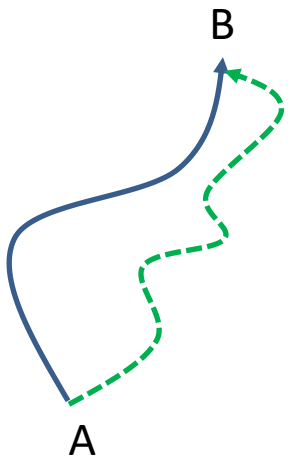
# Modelling of ligand-protein binding
# I. Computation of thermodynamic quantities

# Outline

- Molecular dynamics
  - Theoretical background
- Force fields
  - Energy terms
  - Parameters
- Sampling in molecular dynamics
- Free energy difference and alchemical transformations
- Examples



# Molecular 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)

1208 LETTERS TO THE EDITOR

relatively free diffusion, while in the low $g(\sigma)$ states diffusion is much restricted.

The conjecture[1] that some high-order virial coefficients might be negative is not necessarily supported by the present results, since only to the left of the apparent transition do the latter give lower pressures than the five-term virial expression.

Some further investigation for both 32 molecule and larger systems will be made on the present calculators, but a satisfactory determination of the detailed behavior in the apparent transition region will require higher speed equipment. The possibility that a similar phenomenon for hard spheres in two dimensions may have been missed in the original Monte Carlo calculations[1] will also be investigated.

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] M. N. Rosenbluth and A. W. Rosenbluth, J. Chem. Phys. **22**, 881 (1954).
[2] Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller, J. Chem. Phys. **21**, 1087 (1953).
[3] B. J. Alder and T. Wainwright, J. Chem. Phys. **27**, 1208 (1957).
[4] W. W. Wood and F. R. Parker, J. Chem. Phys. **27**, 720 (1957). This paper discusses the Monte Carlo method in some detail, as well as giving computational results for Lennard-Jones molecules.
[5] Kirkwood, Maun, and Alder, J. Chem. Phys. **18**, 1040 (1950).

**Phase Transition for a Hard Sphere System**

B. J. Alder and T. E. Wainwright
*University of California Radiation Laboratory, Livermore, California*
(Received August 12, 1957)

A CALCULATION of molecular dynamic motion has been designed principally to study the relaxations accompanying various nonequilibrium phenomena. The method consists of solving exactly (to the number of significant figures carried) the simultaneous classical equations of motion of several hundred particles by means of fast electronic computors. Some of the details as they relate to hard spheres and to particles having square well potentials of attraction have been described.[1,2] The method has been used also to calculate equilibrium properties, particularly the equation of state of hard spheres where differences with previous Monte Carlo[3] results appeared.

The calculation treats a system of particles in a rectangular box with periodic boundary conditions.[4] Initially, the particles are in an ordered lattice with velocities of equal magnitude but with random orientations. After a very short initial run[1,3] the system reached the Maxwell-Boltzmann velocity distribution so that the pressure could thereafter be evaluated directly by means of the virial theorem, that is by the rate of change of the momentum of the colliding particles.[1,3] The pressure has also been evaluated from the radial distribution function.[5] Agreement between the two methods is within the accuracy of the calculation.

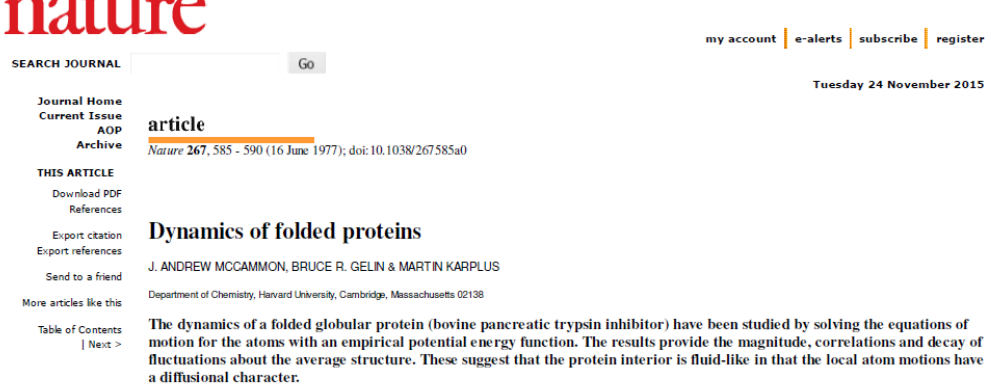
A 32-particle system in a cube and initially in a face-centered cubic lattice proceeded at about 300 collisions an hour on the UNIVAC. For comparison a 96-particle system in a rectangular box and initially in a hexagonal arrangement has been calculated, however only at high densities so far. No differences in the pressures can be detected. It became apparent that some long runs were necessary at intermediate densities, accordingly the IBM-704 was utilized where, for 32 particles, an hour is required for 7000 collisions. Larger systems of 108, 256, and 500 particles can also conveniently be handled; in an hour 2000, 1000, and 500 collisions, respectively, can be calculated. The results for 256 and 500 particles are not now presented due to inadequate statistics.

The equation of state shown in Fig. 1 of the accompanying paper[6] for 32 and 108 particles is for the intermediate region of density, where disagreement was found with the previous Monte Carlo results. The volume, $v$, is given relative to the volume of close packing, $v_0$. Plotted also are the more extended Monte Carlo results; the agreement between these three systems is within the present accuracy of the pressure determination. This agreement provides an interesting confirmation of the postulates of statistical mechanics for this system.

Figure 1 of the accompanying paper shows two separate and overlapping branches. In the overlapping region the system can, at a given density, exist in two states with considerably different pressures. As the calculation proceeds the pressure is seen to jump suddenly from one level to the other. A study of the positions of the particles reveals that as long as the system stays on the lower branch of the curve the particles are all confined to the narrow region in space determined by their neighbors, while on the upper branch of the curve the particles have acquired enough freedom to exchange with the surrounding particles. Since the spheres are originally in ordered positions, the system starts out on the lower branch; the first jump to the upper branch can require very many collisions. The trend, as expected, is that at higher densities more collisions are necessary for the first transition, however, there are large deviations. At $v/v_0=1.60$, 5000 collisions were required; at 1.55, 25 000; while at 1.54 only 400; at 1.535, 7000; at 1.53, 75 000; and at 1.525, 95 000. Runs in excess of 200 000 collisions at $v/v_0$ of 1.55 and 1.53 have not shown any return to the lower branch, while at 1.525 the system has returned several times, however only for relatively few collisions. The lowest density at which the system did not jump to the upper curve is at 1.50, however the run extends only to 50 000 collisions and at that density it might take very many collisions before the appropriate fluctuation occurs for a molecule to escape from its neighborhood. For comparison, the first jump for 108 particles occurred for $v/v_0=1.55$ and 1.60 at about 2000 collisions. This is fewer collisions per particle than for the smaller system and is indicative of larger possible density fluctuations in larger systems. Apparently, the

- McCammon, J. A., Gelin, B. R., and Karplus, M.

