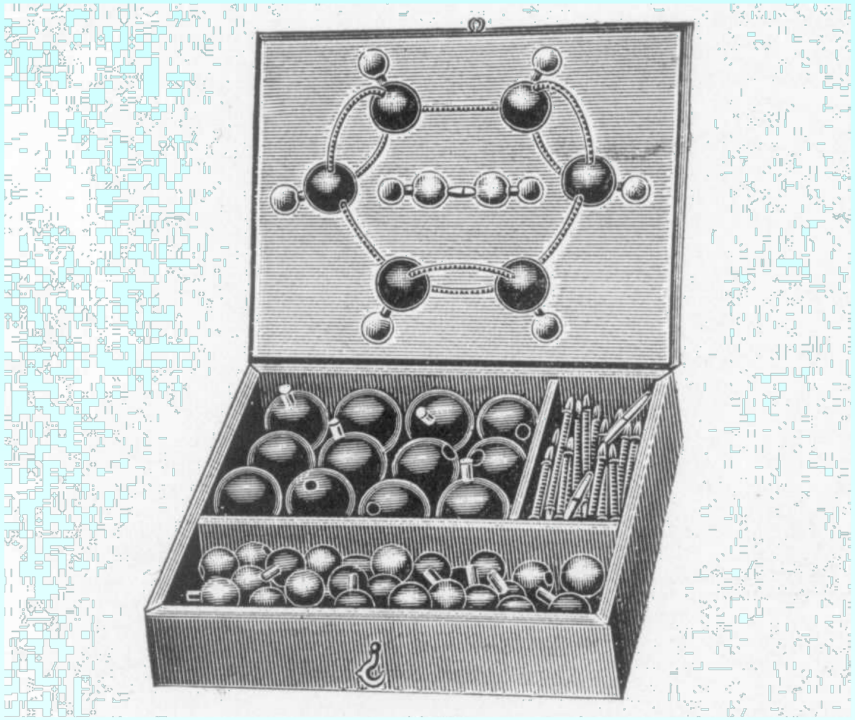
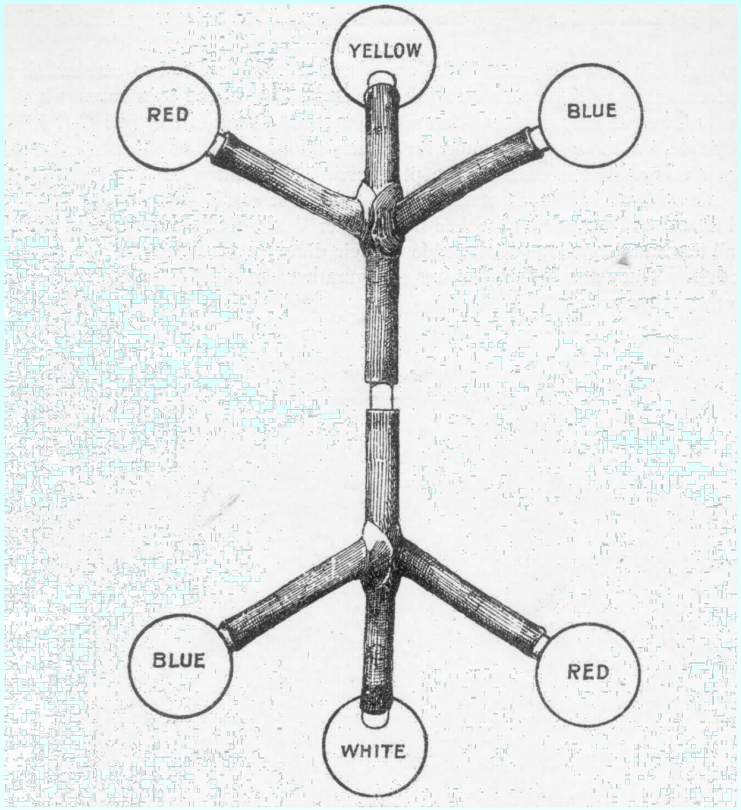
MMTSB and Amber

MMTSB/CTBP Workshop, August 2009

Basics of molecular mechanics and dynamics
Statistical mechanics of liquids
Basic ideas of continuum solvation
The MM/PBSA model

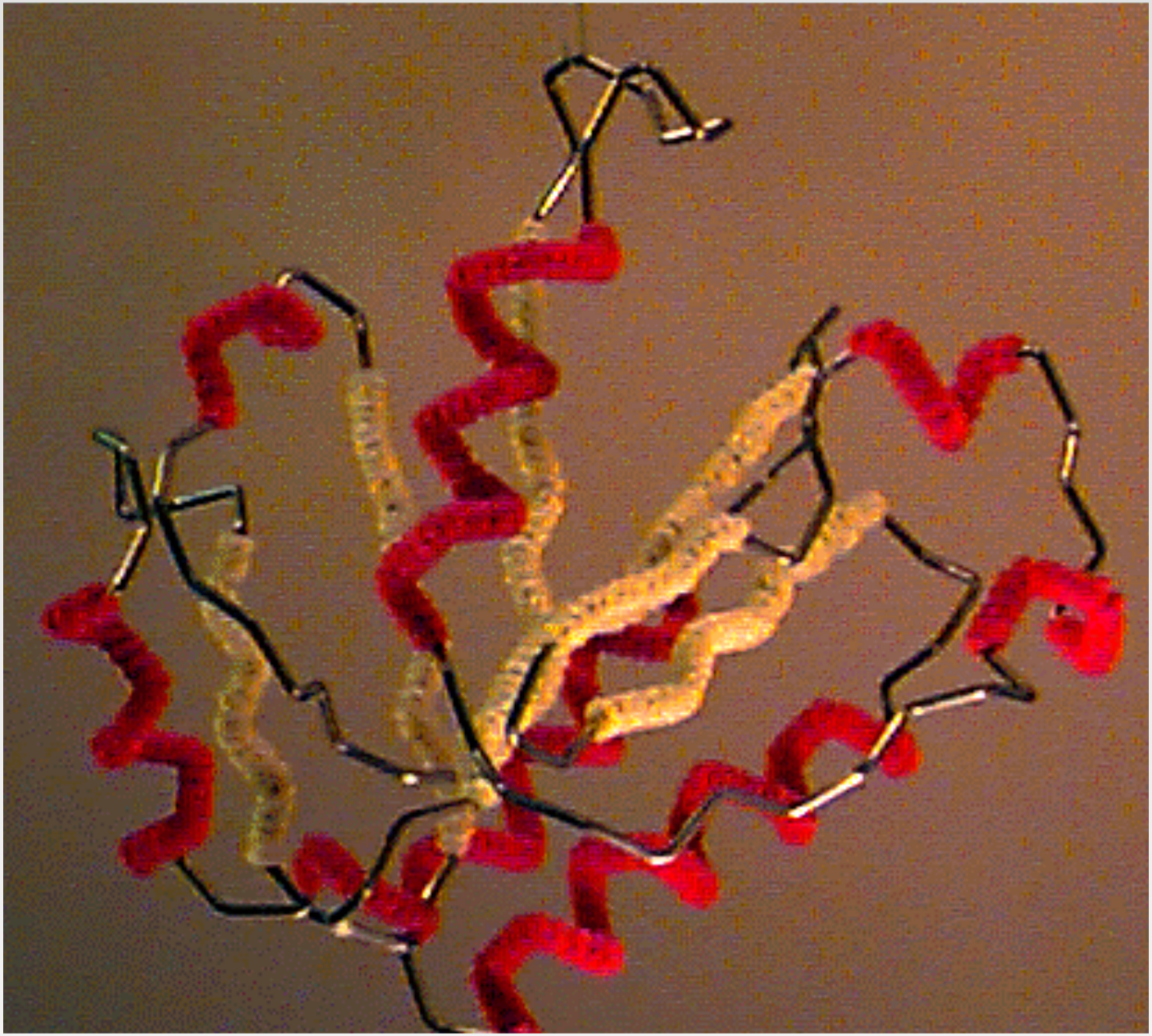
1901 (and earlier?) ball and stick models



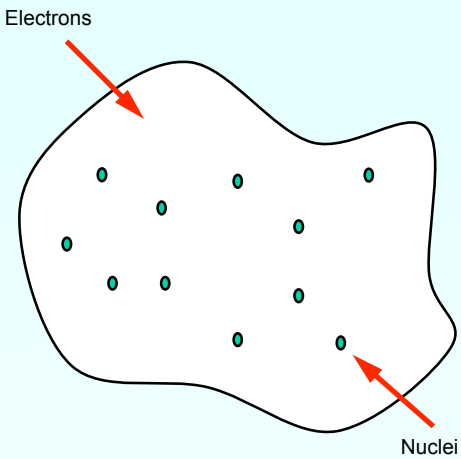


Baird & Tatlock 1901

1950s: wire models of proteins

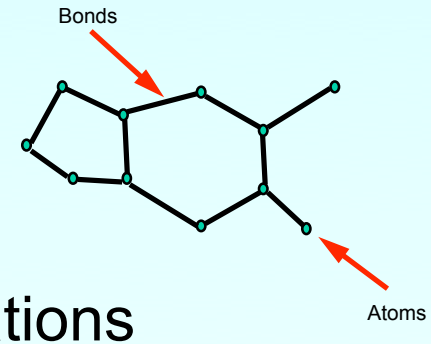


- separate nuclei and electrons
- polarisation, electron transfer and correlation
- can specify electronic state
- can calculate formation energies
- can do chemistry (bond breaking and making)
- variationally bound
- computationally expensive
- typically ~10-100 atoms
- dynamics ~1 ps



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



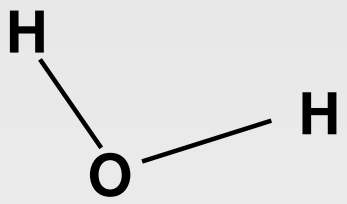
MM MOLECULE

Some force field assumptions

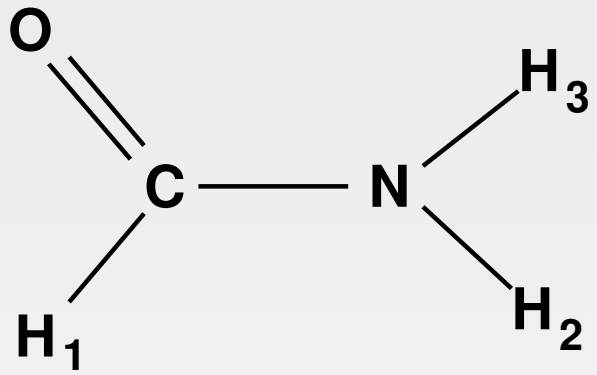
1. Born-Oppenheimer approximation (separate nuclear and electronic motion)
2. Additivity (separable energy terms)
3. Transferability (look at different conformations, different molecules)
4. Empirical (choose functional forms and parameters based on experiment)

What does a force field look like?

