Modelling of ligand-protein binding I. Computation of thermodynamic quantities

Ferenczy György

Semmelweis University
Department of Biophysics and Radiation Biology

ferenczy.gyorgy@med.semmelweis-univ.hu

Outline

- Molecular dynamics
- Sampling in molecular dynamics
- Calculation of thermodynamic quantities along the pathway
  - Techniques for calculating free energy:
    - Thermodynamic integration
    - Free energy perturbation
    - Potential of mean force
    - Non-equilibrium work
  - Enthalpy and entropy
  - Examples

Molekular dynamics

Introduction

- Molecular dynamics – link between microscopic and macroscopic quantities
  - structure
  - dynamics
  - thermodynamics


FnIII_10_preSMD.mpg

History

- Alder, B. J. and Wainwright, T. E.
J. Chem. Phys. **27**, 1208 (1957)


LETTERS TO THE EDITOR

Phase Transition for a Hard Sphere System

- McCammon, J. A., Gelin, B. R., and Karplus, M.
Nature (Lond.) **267**, 585 (1977)


nature

article

Dynamics of folded proteins

Basic terms

- Thermodynamic/Macroscopic state
  - The system is characterized by few macroscopic parameters; e.g.: T, P, N
- Microscopic state
  - The system is characterized by the positions and momenta of atoms (phase space).
- Ensembles
  - Microscopic states corresponding to a macroscopic state
- Molecular dynamics simulations
  - Generation of microscopic states of an ensemble as a function of time

Thermodynamic ensembles

- Microcanonical – NVE (isolated system)
- Canonical – NVT (thermal equilibrium)
- Izotherm-izobar – NPT
- Grand canonical – μVT (equilibrium with a reservoir of particles)

Ergodic hypothesis

- Measurable quantities: ensemble average $\langle A \rangle_{ensemble}$
  - e.g. (non-covalent) binding of two molecules in solution
- Molecular dynamics: time average $\langle A \rangle_{time}$

$$\langle A \rangle_{time} = \langle A \rangle_{ensemble}$$

- „long enough" MD – appropriate sampling

Force field

- Molecular mechanics
  - „classical"
  - simple, fast computations
  - includes parameters
    - Can be applied within the validity of the parameter space
    - Chemical reactions are typically outside the validity
- Quantum mechanics
  - accurate
  - time intensive computations

Molecular mechanics

Atoms are pointlike objects with mass and interactions

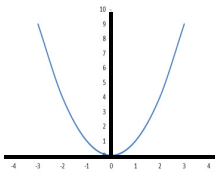
$$E = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

intramolecular — intermolecular

Bond stretching energy

$$E_{str} = k(r - r_0)^2$$

$$F_{str} = -2k(r - r_0)$$



good approximation in the vicinity of r_0
k and r_0 are atom dependent parameters

Bond stretching energy - parameters

$$E_{str} = k(r - r_0)^2$$

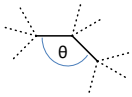
general class	atom type[1]	description
hydrogen types		
	H	amide or imino hydrogen
	HC	explicit hydrogen attached to carbon
	HO	hydrogen on hydroxyl oxygen
	HS	hydrogen attached to sulfur
	HW	hydrogen in water
	H2	amino hydrogen in NH_2
	H3	hydrogen of lysine or arginine (positively charged)
all-atom carbon types[2]		
	C	sp^2 carbonyl carbon and aromatic carbon with hydroxyl substituent in tyrosine
	CA	sp^2 aromatic carbon in 6-membered ring with 1 substituent
	CB	sp^2 aromatic carbon at junction between 5- and 6-membered rings
	CC	sp^2 aromatic carbon in 5-membered ring with 1 substituent and next to a nitrogen
	CK	sp^2 aromatic carbon in 5-membered ring between 2 nitrogens and bonded to 1 hydrogen (in purine)
	CM	sp^2 same as CJ but one substituent
	CN	sp^2 aromatic junction carbon in between 5- and 6-membered rings
	CQ	sp^2 carbon in 6-membered ring of purine between 2 NC nitrogens and bonded to 1 hydrogen
	CR	sp^2 aromatic carbon in 5-membered ring between 2 nitrogens and bonded to 1 H (in his)
	CT	sp^3 carbon with 4 explicit substituents
	CV	sp^2 aromatic carbon in 5-membered ring bonded to 1 N and bonded to an explicit hydrogen
	CW	sp^2 aromatic carbon in 5-membered ring bonded to 1 N-H and bonded to an explicit hydrogen
	C*	sp^2 aromatic carbon in 5-membered ring with 1 substituent

Bond Stretching Potential Parameters

Bond	used for	K_r/ kcal mol^{-1} Å^{-2}	R_{eq}/ Å
CT-CT	*BMI˙.EMI˙*	310.0	1.526
CT-H1	*BMI˙.EMI˙*	340.0	1.090
CT-HC	*BMI˙.EMI˙*	340.0	1.090
CT-NA	*BMI˙.EMI˙*	337.0	1.475
CR-H5	*BMI˙.EMI˙*	367.0	1.080
CR-NA	*BMI˙.EMI˙*	477.0	1.343
CW-H4	*BMI˙.EMI˙*	367.0	1.080
CW-NA	*BMI˙.EMI˙*	427.0	1.381
CW-CW	*BMI˙.EMI˙*	549.0	1.350
AL-Cl	*TCA˙*	116.1	2.170
P-F	*PF_6˙*	260.3	1.646
NN-ON	*NO_3˙*	300.0	1.260

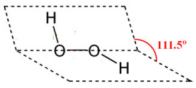
Bending energy

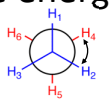
$$E_{bend} = k(\theta - \theta_0)^2$$



k and θ_0 are atom dependent parameters

Torsional/dihedral angle energy



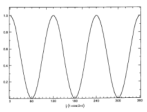


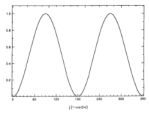
$$E_{tors} = \frac{V_n}{2}[1 + cos(n\varphi - \varphi_0)]$$

V_n – barrier height
n - periodicity

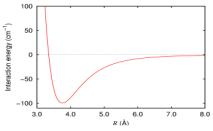
```
Dihedral angle          Vn,      n    ϕ0      comments
(atom types)          kcal/mol

C    CT1  NH1  C        0.2000   1   180.00 ! backbone  phi
NH1  C    CT1  NH1      0.6000   1     0.00 ! backbone  ksi
CT1  C    NH1  CT1      1.6000   1     0.00 ! backbone omega
CA   CA   CA   CA       3.1000   2   180.00 ! Phe side chain
H    OH1  CT2  CT1      0.4200   3     0.00 ! Ser side chain
```

~15000 parameters





van der Waals energy



short range:	repulsive;	exp(-r) or r^{-12}	Pauli repulsion
middle range:	attractive;	r^{-6}	dispersion
long range:	disappears		

$$E_{vdw} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12}\right] - \left[\left(\frac{\sigma}{r}\right)^{6}\right]$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$$

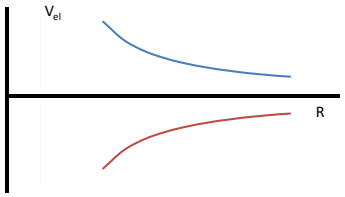
$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$$

$$c * exp\left(\frac{-r}{\sigma}\right)$$

Electrostatic energy

$$V_{el} = \frac{q_i q_j}{\epsilon r_{ij}}$$

Coulomb

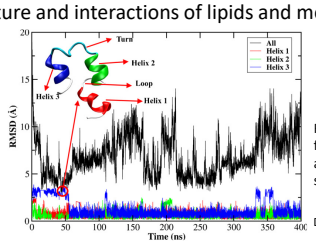


MM parameters

- Derivation
  - Quantum mechanical calculations
  - Experimental data
  - Extension based on analogy
- Validation by comparing computed and experimental data
  - Macromolecular structure
  - NMR data
  - Structure and energy of van der Waals complexes
- Error compensation; mutual interdependence of parameters

Quality of MM force field

- Protein structure
- DNA, RNA structure
- Conformation of organic molecules
- Ligand-protein interactions
- Structure and interactions of lipids and membranes
- …



RMSDs of backbone atoms from the native structure as a function of MD simulation time

Duan J Mol Model 2014,20,2195

Selected MM force fields

- Charmm (Chemistry at HARvard Macromolecular Mechanics)
- AMBER (Assisted Model Building with Energy Refinement)
- OPLS (Optimized Potentials for Liquid Simulations)
- GROMOS (GROningen Molecular Simulation)
- MMFF (Merck Molecular Force Field)

MD algorithm

Newtonian mechanics

$$r^N(r_1, r_2 \dots r_N) \qquad p^N(p_1, p_2 \dots p_N)$$

$$U(\underline{r}) \qquad K\left(\underline{p}\right) = \sum_i \frac{|p_i|^2}{2m_i}$$

$$H = K + U \qquad \dot{r}_i = \frac{p_i}{m_i} \qquad \dot{p}_i = f_i$$

Verlet algorithm

Calculation of p_i and r_i at δt time steps

$$p_i(t), r_i(t) \rightarrow p_i(t+\delta t), r_i(t+\delta t) \qquad \rightarrow f_i(t)$$

$$p_i\left(t + \frac{1}{2}\delta t\right) = p_i(t) + \frac{1}{2}\delta t \cdot f_i(t)$$

$$r_i(t+\delta t) = r_i(t) + \frac{\delta t \cdot p_i\left(t + \frac{1}{2}\delta t\right)}{m_i} \qquad \rightarrow f_i(t+\delta t)$$

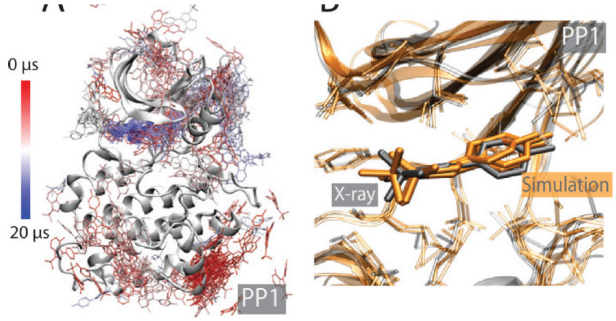
$$p_i(t+\delta t) = p_i\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t \cdot f_i\,(t+\delta t)$$

Typical δt for simulation of biochemical systems: 1-4 fs

Selected applications

Application 1

process of ligand binding



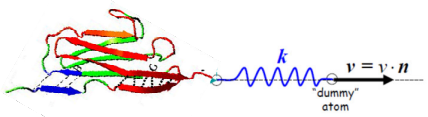
PP1 molecule finds the binding site of Src kinase in a 15µs simulation

JACS 2011 133 9181

Application 2

Steered MD

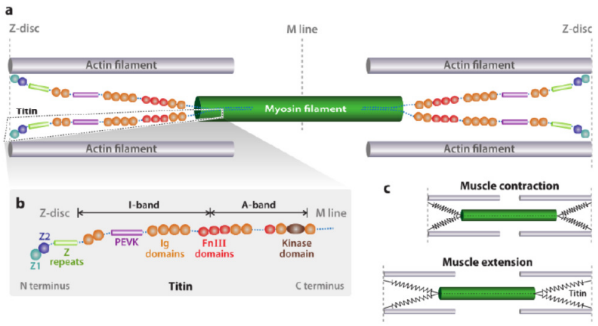
- Constant velocity pulling



$$U = \frac{1}{2}k[vt - (\vec{r} - \vec{r}_0) \cdot \vec{n}]^2$$

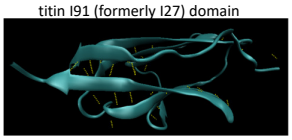
- Constant force pulling

Titin structure and function



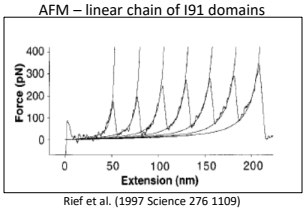
Unfolding and force

titin I91 (formerly I27) domain

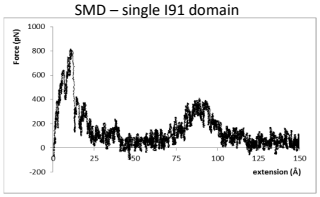


Constant velocity pulling

AFM – linear chain of I91 domains

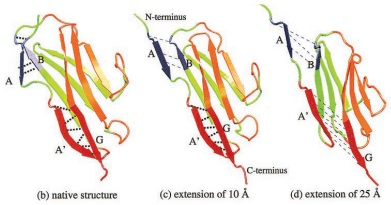


SMD – single I91 domain



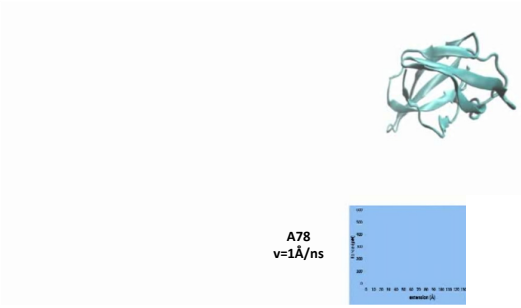
Rief et al. (1997 Science 276 1109)

