

Molecular dynamics simulations

Methods in Molecular Biophysics, Spring 2010

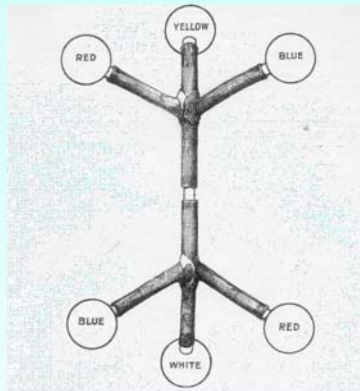
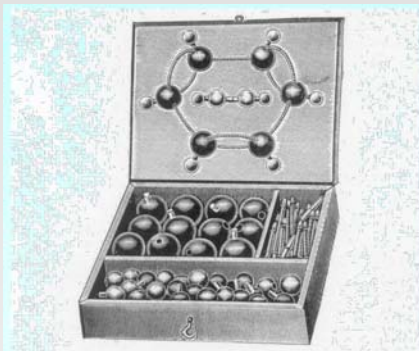
Basics of molecular mechanics and dynamics

Statistical mechanics of liquids

Basic ideas of continuum solvation

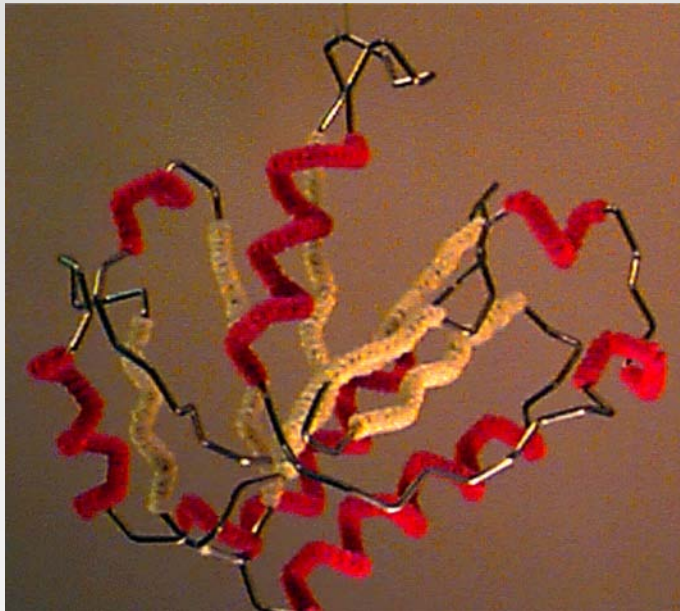
The MM/PBSA model

1901 (and earlier?) ball and stick models

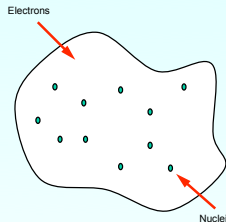


Baird & Tatlock 1901

1950s: wire models of proteins

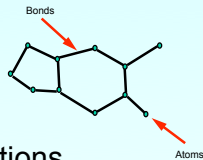


- separate nuclei and electrons
- polarisation, electron transfer and correlation
- can specify electronic state
- can calculate formation energies
- can do chemistry (bond breaking and making)
- variationally bound
- computationally expensive
- typically ~10-100 atoms
- dynamics ~1 ps



QM MOLECULE

- no explicit electrons, net atomic charges
- no polarisation, electron transfer or correlation
- conformational energies for ground state
- no chemistry
- semi-empirical force fields
- not variationally bound
- solvent and counterion representations
- typically ~1000-100000 atoms
- dynamics up to ~100 ns



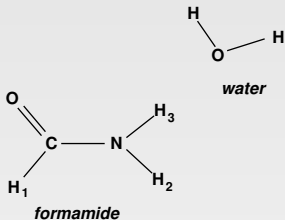
MM MOLECULE

Some force field assumptions

- 1 **Born-Oppenheimer approximation** (separate nuclear and electronic motion)
- 2 **Additivity** (separable energy terms)
- 3 **Transferability** (look at different conformations, different molecules)
- 4 **Empirical** (choose functional forms and parameters based on experiment)

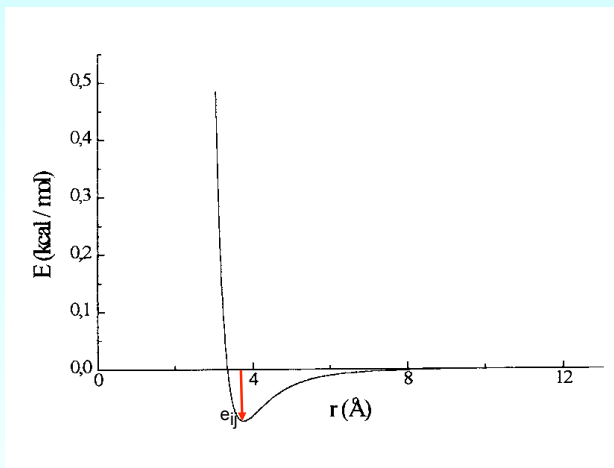
What does a force field look like?

$$U = \sum_{\text{bonds}} K_b(b - b_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{impropers}} K_w w^2 \\ + \sum_{\text{torsions}} K_\phi \cos(n\phi) + \sum_{\text{nonbonded pairs}} \left\{ 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] + \frac{q_i q_j}{r} \right\}$$