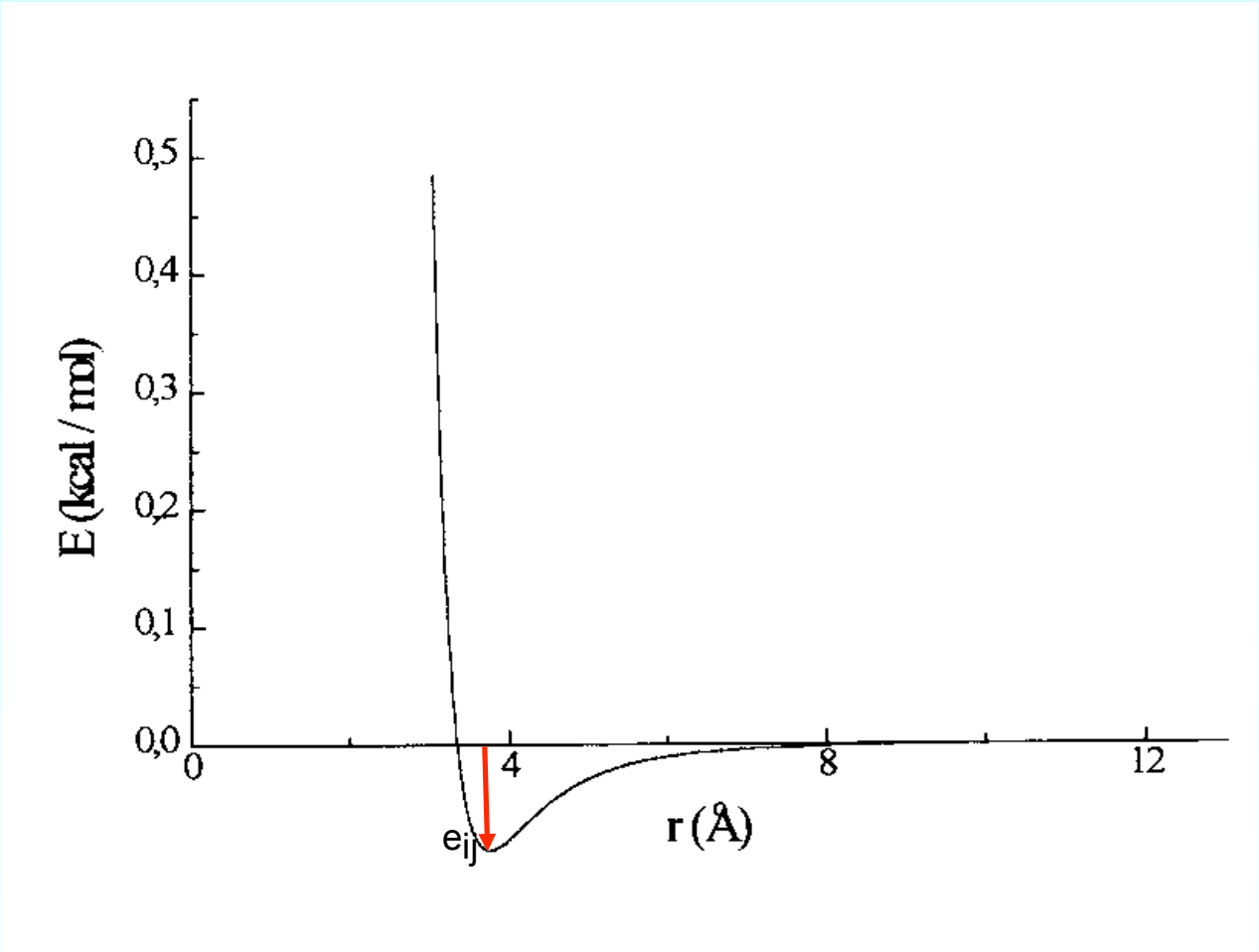
$$
\begin{aligned}
U &= \sum_{\text{bonds}} K_b(b - b_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{impropers} K_w w^2 \\
&+ \sum_{nonbonded\ pairs} \left\{ 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \frac{q_i q_j}{r} \right\} \qquad (1)
\end{aligned}
$$



