80.3 (4 mutations)	3



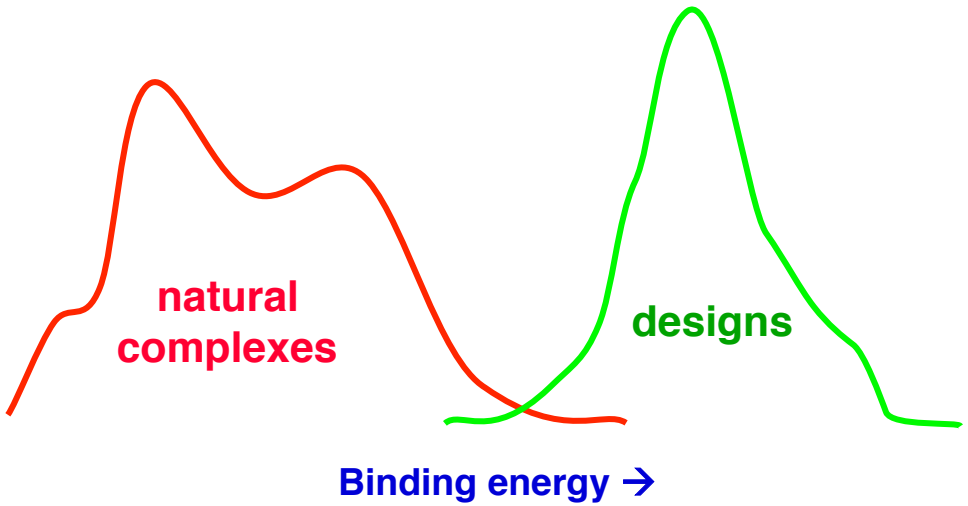
A X-ray structure confirms that the design is correct

Schematic view of the results

Some **100 candidate complexes** were designed and tested in two separate experiments.

Only **three** (Pdar/Prb; Karanicolas et al. 2011; HB36 and HB80, Fleishman et al. 2011) were reproducibly detected in the yeast display/fluorescence assay.

All other candidates have $K_d >> 1\ \mu M$ ($\Delta G_d < 8$ kcal/mol).

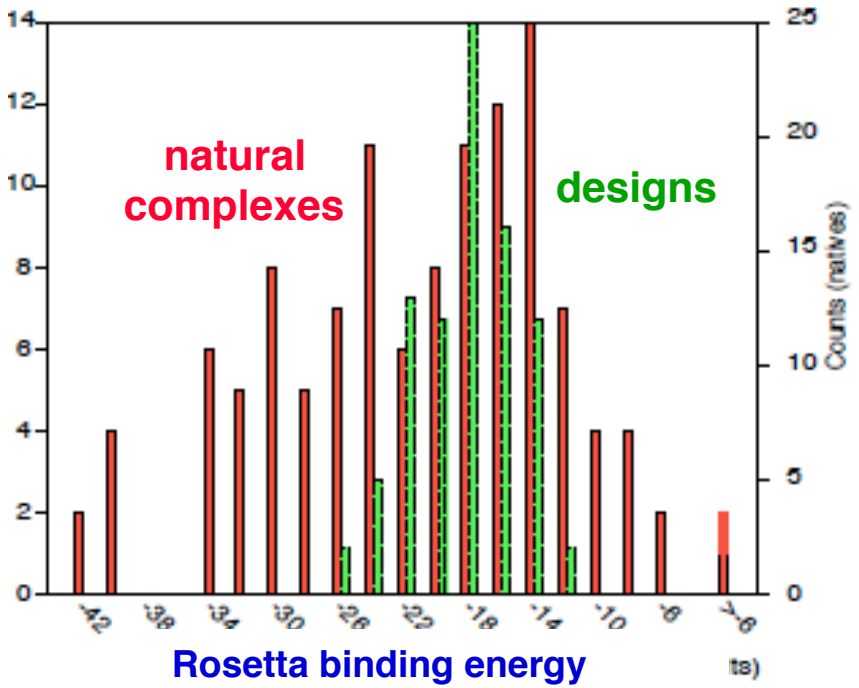


Fleishman et al. (2011) JMB

Why is the success rate of the designs so low ?

Rosetta force field predicts similar binding free energies for the **designs** and a majority of the 120 **natural complexes** taken from the docking benchmark of Hwang et al. (2010).

- **What's wrong with Rosetta?**
- **Can other force fields do better?**



Fleishman et al. (2011) JMB

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 with >100 targets and >500 predictors

➢ **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; about 60 groups participate
- Targets are few: a round of prediction begins any time one is made available

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

The Seattle Challenge to CAPRI: predict affinity

Based on their refined docking models, David Baker and Sarel Fleishman challenged CAPRI groups to

- **predict which designs make a stable complex**
- **rank the designs** relative to the known natural complexes in terms of binding free energy

CAPRI Round 20 (Feb. 2010):
42 designs, one successful

CAPRI Round 21 (April-June 2010):
87 designs, one successful

38 CAPRI groups participated

… and cosigned a *JMB* paper

Fleishman et al., 2011

doi:10.1016/j.jmb.2011.09.031 J. Mol. Biol. (2011) xx, xxx–xxx

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com/jmb

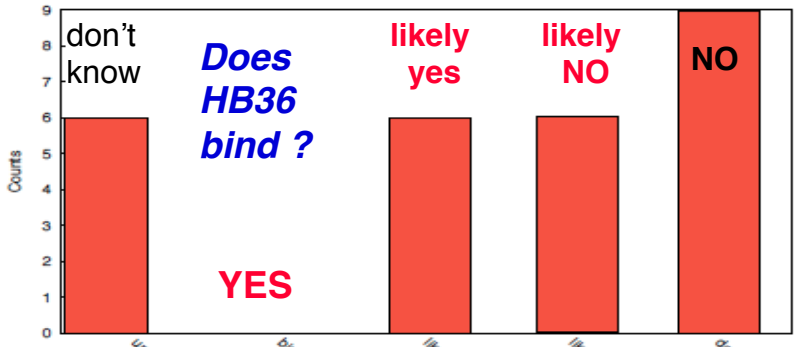
Community-Wide Assessment of Protein-Interface Modeling Suggests Improvements to Design Methodology

Sarel J. Fleishman[1], Timothy A. Whitehead[1], Eva-Maria Strauch[1], Jacob E. Corn[1], Sanbo Qin[2], Huan-Xiang Zhou[2], Julie C. Mitchell[3], Omar N. A. Demerdash[4], Mayuko Takeda-Shitaka[5], Genki Terashi[5], Iain H. Moal[6], Xiaofan Li[6], Paul A. Bates[6], Martin Zacharias[7], Hahnbeom Park[8], Jun-su Ko[8], Hasup Lee[8], Chaok Seok[8], Thomas Bourquard[9,10,11], Julie Bernauer[10], Anne Poupon[12,13,14], Jérôme Azé[10], Seren Soner[15], Şefik Kerem Ovalı[15], Pemra Ozbek[15], Nir Ben Tal[16], Türkan Haliloglu[15], Howook Hwang[17], Thom Vreven[17], Brian G. Pierce[17], Zhiping Weng[17], Laura Pérez-Cano[18], Carles Pons[18], Juan Fernández-Recio[18], Fan Jiang[19], Feng Yang[20], Xinqi Gong[20], Libin Cao[20], Xianjin Xu[20], Bin Liu[20], Panwen Wang[20], Chunhua Li[20], Cunxin Wang[20], Charles H. Robert[21], Mainak Guharoy[21], Shiyong Liu[22], Yangyu Huang[22], Lin Li[22], Dachuan Guo[22], Ying Chen[22], Yi Xiao[22], Nir London[23], Zohar Itzhaki[23], Ora Schueler-Furman[23], Yuval Inbar[24], Vladimir Potapov[24], Mati Cohen[24], Gideon Schreiber[24], Yuko Tsuchiya[25], Eiji Kanamori[26], Daron M. Standley[27], Haruki Nakamura[25], Kengo Kinoshita[28], Camden M. Driggers[29], Robert G. Hall[30], Jessica L. Morgan[29], Victor L. Hsu[29], Jian Zhan[31], Yuedong Yang[31], Yaoqi Zhou[31], Panagiotis L. Kastritis[32], Alexandre M. J. J. Bonvin[32], Weiyi Zhang[33], Carlos J. Camacho[33], Krishna P. Kilambi[34], Aroop Sircar[34], Jeffrey J. Gray[34], Masahito Ohue[35], Nobuyuki Uchikoga[35], Yuri Matsuzaki[35], Takashi Ishida[35], Yutaka Akiyama[35], Raed Khashan[36], Stephen Bush[36], Denis Fouches[36], Alexander Tropsha[36], Juan Esquivel-Rodríguez[37], Daisuke Kihara[37], P. Benjamin Stranges[38], Ron Jacak[38], Brian Kuhlman[38], Sheng-You Huang[39], Xiaoqin Zou[39], Shoshana J. Wodak[40,41,42], Joel Janin[43] and David Baker[1,44]*