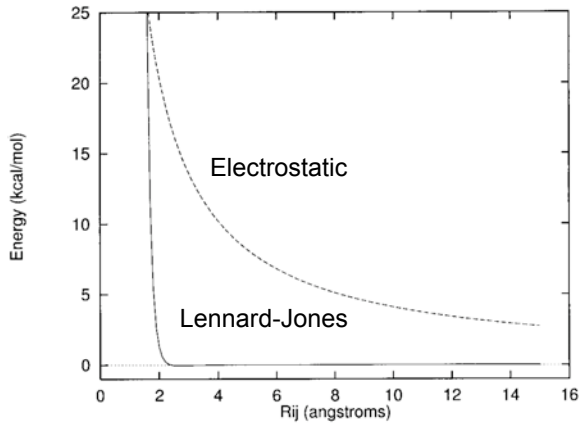


Amber tradition for parameters:

- top line from X-ray structures, quantum calculations, vibrational spectroscopy
- partial charges from fits to electrostatic potential from HF/6-31G*
- van der Waals ϵ , σ from neat liquids (*not* water/solute simulations)
- torsional parameters from quantum calculations



Lennard-Jones energy curve



Distance dependence

H	H bonded to nitrogen atoms
HC	H aliph. bond. to C without electrwd.group
H1	H aliph. bond. to C with 1 electrwd. group
H2	H aliph. bond. to C with 2 electrwd.groups
H3	H aliph. bond. to C with 3 eletrwd.groups
HA	H arom. bond. to C without elctrwd. groups
H4	H arom. bond. to C with 1 electrwd. group
H5	H arom. bond. to C with 2 electrwd. groups
HO	hydroxyl group
HS	hydrogen bonded to sulphur
HW	H in TIP3P water
HP	H bonded to C next to positively charged gr

AMBER parm94 H atom types

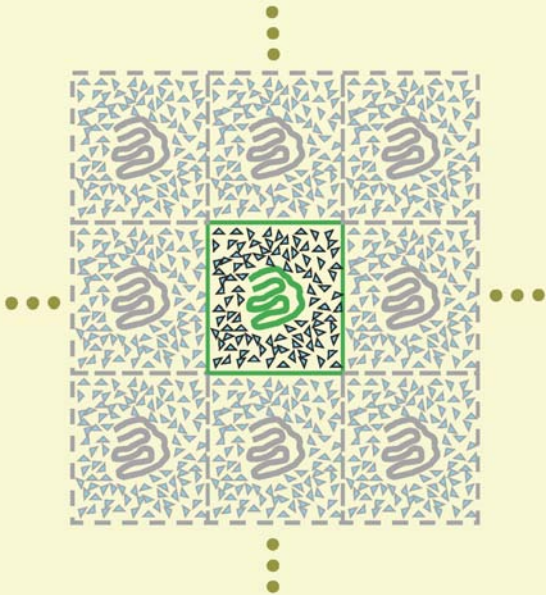
C	sp2 C carbonyl group
CA	sp2 C pure aromatic (benzene)
CB	sp2 aromatic C, 5&6 membered ring junction
CC	sp2 aromatic C, 5 memb. ring HIS
CK	sp2 C 5 memb.ring in purines
CM	sp2 C pyrimidines in pos. 5 & 6
CN	sp2 C aromatic 5&6 memb.ring junct.(TRP)
CQ	sp2 C in 5 mem.ring of purines between 2 N
CR	sp2 arom as CQ but in HIS
CT	sp3 aliphatic C
CV	sp2 arom. 5 memb.ring w/1 N and 1 H (HIS)
CW	sp2 arom. 5 memb.ring w/1 N-H and 1 H (HIS)
C*	sp2 arom. 5 memb.ring w/1 subst. (TRP)

AMBER parm94 C atom types

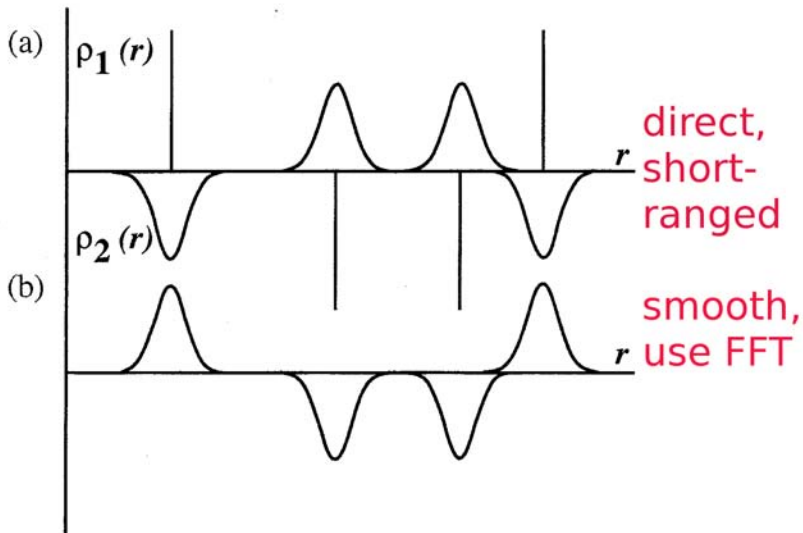
Force fields in Amber

- **ff94**: widely used (“Cornell et al.”), pretty good nucleic acid, too much α -helix for proteins
- **ff99**: major recalibration by Junmei Wang and others; basis for most current Amber ff’s
- **ff99SB**: recalibration of backbone potentials for proteins by Carlos Simmerling (“SB”)
- **ff02r1**: polarizable extension for ff99
- **ff03**: new charge model (Yong Duan) + backbone torsions for proteins
- **ff03ua**: united atom extension