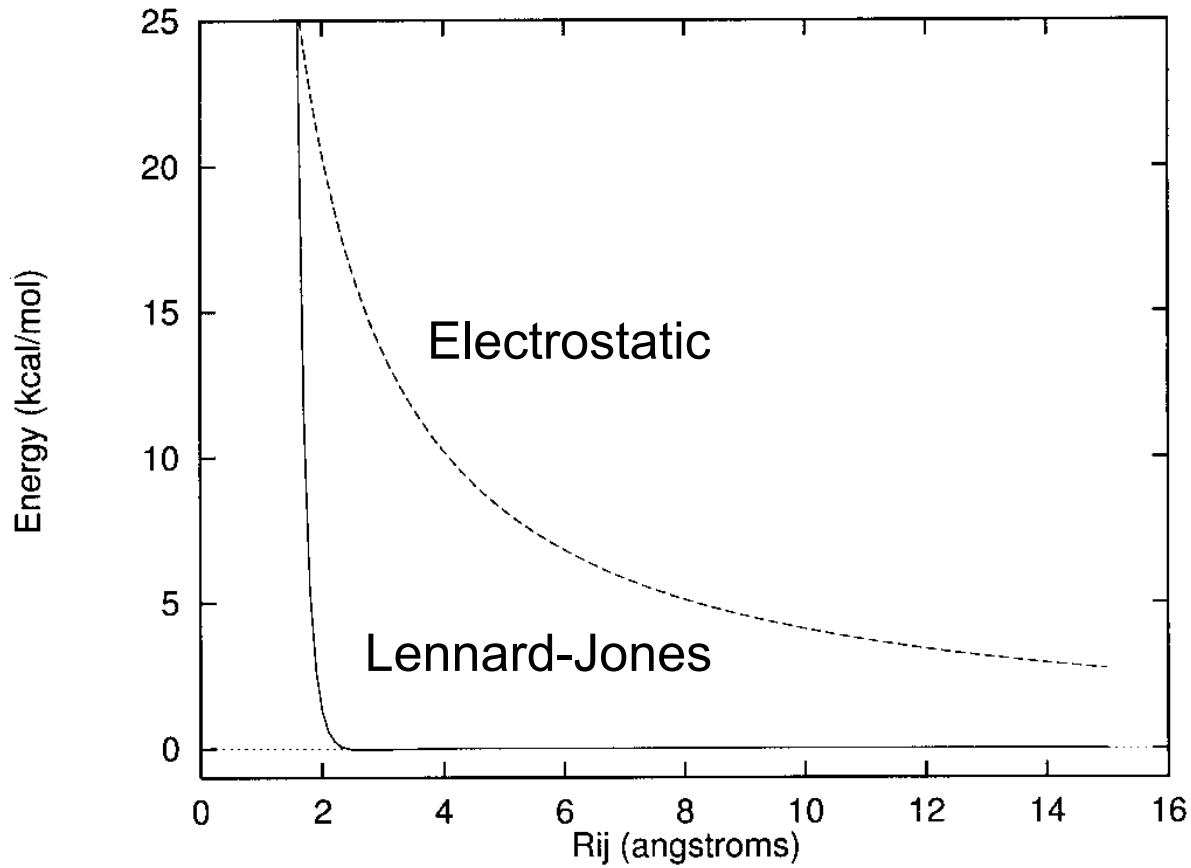
water



formamide



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

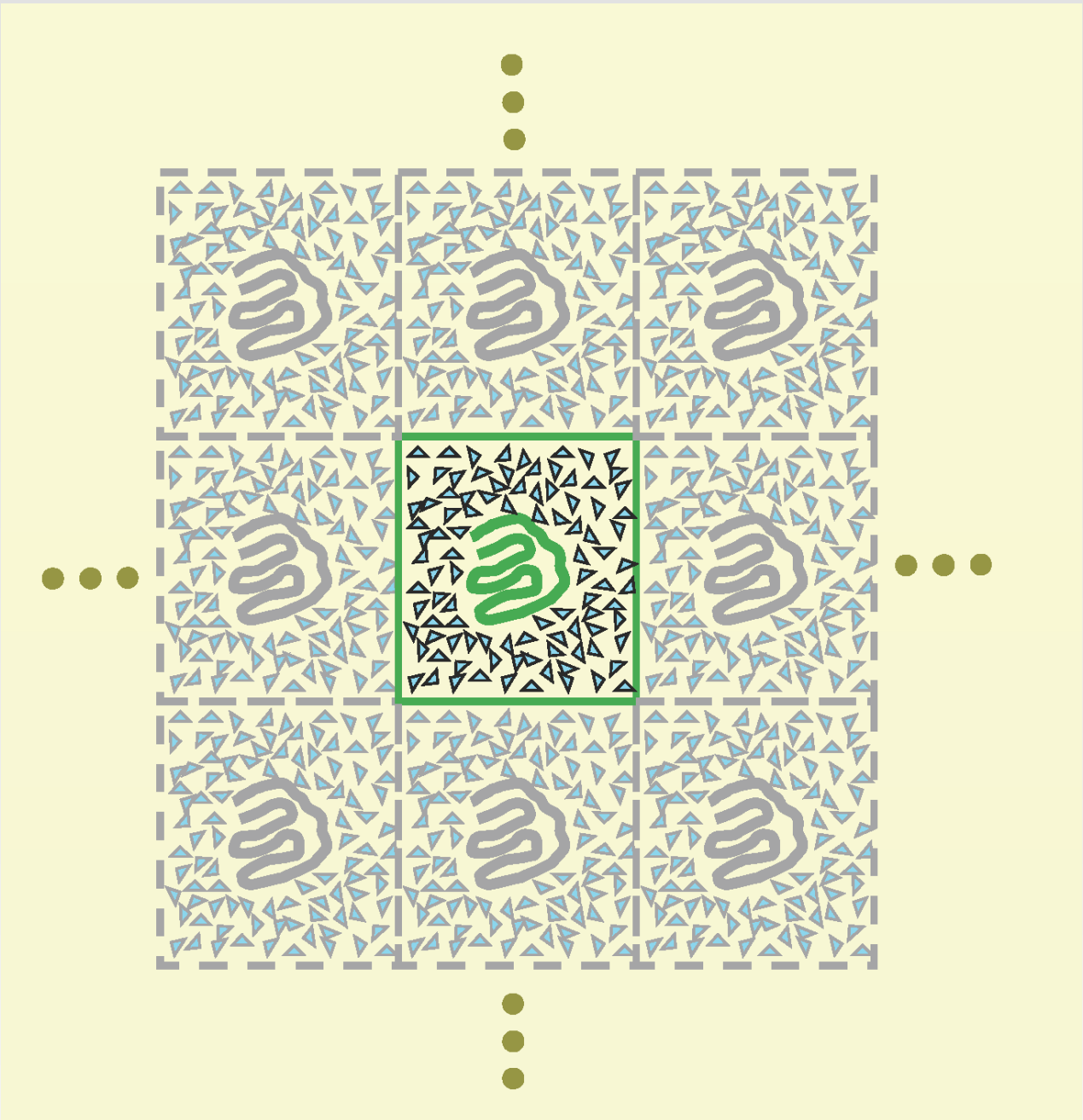
C	sp2 C carbonyl group
CA	sp2 C pure aromatic (benzene)
CB	sp2 aromatic C, 5&6 membered ring junction
CC	sp2 aromatic C, 5 memb. ring HIS
CK	sp2 C 5 memb.ring in purines
CM	sp2 C pyrimidines in pos. 5 & 6
CN	sp2 C aromatic 5&6 memb.ring junct.(TRP)
CQ	sp2 C in 5 mem.ring of purines between 2 N
CR	sp2 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

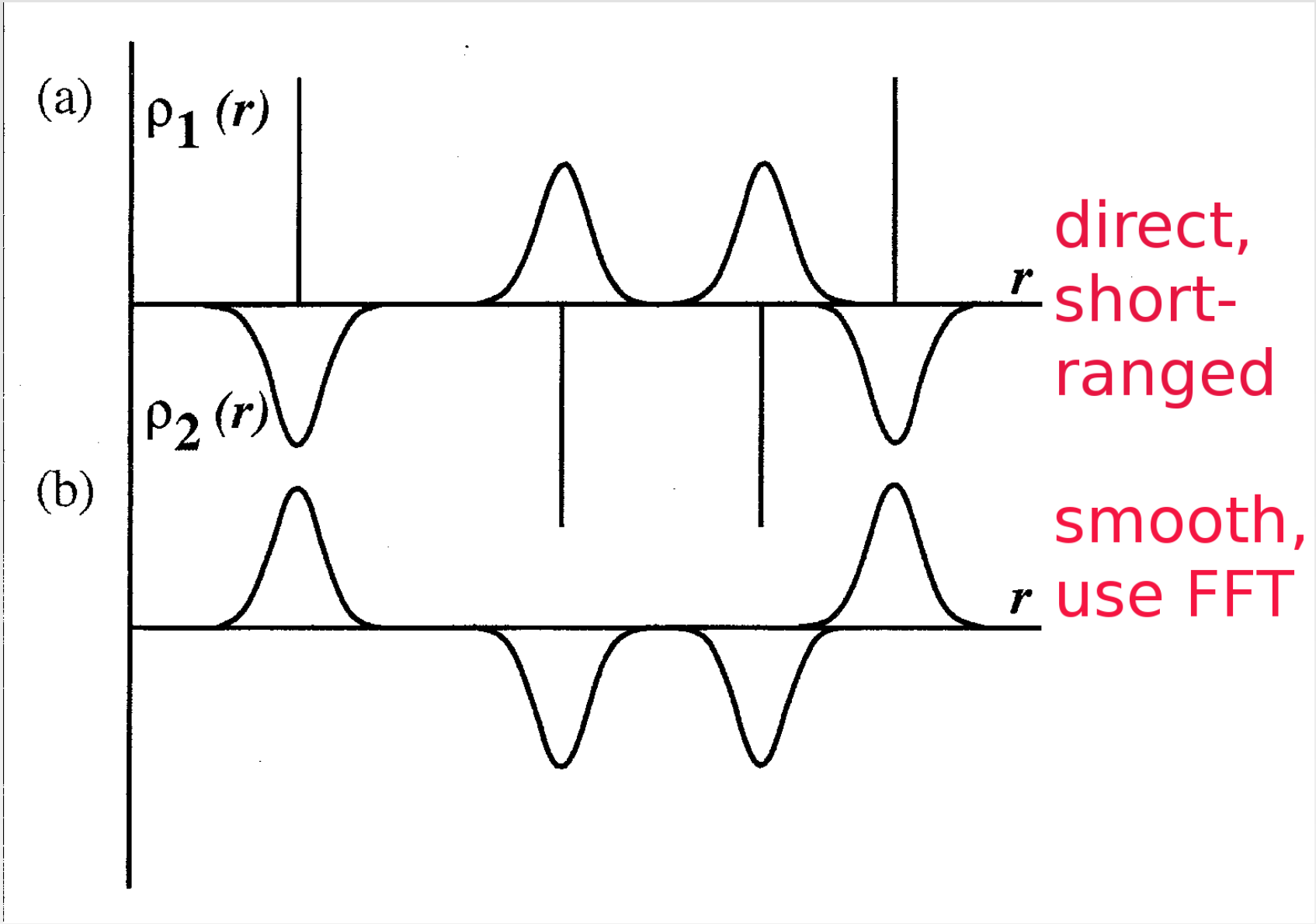
Force fields in Amber

- 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
- ff99bsc0: new torisons for nucleic acids
- ff09: "coming"

