

Free energy simulations: theory and applications

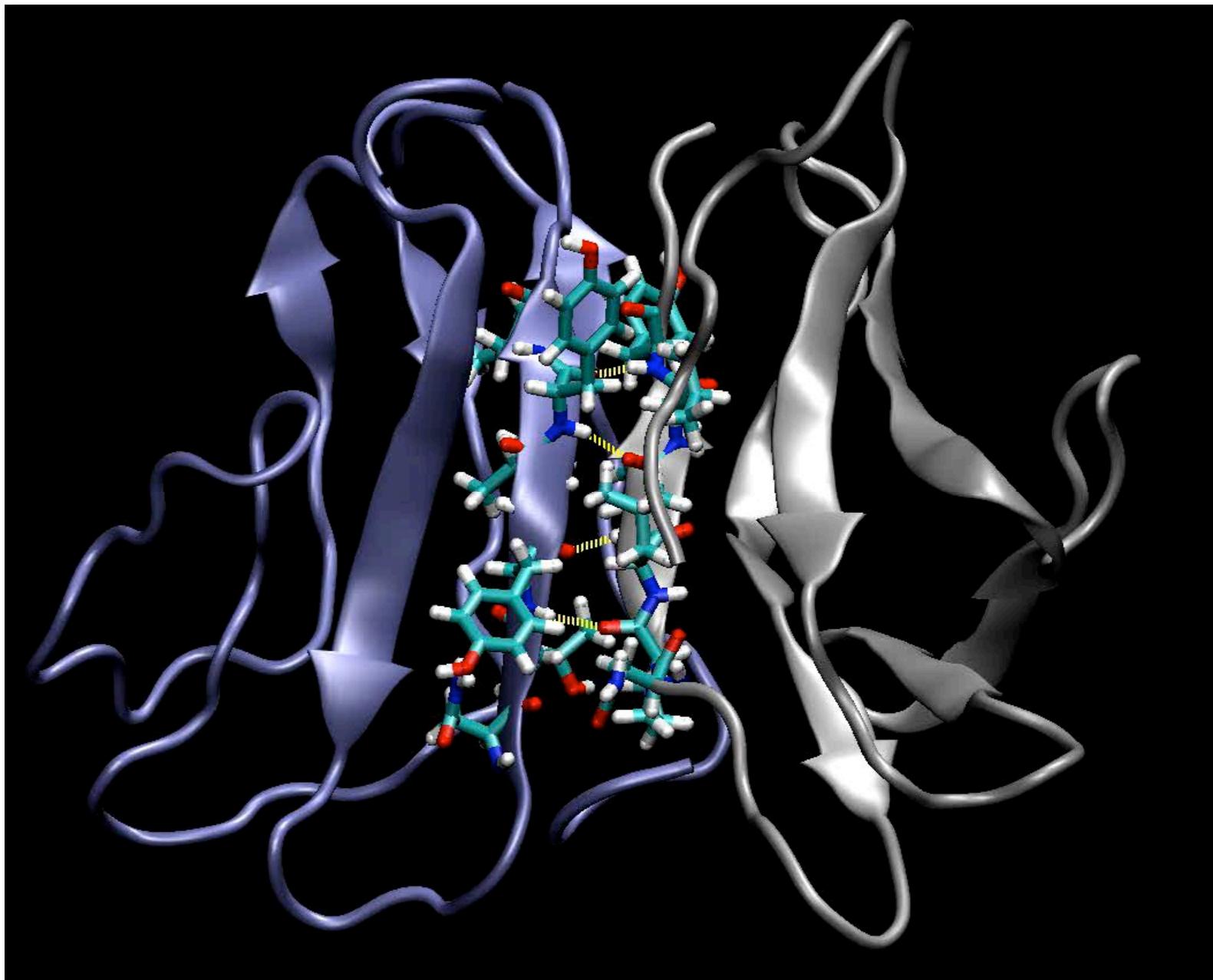
O. Michelin^(1,2,4)

(1) Ludwig Institute for Cancer Research
Epalinges, Switzerland

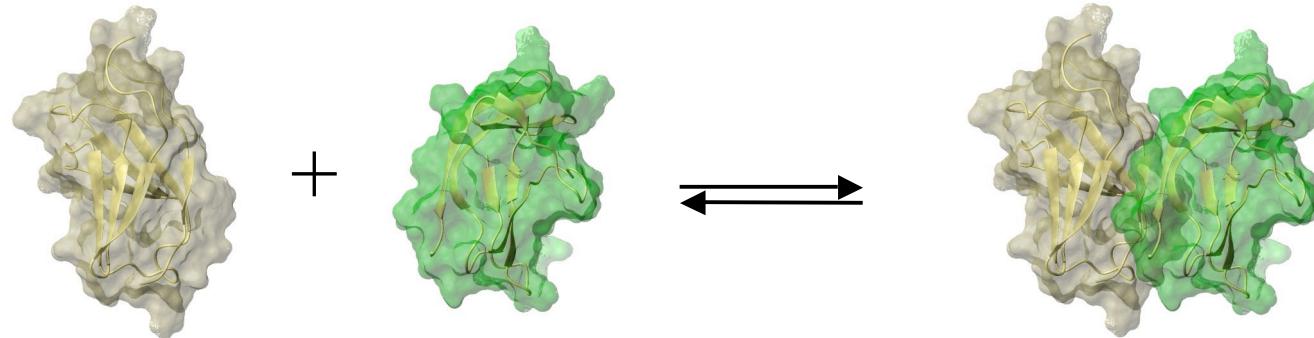
(2) Swiss Institute for Bioinformatics
Epalinges, Switzerland

(4) Centre Pluridisciplinaire d'oncologie
CHUV, Lausanne, Switzerland

Dynamical aspects of molecular recognition



Free energy: classical definition



The free energy is the energy left for once you paid the tax to entropy:

$$\Delta G = \boxed{\Delta H} - T \boxed{\Delta S}$$

Enthalpic

- Hydrogen bonds
- Polar interactions
- Van der Waals interactions
- ...

Entropic

- Loss of degrees of freedom
- Gain of vibrational modes
- Loss of solvent/protein structure
- ...

Theoretical Predictions:

- *Approximate:* empirical formula for all contributions
- *Exact:* using statistical physics definition of G

$$G = -K_B T \ln(Z)$$

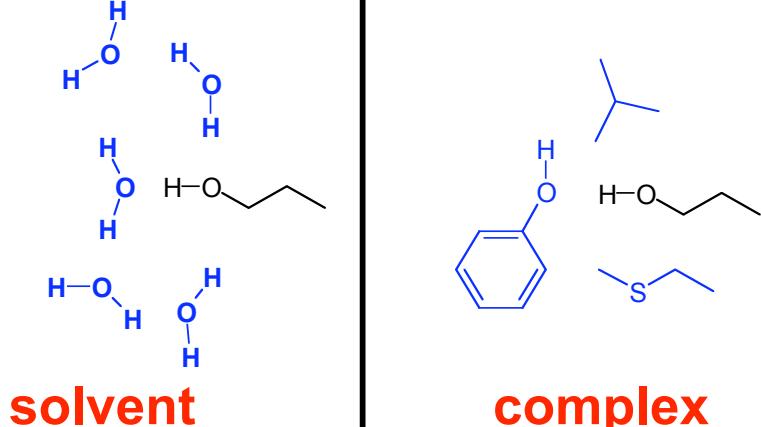
Examples of factors determining the binding free energy

Electrostatic interactions

- Strength depends on microscopic environment (ϵ)
- Case of hydrogen bonds

Neutral : $-1.2 \pm 0.6 \text{ kcal/mol}$

Charge assisted : $-2.4 \text{ to } -4.8 \text{ kcal/mol}$



$E_{\text{H-bond}}(\text{solv.}) - E_{\text{H-bond}}(\text{comp.})$
determines if H-bonds
contributes to affinity or not

Unpaired polar groups upon binding
are detrimental

Strong directional nature

→ Specificity of molecular recognition

Free energy: statistical mechanics definition

$$G = -k_B T \ln(Z)$$

where

$$Z = \sum_i e^{-\beta E_i}$$

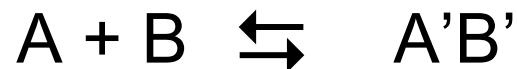
is the partition function

Free energy differences between 2 states (bound/unbound, ...) are, therefore, ratios of partition functions

$$\Delta G = G_A - G_B = -k_B T \ln \left(\frac{Z_A}{Z_B} \right)$$

Free energy simulations aim at computing this ratio using various techniques.

Relation with chemical equilibrium



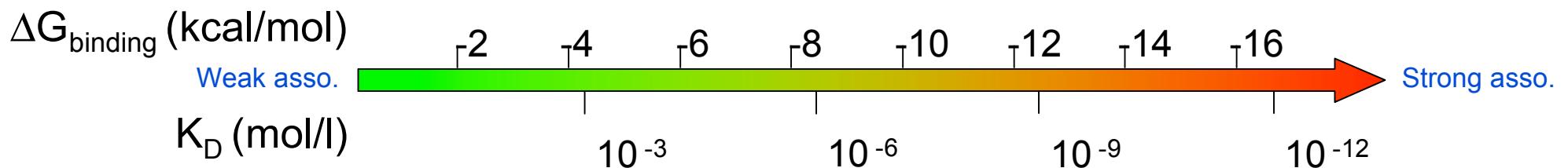
$$K_A = K_b = \frac{[A'B']}{[A][B]}$$



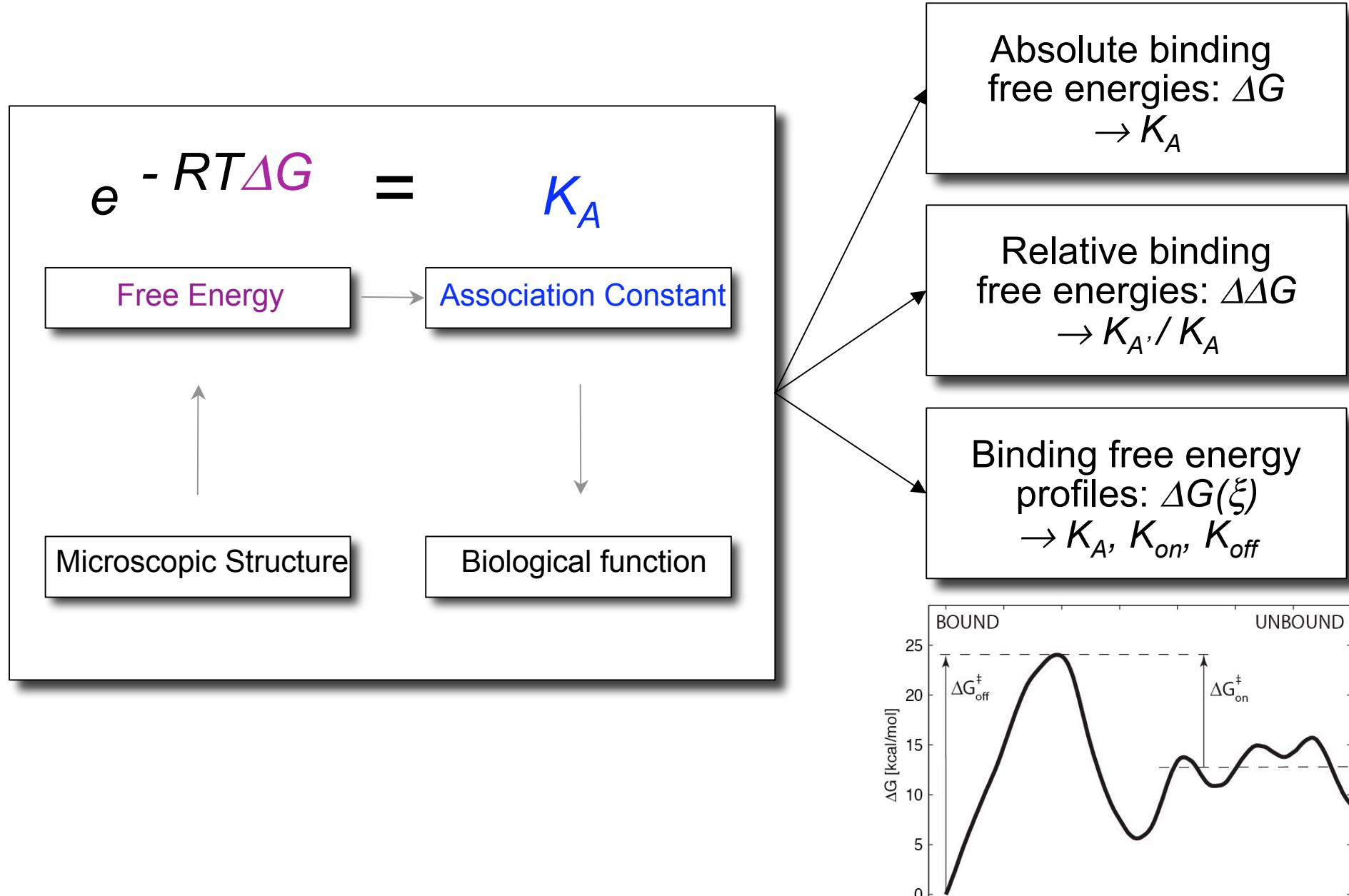
$$K_D = K_i = \frac{[A][B]}{[A'B']}$$

K_b : binding constant, K_d : dissociation constant, K_i : inhibition constant

$$\Delta G_{\text{binding}} = -RT \ln K_A = RT \ln K_D = \Delta H - T\Delta S$$



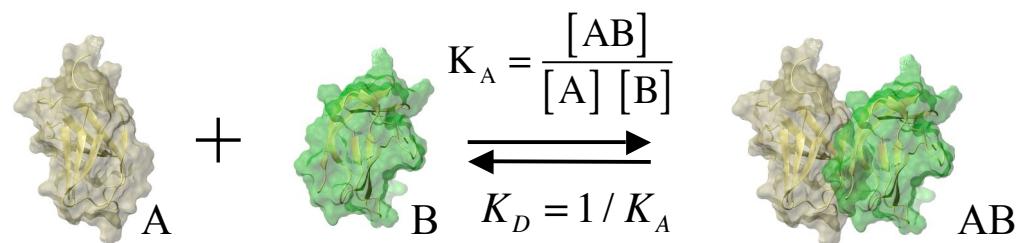
Connection micro/macroscopic: thermodynamics and kinetics



The free energy is the main function behind all process

A) Chemical equilibrium

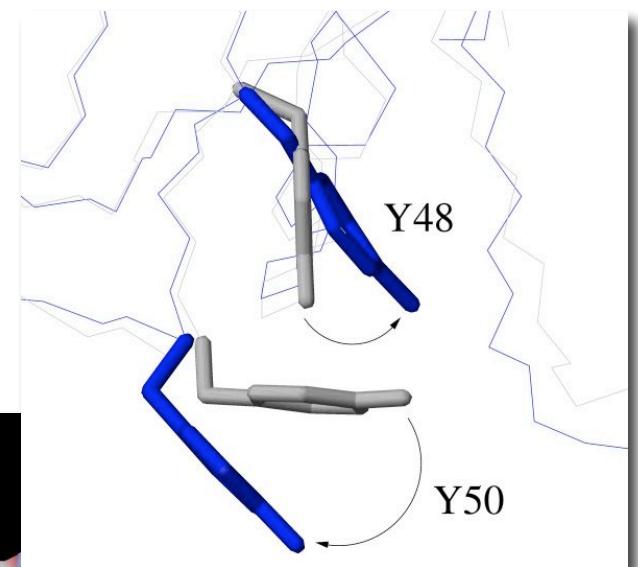
$$\Delta G_{\text{binding}} = RT \ln K_A$$



B) Conformational changes

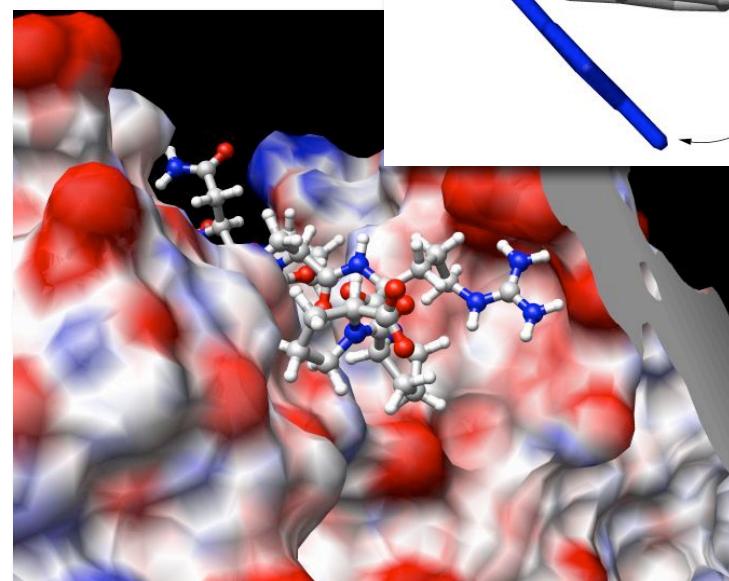
$$\Delta G_{\text{conf}} = k_B T \ln \frac{P_{\text{Conf1}}}{P_{\text{Conf2}}}$$

$$R = k_B N_A$$



C) Ligand binding

$$\Delta G_{\text{binding}} = k_B T \ln \frac{P_{\text{Unbound}}}{P_{\text{Bound}}}$$

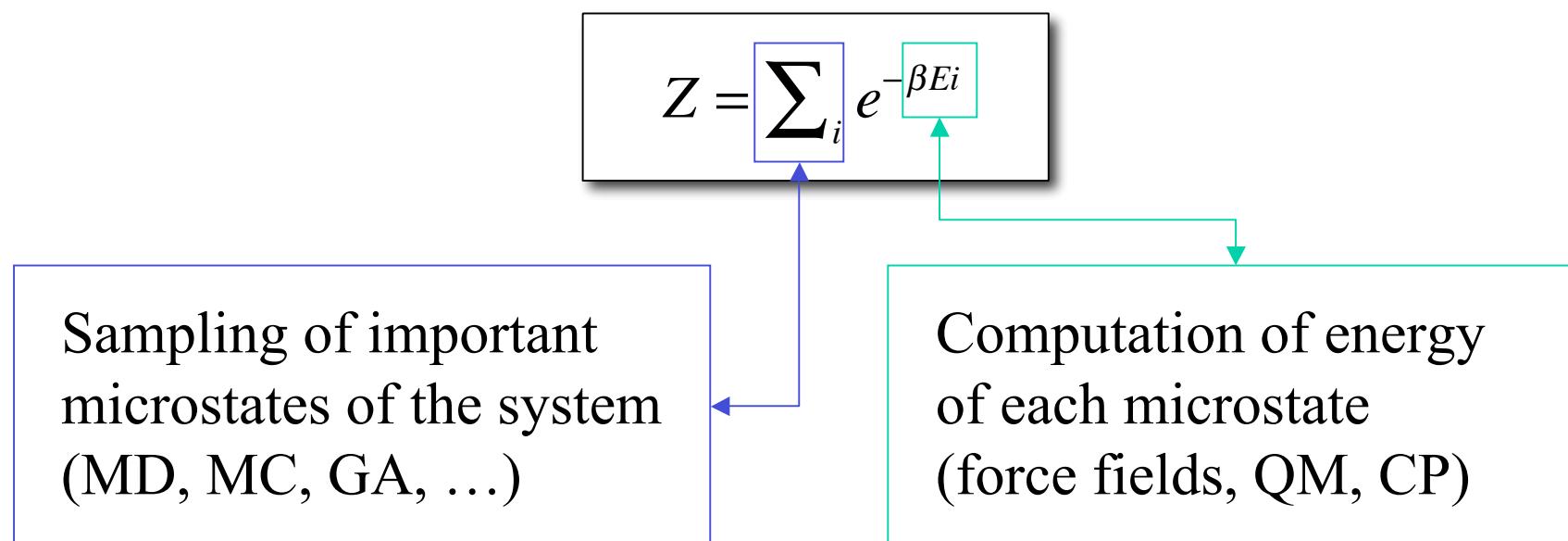


D) ...