The Seattle Challenge: how did CAPRI perform?

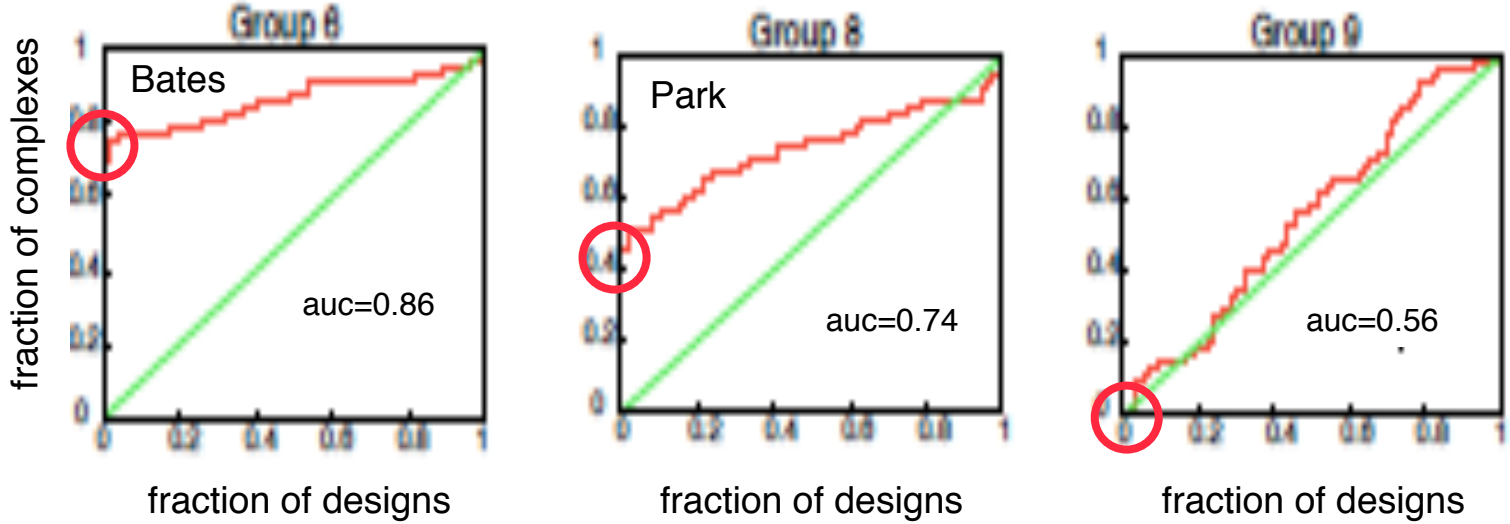
Predict the stable designs ?
(HB36, HB80)

No one could !



Rank designs vs. complexes in terms of stability: ***some predictors did better than others…***

- **Group 6** (Paul Bates, Cancer Research UK, London) ranks 75% of the natural complexes above all the designs
- **Group 8** (Park, Seoul National University) ranks 40% of the natural complexes above all the designs
- **Group 9** (xxx) returns nearly random ranks



Conclusion (2)

What did we learn from CAPRI affinity predictions ?

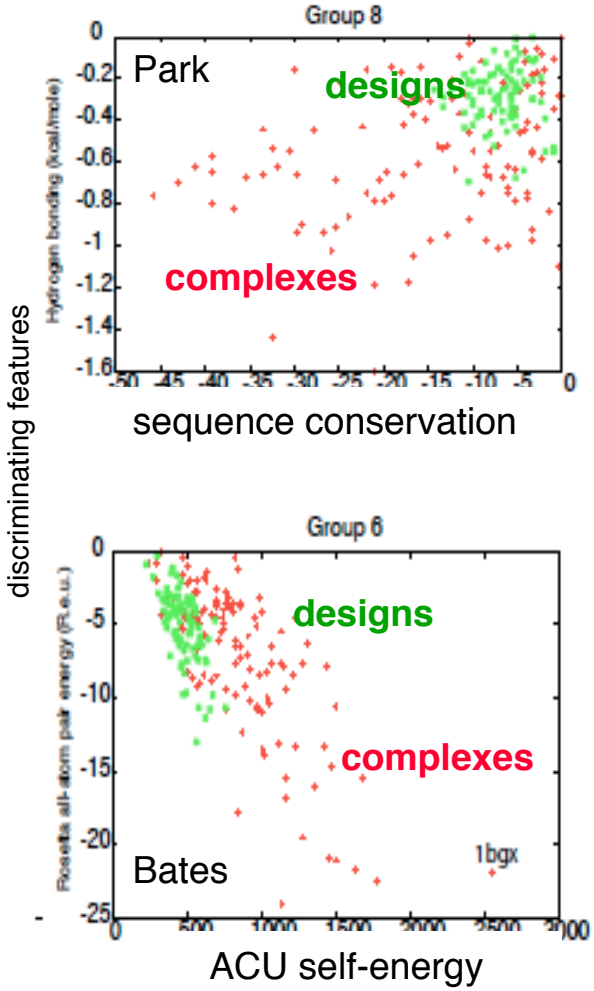
Telling a design from a natural complex may not need a force field…

- Group 8 (Park, Seoul National University) makes out designs by their lack of sequence conservation…
- Other groups trained their procedure on previous designs from the Baker lab…

***… but the goal was to improve* Rosetta**

Commonly used force fields (including Rosetta) contain poorly estimated energy terms, especially electrostatics

Group 6 (Paul Bates, Cancer Research UK, London) uses a **solvation self-energy** (ACE: Analytical Continuum solvent Electrostatics) → discriminates between natural complexes and designs much better than the Coulombic energy in Rosetta.