Unfolding and structure



Lu, H., and Schulten, K. (2000).. Biophys. J. 79, 51–65

Titin Fn domain unfolding mechanism



A78
v=1Å/ns

MD scope and limitation

- Scope
  - Structural study; structure refinement
  - Dynamics
    - conformations, ligand-protein binding, steered processes,…
  - Thermodynamics see later
    - Free energy changes
      - solvation, ligand-protein binding,…
- Limitations
  - Sampling see later
  - Accuracy of force field
  - Chemical reactions cannot be routinely studied

MD - sampling

- Microstates appear according to Boltzmann distribution
  - $exp\left(-\frac{E}{kT}\right)$
- Simulation time is limited by computational capacity
  - Time scale for proteins: ~µs
- Rare events with high energy barrier cannot be straightforwardly simulated

MD - Sampling

Time scale (s)	Amplitude (Å)	Description	# MD steps (step ~ fs)
10^{-15}-10^{-12}	0.001-0.1	Bond stretching, bond angle deformation	*1-1000*
10^{-12}-10^{-9}	0.1-10	Protein sidechain, loop and collective motions	*10^{3}-10^{6}*
10^{-9}-10^{-6}	1-100	Folding of small proteins	*10^{6}-10^{9}*
10^{-6}-10^{-1}	10-100	Protein folding, Ligand-protein binding	*10^{9}-10^{14}*

Free energy - Sampling

$$F = -kT\ln\left[h^{-3N}\iint exp\left(-\frac{E(r,p)}{kT}\right)dpdr\right]$$ (1) Formula for free energy

Free energy calculation with MD sampling is problematic

phase space incomplete in (1)
positive integral
ln function increases monotonically
negative contribution missing
F overestimated

Free energy - Sampling

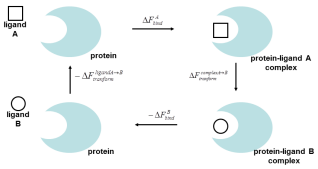
$$F = -kT\ln\left[h^{-3N}\iint exp\left(-\frac{E}{kT}\right)dpdr\right]$$ Formula for free energy

$$F' = kT\ln\left[\frac{\int exp\left(+\frac{E}{kT}\right)exp\left(-\frac{E}{kT}\right)dr}{\int exp\left(-\frac{E}{kT}\right)dr}\right] = kT\ln\left\langle exp\left(\frac{E}{kT}\right)\right\rangle$$ Formulated as expected value

Terms with high E contribute to F significantly, but their sampling has a low probability

Free energy difference

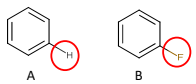
- Sampling issue hampers the calculation of F and $\Delta F=F_{Bound}-F_{Free}$
- Special techniques for calculating $\Delta F=F_B-F_A$ (A similar to B) for similar systems (see later)
- Thermodynamic cycle: binding free energy difference ($\Delta\Delta F$) of two similar ligands is obtained from the free energy difference of similar systems



– $\Delta\Delta F = \Delta F^{A}_{bind} - \Delta F^{B}_{bind} = \Delta F^{complexA\to B}_{transform} - \Delta F^{ligandA\to B}_{transform}$

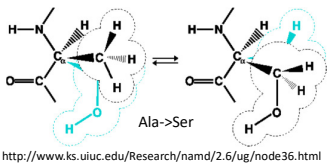
– „alchemical" transformations: $\Delta F^{complexA\to B}_{transform}$ and $\Delta F^{ligandA\to B}_{transform}$
  - 2 transformations to obtain $\Delta\Delta F$

Alchemical transformation- coupling parameter



H(r,p,λ=0) H(r,p,λ=1)