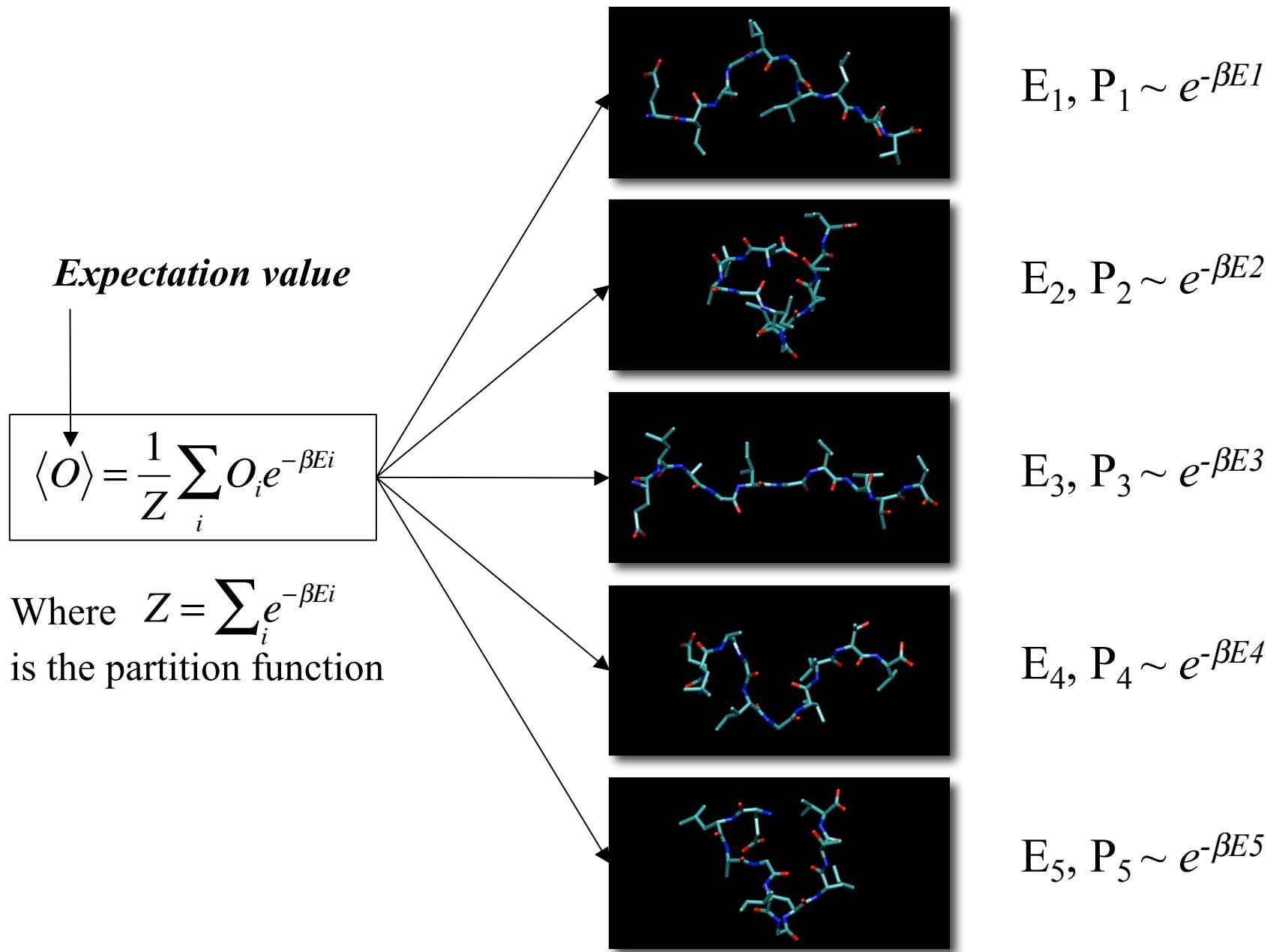
Free energy: computational approaches

$$\Delta G = G_A - G_B = -k_B T \ln \left(\frac{Z_A}{Z_B} \right)$$

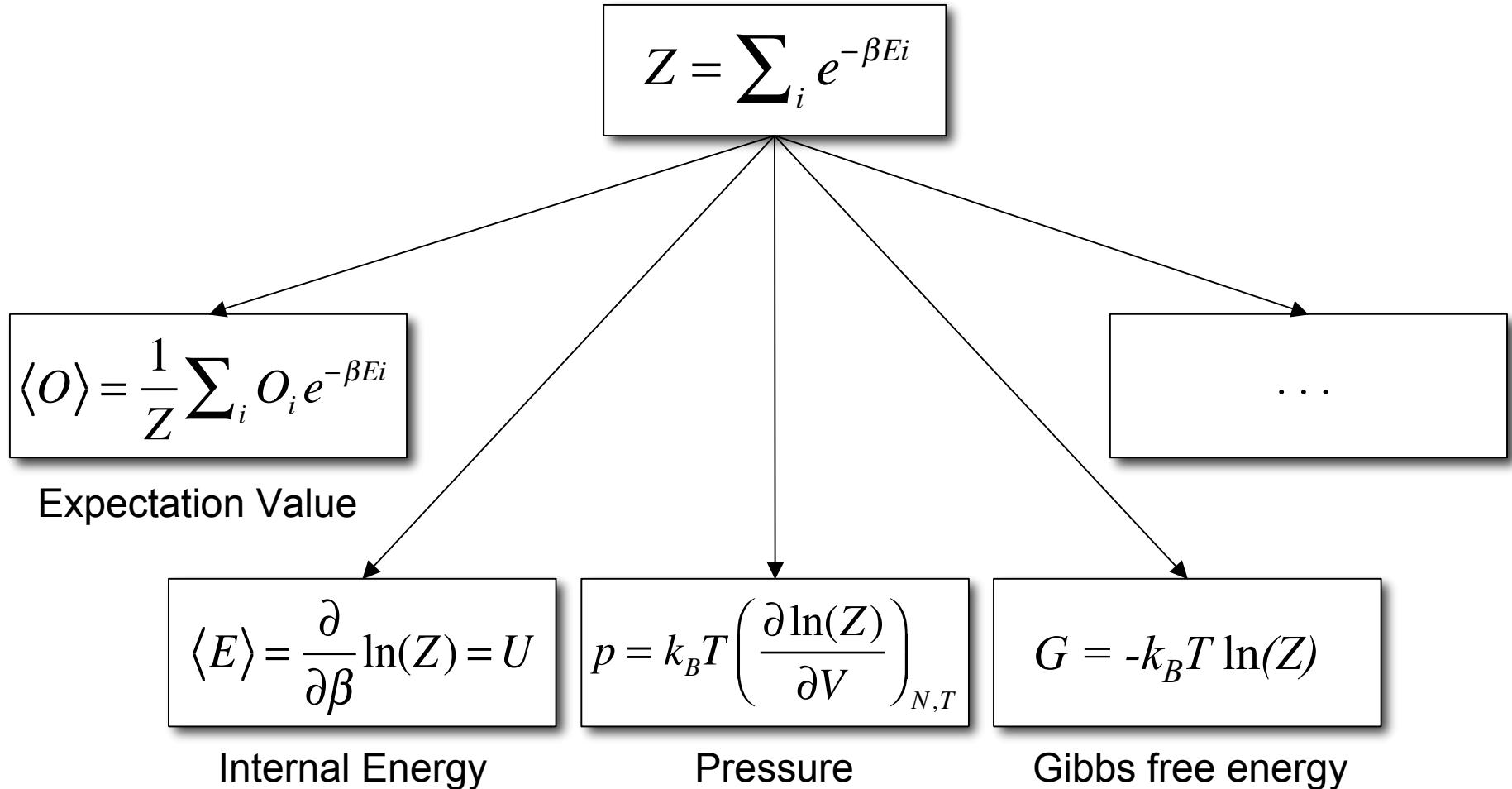
Free energy simulations techniques aim at computing ratios of partition functions using various techniques.



Connection micro/macrosopic: intuitive view

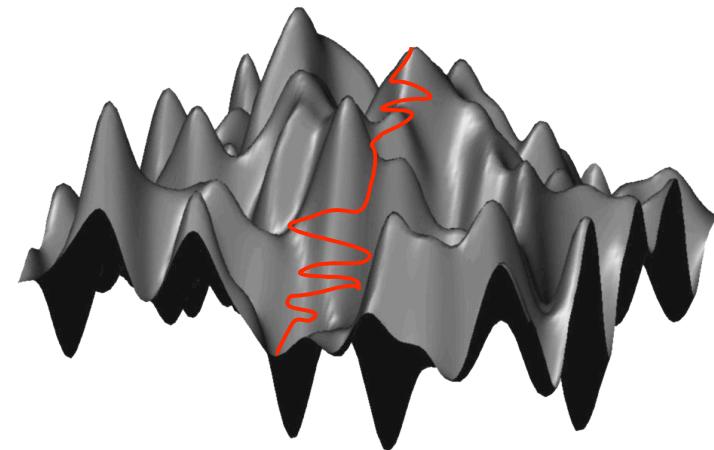
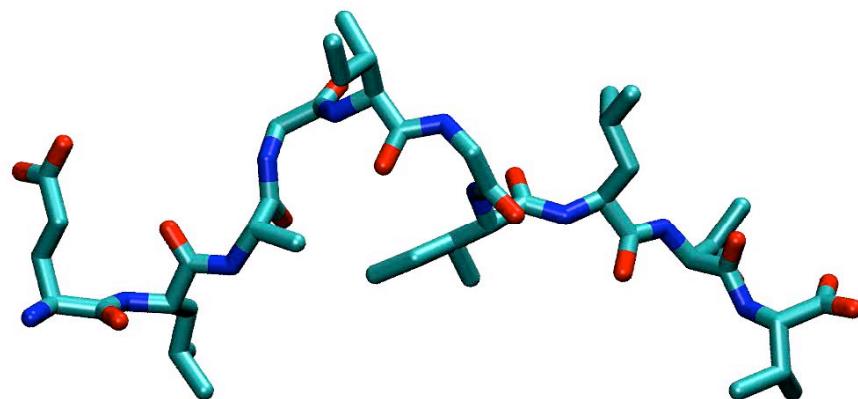


Central Role of the Partition Function



Molecular Modeling Principles

1) Modeling of molecular interactions



Free energy landscape

2) Simulation of time evolution (Newton)

3) Computation of average values

$$O = \langle O \rangle_{\text{Ensemble}} = \langle O \rangle_{\text{Temps}} \quad (\text{Ergodicity})$$

↑
Macroscopic value ↑
Average simulation value

Connection
microscopic/
macroscopic

The CHARMM Force Field

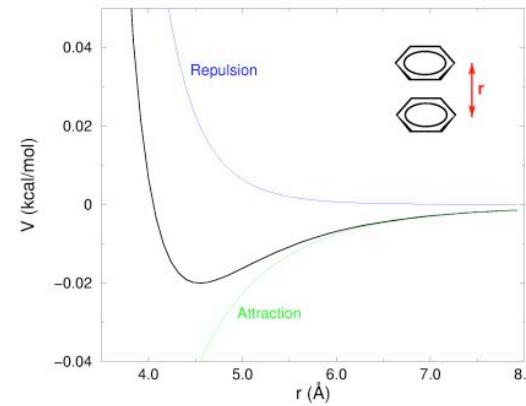
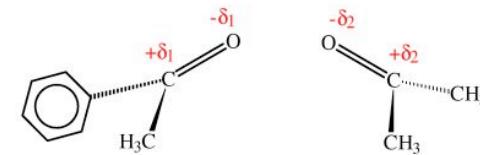
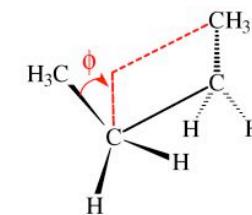
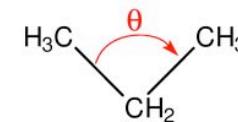
$$V = \sum_{\text{Bonds}} K_b (b - b_0)^2 + \sum_{\text{Angles}} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{Improper}} K_\omega (\omega - \omega_0)^2$$

$$+ \sum_{\text{Dihedrals}} K_\phi [1 - \cos(n_\phi \phi - \delta_\phi)]$$

$$+ \sum_{i>j} \frac{q_i q_j}{4\pi\epsilon} \frac{1}{r_{ij}}$$

$$+ \sum_{i>j} 4\epsilon_{ij} \left[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6 \right]$$



MD Techniques: Microcanonical sampling

For an Hamiltonian of the form $H(\mathbf{p}, \mathbf{r}) = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + \phi(r_1, \dots, r_{3N})$

in cartesian coordinates, the Hamilton equations of motion reduce to the Newton equations

$$\dot{r}_i = \frac{p_i}{m_i} \quad m_i a_i = \frac{-\partial}{\partial r_i} \phi(\mathbf{r}) \quad i=1, \dots, N$$

Several numerical methods have been developed to integrate these equations. One of the most stable integrator is that of *Verlet*: for a small time increment Δt , one can use a Taylor expansion of the function $\mathbf{r}(t)$:

$$\begin{aligned}\mathbf{r}_i(t+\Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \Delta t + 1/2 \mathbf{a}_i(t) \Delta t^2 + \dots \\ \mathbf{r}_i(t-\Delta t) &= \mathbf{r}_i(t) - \mathbf{v}_i(t) \Delta t + 1/2 \mathbf{a}_i(t) \Delta t^2 + \dots\end{aligned}$$

Adding those equations, one gets $\mathbf{r}(t+\Delta t)$ as a function of $\mathbf{r}(t)$ and $\mathbf{r}(t-\Delta t)$.

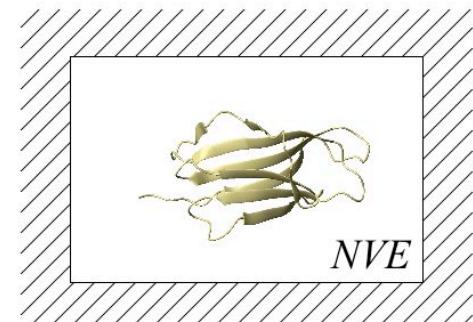
$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \mathbf{a}_i(t) \Delta t^2$$

- In practice, this scheme is applied iteratively, starting from the initial conditions.
- Velocities are postcomputed as $\mathbf{v}(t) = [\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)] / 2\Delta t$.
- Positions are correct up to Δt^4 and velocities to Δt^2 .
- This scheme conserves energy with very good accuracy.

MD Techniques: Sampling of the various ensembles

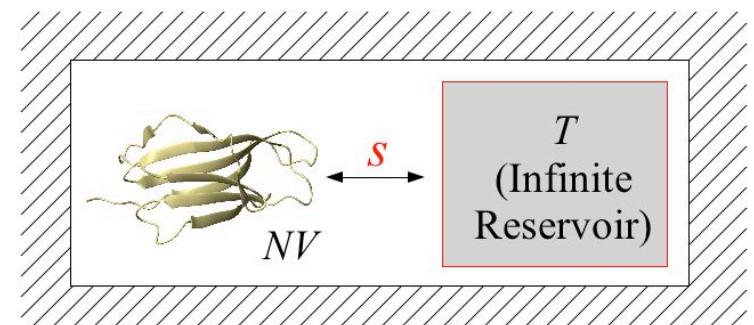
1) Microcanonical ensemble (constant N, V, E)

$$H(\mathbf{p}, \mathbf{q}) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i s^2} + \phi(\mathbf{q})$$



2) Canonical ensemble (constant N, V, T)

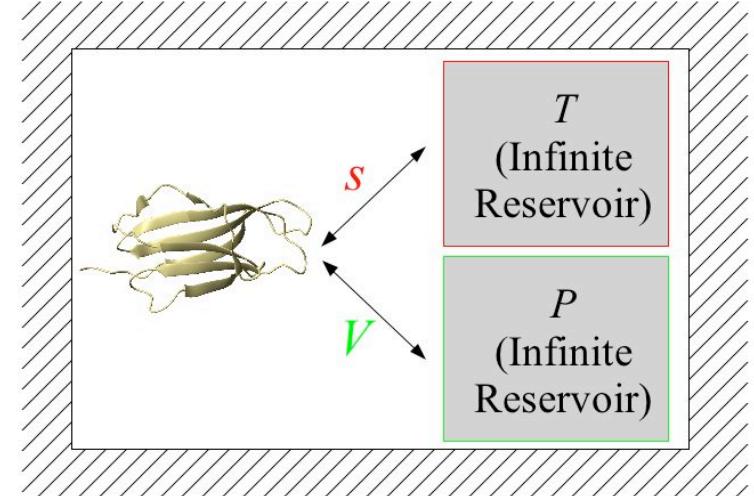
$$H(\mathbf{p}, \mathbf{q}, p_s, s) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i s^2} + \phi(\mathbf{q}) + \boxed{\frac{p_s^2}{2Q} + (3N+1)kT \ln s}$$



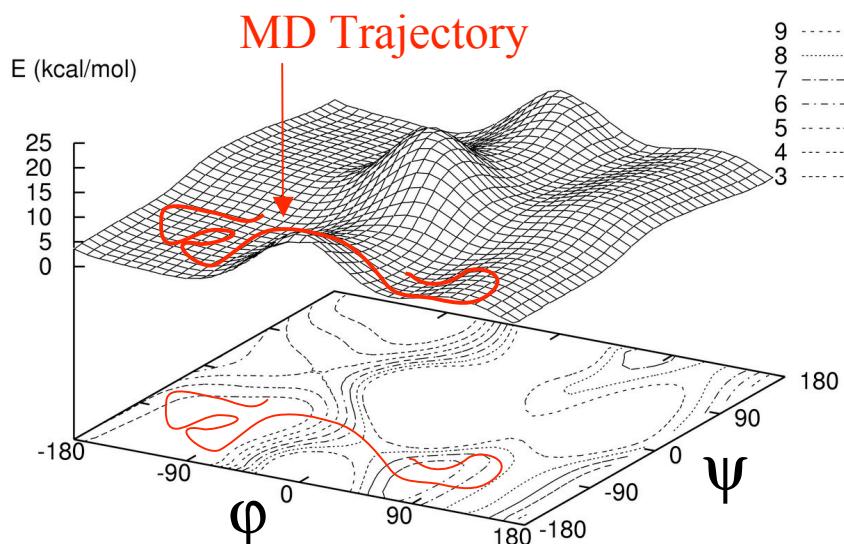
3) Isothermic-isobaric ensemble (constant N, P, T)

$$H = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i s^2 V^{2/3}} + \phi(V^{1/3} \mathbf{q}) + \boxed{\frac{p_s^2}{2Q} + (3N+1)kT \ln s}$$