*Nature (Lond.)* **267**, 585 (1977)



nature

my account | e-alerts | subscribe | register

SEARCH JOURNAL Go

Tuesday 24 November 2015

Journal Home
Current Issue
AOP
Archive

THIS ARTICLE
Download PDF
References
Export citation
Export references
Send to a friend
More articles like this
Table of Contents
| Next >

article

*Nature* **267**, 585 - 590 (16 June 1977); doi:10.1038/267585a0

**Dynamics of folded proteins**

J. ANDREW MCCAMMON, BRUCE R. GELIN & MARTIN KARPLUS

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

**The dynamics of a folded globular protein (bovine pancreatic trypsin inhibitor) have been studied by solving the equations of motion for the atoms with an empirical potential energy function. The results provide the magnitude, correlations and decay of fluctuations about the average structure. These suggest that the protein interior is fluid-like in that the local atom motions have a diffusional character.**

# 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).
- Ensemble
  - 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

- Experiment: measure the average for ~$10^{23}$ molecules at a certain time
- MD: calculate the time average for one (few) molecule(s)

- 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

- Quantum mechanics
  - basic entities: nuclei and electrons
  - accurate
  - time intensive computations
- Molecular mechanics
  - basic entities: atoms
  - „classical”
  - simple, fast computations
  - includes parameters
    - Can be applied within the validity of the parameter space
    - Chemical reactions are typically outside the validity

# Molecular mechanics

Atoms are pointlike objects with mass and interactions

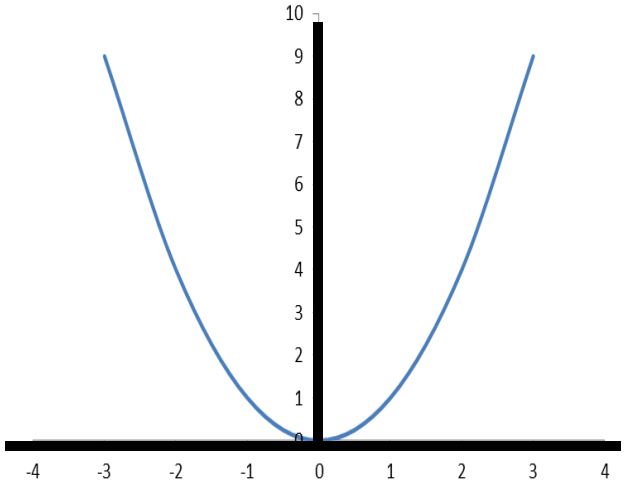
$$E = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

intramolecular intermolecular

# Bond stretching energy

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

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

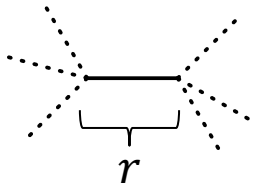



good approximation in the vicinity of $r_0$

$k$ and $r_0$ are atom dependent parameters

# Bond stretching energy - parameters

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



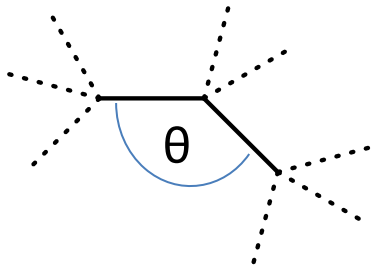

| general class | atom type[1] | description |
|---|---|---|
| *hydrogen types* | | |
| | H | amide or imino hydrogen |
| | HC | explicit hydrogen attached to carbon |
| | HO | hydrogen on hydroxyl oxygen |
| | HS | hydrogen attached to sulfur |
| | HW | hydrogen in water |
| | H2 | amino hydrogen in $NH_2$ |
| | 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 |