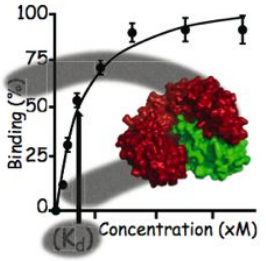


WHY (OR HOW) 1

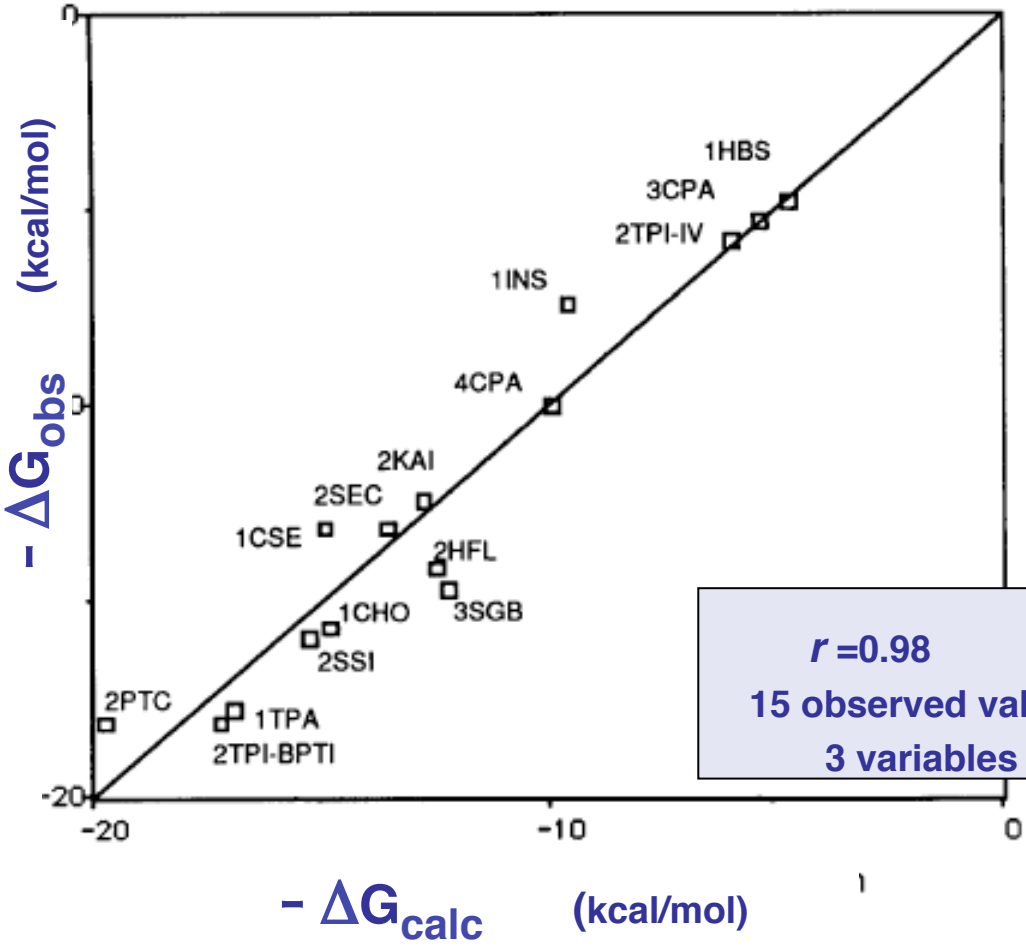
Modelling of affinity

Modeling affinity from structure

Horton & Lewis (1992, *Protein Sci.* 1:169)



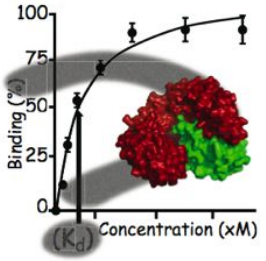
$$\Delta G_{calc} = \alpha \Delta G_{np} + \beta \Delta G_{pol} + \gamma$$



***r* =0.98**
15 observed values
3 variables

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

Later attempts to fit ΔG had more parameters, yet they did far less well !

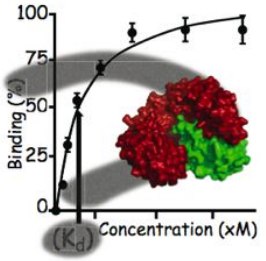


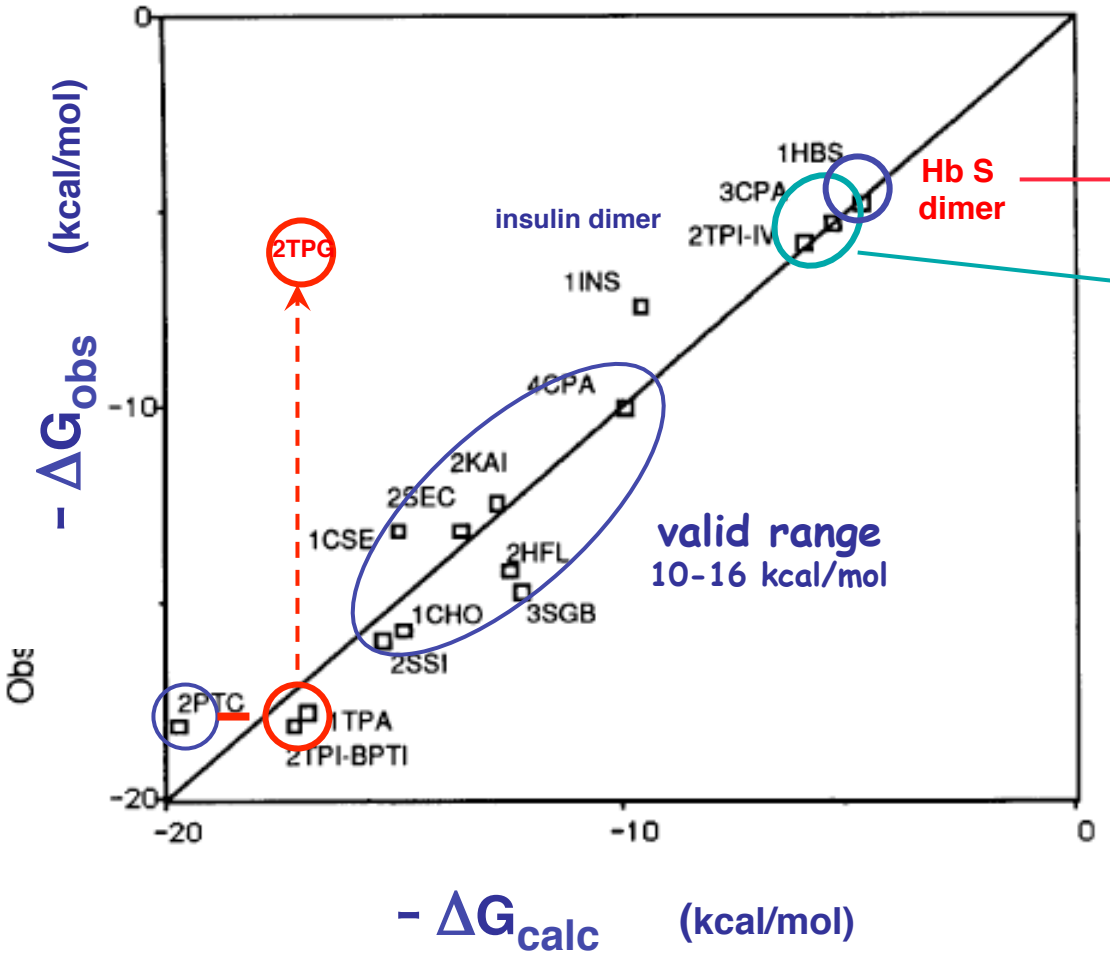
		Sample size	*r* correl. coeff	$\langle \Delta G_{calc} - \Delta G_{obs} \rangle$ (kcal/mol)
Horton & Lewis (1992)		15	0.98	0.8
Audie & Scarlata (2007)				
overfitting ?	training set	24	0.98	0.6
	test set	35	0.73	2.4
Zhang et al. (2005)		82	0.73	2.2
Su et al. (2009)	test set 5	82	0.73	2.2
	test set 6	86	0.76	2.2

- **Wrong model** the reaction has a product, but no reactants !
- **Wrong data** and errors propagating from one paper to the next

Problems with the experimental data

in Horton & Lewis (1992)





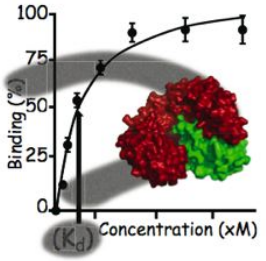
1HBS Hemoglobin S

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

dipeptides

BPTI / trypsin or trypsinogen		ΔG_{obs} (kcal/mol)
2PTC	trypsin	18
2TPI	~~trypsinogen/IleVal~~	*no data*
2TPG	trypsinogen	7

Building a structure-affinity benchmark



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

Did their best NOT to

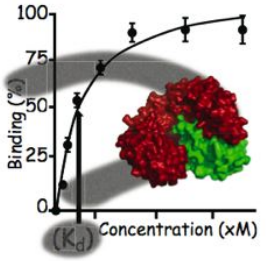
- associate a K_d with the **wrong proteins** or the wrong complex
- use **second hand data** that can't be traced to an actual measurement
- or data obtained *in vivo*, or under poorly defined conditions (IC_{50})
- copy typos (including typos in original papers)

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

"A structure-based benchmark for protein–protein binding affinity" Kastritis, Moal, Hwang, Weng, Bates, Bonvin & Janin. (2011) Prot Sci 20, 482-91.

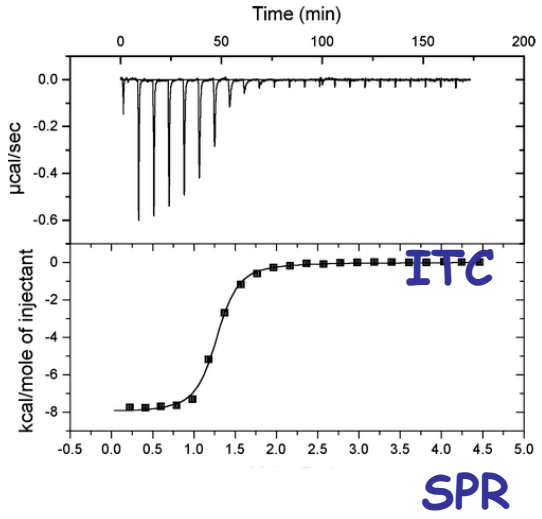
Benchmark composition: Measuring K_d



144 experimental values:

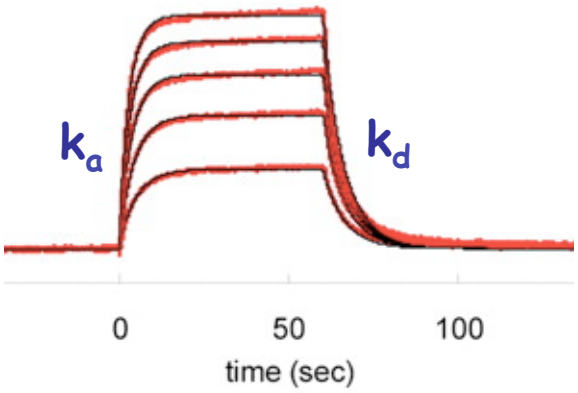
40% Titration

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



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

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



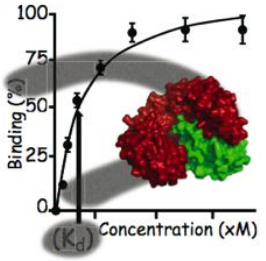
15% Enzyme inhibition

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

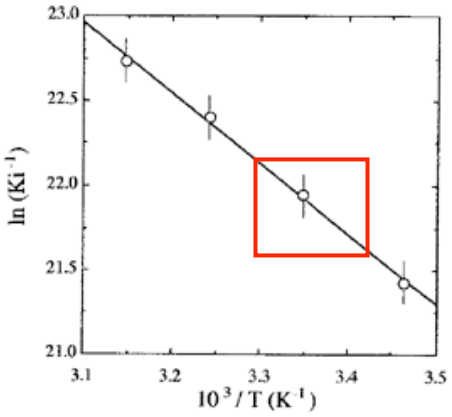
5% Other methods

- **Analytical ultracentrifugation, …**

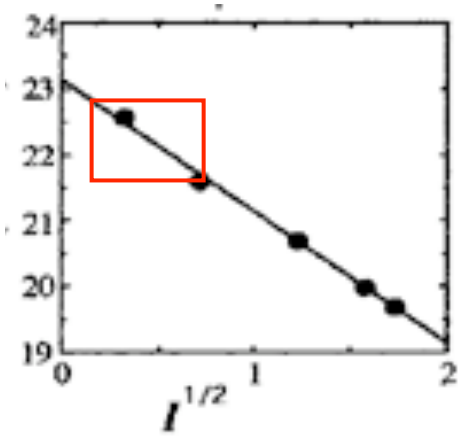
How experimental conditions affect K_d



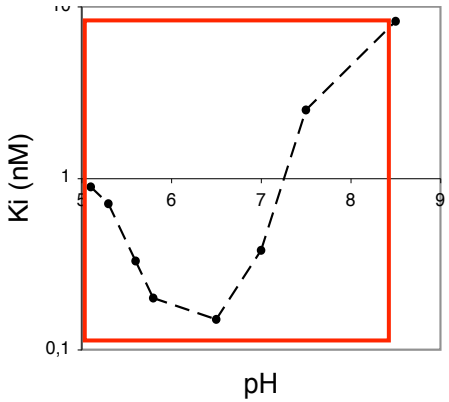
Temperature



Ionic strength



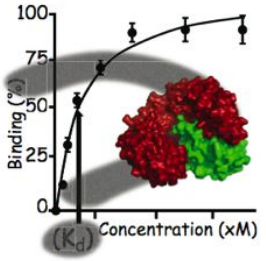
pH



	range	K_d ratio	$\sigma(\Delta G)$ (kcal/mol)
Temperature	20 - 35 °C	2	0.4
Ionic strength	0.1 - 0.5 M	3	0.7
pH	5 - 8.5	53	2.4

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)	10-90	1.4-2.7

Conclusion:

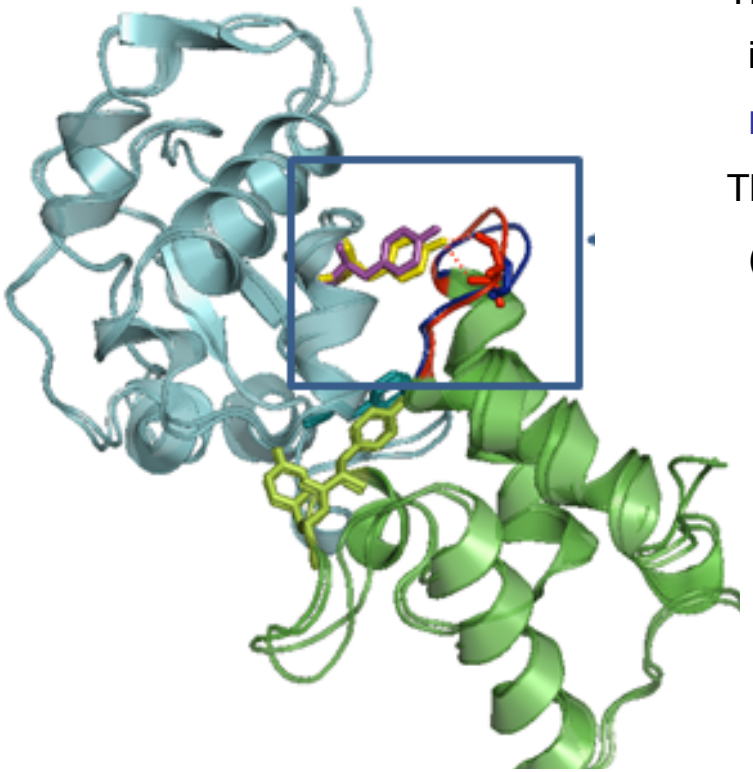
- **Most K_d values in our set are defined to within one order of magnitude**
- **It makes no sense to model or predict ΔG to within better than 1.4 kcal/mol unless one can also model its pH dependence**

Similar structures, different affinities:

Colicin Dnase domain / 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** (1EMV) and the **non-cognate Im2** (68% seq id; 2WPT).