Periodic boundary conditions

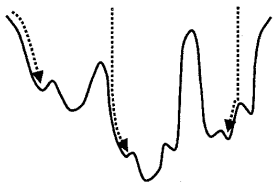


Basics of the Ewald approach

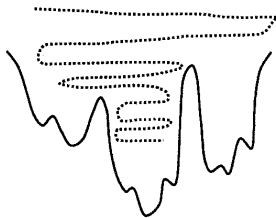


Minimization and simulated annealing

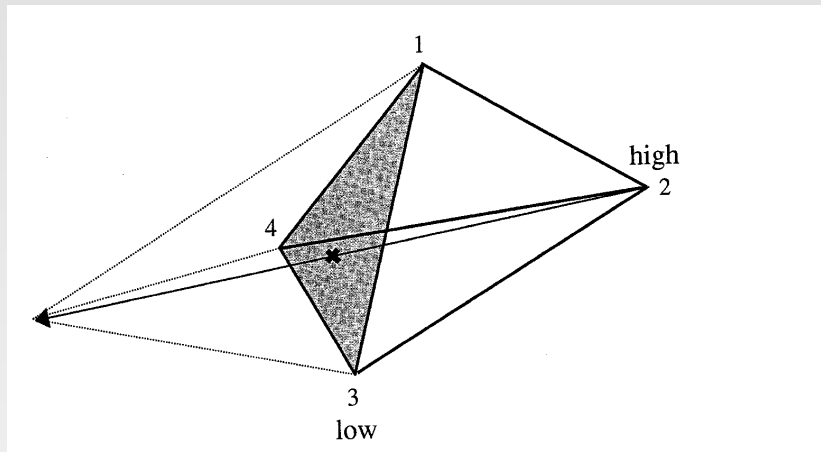
(a)



(b)



The Simplex algorithm



See Numerical Recipes, by Press, Teukolsky, Vetterling and Flannery

Molecular dynamics algorithms

$$x(t+h) = x(t) + v(t)h + \frac{1}{2}a(t)h^2 + \frac{1}{6}\frac{d^3x}{dt^3}h^3 + O(h^4)$$

$$x(t-h) = x(t) - v(t)h + \frac{1}{2}a(t)h^2 - \frac{1}{6}\frac{d^3x}{dt^3}h^3 + O(h^4)$$

$$x(t+h) = 2x(t) - x(t-h) + a(t)h^2 + O(h^4) \quad (1)$$

$$x(t+h) - x(t) = x(t) - x(t-h) + a(t)h^2 + O(h^4)$$

$$v(t + \frac{1}{2}h) = v(t - \frac{1}{2}h) + a(t)h + O(h^3) \quad (2)$$

$$x(t+h) = x(t) + v(t + \frac{1}{2}h)h + O(h^4) \quad (3)$$

Eq. (1) is the original Verlet propagation algorithm; Eqs. 2 and 3 are the “leap-frog” version of that. Remember that $a = d^2x/dt^2 = F/m = (\partial V/\partial x)/m$. See pp. 42-47 in Becker & Watanabe.

Regulating temperature

“Temperature” is a measure of the mean kinetic energy. The instantaneous KE is

$$T(t) = \frac{1}{k_B N_{dof}} \sum_i^{N_{dof}} m_i v_i^2$$

(cf. classical rule of thumb: “ $k_B T/2$ of energy for every squared degree of freedom in the Hamiltonian”)

Suppose the temperature is not what you want. At each step, you could scale the velocities by:

$$\lambda = \left[1 + \frac{h}{2\tau} \left(\frac{T_0}{T(t)} - 1 \right) \right]^{1/2}$$

This is the “Berendsen” or “weak-coupling” formula, that has a minimal disruption on Newton’s equations of motion. But it does not guarantee a canonical distribution of positions and velocities. See Morishita, J. Chem. Phys. 113:2976, 2000; and Mudi and Chakravarty, Mol. Phys. 102:681, 2004.

Consider the stochastic, Langevin equation:

$$d\mathbf{v}/dt = -\zeta\mathbf{v} + \mathbf{A}(t)$$

By Stokes' law, the friction coefficient is related to the viscosity of the environment: $\zeta = 6\pi a\eta/m$. At long times, we want this system to go to equilibrium at a temperature T , which is a Maxwell-Boltzmann distribution:

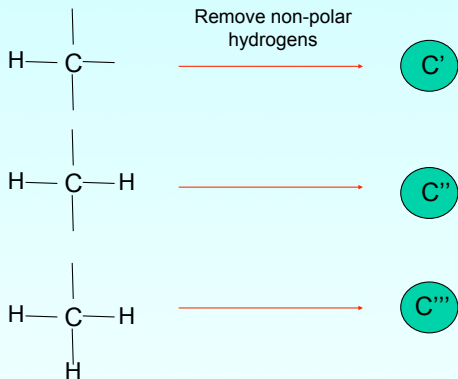
$$W(\mathbf{v}, t; \mathbf{v}_0) \sim \exp[-m\mathbf{v}^2/2k_B T]$$

for every value of \mathbf{v}_0 . This places restraints on the properties of the stochastic force $\mathbf{A}(t)$. It can be shown that

$$\zeta = (\beta/m) \langle A^2 \rangle$$

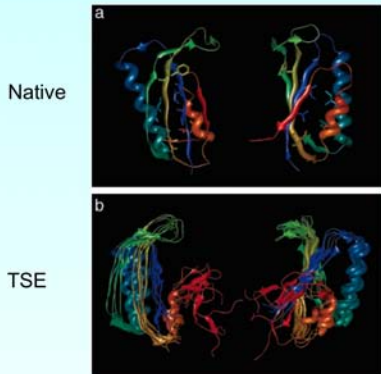
where we have assumed that $\langle A \rangle = 0$ and $\langle A(0).A(t) \rangle = \langle A^2 \rangle \delta(t)$.