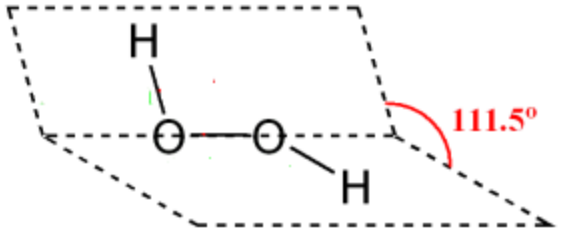
# Bending energy

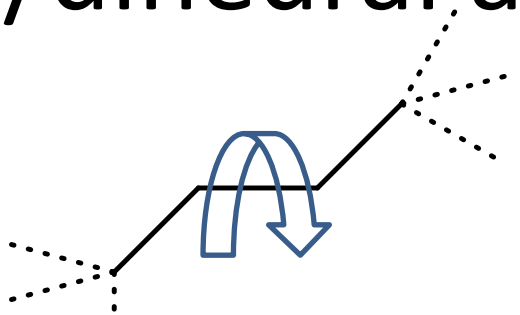
$$E_{bend} = k(\theta - \theta_0)^2$$

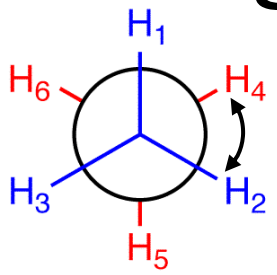



$k$ and $\theta_0$ are atom dependent parameters

# Torsional/dihedral angle energy





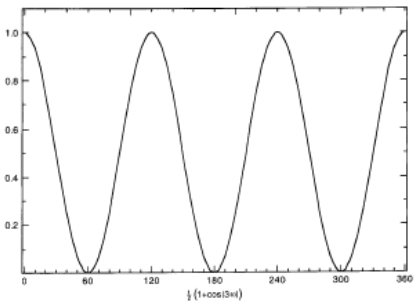

$$E_{tors} = \frac{V_n}{2}[1 + cos(n\varphi - \varphi_0)]$$

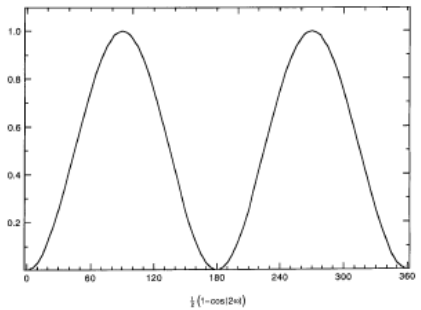
$V_n$ – barrier height

n - periodicity

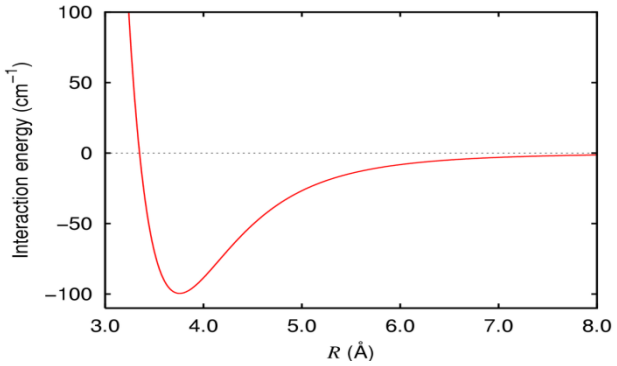
| Dihedral angle (atom types) | | | | $V_n$, kcal/mol | n | $\phi_0$ | comments |
|---|---|---|---|---|---|---|---|
| C | CT1 | NH1 | C | 0.2000 | 1 | 180.00 | ! backbone phi |
| NH1 | C | CT1 | NH1 | 0.6000 | 1 | 0.00 | ! backbone ksi |
| CT1 | C | NH1 | CT1 | 1.6000 | 1 | 0.00 | ! backbone omega |
| 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 | | |

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

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

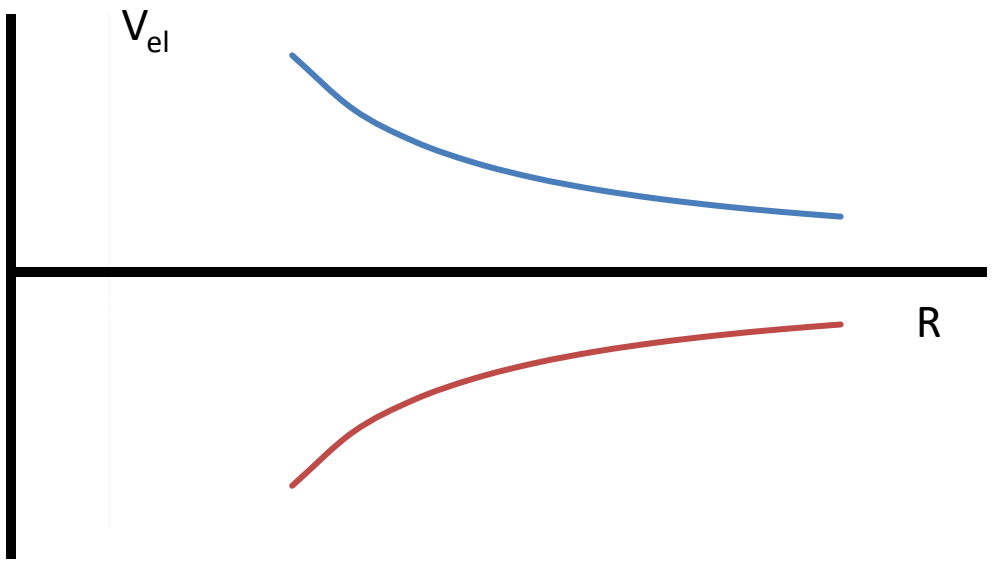
$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$$

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

# Electrostatic energy

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

Coulomb

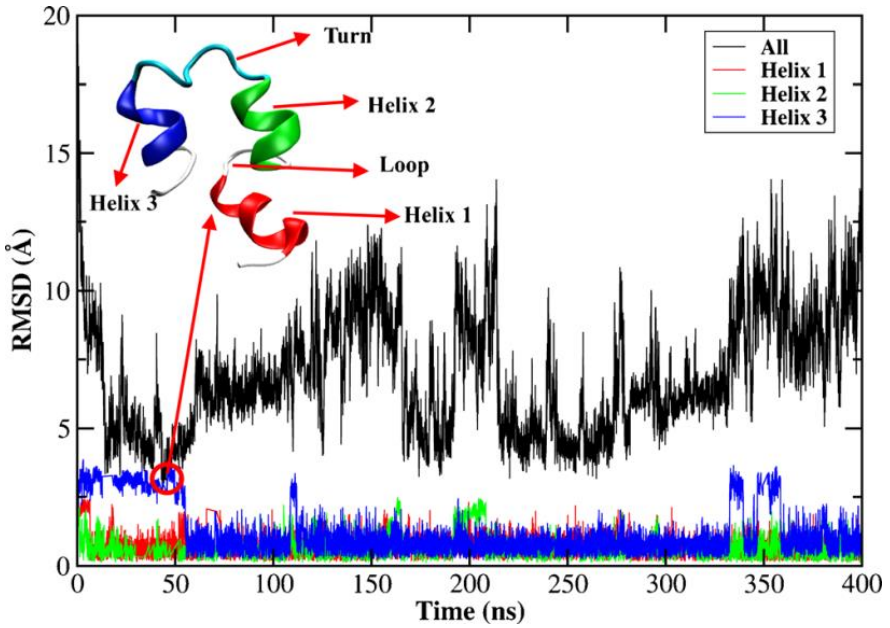


# MM parameters

- 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

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

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

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

# Verlet algorithm

Calculation of $p_i$ and $r_i$ at $\delta t$ time steps

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

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

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

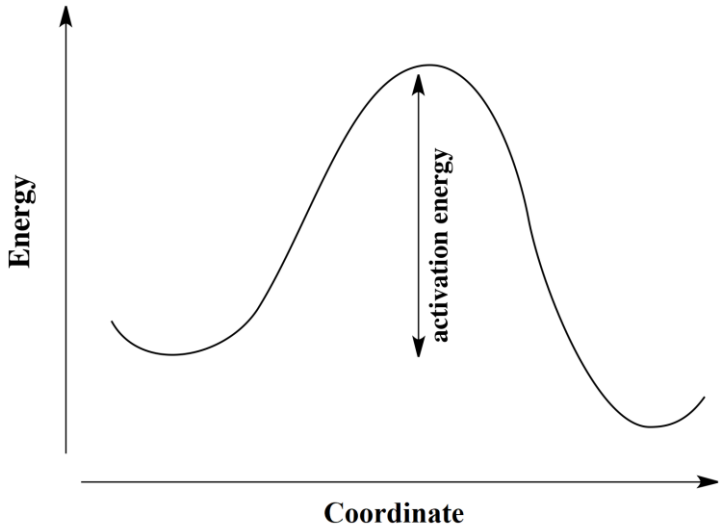
$$p_i(t+\delta t) = p_i\left(t+\frac{1}{2}\delta t\right) + \frac{1}{2}\delta t \cdot f_i\,(t+\delta t)$$