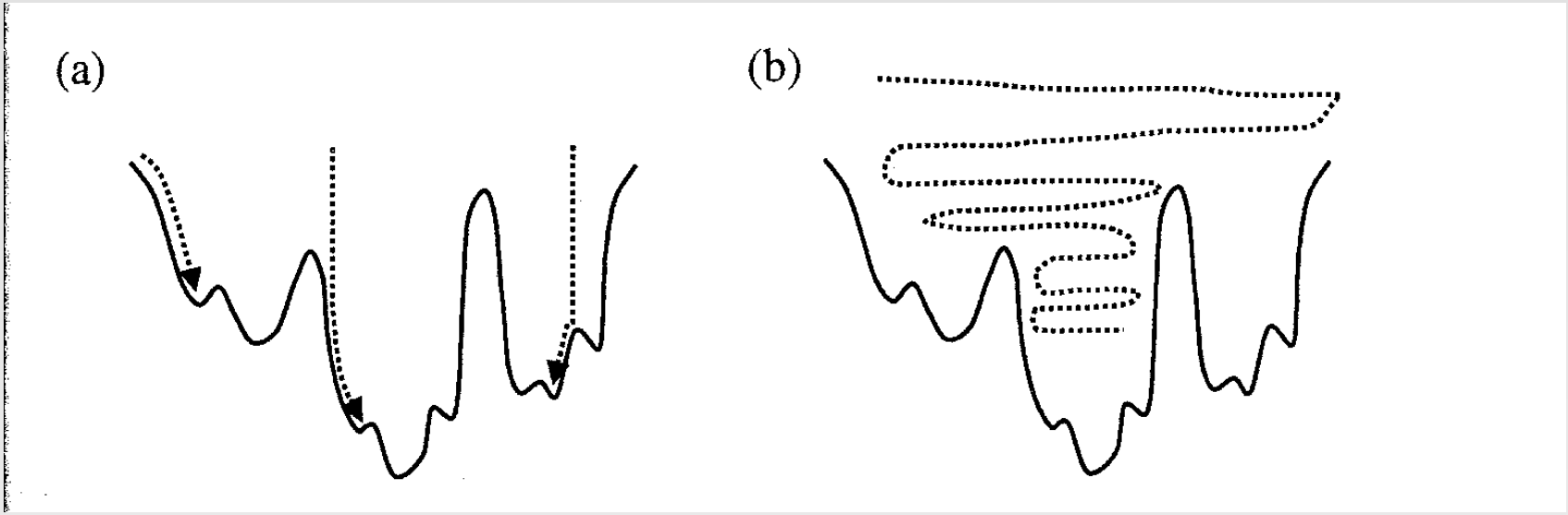
Periodic boundary conditions



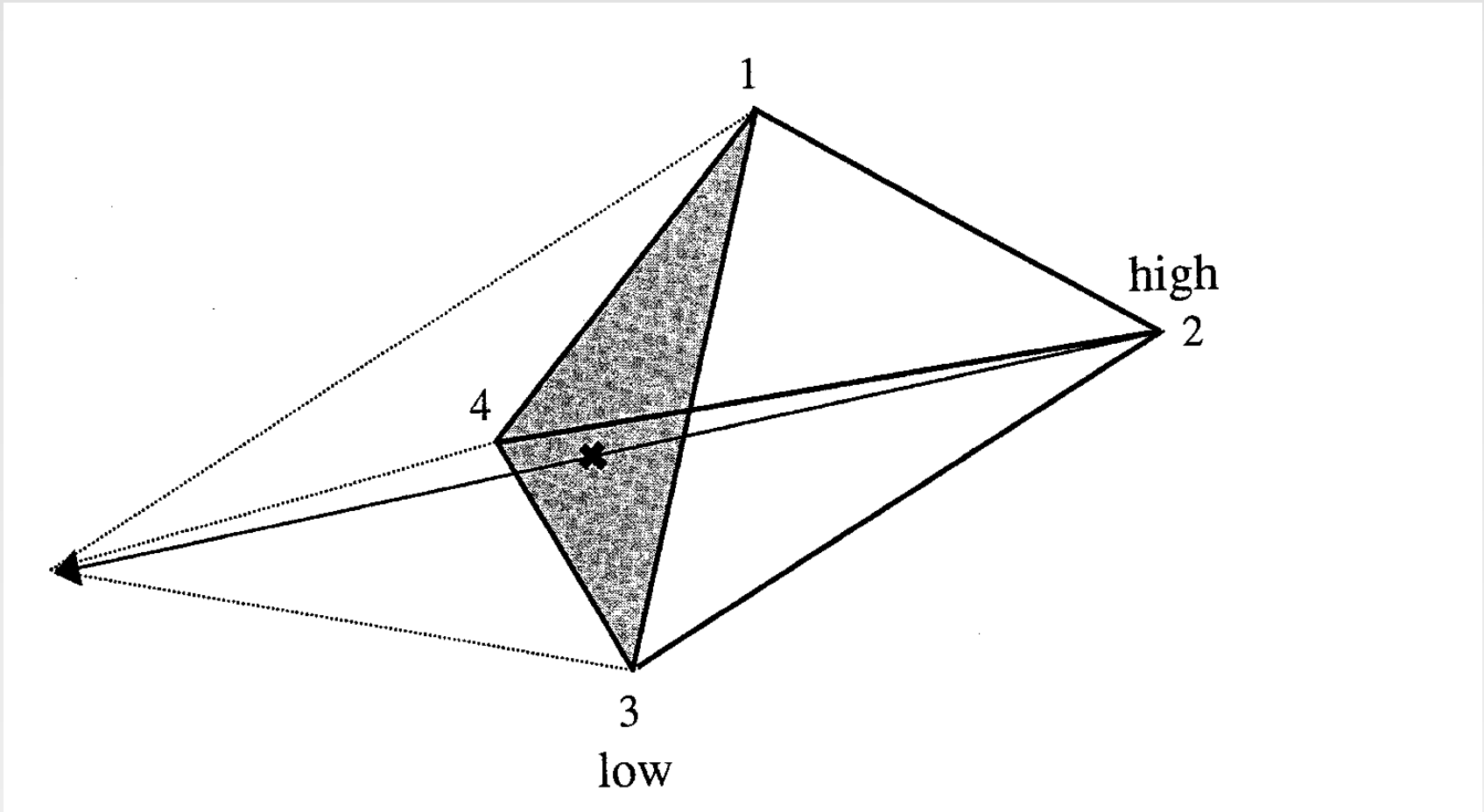
Basics of the Ewald approach



Minimization and simulated annealing



The Simplex algorithm



Molecular dynamics algorithms

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

Eq. (2) is the original Verlet propagation algorithm; Eqs. 3 and 4 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.

Langevin dynamics

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 vicsocity 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:

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

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

$$\zeta = (\beta/m) < A^2 >$$

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

(good reading: J.C. Slater, “Introduction to Chemical Physics”; Dover, pp. 3-51)

- First law of thermodynamics:

$$dU = dQ - dW \text{ or } \Delta U = \int dU = \int dQ - \int dW \tag{5}$$

- Second law of thermodynamics:

$$dS \geq dQ/T \text{ or } TdS \geq dU + dW \tag{6}$$

Connections to microscopic properties

Let p_i be the probability (fraction) of micro-state i. Then we can postulate a connection to the entropy:

$$\boxed{S = -k \sum_i p_i \ln p_i} \tag{7}$$

This is large when the system is "random". For example, if $p_i = 1/W$ (same for all i), then $S = k \ln W$. This entropy is also additive (or "extensive"). Consider two uncorrelated systems that have a total number of states W_1 and W_2. The total number of possibilities for the combined system is $W_1 W_2$. Then:

$$S = k \ln(W_1 W_2) = k \ln W_1 + k \ln W_2 = S_1 + S_2 \tag{8}$$

The canonical ensemble: temperature

Now consider dividing an isolated system (whose total energy U is therefore fixed) into a number of subsystems, each of which could have its own internal energy E_i, but where there is thermal contact between the subsystems, so that energy can be transferred among them. The fixed total energy is

$$U = \sum_i E_i p_i$$

where p_i is the probability that subsystem i will have energy E_i. Let us find the most probable configuration by maximizing the entropy, subject to the constraint of constant total energy and that $\sum p_i = 1$:

$$dS = 0 = -k \sum dp_i (\ln p_i) + k\beta \sum E_i dp_i - ka \sum dp_i \tag{9}$$

Here a and β are undetermined multipliers. The only general solution is when the coefficients of the dp_i terms add to zero:

$$\ln p_i = a - \beta E_i$$

$$\boxed{p_i = \frac{\exp(-\beta E_i)}{\sum \exp(-\beta E_i)}} \tag{10}$$

Connections to clasical thermodynamics

The Lagrange multiplier a is just the denominator of Eq. 10. To figure out what β is, we connect this back to thermodynamics:

$$dS = k\beta \sum_i dp_i E_i = k\beta\, dQ \;\Rightarrow \beta = 1/kT$$

The denominator of Eq. 10 is called the partition function, and all thermodynamic quantities can be determined from it and its derivatives:

$$\boxed{Z \equiv \sum \exp(-\beta E_i)}$$

$$\begin{aligned} A &= U - TS = -kT \ln Z \\ S &= -(\partial A/\partial T)_V = k \ln Z + kT(\partial \ln Z/\partial T)_V \\ U &= -(\partial \ln Z/\partial \beta);\; C_V = T\left(\frac{\partial^2 (kT \ln Z)}{\partial T^2}\right) \end{aligned}$$

Connections to classical mechanics

We have implicitly been considered a discrete set of (quantum) states, E_i, and the dimensionless partition function that sums over all states:

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

How does this relate to what must be the classical quantity, integrating over all phase space:

$$Z_C = \int e^{-\beta H(p,q)} dpdq$$

Z_c has units of $(energy \cdot time)^{3N}$ for N atoms. The Heisenberg principle states (roughly): $\Delta p \Delta q \simeq h$, and it turns out that we should "count" classical phase space in units of h:

$$Z_Q \simeq Z_c / h^{3N}$$

For M indistinguishable particles, we also need to divide by $M!$. This leads to a discussion of *Fermi, Bose* and *Boltzmann* statistics....

Separation of coordinates and momenta

In classical mechanics, with ordinary potentials, the momentum integrals always factor out:

$$Z = h^{-3N} \int e^{-\beta p^2/2m} dp \int e^{-\beta V(q)} dq$$

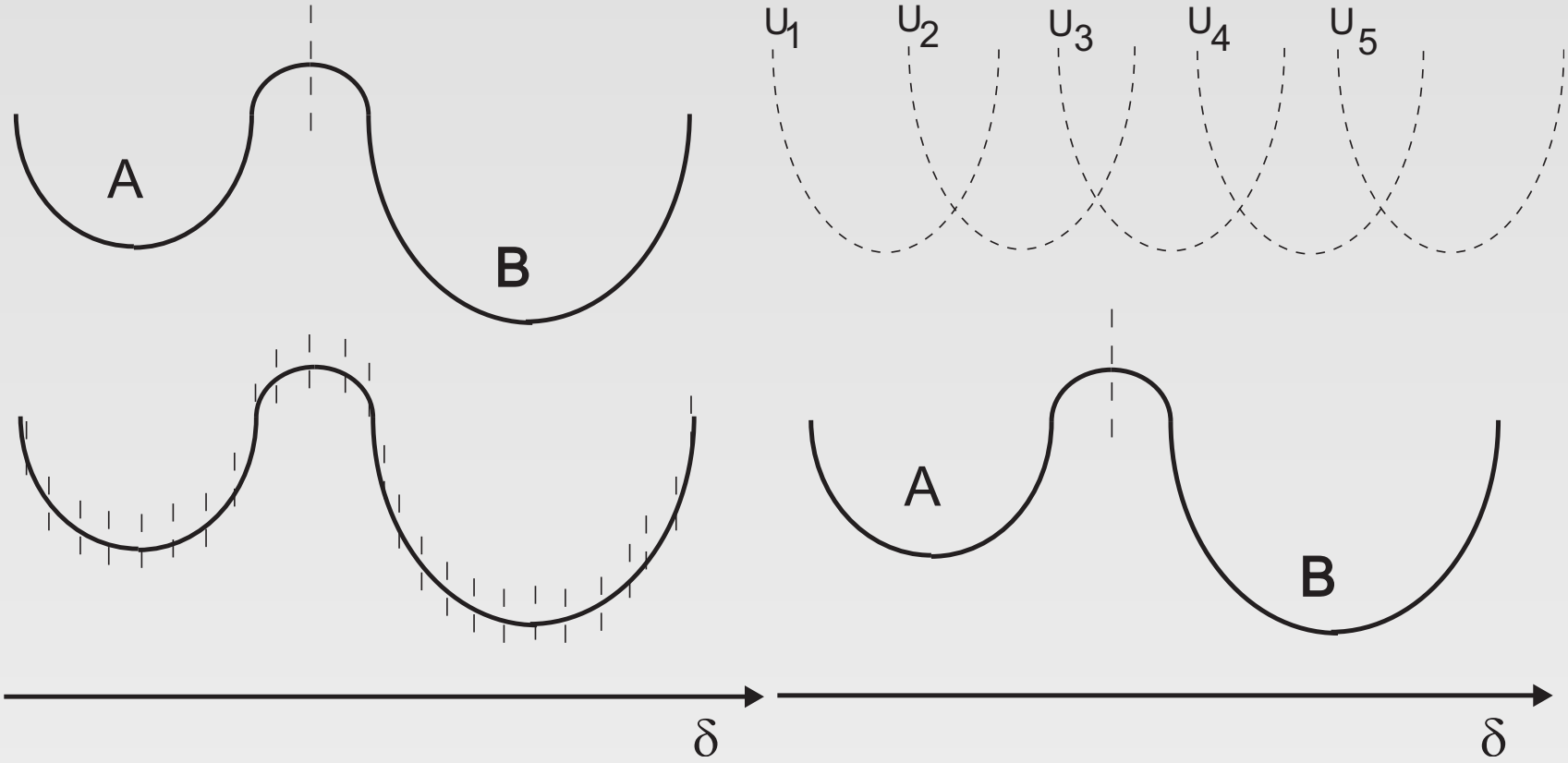
The momentum integral can be done analytically, but will always cancel in a thermodynamic cycle; the coordinate integral is often called the configuration integral, Q. The momentum terms just give ideal gas behavior, and the excess free energy (beyond the ideal gas) is just