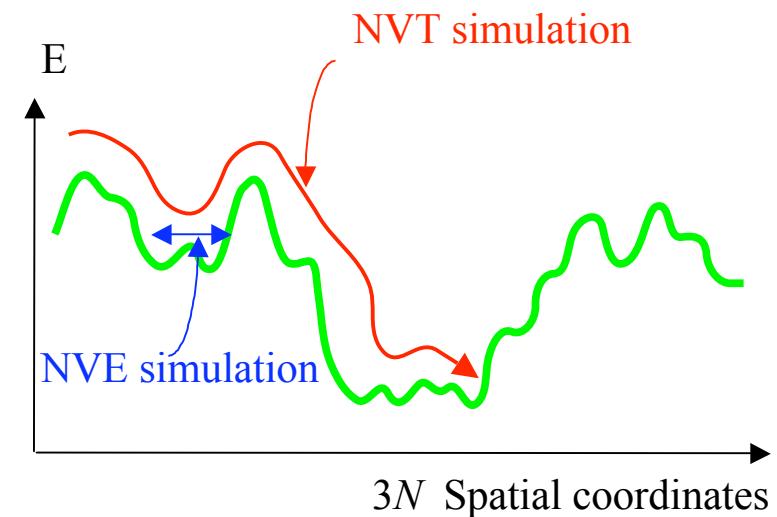
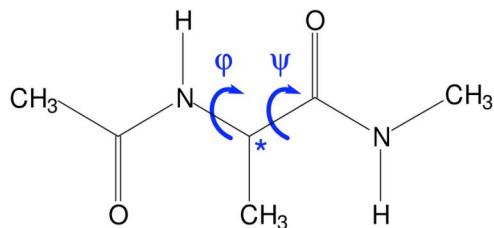
$$+ \boxed{\frac{p_V^2}{2W} + P_{ex} V}$$



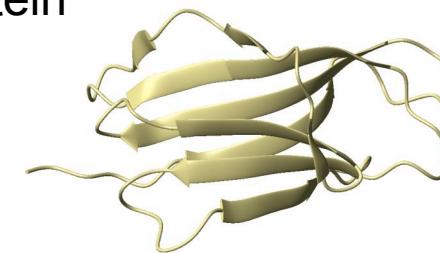
Ergodic Hypothesis



Dialanine



Protein



$$\langle O \rangle_{Ensemble} = \frac{1}{Z} \int O(\varphi, \psi) e^{-\beta E(\varphi, \psi)} d\varphi d\psi = \overset{?}{=} \frac{1}{\tau} \int_0^\tau O(t) dt = \langle O \rangle_{Time}$$

Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle \quad \Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

Quantitative Structure Activity Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i$$

$$\Delta G = F(\mathbf{X}) \quad (\mathbf{X} \text{ is a descriptor})$$

CPU Time



Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle \quad \Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

Quantitative Structure Activity Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i$$

$$\Delta G = F(\mathbf{X})$$

(\mathbf{X} is a descriptor)

CPU Time

Theoretical approaches for the estimation of binding affinities

Without 3D structure of the complex

- 2D - QSAR
- 3D - QSAR

With 3D structure of the complex

- Knowledge based approaches
 - regression based methods
 - potential of mean force
- Free energy simulation
- Linear interaction energy
- Binding free energy decomposition (MM-PBSA, MM-GBSA)

Quantitative Structure Activity Relationships (QSAR)

(Drug design)

Assumption

Chemical similarity of ligands → Similarity of biological response

Affinity is a function of the ligand physico-chemical properties

Advantage

No need for structural information about the target

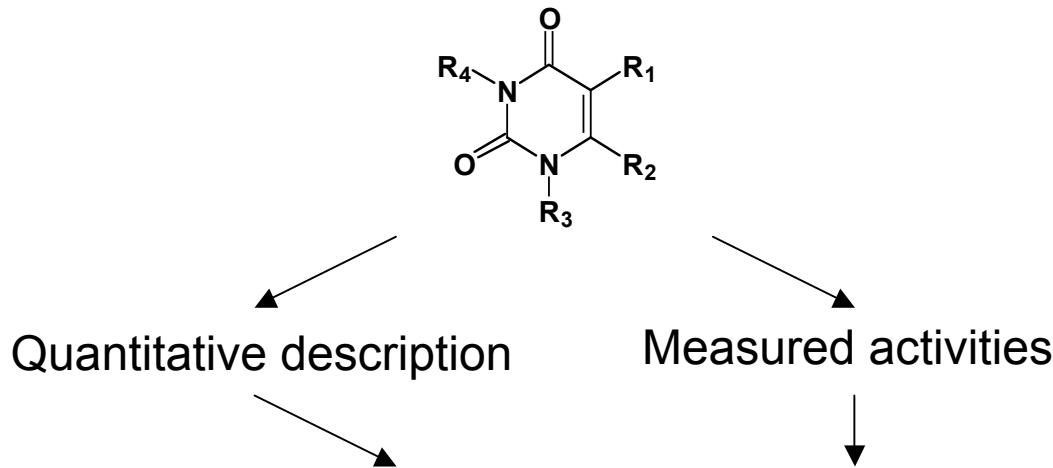
Requirements (drawbacks)

Known affinities for a series of ligands

Structurally related ligands or similar binding modes

2D - QSAR

n structurally related molecules



Descriptors (X_i):

- **Substituents**: surface, volume, electrostatics (σ), hydrophobicity (π), partial charges, etc...
- **Molecule** : volume, MR, HOMO, dipole moment, etc...

$$\Delta G_{\text{bind}} = k_0 + \sum k_i X_i$$

$$\Delta G_{\text{bind}} = \text{neural network } (X_i)$$

C. Hansch and T. Fujita, JACS, 1964, 86, 1616

S.S. So and M. Karplus, J. Med. Chem., 1996, 39, 1521

2D - QSAR

Limitations

Limited to structurally related molecules

Needs the experimental activity of a series ligands

➡ Not for *ab initio* studies

Overfitting

➡ - Method for selecting the descriptor (genetic algorithm)
- Estimation of the predictive ability
(test set, randomization test, Leave-One-Out method, ...)

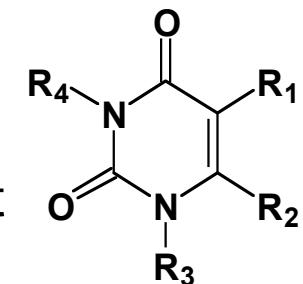
Use limited to the descriptor's range in the training set :

If only hydrophobic groups at R₁ in the training set

➡ Influence of a hydrophilic group at R₁ ?

If only methyl, ethyl, propyl, butyl at R₁ in the training set

➡ Contribution of pentyl, hexyl, etc... ?



3D - QSAR

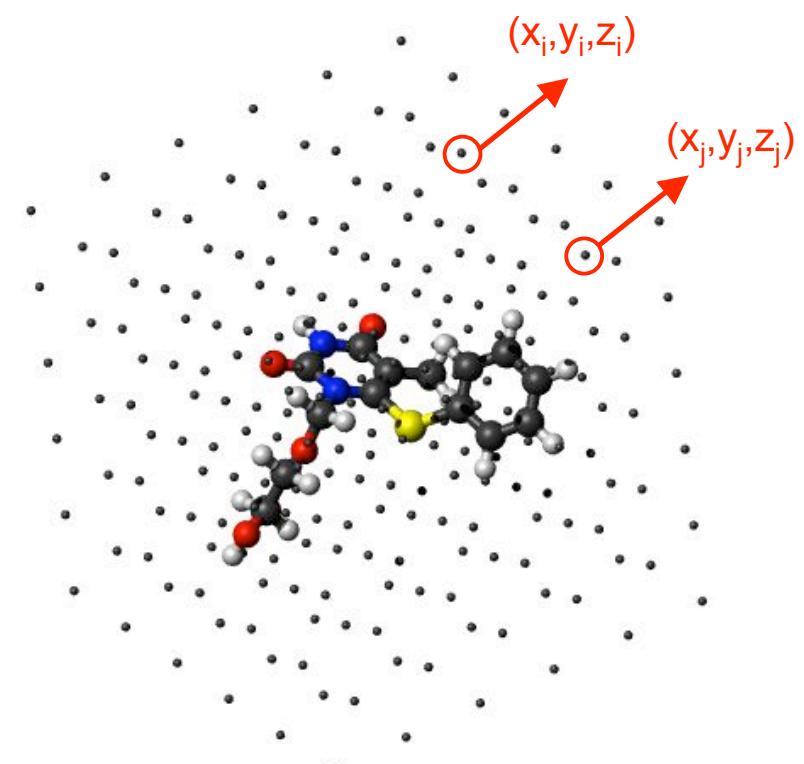
Example : Comparative molecular field analysis (CoMFA)

R.D. Cramer *et al.*, JACS, 1988, 110, 5959

	(x_1, y_1, z_1)	(x_2, y_2, z_2)		(x_1, y_1, z_1)	(x_2, y_2, z_2)	
	↓	↓	↓	↓	↓	
	S	S	.	E	E	.
1	1	2	.	1	2	.
2						b i n d
3						
.						
n						

Steric field Electrostatic field

Ligands mutually aligned
Common 3D lattice



PLS

$$\Delta G_{\text{bind}} = k_0 + \sum \alpha_i S_i + \sum \beta_i E_i$$

3D - QSAR

Limitations

Needs the experimental activity of a series of ligands

Not limited to structurally related molecules

Alignment of the molecules in their bioactive conformation. Use of:

- known structure of a complex (QSAR?)
- conformationally rigid example in the dataset
- functional groups in agreement with a pharmacophore hypothesis

Others : CoMSIA, HASL, Compass, APEX-3D, YAK, ...

Knowledge based approaches. Regression based methods

Example : LUDI score

H.J. Bohm, *J. Comput.-Aided Mol. Des.*, **1994**, 8, 623

H.J. Bohm, *J. Comput.-Aided Mol. Des.*, **1998**, 12, 309

$$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{polar}} + \Delta G_{\text{apolar}} + \Delta G_{\text{solv}} + \Delta G_{\text{flex}}$$

**Trained using a 82 protein-ligand complexes dataset
with known experimental ΔG_{bind}**

Polar interactions

$$\Delta G_{\text{polar}} = \Delta G_{\text{hb}} \sum_{\text{hb}} f(\Delta R, \Delta \alpha) \times f(N_{\text{neighb}}) \times \text{fpcs}$$

$$+ \Delta G_{\text{ion}} \sum_{\text{ion}} f(\Delta R, \Delta \alpha) \times f(N_{\text{neighb}}) \times \text{fpcs}$$

$$+ \Delta G_{\text{esrep}} N_{\text{esrep}}$$

$$\Delta G_{\text{hb}} = -0.81, \Delta G_{\text{ion}} = -1.41 \text{ and } \Delta G_{\text{esrep}} = +0.10 \text{ kcal/mol}$$

Apolar interactions

$$\Delta G_{\text{apolar}} = \Delta G_{\text{lipo}} A_{\text{lipo}} + \Delta G_{\text{aro}} \sum_{\text{aro}} f(R)$$

$$\Delta G_{\text{lipo}} = -0.81 \text{ and } \Delta G_{\text{aro}} = -0.62 \text{ kcal/mol}$$

Desolvation effect

Active site filled with water molecules



Unbound water molecules

$$\Delta G_{\text{solv}} = \Delta G_{\text{lipo wat}} \sum \text{unbound water}$$

$$\Delta G_{\text{lipo water}} = -0.33 \text{ kcal/mol}$$

Ligand flexibility

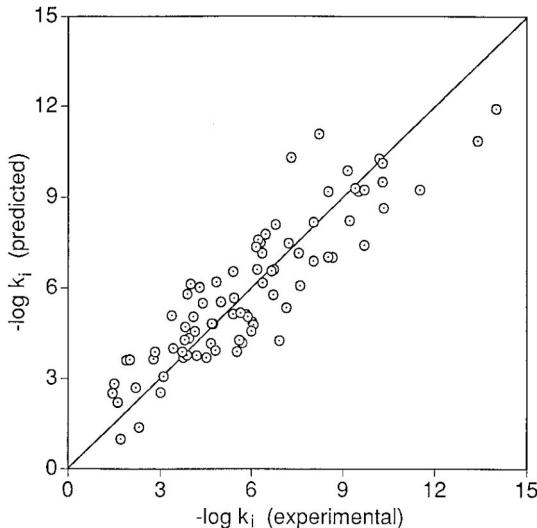
$$\Delta G_{\text{flex}} = \Delta G_{\text{rot}} N_{\text{rot}}$$

$$\Delta G_{\text{rot}} = +0.26 \text{ kcal/mol}$$

Nrot : number of rotatable bonds
(cyclic en3 en3 en3 en3 en2)

Knowledge based approaches. Regression based methods

Example : LUDI score



82 complexes of the training set

SD ~2 kcal/mol

Advantages :

- Allows identification of high affinity ligands
- Rapid estimation of the affinities
- Structurally different ligands
- “Universal” (different proteins)

Drawbacks :

- Somewhat large errors
- Method biased:
 - certain type of proteins
 - only good complementarity protein/ligand
- Some interactions ignored (cation – π)

Knowledge based approaches. Potential of mean force

Example : PMF score

I. Muegge et al., *J. Med. Chem.*, **1999**, 42, 791

I. Muegge et al., *Persp. In Drug Disc. And Des.*, **2000**, 20, 99

**Trained using 697 complexes from the PDB.
No need for experimental ΔG_{bind}**

$$A_{ij}(r) = -k_b T \ln \left[f_{\text{vol_corr}}^j(r) \frac{\rho^{ij}(r)}{\rho_{\text{bulk}}^{ij}} \right]$$

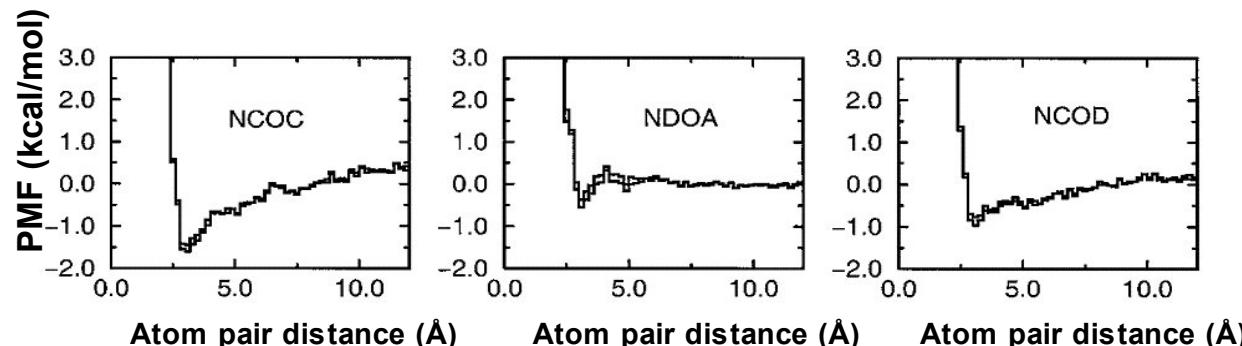
atom type **i** for protein and **j** for ligand

$\rho^{ij}(r)$: number density of atom pairs of type ij at distance r

$\rho_{\text{bulk}}^{ij}(r)$: number density of atom pairs of type ij in a sphere with radius R

$f_{\text{volv_corr}}^j$: ligand volume correction

16 protein atom types, 34 ligand atom types



NC positively charged nitrogen

OC negatively charged oxygen

ND nitrogen as hydrogen bond donor

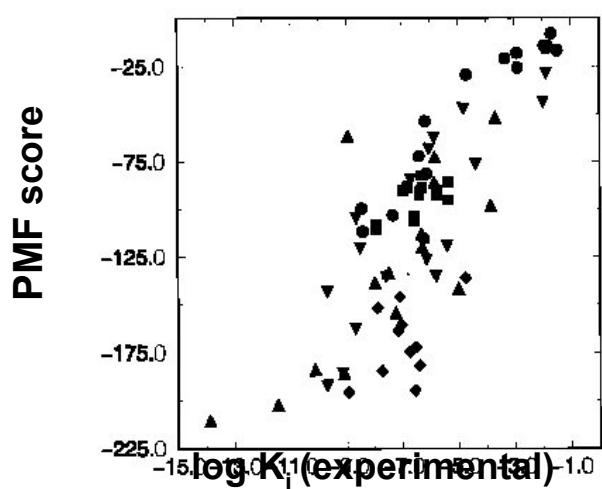
OA oxygen as hydrogen bond acceptor

OD oxygen as hydrogen bond donor

$$\text{PMF score} = \sum_{ii} \sum_{kl \text{ of type } ii} A_{ij}(r)$$

Knowledge based approaches. Potential of mean force

Example : PMF score



77 complexes, 5 different proteins
SD ~2 kcal/mol

Advantages :

- Allows identification of high affinity ligands
- Rapid estimation of the affinities
- Structurally different ligands
- “Universal” (different ligand and protein types)
- No fitting parameters to measured ΔG_{bind}

Drawbacks :

- Somewhat large errors
- No measure of directionality of H-bonds

Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle$$

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

Quantitative Structure Activity Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i$$

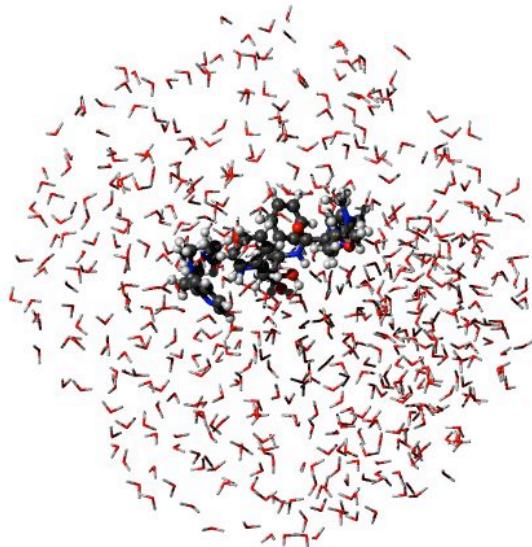
$$\Delta G = F(\mathbf{X}) \quad (\mathbf{X} \text{ is a descriptor})$$

CPU Time

Linear interaction energy (LIE)

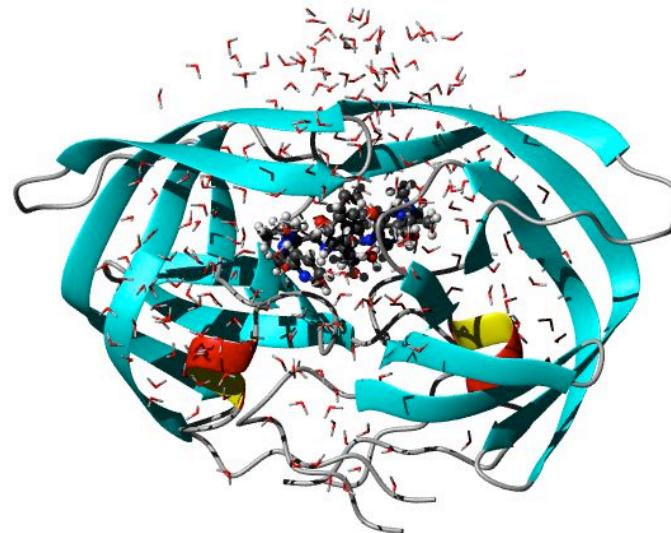
J. Åqvist, *J. Phys. Chem.*, **1994**, 98, 8253

Two MD runs : free state and bound state



Free state

“solvent” = water



Bound state

“solvent” = water + protein

$$\Delta G_{\text{bind}} = \alpha \left(\langle E_{l-s}^{\text{vdw}} \rangle_{\text{bound}} - \langle E_{l-s}^{\text{vdw}} \rangle_{\text{free}} \right) + \beta \left(\langle E_{l-s}^{\text{elec}} \rangle_{\text{bound}} - \langle E_{l-s}^{\text{elec}} \rangle_{\text{free}} \right)$$

J. Åqvist, *J. Phys. Chem.*, **1994**, 98, 8253

$\alpha=0.165$ and $\beta=0.5$

T. Hansson *et al.*, *J. Comp.-Aided Molec. Design*, **1998**, 12, 27

$\alpha=0.181$ and $\beta=0.5, 0.43, 0.37, 0.33$

W. Wang, *Proteins*, **1999**, 34, 395

Linear interaction energy (LIE)

Advantages :

- Faster than free energy simulation
- More structurally different ligands than for free energy simulation.
But generally restricted to rather similar ligands.