Typical $\delta t$ for simulation of biochemical systems: 1-4 fs

1 μs MD ~ $10^9$ evaluations

# 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



# Free energy - Sampling

$$F = -kTln\left[h^{-3N}\iint exp\left(-\frac{E(r,p)}{kT}\right)dpdr\right]$$

(1) Formula for free energy

Free energy calculation with MD sampling is problematic

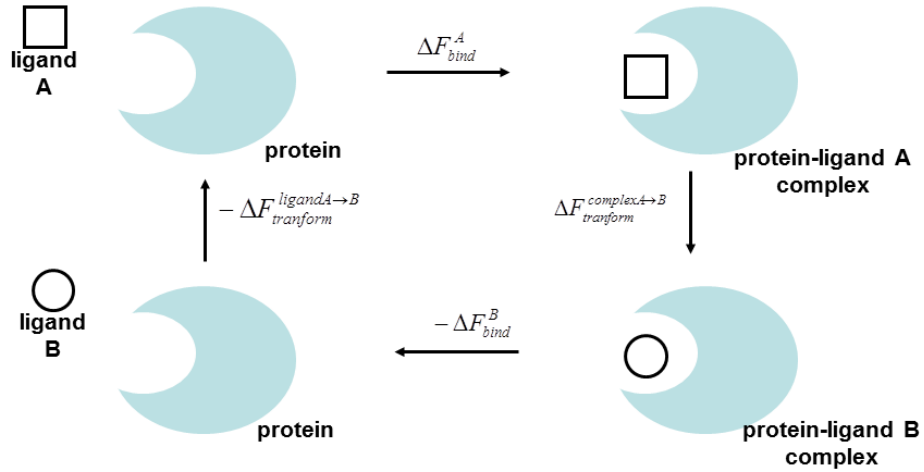
phase space incomplete in (1)
positive integral
*ln* function increases monotonically
negative contribution missing
F overestimated

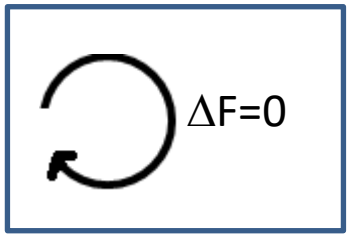
# MD - Sampling

| 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 difference

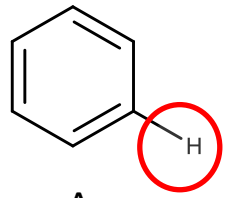
- Sampling issue hampers the calculation of F and $\Delta F=F_{Bound}-F_{Free}$
- Special techniques for calculating $\Delta F=F_B-F_A$ (A similar to B) for similar systems
  - Free energy perturbation, Thermodynamic integration
- Thermodynamic cycle: binding free energy difference ($\Delta\Delta F$)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



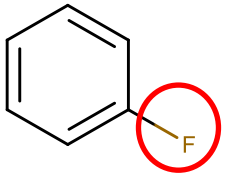

- $\Delta\Delta F = \Delta F^{A}_{bind} - \Delta F^{B}_{bind} = \Delta F^{complexA\rightarrow B}_{transform} - \Delta F^{ligandA\rightarrow B}_{transform}$
- „alchemical" transformations: $\Delta F^{complexA\rightarrow B}_{transform}$ and $\Delta F^{ligandA\rightarrow B}_{transform}$
  - 2 transformations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

# Alchemical transformation- coupling parameter




A

H(r,p,λ=0)




B

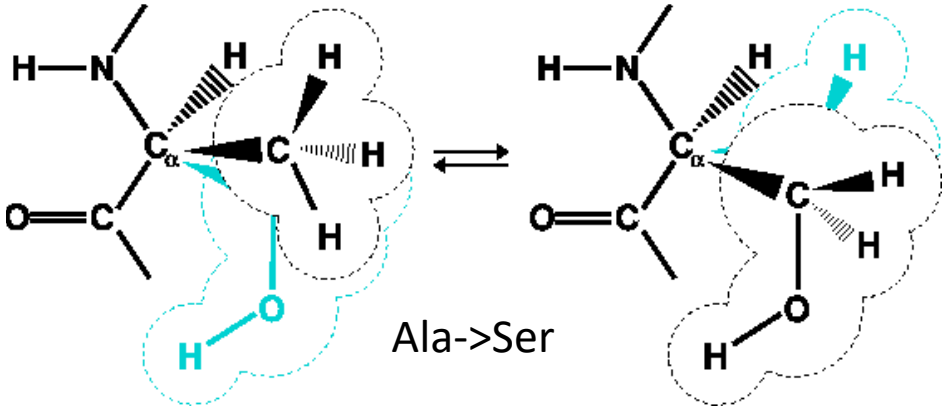
H(r,p,λ=1)