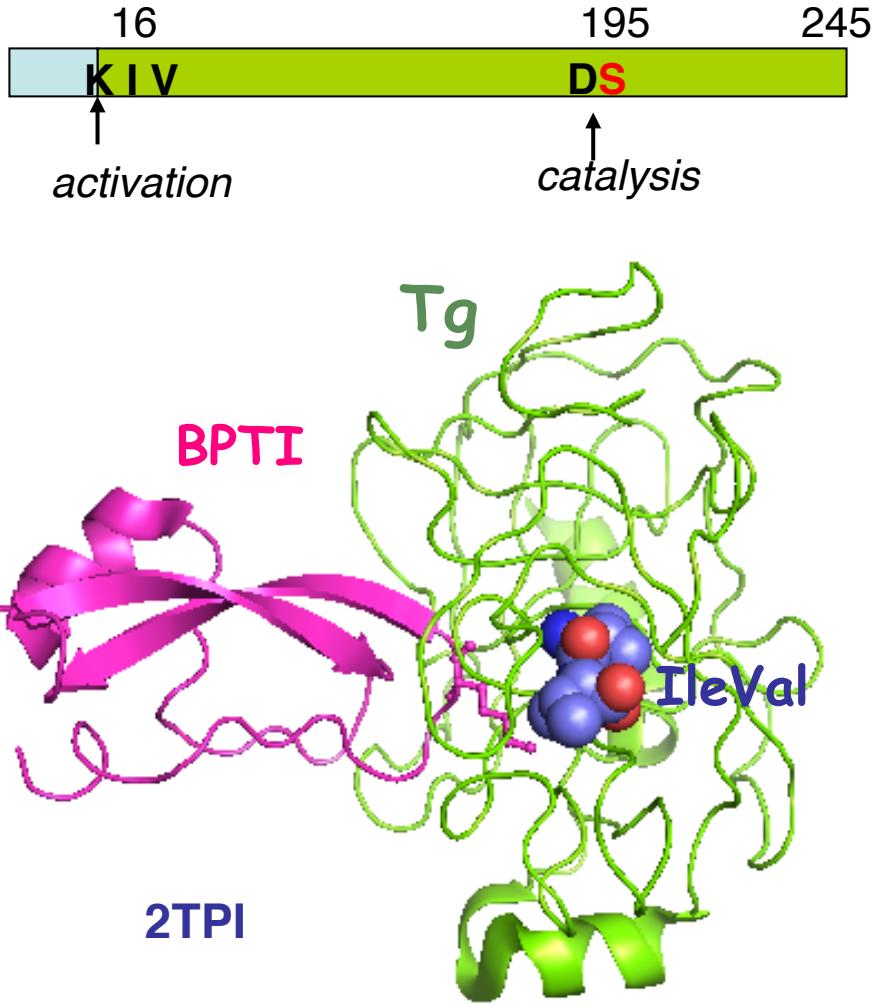
The two complexes have **very similar structures** (rmsd = 0.4 Å), and very **different affinities**

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

Similar structures, different affinities:

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

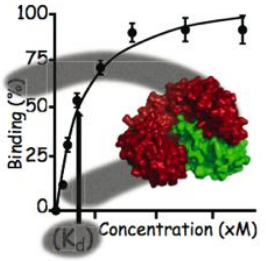
- **BPTI binding induces the same change**
- **addition of the *IleVal* dipeptide also !**

Allosteric interaction:

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

Conclusion (3)

What is new in the structure/affinity benchmark ?



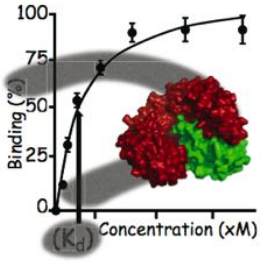
Reliable K_d values for ≈ 80% of the complexes in the Docking Benchmark and built the first version of a database

- Along with the complexes, it contains **unbound structures**
- Nine entries represent **cognate/non-cognate pairs** of complexes,
- Many proteins are **allosteric**
  (trypsinogen, G-proteins, receptors etc…)
- Many displays **large conformation changes** …
- Empirical models must account for their free energy cost !

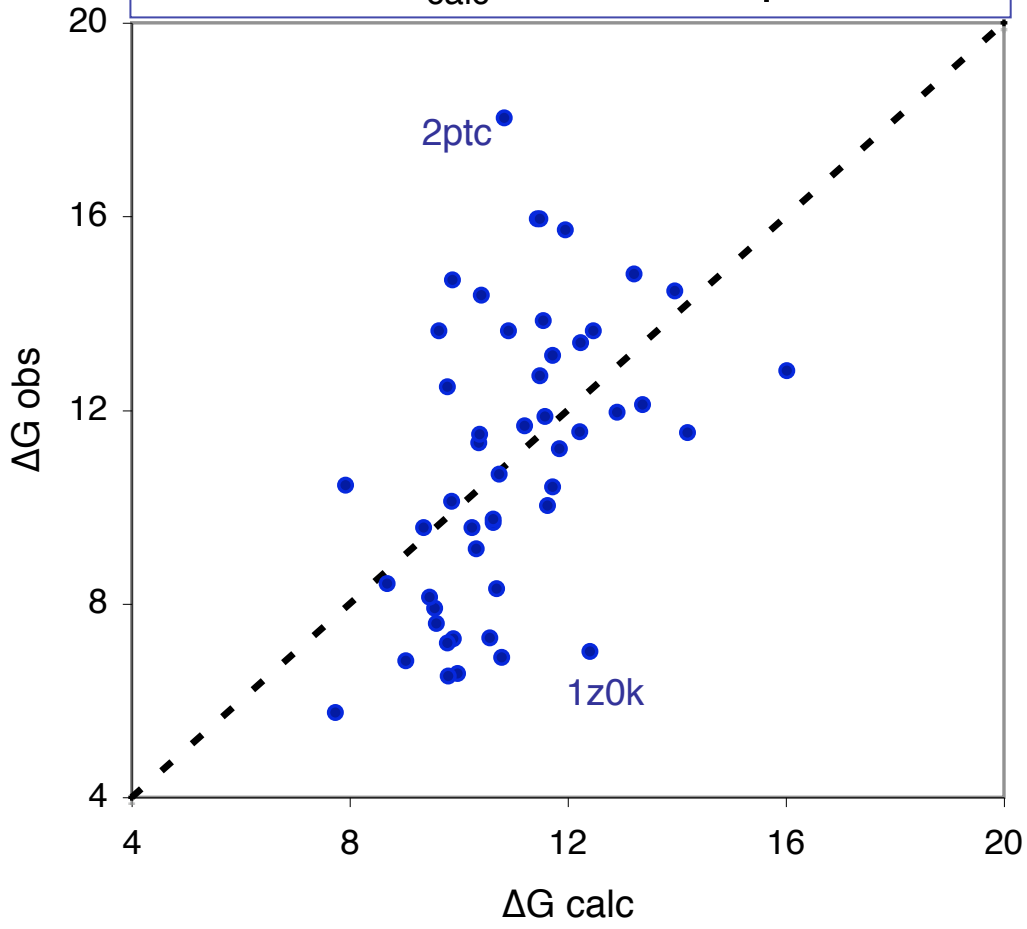
http://bmm.cancerresearchuk.org/~bmmadmin/Affinity