More coarse-grained potentials



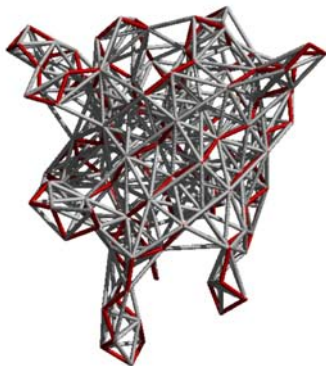
United atom approximation

Go model for protein folding



- square-well potential
- native contacts "+1"
- non-native contacts "-1"
- cannot represent frustration during folding

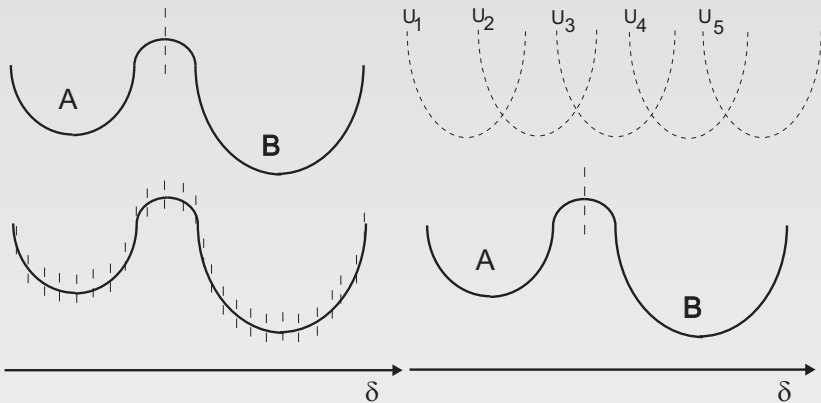
Gaussian network model



$$r_{ij}^n < R_{cut}$$

$$E_{ij} = k(r_{ij} - r_{ij}^n)^2$$

Getting free energies

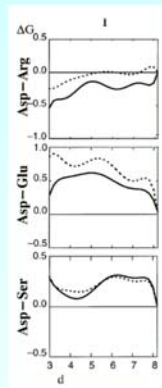


$$\Delta A = -k_B T \ln \frac{\rho(B)}{\rho(A)}$$

$$W = -k_B T \ln \rho(\delta) \quad (4)$$

Knowledge-based potentials

- Start from set of known protein structures
- Assume energy can be decomposed into residue pair interactions
- Assume that frequency of interactions within the ensemble \equiv frequency of interactions within the equilibrium ensemble of a single protein
- Derive potential of mean force for residue pairs from observed occurrence probabilities
- Knowledge-based potentials are used in both threading and folding



- - - small proteins
- large proteins

$$\rho(\delta) = \frac{\int \exp(-\beta U) d\Sigma}{\int \exp(-\beta U) d\delta d\Sigma} \quad (5)$$

Here $\beta = 1/k_B T$ and $d\Sigma$ represents an integration over all remaining degrees of freedom except δ . Now add a biasing potential $U^*(\delta)$ which depends only upon δ :

$$\begin{aligned} \rho^*(\delta) &= \exp[-\beta U^*(\delta)] \frac{\int \exp(-\beta U) d\Sigma}{\int \exp(-\beta [U + U^*]) d\delta d\Sigma} \\ &= \rho(\delta) \exp[-\beta U^*(\delta)] / \langle \exp(-\beta U^*) \rangle \end{aligned} \quad (6)$$

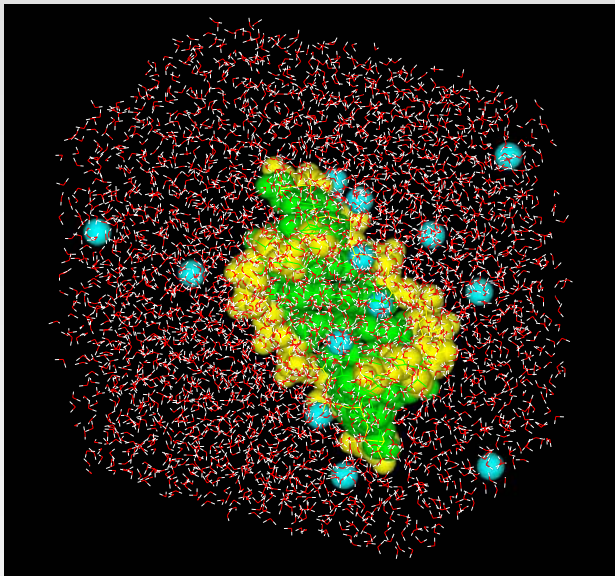
$$\langle \exp(-\beta U^*) \rangle = \frac{\int \exp(-\beta U^*) \exp(-\beta U) d\delta d\Sigma}{\int \exp(-\beta U) d\delta d\Sigma} \quad (7)$$

Taking logarithms, the potential of mean force in the presence of the umbrella potential, W^* , is related to that in an unbiased simulation by:

$$W^*(\delta) = W(\delta) + U^*(\delta) - C' \quad (8)$$

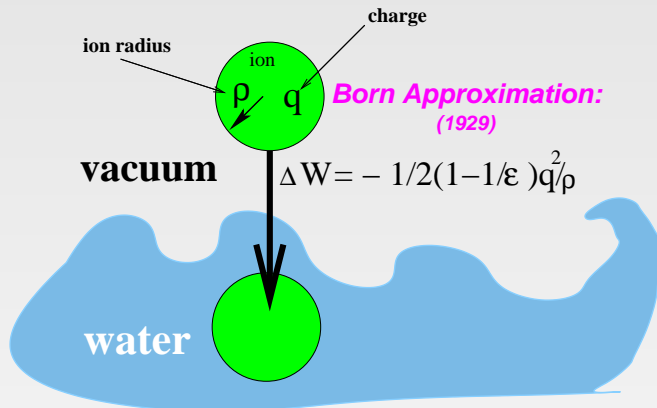
where $C' = -k_B T \ln \langle \exp(-\beta U^*) \rangle$ is a constant independent of δ .

