# SAMPL4 blind challenge
# Hydration free energy predictions

## Lars Sandberg            20/09/2013

Biological Chemistry and Drug Discovery

College of Life Sciences

University of Dundee

United Kingdom

Drug
Discovery
Unit

Neglected tropical diseases
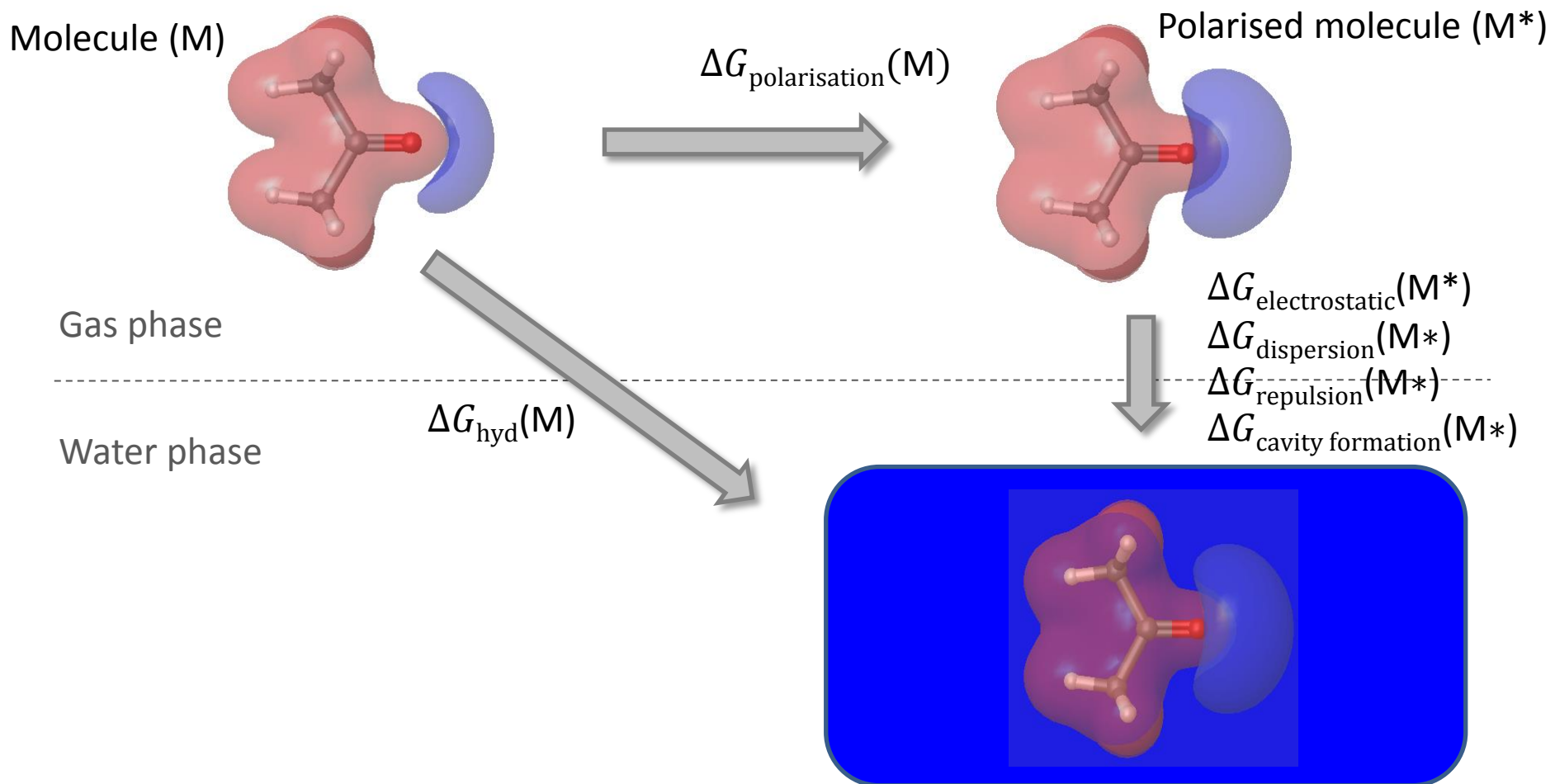Innovative targets and pathways

# Implicit solvent model - objectives

1. Include all physical effects  liable to affect molecular solvation

2. As few empirical parameters and additive corrections, derived from experimental hydration data, as possible

3. Applicable to molecules found in biological and organic chemistry

4. Transferable parameters with, and with comparable accuracy to, the golden standard of explicit solvent free energy calculations

5. Chemical accuracy – solvation free energy errors in the order of 1 kcal/mol or less

# Molecular preparation

- Preparation using the Schrödinger Suite
  - 2D to 3D conversion, ionization state, chirality, and tautomer form (Ligprep)
  - Energy minimization (Macromodel/MMFF94s)
  - 5 lowest energy conformers (Macromodel/MMFF94s)
  - Gas phase geometry optimisation (Jaguar/PBE1PBE/6-311++G(d,p))
  - Re-rank and select the lowest energy conformer out of the 5

NB: The gas phase geometry is used since the primary component to the solute-solvent electrostatic energy comes from the gas phase, and not from the water phase (cf. the Born model). For most small molecules the difference in conformation is negligible. Hence the computational effort is minimised when only the gas phase geometry is used.