$H_\lambda = H(r,p,\lambda) = (1-\lambda)\, H_A + \lambda H_B$

$\lambda$ – coupling parameter

$H_\lambda$ may be other function of $\lambda$

$H_A$, $H_B$ may depend on $\lambda$

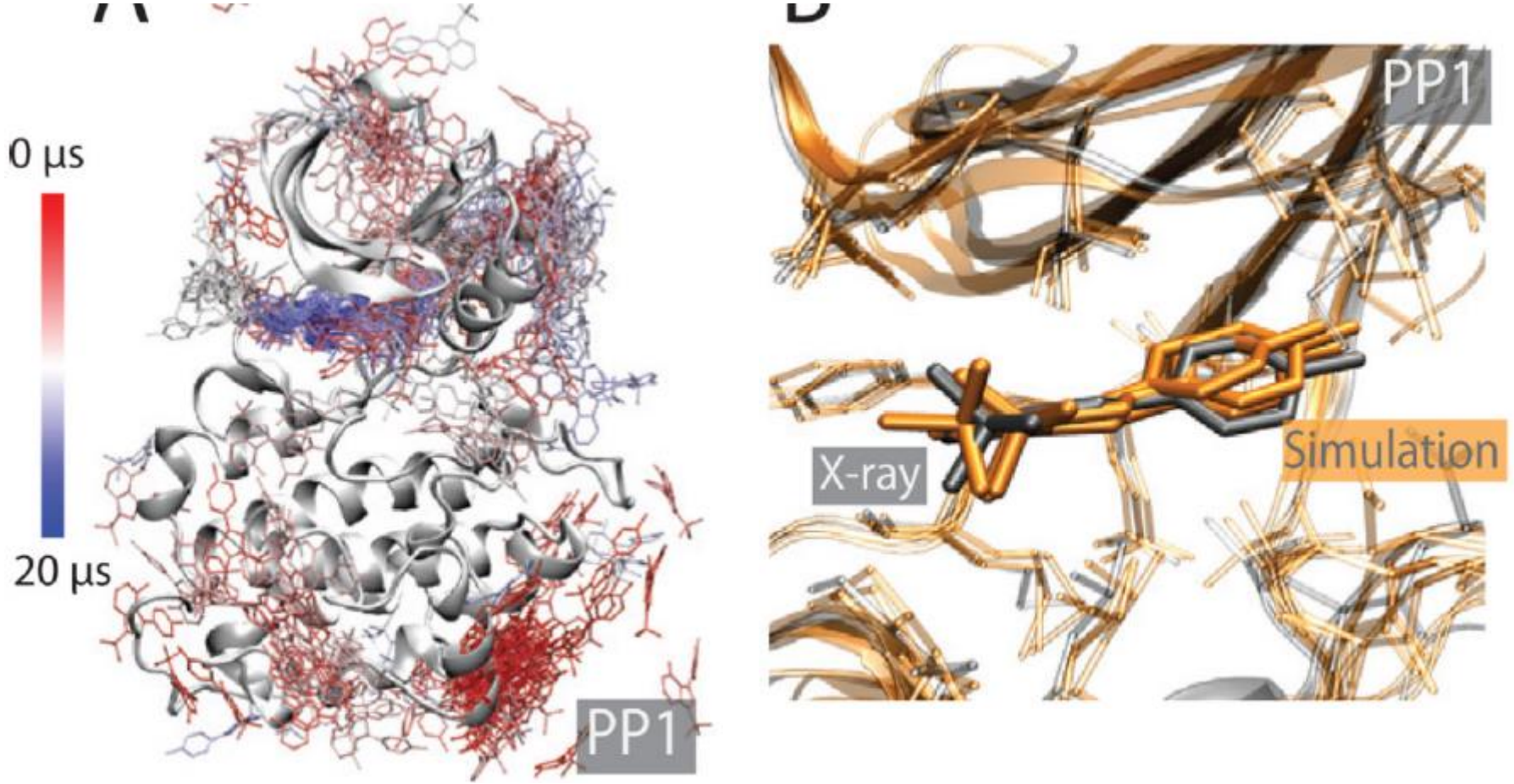



http://www.ks.uiuc.edu/Research/namd/2.6/ug/node36.html

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

# Selected applications

# process of ligand binding

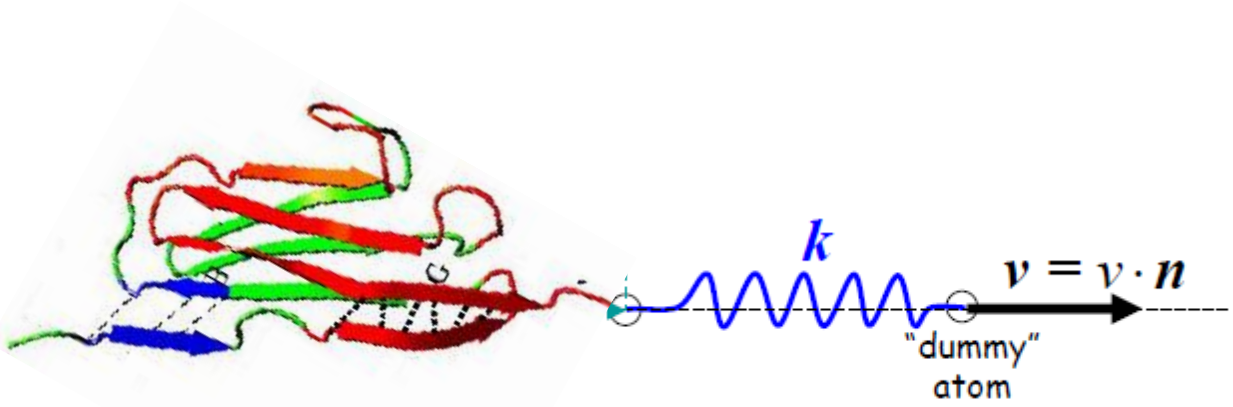



PP1 molecule finds the binding site of Src kinase in a 15μs simulation