Example of explicit solvation setup



Basic ideas of continuum solvent models

- Tomasi & Persico, *Chem Rev.* **94**, 2027 (1994)
- Simonson, *Rep. Prog. Phys.* **66**, 737 (2003)
- Bashford & Case, *Annu. Rev. Phys. Chem.* **51**, 129 (2000)
- Gallicchio & Levy, *J. Comput. Chem.* **25**, 479 (2004)



Conductor-like Screening Model

Klammt
JPC 99
2224 (1995)



Ponder + Case pp 65-67

$$E = E_{\text{gas}} + \int z \frac{1}{r_i - r_j} \rho + \frac{1}{2} \int \rho \frac{1}{r_i - r_j} \rho'$$

$\frac{\epsilon}{\epsilon - 1}$

$$= E_{\text{gas}} + z B \rho + \frac{1}{2} \rho A \rho$$

$$\frac{\partial E}{\partial \rho} = 0 \Rightarrow A \rho = -B z \quad \text{or} \quad \boxed{\rho = -A^{-1} B z}$$

molecule-solvent
interaction:

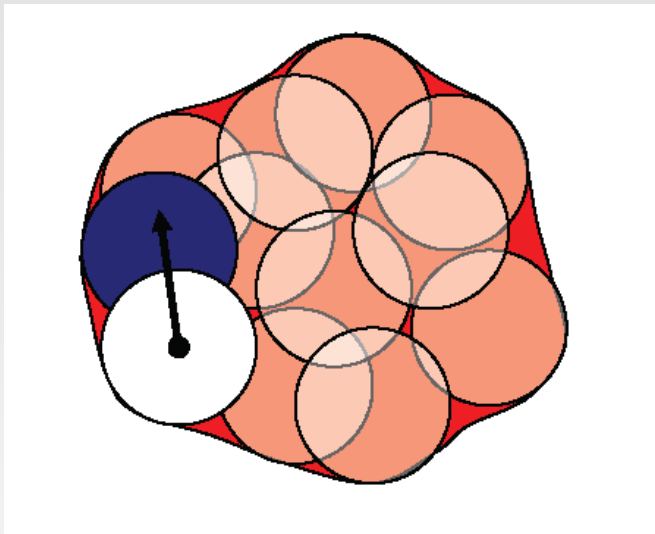
$$-z B A^{-1} B z = -z \phi^{\text{RF}}$$

solvent-solvent
interaction:

$$\frac{1}{2} z B A^{-1} A A^{-1} B z = \frac{1}{2} z B \bar{A} B z$$

Defining the continuum solvent model

Simplest model has “high” ϵ_{ext} outside (white) and “low” ϵ_{in} where solvent is excluded:



Generalized Born model

The solvation energy can be computed by quadrature if one adopts the **Coulomb field approximation**:

$$W = \frac{1}{8\pi} \int \mathbf{E} \cdot \mathbf{D} dV = \frac{1}{8\pi} \left[\int_{in} \frac{q^2}{\epsilon_{in} r^4} dV + \int_{ext} \frac{q^2}{\epsilon_{ext} r^4} dV \right]$$

$$\Delta G = W(\epsilon_{ext} = 80) - W(\epsilon_{ext} = 1)$$

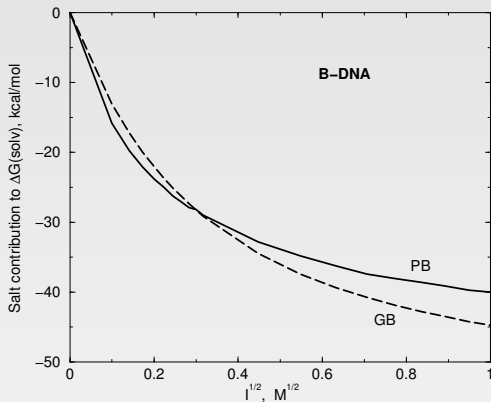
$$\Delta G_{GB} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon_{ext}} \right) \frac{q^2}{R_{eff}}; \quad \text{or} \quad -\frac{1}{2} \left(1 - \frac{1}{\epsilon_{ext}} \right) \frac{q_i q_j}{f^{GB}(R_{eff}^i, R_{eff}^j, r_{ij})}$$

$$R_{eff}^{-1} = \frac{1}{4\pi} \int_{ext} r^{-4} dV$$

Bashford & Case, *Annu. Rev. Phys. Chem.* **51**, 129 (2000)

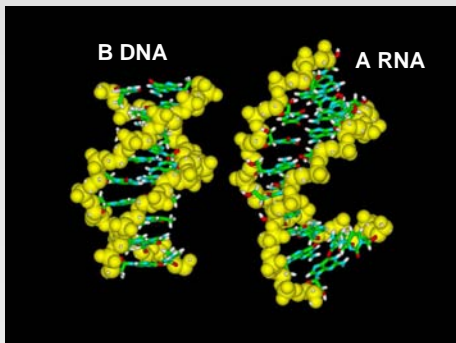
Effects of added salt

$$\left(1 - \frac{1}{\epsilon}\right) \rightarrow \left(1 - \frac{e^{-\kappa r^{GB}(d_{ij})}}{\epsilon}\right)$$



Srinivasan, Trevathan, Beroza, Case, *Theor. Chem. Accts.* **101**, 426 (1999)

B-A energy differences for r,d(CCAACGTTGG)₂



	DNA	RNA
Couomb	-293.0	-266.9
PB	286.6	240.2
GB	288.1	242.2
vdW	-7.7	18.7
bad	-7.0	17.6
$-T\Delta S$	2.9	0.5
total	-21.0	9.8
0.1M salt	5.2	3.4
1.0M salt	6.0	3.9

Srinivasan, Cheatham, Kollman, Case, JACS 120, 9401 (1998)