# Hydration free energy thermocycle

Molecule (M)

$\Delta G_{\text{polarisation}}(\text{M})$

Polarised molecule (M*)

Gas phase

$\Delta G_{\text{hyd}}(\text{M})$

Water phase

$\Delta G_{\text{electrostatic}}(\text{M*})$
$\Delta G_{\text{dispersion}}(\text{M}*)$
$\Delta G_{\text{repulsion}}(\text{M}*)$
$\Delta G_{\text{cavity formation}}(\text{M}*)$

$$\Delta G_{\text{hyd}}(\text{M}) = \Delta G_{\text{pol}}(\text{M}) + \Delta G_{\text{elec}}(\text{M}*) + \Delta G_{\text{disp}}(\text{M}*) + \Delta G_{\text{rep}}(\text{M}*) + \Delta G_{\text{cavity}}(\text{M}*)$$

# Molecular polarisation $\Delta G_{polarisation}$

- Reaction field single point calculation
  - Jaguar PBE1PBE/Aug-cc-pVTZ using solvation model PBF/water

  - Electron reorganisation:

$$\Delta G_{polarization}(M) = \langle \psi(F)|\mathcal{H}|\psi(F)\rangle - \langle \psi(0)|\mathcal{H}|\psi(0)\rangle$$

  <span style="color:blue">Total solute energy</span>       <span style="color:blue">Gas phase energy</span>

  where $\mathcal{H}$ is the electronic Hamiltonian and $F$ is the local reaction-field.

  - Molecular geometry, ESP atomic partial charges and $\Delta G_{polarisation}$ are passed on to the implicit solvent model to calculate the hydration free energy.

# The implicit solvent model

- $\Delta G_{elec}$: Maxwell's equations
  - Advanced dielectric continuum model including nonlinear response effects [1]
  - Finite difference method, lattice spacing 0.2 Å [2]

- $\Delta G_{disp}+\Delta G_{rep}$: Lennard-Jones 12-6 (LJ) potential using the Liquid Simulation Force-Field (LSFF) [3]
  - One pair of transferable LJ parameters per chemical element
    - Solute parameters from molecular models simulating vapour-liquid equilibria (VLE) of pure substances
    - Water solvent LJ parameters from the TIP5P-E water model.
    - Lorentz-Berthelot mixing rules

- $\Delta G_{cav}$: The solvent accessible surface area times the surface tension of water [3]
    - For smaller cavities higher order curvature terms are needed. These are the only implicit solvent model parameters fitted to experimental hydration free energies.
    - Curvature terms are fitted to noble gases, small molecules, and alkanes (negligible electrostatic contribution)

[1] L. Sandberg, and O. Edholm, Nonlinear response effects in continuum models of the hydrations of ions, *J.Chem.Phys.* (2002) **116**: 2936-2944
[2] L. Sandberg, R. Casemyr, and O. Edholm, Calculated Hydration Free Energies of Small Organic Molecules Using a Nonlinear Dielectric Continuum Model, *J.Phys.Chem. B* (2002) 106: 7889-7897
[3] L. Sandberg unpublished data, to be published

# What's different with an advanced dielectric continuum model?

- The simple medium approximation of water: relative permittivity $\varepsilon_r$=78.5 (the dielectric constant)

- However, the thermally fluctuating water dipoles align in the electric field $E$ of the solute. In a strong field the dipole orientational polarisation approaches total alignment, the dielectric medium saturates, and the relative permittivity $\varepsilon_r=\varepsilon(E)$:
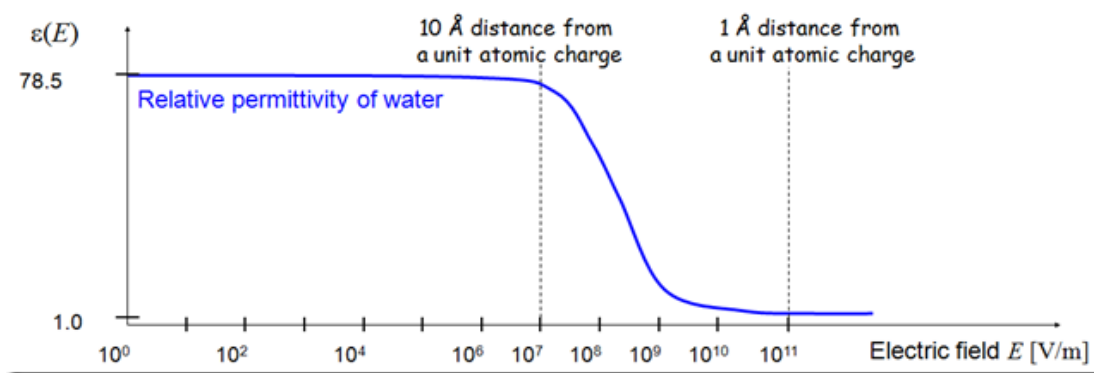


- An advanced dielectric continuum model includes nonlinear dielectric response effects: relative permittivity $\varepsilon_r=\varepsilon(E)$
  - Normal dielectric saturation
  - Electrostriction (compression of the dielectric medium due to the applied electric field)

# The implicit solvent model – validation

- What is the interaction energy between a TIP5P-E water model molecule and the implicit solvent model? Compare with published liquid simulation data.

- The average potential energy of the TIP5P-E model at 25$^o$ [4]

$$U(\text{liquid}) = -9.78 \text{ kcal/mol}$$

- Advanced dielectric continuum theory at 25$^o$ (lattice spacing 0.2 Å)

$$U(\text{liquid}) = \underbrace{\Delta G_{pol}}_{= 0 \text{ for TIP5P-E}} + \Delta G_{elec} + \Delta G_{LJ} = -9.74 \text{ kcal/mol}$$

[4] S.W. Rick, A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums, *J.Chem.Phys.* (2004) **120**: 6085-6093

# Post-processing - corrections

- ## Why are corrections needed?
  - Polarisation stress is relieved through charge transfer between the solute and the 1$^{st}$ solvation shell solvent molecules. This physical effect is not a part of a classical implicit solvent model (nor any classical explicit solvent models), and that is why corrections are needed.

- ## Ways to address it
  - One solution would be to solvate the solute-water complex and include the polarisation stress relief from first principle. No corrections are needed.
  - A simpler way is to introduce empirical functional group (FG) corrections, determined from the difference between known experimental and calculated hydration free energy data for each FG. This approach was used in the SAMPL4 blind challenge.

# Empirical functional group corrections

- Empirical correction factors for FG with a correction > 0.5 kcal/mol are applied (true for 7 FG in SAMPL4, see below)

  - More important (> 1.0 kcal/mol)

    | | | |
    |---|---|---|
    | Alcohol | correction = -1.8 kcal/mol | used in SAMPL4 |
    | Ether | correction = -2.0 kcal/mol | used in SAMPL4 |
    | Primary amine | correction = -1.9 kcal/mol | used in SAMPL4 |
    | Secondary amine | correction = -3.7 kcal/mol | used in SAMPL4 |
    | Tertiary amine | correction = -3.9 kcal/mol | used in SAMPL4 |
    | Aromatic sp2 amine | correction = -2.4 kcal/mol | used in SAMPL4 |

  - Less pronounced (< 1.0 kcal/mol)

    | | | |
    |---|---|---|
    | Ketone | correction = -0.4 kcal/mol | Not applied |
    | Carboxylic acid | correction = -0.2 kcal/mol | Not applied |
    | Ester | correction = +0.7 kcal/mol | used in SAMPL4 |

  - Groups connected to an aryl have a lesser need for FG correction by half per aryl due to a weaker hydrogen bond acceptor functionality, e.g.

    - 1,2-Dimethoxybenzene: only half the ether correction per methoxy group is needed
    - Dibenzo-p-dioxine: no ether correction needed
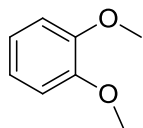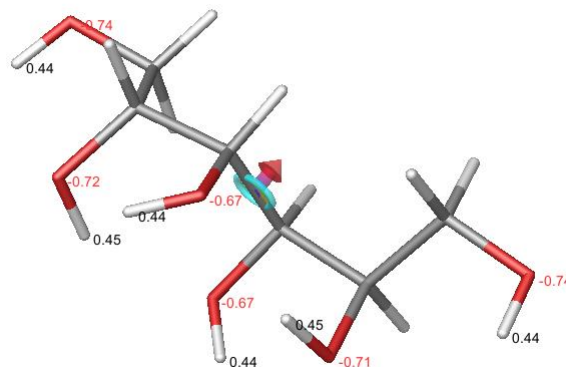


SAMPL4_005: Exp  -5.33 ± 0.10, Calc -4.34 ± 0.78
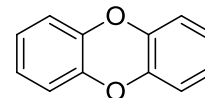


SAMPL4_049: Exp  -3.16 ± 0.10 , Calc -3.09 ± 0.54

# An example – mannitol (SAMPL4_001)

- New lowest energy conformer (kindly provided by Prof. Dr. Andreas Klamt)
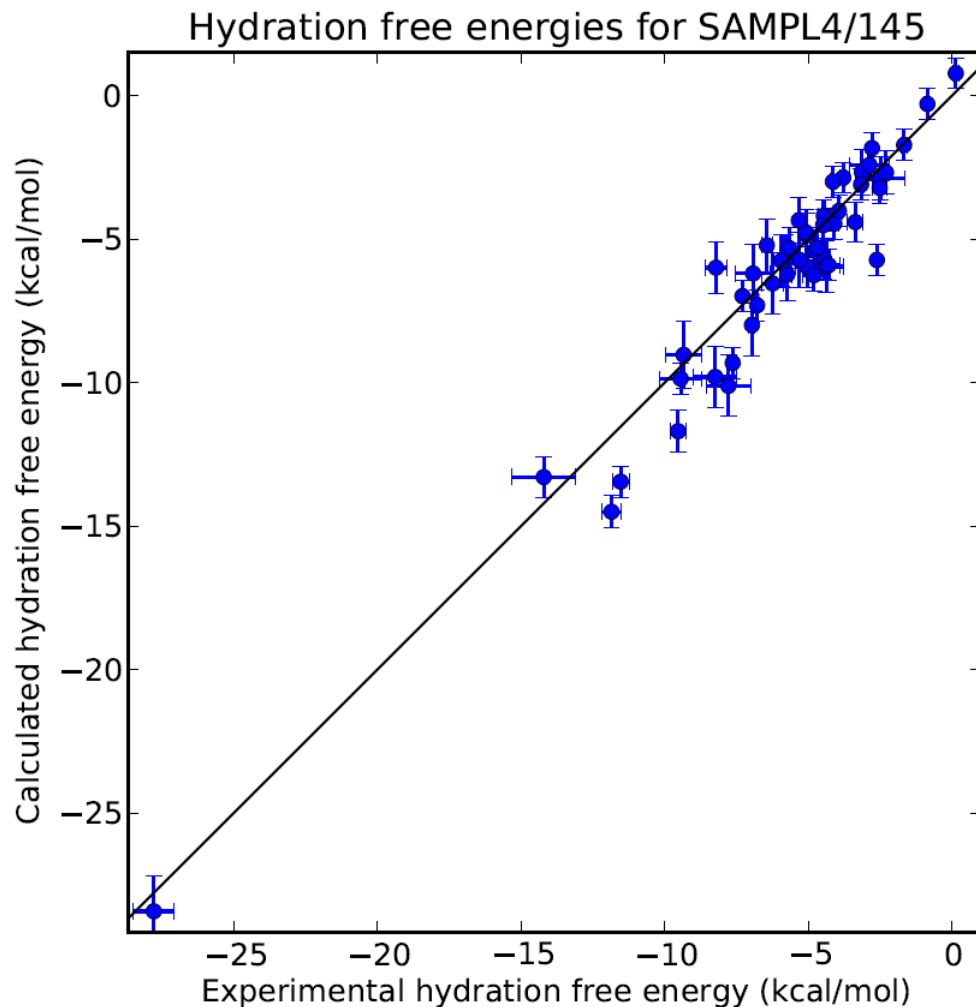


- Calculated hydration free energy

$$\Delta G_{hyd}(M) = \Delta G_{pol}(M) + \Delta G_{elec}(M*) + \Delta G_{LJ}(M*) + \Delta G_{cavity}(M*) = \text{-16.5 kcal/mol}$$

$$+4.6 \qquad -24.0 \qquad -10.9 \qquad +13.8$$

- Post-processing: alcohol FG correction $6 \times (\text{-1.8}) = \text{-10.8}$ kcal/mol

- New prediction:

| | | |
|---|---|---|
| Calc | $\Delta G_{hyd}$(mannitol) = | -27.3 ± 1.24  kcal/mol |
| Exp | | -27.8 ± 0.71 |

# SAMPL4 results



Hydration free energies for SAMPL4/145

Blind test:

- AE = - 0.27 ± 0.25
- AUE = 0.94 ± 0.14
- R = 0.98 ± 0.02
- Tau = 0.71 ± 0.05

Supplementary set:

- AE = - 0.40 ± 0.22
- AUE = 0.82 ± 0.16
- R = 0.96 ± 0.02
- Tau = 0.79 ± 0.05

Total set:

- AE = -0.34 ± 0.15
- AUE = 0.87 ± 0.11
- R = 0.97 ± 0.02
- Tau = 0.77 ± 0.05