JACS 2011 133 9181

# Steered MD

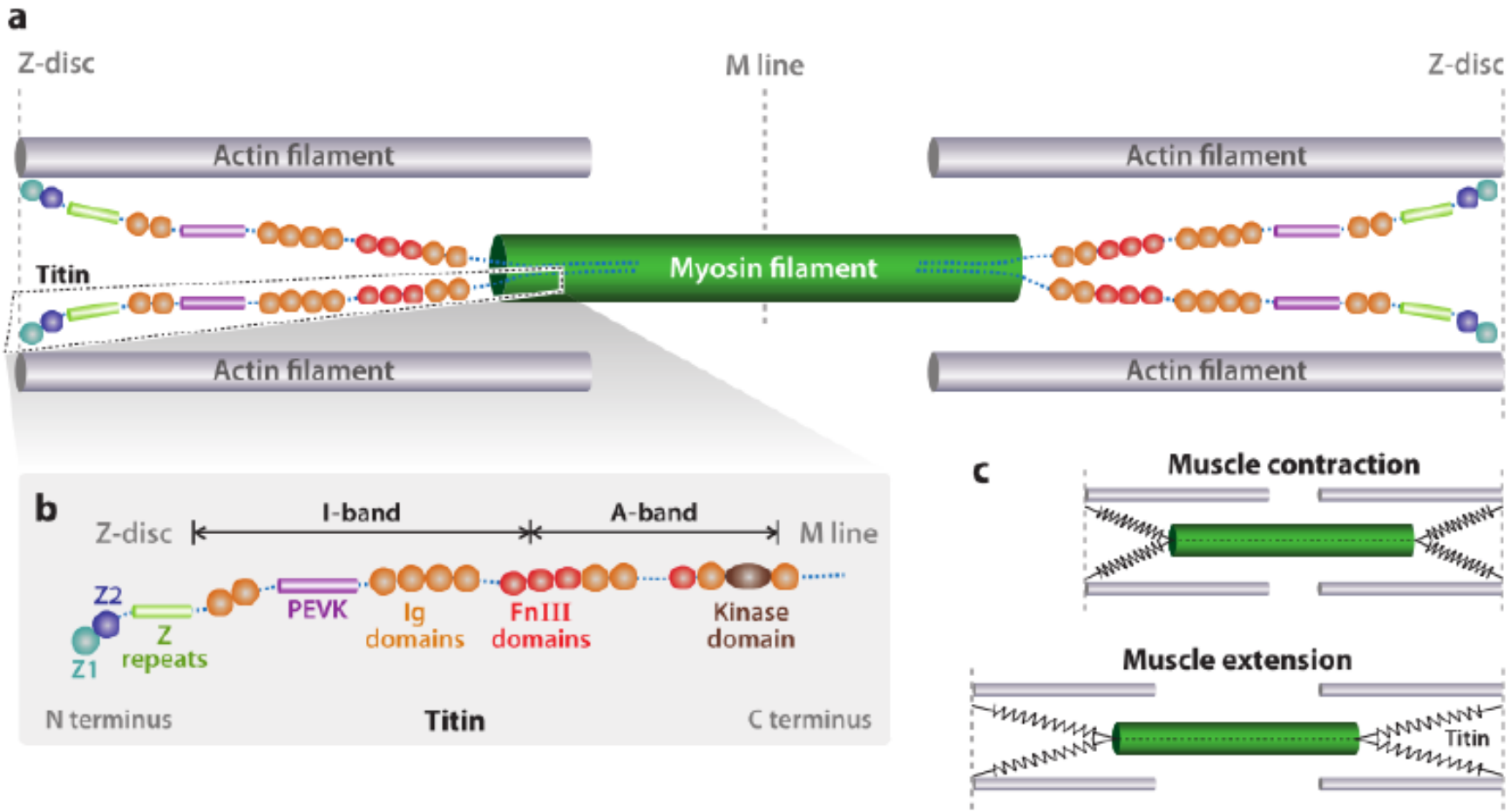
- Constant velocity pulling




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

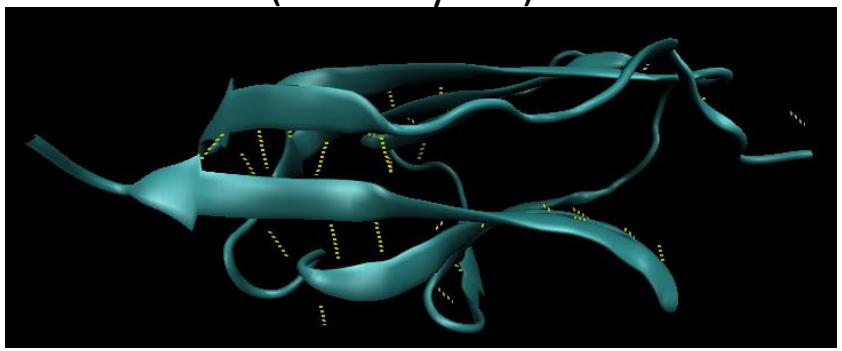
- Constant force pulling

# Titin structure and function



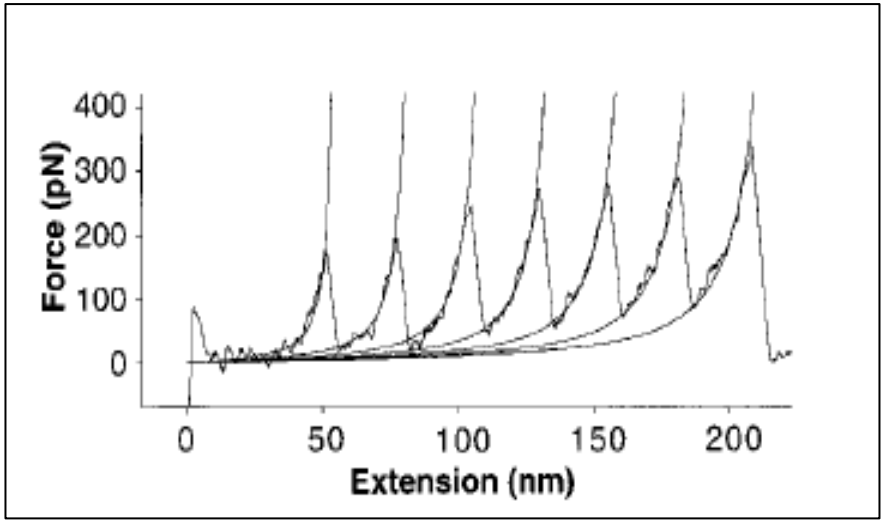
# Unfolding and force

titin I91 (formerly I27) domain



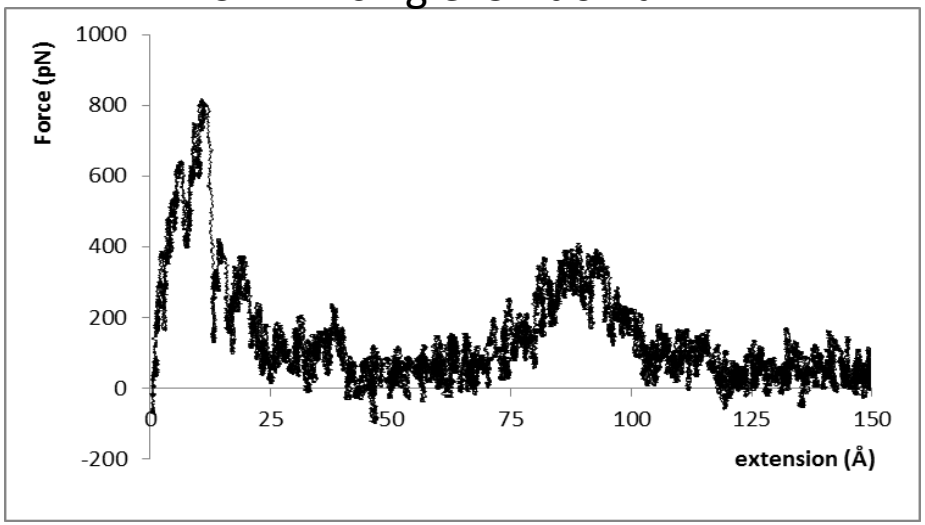
## Constant velocity pulling

AFM – linear chain of I91 domains

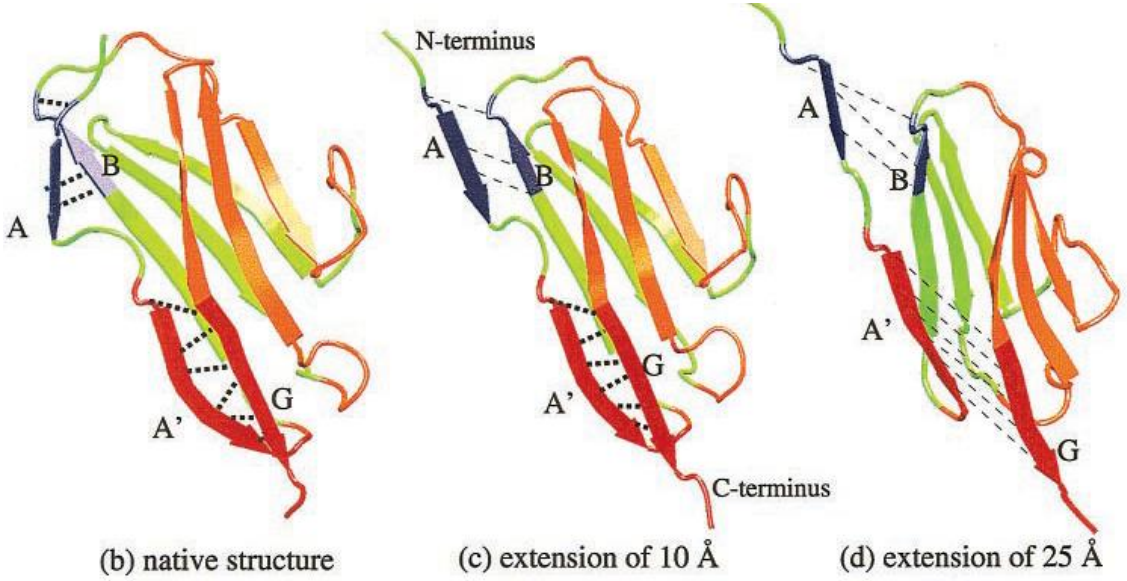



Rief et al. (1997 Science 276 1109)

SMD – single I91 domain

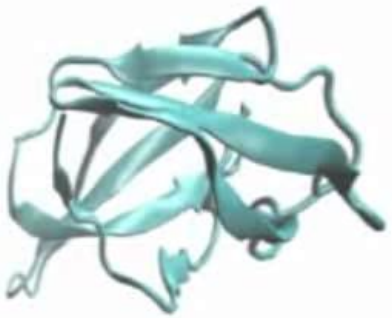


# Unfolding and structure

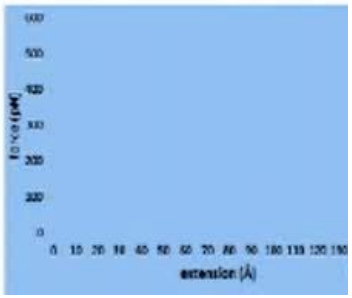



Lu, H., and Schulten, K. (2000).. Biophys. J. 79, 51–65

# Titin Fn domain unfolding mechanism



**A78**
**v=1Å/ns**



# Example – binding to lysozyme

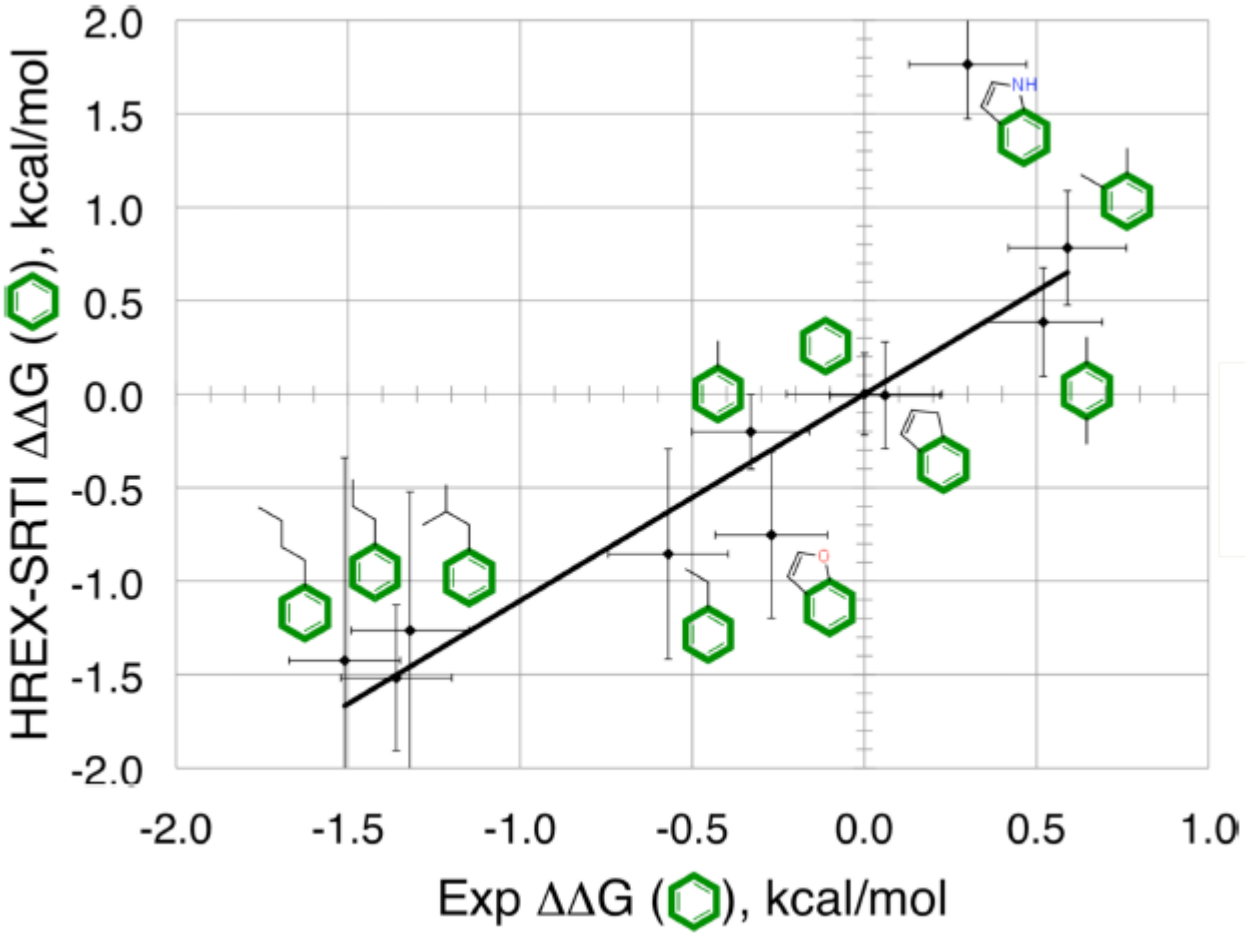



Figure 6. Comparison of the HREX-SRTI relative binding free energy predictions to experiment