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

Fitting ΔG with one parameter: rigid-body recognition



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



48 of the 145 complexes (33%) display small changes at the interface ($\Sigma\ \delta x^2$ <35 Å^2, I_rmsd below ≈1 Å)

For 46 of them, ΔG_d correlates well with the interface size: the *BSA* accounts for ≈1/3 of the variance

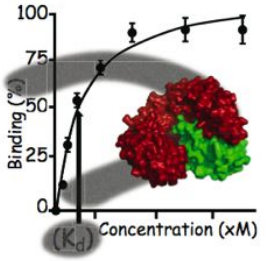
N	*r*	<ΔG_{calc}-ΔG_{obs}> (kcal/mol)	Outliers
48	0.55	2.4	2

The outliers

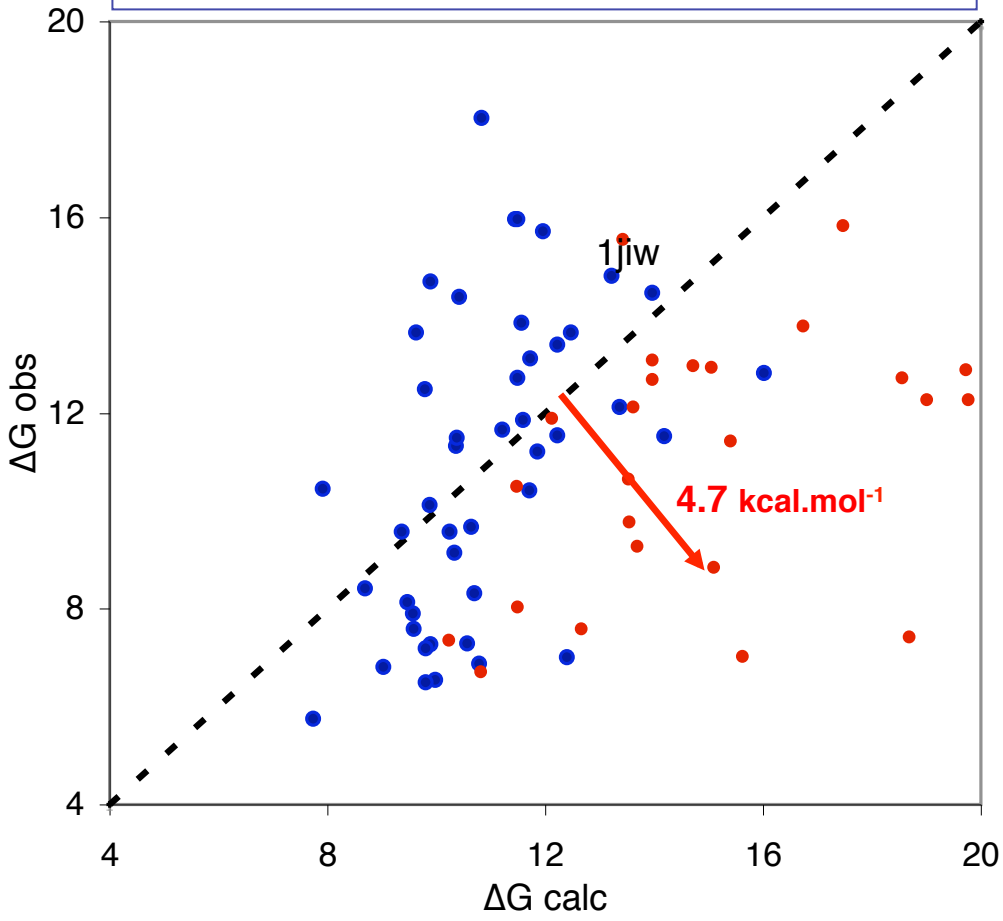
2ptc (trypsin/BPTI) *electrostatics?*

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 Å^2, I_rmsd = 1.5 to 9 Å)

They all yield $\Delta G_{calc} > \Delta G_{obs}$

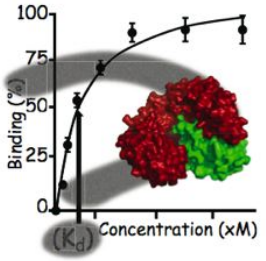
except 1jiw (UEV/ubiquitin), which has a Zn metal bond at the interface.

Taking $\Delta G_{calc} - \Delta G_{obs}$ to be an estimate of the free energy cost of the conformation changes, the maximum is 34 kcal/mol and the mean:

$<\Delta G_{conf}>$ = 4.7 kcal/mol

()

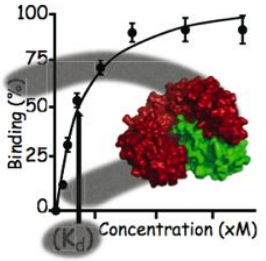
Fitting observed ΔG's with protein-protein docking potentials



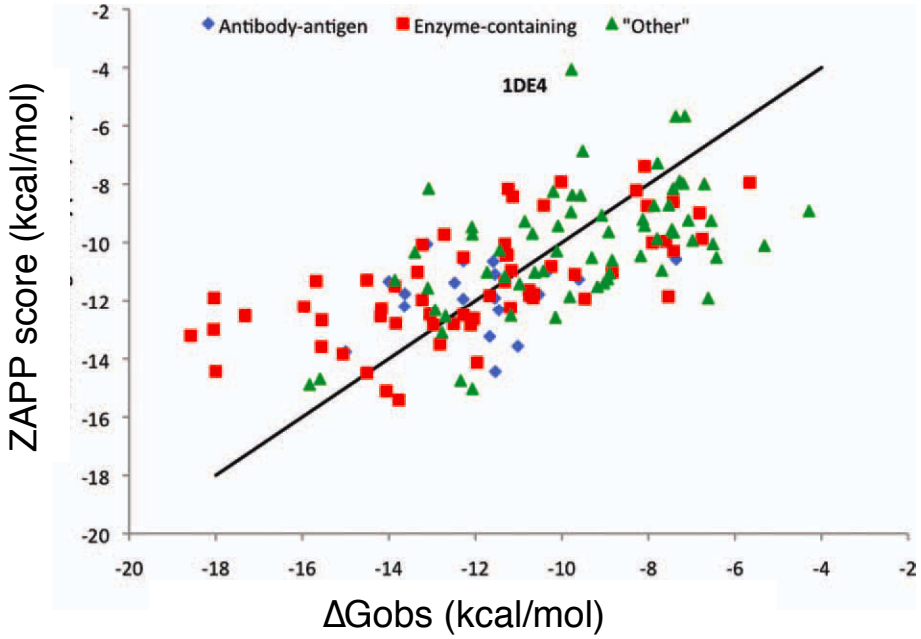
Vreven, Hwang, Pierce & Weng (2012) *Protein Sci.* 21:396

I-rmsd Conformation changes	<1Å small	1-2Å medium	>2Å large	All
Potential	*r* (correlation coefficient to observed ΔG)			
AffinityScore (Audie 2009)	0.46	0.07	0.28	0.25
PyDock (Cheng, Blundell, Fernandez-Recio 2007)	0.21	0.42	0.06	0.26
Rosetta (Gray et al., Baker 2003)	0.61	0.24	0.36	0.41
ZRANK (Pierce & Weng 2007)	0.51	0.11	0.20	0.22
new ZAPP (Vreven et al 2012)	0.66	0.61	0.62	0.63

A multi-parameter fit of ΔG



Vreven et al. (2012) *Protein Sci.* 21:396



Descriptors used in ZAPP	
TB	Tobi-Bahar (2006) residue pair potential
Ros_Sol	Rosetta solvation potential
Ros_HB	Rosetta H-bonding potential
Elec_lrA	Zrank long-range electrostatics
Elec_lrR	(attractive and repulsive terms)
Bur_CS	#buried hydrophobic groups
Loop, helix	#loop and helix residues at interface
MisRes	#residues ordered at interface