Thermodynamic integration: computational alchemy

Now suppose that V (and hence Q and A) are parameterized by λ :
 $V \rightarrow V(\lambda)$. Then, since $A = -kT \ln Q$:

$$\frac{\partial A(\lambda)}{\partial \lambda} = -kT \int \frac{\partial}{\partial \lambda} e^{-\beta V(\lambda)} dq / Q = \frac{1}{Q} \int \left(\frac{\partial V}{\partial \lambda} \right) e^{-\beta V(\lambda)} dq = \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda}$$

The total change in A on going from $\lambda = 0$ to $\lambda = 1$ is:

$$\Delta A = A(1) - A(0) = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (9)$$

This is called thermodynamic integration, and is a fundamental connection between macroscopic free energies, and microscopic simulations. The integral over λ can be done by quadrature, and the Boltzmann averages $\langle \partial V / \partial \lambda \rangle_{\lambda}$ can be carried out by molecular dynamics or Monte Carlo procedures.

Thermodynamic integration: linear mixing

Consider the special case of **linear mixing**, where

$$V(\lambda) = (1 - \lambda) V_0 + \lambda V_1$$

Then $\partial V / \partial \lambda = V_1 - V_0 \equiv \Delta V$ (often called the **energy gap**), and

$$\Delta A = \int_0^1 \langle \Delta V \rangle_\lambda d\lambda \quad (10)$$

The simplest numerical approximation to the λ integral is just to evaluate the integrand at the midpoint, so that $\Delta A = \langle \Delta V \rangle_{1/2}$. This says that the free energy difference is approximately equal to the average potential energy difference, evaluated for a (hypothetical) state half-way between $\lambda = 0$ and $\lambda = 1$.

It is often convenient for other purposes to perform simulations only at the endpoints. In this case, a convenient formula would be:

$$\Delta A \simeq \frac{1}{2} \langle \Delta V \rangle_0 + \frac{1}{2} \langle \Delta V \rangle_1$$

And more elaborate formulas (e.g. from Gaussian integration) are feasible (and often used). See Hummer & Szabo, *J. Chem. Phys.* **105**, 2004 (1996) for a fuller discussion.

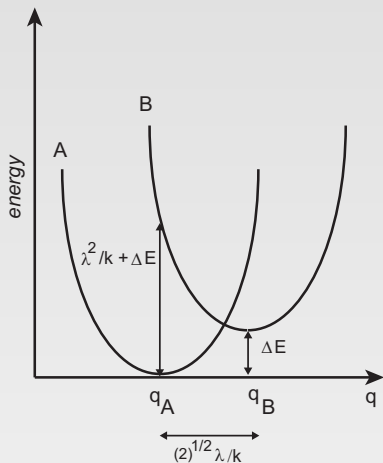
Free energy perturbation theory

Here is an (initially) completely different approach:

$$\begin{aligned}\Delta A &= -kT \ln \left(\frac{Q_1}{Q_0} \right) \\ &= -kT \ln \left(\frac{\int \exp(-\beta E_1) \exp(\beta E_0) \exp(-\beta E_0) dq}{\int \exp(-\beta E_0) dq} \right) \\ &= -kT \ln \left(\frac{1}{Q_0} \int \exp(-\beta [E_1 - E_0]) \exp(-\beta E_0) dq \right) \\ &= -kT \ln \langle \exp(-[E_1 - E_0]/kT) \rangle_0 \\ &= -kT \ln \langle \exp(-[E_0 - E_1]/kT) \rangle_1\end{aligned}$$

This is generally called “**perturbation theory**”, and involves averaging the exponential of the energy gap, rather than the energy gap itself.

A simple model: "Marcus theory"



$$V_A(q) = \frac{1}{2}k(q - q_A)^2$$

$$V_B(q) = \frac{1}{2}k(q - q_B)^2$$

$$\Delta V(q) = \sqrt{2}\lambda(q - q_A) + \frac{\lambda^2}{k} + \Delta E$$

Marcus theory thermodynamic integration

$$\langle V_B - V_A \rangle_A = Q_A^{-1} \int \left[\sqrt{2}\lambda(q - q_A) + \frac{\lambda^2}{k} + \Delta E \right] e^{-\beta V_A(q)} dq = \frac{\lambda^2}{k} + \Delta E$$

$$\langle V_B - V_A \rangle_B = -\frac{\lambda^2}{k} + \Delta E; \quad \Delta A \simeq \frac{1}{2} [\langle \Delta V \rangle_A + \langle \Delta V \rangle_B] = \Delta E$$

What is the distribution of ΔV in the V_A state?

$$\rho(\Delta V) = \rho(q) \left| \frac{dq}{d\Delta V} \right| \quad \text{where} \quad q(\Delta V) = \left(\frac{\lambda^2 + k\Delta E}{\sqrt{2k\lambda}} \right) - \frac{\Delta V}{\sqrt{2}\lambda}$$

$$\rho(\Delta V) \sim \frac{1}{\sqrt{2}\lambda} \exp \{ -\beta V_A [q(\Delta V)] \} \simeq \exp \left\{ -\frac{(\Delta V - \lambda^2/k - \Delta E)^2}{2\sigma^2} \right\} \quad \text{with} \quad \sigma^2 = \frac{2\lambda^2}{k\beta}$$

Hence, the **mean** of the distribution gives $\lambda^2/k + \Delta E$, and the **width** of the distribution gives λ^2/k (the “relaxation”); knowing both allows you to get ΔE and λ separately.

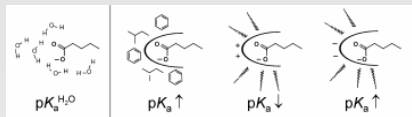
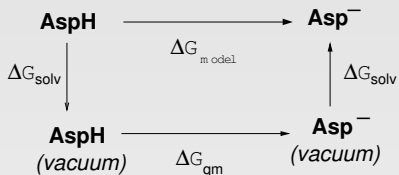
For **perturbation theory**:

$$\Delta A = -kT \ln \left\langle e^{-\beta \Delta V} \right\rangle_A = \Delta E$$

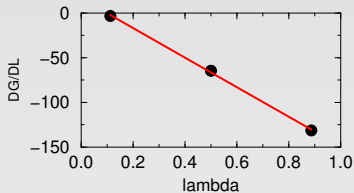
Application: pKa behavior in proteins



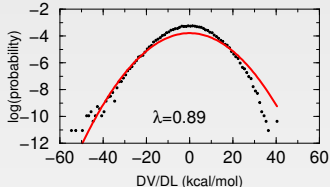
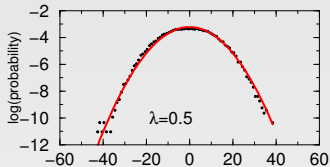
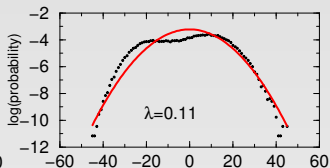
$$\Delta\Delta G = \Delta G_{\text{prot}} - \Delta G_{\text{model}}$$



Energy gap distributions

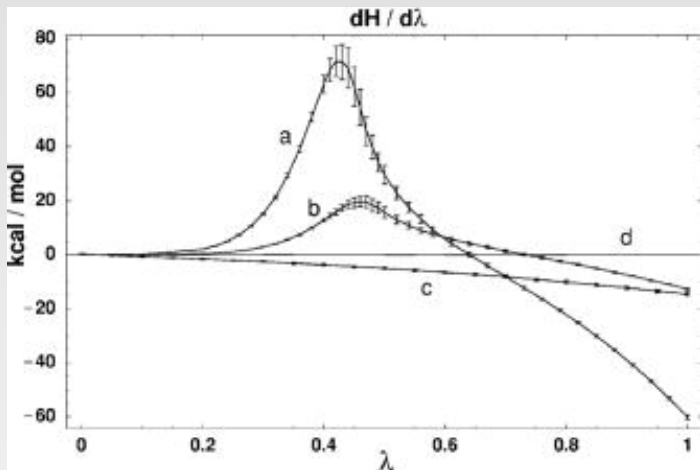


Lambda	DG/DL (kcal/mol)
0.11270	-3.1
0.50000	-64.5
0.88729	-131.4



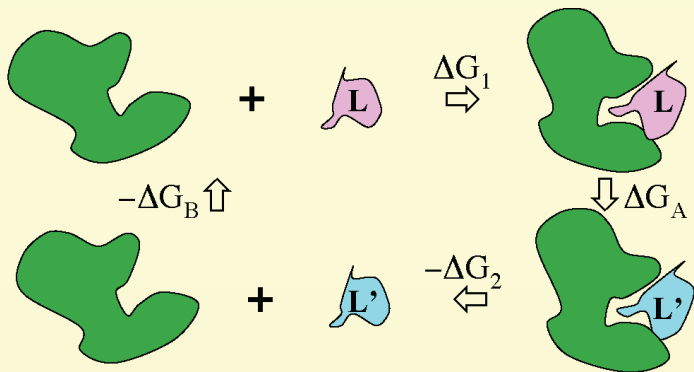
Simonson, Carlson, Case, *JACS* **126**:4167 (2004)

Not everything is linear!



Shirts, Pitera, Swope, Pande, *J. Chem. Phys.* **119**, 5740 (2003).

Thermodynamics cycles in ligand binding

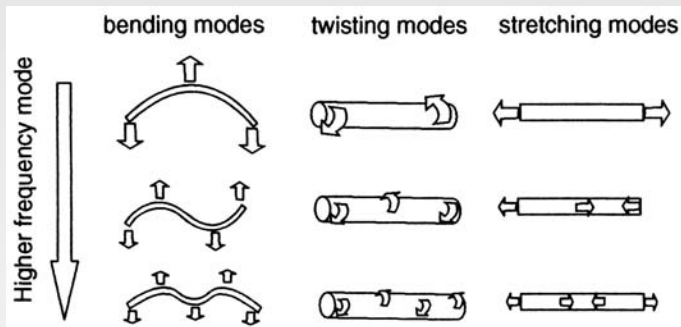


Duplex DNA: Comparison of atomic and elastic rod models

$$\text{Bending rigidity: } A = \frac{M(2\pi\nu_n)^2}{LP_n^4}; \quad a = A/(k_B T)$$

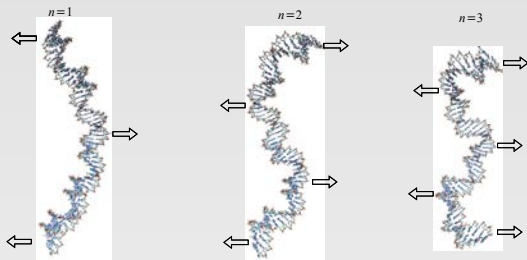
$$\text{Twisting rigidity: } C = \frac{I(2L\nu_n)^2}{n^2}$$

$$\text{Stretching rigidity: } Y = \frac{ML(2\nu_n)^2}{n^2}$$



Lord Rayleigh, *The Theory of Sound*, 1894

Bending rigidity for linear duplex DNA

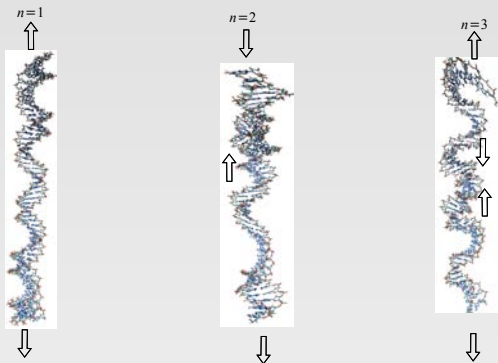


d(GACT) 60 base pairs			
n	GB frequencies		Analytical frequencies
1	0.114	0.116	0.100
2	0.294	0.295	0.275
3	0.522	0.532	0.539

$$A = 2.44 \cdot 10^{-19} \text{ erg.cm} \quad (2.26 \cdot 10^{-19} \text{ erg.cm}) \quad a = 594 \text{ \AA} \quad (550 \text{ \AA})$$