$$A = -kT \ln Q$$

The momentum integrals can be done analytically:

$$Z = Q \prod_{i=1}^{N} \Lambda_i^{-3}; \quad \Lambda_i = h/(2\pi m_i k_B T)$$

Getting conformational free energies



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

$$W = -k_B T \ln \rho(\delta) \qquad (11)$$

Free energy profiles

$$\rho(\delta) = \frac{\int \exp(-\beta U)\, d\Sigma}{\int \exp(-\beta U)\, d\delta\, d\Sigma} \tag{12}$$

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:

$$\begin{aligned} \rho^*(\delta) &= \exp[-\beta U^*(\delta)] \frac{\int \exp(-\beta U)\, d\Sigma}{\int \exp(-\beta [U + U^*])\, d\delta\, d\Sigma} \\ &= \rho(\delta) \exp[-\beta U^*(\delta)] / \langle \exp(-\beta U^*) \rangle \end{aligned} \tag{13}$$

$$\langle \exp(-\beta U^*) \rangle = \frac{\int \exp(-\beta U^*) \exp(-\beta U)\, d\delta\, d\Sigma}{\int \exp(-\beta U)\, d\delta\, d\Sigma} \tag{14}$$

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

$$W^*(\delta) = W(\delta) + U^*(\delta) - C' \tag{15}$$

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

Thermodynamic integration: computational alchemy

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

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

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

$$\boxed{\Delta A = A(1) - A(0) = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} d\lambda} \qquad (16)$$

This is called thermodynamic integration, and is a fundamental connection between macroscopic free enegies, 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

$$\boxed{\Delta A = \int_0^1 \langle \Delta V \rangle_\lambda \, d\lambda} \tag{17}$$

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.

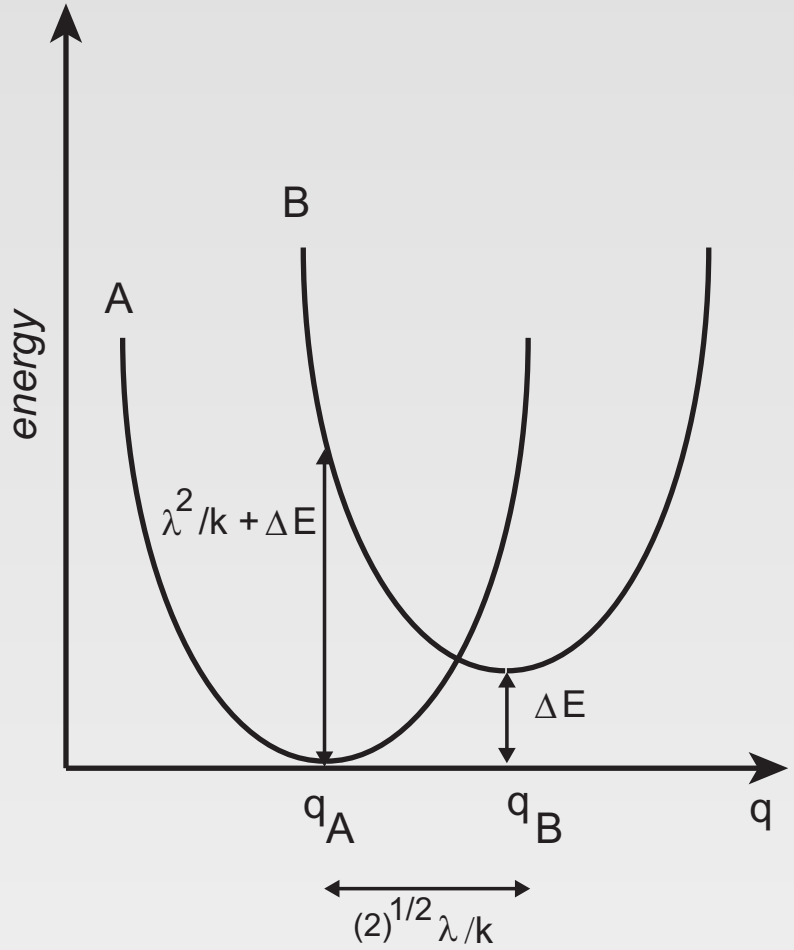
Free energy perturbation theory

Here is an (initially) completely different approach:

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

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”



$$
\begin{aligned}
V_A(q) &= \frac{1}{2}k(q-q_A)^2 \\
V_B(q) &= \frac{1}{2}k(q-q_B)^2 \\
\Delta V(q) &= \sqrt{2}\lambda(q-q_A)+\frac{\lambda^2}{k}+\Delta E
\end{aligned}
$$

Marcus theory thermodynamic integration

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

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

What is the distribution of ΔV in the V_A state?

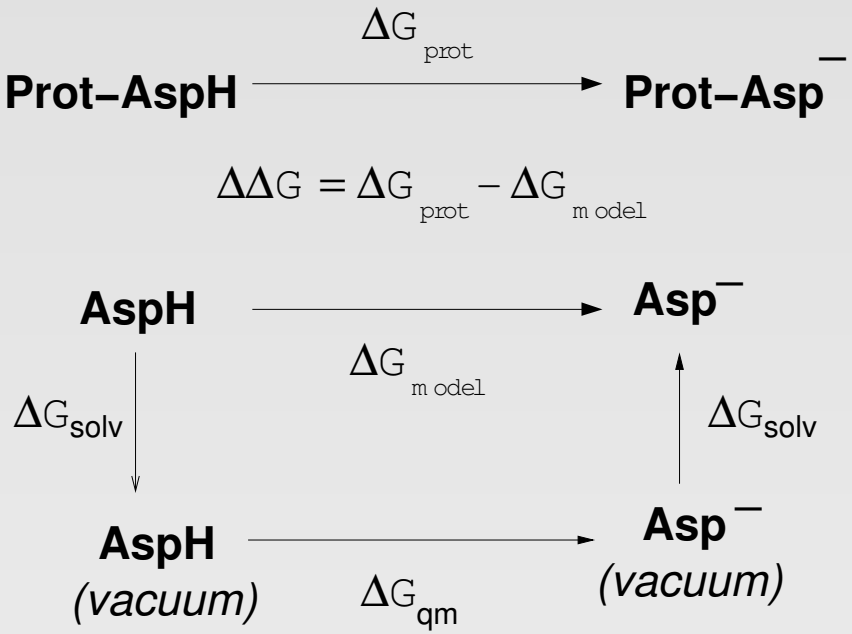
$$\rho(\Delta V) = \rho(q) \left| \frac{dq}{d\Delta V} \right| \quad \textit{where} \quad q(\Delta V) = \left(\frac{\lambda^2 + k\Delta E}{\sqrt{2}k\lambda} \right) - \frac{\Delta V}{\sqrt{2}\lambda}$$

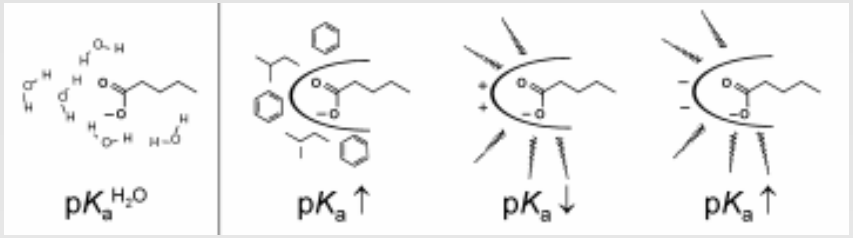
$$\rho(\Delta V) \sim \frac{1}{\sqrt{2}\lambda} \exp\{-\beta V_A[q(\Delta V)]\} \simeq \exp\left\{ -\frac{(\Delta V - \lambda^2/k - \Delta E)^2}{2\sigma^2} \right\} \quad \textit{with} \quad \sigma^2 = \frac{2\lambda^2}{k\beta}$$

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:

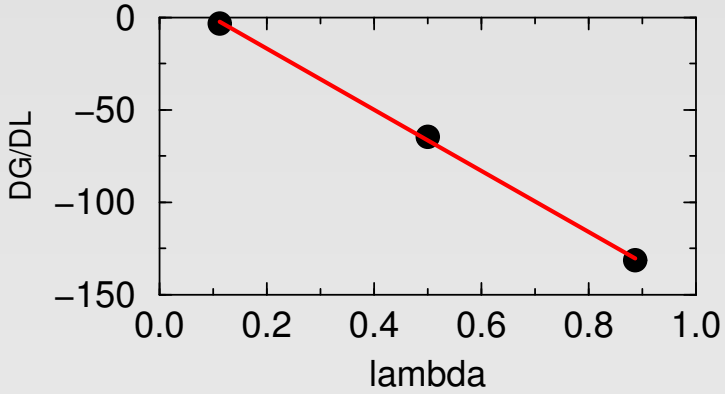
$$\Delta A = -kT \ln \left\langle e^{-\beta \Delta V} \right\rangle_A = \Delta E$$

Application: pKa behavior in proteins

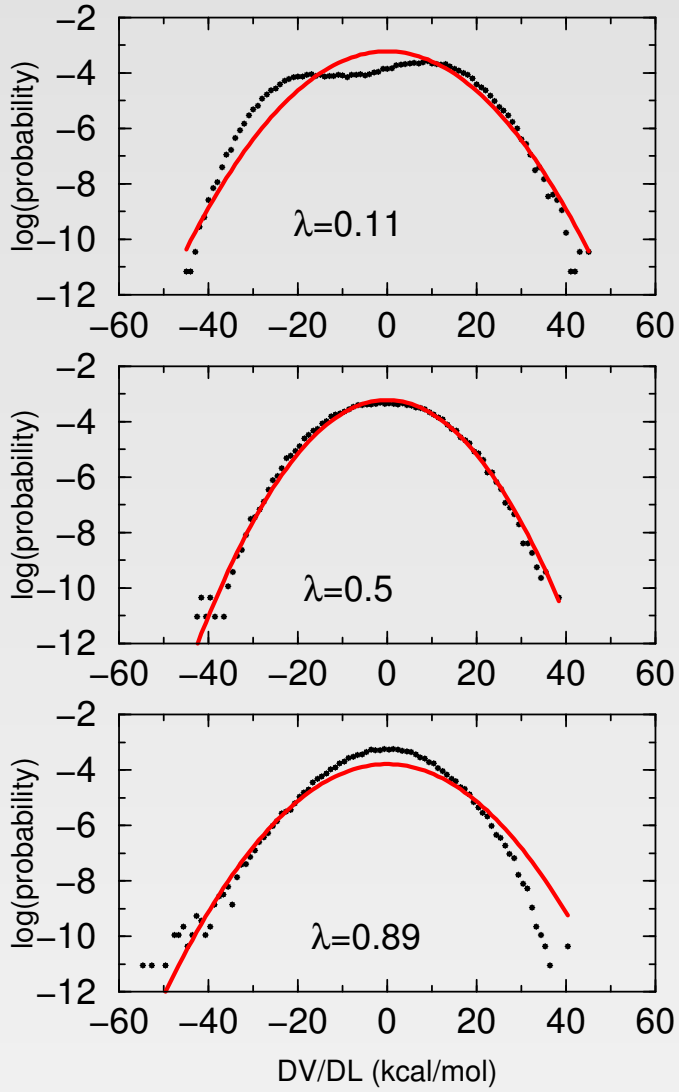




Energy gap distributions

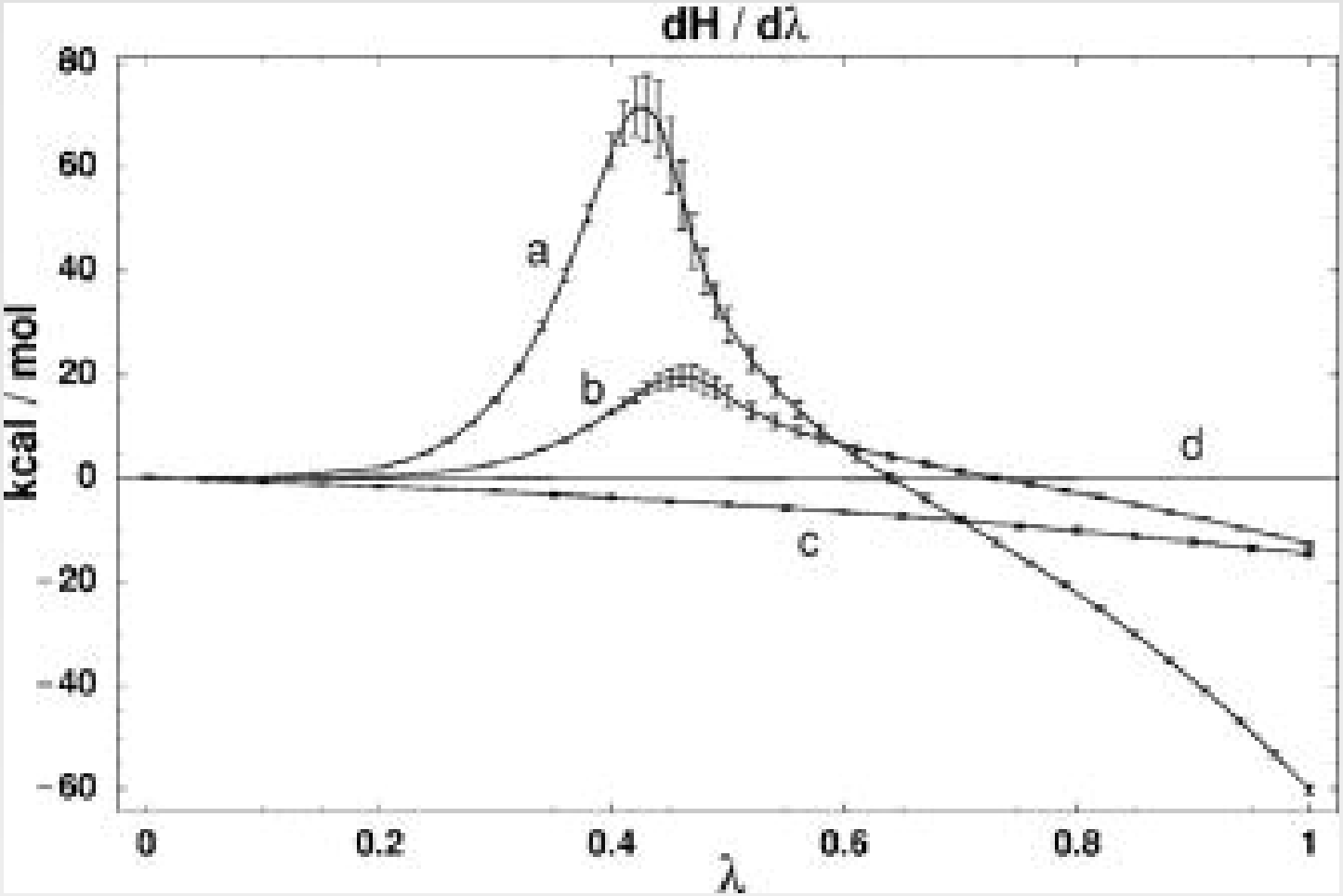


Lambda	DG/DL	(kcal/mol)
0.11270	−3.1	
0.50000	−64.5	
0.88729	−131.4	



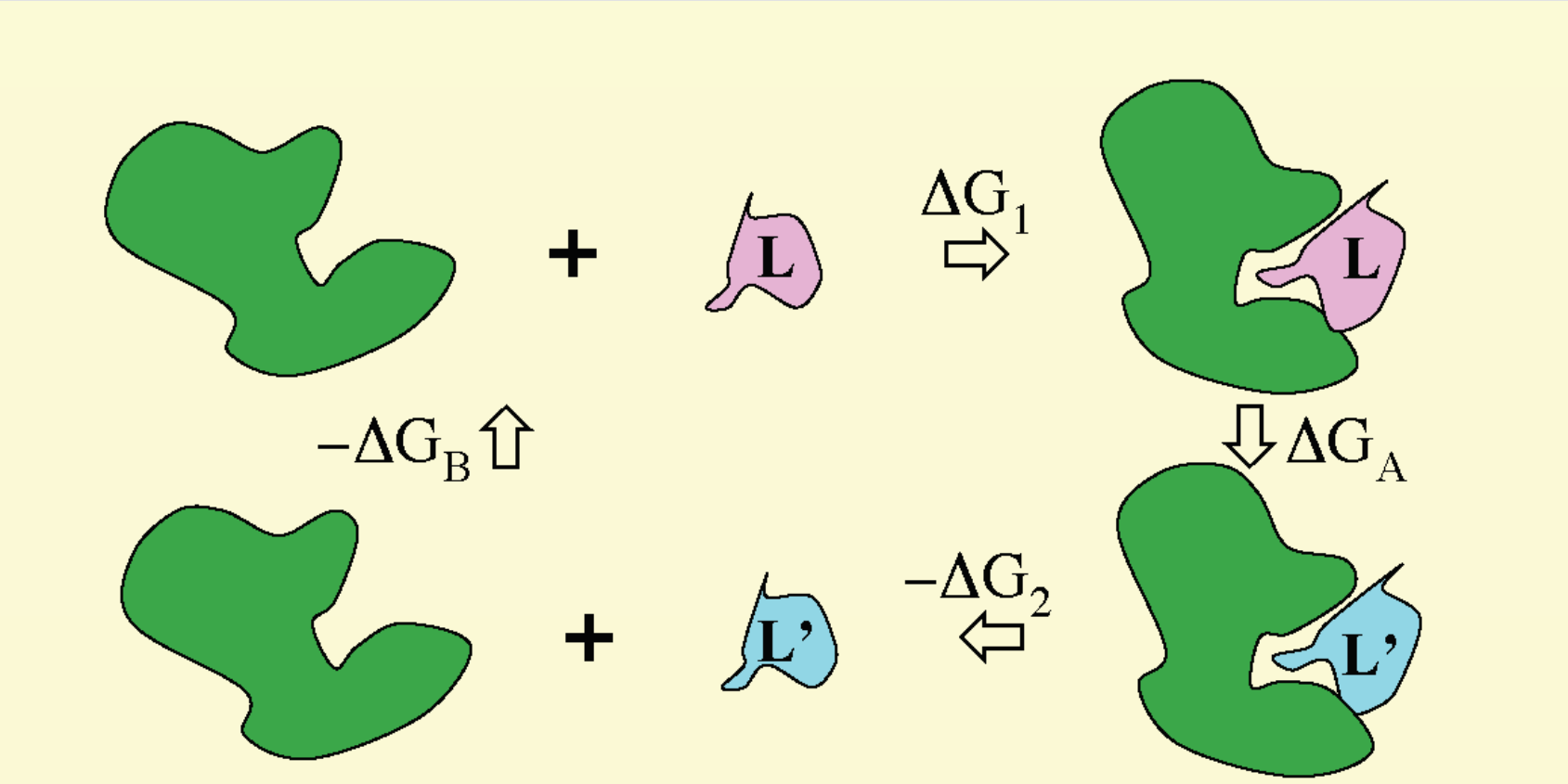
Simonson, Carlson, Case, *JACS* **126**:4167 (2004)

Not everything is linear!