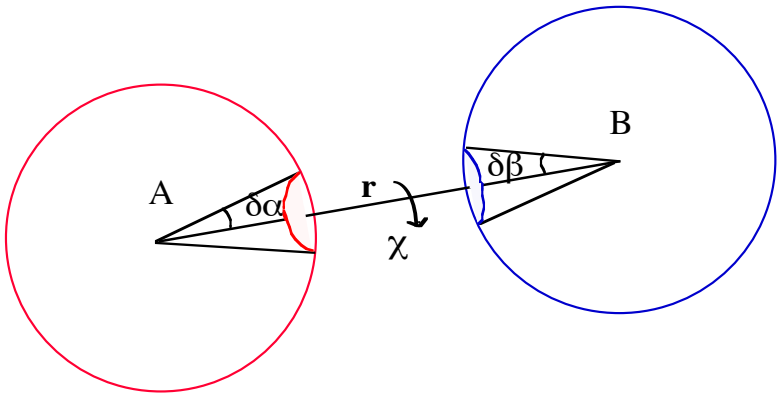
$r = 0.63$ rms $[\Delta G_{calc} - \Delta G_{obs}] = 2.25$ kcal/mol

Kinetics of rigid-body protein-protein recognition

$$k_{on} = \kappa\, q_t\, q_r\, p_r\, k_{coll}$$

$p_r = \pi/16\ \ \delta\alpha^2\, \delta\beta^2\, \delta\chi$

$\approx 10^{-4}$ for $\delta\alpha \approx \delta\beta \approx \delta\chi \approx 20°$



Janin (1997) *Proteins* 28:153

$\kappa \approx 0.5$ transmission coefficient

$k_{coll} \approx 6.6.10^9\ M^{-1}.s^{-1}$

(Einstein-Schmoluchowski, 300 K in water)

For many **antibody-antigen** (including lysozyme-HyHEL5) and **protease-inhibitor** complexes, electrostatics play a minor part:

$k_{on} = 10^5\text{-}10^6\ M^{-1}.s^{-1}$

with $q_t \approx q_r \approx 1$, $p_r = 10^{-4}\text{-}10^{-5}$

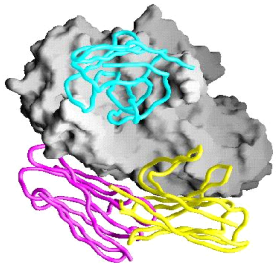
In **barnase-barstar**, electrostatic steering is important at low/moderate ionic strength:

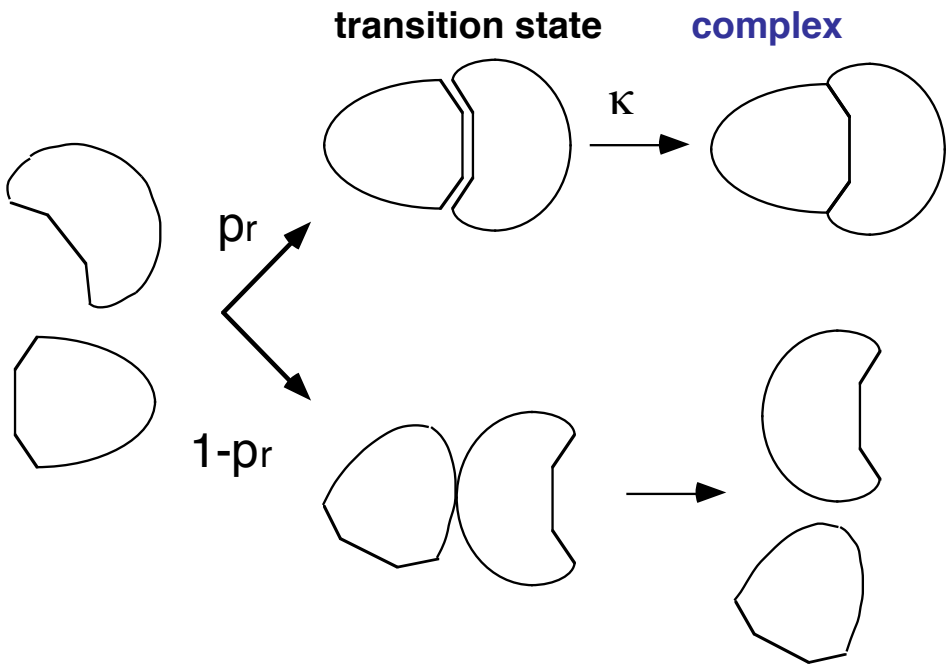
$q_t\, q_r = 10^3\text{-}10^6$

$k_{on} = 10^7\text{-}10^{10}\ M^{-1}.s^{-1}$

(Schreiber & Fersht, 1998)

Modeling the rigid-body association reaction





$$k_a = \kappa\, q_{lr}\, p_r\, k_{coll}$$

- κ transmission coefficient
- $k_{coll} \approx 10^{10}\ M^{-1}.s^{-1}$ collision rate (Einstein-Smoluchowski)
- p_r probability of the correct orientation
- q_{el} long-range electrostatics

predicts $k_a = 10^5\text{-}10^6\ M^{-1}.s^{-1}$

assuming

- efficient conversion of the transition state to the product complex ($\kappa \approx 0.5$)
- weak long-range interactions ($q_{lr} \approx 1$)
- that in the transition state, the subunit orientation is determined to within 15-20° ($p_r \approx 10^{-4}\text{-}10^{-5}$)

Janin (1997) *Proteins* 28:153

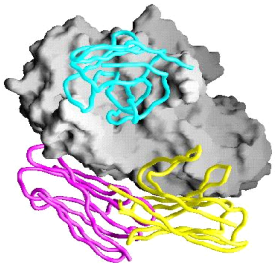
Northup & Erickson (1992) *PNAS* 89:3338

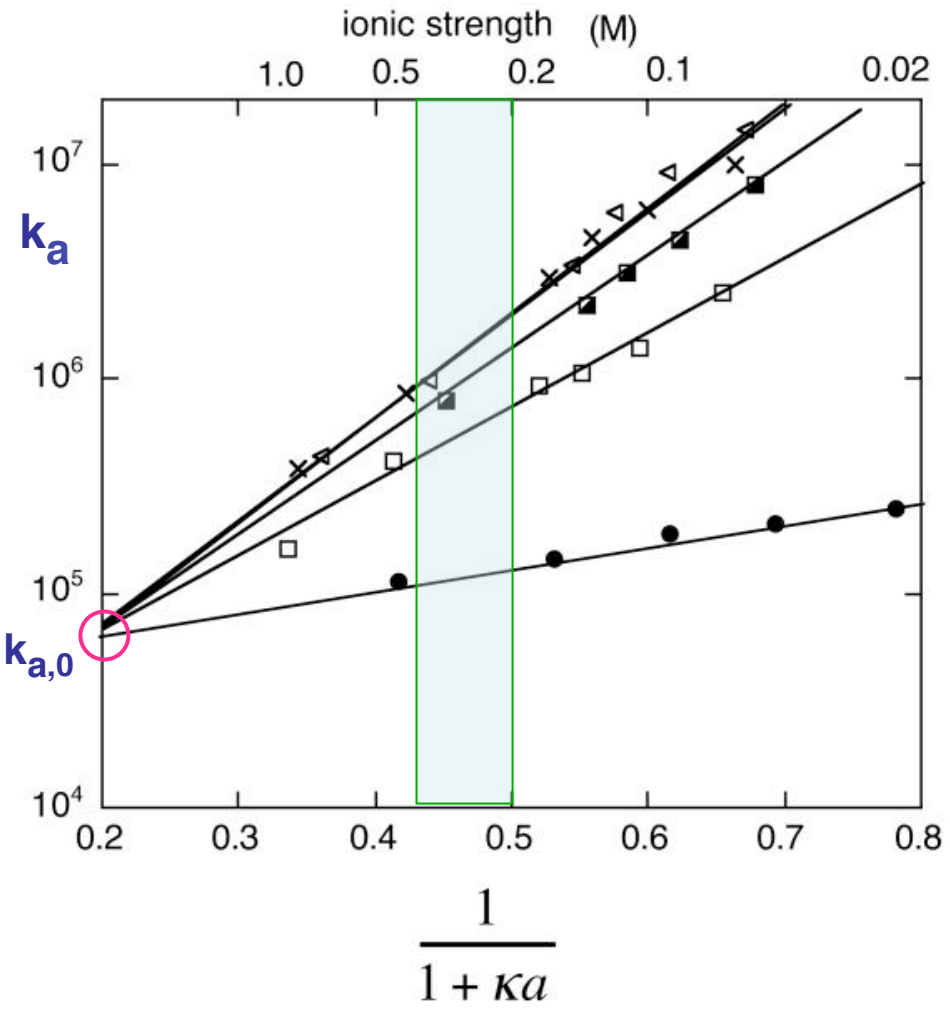
Zhou (1993) *Biophys. J.* 64:1711

Gabdoulline &Wade (2001) *JMB* 306:1139

Modeling long-range electrostatics effects on k_a

Schreiber, Haran & Zhou (2009) *Chem. Rev.* 109:839

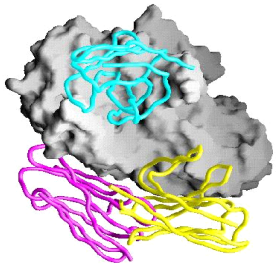




- At moderate ionic strength (0.2 M), long-range effects change k_a by a factor $q_{lr} < 20$.
- They are modeled accurately by Debye-Hückel screening

The association rate constants of wild-type and mutant TEM1-BLIP complexes at different salt concentrations.

Flexible recognition:
Conformer selection vs. Induced fit



Conformer selection

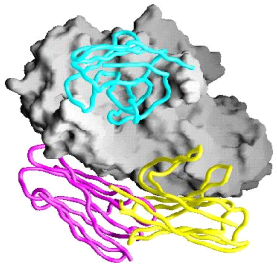
- the free protein is in equilibrium between two or more conformations
- only **bound-like** conformers can make productive collisions; if they form a fraction α (probably <<1) of the population
- the association rate becomes $\alpha k_a << k_a$

Induced fit

Koshland (1958, *PNAS* 44:98)

- the free protein is in the **unbound** conformation
- interactions made in the transition state induce a change to the **bound** conformation;
- the change has a high probability β to be occur before the transition state dissociates; the transmission coefficient becomes $\beta\kappa$
- the association rate becomes βk_a (possibly $\approx k_a$)

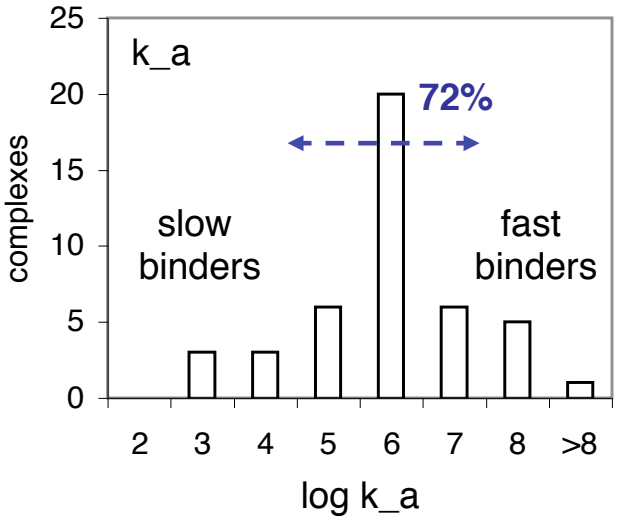
Rate constants in the structure/affinity benchmark

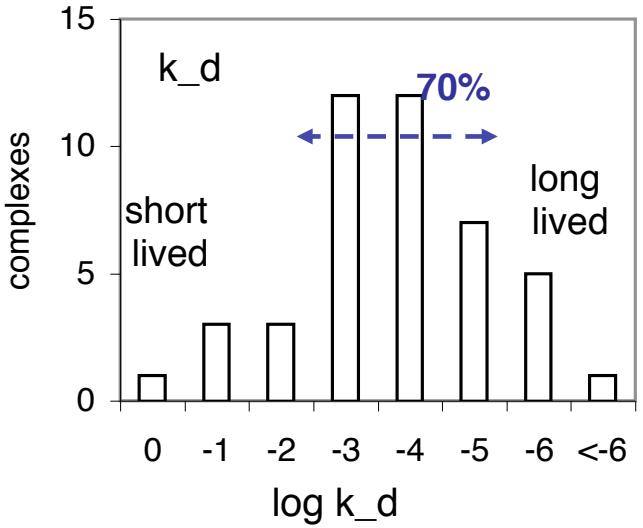


Moal & Bates (2012) *PLOS Comp Biol* 8:e1002351

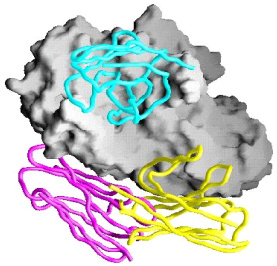
Kinetic data are available for 44 out of the 144 complexes of the structure/affinity benchmark.

- k_a is in the range 10^4-10^7 $M^{-1}s^{-1}$ for 72% of the complexes fast binders
- k_d is in the range 10^{-5}-10^{-2} s^{-1} for 70% " "





Fitting observed rate constants



Moal & Bates (2012) *PLOS Comp Biol* 8:e1002351

- 23 descriptors for 44 k_a and 44 k_d values.
- models evaluated on all 144 K_d values (= k_d/k_a) of the structure/affinity benchmark.

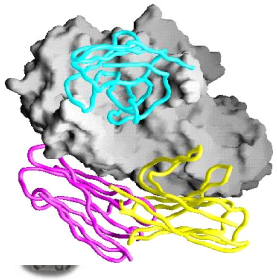
Molecular descriptors with a high correlation to log k_a

Descriptor	*r*	
BSA	*0.24*	
DFIRE_EBU	-0.47	Energy change btween bound and unbound (DFIRE)
OPUS_PSP_EBU	-0.40	id. (OPUS force field)
NUM_HB	0.39	Number of interface H-bonds
H_BOND_ENS	-0.35	H-bonding potential (FireDock)
ROS_HBOND_UB	-0.35	id. (PyRosetta)
ATOM_P	0.39	Fraction of polar atoms at interface

EBU (from DFIRE of OPUS) represents the energy cost of conformation changes. Its high correlation to log k_a suggests a predominance of **conformer selection**.

A model with only 3 parameters (the two descriptors DFIRE_EBU and NUM_HB and a constant) **predicts log k_a** to within 0.8 RMS.

Conclusion (4)



Rigid-body recognition:

a simple geometric model of translational/rotational diffusion accounts for observed rates of association; except at low ionic strength, long-range electrostatics plays only a minor role. Thus:

- The rate of dissociation largely determines K_d;
- Short-range interactions govern affinity and specificity.

Flexible recognition:

slow binding is the exception, either:

- the conformation changes are fast (induced fit mechanism), or
- the competent species are highly populated (conformer selection)
- the correlation with EBU suggests that the second mechanism may be rather common