Bomble & Case, *Biopolymers*, **89**, 722 (2008)

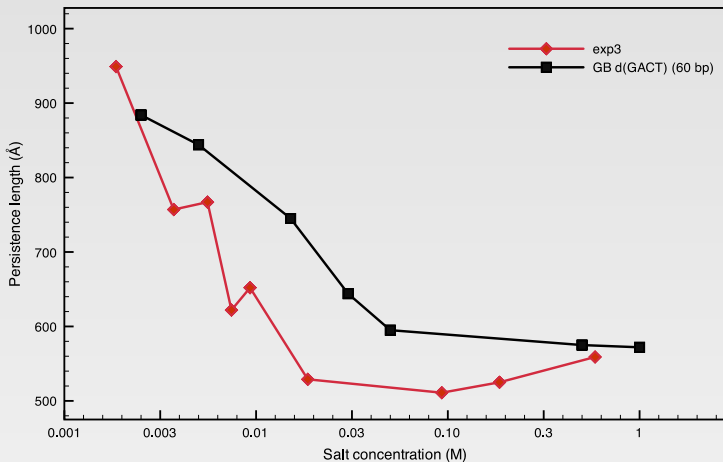
Stretching rigidity for linear duplex DNA



d(GACT) 60 base pairs		
n	GB frequencies	Analytical frequencies
1	0.664	0.619
2	1.289	1.237
3	1.807	1.846

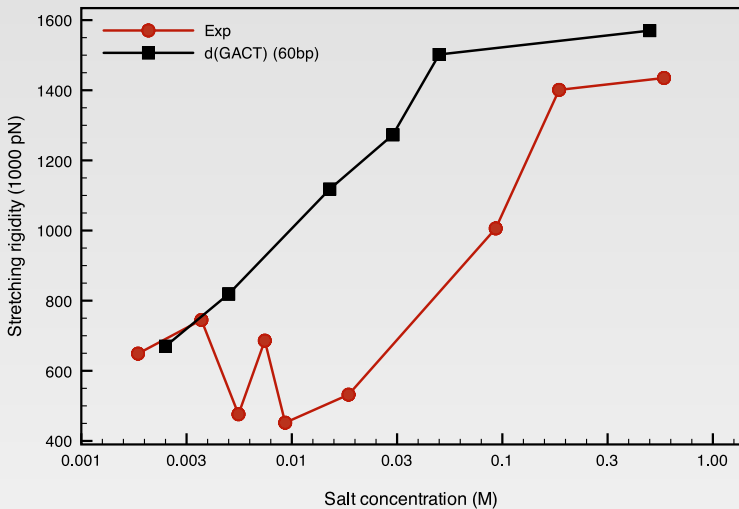
$$Y = 1502 \text{ pN} \quad (1000 - 1500 \text{ pN})$$

Salt dependence of bending



Baumann, Smith, Bloomfield, Bustamante, *PNAS* **94**, 6185 (1997)

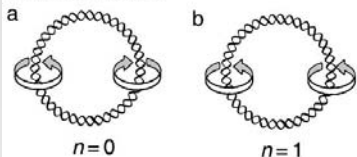
Salt dependence of stretching



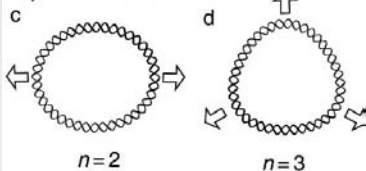
Baumann, Smith, Bloomfield, Bustamante, *PNAS* **94**, 6185 (1997)

Now consider circular DNA

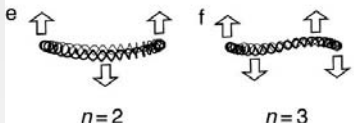
Torsional modes



In-plane modes



Out-of-plane modes



$$v_n = f(\Omega, R, \Delta Tw, n, \rho)$$

- R is the circle radius
- ΔTw is the excess twist
- $\Omega = C/A$
- ρ is the mass density

Matsumoto, Tobias, Olson,
JCTC 1, 117 (2005)

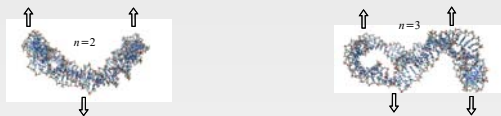
In-plane and out-of-plane modes for circular DNA

In plane bending motions



Relaxed minicircle with 94 base pairs			Overtwisted minicircle with 94 base pairs			
n	GB frequencies		Analytical frequencies	GB frequencies		Analytical frequencies
2	0.165	0.172	0.121	0.166	0.178	0.116
3	0.394	0.452	0.366	0.443	0.471	0.377
4	0.672	0.724	0.697	0.704	0.778	0.734

Out of plane bending motions



Relaxed minicircle with 94 base pairs			Overtwisted minicircle with 94 base pairs			
n	GB frequencies		Analytical frequencies	GB frequencies		Analytical frequencies
2	0.217	0.231	0.211	0.211	0.248	0.235
3	0.525	0.545	0.496	0.520	0.542	0.551
4	0.793	0.851	0.864	0.802	0.901	0.958

DNA binding to the histone core proteins