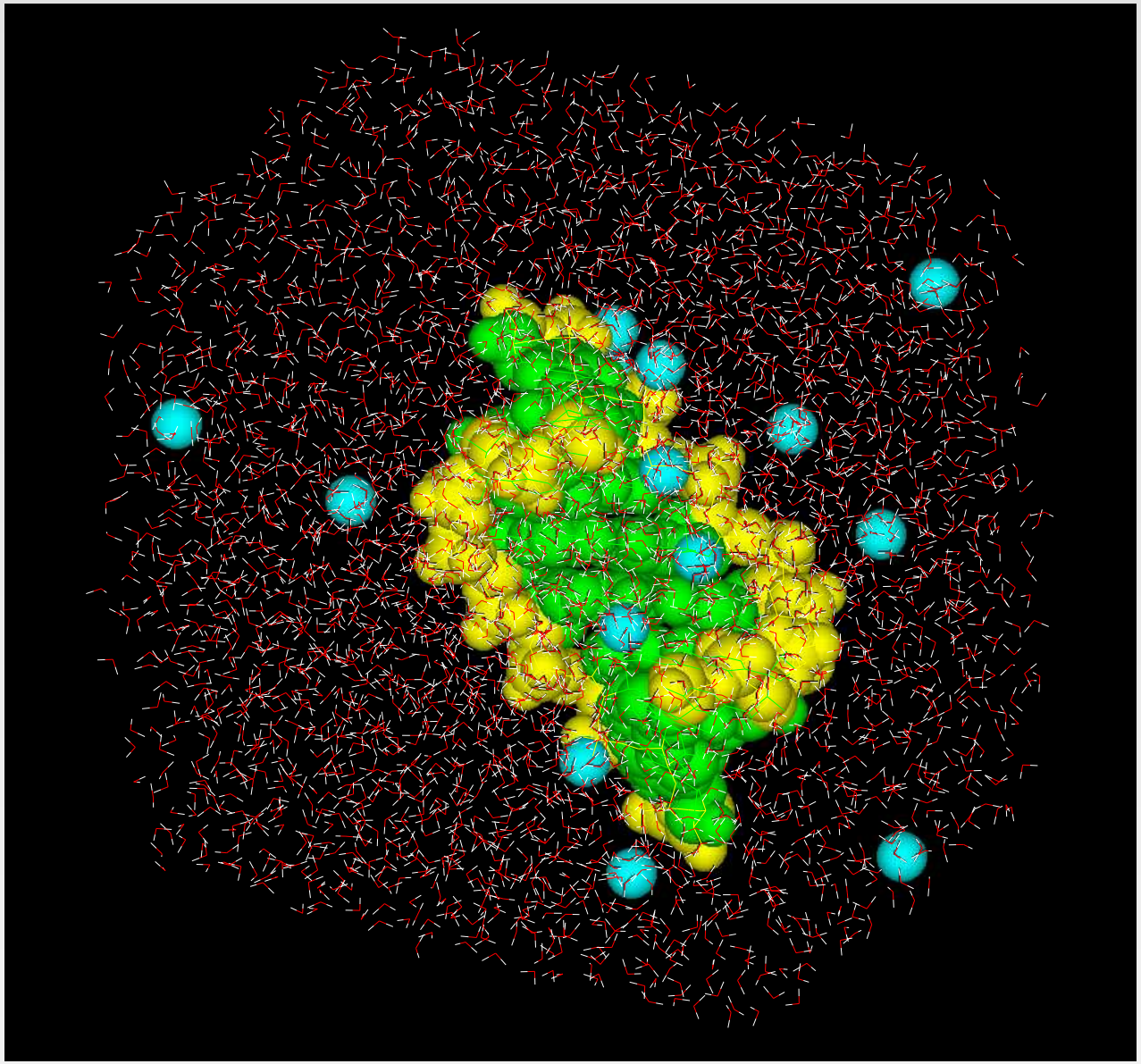


Shirts, Pitera, Swope, Pande, *J. Chem. Phys.* **119,** 5740 (2003).

Thermodynamics cycles in ligand binding

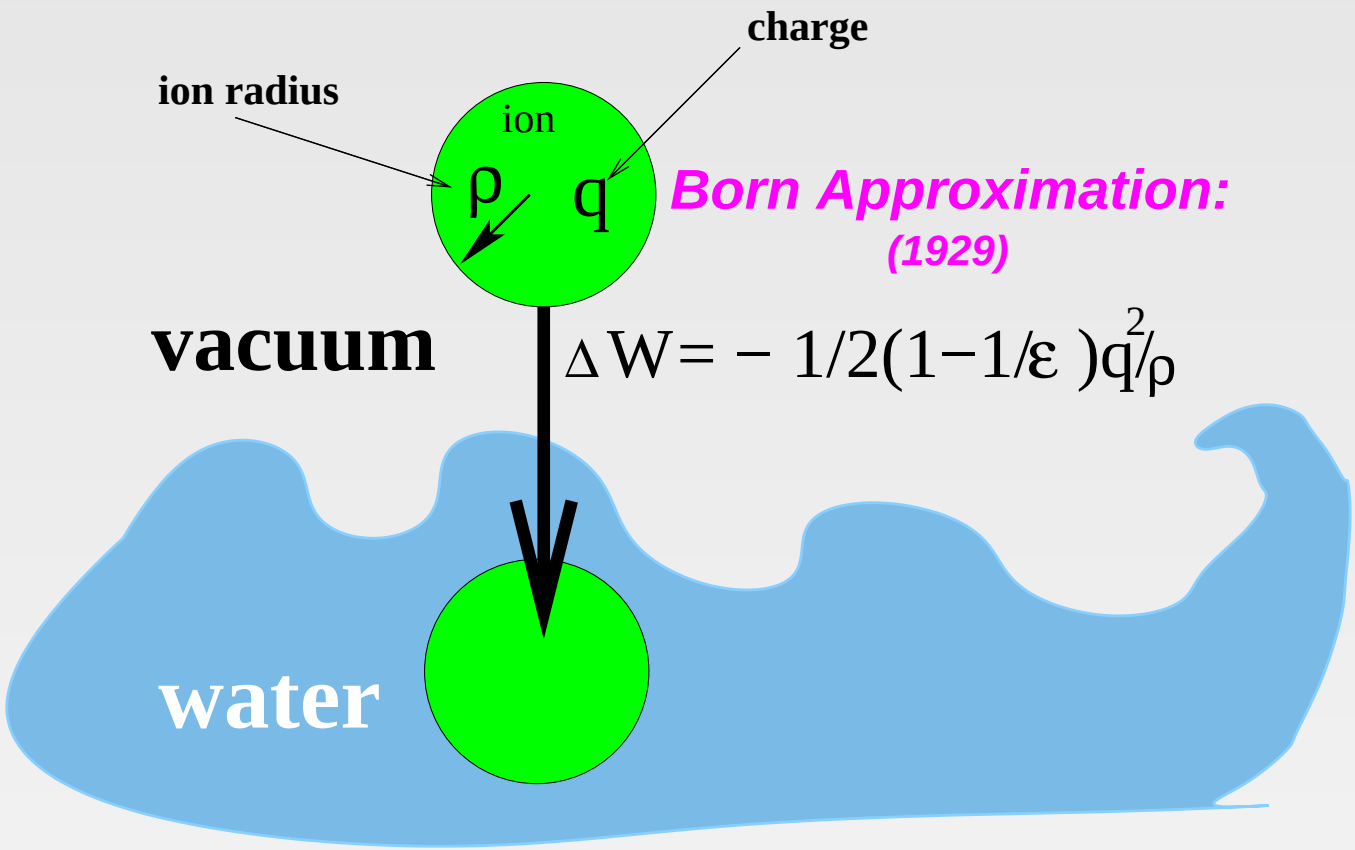


Example of explicit solvation setup



Basic ideas of continuum solvent models

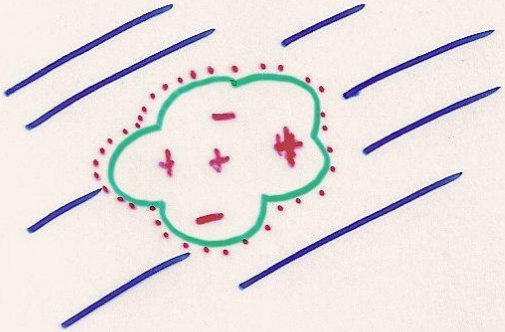
- Tomasi & Persico, *Chem Rev.* **94,** 2027 (1994)
- Simonson, *Rep. Prog. Phys.* **66**, 737 (2003)
- Bashford & Case, *Annu. Rev. Phys. Chem.* **51**, 129 (2000)
- Gallicchio & Levy, *J. Comput. Chem.* **25,** 479 (2004)



Conductor-like Screening Model

Klamt JPC 99 2224 (1995)

Ponder + Case pp 65-67



$$\frac{\varepsilon}{\varepsilon-1}$$

$$E = E_{gas} + \int z \frac{1}{r_z - r_s} q + \frac{1}{2}\int q \frac{1}{r_s - r_{s'}} q'$$

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

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

molecule-solvent interaction:

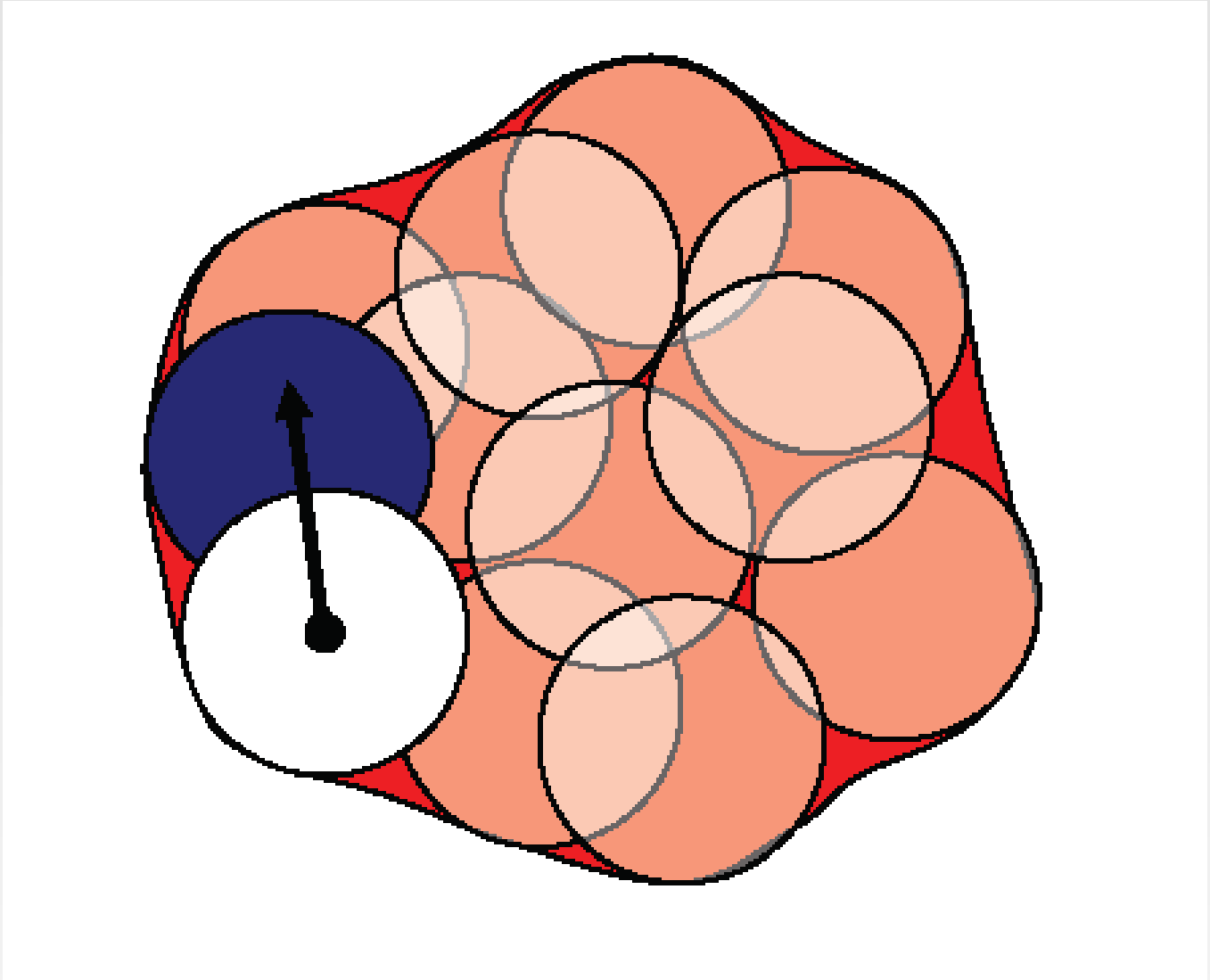
$$-z B A^{-1} B z \equiv -z \phi^{RF}$$

solvent-solvent interaction:

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

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:

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

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

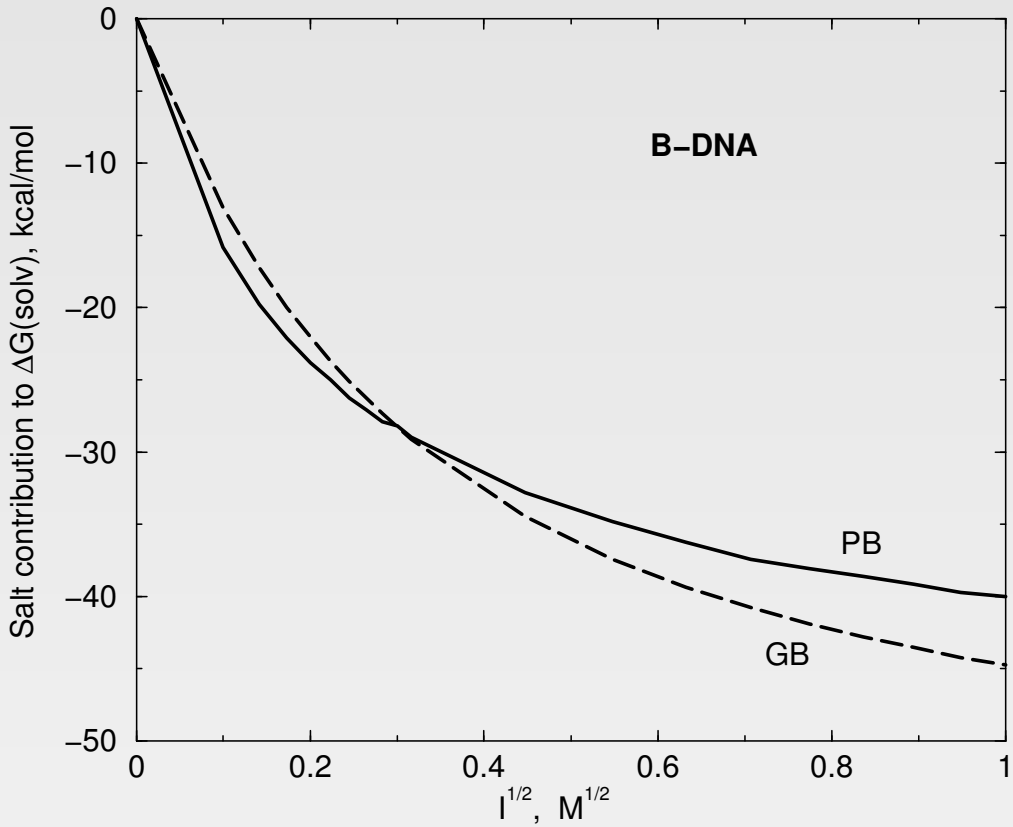
$$\Delta G_{GB} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon_{ext}}\right)\frac{q^2}{R_{eff}}; \quad or \quad -\frac{1}{2}\left(1-\frac{1}{\varepsilon_{ext}}\right)\frac{q_i q_j}{f^{GB}(R^i_{eff}, R^j_{eff}, r_{ij})}$$

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

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

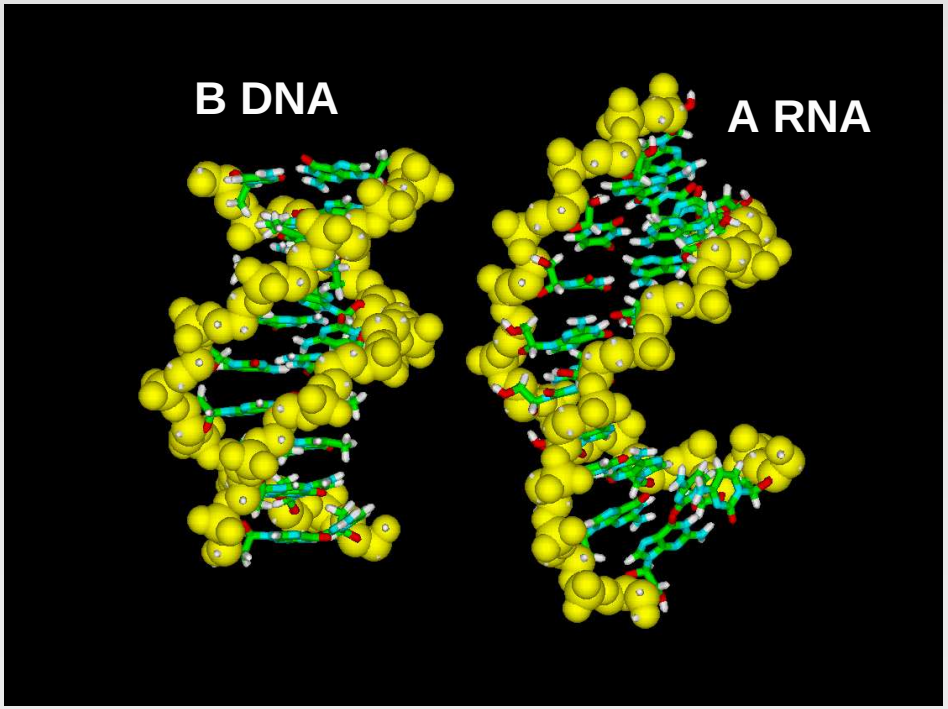
Effects of added salt

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



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)