

# Minimal-parameter implicit solvent model for large-scale DFT calculations

Jacek Dziedzic<sup>123</sup>, Arihant Bhandari<sup>12</sup>, Chris-Kriton Skylaris<sup>12</sup>

with previous work by

James Womack, Lucian Anton, Hatem H. Helal, Arash A. Mostofi, and Mike Payne

1



2



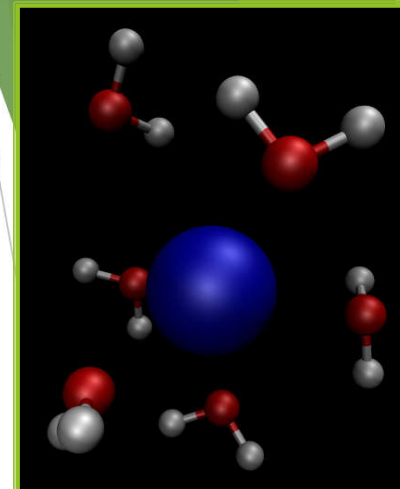
3



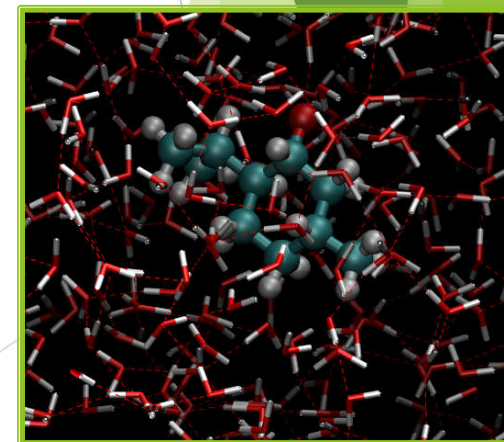
ONETEP Masterclass 2024

## Solvation - why bother?

- ▶ Many important biochemical reactions occur in aqueous solution.
- ▶ Performing calculations *in vacuo* often leads to greatly inaccurate results. Especially sensitive properties and phenomena include, among others:
  - ▶ energy differences between molecular conformers [1],
  - ▶ rates of reactions [2],
  - ▶ tautomeric equilibria [1],
  - ▶ molecular (esp. protein-protein) associations [1,3],
  - ▶ protein structures [4],
  - ▶ ligand binding free energies [5].
- ▶ Thus, it is crucial to include the solvent environment in simulations of biological molecules.



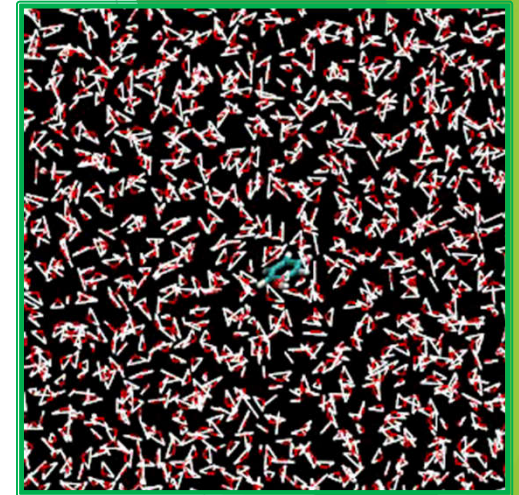
Na<sup>+</sup> and its  
1<sup>st</sup> solvation shell



menthol molecule in water

# Explicit solvent

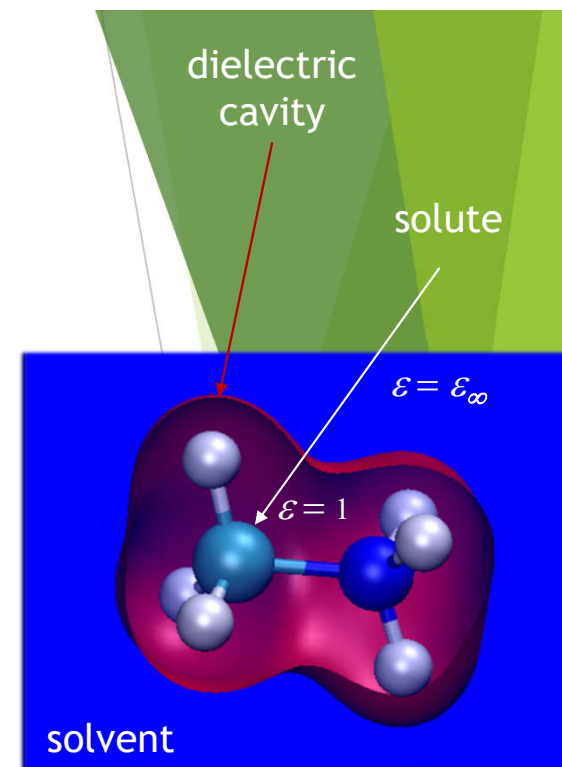
- ▶ In *explicit solvent* methods we introduce the solvent in full atomic detail.
- ▶ (+) Accurate treatment of solute-solvent interactions.
- ▶ (-) Increase in system size, possibly by an order of magnitude.
- ▶ (-) Must average out instantaneous interactions (integrate out the degrees of freedom of solvent).
  - ▶ How to orient the solvent molecules?
  - ▶ How many configurations for averaging?
  - ▶ How to generate these configurations?



Phenol in explicit water  
Animation by Chris Pittock  
(priv. comm.)

# Implicit solvent


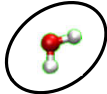



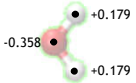

- ▶ Only the solute is treated quantum-mechanically. We place it inside a suitably constructed **dielectric cavity**, whose inside is inaccessible to the solvent.
- ▶ The solvent is represented by an **unstructured dielectric continuum**. We only model its mean effect on the solute.
- ▶ (+) No solvent atoms (low cost).
- ▶ (+) Eliminates the costly sampling of solvent motions.
- ▶ (-) Simplified, mean-field model. All specific interactions between solute and solvent are lost.



Methylammonium in implicit solvent

General idea of *implicit solvent*. The solute is treated quantum-mechanically, while the solvent is represented by a dielectric continuum.

# Implicit solvent pick and mix

cavity shape	 spherical	 ellipsoidal	 interlocking spheres	 defined by charge density
charge representation (of the solute)	 dipole	$\sum_{l,m} c_l^m Y_l^m(\theta, \phi)$ multipole expansion	 collection of point charges	 charge density
method of obtaining the reaction field of the dielectric	analytical solution of GPE	discretising the apparent surface charge (ASC) on the cavity surface	Numerical solution of GPE in the bulk	

Generalised Poisson Equation (GPE)  $\nabla (\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -4\pi \rho(\mathbf{r})$

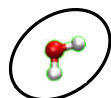


# PCM-type *implicit solvent* models

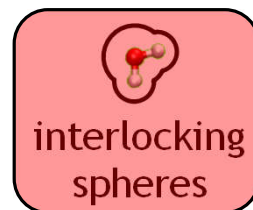
cavity  
shape



spherical



ellipsoidal



interlocking  
spheres



defined by  
charge density

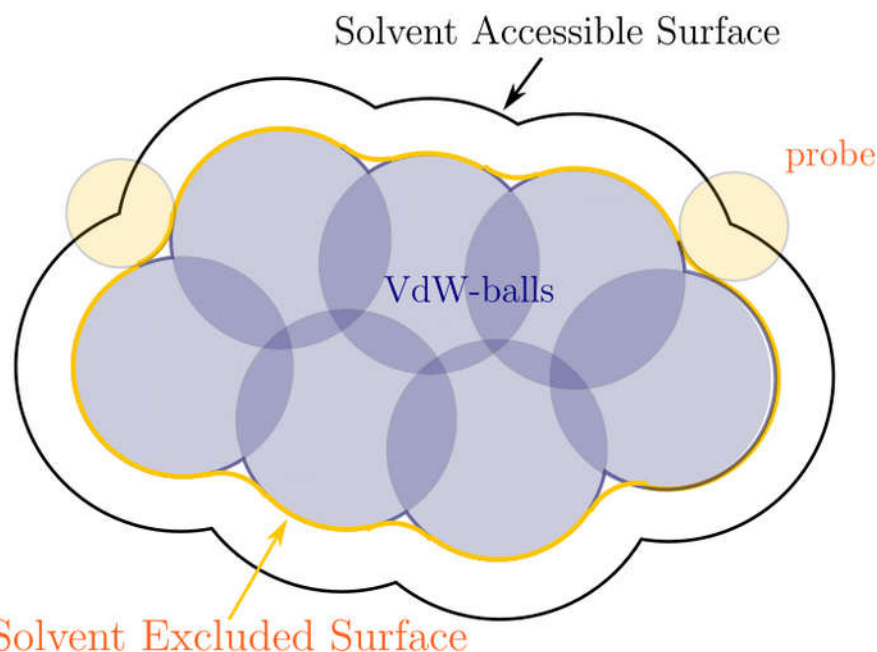


Image source: C. Quan, B. Stamm, *Meshing Molecular Surfaces Based on Analytical Implicit Representation*, *Journal of Molecular Graphics and Modelling* 71 (2016).

Here the dielectric cavity is built from interlocking spheres centered on atoms. Subsequently, a spherical model of the solvent ("probe") is used to establish the solvent-accessible surface (SAS) and solvent-excluded surface (SES).

## Drawbacks:

- Parameterization of the sphere radii is needed for all chemical species in the solvent.
- Insensitive to the oxidation state.

# Elegant *implicit solvent* model

cavity  
shape



spherical



ellipsoidal

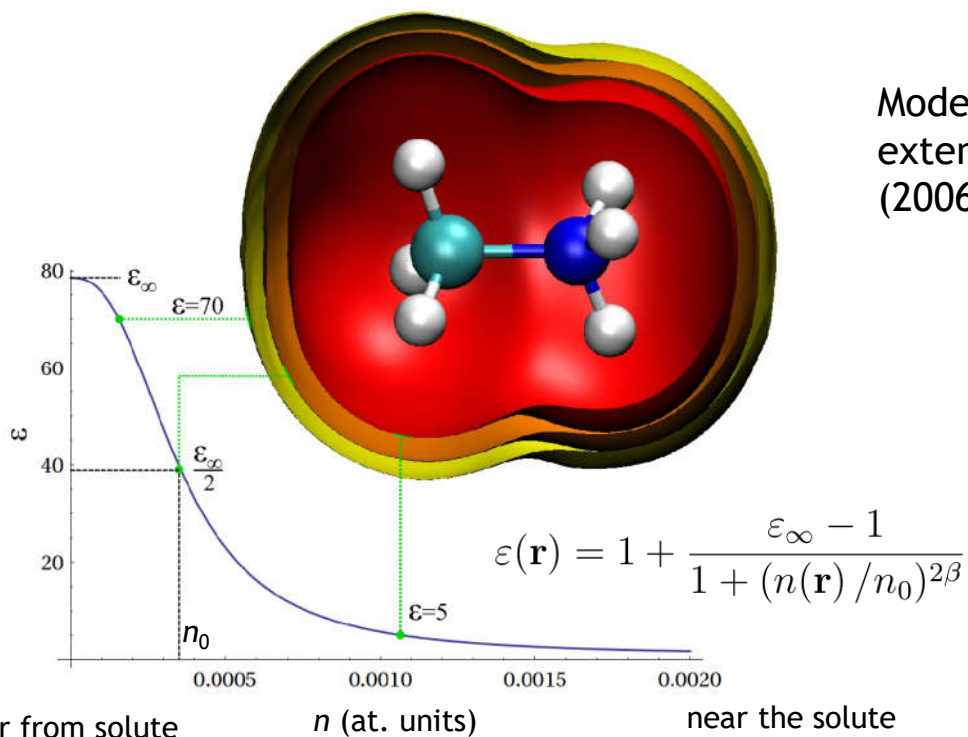


interlocking  
spheres



defined by  
charge density

Model proposed by Fattebert and Gygi (2002),  
extended by Scherlis, Cococcioni and Marzari  
(2006) and revised by us.



far from solute

$n$  (at. units)

near the solute

Dielectric permittivity  $\epsilon(\mathbf{r})$  is defined  
as a function of the charge density  
 $n(\mathbf{r})$ .

Smooth, rather than discrete,  
transition between the permittivity  
of the bulk dielectric and 1.

Only two parameters:  $n_0$  i  $\beta$ .

# Elegant *implicit solvent* model

- ▶ We solve either the Generalized Poisson Equation:

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

- ▶ or the Poisson-Boltzmann equation (when there is electrolyte):

$$\nabla \cdot (\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -4\pi \left[ \rho(\mathbf{r}) + \sum_{i=1}^p z_i c_i(\mathbf{r}) \right]$$

$\{z_i\}_{i=1}^p$  - charges of ion types  
 $\{c_i(\mathbf{r})\}_{i=1}^p$  - their concentration distributions

- ▶ ... to get the electrostatic potential in solvent.



## Two terms in free energy of solvation

only the total  
can be  
obtained  
experimentally  $\rightarrow$

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{npol}$$

(Arrows point from the text above to  $\Delta G_{pol}$  and  $\Delta G_{npol}$  in the equation.)

- ▶ The **electrostatic** or **polar term** describes the response of the solvent to the charge distribution of the solute [7].

- ▶ It is the difference between the electrostatic energy

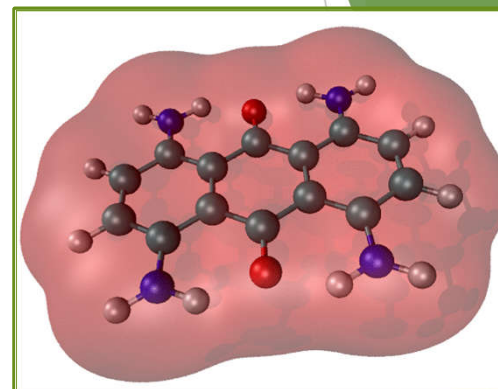
$$\frac{1}{2} \int \rho(\mathbf{r}) \phi(\mathbf{r}) d\mathbf{r}$$

in solvent and in vacuum.

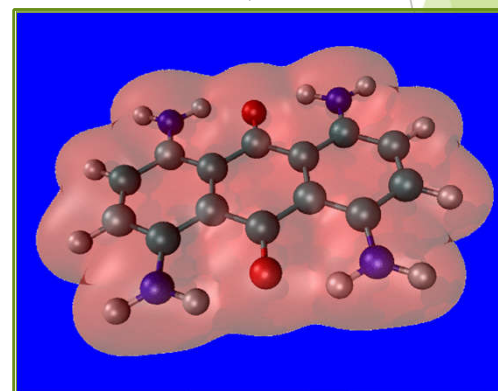
- The **nonpolar term** accounts for
  - the entropic cost of forming a cavity within the solvent (**cavitation energy**),
  - for the van der Waals interaction of the solute with the solvent [4] (**dispersion-repulsion energy**).
- Difficult to describe rigorously.
- A widely used approach is to represent it as a linear function of the molecular surface area [7]:  $\Delta G_{npol} = \gamma A_{SA}$ .

## The procedure

- ▶ Perform a calculation in vacuo to obtain  $E_{vac}$  and the charge density in vacuum.
- ▶ Start a calculation in solvent, using the charge density in vacuum as initial guess.
- ▶ Generate the cavity basing on current charge density.
- ▶ Solve  $\nabla \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r})$  with a multigrid solver to obtain  $\phi(\mathbf{r})$  in solvent. Use this in the electrostatic energy terms.
- ▶ Repeat until convergence in solvent.



First achieve self-consistency in vacuum.



... then in solution.  
NB how the density hardly changes at all in solvent.

# In practice, it's as simple as that

! Turn on auto-solvation

**is\_auto\_solvation T**

! Define the permittivity of your solvent (default: water)

**is\_bulk\_permittivity 78.54**

! Define the Surface tension of your solvent (default: water)

**is\_solvent\_surf\_tension 0.07415 N/m**

Calculation of free energy of solvation:	hartree	kcal/mol
- Total energy <b>in</b> solvent: (+)	-23.22902755568246	-14576.434548
- Total energy <b>in</b> vacuum: (-)	-23.20990671966879	-14564.436043
-----		
- Total free energy of solvation:	-0.01912083601367	-11.998505

The above is a direct calculation of the free energy of solvation as a difference of the in-solvent and in-vacuum energies.

Components of polar term <b>in</b> f.e. of solvation:	hartree	kcal/mol
- Electrostatic:	-0.06759943752720	-42.419287
- Change <b>in</b> nonelectrostatic DFT terms:	0.04263084483500	26.751258
-----		
- Polar term <b>in</b> f.e. of solvation:	-0.02496859269221	-15.668028

The above is the calculation of the polar term to solvation, as a sum of the change in electrostatic energy between in-solvent and in-vacuum and the change in the remaining DFT terms.

Components of free energy of solvation:	hartree	kcal/mol
- Polar term <b>in</b> f.e. of solvation: (+)	-0.02496859269221	-15.668028
- Apolar (cavitation, dis., rep.): (+)	0.00584775667854	3.669523
-----		
- Total free energy of solvation:	-0.01912083601367	-11.998505

Finally, the total free energy of solvation is calculated as the sum of the polar and apolar terms calculated earlier. This is usually what you are after.

# Read the manual!

ONETEP Documentation  
7.1

Search docs

CONTENTS:

- Starting with ONETEP
- Working with GitHub

☰ Ground State Calculation Setup

- Using the Pseudoatomic Solver to Generate NGWFs
- Conduction NGWF optimisation and optical absorption spectra
- Finite-temperature DFT calculations using the Ensemble-DFT method
- Running linear-scaling DFT calculations for metallic systems with the AQUA-FOE method
- Density mixing (Kernel-DIIS)
- Empirical Dispersion Correction
- Using van der Waals Density Functionals
- Realspace local pseudopotential in ONETEP
- ☰ Solvent and Electrolyte Model**
  - Overview of capabilities
  - ☰ The models

🏠 / [Ground State Calculation Setup](#) / Solvent and Electrolyte Model

[View page source](#)

## Solvent and Electrolyte Model

**Author:** Jacek Dziedzic, University of Southampton  
**Author:** James C. Womack, University of Southampton  
**Author:** Arihant Bhandari, University of Southampton  
**Author:** Gabriel Bramley, University of Southampton  
**Date:** September 2022

**This manual pertains to ONETEP versions v6.0.0 and later.**

For older versions, see separate documentation on the ONETEP website.

**Major changes relative to v6.0.0:**

- **Soft-sphere model added in v6.1.1.8**
- **Surface Accessible Volume added in v6.1.3.0**
- **Conjugate gradient solver added in v6.1.3.6**
- **Self-consistent Continuum Solvation (SCCS) model added in v6.1.11.0**
- **Solvation forces in PBC added in v6.1.15.0**
- **Electrolyte forces added in v6.1.15.5**
- **Forces for soft-spheres solvation model added in v6.1.15.9**

**WARNING to users of v6.1.3.0 and later.**

The method used to calculate the surface area of the dielectric cavity was changed in version 6.1.3.0. The surface area is used to calculate the  $\Delta G_{\text{npol}}$  component of the solvation. The new method is more mathematically consistent, but gives approximately 20% smaller values for the surface area. By default, we use the new method, which means the value of  $\Delta G_{\text{solv}}$  and may not

# Do the tutorial!

Search docs

**CONTENTS:**

- Introduction
- Tutorial 1: Setting up Simple ONETEP Calculations
- Tutorial 2: ASE ONETEP interface
- Tutorial 3: Setting up Spin-Polarised Calculations
- Tutorial 4: Geometry optimization
- Tutorial 5: Analysis and visualization
- Tutorial 6: Time-Dependent DFT
- Tutorial 7: Spectral Function Unfolding
- Tutorial 8: Implicit solvation, visualisation and properties: Protein-ligand free energy of binding for the T4 lysozyme**
- Introduction
- Setting up the calculations
- Evaluating the Outputs
- Properties
- