

Solvation (Condensed Phase) Models

Chemical Phenomena

Video VI.i

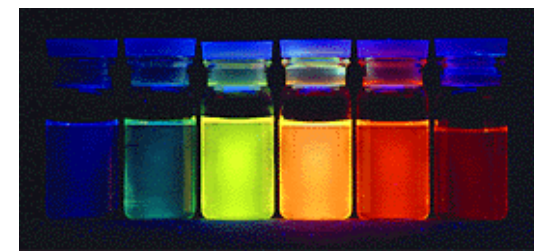
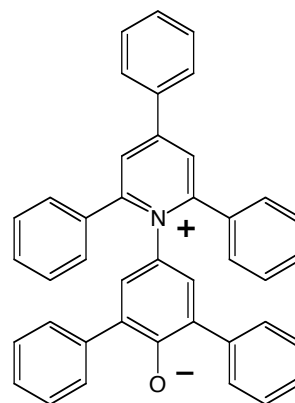
What Condensed Phases are Important?

- Homogeneous liquid solutions are the most common condensed phases found in experimental chemistry. We'll focus on these, but make connections to others
- Solids
- Surfaces
- Liquid crystal solutions
- Supercritical fluids
- Membranes
- Note, in certain instances, the fine line between a condensed phase and a supermolecular complex

Why is Solvation Important?

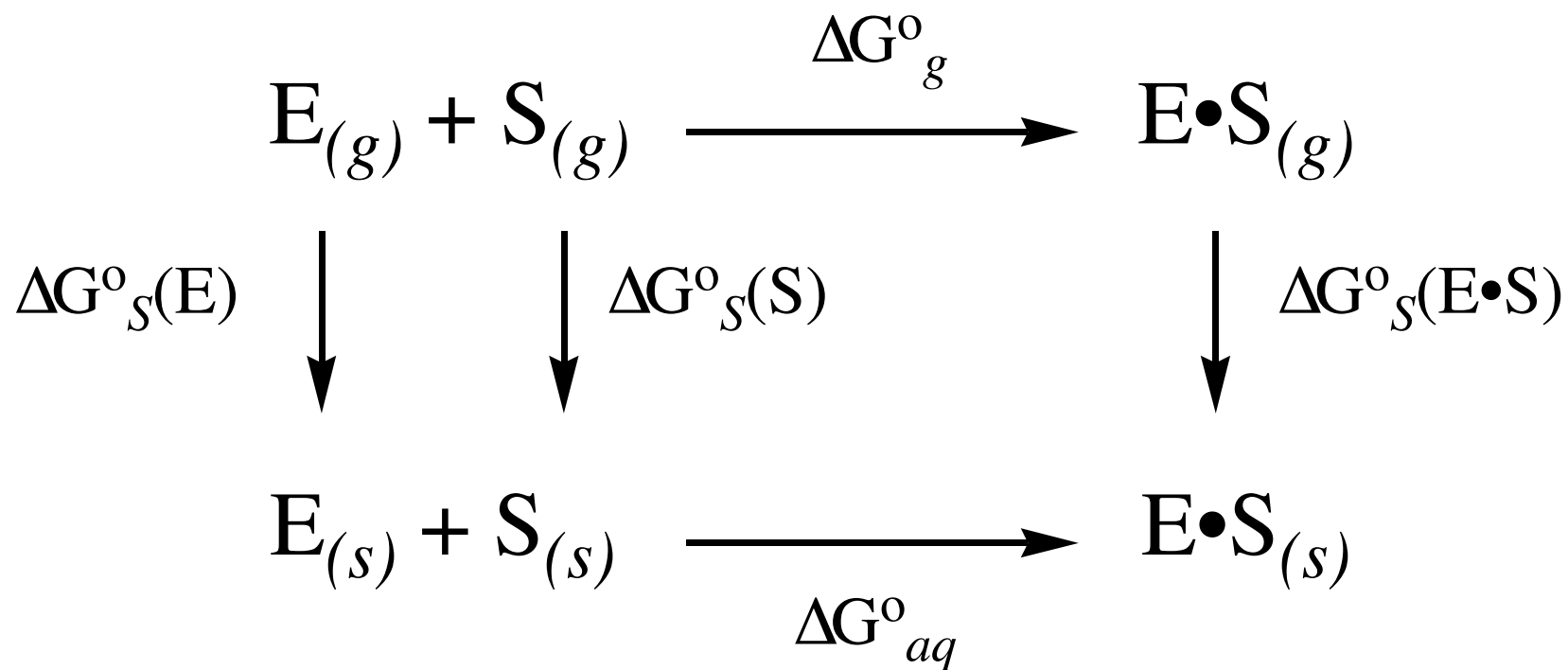
- **Condensed-phase properties depend on the condensed-phase wave function, and $\langle \Psi_{\text{gas}} | A | \Psi_{\text{gas}} \rangle$ may be very different from $\langle \Psi_{\text{sol'n}} | A | \Psi_{\text{sol'n}} \rangle$**
- **Interactions between two (or more) molecules in solution depend on partial desolvation of each**
- **Potential energy hypersurfaces (and hence kinetics and equilibria) may be quantitatively and qualitatively different in solution by comparison to the gas phase**

Solvatochromism of Dye E_T-30 (S₁ – S₀)

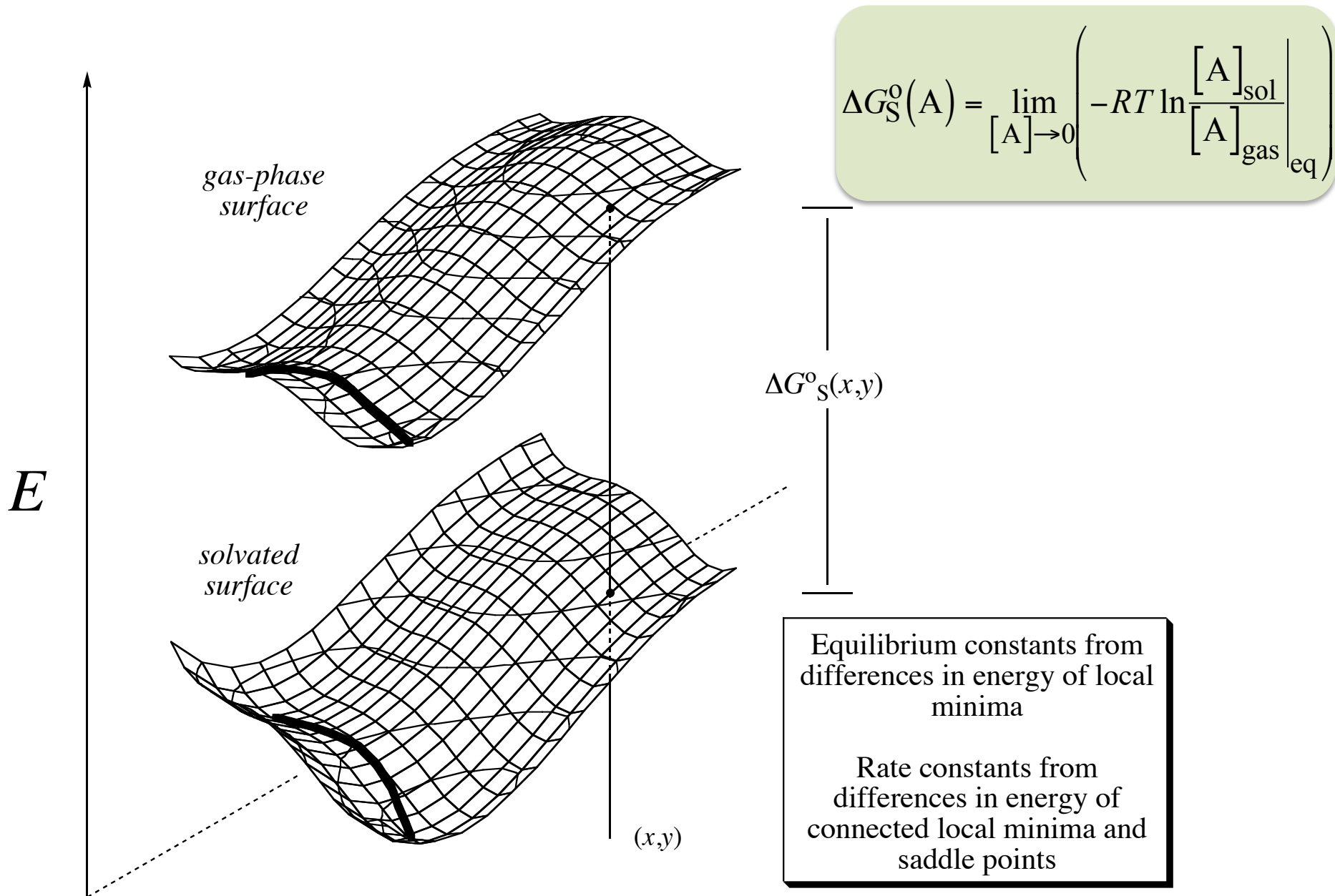


Solvent	Color	λ_{\max} , nm
anisole	yellow	769
acetone	green	677
2-pentanol	blue	608
ethanol	violet	550
methanol	red	515

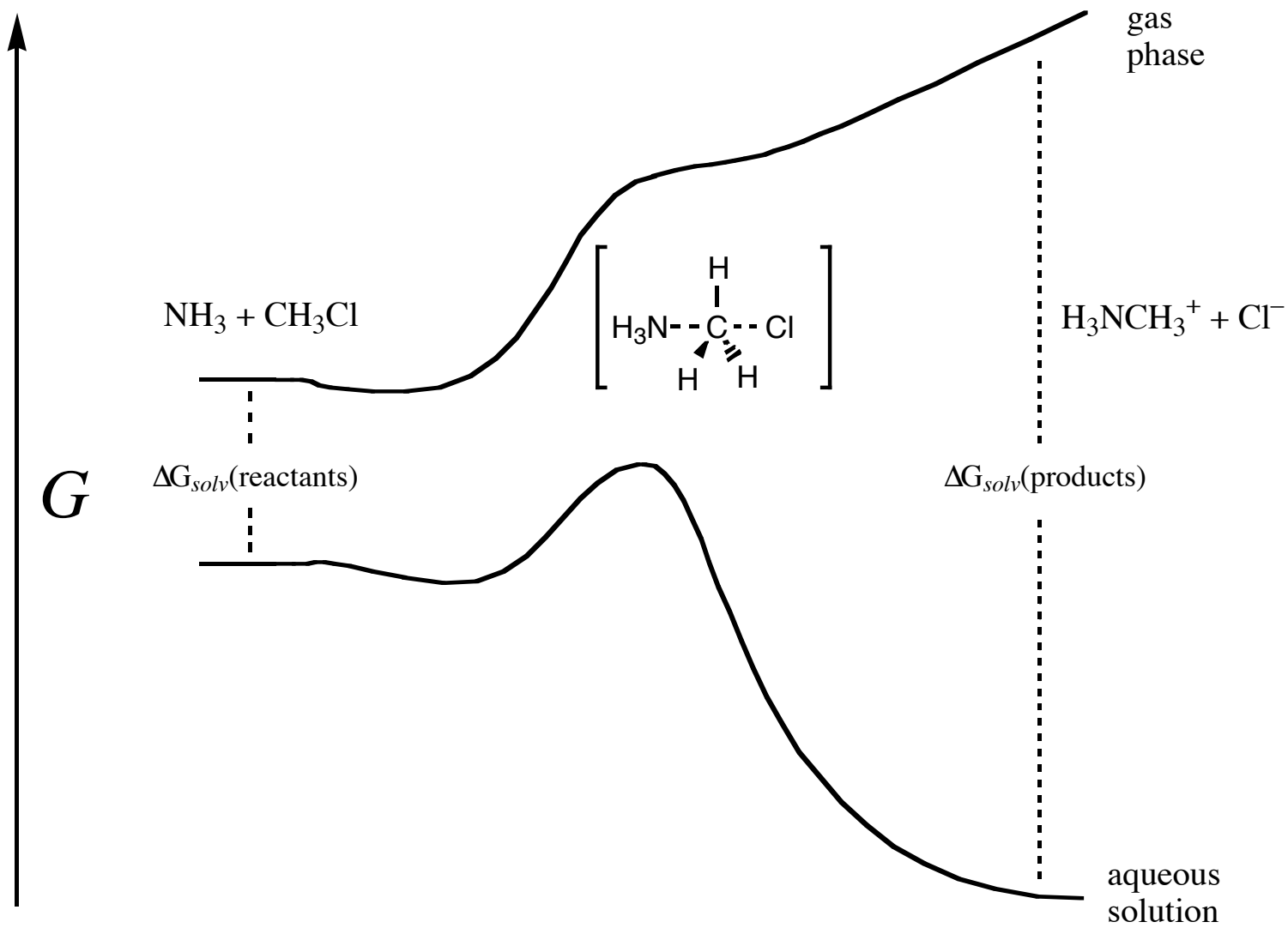
Enzyme–Substrate Interactions



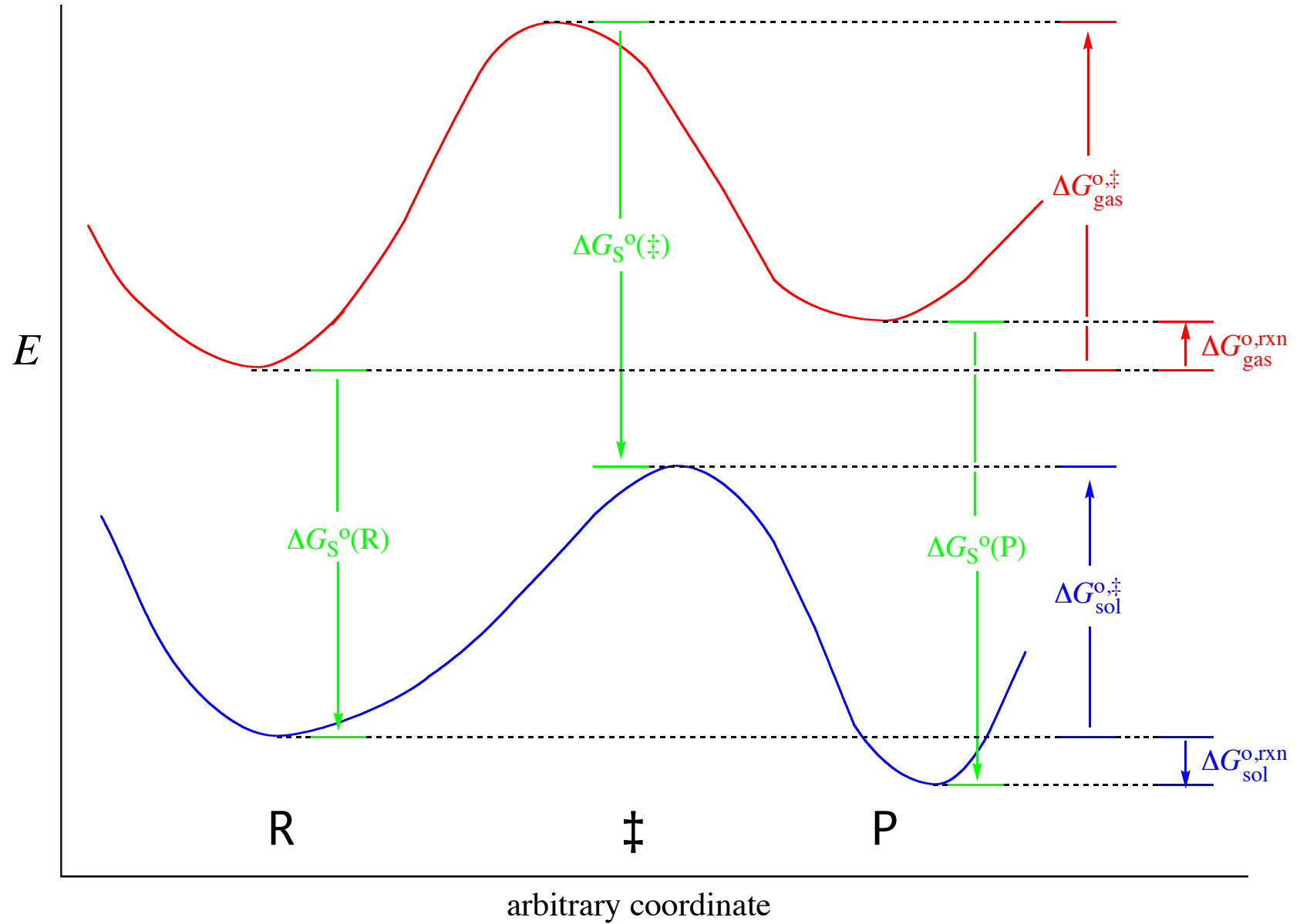
Related Potential Energy Surfaces



PES Slice — The Menshutkin Reaction



The Generic Free-Energy Cycle



Explicit and Implicit Solvent Modeling: Two Alternative Approaches

- Explicit solvent modeling is conceptually obvious: add lots of solvent molecules and compute blue curve on last slide
- Implicit solvent modeling is more subtle: forget about molecular representation of solvent; compute gas-phase curve by one method, then compute free energies of solvation at points of interest by a different method (this can actually be done in a blue-curve fashion too, but the solvent remains implicit)

(Equilibrium) Free Energy of Solvation

$$\Delta G_S^{\circ} = \Delta G_{ENP} + G_{CDS}^{\circ}$$

SOLVENT DEPENDENCE

E \Rightarrow Electronic Energy

N \Rightarrow Nuclear Repulsion

P \Rightarrow Solute-Solvent Polarization

} Solvent Dielectric

C \Rightarrow Cavitation

D \Rightarrow Dispersion

S \Rightarrow Structural etc.

} Other Solvent Properties
(also called “non-electrostatic” component)

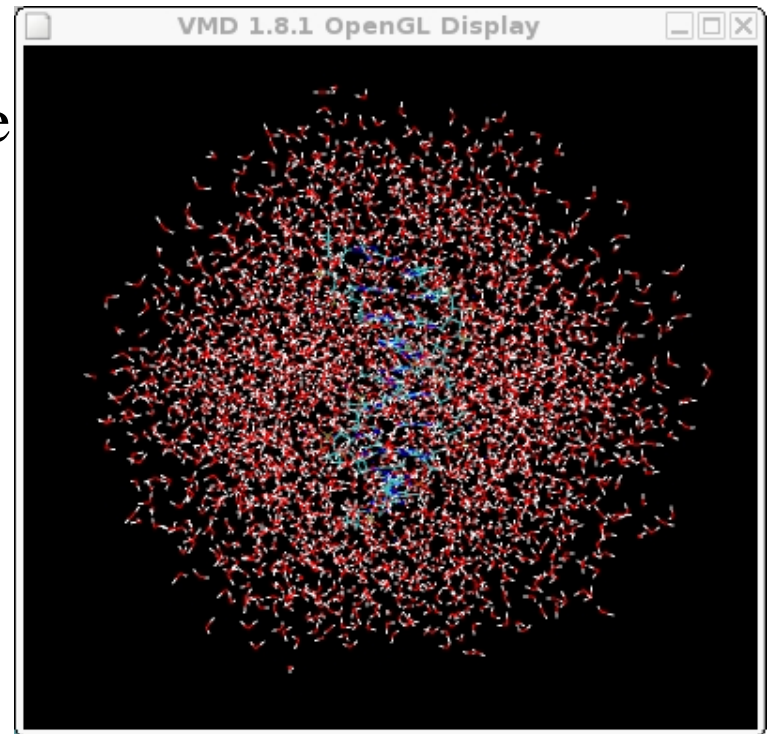
Solvation (Condensed Phase) Models

Explicit Solvent—Atomistic Analysis

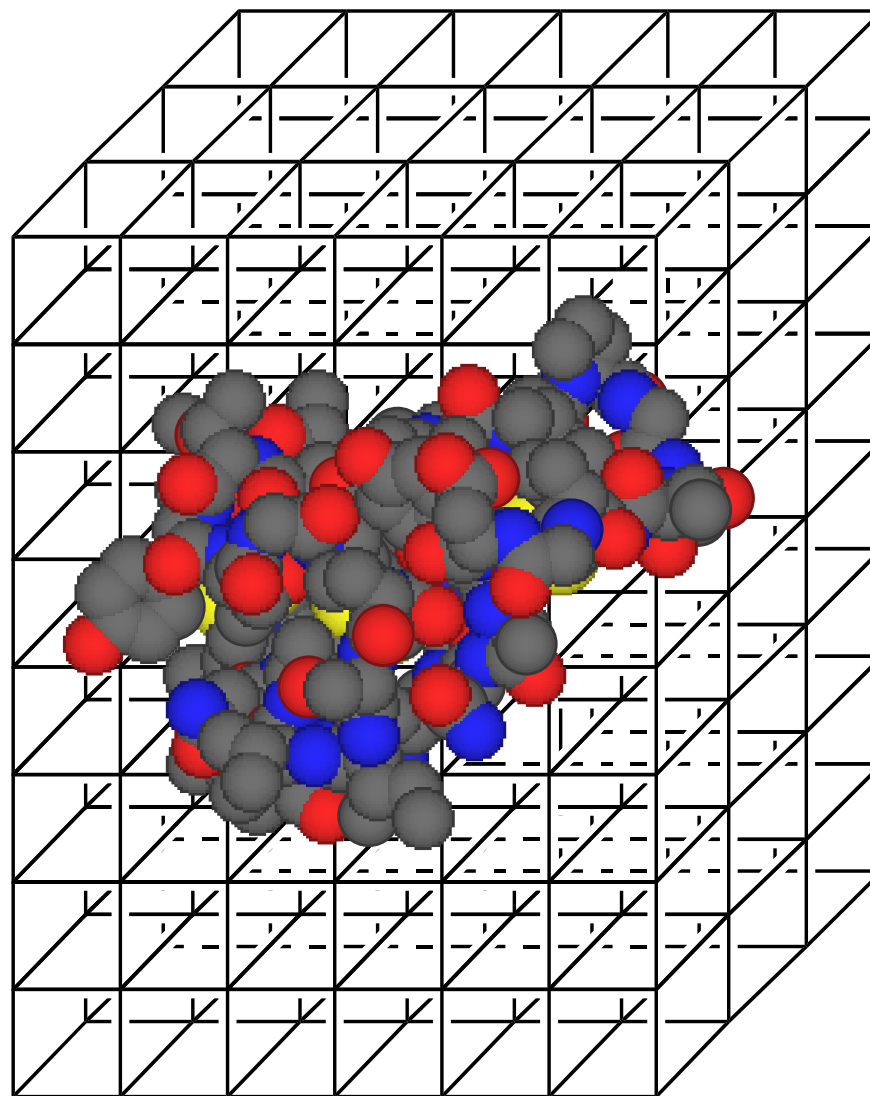
Video VI.ii

Some Rules for Explicit Solvent Modeling

- Rule 1: It takes a *lot* of solvent molecules to look like a solution. Put differently, clusters only tell one about clusters.
- Consequence: Quantum mechanics is very, very expensive (although use of Car-Parinello approach is ongoing). Instead, molecular mechanics (i.e., force field) approaches tend to be used for at least *some* of the system.
- Tools: Periodic boxes or Ewald sums to limit system size. QM/MM implementations.

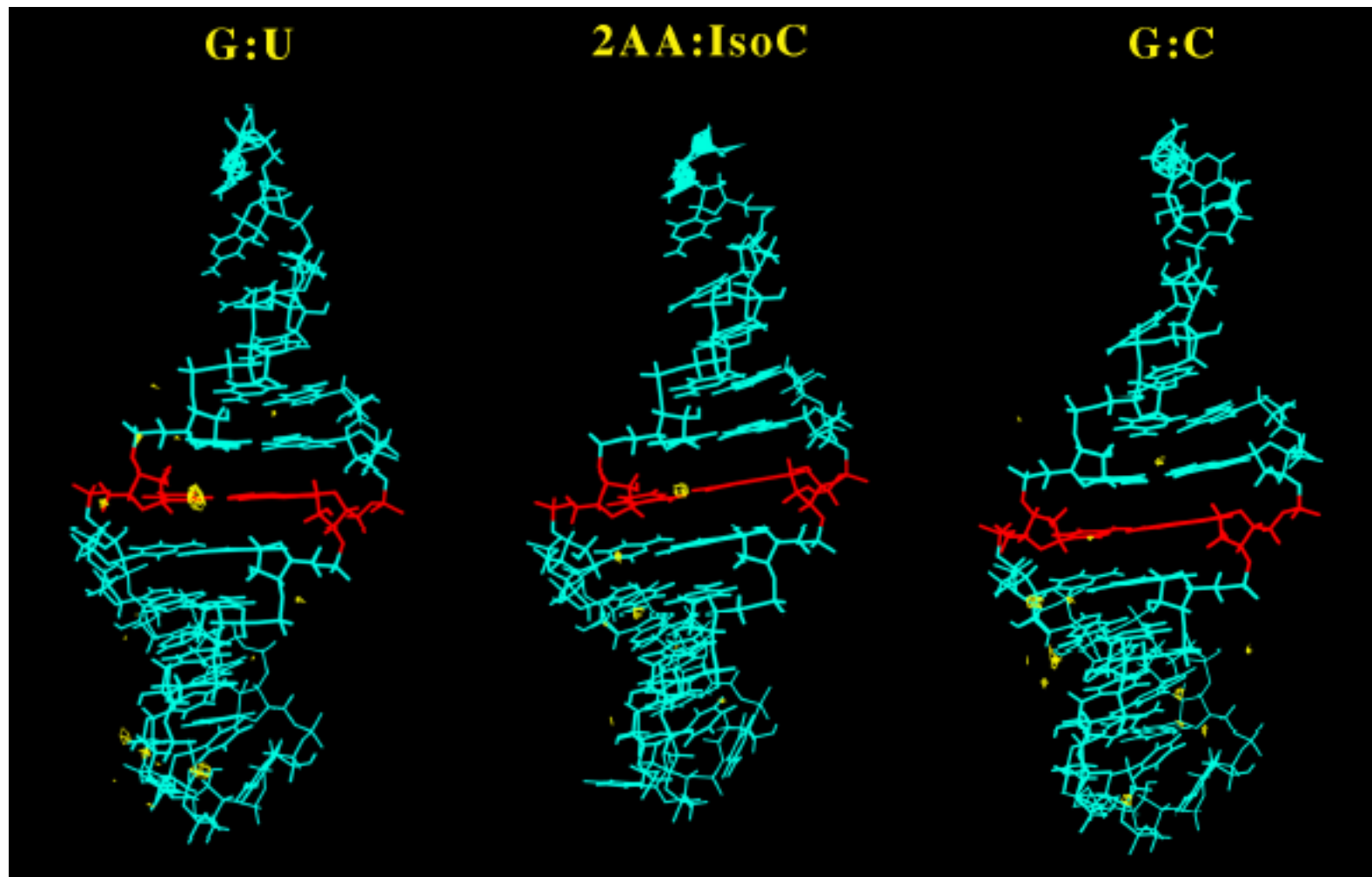


But It *Is* There! Solvent Density Analysis



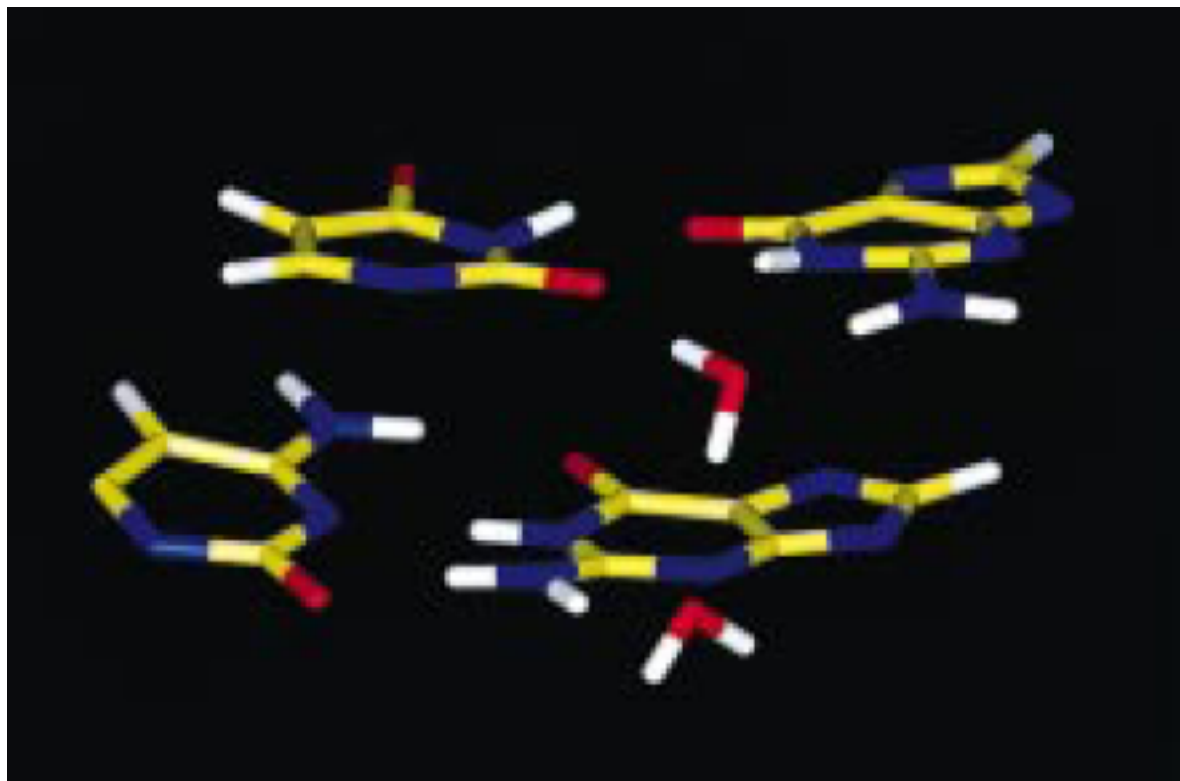
Nagan, et al. *J. Am. Chem. Soc.* **1999**, *121*, 7310.

But It *Is* There! Solvent Density Analysis



Nagan, et al. *J. Am. Chem. Soc.* **1999**, *121*, 7310.

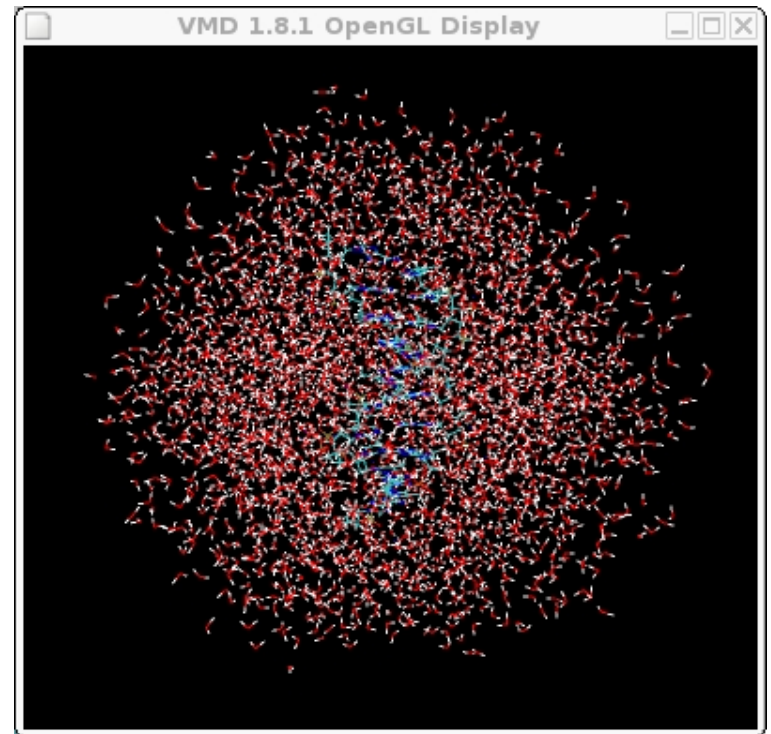
But It *Is* There! Solvent Density Analysis



Nagan, et al. *J. Am. Chem. Soc.* **1999**, *121*, 7310.

Some Rules for Explicit Solvent Modeling

- Rule 2: Equilibrium properties (e.g., free energies) require averaging over phase space.
- Consequence: Sampling phase space becomes a key issue. Brute force is impossible in a real system. Free energy convergence can be very slow.
- Tools: Monte Carlo or Molecular Dynamics sampling until apparent (ergodic) convergence. More robust if multiple trajectories are run from different starting points.



Integrating over Phase Space

$$\langle \Xi \rangle = \frac{\int_{\text{PS}} \Xi(\mathbf{r}) P(\mathbf{r}) d\mathbf{r}}{\int_{\text{PS}} P(\mathbf{r}) d\mathbf{r}}$$

Expectation values (whether quantum or classical) are dictated by the relative probabilities of being in different regions of phase space

$$P(\mathbf{r}) = e^{-E(\mathbf{q}, \mathbf{p})/k_B T} \quad Q = \int_{\text{PS}} P(\mathbf{r}) d\mathbf{r}$$

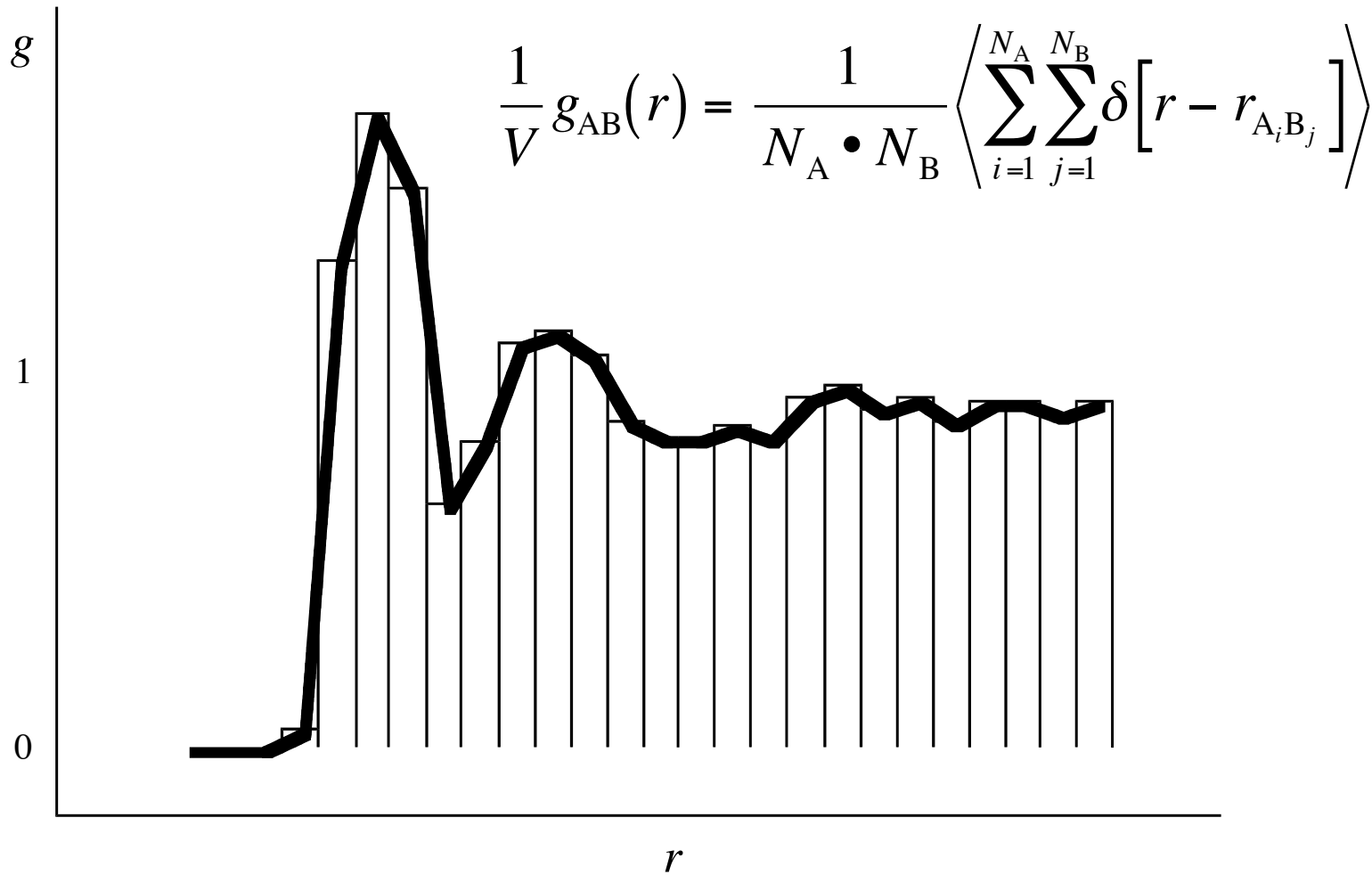
Key point: Don't waste time evaluating $\Xi(\mathbf{r})$ if $P(\mathbf{r})$ is zero.

Difficulty: Phase space is $6N$ -dimensional. If you only want to sample all possible combinations of either positive or negative values for each coordinate (i.e., hit every “hyperoctant” in phase space *once*), you need 2^{6N} points!

MC/MD Provides Ready Access to Many Properties in Solution

- Average structures (with standard deviations); note that these structures may be determined with constraints imposed (e.g., from NMR NOE measurements)
- All kinds of quantum mechanical properties (albeit usually averaged over fewer snapshots than the entire trajectory)
- Structural details associated with the solvation shell (typically NOT available from implicit models)

Radial Distribution Function



Solvation (Condensed Phase) Models

Explicit Solvent—Free Energy Perturbation

Video VI.iii

Integrating over Phase Space

$$\langle \Xi \rangle = \frac{\int_{\text{PS}} \Xi(\mathbf{r}) P(\mathbf{r}) d\mathbf{r}}{\int_{\text{PS}} P(\mathbf{r}) d\mathbf{r}}$$

Expectation values are dictated by the relative probabilities of being in different regions of phase space

$$P(\mathbf{r}) = e^{-E(\mathbf{q}, \mathbf{p})/k_B T} \quad Q = \int_{\text{PS}} P(\mathbf{r}) d\mathbf{r}$$

Key point: Don't waste time evaluating $\Xi(\mathbf{r})$ if $P(\mathbf{r})$ is zero.

Difficulty: Phase space is $6N$ -dimensional. If you only want to sample all possible combinations of either positive or negative values for each coordinate (i.e., hit every “hyperoctant” in phase space *once*), you need 2^{6N} points!

What About Free Energy as the Average Property?

$$\begin{aligned} A &= k_B T \ln \frac{1}{Q} \\ &= k_B T \ln \left[\frac{\iint e^{E(\mathbf{q}, \mathbf{p}) / k_B T} e^{-E(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}}{\iint e^{-E(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}} \right] \\ &= k_B T \ln \left[\iint e^{E(\mathbf{q}, \mathbf{p}) / k_B T} P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} \right] \\ &= k_B T \ln \left\langle e^{E / k_B T} \right\rangle \end{aligned}$$

$$\begin{aligned} \langle A \rangle_B - \langle A \rangle_A &= k_B T \ln \left\langle e^{E_B / k_B T} \right\rangle_B - k_B T \ln \left\langle e^{E_A / k_B T} \right\rangle_A \\ &= k_B T \ln \left(\frac{1}{M_B} \sum_i^{M_B} e^{E_i / k_B T} \right) - k_B T \ln \left(\frac{1}{M_A} \sum_i^{M_A} e^{E_i / k_B T} \right) \end{aligned}$$

Horribly slowly convergent for two structures A and B because sampling procedures are designed to sample predominantly low-energy points, but high-energy points contribute exponentially more to sum.

A Trick With the Integral

$$\begin{aligned}
 A_B - A_A &= k_B T \ln \frac{Q_A}{Q_B} = -k_B T \ln \frac{Q_B}{Q_A} = -k_B T \ln \left[\frac{\iint e^{-E_B(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}}{\iint e^{-E_A(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}} \right] \\
 &= -k_B T \ln \left[\frac{\iint e^{-E_B(\mathbf{q}, \mathbf{p}) / k_B T} \left[e^{E_A(\mathbf{q}, \mathbf{p}) / k_B T} e^{-E_A(\mathbf{q}, \mathbf{p}) / k_B T} \right] d\mathbf{q} d\mathbf{p}}{\iint e^{-E_A(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}} \right] \\
 &= -k_B T \ln \left[\frac{\iint \left[e^{-E_B(\mathbf{q}, \mathbf{p}) / k_B T} e^{E_A(\mathbf{q}, \mathbf{p}) / k_B T} \right] e^{-E_A(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}}{\iint e^{-E_A(\mathbf{q}, \mathbf{p}) / k_B T} d\mathbf{q} d\mathbf{p}} \right] \\
 &= -k_B T \ln \left[\iint e^{-[E_B(\mathbf{q}, \mathbf{p}) - E_A(\mathbf{q}, \mathbf{p})] / k_B T} P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} \right] = -k_B T \left\langle e^{-(E_B - E_A) / k_B T} \right\rangle_A
 \end{aligned}$$

Thus, sampling over coordinates for A, determine exponential of energy difference between A and B.

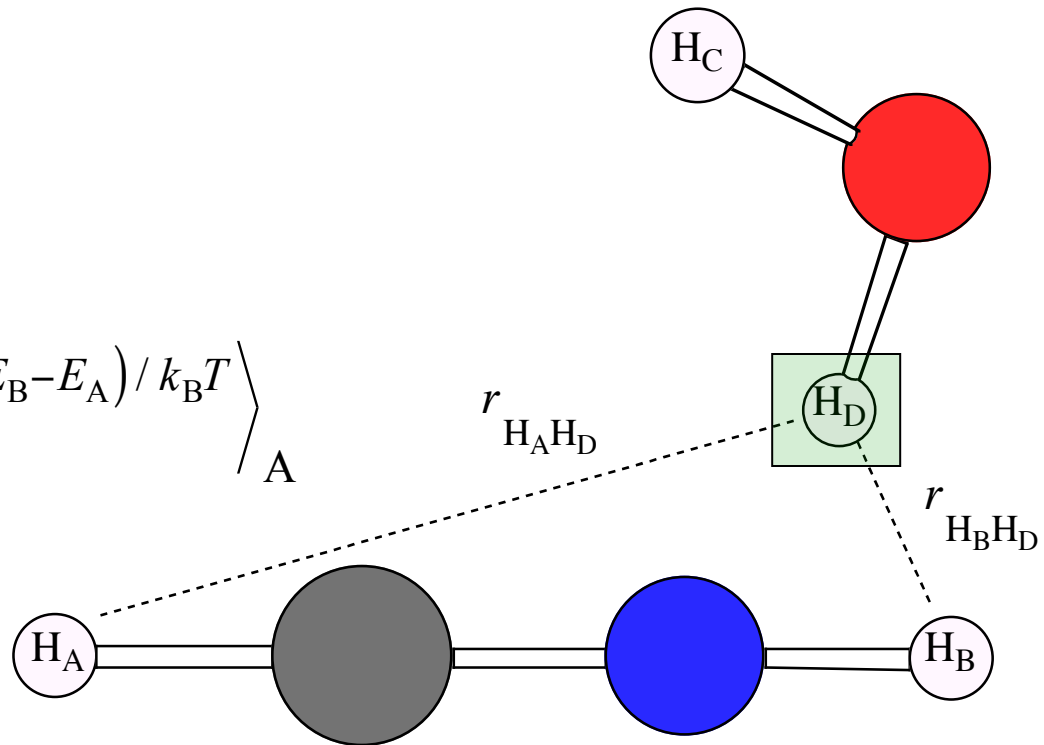
Simpler because only a single ensemble, but what if sample over A is not ergodic for B?

How Can We Ensure That A and B Are “Similar”?

e.g., HCN vs. HNC

$$\Delta E_{H_D} = \left(\frac{a_{HH}}{r_{H_B H_D}^{12}} - \frac{b_{HH}}{r_{H_B H_D}^6} + \frac{q_{H_B} q_{H_D}}{\epsilon r_{H_B H_D}} \right) - \left(\frac{a_{HH}}{r_{H_A H_D}^{12}} - \frac{b_{HH}}{r_{H_A H_D}^6} + \frac{q_{H_A} q_{H_D}}{\epsilon r_{H_A H_D}} \right)$$

$$\langle A \rangle_B - \langle A \rangle_A = -k_B T \ln \left\langle e^{-(E_B - E_A)/k_B T} \right\rangle_A$$



$$E(\lambda) = \lambda E_B + (1 - \lambda) E_A$$

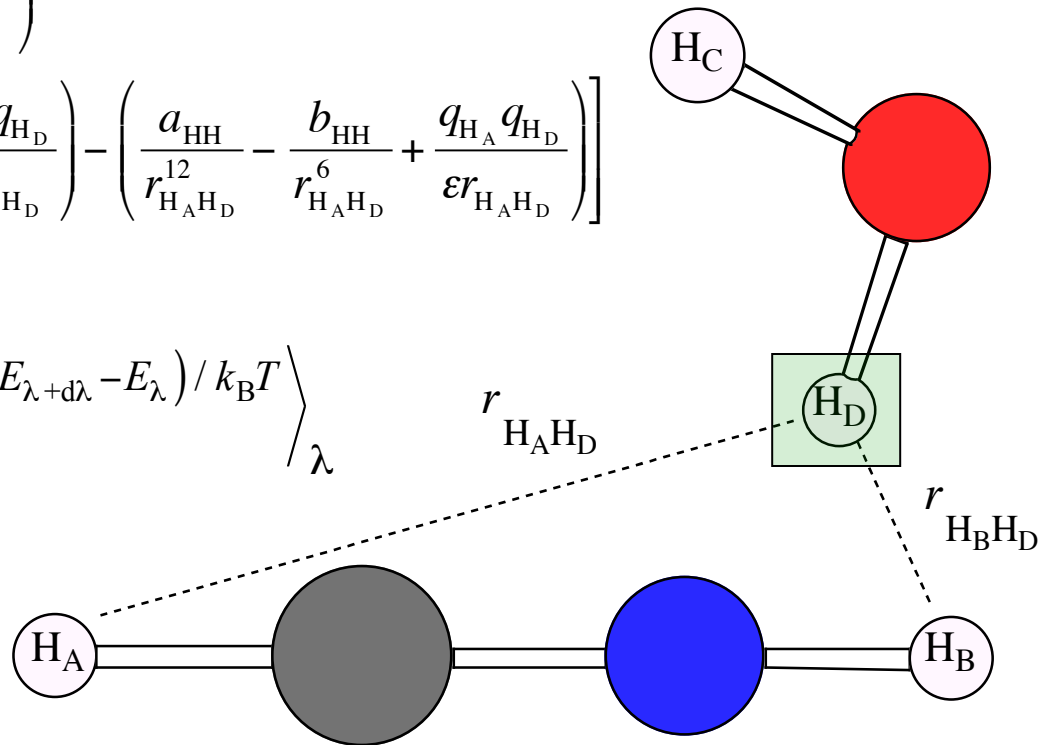
Free Energy Perturbation

How Can We Ensure That A and B Are “Similar”?

e.g., HCN vs. HNC

$$\begin{aligned} \Delta E_{H_D} &= 0.05 \left(\frac{a_{HH}}{r_{H_B H_D}^{12}} - \frac{b_{HH}}{r_{H_B H_D}^6} + \frac{q_{H_B} q_{H_D}}{\epsilon r_{H_B H_D}} \right) + 0.95 \left(\frac{a_{HH}}{r_{H_A H_D}^{12}} - \frac{b_{HH}}{r_{H_A H_D}^6} + \frac{q_{H_A} q_{H_D}}{\epsilon r_{H_A H_D}} \right) \\ &\quad - \left(\frac{a_{HH}}{r_{H_A H_D}^{12}} - \frac{b_{HH}}{r_{H_A H_D}^6} + \frac{q_{H_A} q_{H_D}}{\epsilon r_{H_A H_D}} \right) \\ &= 0.05 \left[\left(\frac{a_{HH}}{r_{H_B H_D}^{12}} - \frac{b_{HH}}{r_{H_B H_D}^6} + \frac{q_{H_B} q_{H_D}}{\epsilon r_{H_B H_D}} \right) - \left(\frac{a_{HH}}{r_{H_A H_D}^{12}} - \frac{b_{HH}}{r_{H_A H_D}^6} + \frac{q_{H_A} q_{H_D}}{\epsilon r_{H_A H_D}} \right) \right] \end{aligned}$$

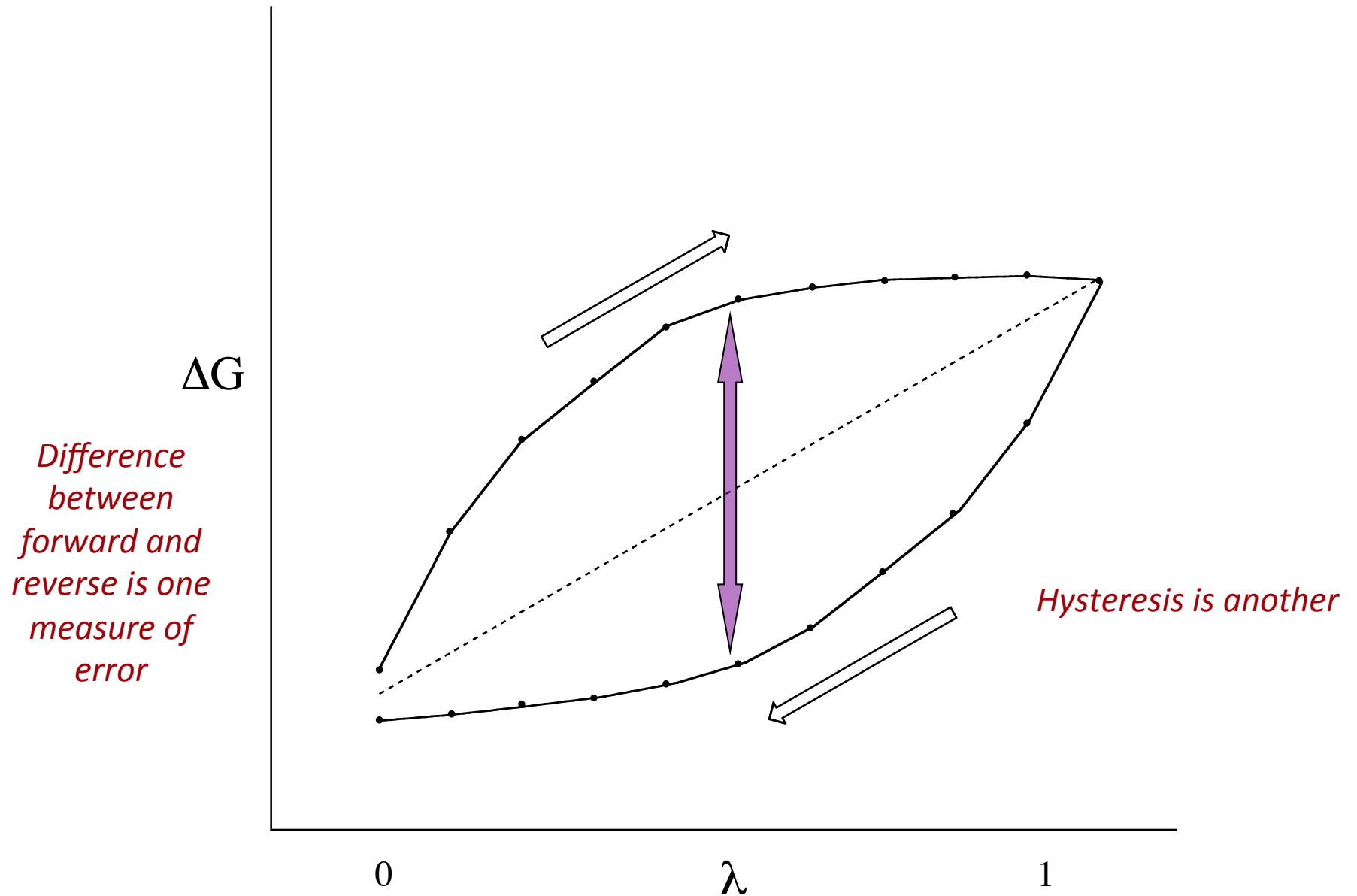
$$\langle A \rangle_B - \langle A \rangle_A = -k_B T \sum_{\lambda=0}^1 \ln \left\langle e^{-\frac{(E_{\lambda+d\lambda} - E_{\lambda})}{k_B T}} \right\rangle_{\lambda}$$



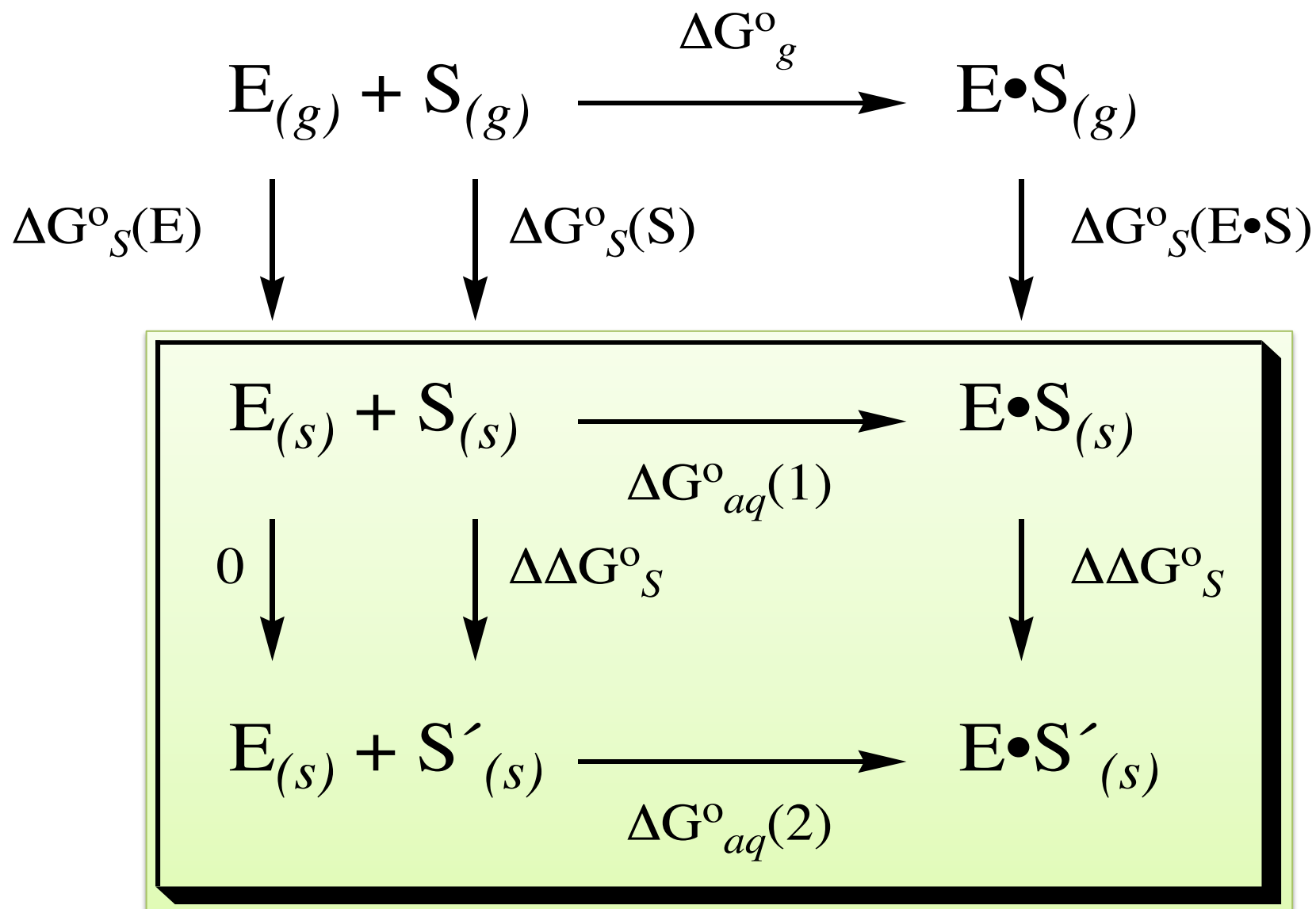
$$E(\lambda) = \lambda E_B + (1 - \lambda) E_A$$

Free Energy Perturbation

Monitoring the Alchemical Change



Computational Alchemy



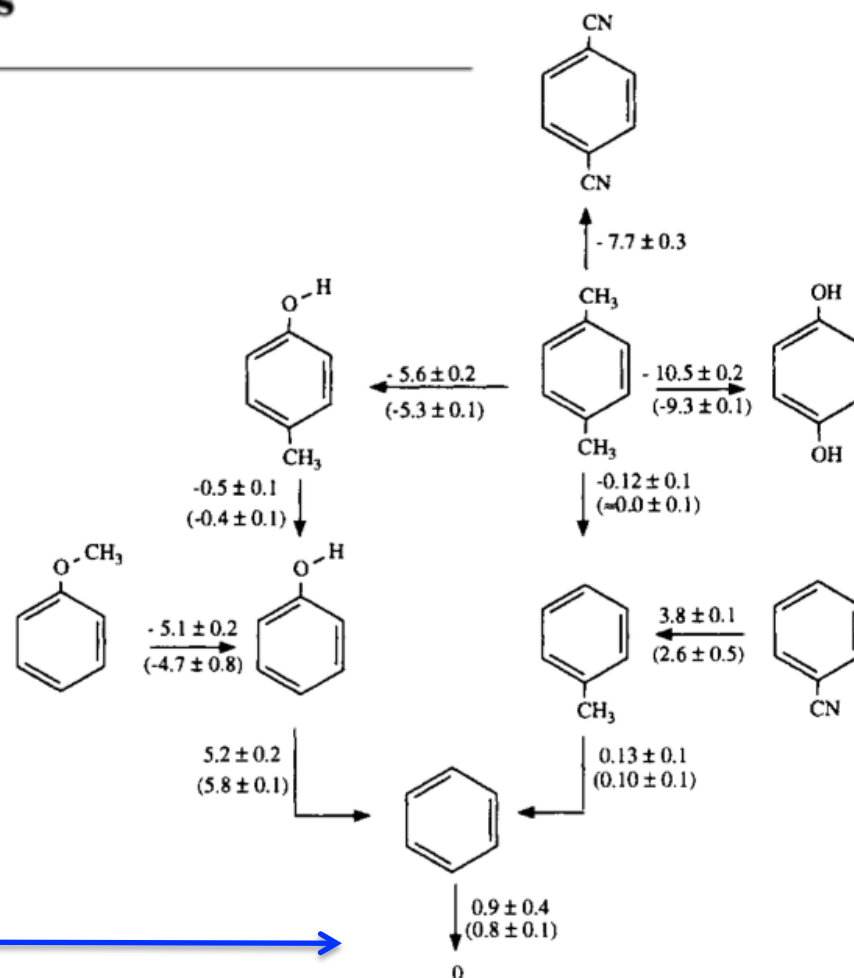
Computational Alchemy

Monte Carlo Simulations of the Hydration of Substituted Benzenes with OPLS Potential Functions

William L. Jorgensen* and Toan B. Nguyen

Department of Chemistry, Yale University, New Haven, Connecticut 06511

Journal of Computational Chemistry, Vol. 14, No. 2, 195–205 (1993)
© 1993 by John Wiley & Sons, Inc.



Note that perturbation to nothing (annihilation) is allowed, although some care must be taken for so drastic a perturbation.

Figure 1. Computed and experimental (in parentheses) relative free energies of hydration in kcal/mol.

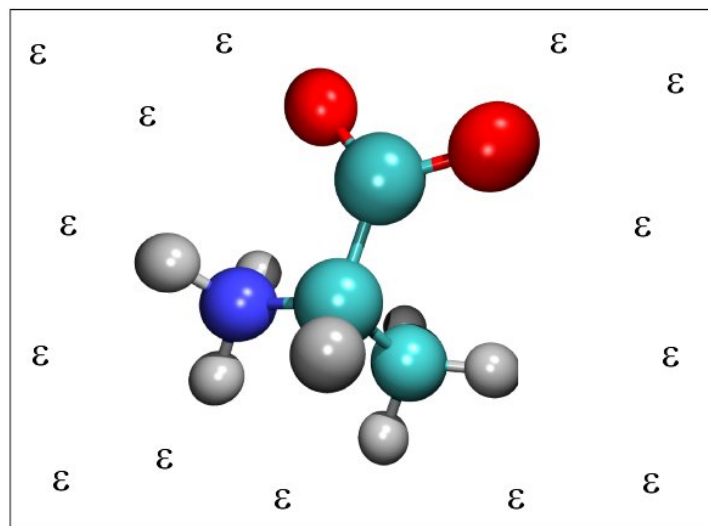
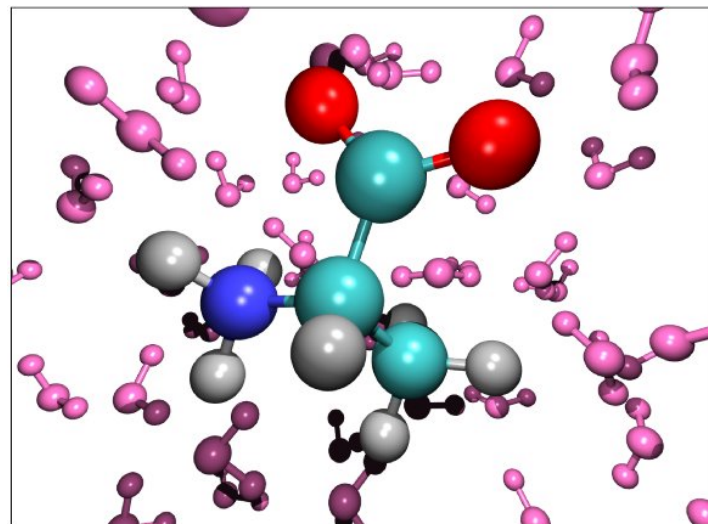
Solvation (Condensed Phase) Models

Implicit Solvent—Electrostatics

Video VI.iv

Some Rules for Implicit Solvent Modeling

- Rule 1: If one replaces the solvent molecules with a continuum dielectric (an equilibrium averaging, in essence) one is left with a system no “larger” than the solute.
- Consequence: All solvent-structural information is lost, but if one can afford QM for gas phase, one can afford QM for solution. Polarization now arises *from first principles* (self-consistent reaction field—SCRF—induced in the medium).
- Tools: Numerous methods to solve or approximate the Poisson equation for continuous or discrete charge representations.



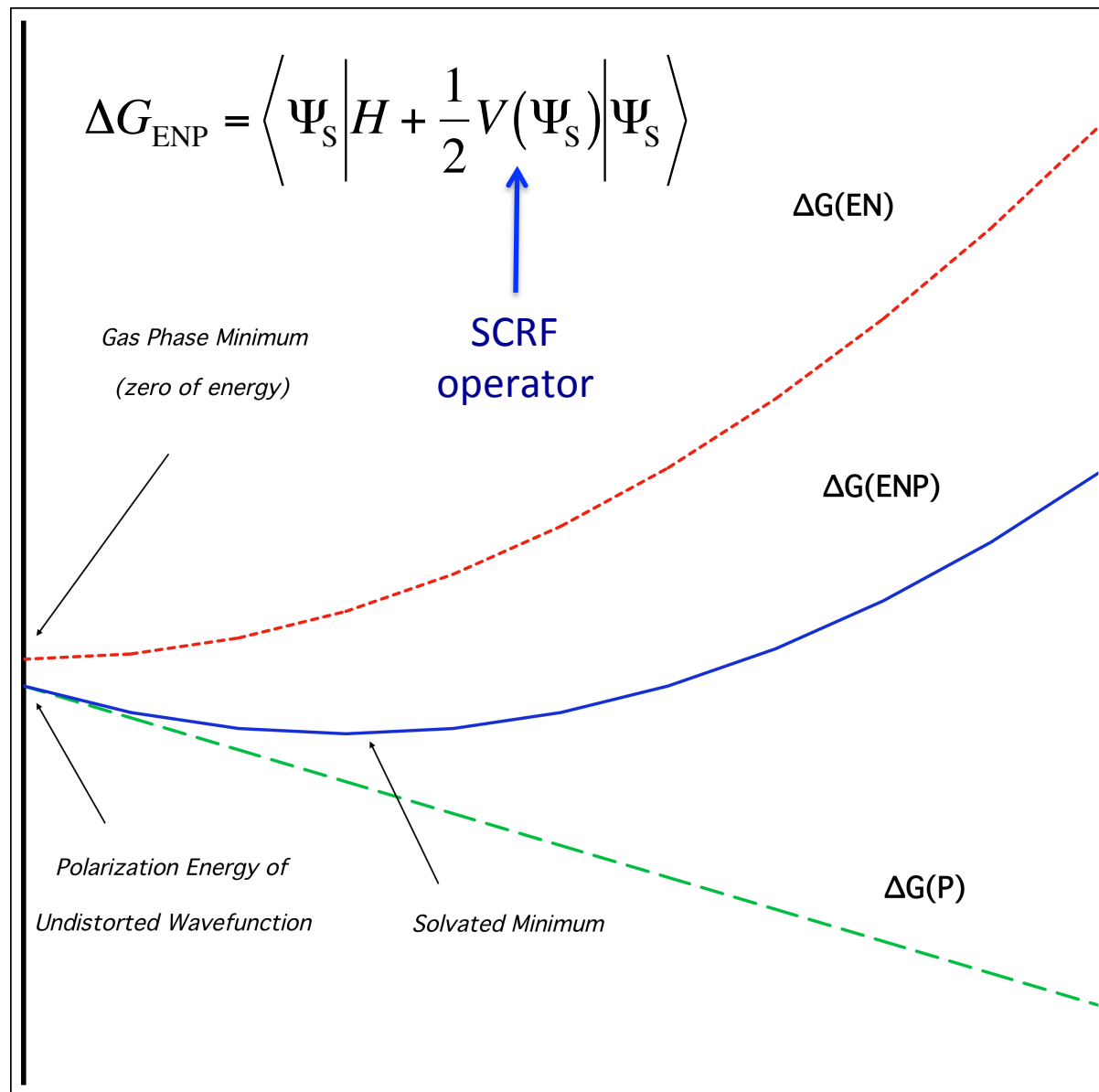
Some Rules for Implicit Solvent Modeling

- Rule 1: If one replaces the solvent molecules with a continuum dielectric (an equilibrium averaging, in essence) one is left with a system no “larger” than the solute.
- Consequence: All solvent-structural information is lost, but if one can afford QM for gas phase, one can afford QM for solution. Polarization now arises *from first principles* (self-consistent reaction field—SCRF—induced in the medium).
- Tools: Numerous methods to solve or approximate the **Poisson equation** for continuous or discrete charge representations.

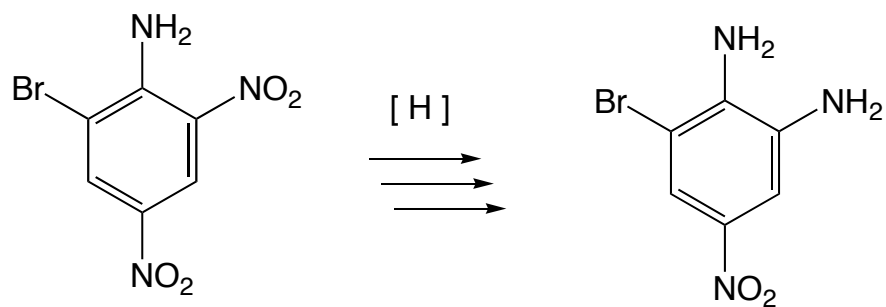
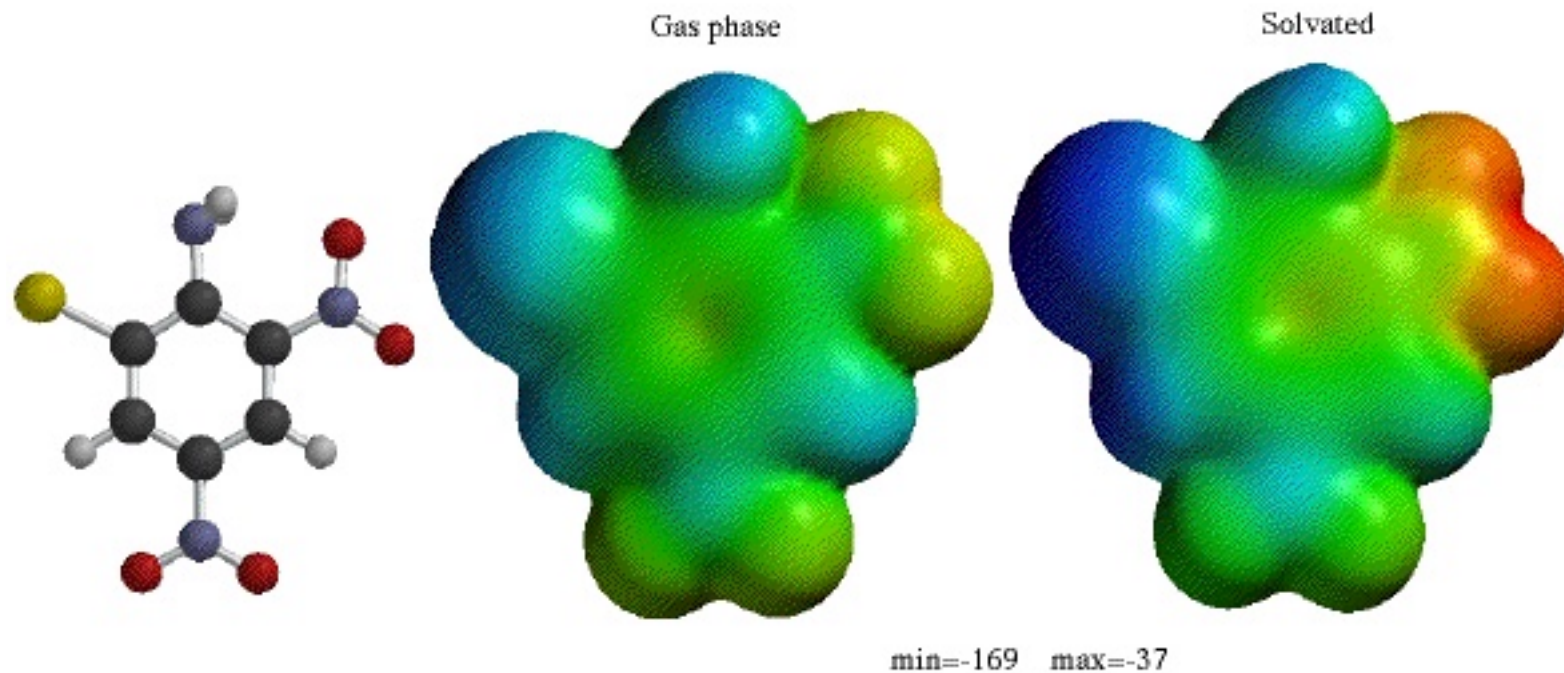
$$G = -\frac{1}{2} \int \rho(\mathbf{r})\phi(\mathbf{r})d\mathbf{r}$$

$$\nabla\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r})$$

Solvent-induced Polarization



Polarization Example—Nitroaromatic Radical Anion



Conducting Sphere (Ion) Example



Sphere carries charge q and has radius α .

Charge distribution on a conducting sphere is $\rho(\mathbf{s}) = \frac{q}{4\pi\alpha^2}$, \mathbf{s} on surface

Potential outside sphere is $\phi(\mathbf{r}) = -\frac{q}{\epsilon|\mathbf{r}|}$

Thus, solving for *the work of charging* requires integrating only over the surface of the sphere:

$$\begin{aligned} G &= -\frac{1}{2} \int_s \rho(\mathbf{s}) \phi(\mathbf{s}) d\mathbf{s} \\ &= \frac{1}{2} \int_s \frac{q}{4\pi\alpha^2} \frac{q}{\epsilon\alpha} d\mathbf{s} \\ &= \frac{1}{2} \frac{q^2}{\epsilon\alpha} \end{aligned}$$

and the free energy of polarization G_p is the difference in the work of charging in solution ($\epsilon \neq 1$) and the gas phase ($\epsilon = 1$)

$$G_p = -\frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \frac{q^2}{\alpha} \quad \text{the Born equation}$$

Dipole in a Sphere Example (Kirkwood-Onsager)



Sphere carries dipole μ at center and has radius α .

Analogous analysis leads to
$$G_P = -\frac{1}{2} \left(\frac{2(\epsilon - 1)}{(2\epsilon + 1)} \right) \frac{\mu^2}{\alpha^3}$$

This leads to the Schrödinger equation:
$$\left\{ H - \frac{1}{2} \left[\frac{2(\epsilon - 1)}{(2\epsilon + 1)} \right] \frac{\langle \Psi | \mu | \Psi \rangle}{\alpha^3} \mu \right\} \Psi = E \Psi$$

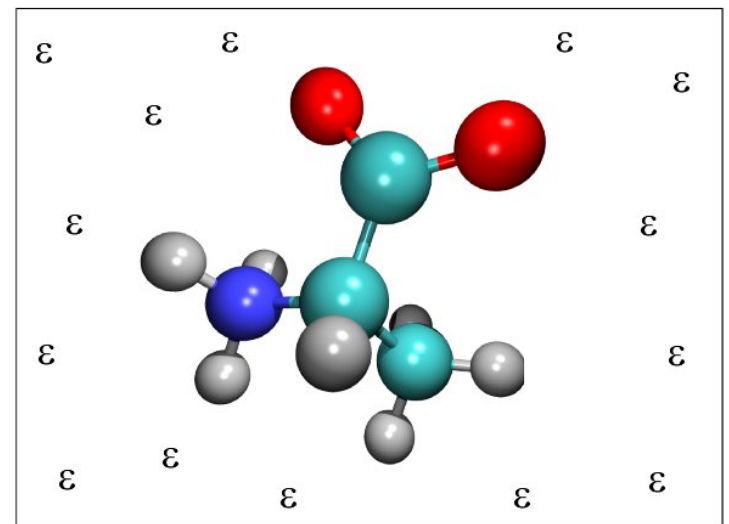
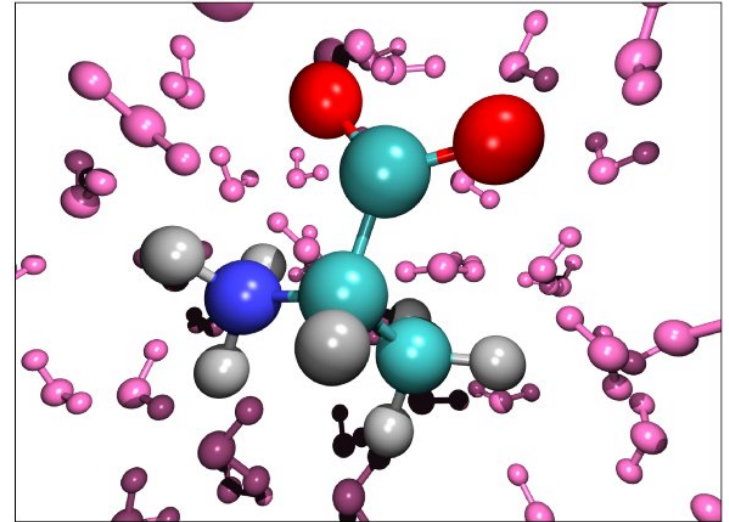
which is variationally minimized by a Slater determinant formed from solvated orbitals that are eigenfunctions of:

$$\left\{ F_i - \left[\frac{2(\epsilon - 1)}{(2\epsilon + 1)} \right] \frac{\langle \Psi | \mu | \Psi \rangle^2}{\alpha^3} \right\} \psi_i = \epsilon_i \psi_i \quad \text{SCRF}$$

Note that the solvated orbitals and associated eigenvalues will be different than their gas-phase counterparts

Some Rules for Implicit Solvent Modeling

- Rule 2: A continuum dielectric is a fiction, so it is best not to get too caught up in theoretical rigor (ΔG_{ENP} is not even a physical observable...)
- Consequence: Construction of the dielectric cavity and/or methods for solving or approximating the Poisson equation can vary significantly from one model to the next.
- Tools: Parameterization just as important as for force-field models (even if occasionally it is stealth-parameterization).



Cavities and Their Fillings

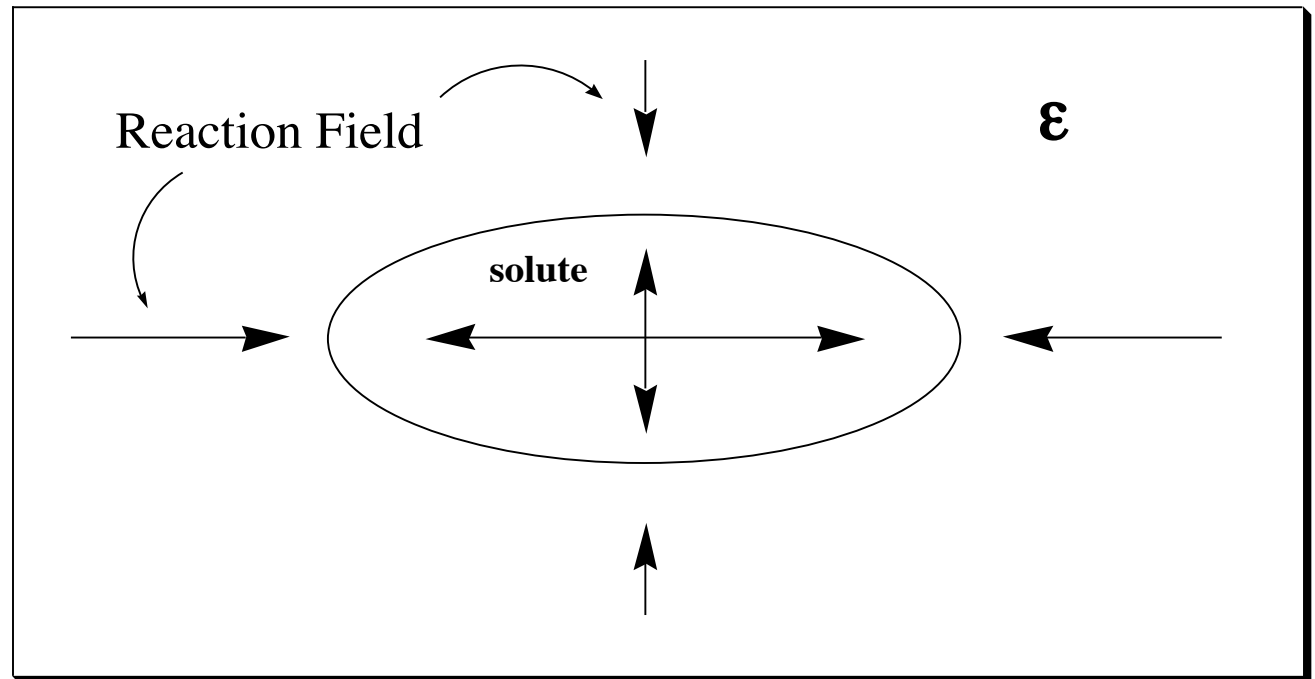
Ideal Cavities

Spheres and ellipsoids: Permit analytic solution of Poisson equation for interior charge expressed as multipole expansion

Charge in a sphere: Born equation

$$\Delta G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \frac{q^2}{\alpha}$$

q is charge and α is radius of cavity sphere



Cavities and Their Fillings

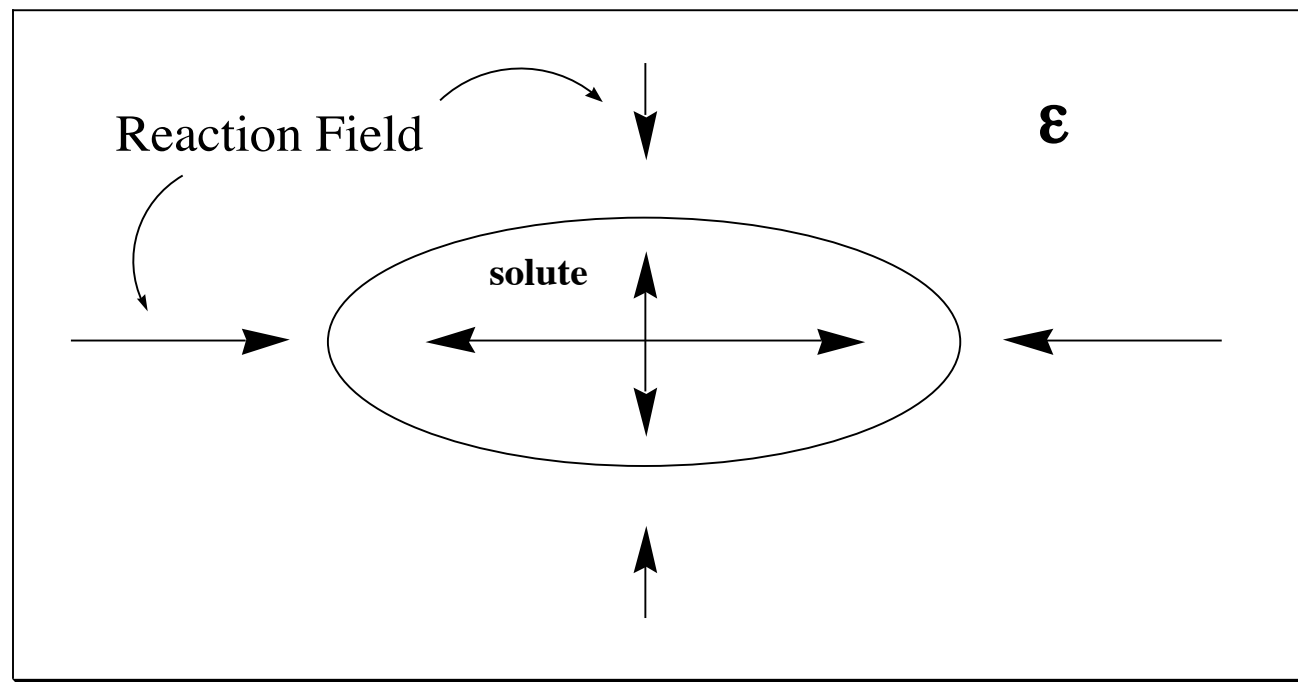
Ideal Cavities

Spheres and ellipsoids: Permit analytic solution of Poisson equation for interior charge expressed as multipole expansion

Dipole in a sphere: Kirkwood-Onsager equation

$$\Delta G_P = -\frac{1}{2} \left[\frac{2(\epsilon - 1)}{(2\epsilon + 1)} \right] \frac{\mu^2}{\alpha^3}$$

μ is dipole moment
and α is radius of
cavity sphere;
requires SCRF since
relaxation of Ψ affects
 μ



Cavities and Their Fillings

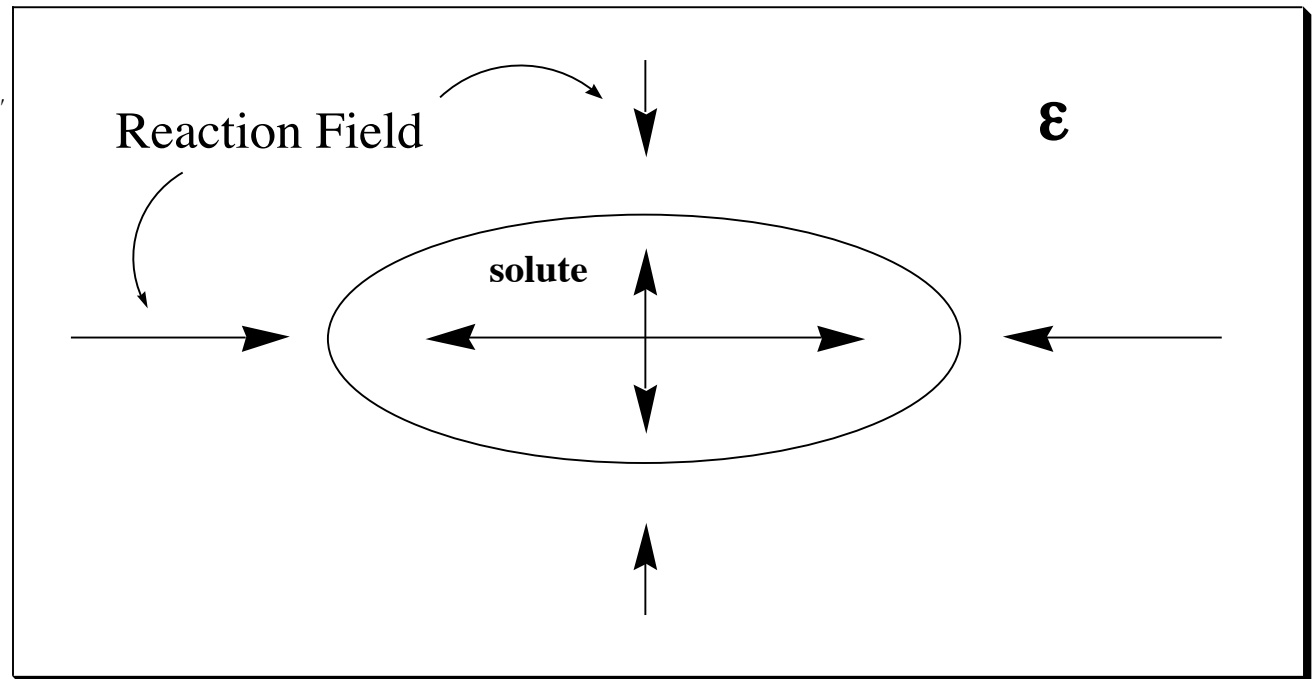
Ideal Cavities

Spheres and ellipsoids: Permit analytic solution of Poisson equation for interior charge expressed as multipole expansion

Any number of multipoles in a sphere or ellipse:
Rivail and Rinaldi

$$G_P = -\frac{1}{2} \sum_{l=0}^L \sum_{m=-l}^l \sum_{l'=0}^L \sum_{m'=-l'}^{l'} M_l^m f_{ll'}^{mm'} M_{l'}^{m'}$$

M is solute or reaction field multipole moment (order *l* component *m*) and *f* is a “reaction-field factor”



Cavities and Their Fillings

Ideal Cavities

Spheres and ellipsoids: Permit analytic solution of Poisson equation for interior charge expressed as multipole expansion

But such cavities tend to be very unrealistic for arbitrarily shaped molecules

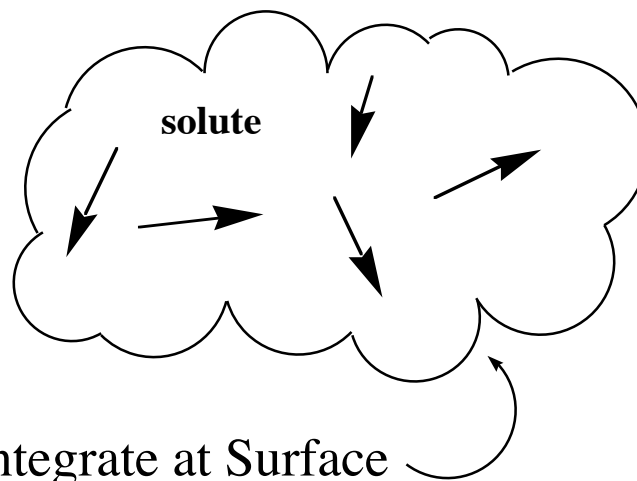
Cavities and Their Fillings

Arbitrary Cavities

Arbitrary cavities require non-analytic or approximate solutions of the Poisson equation for interior charge expressed as either a continuous charge distribution or a single- or multicenter multipole expansion

Continuous
distribution: PCM
(MST) of Tomasi and
many coworkers (also
COSMO)

Can Integrate through Volume



Can Integrate at Surface

ϵ

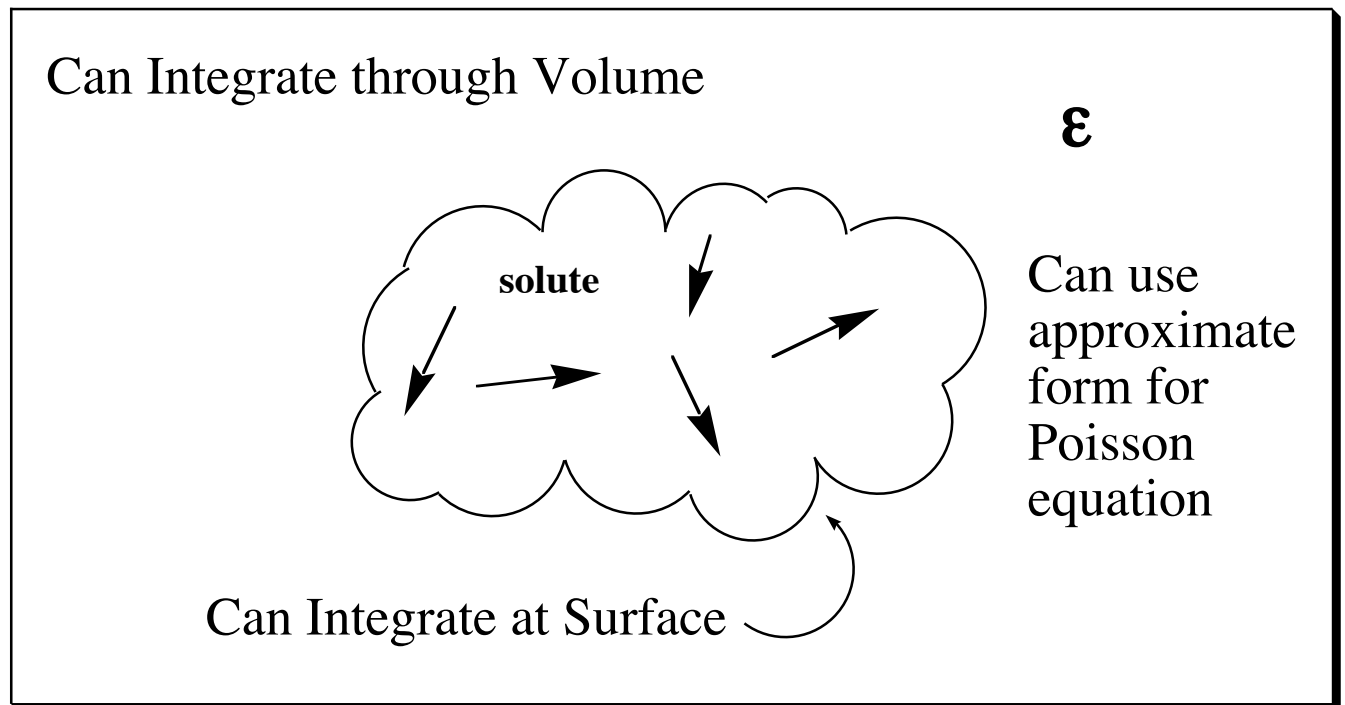
Can use
approximate
form for
Poisson
equation

Cavities and Their Fillings

Arbitrary Cavities

Arbitrary cavities require non-analytic or approximate solutions of the Poisson equation for interior charge expressed as either a continuous charge distribution or a single- or multicenter multipole expansion

Multipole expansion:
Rivail and many
coworkers



Cavities and Their Fillings

Arbitrary Cavities

Arbitrary cavities require non-analytic or approximate solutions of the Poisson equation for interior charge expressed as either a continuous charge distribution or a single- or multicenter multipole expansion

$$G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{k,k'}^{\text{atoms}} q_k q_{k'} \gamma_{kk'}$$

Generalized Born: GB
(Still and subsequent
others, classical) and
SMx (Cramer &
Truhlar, quantum
SCRF)

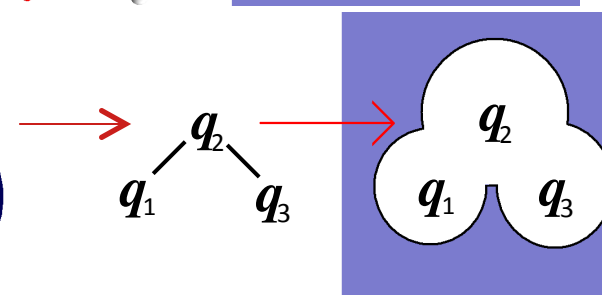
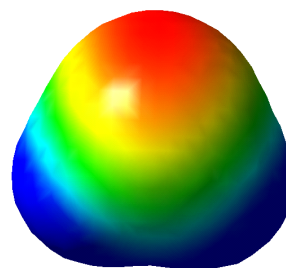
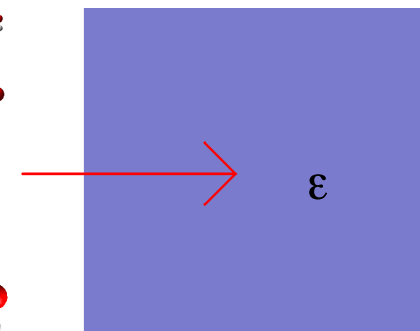
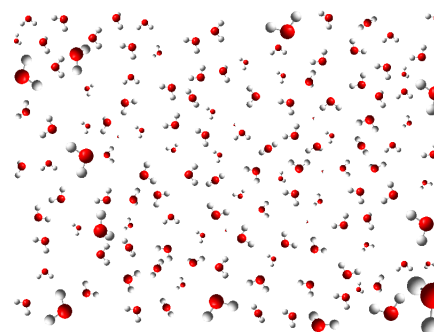
$$\gamma_{kk'} = \left(r_{kk'}^2 + \alpha_k \alpha_{k'} e^{-r_{kk'} / d_{kk'}} \right)^{-1/2}$$

Bulk Electrostatic Effects

Generalized Born (GB) equation

$$G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \sum_{kk'}^{\text{atoms}} q_k \gamma_{kk'} q_{k'}$$

$$\gamma_{kk'} = \left(r_{kk'}^2 + \alpha_k \alpha_{k'} e^{-r_{kk'}/d} \right)^{-1/2}$$



Limiting behaviors...

$$r_{kk'} \gg 0$$

$$G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \frac{q_k q_{k'}}{r_{kk'}}$$

Coulomb's Law
 $^{1/2}(-\text{gas} + \text{solution})$

$$r_{kk'} = 0$$

$$G_P = -\frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \frac{q_k^2}{\alpha_k}$$

Born's Equation
 monatomic ion

Avoids charge penetration and can be put into pairwise form (advantages) but requires partial atomic charges (disadvantage). Now heavily used for classical simulations, less for quantum calculations.

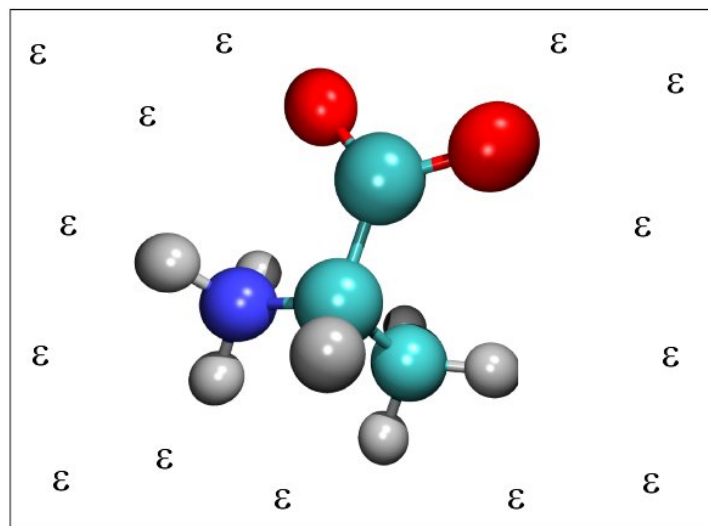
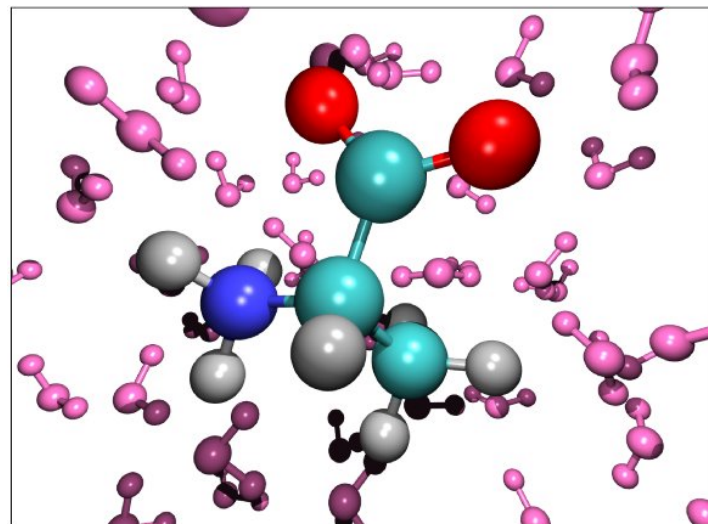
Solvation (Condensed Phase) Models

Implicit Solvent—Non-electrostatics

Video VI.v

Some Rules for Implicit Solvent Modeling

- Rule 3: Electrostatics are only *part* of the free energy of solvation.
- Consequence: One needs to somehow account for cavitation, dispersion, solvent-structural changes, etc., if one wants to make contact with the experimental observable.
- Tools: It's nice if you can also fix up your electrostatic approximations at the same time

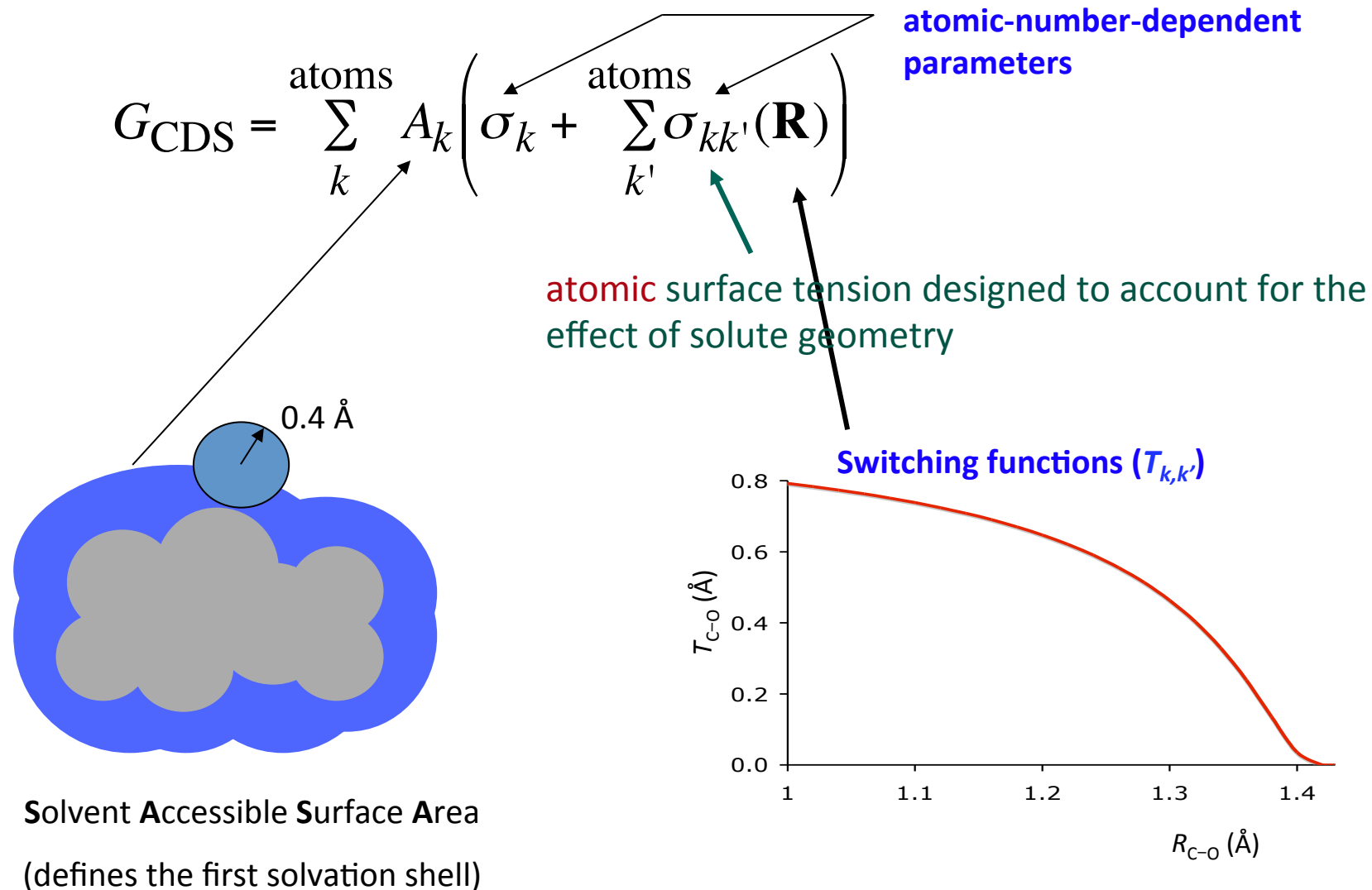


How to Account for Nonelectrostatic Terms?

- Ignore them completely (potentially valid for polyelectrolytes, where electrostatic effects will be expected to dominate in any case)
- Attempt to compute separately using, e.g., scaled-particle theory to estimate cavitation costs, dispersion from models employing atomic or group polarizabilities, others from...?
- Assume proportionality to solvent-accessible surface area of atoms or groups and parameterize microscopic surface tensions (or surface tension functionals)
- Continuum solvation is semiempirical from the outset, so parameterization is no sin (CJC personal opinion...)

First-Solvation-Shell Contributions

Cavitation, Dispersion, Structural rearrangement of solvent



Microscopic Surface Tensions

$$G_{\text{CDS}} = \sum_k A_k \sigma_k$$

Example: The SMx universal solvation models

$$\sigma_i = \sum_j^{\text{descr}} \hat{\sigma}_{Z_i}^{(\xi_j)} \xi_j \quad Z \text{ is atomic number}$$

Some Descriptors ξ :

n is solvent index of refraction

γ is solvent *macroscopic* surface tension

α is solvent hydrogen-bonding acidity (Abraham)

β is solvent hydrogen-bonding basicity (Abraham)

Microscopic Surface Tensions

$$\Delta G_{S,\text{expt}}^{\circ} - \Delta G_{\text{ENP}} = G_{\text{CDS}} = \sum_k A_k \sigma_k$$

Example: The SMx universal solvation models

$$\sigma_i = \sum_j^{\text{descr}} \hat{\sigma}_{Z_i}^{(\xi_j)} \xi_j \quad Z \text{ is atomic number}$$

Some Descriptors ξ :

n is solvent index of refraction

γ is solvent *macroscopic* surface tension

α is solvent hydrogen-bonding acidity (Abraham)

β is solvent hydrogen-bonding basicity (Abraham)

After parameterization (about 72 parameters for 2500 data {H,C,N,O,F,S,P,Cl,Br-compounds} in 91 solvents including water) SM8 has a mean unsigned error of approximately 0.6 kcal mol⁻¹ for neutrals and 3-6 kcal mol⁻¹ for ions, depending on solvent

Examples of Solvent Descriptors

	H ₂ O	C ₆ H ₆	CH ₂ Cl ₂
dielectric constant	78.36	2.27	8.93
Abraham's hydrogen bond acidity	0.82	0.00	0.10
Abraham's hydrogen bond basicity	0.38	0.14	0.05
refractive index	1.33	1.50	1.42
surface tension (cal mol ⁻¹ Å ⁻²)	104.71	40.62	39.15
carbon aromaticity	0.00	1.00	0.00
electronegative halogenicity	0.00	0.00	0.67

SM8 Performance

Mean unsigned errors (kcal/mol) for SM8 and some other popular continuum solvation models

Solute class	Data	SM8	<u>IEF-PCM</u>	C-PCM	PB	All equal
	<i>N</i>		<i>G03/UA0</i>	<i>GAMESS</i>	<i>Jaguar</i>	to mean
aqueous neutrals	274	0.5	4.9	1.6	0.9	2.7
nonaq. neutrals	666	0.6	6.0	2.8	2.3	1.5
aqueous ions	112	3.2	12.4	8.4	4.0	8.6
nonaqueous ions	220	4.9	8.4	8.4	8.1	8.6

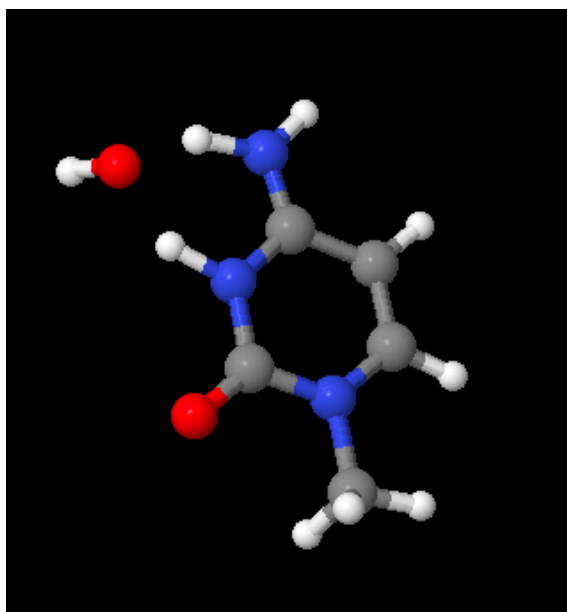
Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. "Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies" *J. Phys. Chem. B* **2009**, *113*, 4538.

Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. "Prediction of SAMPL2 Aqueous Solvation Free Energies and Tautomeric Ratios Using the SM8, SM8AD, and SMD Solvation Models" *J. Comput.-Aid. Mol. Des.* **2010**, *24*, 317.

SMD Example from G09

TS structure for water-assisted tautomerization of 1-methylthymine



Gas phase

SCF Done: E(RM06) = -510.336253045 A.U. after 1 cycles

Dipole moment (Debye): Tot = 5.3667

Aqueous

SCF Done: E(RM06) = -510.358270799 A.U. after 12 cycles

Convg = 0.4387D-08 $-V/T = 2.0099$

SMD-CDS (non-electrostatic) energy (kcal/mol) = 4.81
(included in total energy above)

Dipole moment (Debye): Tot = 7.5777

Free energy of solvation = $-510.358271 - (-510.336253) = -0.022018$ a.u. = -13.8 kcal/mol

Note (i) significant non-electrostatic component and (ii) significant polarization as judged by increased dipole moment.

Introduction of Temperature Dependence into G_{CDS}

SM6
↓
SM6T

$$G_{\text{CDS}} = \sum_k^{\text{atoms}} A_k \left(\sigma_k + \sum_{k'}^{\text{atoms}} \sigma_{kk'}(\mathbf{R}) \right)$$


↓

$$G_{\text{CDS}} = G_{\text{CDS}}(T_0) + \left[\begin{array}{l} \Delta S_S^\circ(T - T_0) + \\ \Delta C_P^\circ[(T - T_0) - T \ln(T/T_0)] \end{array} \right]$$

← contribution at 298 K from SM6

temperature dependence of the non-electrostatic contributions to the free energy of solvation relative to the value at T_0 (298 K)

Introduction of Temperature Dependence into G_{CDS}

SM6

 SM6T

$$G_{\text{CDS}} = \sum_k^{\text{atoms}} A_k \left(\sigma_k + \sum_{k'}^{\text{atoms}} \sigma_{kk'}(\mathbf{R}) \right)$$

$$G_{\text{CDS}} = G_{\text{CDS}}(T_0) + \left(\sum_k^{\text{atoms}} A_k \left(\sigma_k^{\Delta S} + \sum_{k'}^{\text{atoms}} \sigma_{kk'}^{\Delta S}(\mathbf{R}) \right) \right) (T - T_0) + \left(\sum_k^{\text{atoms}} A_k \left(\sigma_k^{\Delta C_P} + \sum_{k'}^{\text{atoms}} \sigma_{kk'}^{\Delta C_P}(\mathbf{R}) \right) \right) [(T - T_0) - T \ln(T/T_0)]$$

entropy-like component

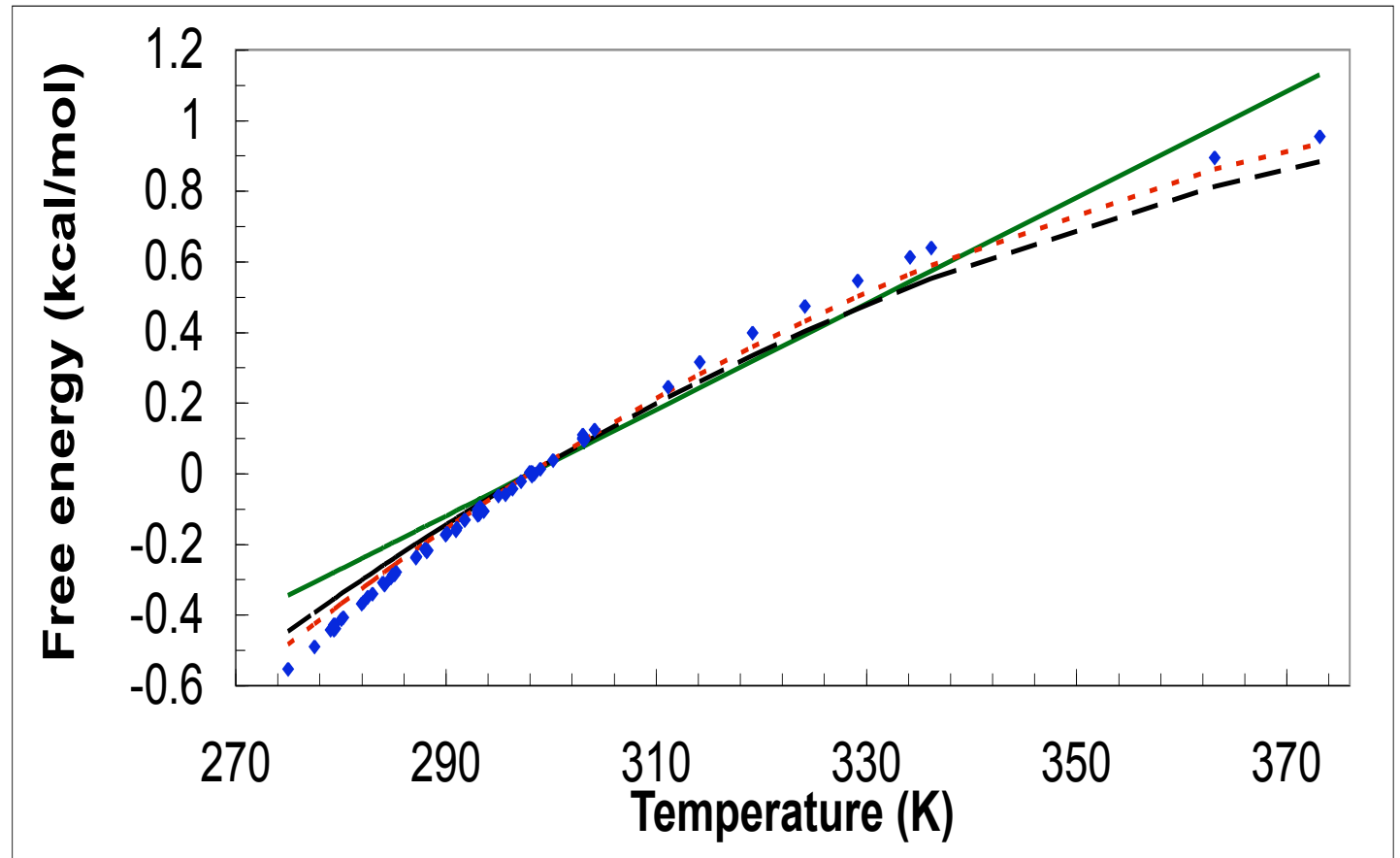
heat capacity-like component

temperature dependence of the non-electrostatic contributions to the free energy of solvation relative to the value at T_0 (298 K)

Quality of Parameterization for Benzene

Variation of the free energy of hydration of benzene relative to 298 K

- Experiment
- Linear
- Unrestricted -
14 parameters
- Final -
7 parameters

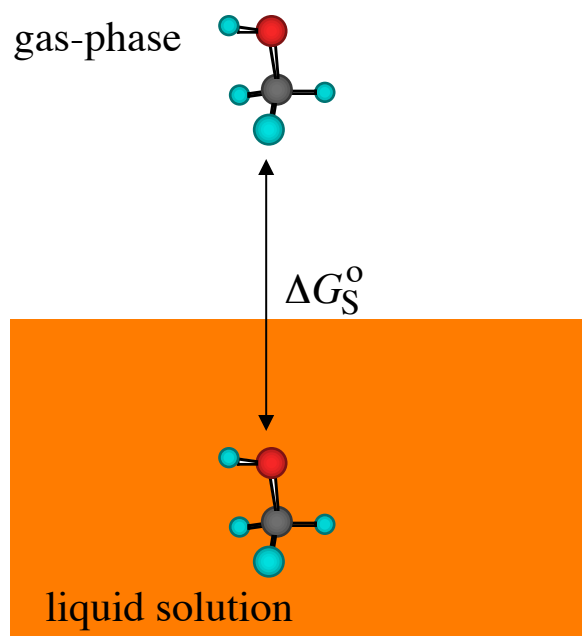


Solvation (Condensed Phase) Models

Implicit Solvent—Phase-phase Partitioning

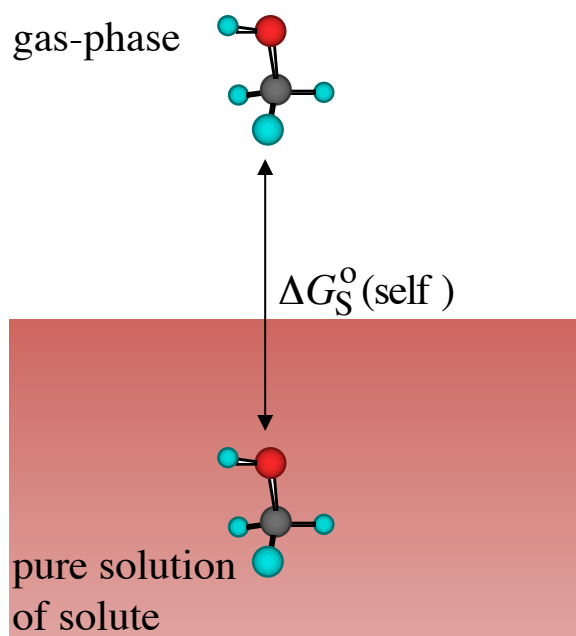
Video VI.vi

What Do We Predict with SMx Solvation Models?



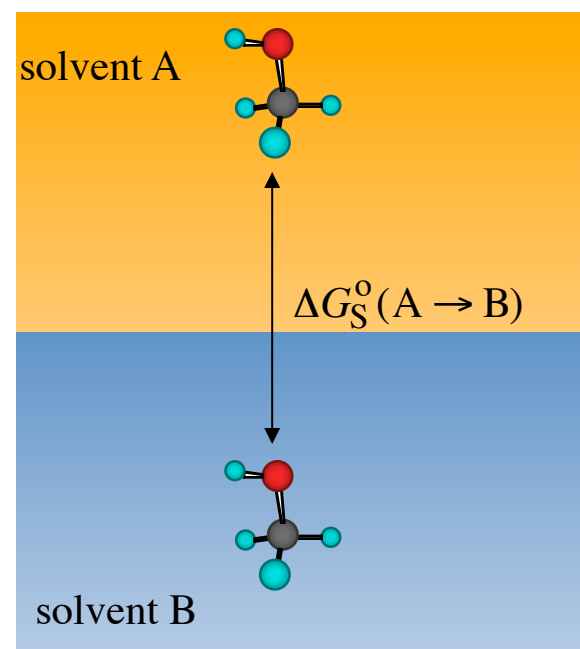
Absolute free energy of solvation

Solvation free energy —
all solvents, no types



Free energy of self-solvation

Vapor pressure



Transfer free energy of solvation

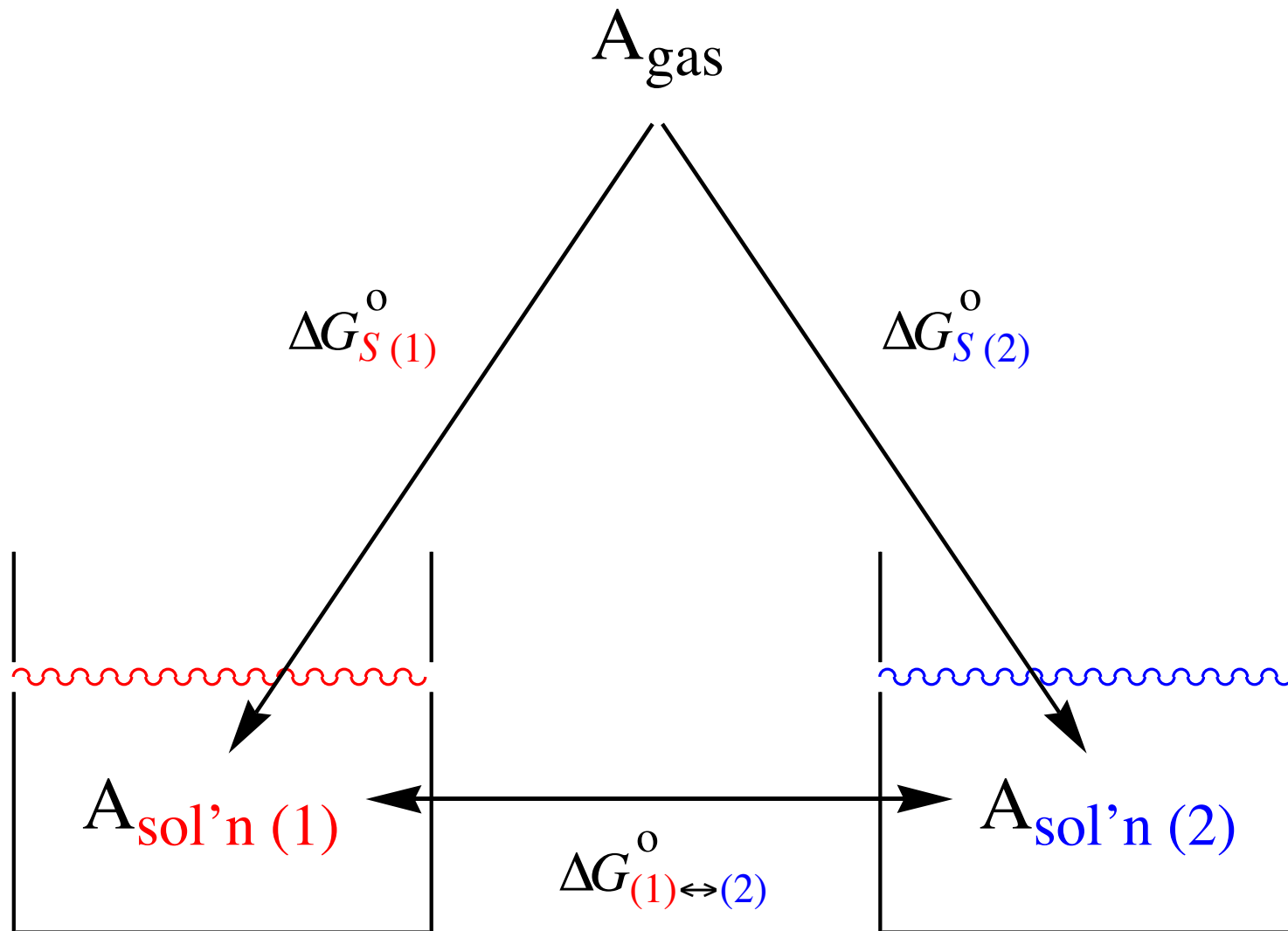
Partition coefficient

By combining these, we also calculate solubility.

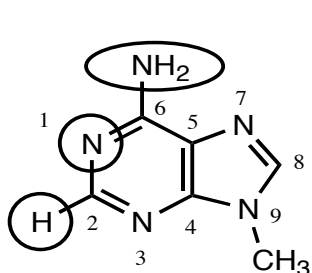
Have also extended to:

- *Interface adsorption*
- *Membrane permeability*

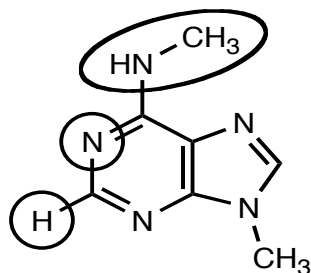
Free Energies of Solvation and Partition Coefficients



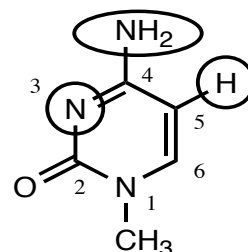
Natural and Unnatural Nucleic Acid Bases



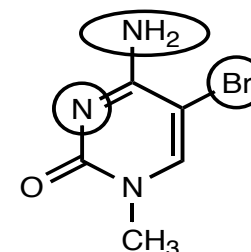
9-methyladenine



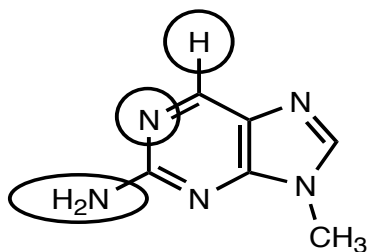
9,*N*⁶-dimethyladenine



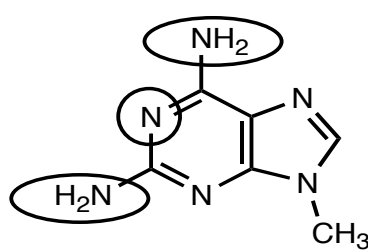
1-methylcytosine



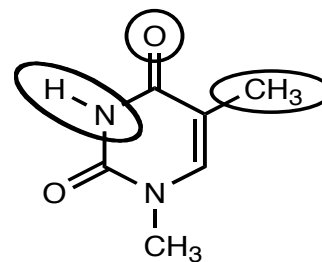
5-bromo-1-methylcytosine



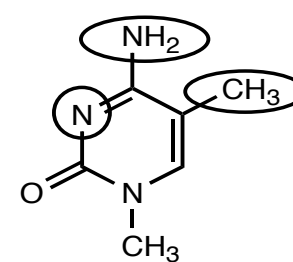
2-amino-9-methylpurine



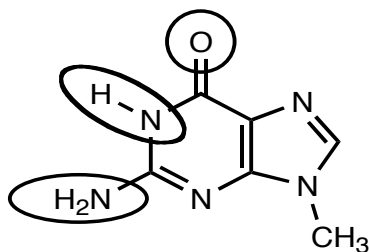
2,6-diamino-9-methylpurine



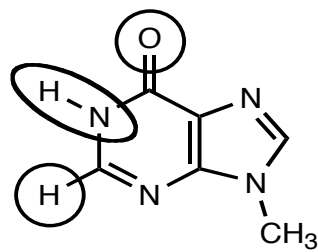
1-methylthymine



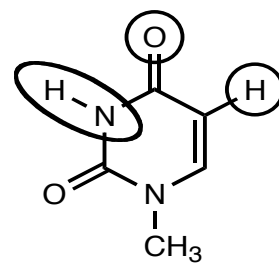
1,5-dimethylcytosine



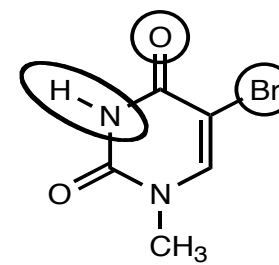
9-methylguanine



9-methylhypoxanthine



1-methyluracil

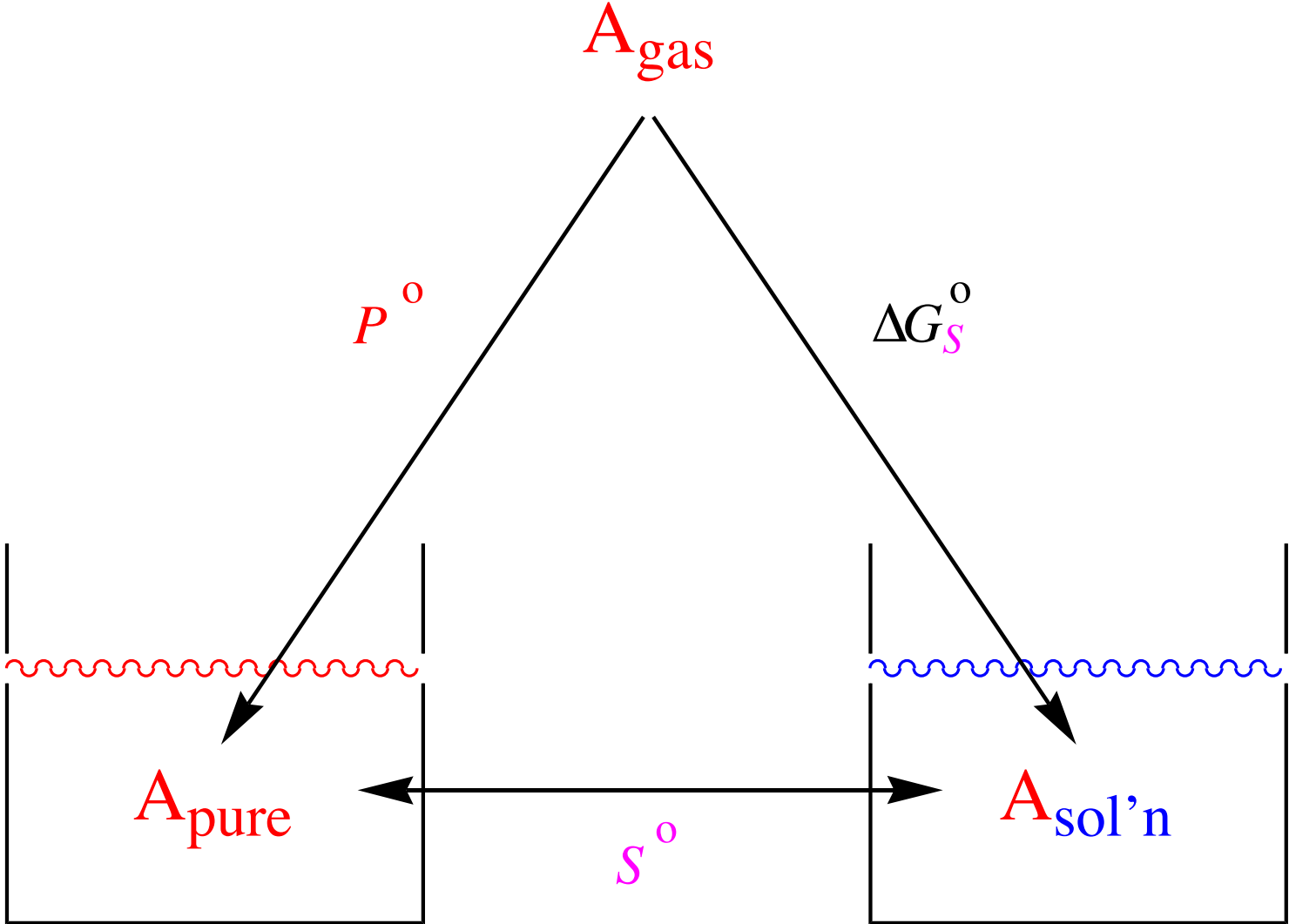


5-bromo-1-methyluracil

Calculated Chloroform/Water Partition Coefficients

Solute	SM5.4/A $\log K_{\text{CHCl}_3/\text{H}_2\text{O}}$	Experiment $\log K_{\text{CHCl}_3/\text{H}_2\text{O}}$	
9-Methyladenine	-1.6	-0.8	<i>* = predicted before measurement</i>
9,N ⁶ -Dimethyladenine	-0.3	0.4*	
2-Amino-9-methylpurine	-1.9	-0.5*	
2,6-Diamino-9-methylpurine	-2.4		
9-Methylguanine	-4.1	-3.5	
9-Methylhypoxanthine	-3.5	-2.5	
1-Methylcytosine	-4.3	-3.0	
5-Bromo-1-methylcytosine	-2.4		
1-Methylthymine	-0.3	-0.5	
1,5-Dimethylcytosine	-3.1		
1-Methyluracil	-1.2	-1.2	
5-Bromo-1-methyluracil	-0.3	-0.7*	
<i>Mean unsigned error:</i>	0.7		

Solubility from Solvation Free Energy and Vapor Pressure



Mean-unsigned error (MUE) in predicted $\log P$

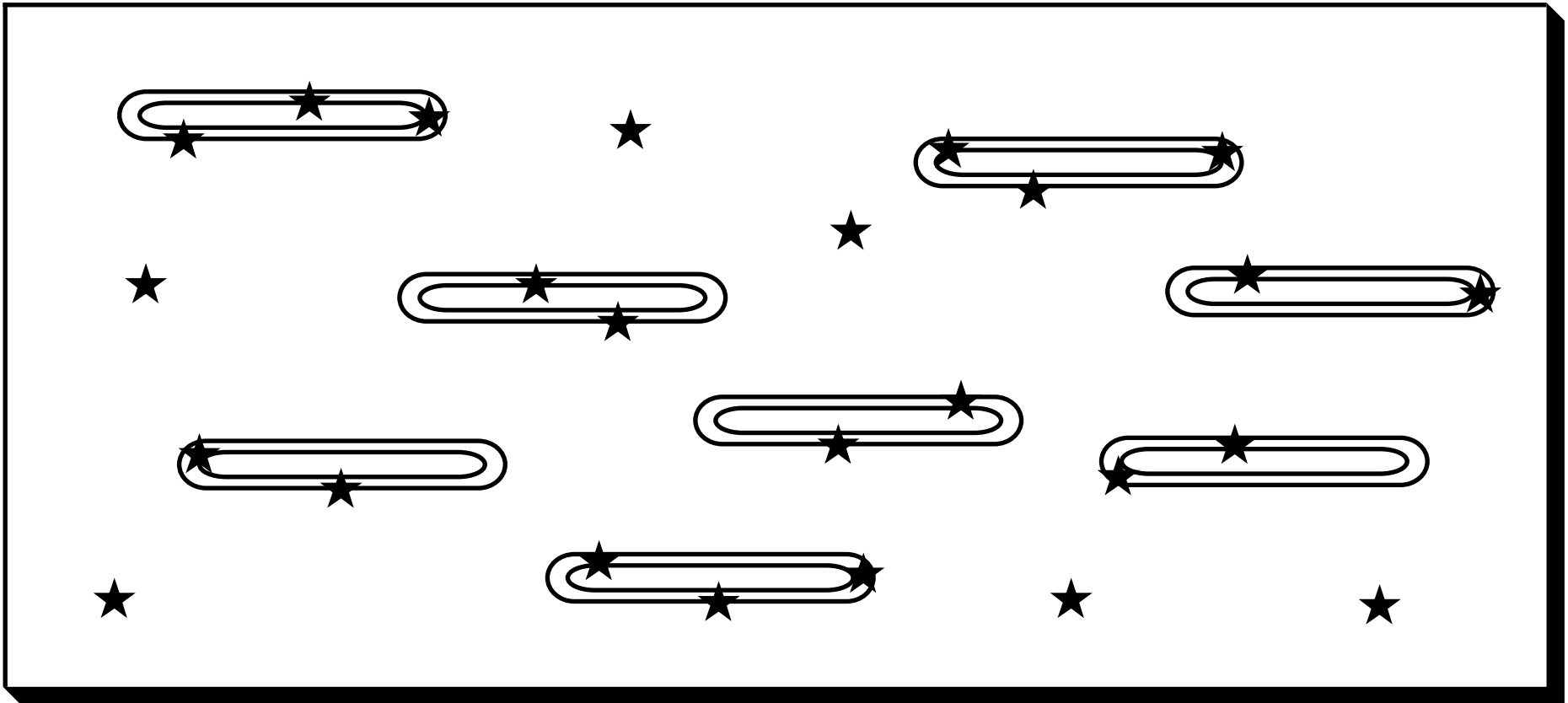
Solute class	No. data	SM5.42R		
		HF	B3LYP	AM1
hydrocarbons	11	0.3	0.3	0.2
aromatics	6	0.2	0.3	0.2
alcohols/phenols	9	0.3	0.2	0.3
ethers	4	0.2	0.2	0.2
carbonyls	11	0.7	0.5	0.8
esters	7	0.3	0.2	0.6
CHN compounds	7	0.2	0.2	0.4
nitro compounds	5	0.1	0.2	0.3
HCNO compounds	60	0.3	0.3	0.4
halocarbons	15	0.4	0.4	0.5
all liquid solutes	75	0.3	0.3	0.4

Winget et al. *J. Phys. Chem. B* **2000**, *104*, 4726

Mean-unsigned error (MUE) in predicted log *S*

Solute class	No. data	SM5.42R			
		HF	B3LYP	AM1	UNIFAC
hydrocarbons	11	0.5	0.4	0.4	1.4
aromatics	6	0.1	0.0	0.1	0.2
alcohols/phenols	9	0.3	0.2	0.3	0.6
ethers	4	0.5	0.5	0.5	0.5
carbonyls	11	0.4	0.4	0.5	0.3
esters	7	0.3	0.3	0.2	0.1
CHN compounds	7	0.7	0.5	0.6	0.5
nitro compounds	5	0.4	0.4	0.2	0.2
HCNO compounds	60	0.4	0.4	0.4	0.6
halocarbons	10	0.2	0.3	0.2	0.4
all liquid solutes	70	0.4	0.3	0.3	0.5
solid solutes	13	0.3	0.4	0.5	0.8

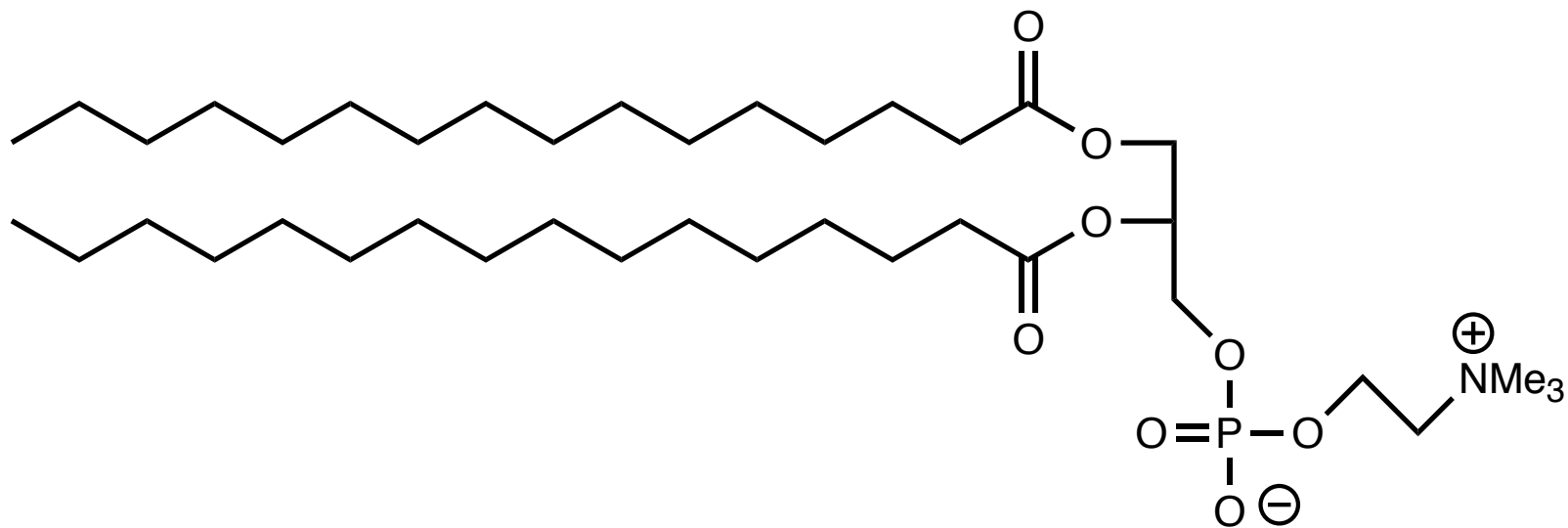
Analysis of Membrane/Water Partitioning



★ = Solute

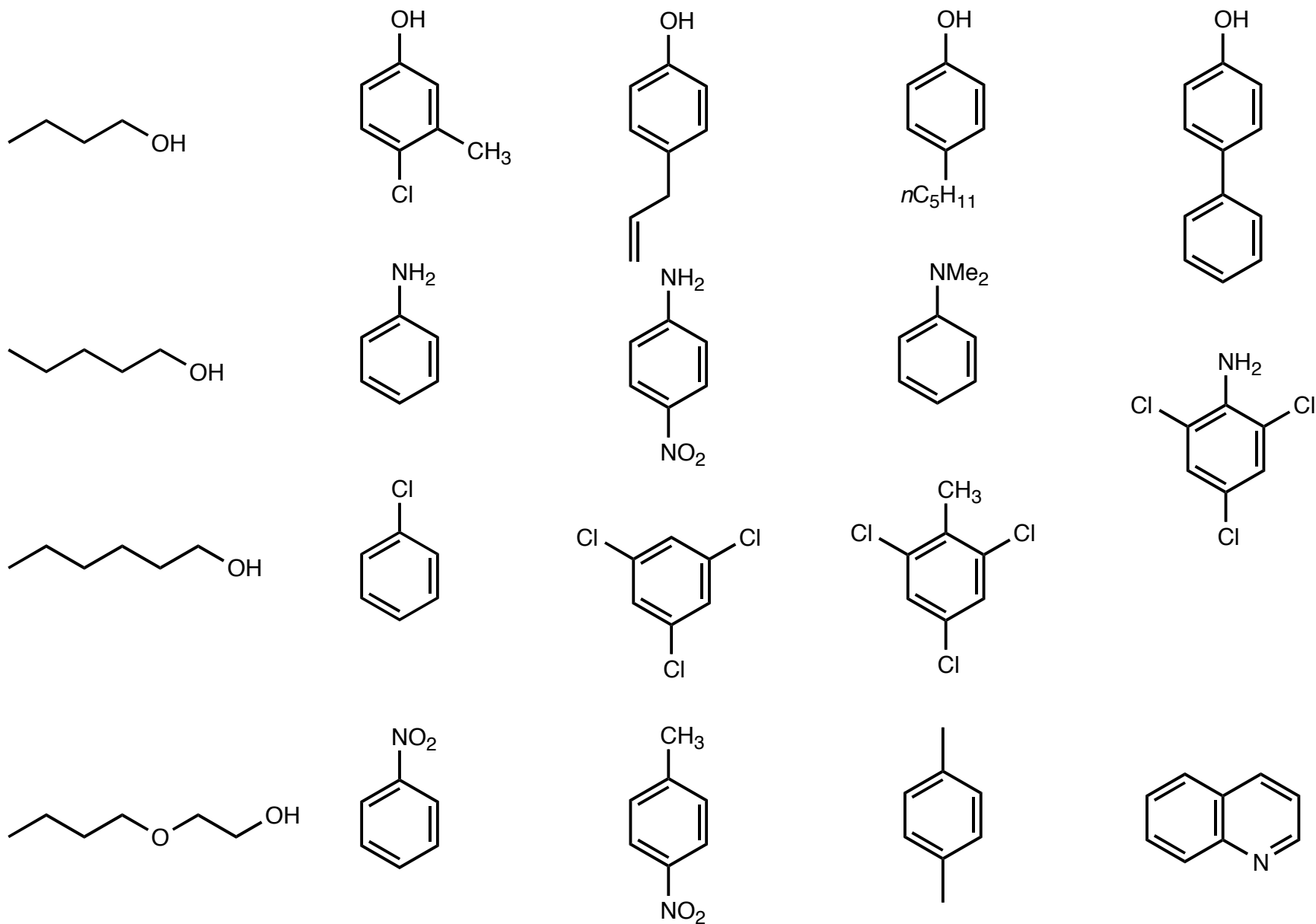
*Known starting concentrations
Microsyringe to sample aqueous phase
Interior vesicle volume negligible*

“Solvent” Model for Phosphatidyl Choline

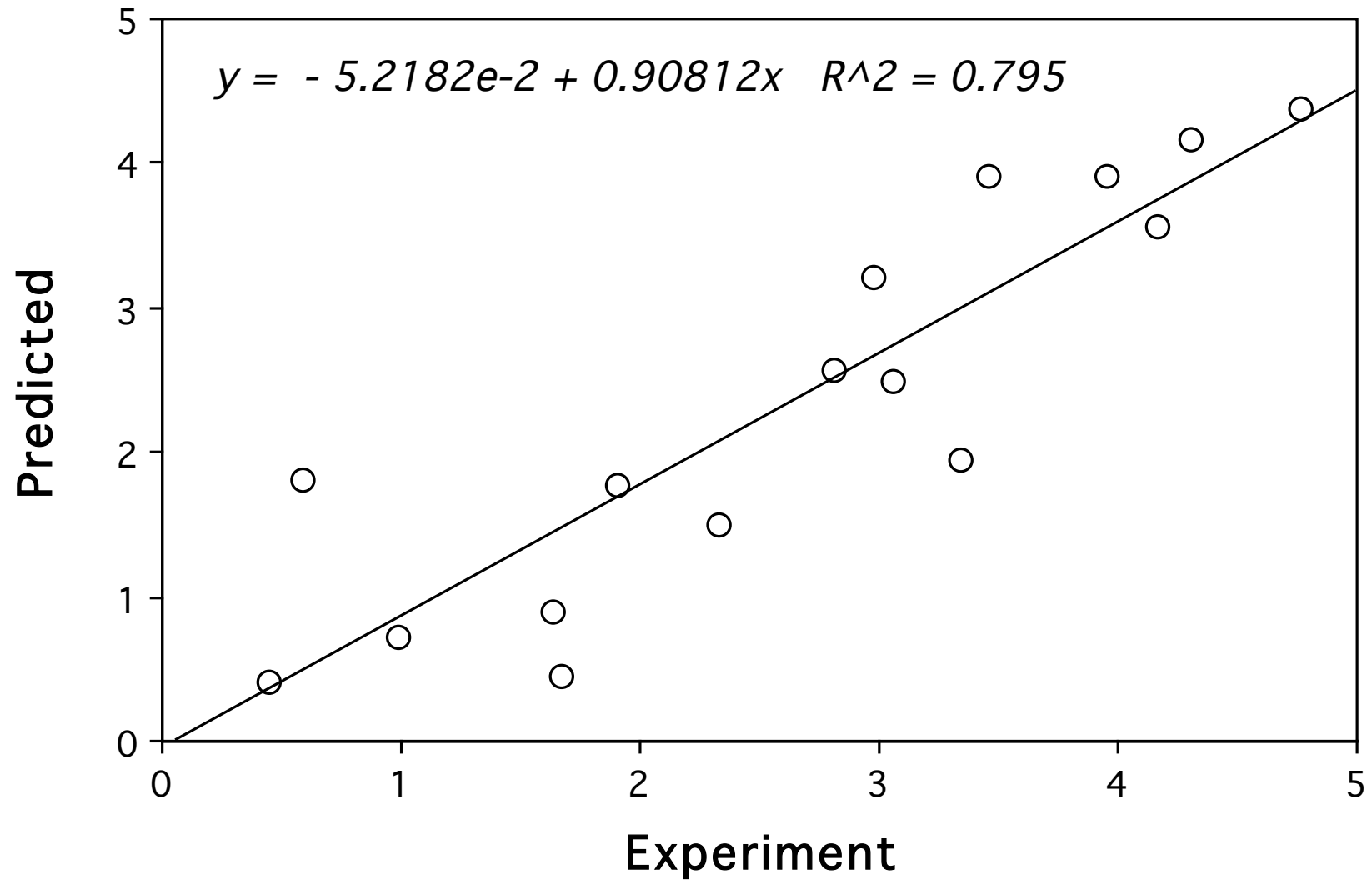


<i>SM5.4 requires:</i>	$\underline{\epsilon}$	\underline{n}	$\underline{\alpha}$	$\underline{\beta}$	$\underline{\gamma}$	
<i>Best guess:</i>	5.0	1.37	0.0	0.9	27	(Const.)
<i>Regression fit:</i>	(5.0)	1.40	(0.0)	1.15	25	0.59

Water/Phosphatidyl Choline Partitioning



Predicted vs Experimental log K (m/w)

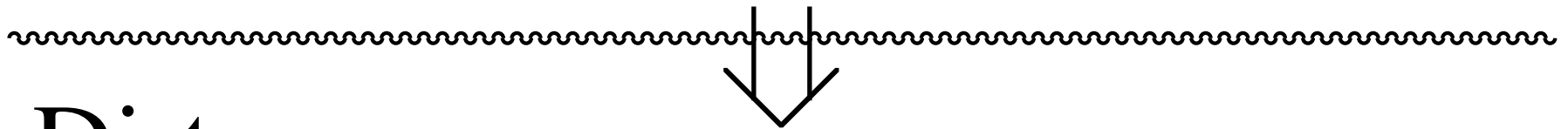


Chambers et al. in *Rational Drug Design*, Truhlar, et al. Eds.; Springer: New York, 1999, p. 51.

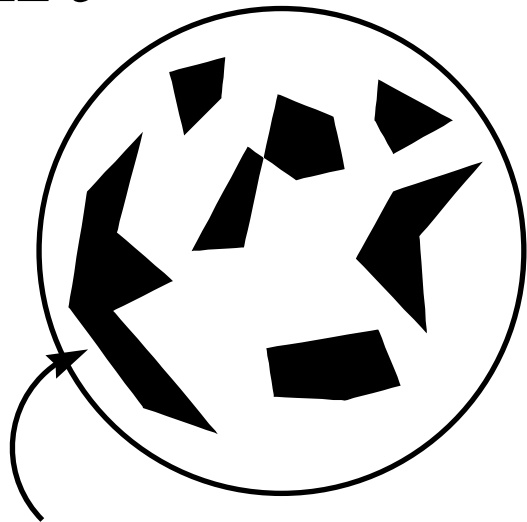
Soil/Water Partitioning

Important factor controlling the persistence of environmental contaminants

carbamates, phosphonothioates, polyhalogenated aromatics, ureas, horrible molecules



Dirt



*organic
carbon*

$$K_{OC} = \frac{[X]_{\text{soil}} / \%OC}{[X]_{\text{aqueous}}}$$

remarkably constant from clay to loam to peat

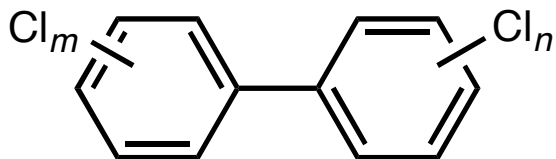
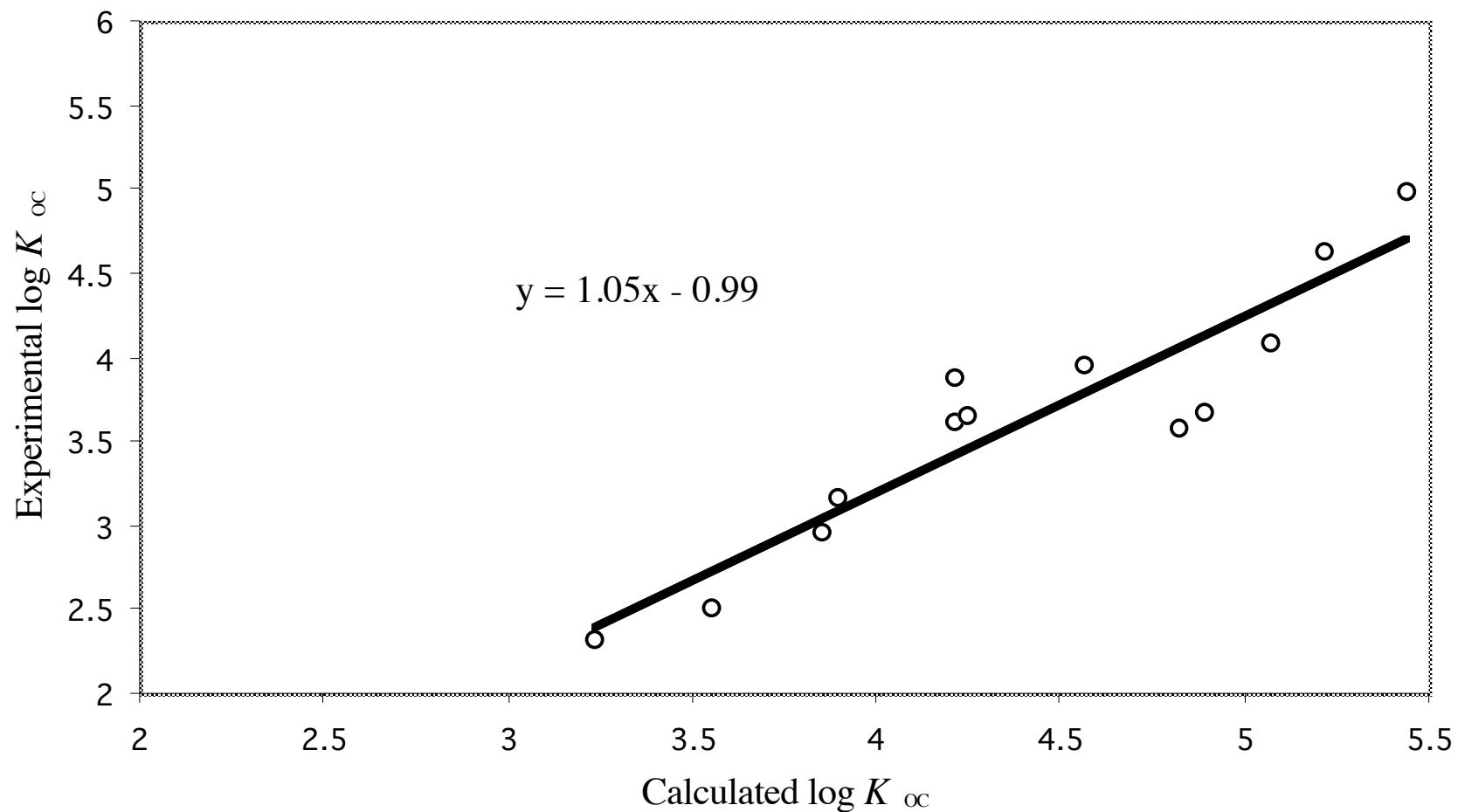
“Solvent” Model for Soil

"Dirt/MIDI!"

<i>SM5.42R requires:</i>	$\underline{\varepsilon}$	\underline{n}	$\underline{\alpha}$	$\underline{\beta}$	$\underline{\gamma}$
<i>Best guess:</i>	?	?	?	?	?
<i>Regression fit:</i>	(15.0)	1.379	0.61	0.60	46.0

Mean unsigned error over 387 compounds = 1 log unit

Correlation of a Subset of Chlorinated Biphenyls



Theory allows for an atom-by-atom decomposition of the partitioning energies to better understand factors affecting them

Solvation (Condensed Phase) Models

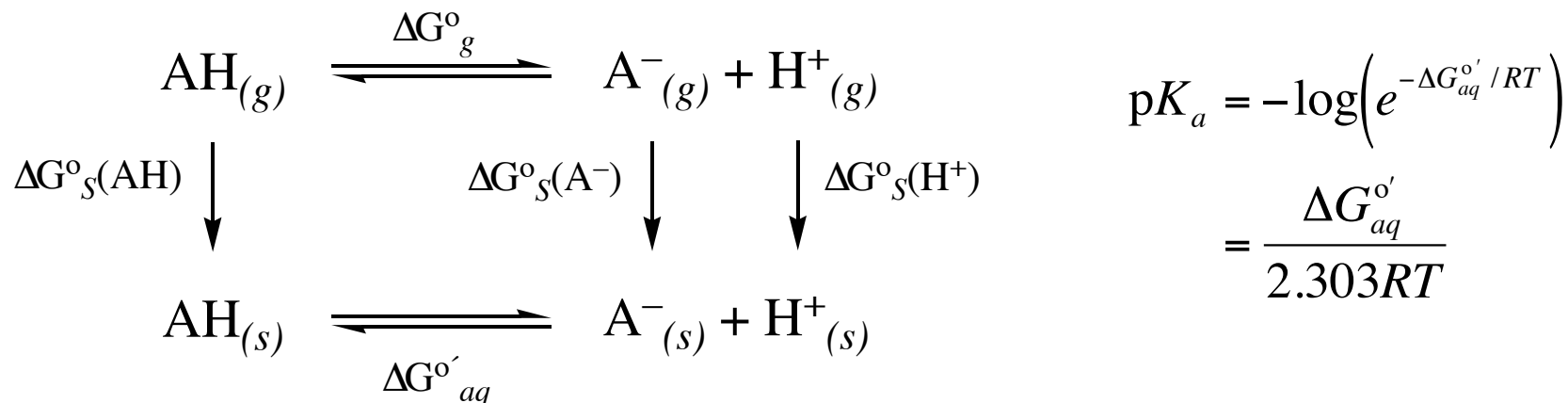
Hybrid Explicit/Implicit Solvent

“it was the best of models, it was the worst of models”

Video VI.vii

Equilibria: Thema mit Veränderungen

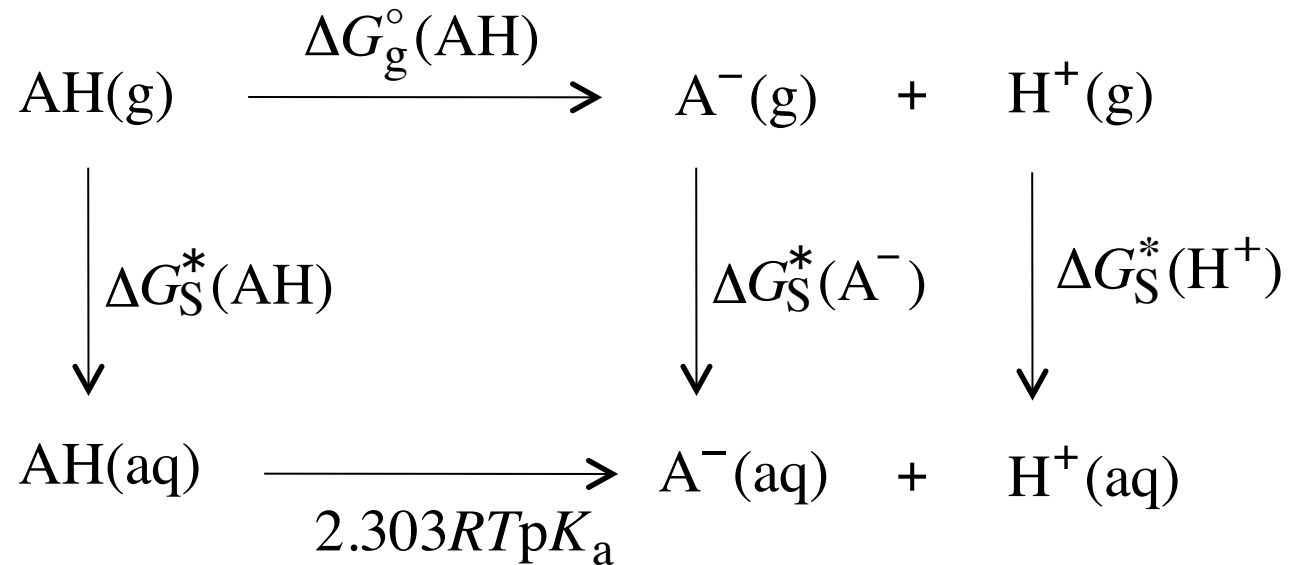
pK_a (Born-Haber cycle)



- 1) Need diffuse functions in basis set and good theory to get accurate gas-phase deprotonation free energy
- 2) Can't compute E for H^+ (no electrons!) so electronic structure programs are reluctant to compute thermal contributions to G (but a good spreadsheet will)
- 3) ΔG°_S of proton is an experimental quantity ($-264.0 \text{ kcal mol}^{-1}$)
- 4) Standard-state concentration-change free energy must be included
- 5) Each non-cancelled error of $1.4 \text{ kcal mol}^{-1}$ in any step will lead to an error in pK_a of 1 pK unit — errors in ionic solvation free energies are potentially *much* larger than this...
- 6) Can correct for functional-group systematic errors, see: Klicic et al. *J. Phys. Chem. A* **2002**, *106*, 1327.

Free Energy Cycles and Ions

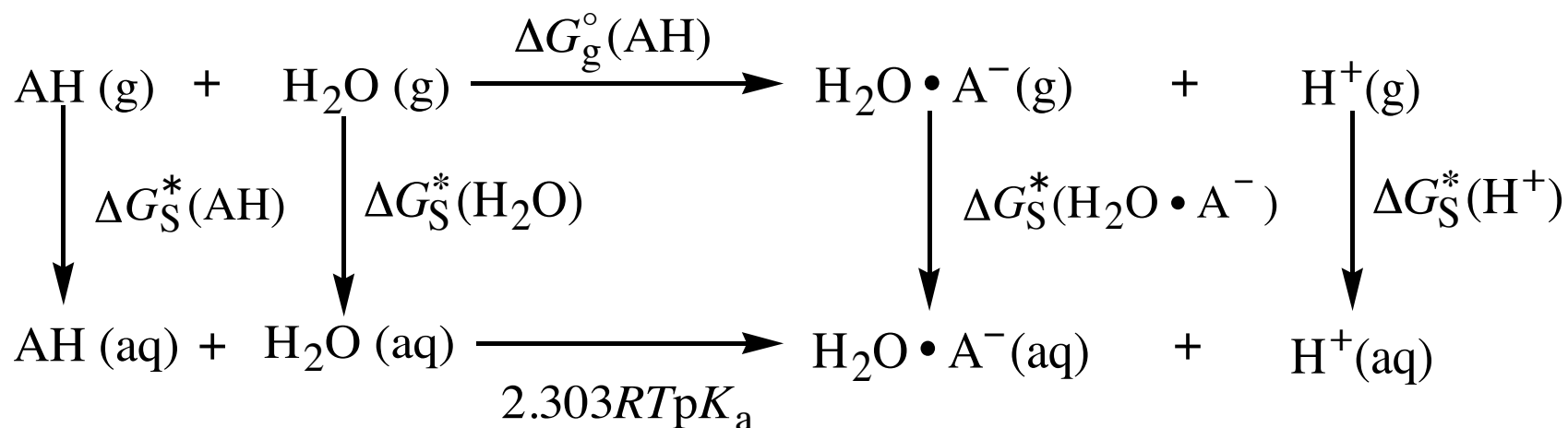
No explicit water molecules



$$2.303RT\text{p}K_a = \Delta G_g^\circ(\text{AH}) - \Delta G_S^*(\text{AH}) + \Delta G_S^*(\text{A}^-) + \Delta G_S^*(\text{H}^+)$$

Free Energy Cycles and Ionic Clusters

Treat the ion as a cluster



$$2.303RT\text{p}K_a = \Delta G_g^\circ(\text{AH}) - \Delta G_S^*(\text{AH}) - \Delta G_S^*(\text{H}_2\text{O}) + \Delta G_S^*(\text{H}_2\text{O} \cdot \text{A}^-) + \Delta G_S^*(\text{H}^+)$$

*For additional details on cluster solvation free energies and their use to set the absolute solvation free energy of the proton and the absolute potential of the normal hydrogen electrode (NHE), see Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. J. Phys. Chem. B **2006**, 110, 16066.*

Example: pK_a of Methanol (Expt. 15.5)

MeOH / MeO⁻

(Cycle 1)

Experimental data:

$$\Delta G_g^\circ = +375.0 \text{ kcal/mol}$$

$$\Delta G_S^*(H^+) = -265.9 \text{ kcal/mol}$$

$$\Delta G_S^*(\text{MeOH}) = -5.11 \text{ kcal/mol}$$

Calculated (SM6) data:

$$\Delta G_S^*(\text{MeO}^-) = -88.3 \text{ kcal/mol}$$

$$pK_a = 20.4$$

MeOH / H₂O • MeO⁻

(Cycle 2)

Experimental data:

$$\Delta G_g^\circ = +358.0 \text{ kcal/mol}$$

$$\Delta G_S^*(H^+) = -265.9 \text{ kcal/mol}$$

$$\Delta G_S^*(\text{MeOH}) = -5.11 \text{ kcal/mol}$$

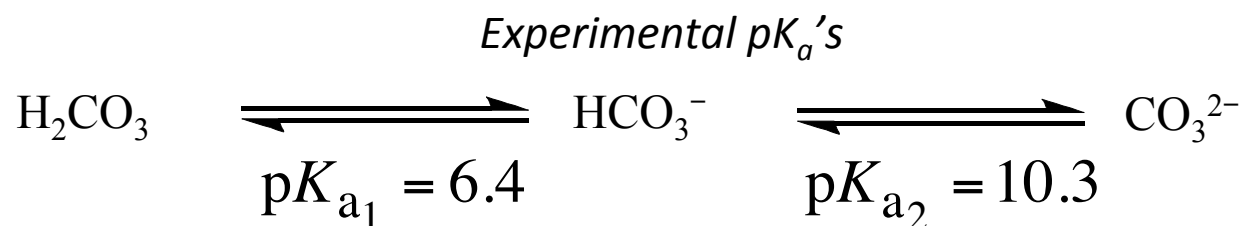
$$\Delta G_S^*(\text{H}_2\text{O}) = -6.32 \text{ kcal/mol}$$

Calculated (SM6) data:

$$\Delta G_S^*(\text{H}_2\text{O} \cdot \text{MeO}^-) = -81.8 \text{ kcal/mol}$$

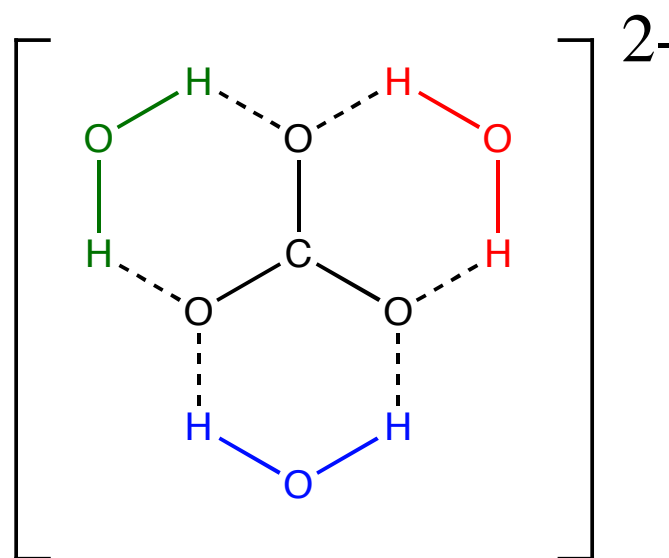
$$pK_a = 16.0$$

Adding More Waters



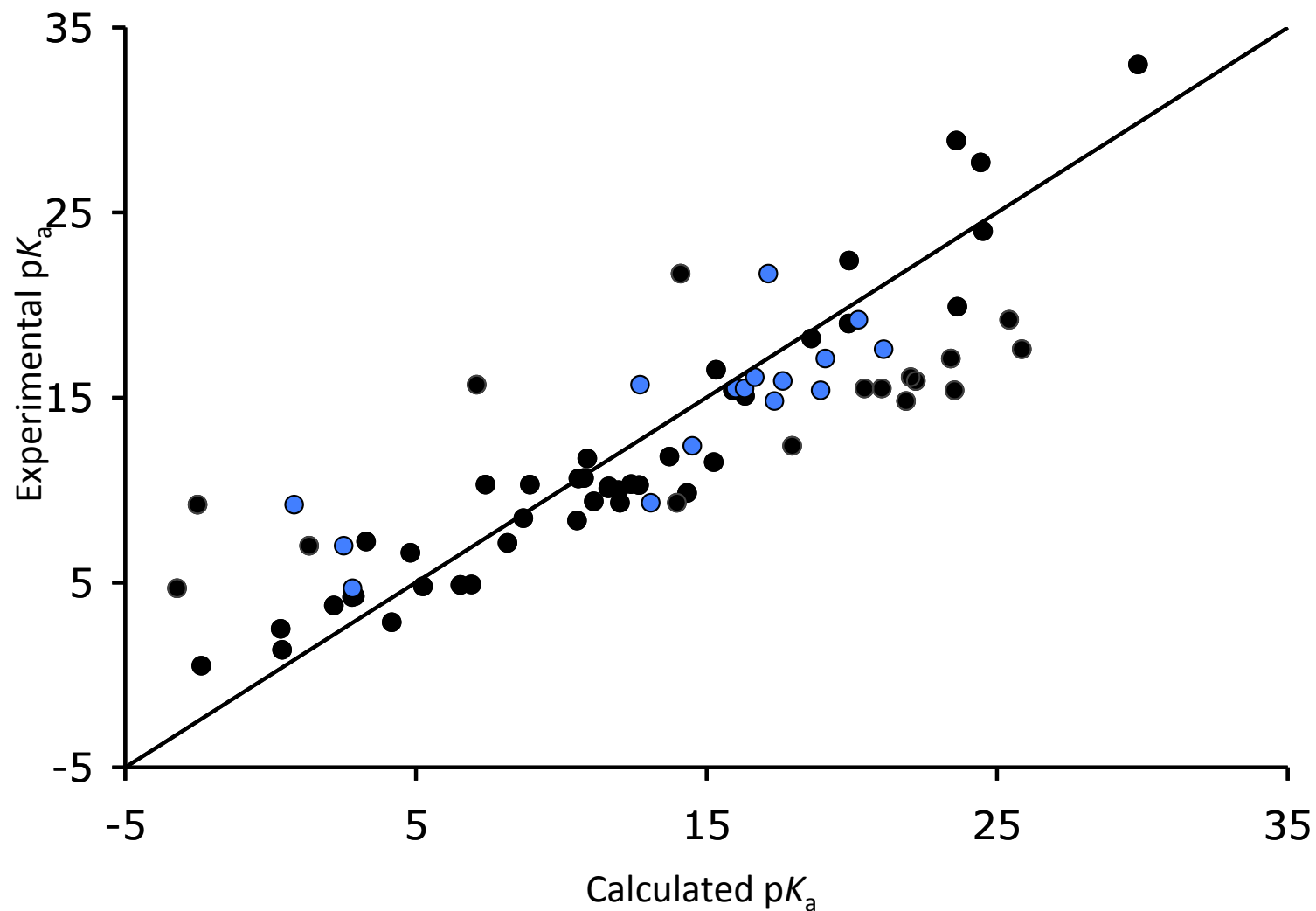
Calculated pK_a 's

No. H ₂ O	pK_{a1}	pK_{a2}
0	-0.6	1.6
1	1.3	5.0
2	2.3	7.8
3	4.2	9.0



Adding explicit water molecules improves the accuracy of the calculation

Clustering Other Ions



● , ○ : bare ion used

● : clustered ion used

Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 2493

Ru(bpy)(damp) Catalyst

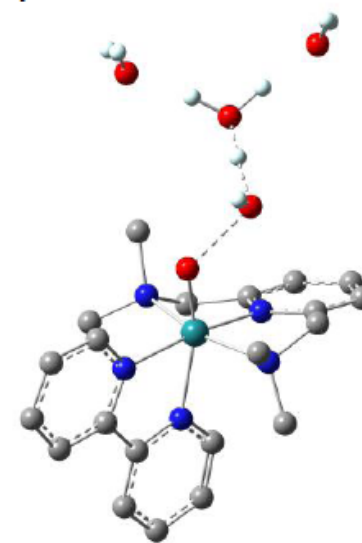
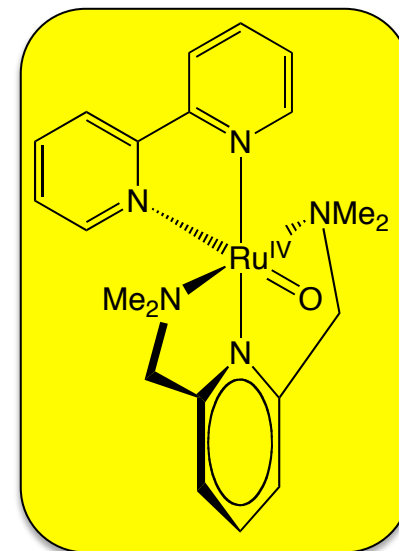
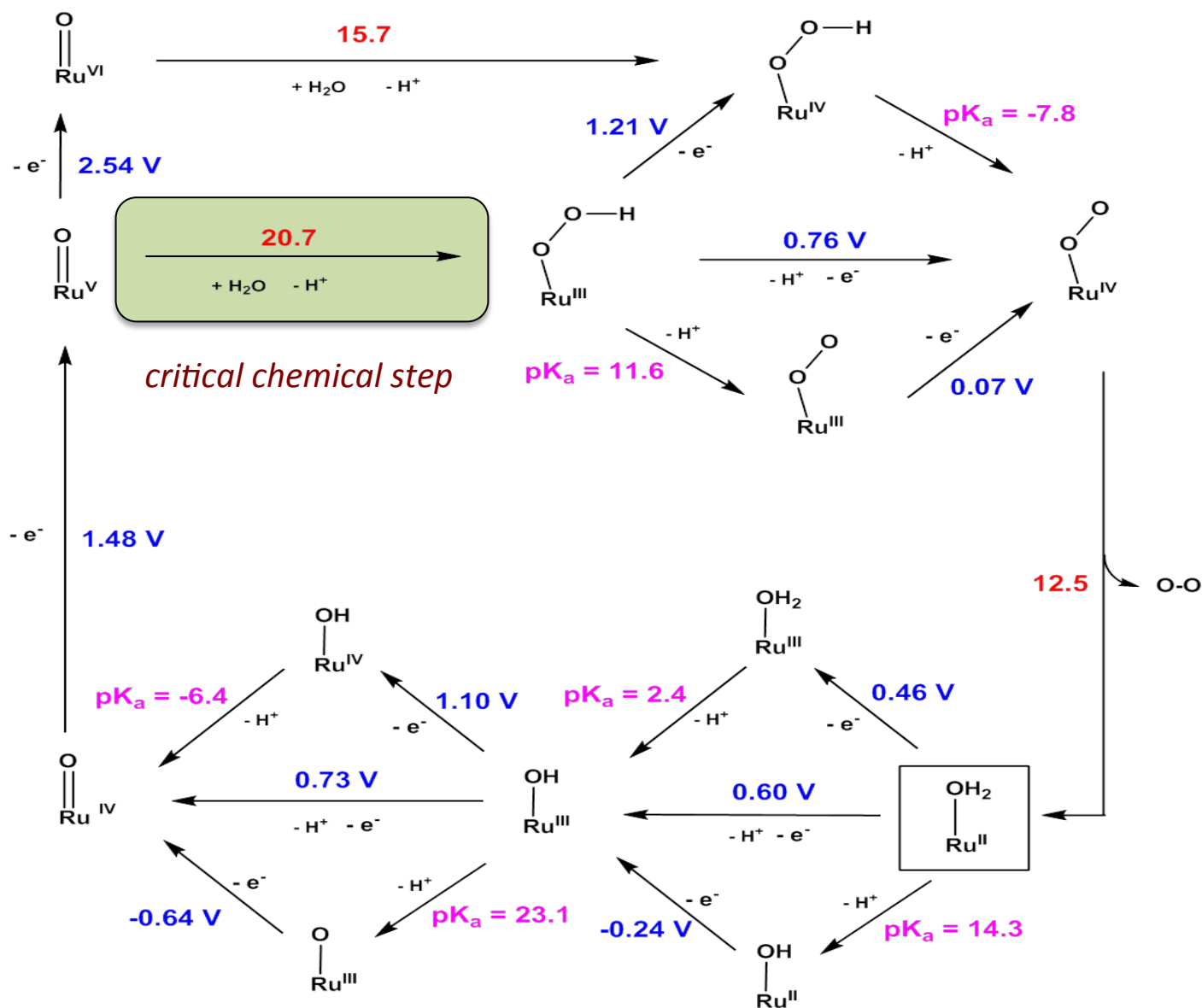




Photo courtesy Natalie Lucier of Flickr under Creative Commons license