Drawbacks :

- Slower than scores based on a single conformation (LUDI, PMF, ...)
- Not really universal (α and β system dependent)
- Need experimental binding affinities of known complexes

Modifications :

- Additional term proportional to buried surface upon complexation

D.K. Jones-Hertzog and W.L. Jorgensen, *J. Med. Chem.*, **1997**, *40*, 1539

- Use of continuum solvent model instead of explicit solvent

R. Zhou and W.L. Jorgensen *et al.*, *J. Phys. Chem.*, **2001**, *105*, 10388

New LIE-like method

V. Zoete, O. Michielin and M. Karplus, *J. Comput.-Aided Molec. Design*, **2004**, in press

$$\Delta G_{\text{bind}} = \beta \langle E_{\text{prot-lig}}^{\text{elec}} \rangle + \gamma \langle \Delta G_{\text{desolv}} \rangle + \sigma \langle \text{SASA}_{\text{buried}} \rangle$$

MD simulation of the complex without explicit solvent

$E_{\text{prot-lig}}^{\text{elec}}$

Electrostatic interactions between protein and ligand

ΔG_{desolv}

Electrostatic contribution to desolvation energy (continuum model)

$\text{SASA}_{\text{buried}}$

Solvent accessible surface of protein and ligand
buried upon complexation

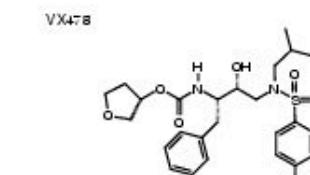
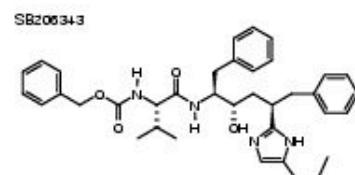
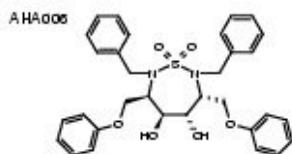
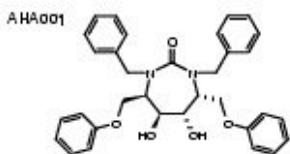
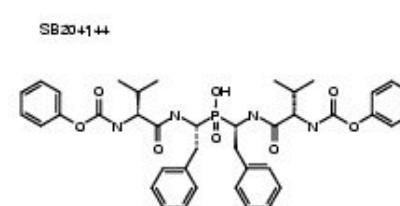
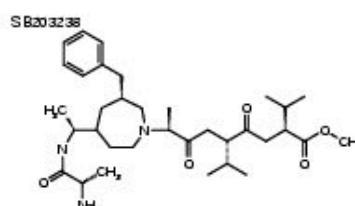
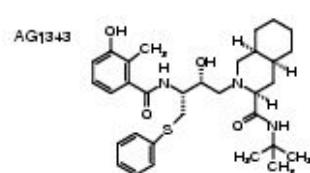
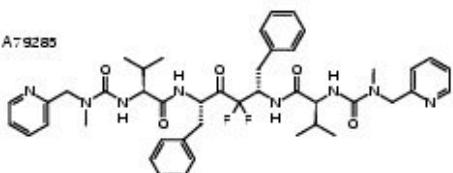
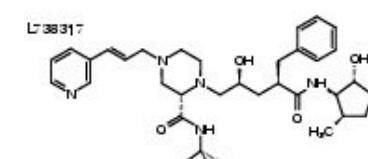
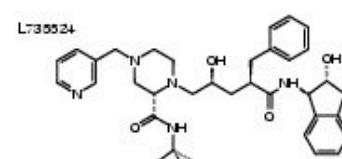
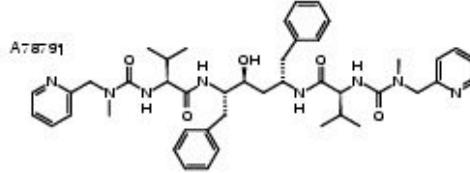
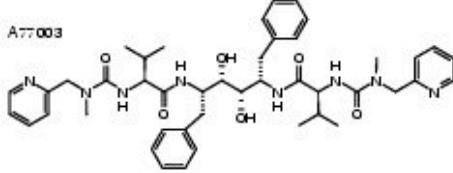
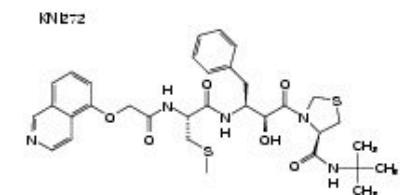
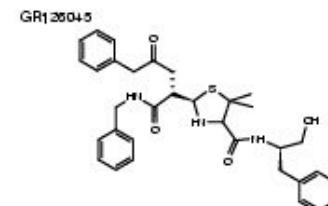
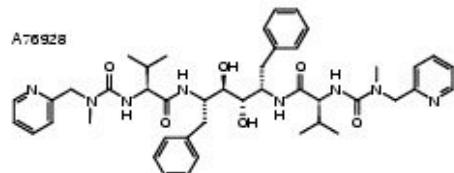
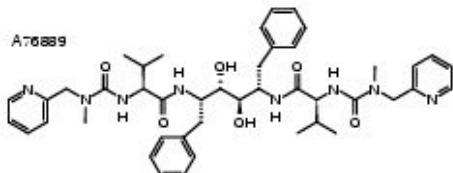
β , γ and σ fitted using a training set

New LIE-like method

Trained with a set of 16 known inhibitors of the HIV-1 protease belonging to different chemical families

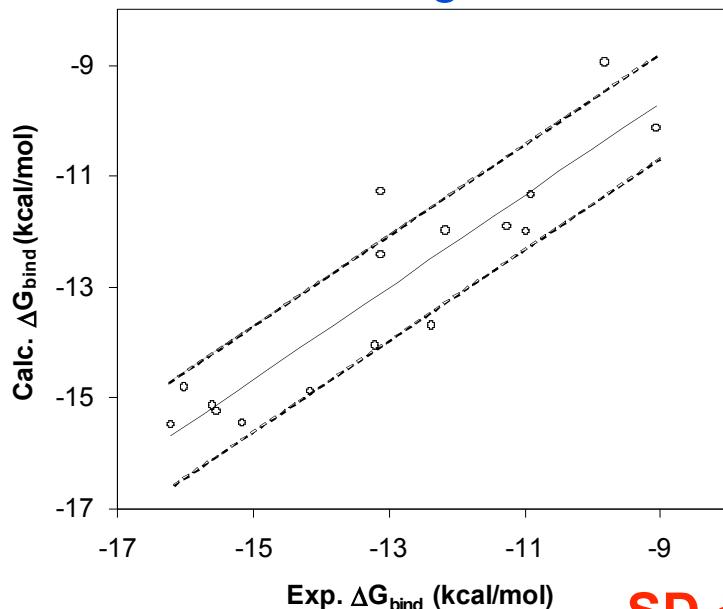
Used to rank new HIV-1 protease ligands

Training set :

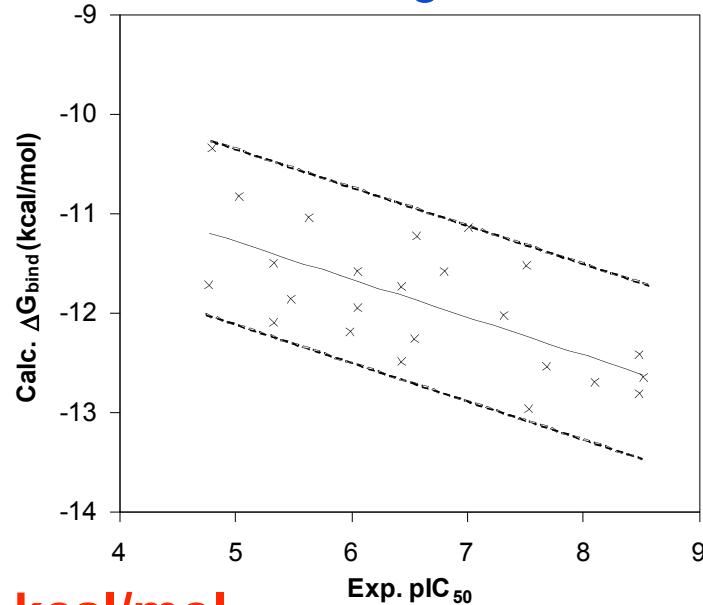


New LIE-like method

Training set



New ligands



SD ~ 0.85 kcal/mol

Advantages :

- Faster than free energy simulation LIE
- Ligands with very different chemical structures

Drawbacks :

- Slower than scores based on a single conformation
- Not universal
- Need experimental binding affinities of known complexes
- Restricted to ligands uncharged when bound

Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle$$

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

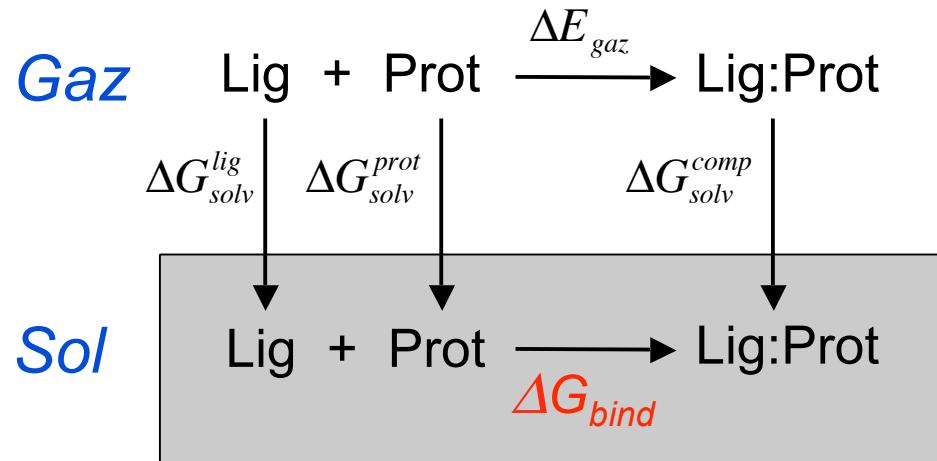
Quantitative Structure Activity Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i$$

$$\Delta G = F(\mathbf{X}) \quad (\mathbf{X} \text{ is a descriptor})$$

CPU Time

Binding free energy decomposition: MM-PBSA, MM-GBSA



Averaged over an MD simulation trajectory
of the complex (and isolated parts)

$$\Delta G_{bind} = \langle \Delta E_{gaz} \rangle + \langle \Delta G_{desolv} \rangle - T \langle \Delta S \rangle$$

$$E_{gaz} = E_{elec} + E_{vdw} + \Delta E_{int\ ra}$$

$$\Delta G_{desolv} = \Delta G_{solv}^{comp} - (\Delta G_{solv}^{lig} + \Delta G_{solv}^{prot})$$

$$-T\Delta S = -T(S^{comp} - (S^{prot} + S^{lig}))$$

$$S = S_{trans} + S_{rot} + S_{vib}$$

B. Tidor and M. Karplus, *J. Mol. Biol.*, **1994**, 238, 405

$$\Delta G_{solv} = \Delta G_{solv,elec} + \Delta G_{solv,np}$$

$$\Delta G_{desolv} = \Delta G_{solv,elec}^{comp} - (\Delta G_{solv,elec}^{lig} + \Delta G_{solv,elec}^{prot}) + \sigma (SASA^{comp} - (SASA^{lig} + SASA^{prot}))$$

Depending on the way $\Delta G_{solv,elec}$ is calculated:

Molecular mechanics – Poisson-Boltzmann Surface Area (MM- PBSA)

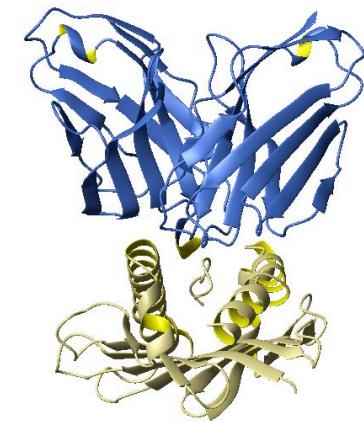
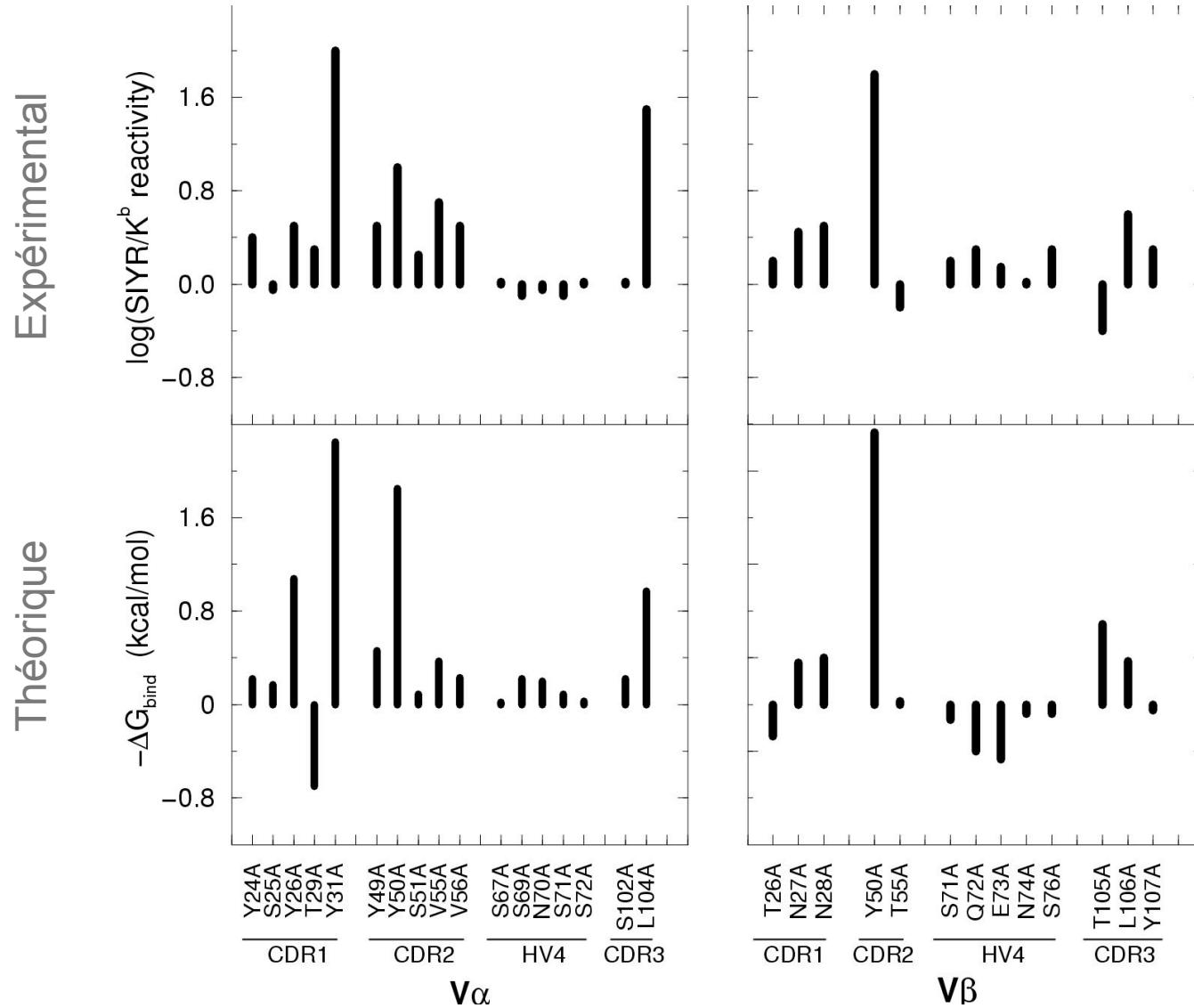
J. Srinivasan, P.A. Kollmann *et al.*, *J. Am. Chem. Soc.*, **1998**, 120, 9401

Molecular mechanics – Generalized Born Surface Area (MM- GBSA)

H. Gohlke, C. Kiel and D.A. Case, *J. Mol. Biol.*, **2003**, 330, 891

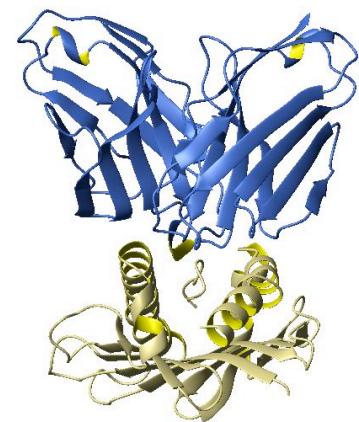
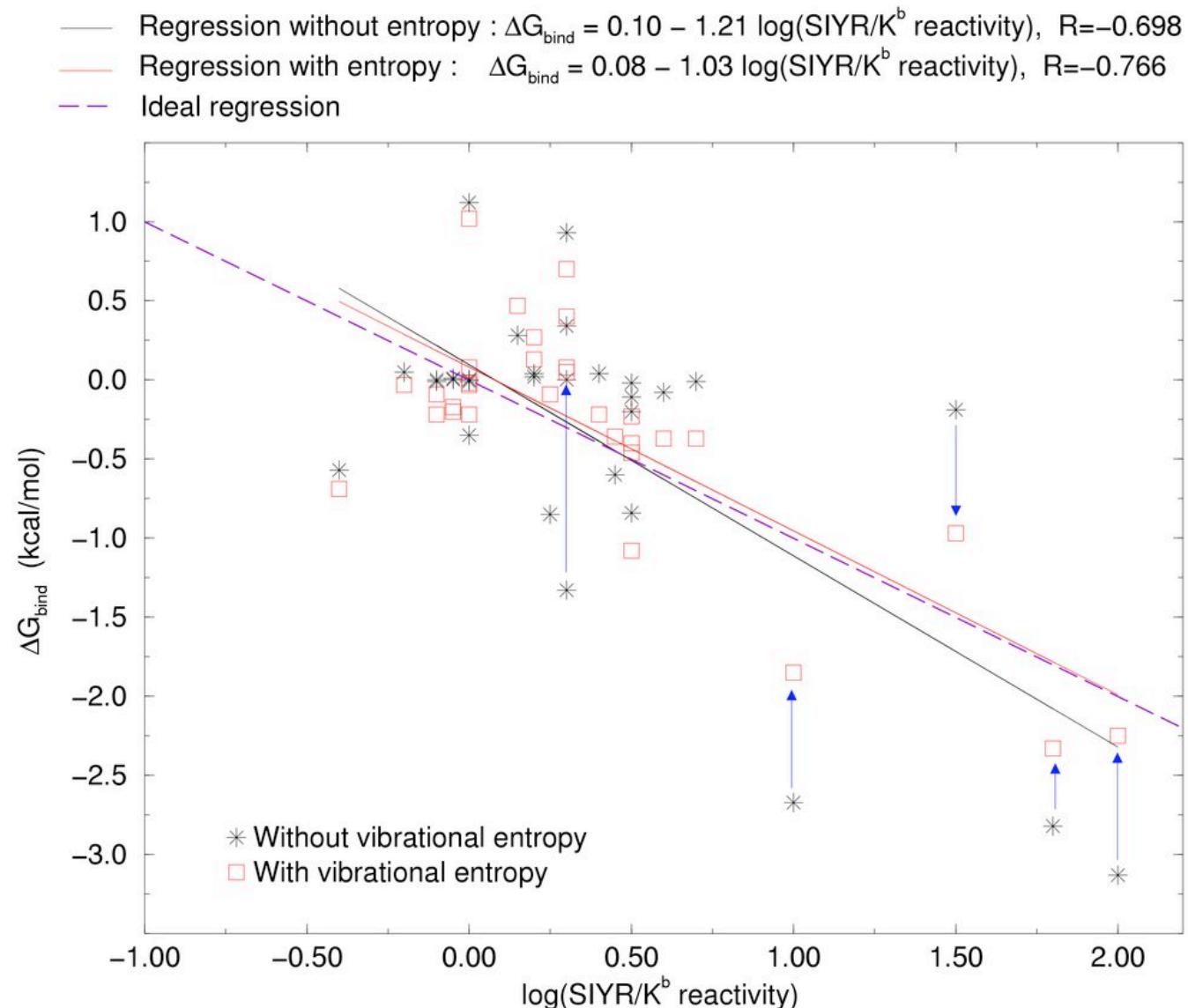
MM-GBSA Method: application to TCR-p-MHC

$$\Delta G_{bind} = \langle \Delta E_{gaz} \rangle + \langle \Delta G_{desolv} \rangle - T \langle \Delta S \rangle$$



MM-GBSA Method: application to TCR-p-MHC

$$\Delta G_{bind} = \langle \Delta E_{gaz} \rangle + \langle \Delta G_{desolv} \rangle - T \langle \Delta S \rangle$$



Examples of TCR optimization: 2C TCR

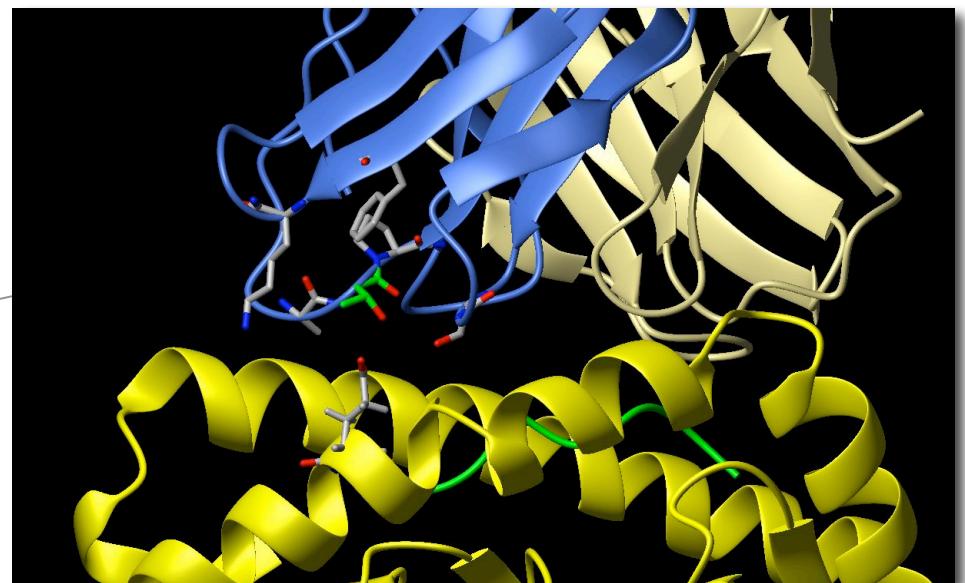
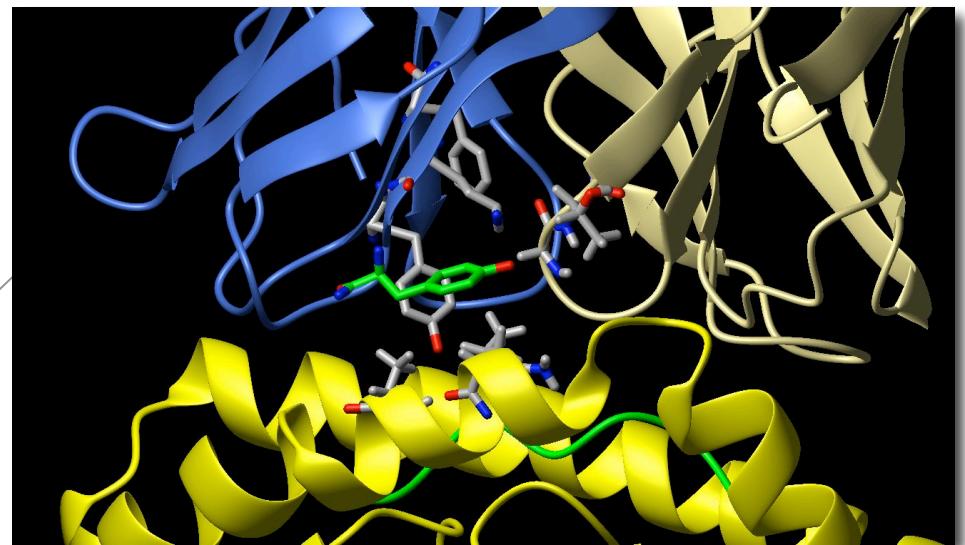
Residue	Domain	$\langle E_{vdW}^{sc} \rangle$	$\langle E_{elec}^{sc} \rangle$	$\langle \Delta G_{elec,solv}^{sc} \rangle$	$\langle \Delta G_{np,solv}^{sc} \rangle$	$-\langle TS_{vib}^{sc} \rangle$	$\langle \Delta G_{bind}^{sc} \rangle$
Ser93	CDR3	0.63	-9.06	3.04	-0.01	0.17	-5.23
Phe100	CDR3	-3.59	-0.74	1.36	-0.72	0.65	-3.04
Tyr31	CDR1	-3.46	-2.37	3.31	-0.61	0.75	-2.38
Tyr50	CDR2	-3.70	-4.02	5.63	-0.57	0.58	-2.08
Lys68	HV4	0.87	-56.18	53.34	-0.34	0.59	-1.72
Ser27	CDR1	-0.52	-5.14	4.32	-0.32	0.06	-1.60
Lys48	CDR2	0.63	-65.57	62.44	-0.26	1.25	-1.51
Tyr26	CDR1	-1.06	-1.46	1.79	-0.11	-0.11	-0.95
Ala28	CDR1	-0.73	0.49	-0.37	-0.33	0.13	-0.81
Ala101	CDR3	-0.40	-0.38	0.24	-0.08	0.02	-0.60
Leu104	CDR3	-0.10	-0.15	0.06	-0.00	-0.33	-0.52
Ser51	CDR2	-0.57	-1.58	1.77	-0.46	0.47	-0.37
Gln1	-	-0.16	-0.44	0.28	-0.00	0.00	-0.32
Ala103	CDR3	-0.04	-0.17	0.11	0.00	-0.19	-0.29
Pro30	CDR1	-0.10	-3.20	3.14	0.00	-0.12	-0.28
Phe66	-	-0.08	-0.10	0.18	0.00	-0.26	-0.26
Tyr49	CDR2	-0.21	0.15	-0.05	-0.00	-0.13	-0.24
Ser102	CDR3	-0.96	-4.31	5.08	-0.17	0.13	-0.23
...							
Thr29	CDR1	-0.37	-1.55	3.03	-0.18	-0.12	0.81
Asp53	CDR2	-0.10	28.56	-27.32	-0.01	0.02	1.15

Improvement of favorable TCR residues

Replacement of unfavorable TCR residues

Examples of TCR optimization: 2C TCR

Residue	Domain	$\langle E_{vdW}^{sc} \rangle$	$\langle E_{elec}^{sc} \rangle$	$\langle \Delta G_{elec,solv}^{sc} \rangle$	$\langle \Delta G_{np,solv}^{sc} \rangle$	$-\langle TS_{vib}^{sc} \rangle$	$\langle \Delta G_{bind}^{sc} \rangle$
Ser93	CDR3	0.63	-9.06	3.04	-0.01	0.17	-5.23
Phe100	CDR3	-3.59	-0.74	1.36	-0.72	0.65	-3.04
Tyr31	CDR1	-3.46	-2.37	3.31	-0.61	0.75	-2.38
Tyr50	CDR2	-3.70	-4.02	5.63	-0.57	0.58	-2.08
Lys68	HV4	0.87	-56.18	53.34	-0.34	0.59	-1.72
Ser27	CDR1	-0.52	-5.14	4.32	-0.32	0.06	-1.60
Lys48	CDR2	0.63	-65.57	62.44	-0.26	1.25	-1.51
Tyr26	CDR1	-1.06	-1.46	1.79	-0.11	-0.11	-0.95
Ala28	CDR1	-0.73	0.49	-0.37	-0.33	0.13	-0.81
Ala101	CDR3	-0.40	-0.38	0.24	-0.08	0.02	-0.60
Leu104	CDR3	-0.10	-0.15	0.06	-0.00	-0.33	-0.52
Ser51	CDR2	-0.57	-1.58	1.77	-0.46	0.47	-0.37
Gln1	-	-0.16	-0.44	0.28	-0.00	0.00	-0.32
Ala103	CDR3	-0.04	-0.17	0.11	0.00	-0.19	-0.29
Pro30	CDR1	-0.10	-3.20	3.14	0.00	-0.12	-0.28
Phe66	-	-0.08	-0.10	0.18	0.00	-0.26	-0.26
Tyr49	CDR2	-0.21	0.15	-0.05	-0.00	-0.13	-0.24
Ser102	CDR3	-0.96	-4.31	5.08	-0.17	0.13	-0.23
...							
Thr29	CDR1	-0.37	-1.55	3.03	-0.18	-0.12	0.81
Asp53	CDR2	-0.10	28.56	-27.32	-0.01	0.02	1.15



Binding free energy decomposition

MM- PBSA, MM-GBSA

Advantages :

- Used for ligand:protein and protein:protein complexes
- Could be applied to structurally different ligands (but in fact applied to similar ones)
- “Universal” (no parameter to be fitted)
- MM-GBSA allows a per-atom decomposition of ΔG_{bind} (e.g. contribution of side chains)

Drawbacks :

- Rather time consuming
- In some cases, found unable to rank ligands

- $T\Delta S$ is necessary to find the order of magnitude of the absolute binding free energies but, in some cases, it is not necessary to estimate relative binding free energies

W. Wang and P.A. Kollman, *J. Mol. Biol.*, **2000**, 303, 567

H. Gohlke, C. Kiel and D.A. Case, *J. Mol. Biol.*, **2003**, 330, 891

Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle \quad \Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

Quantitative Structure Activity Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i$$

$$\Delta G = F(\mathbf{X}) \quad (\mathbf{X} \text{ is a descriptor})$$

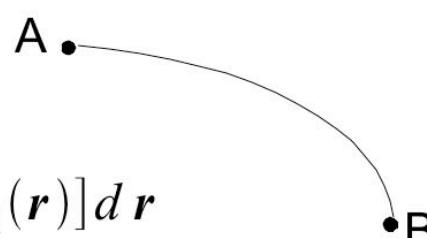
CPU Time

Free energy calculation: FEP formalism

Statistical Mechanics Definition of G:

$$G = -k_B T \ln Z = -k_B T \ln \int \exp[-\beta H(\mathbf{r})] d\mathbf{r}$$

Free energy difference between two states A and B:

$$\Delta G = G_A - G_B = -k_B T \ln \frac{Z_A}{Z_B} = -k_B T \ln \frac{\int \exp[-\beta H_A(\mathbf{r})] d\mathbf{r}}{\int \exp[-\beta H_B(\mathbf{r}')] d\mathbf{r}'}$$


Multiplying by $(\exp[+bH_B] \exp[-bH_B])$ the numerator, one gets

$$\Delta G = -k_B T \ln \frac{\int \exp[-\beta H_A(\mathbf{r})] \exp[+\beta H_B(\mathbf{r})] \exp[-\beta H_B(\mathbf{r})] d\mathbf{r}}{\int \exp[-\beta H_B(\mathbf{r}')] d\mathbf{r}'}$$

$$\Delta G = -k_B T \ln \langle \exp[-\beta H_A] \exp[+\beta H_B] \rangle_B = -k_B T \ln \langle \exp[-\beta \Delta H] \rangle_B$$

Note that the ensemble average is performed on the system in state B:

- if $|H_A - H_B| \ll k_B T$, intermediate states should be introduced: H_1, H_2, \dots, H_n .

Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle \quad \Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

Quantitative Structure Activity Relationship (QSAR)

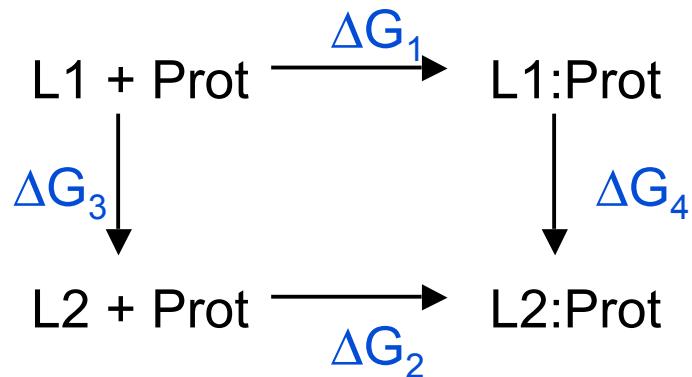
$$\Delta G = k_0 + \sum k_i X_i$$

$$\Delta G = F(\mathbf{X}) \quad (\mathbf{X} \text{ is a descriptor})$$

CPU Time



Use of thermodynamical cycles



Thermodynamic cycle perturbation approach:

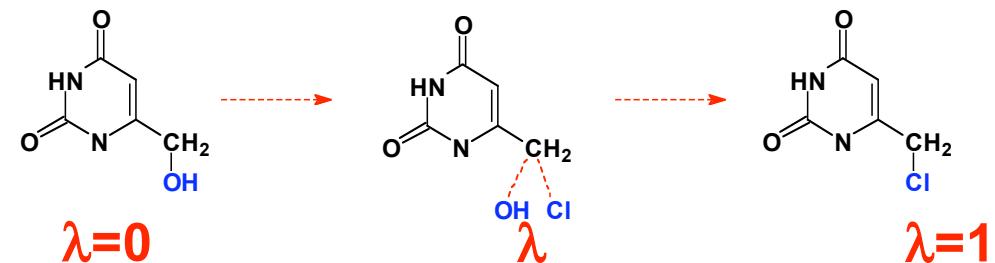
$$\Delta\Delta G_{\text{bind}} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

$\Delta G_4 - \Delta G_3$ is computationally accessible

“Alchemical” reaction. MD or MC at different λ .

Coupling parameter λ

$$H_\lambda = H_0 + \lambda H_{L_1} + (1 - \lambda) H_{L_2}$$



Free energy perturbation (FEP)

$$\Delta\Delta G_{\text{bind}} = -RT \sum_{i=0}^{n-1} \ln \left\langle \exp \left(- \left(H_{\lambda_i} - H_{\lambda_{i+1}} \right) / RT \right) \right\rangle_{\lambda_i}$$

Thermodynamic integration (TI)

$$\Delta\Delta G_{\text{bind}} = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Alchemical free energy formalism

Statistical mechanics definition of the free energy:

$$A(N, V, T) = -kT \ln Z(N, V, T) = -kT \ln \int e^{-\beta U(\mathbf{r}, \lambda)} d\mathbf{r} \quad (1)$$

with

$$U = U(\mathbf{r}, \lambda) = U_0(\mathbf{r}) + \lambda U_{Ala}(\mathbf{r}) + (1 - \lambda) U_{Pro}(\mathbf{r}) \quad (2)$$

The derivative of the free energy with respect to λ is

$$\frac{\partial A}{\partial \lambda}(\lambda) = \frac{\int \frac{\partial U(\mathbf{r}, \lambda)}{\partial \lambda} e^{-\beta U(\mathbf{r}, \lambda)} d\mathbf{r}}{\int e^{-\beta U(\mathbf{r}, \lambda)} d\mathbf{r}} = \langle \frac{\partial U(\mathbf{r}, \lambda)}{\partial \lambda} \rangle_\lambda \quad (3)$$

The total free energy difference between $\lambda = 0$ (Pro) and $\lambda = 1$ (Ala) is obtained by integrating Eq. 3,

$$\Delta A_{\lambda=0 \rightarrow 1} = \int_0^1 \langle \frac{\partial U(\mathbf{r}, \lambda)}{\partial \lambda} \rangle_\lambda d\lambda \quad (4)$$

In the case of a linear λ dependence (Eq. 2), the derivative of the potential energy with respect to λ is simply

$$\frac{\partial U(\mathbf{r}, \lambda)}{\partial \lambda} = U_{Ala}(\mathbf{r}) - U_{Pro}(\mathbf{r}) \quad (5)$$

and the change in free energy is given by

$$\Delta A_{\lambda=0 \rightarrow 1} = \int_0^1 \langle U_{Ala}(\mathbf{r}) - U_{Pro}(\mathbf{r}) \rangle_\lambda d\lambda \quad (6)$$

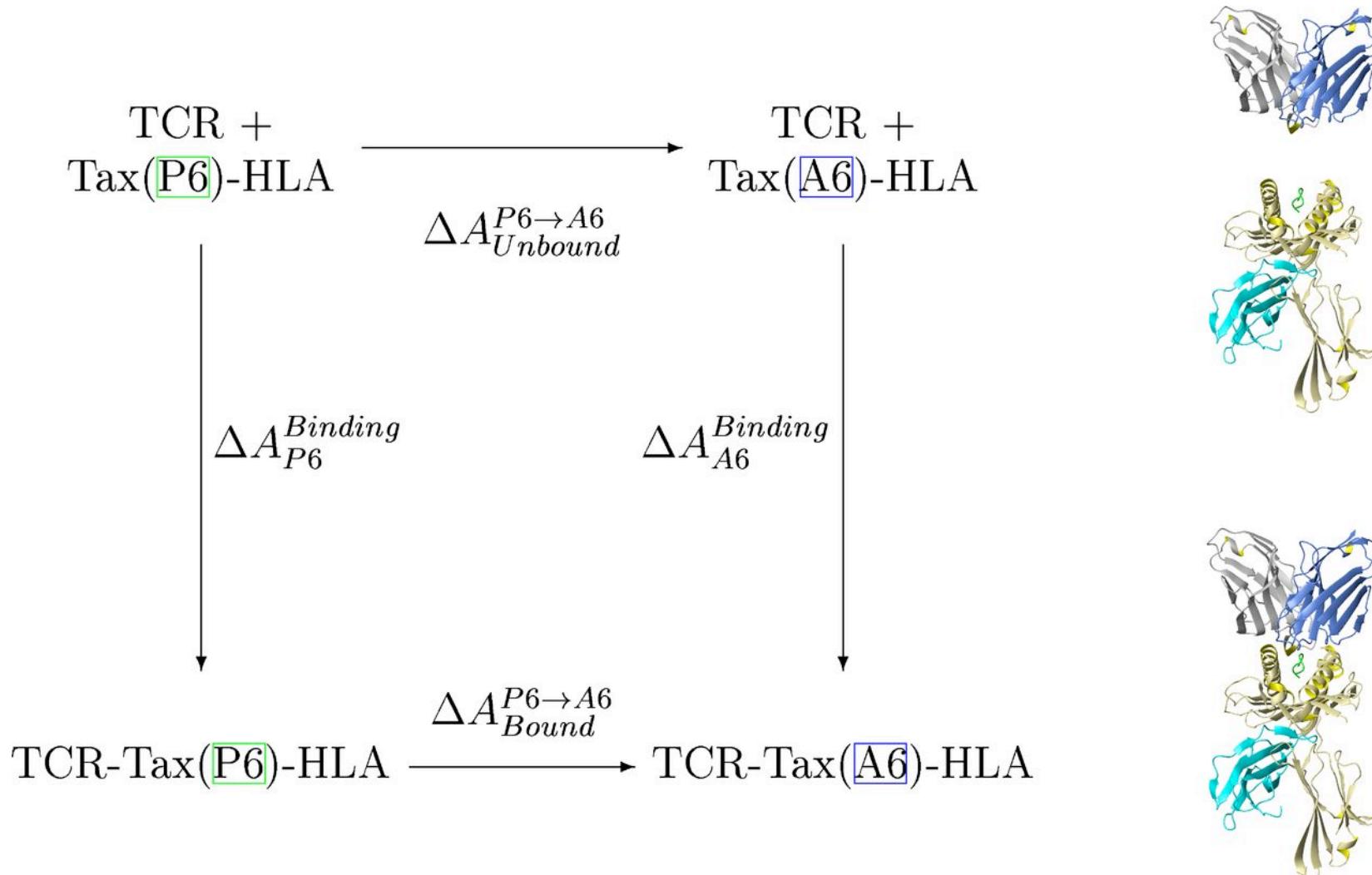
Free energy component analysis:

$$\Delta A_{\lambda=0 \rightarrow 1} = \sum_i \Delta A_{\lambda=0 \rightarrow 1}^{Group i} \quad (7)$$

with

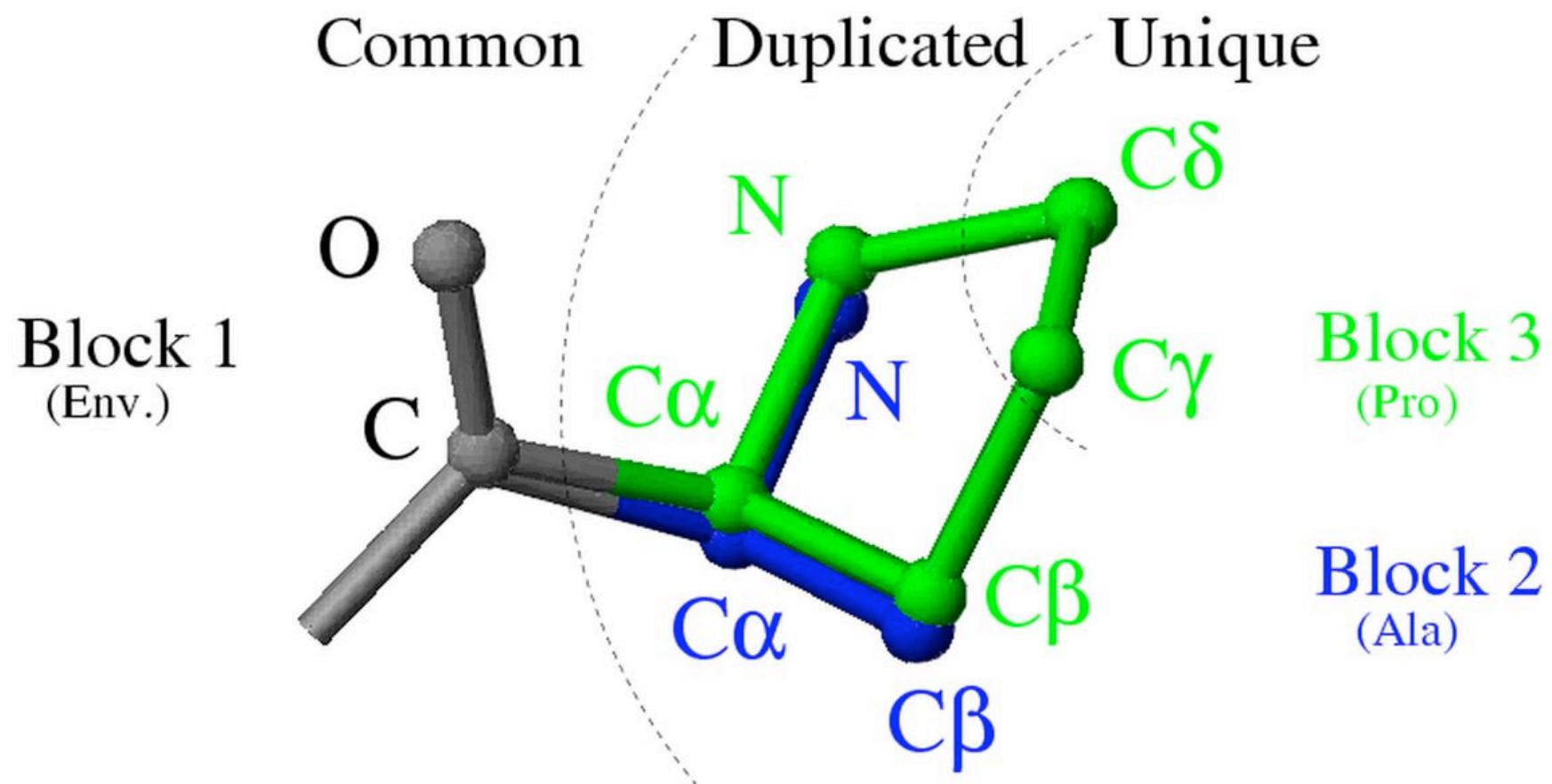
$$\Delta A_{\lambda=0 \rightarrow 1}^{Group i} = \int_0^1 \langle \frac{\partial U(\mathbf{r}, \lambda)^{Group i}}{\partial \lambda} \rangle_\lambda d\lambda = \int_0^1 \langle U(\mathbf{r})_{Ala}^{Group i} - U(\mathbf{r})_{Pro}^{Group i} \rangle_\lambda d\lambda \quad (8)$$

“Alchemical” Free Energy Calculations



$$\Delta\Delta A_{P6A}^{Binding} = \Delta A_{A6}^{Binding} - \Delta A_{P6}^{Binding} = \Delta A_{Bound}^{P6 \rightarrow A6} - \Delta A_{Unbound}^{P6 \rightarrow A6}$$

Hybrid Side Chain for P → A Mutation



$$U(\mathbf{r}, \lambda) = U_0(\mathbf{r}) + \lambda U_{Ala}(\mathbf{r}) + (1 - \lambda) U_{Pro}(\mathbf{r})$$

$$\longrightarrow \Delta A_{\lambda=0 \rightarrow 1} = \int_0^1 \langle U_{Ala}(\mathbf{r}) - U_{Pro}(\mathbf{r}) \rangle_\lambda d\lambda$$

Results of the Free Energy Simulations

$\Delta\Delta A$ Component Analysis⁽¹⁾

Group	Total ⁽²⁾	van der Waals ⁽²⁾	Electrostatic ⁽²⁾
Solvent	0.642(0.305)	1.381(0.336)	-0.736(0.047)
Peptide	0.531(1.001)	-0.341(1.184)	0.882(0.194)
HLA A2	1.264(0.250)	1.124(0.353)	0.142(0.107)
TCR	0.794(0.267)	0.640(0.163)	0.154(0.109)
CDR1 α	0.057(0.014)	0.015(0.002)	0.042(0.016)
α N30	0.029(0.027)	0.015(0.002)	0.014(0.027)
CDR2 α	0.000(0.000)	0.000(0.000)	0.000(0.000)
CDR3 α	0.581(0.090)	0.355(0.153)	0.226(0.080)
α T98	-0.048(0.012)	0.006(0.002)	-0.054(0.014)
α D99	0.418(0.047)	0.052(0.012)	0.366(0.040)
α S100	0.207(0.068)	0.241(0.131)	-0.034(0.078)
CDR1 β	0.036(0.030)	0.005(0.001)	0.030(0.031)
β E30	0.073(0.039)	0.005(0.001)	0.068(0.040)
β Y31	-0.037(0.010)	0.001(0.000)	-0.038(0.010)
CDR2 β	0.040(0.013)	0.000(0.000)	0.040(0.012)
β V50	0.041(0.013)	0.000(0.000)	0.040(0.013)
CDR3 β	0.023(0.285)	0.221(0.291)	-0.198(0.006)
β R95	-0.177(0.011)	0.004(0.001)	-0.181(0.011)
β L98	0.084(0.302)	0.122(0.301)	-0.037(0.006)
β A99	0.023(0.004)	0.012(0.000)	0.011(0.004)
β G100	0.020(0.007)	0.016(0.002)	0.003(0.006)
β G101	0.033(0.004)	0.019(0.003)	0.014(0.003)

⁽¹⁾Only components with an absolute value of more than 0.02 kcal/mol are given.

⁽²⁾All energies are in kcal/mol.

Total (Path Independent)

Experimental: 2.9 (0.2) kcal/mol
 Theoretical: 2.9 (1.1) kcal/mol

Components (Path Dependent)

Experimental: -
 Theoretical:

TCR 25% }
 Solvent 20% } 45%

HLA A2 40% }
 Peptide 15% } 55%

Free energy formalism

From the statistical definition of the free energy,

$$G = -kT \ln Z = -kT \ln \int e^{-\beta U(\mathbf{r}, \lambda)} d\mathbf{r}$$

and with a λ dependent potential, $U = U(\mathbf{r}, \lambda)$,
the derivative of G with respect to λ is simply

$$\frac{\partial G}{\partial \lambda}(\lambda) = \left\langle \frac{\partial U(\mathbf{r}, \lambda)}{\partial \lambda} \right\rangle_{\lambda}$$

And the total free energy change is obtained by integration

$$\Delta G_{\lambda=0 \rightarrow 1} = \int_0^1 \left\langle \frac{\partial U(\mathbf{r}, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

(Exact solution)

Linear λ dependence

Component analysis

$$U = U(\mathbf{r}, \lambda) = U_0(\mathbf{r}) + \lambda U_{Ala}(\mathbf{r}) + (1-\lambda) U_{Pro}(\mathbf{r})$$

$$\Delta G_{\lambda=0 \rightarrow 1} = \sum_i \Delta G_{\lambda=0 \rightarrow 1}^{Group i}$$

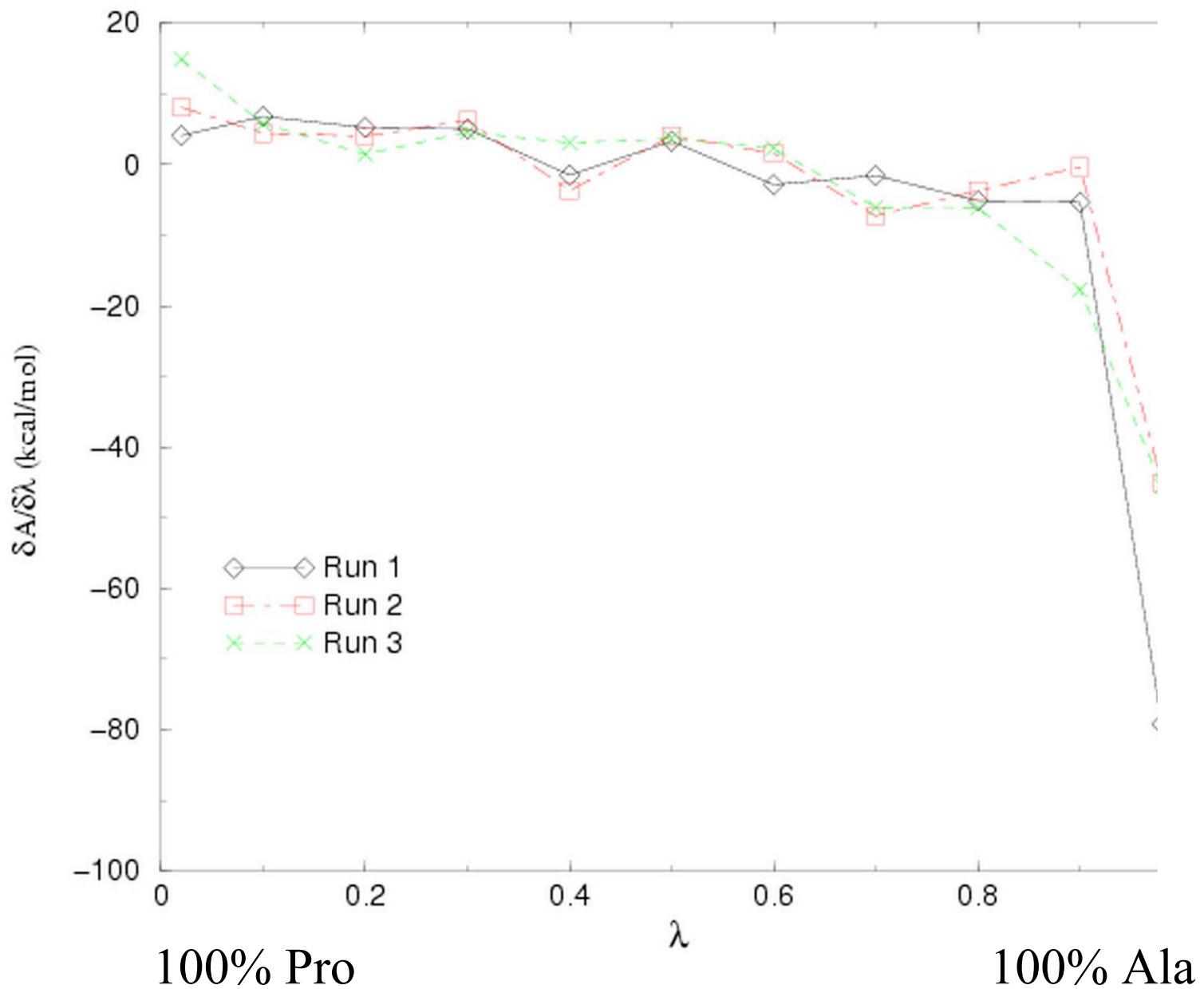
And the total free energy change is

$$\Delta G_{\lambda=0 \rightarrow 1} = \int_0^1 \langle U_{Ala}(\mathbf{r}) - U_{Pro}(\mathbf{r}) \rangle_{\lambda} d\lambda$$

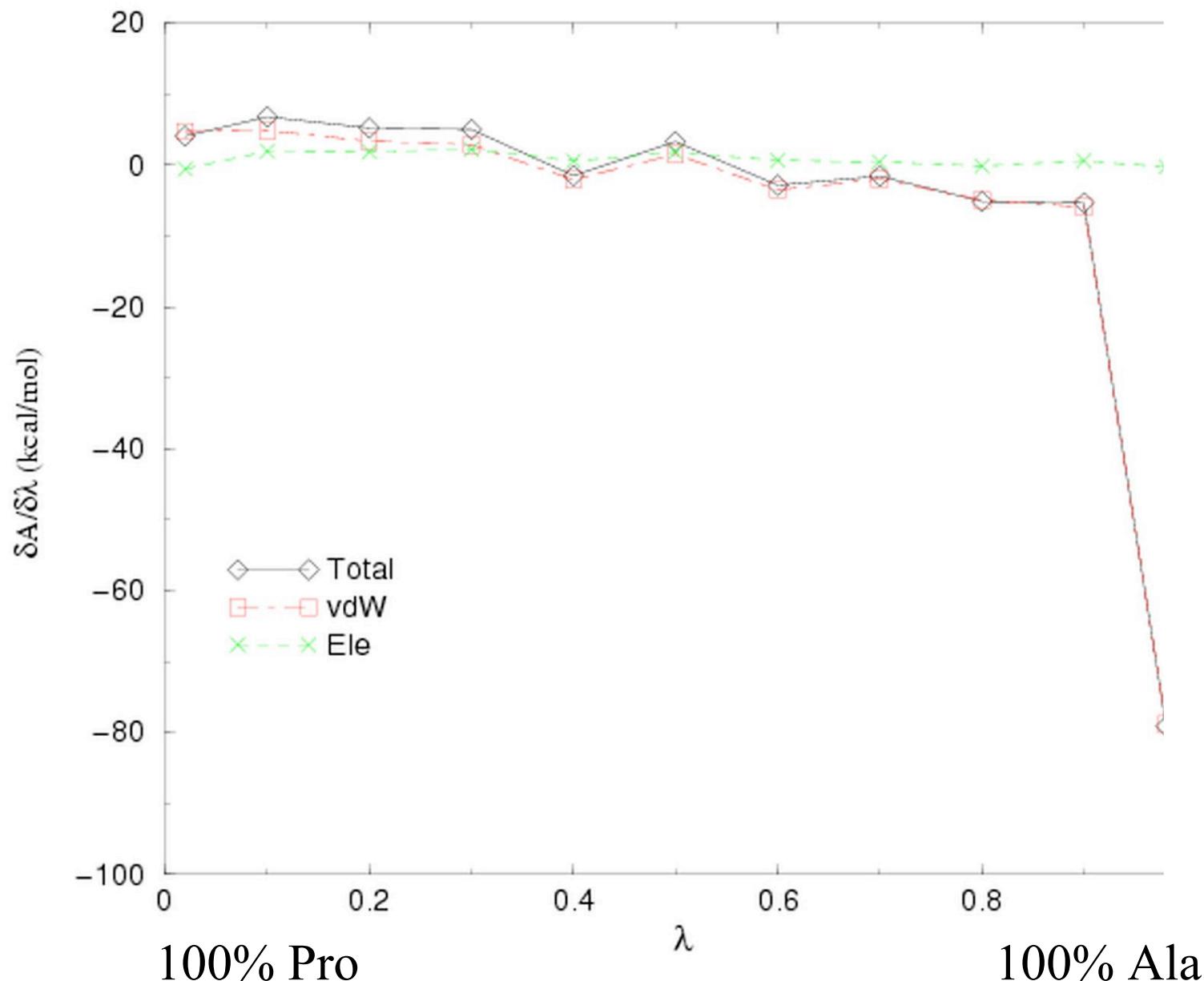
With

$$\Delta G_{\lambda=0 \rightarrow 1}^{Group i} = \int_0^1 \left\langle \frac{\partial U(\mathbf{r}, \lambda)^{Group i}}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

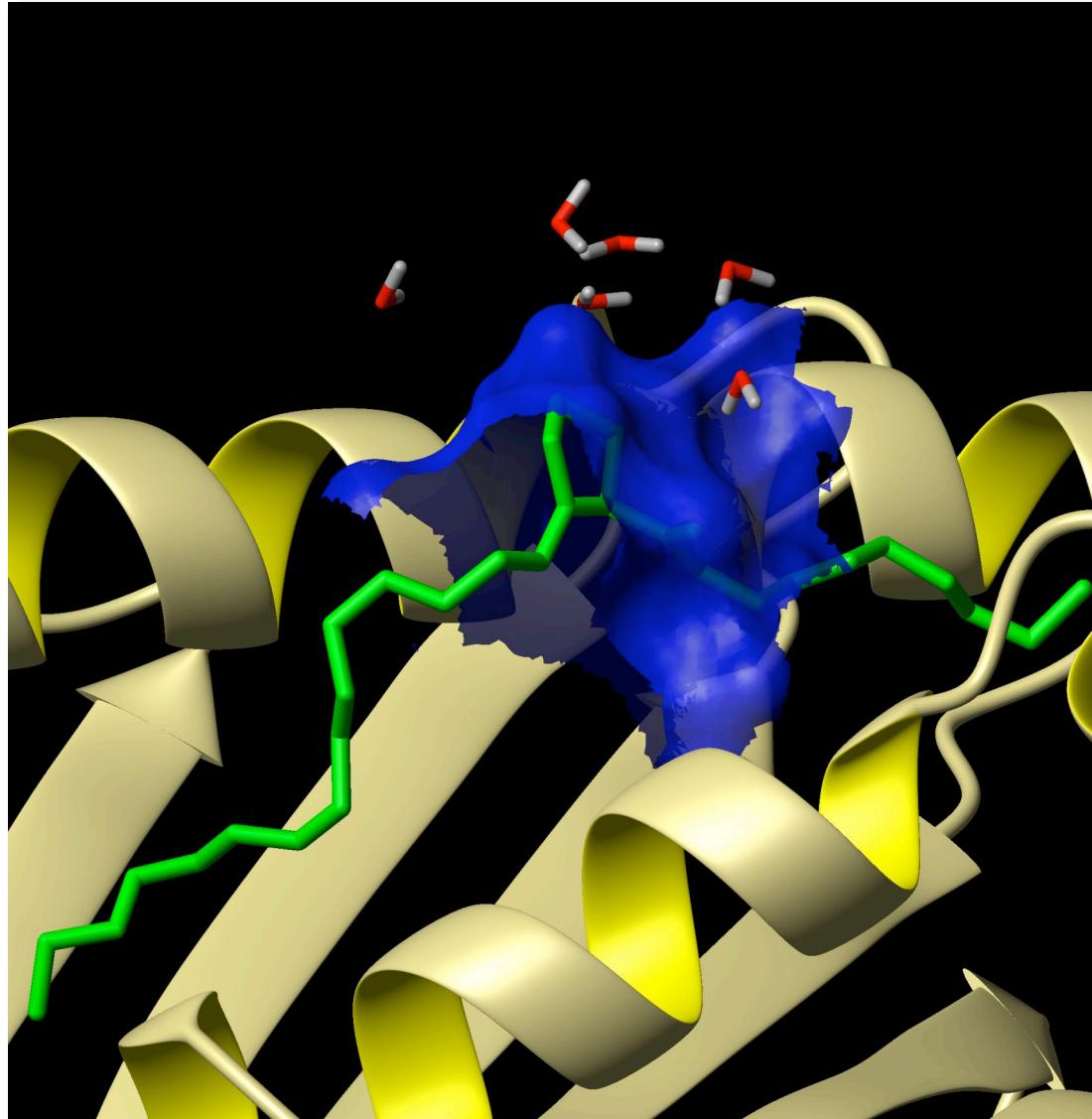
Free energy derivative



Free energy derivative components



Solvent Contribution

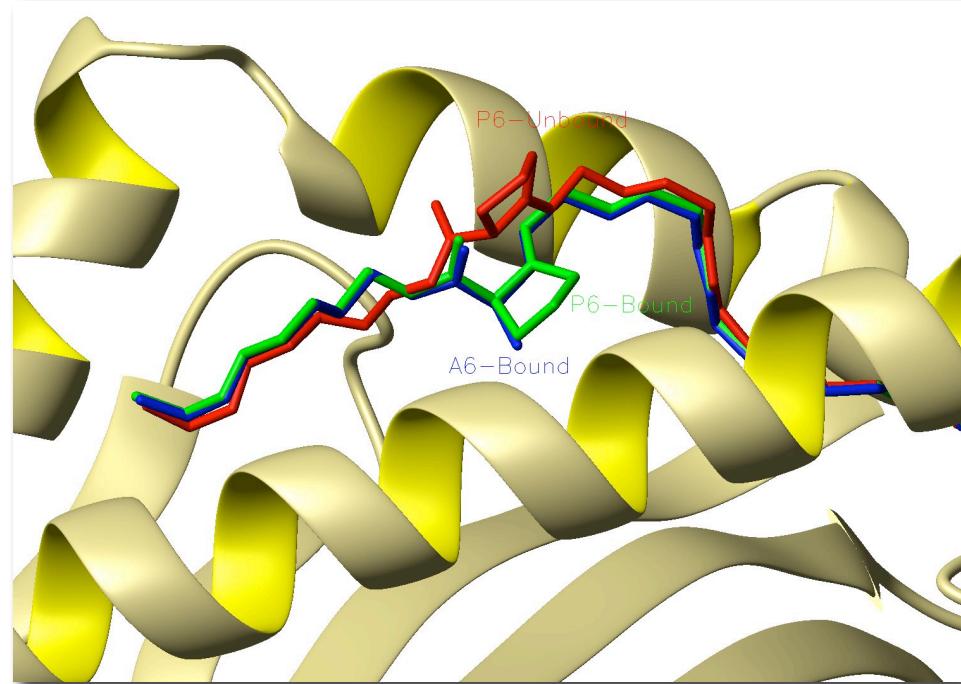
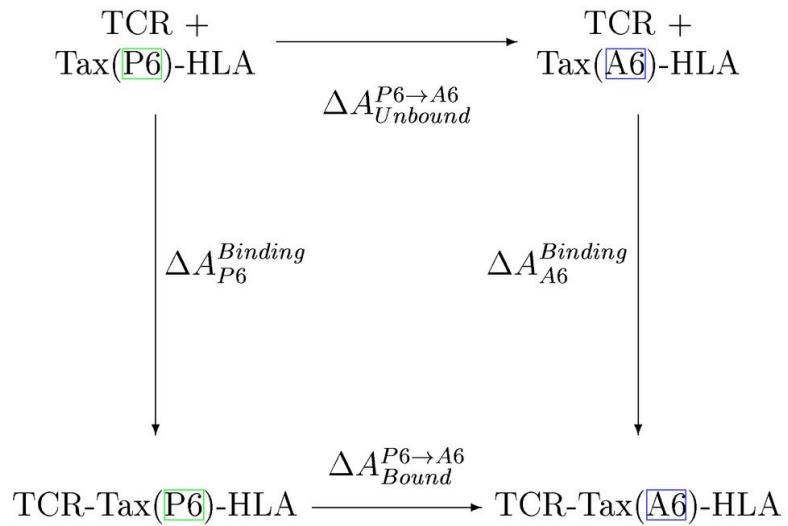


Solvent contribution to $\Delta\Delta A$:

- vdW :	+1.4 kcal/mol
- $Elec$:	-0.7 kcal/mol
<hr/>	
- <i>Total</i> :	+0.7 kcal/mol

The solvent favors association with Tax(P6)-A2 compared to the A6 mutant that is more soluble.

p-MHC conformational change contribution

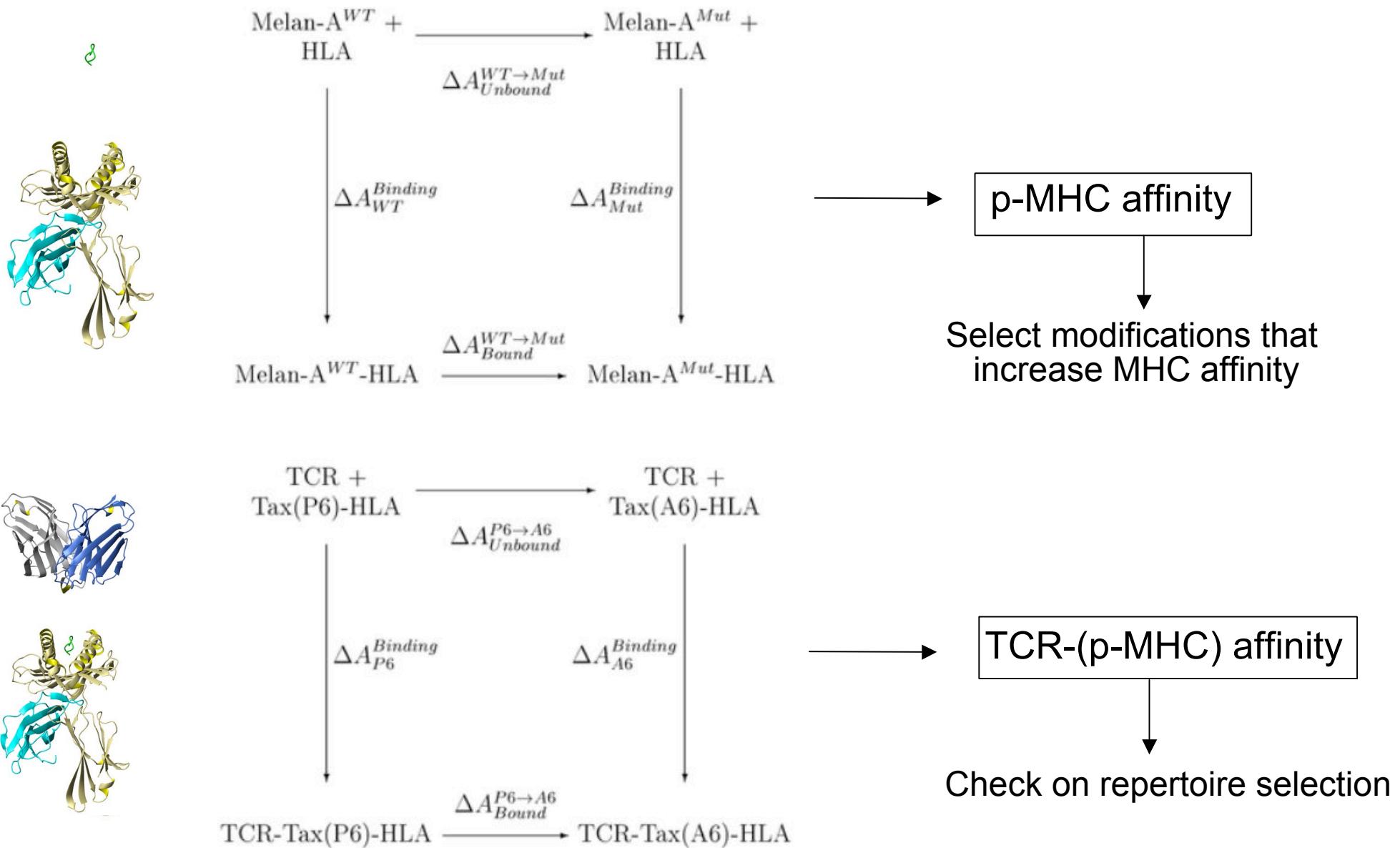


$$\Delta\Delta A_{P6A}^{Binding} = \Delta A_{A6}^{Binding} - \Delta A_{P6}^{Binding} = \Delta A_{Bound}^{P6 \rightarrow A6} - \Delta A_{Ubound}^{P6 \rightarrow A6}$$

p-MHC conformational change observed upon TCR binding is indirectly computed along the horizontal paths:

- the free energy cost of the P→A mutation is higher in the bound simulation because of the numerous favorable interaction of P6 with HLA A2. This brings a net addition of +1.2 kcal/mol to $\Delta\Delta A$.

Combined approach for peptide design



Concluding Remarks

- 1) Free energy simulations can reproduce accurately experimental changes in association constant between two closely related protein systems if detailed structural knowledge is available (X-ray, NMR or model)
- 2) The formalism is exact from a statistical physics stand point and accurate treatment of entropic terms, solvent effect or conformational changes can be obtained
- 3) Convergence of the free energy derivative is still problematic. The situation should improve with new methodological enhancements as well as longer simulation time
- 4) Absolute free energies can also be computed but the convergence is even more difficult
- 5) Much details about the specificity of the association can be gained using component analysis, opening the door to rational peptide or protein design

Free energy calculation: Main approaches

Sampling, Exact

Free Energy Perturbation (FEP)

$$\Delta G = -k_B T \ln \langle \exp(-\beta \Delta V) \rangle$$

Thermodynamical Integration (TI)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Non Equilibrium Statistical Mechanics (Jarzynski)

$$\Delta G = -k_B T \ln \langle \exp(-\beta W) \rangle$$

Sampling, Approx.

Linear Interaction Energy (LIE)

Molecular Mechanics/Poisson-Boltzmann/Surface area (MM-PBSA)

$$\Delta G = \alpha \Delta \langle V_{l \text{ env}}^{VdW} \rangle + \beta \Delta \langle V_{l \text{ env}}^{Elec} \rangle \quad \Delta G = \langle \Delta G_{Gas} \rangle + \langle \Delta G_{Desolv}^{PBSA} \rangle - \langle T \Delta S \rangle$$

Approx.

Quantitative Structure Activity Relationship (QSAR)

$$\Delta G = k_0 + \sum k_i X_i$$

$$\Delta G = F(\mathbf{X}) \quad (\mathbf{X} \text{ is a descriptor})$$

CPU Time



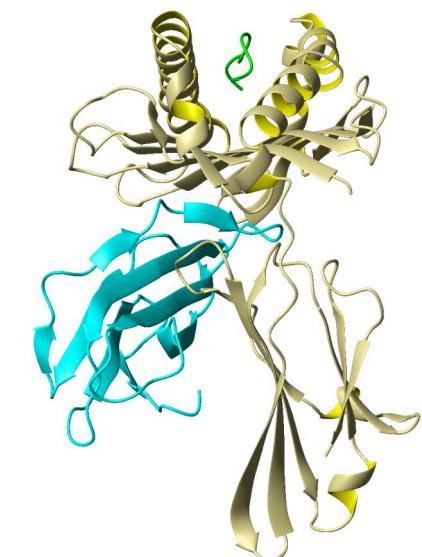
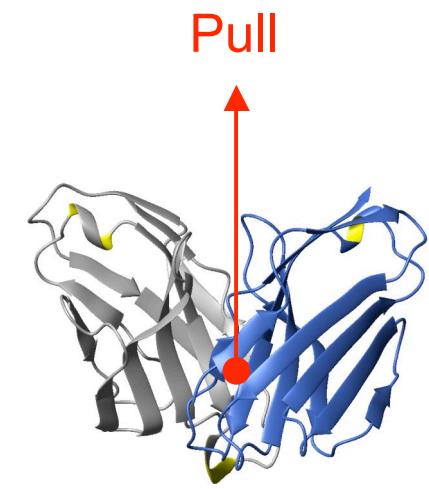
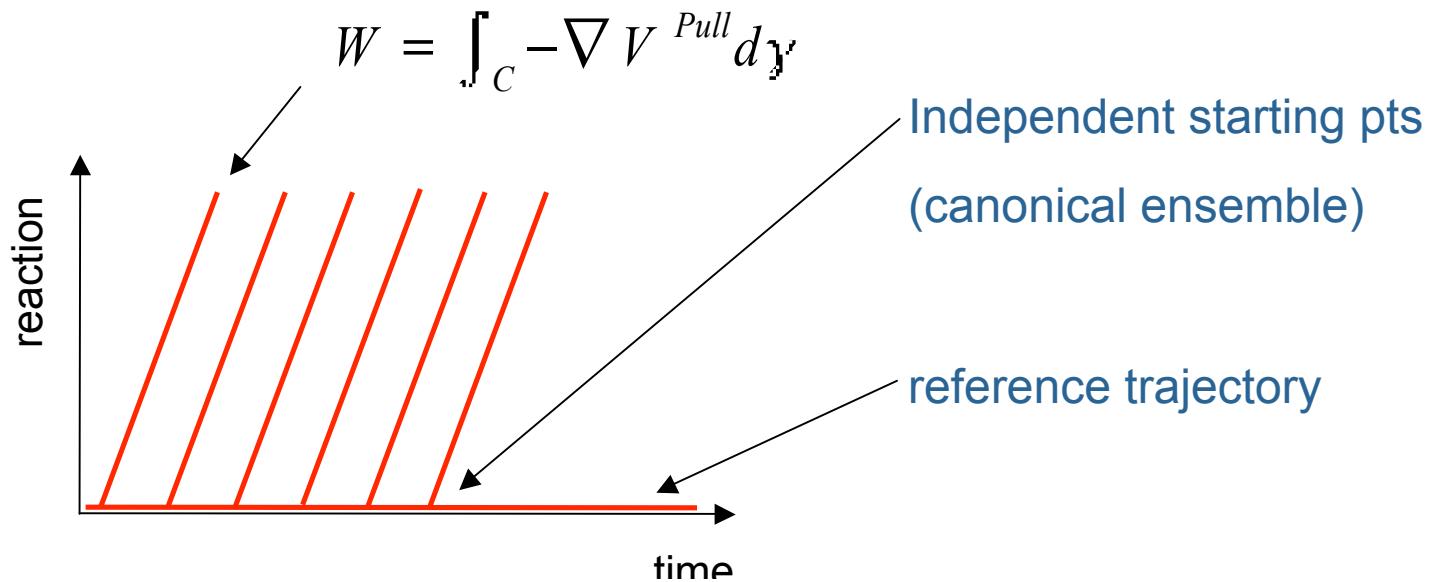
Computation of absolute TCR binding free energy

Let G be the free energy and W the work,

$$W_{\text{adia}} = \Delta G \longrightarrow K_A \quad (\text{Infinitely slow})$$

$$W = W_{\text{adia}} + W_{\text{diss}} \quad (\text{Finite rate})$$

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G} \quad (\text{Jarzynsky})$$



“Proof” of the Jarzynski equation: Park & al. 2004

Suppose a system S in equilibrium with a thermal bath B at temperature T , and

- composite system SB is thermally isolated
- interaction energy of SB is negligible: $H_{\lambda}^{SB}(\Gamma, \Theta) = H_{\lambda}^S(\Gamma) + H^B(\Theta)$
where Γ and Θ denotes the phases (p, q) of S and B , resp.

then

$$Z_{\lambda}^{SB} = \int d\Gamma d\Theta \exp[-\beta H_{\lambda}^{SB}(\Gamma, \Theta)] = \int d\Gamma \exp[-\beta H_{\lambda}^S(\Gamma)] \int d\Theta \exp[-\beta H^B(\Theta)] = Z_{\lambda}^S Z^B$$

Since SB is an isolated system, it follows a microcanonical distribution. However, for large system, one can use the *thermodynamical limit* and use a canonical distribution:

$$\frac{1}{Z_0^{SB}} \exp[-\beta H_0^{SB}(\Gamma_0, \Theta_0)]$$

Let's now consider a process for which the (time dependent) λ parameter goes from 0 to Λ and drives the system from (Γ_0, Θ_0) to $(\Gamma_{\tau}, \Theta_{\tau})$. The work done during the process is

$$W = H_{\Lambda}^{SB}(\Gamma_{\tau}, \Theta_{\tau}) - H_0^{SB}(\Gamma_0, \Theta_0)$$

“Proof” of the Jarzynski equation: Park & al. 2004

The expectation value for the exponential work $W = H_{\Lambda}^{SB}(\Gamma_{\tau}, \Theta_{\tau}) - H_0^{SB}(\Gamma_0, \Theta_0)$ is

$$\begin{aligned}\langle e^{-\beta W} \rangle &= \int d\Gamma_0 d\Theta_0 \frac{1}{Z_0^{SB}} \exp[-\beta H_0^{SB}(\Gamma_0, \Theta_0)] \times \exp\{-\beta [H_{\Lambda}^{SB}(\Gamma_{\tau}, \Theta_{\tau}) - H_0^{SB}(\Gamma_0, \Theta_0)]\} \\ &= \int d\Gamma_0 d\Theta_0 \frac{1}{Z_0^{SB}} \exp[-\beta H_{\Lambda}^{SB}(\Gamma_{\tau}, \Theta_{\tau})]\end{aligned}$$

The last step consists in transforming the integration variable from (Γ_0, Θ_0) to $(\Gamma_{\tau}, \Theta_{\tau})$. According to Liouville’s theorem (cave!) the Jacobian of the transformation is unity and $d\Gamma_0 d\Theta_0 = d\Gamma_{\tau} d\Theta_{\tau}$

$$\langle e^{-\beta W} \rangle = \int d\Gamma_{\tau} d\Theta_{\tau} \frac{1}{Z_0^{SB}} \exp[-\beta H_{\Lambda}^{SB}(\Gamma_{\tau}, \Theta_{\tau})] = \frac{Z_{\Lambda}^{SB}}{Z_0^{SB}}$$

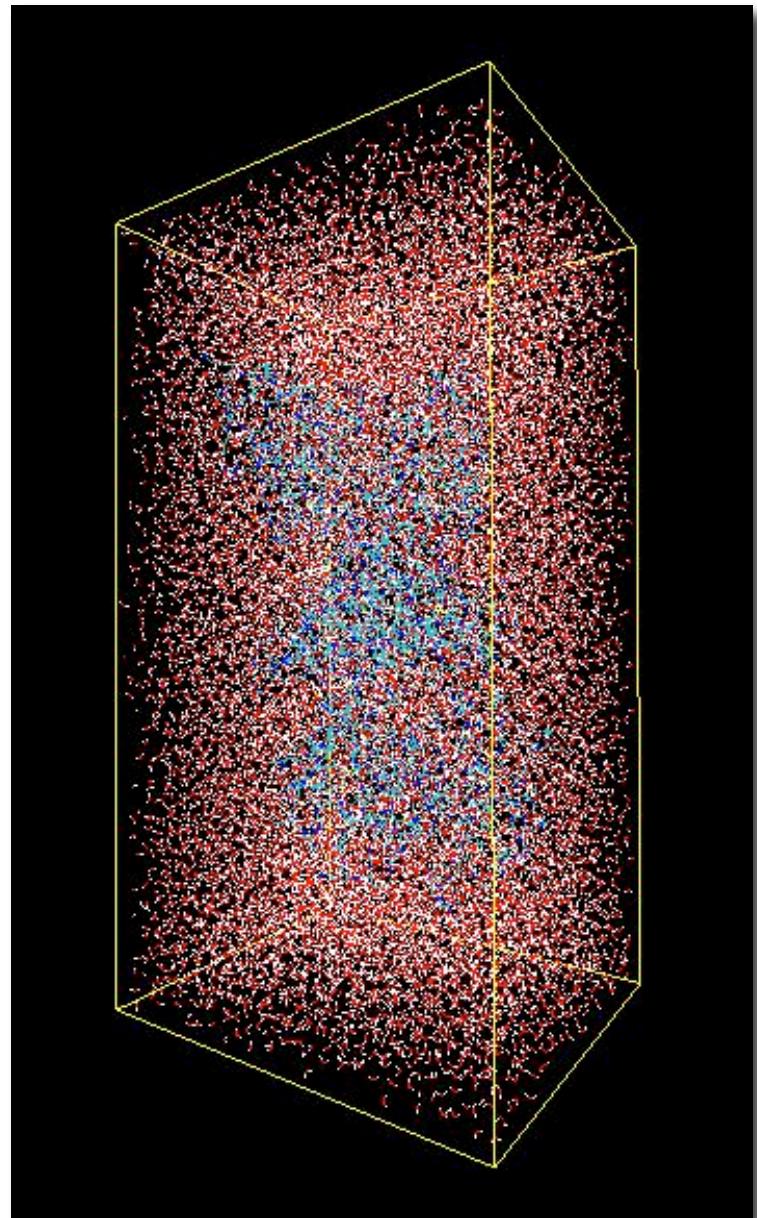
Using the negligible interaction between S and B (see above)

$$\langle e^{-\beta W} \rangle = \frac{Z_{\Lambda}^{SB}}{Z_0^{SB}} = \frac{Z_{\Lambda}^S Z^B}{Z_0^S Z^B} = \frac{Z_{\Lambda}^S}{Z_0^S} = \exp(-\beta \Delta G^S) \quad \text{i.e.}$$

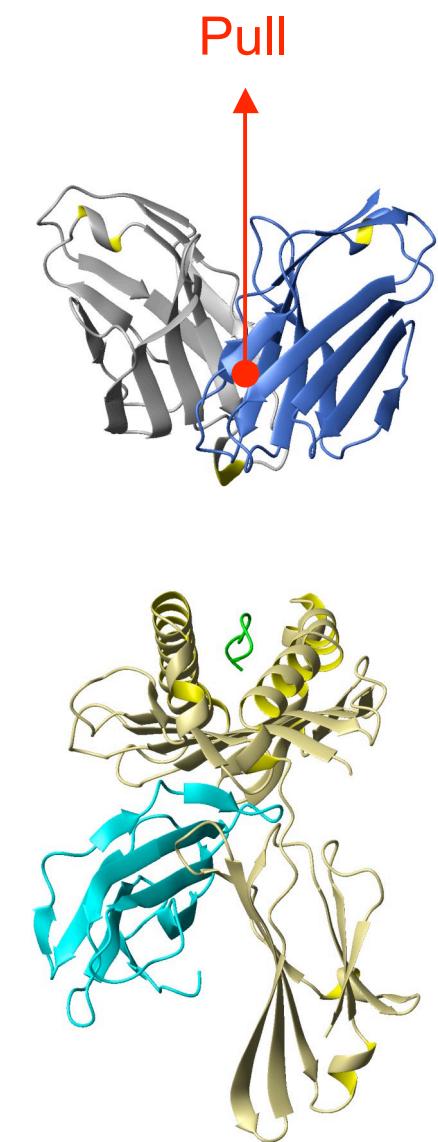
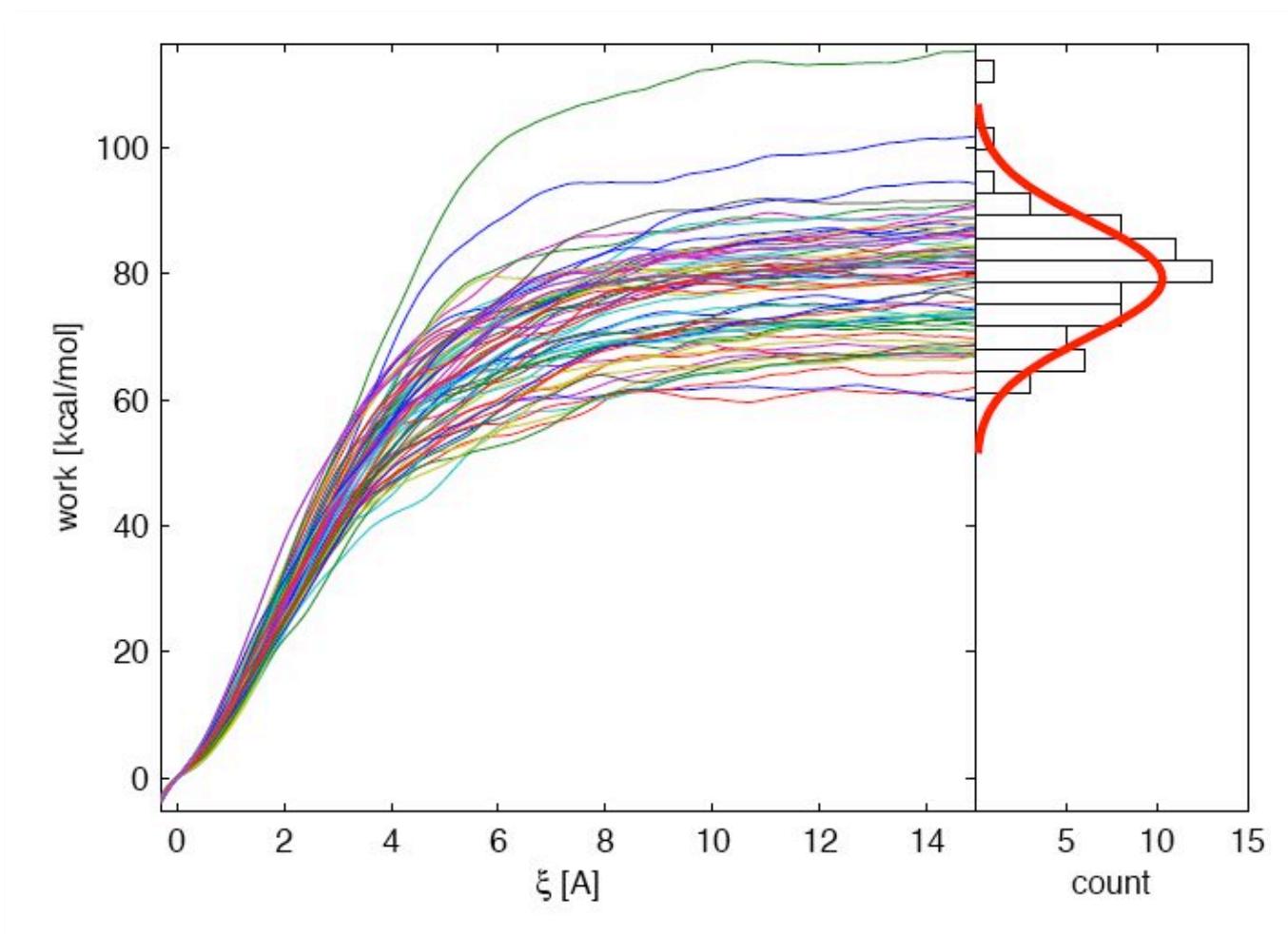
$$\boxed{\langle e^{-\beta W} \rangle = \exp(-\beta \Delta G^S)}$$

Simulation setup

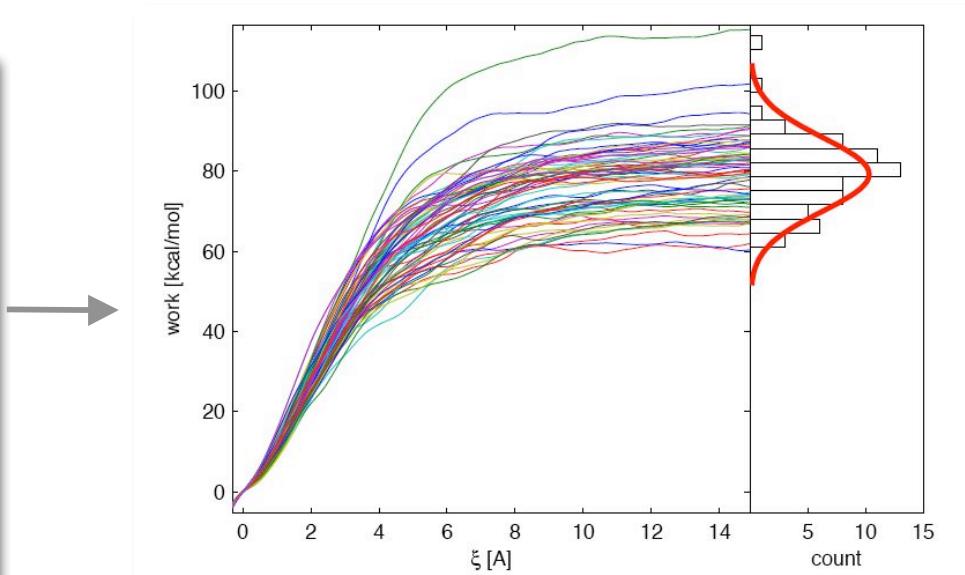
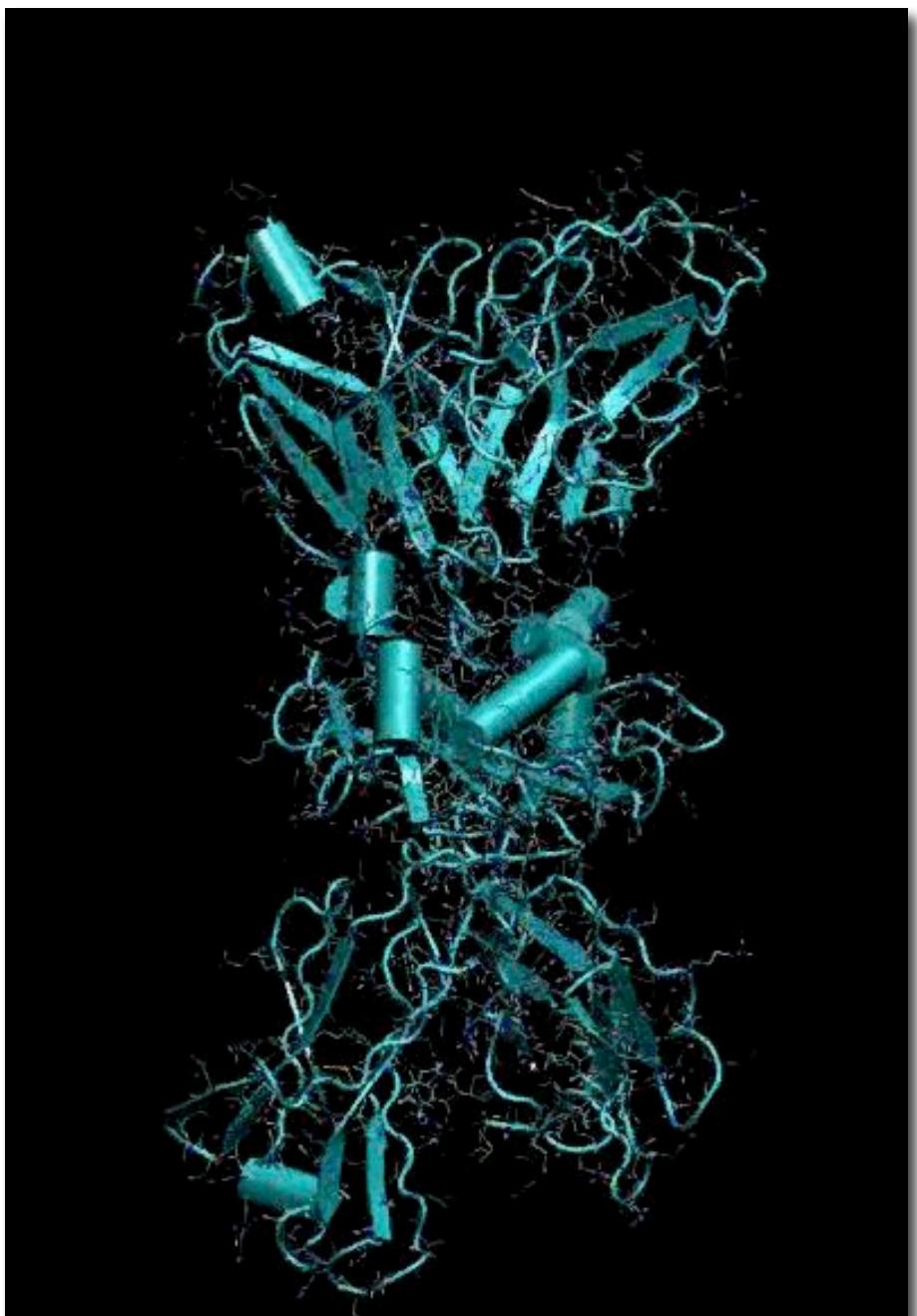
- Gromos96 Force Field
- Gromacs Engine
- Particle Mesh Ewald (PME)
- Periodic boundary conditions
- Box: 80x80x150 Å
- 26000 Water molecules
- 85000 Atoms
- Hydrogen shaken
- 2 fs timestep
- 0.5 ns / 24h on 4 alpha CPU



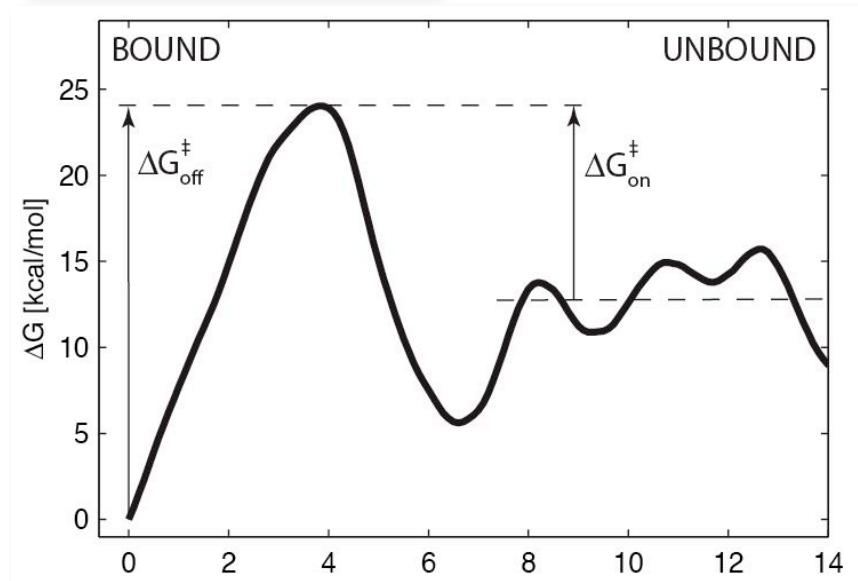
Absolute free energy results



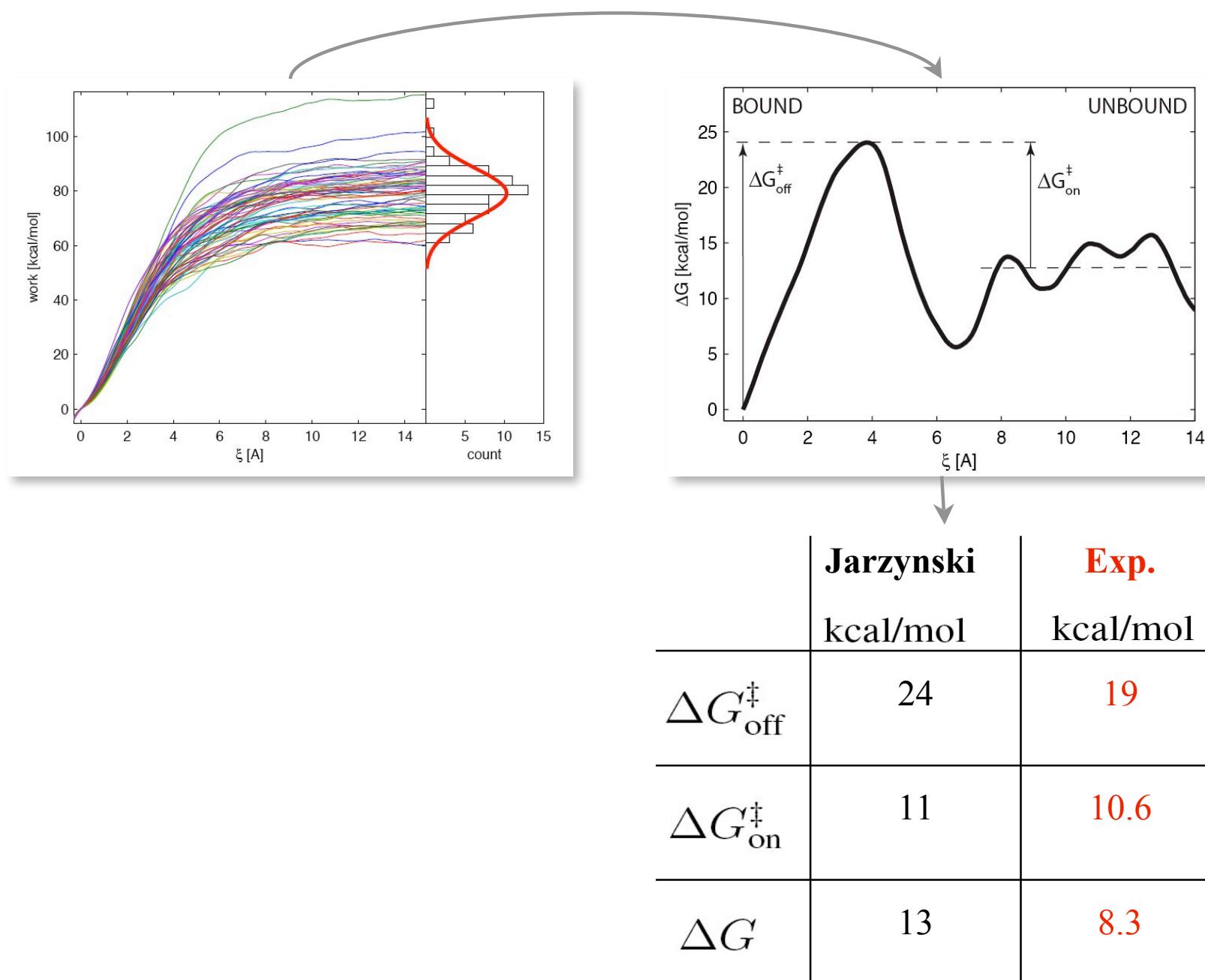
Binding free energy profile:



$$\langle e^{-\beta W} \rangle = e^{-\Delta G}$$



Agreement with experimental values



Free energy simulation: conclusion

Advantages :

- Rigorous
- Estimates influence of small modifications
- No parameter to be fitted
- Partitioning of the free energy (TI)

Drawbacks :

- Restricted to small mutations of ligand or protein
- Most often: relative ΔG_{bind}
- Time consuming

$$\Delta\Delta G_{\text{bind}} = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H_{\text{bonds}}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H_{\text{angles}}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \dots$$

$$\Delta\Delta G_{\text{bind}} = \Delta\Delta G_{\text{bond}} + \Delta\Delta G_{\text{angles}} + \dots$$

S. Boresch *et al.*, *Proteins*, **1994**, 20, 25

S. Boresch and M. Karplus, *J. Mol. Biol.*, **1995**, 254, 801