$H_\lambda = H(r,p,\lambda) = (1-\lambda)\,H_A + \lambda H_B$
λ – coupling parameter
H_λ may be other function of λ
H_A, H_B may depend on λ



http://www.ks.uiuc.edu/Research/namd/2.6/ug/node36.html

- Large perturbation – important change in the environment
- Large perturbation is computationally impractical

MD techniques to calculate ΔF

- Thermodynamic integration (TI)
- Free energy perturbation (FEP)
- Potential of mean force (PMF)
- Non-equilibrium work (Jarzynski equation)

Thermodynamic integration

$$\Delta F = F_B - F_A = F(\lambda=1) - F(\lambda=0) = \int_{\lambda=0}^{\lambda=1} F'(\lambda)\,d\lambda$$

$$F' = \frac{d}{d\lambda}[-kT \ln Z(\lambda)] = \left\langle \frac{\partial E}{\partial \lambda} \right\rangle$$

$$F' = \left\langle \frac{\partial E}{\partial \lambda} \right\rangle \rightarrow \frac{dF(x,\lambda)}{d\lambda} = \frac{d\left(-kT \ln \int e^{-\frac{E(x,\lambda)}{kT}} dx\right)}{d\lambda}$$

$$= -kT \frac{1}{\int e^{-\frac{E(x,\lambda)}{kT}} dx} \int e^{-\frac{E(x,\lambda)}{kT}} \left(-\frac{1}{kT}\frac{dE(x,\lambda)}{d\lambda}\right) dx = \left\langle \frac{dE(x,\lambda)}{d\lambda} \right\rangle$$

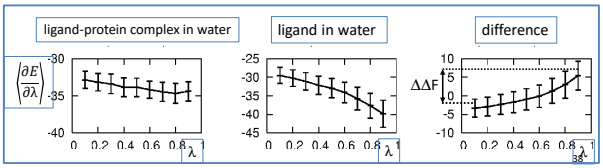
P

E.g.: $E(x,\lambda) = (1-\lambda)\,E_A(x) + \lambda E_B(x)$;
$dE/d\lambda = E_B - E_A$

Thermodynamic integration

- simulation
  1. λ evolves in time
  2. MD at several fixed λ value
  - Calculation of $\left\langle \frac{\partial E}{\partial \lambda} \right\rangle$
    - E.g.: $\frac{\partial E}{\partial \lambda} = E_B - E_A$ -> $\langle E_B - E_A \rangle_\lambda$ depends on λ; $\frac{\partial E}{\partial \lambda}$ may be more involved
  - Numerical integration

$$\Delta F = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_\lambda d\lambda$$



Free energy perturbation

$$F = -kT \ln\left[\int exp\left(-\frac{E}{kT}\right) dr\right]$$ (1) Formula for free energy

$$F_B - F_A = -kT \ln\left[\frac{\int exp\left(-\frac{E_B}{kT}\right) dr}{\int exp\left(-\frac{E_A}{kT}\right) dr}\right]$$ (2) Free energy difference for two systems

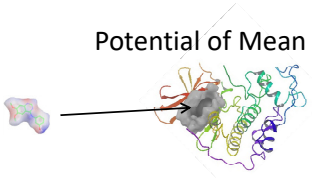
$$F_B - F_A = -kT \ln\left[\frac{\int exp\left(-\frac{E_A}{kT}\right) exp\left(\frac{E_A}{kT}\right) exp\left(-\frac{E_B}{kT}\right) dr}{\int exp\left(-\frac{E_A}{kT}\right) dr}\right]$$ (3) $1 = exp\left(-\frac{E_A}{kT}\right) exp\left(\frac{E_A}{kT}\right)$

$$F_B - F_A = -kT \ln\left[\frac{\int exp\left(-\frac{E_A}{kT}\right) exp\left(-\frac{\Delta E}{kT}\right) dr}{\int exp\left(-\frac{E_A}{kT}\right) dr}\right]$$ (4) $\Delta E = E_B - E_A$

P

$$F_B - F_A = -kT \ln\left[\left\langle exp\left(-\frac{\Delta E}{kT}\right)\right\rangle\right]_A$$ (5) Expectation value

Potential of Mean Force(PMF)



x – intermolecular coordinates
y – intramolecular coordinates

$$F = -RT \ln\left[\int exp\left(-\frac{E(\underline{x},\underline{y})}{RT}\right) d\underline{x}d\underline{y}\right]$$ Free energy

$$F(\underline{x}) \approx -RT \ln\left[\int exp\left(-\frac{E(\underline{x},\underline{y})}{RT}\right) d\underline{y}\right]$$ Potential of mean force

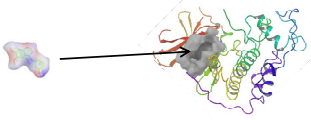
explanation

$$\frac{dF(\underline{x})}{d\underline{x}} = \frac{\int \frac{dE(\underline{x},\underline{y})}{d\underline{x}} e^{-\frac{E(\underline{x},\underline{y})}{RT}} d\underline{y}}{\int e^{-\frac{E(\underline{x},\underline{y})}{RT}} d\underline{y}} = -\langle \varphi(\underline{x}) \rangle_{\underline{y}}$$

P

F – potential; φ – force

Potential of Mean Force(PMF)



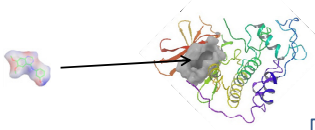
x – intermolecular coordinates
y – intramolecular coordinates

Relation between probability ($P(\underline{x})$) and PMF ($F(\underline{x})$) of $\underline{x}$

$$P(\underline{x}) = \int P(\underline{x},\underline{y})\,d\underline{y} = \frac{\int e^{-\frac{E(\underline{x},\underline{y})}{RT}} d\underline{y}}{\int e^{-\frac{E(\underline{x},\underline{y})}{RT}} d\underline{x}d\underline{y}} = \frac{e^{-\frac{F(\underline{x})}{RT}}}{\int e^{-\frac{E(\underline{x},\underline{y})}{RT}} d\underline{x}d\underline{y}}$$ $P(\underline{x})$ – $\underline{x}$ probability

$$F(\underline{x}) - F_{Ref} = -RT \ln \frac{P(\underline{x})}{P(Ref)}$$

Potential of Mean Force (PMF)

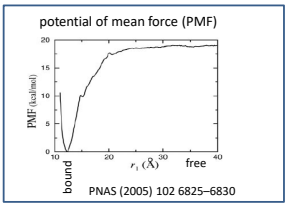


x – intermolecular coordinates

Computation:
$P(\underline{x})$ is calculated along $\underline{x}$

$$F(\underline{x}) - F_{Ref} = -RT \ln \frac{P(\underline{x})}{P(Ref)}$$

$$F(\underline{x}) = -RT \ln P(\underline{x}) + const.$$



Special sampling techniques needed
Standard binding free energy can be calculated

Non-equilibrium work

$$\Delta F = F_2 - F_1 = -kT \ln\left\langle \exp\left(-\frac{W}{kT}\right)\right\rangle$$ (Jarzynski)

- Expected value of work (W) obtained along non-equilibrium paths
- Fast transformation between the states is possible
- Appropriate sampling is challenging
- Current techniques are not superior to equilibrium methods

Calculation of Enthalpy and Entropy

- ΔF can be calculated as the <u>expected value of</u> some function of energy <u>differences</u>
  - *cf.* TI and FEP
- ΔH, TΔS can be calculated as the <u>difference of the expected values</u> of state functions
- ΔH, TΔS can be calculated with significantly lower accuracy than ΔF

Example – binding to lysozyme

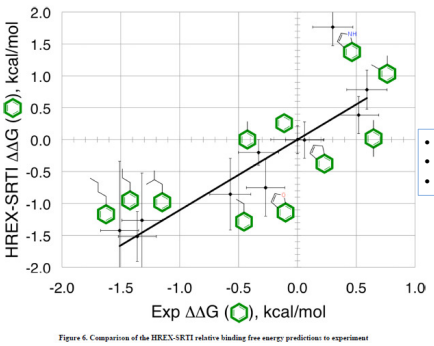
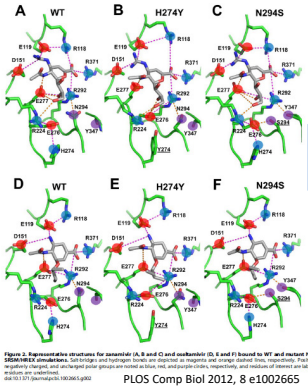


Figure 6. Comparison of the HREX-SRTI relative binding free energy prediction to experiment

- ΔΔG
- Thermodynamic integration
- Small structural changes

J. Chem. Theor. Comput. 2011, 7 3001

Example – neuraminidase inhibitors 1



- 2 ligands
- 3 proteins: wild type + 2 mutants

Y252H not shown

PLOS Comp Biol 2012, 8 e1002665

Example – neuraminidase inhibitors 2

Table 1. Comparison of experimental ΔΔG in oseltamivir and zanamivir for three NA mutations with estimates obtained using different computational approaches.