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

# Example – neuraminidase inhibitors 1

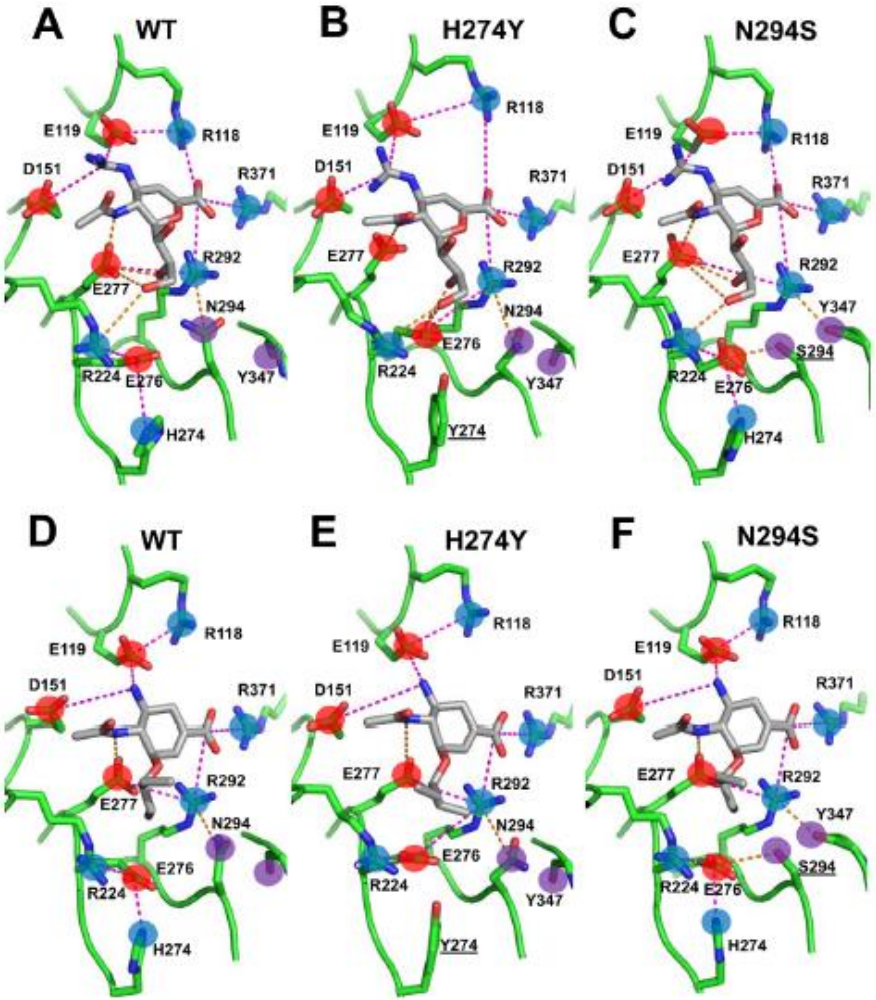



**Figure 2. Representative structures for zanamivir (A, B and C) and oseltamivir (D, E and F) bound to WT and mutant NAs from the SRSM/HREX simulations.** Salt-bridges and hydrogen bonds are depicted as magenta and orange dashed lines, respectively. Positively charged, negatively charged, and uncharged polar groups are noted as blue, red, and purple cirdes, respectively, and residues of interest are labeled. Mutated residues are underlined.
doi:10.1371/journal.pcbi.1002665.g002

PLOS Comp Biol 2012, 8 e1002665

# Example – neuraminidase inhibitors 2

**Table 1.** Comparison of experimental $\Delta\Delta G$ in oseltamivir and zanamivir for three NA mutations with estimates obtained using different computational approaches.

| Method | H274Y | | N294S | | Y252H | | RMSE |
|---|---|---|---|---|---|---|---|
| | $\Delta\Delta G$, kcal/mol | | $\Delta\Delta G$, kcal/mol | | $\Delta\Delta G$, 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

# Example – FKBP12-ligand

Standard binding free energy
Double decoupling
FEP
Free energy components

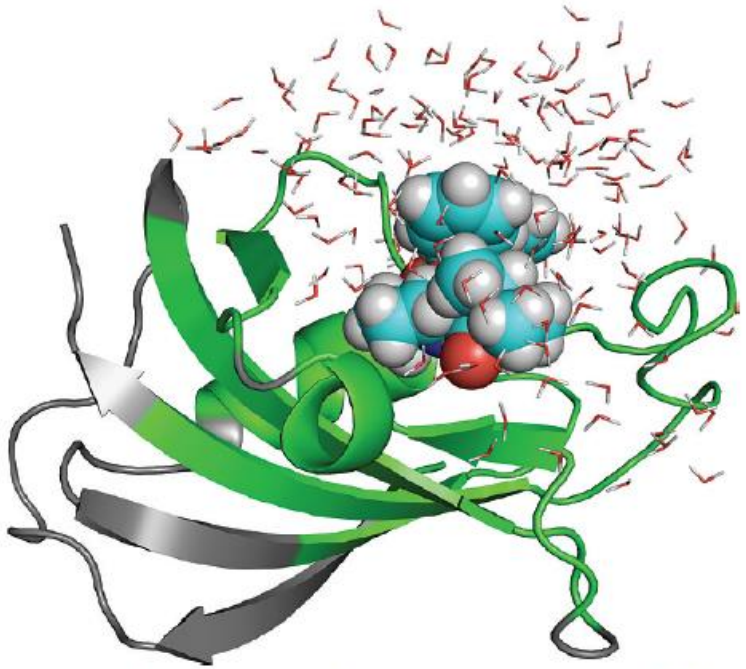


**Figure 4.** FKBP12 bound with ligand #8 studied previously.[40,42] The gray parts are treated as a mean-field approximation with generalized solvent boundary potential.[86] See ref 42 for computational details.

| $\Delta\Delta G_{rep}$ | $\Delta\Delta G_{dis}$ | $\Delta\Delta G_{elec}$ | $\Delta\Delta G_{c}$ | $\Delta\Delta G_{t}^{\circ}$ | $\Delta\Delta G_{r}$ | $\Delta G_{bind}^{\circ}$ | exptl |
|---|---|---|---|---|---|---|---|
| −1.1 | −21.1 | −3.7 | 6.9 | 3.4 | 5.4 | −10.2 | −10.9 |

# MD scope and limitation

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