Method	H274Y		N294S		Y252H		RMSE
	ΔΔG, kcal/mol		ΔΔG, kcal/mol		ΔΔG, kcal/mol		(RMSD), kcal/mol
	zanamivir	oseltamivir	zanamivir	oseltamivir	zanamivir	oseltamivir	
Experimental[a]	**0.4 (0.1)**	**3.3 (0.2)***	**1.2 (0.1)***	**2.6 (0.2)***	**0.1 (0.2)**	**−1.4 (0.1)**	N/A (0.2)
SRMM	−5.8 (7.4)	0.7 (7.0)	8.2 (7.7)	5.8 (6.2)	−0.1 (8.7)	−0.9 (7.4)	4.2 (7.4)
SRSM	1.7 (2.9)	1.2 (3.0)	0.6 (2.0)	1.7 (1.9)	1.5 (1.7)	0.5 (1.5)	1.5 (2.2)
SRSM/HREX	1.3 (0.8)	4.1 (2.4)	2.3 (0.4)	2.2 (0.9)	0.6 (0.8)	0.7 (1.4)	1.1 (1.1)
MM-GBSA	6.2 (8.1)	0.9 (3.8)	5.7 (6.1)	−5.9 (3.6)	2.1 (2.9)	−1.9 (3.0)	4.8 (4.6)
MM-PBSA	8.4 (10.1)	3.0 (3.9)	5.8 (4.5)	−4.7 (3.2)	2.8 (3.1)	0.2 (2.6)	5.0 (4.6)
Rosetta	−0.4 (0.5)	0.8 (0.4)	−0.4 (0.3)	0.3 (0.2)	−0.1 (0.4)	0.0 (0.0)	1.7 (0.3)

[a]Values were derived from the data reported by Collins et al [10].
Standard deviations are shown in parentheses. Root mean squared error (RMSE) and the RMS Standard Deviation (RMSD) are provided.
'*'indicates experimentally determined drug resistant mutation. 'N/A' stands for *not applicable*.
doi:10.1371/journal.pcbi.1002665.t001

PLOS Comp Biol 2012, 8 e1002665

- ΔΔG; TI
- Wild type/mutant
- Experimental/calculated
- (Standard deviation)

Relative affinities toward various sequences are correctly predicted for both ligands

Relative affinities of ligands toward a protein are not always predicted correctly

Example – FKBP12-ligand

Standard binding free energy
Double decoupling
FEP
Free energy components

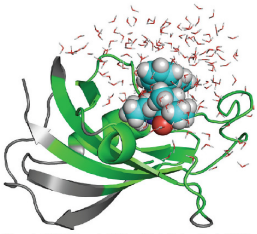


Figure 4. FKBP12 bound with ligand #8 studied previously.[40,42] The gray parts are treated as a mean-field approximation with generalized solvent boundary potential.[40] See ref 42 for computational details.

$\Delta\Delta G_{rep}$	$\Delta\Delta G_{dis}$	$\Delta\Delta G_{elec}$	$\Delta\Delta G_c$	$\Delta\Delta G_t^\circ$	$\Delta\Delta G_r$	ΔG°_{bind}	exptl
−1.1	−21.1	−3.7	6.9	3.4	5.4	−10.2	−10.9
			conf	transl	rot		

J. Phys. Chem. B **2009,** *113,* 2234

Summary

- Computational modelling of protein-ligand interactions – 1st part
  - MD based methods – potentially accurate
  - Major challenge: sampling
  - $\Delta\Delta F$ ($\Delta\Delta G$) can be calculated efficiently; „Alchemical" transformations
  - Techniques to calculate ΔF (between similar states)
    - Thermodynamic integration
    - Free energy perturbation
    - Potential of mean force
    - Non-equilibrium work
  - Non-routine applications; varying accuracy
  - Enthalpy and entropy can be calculated with lower accuracy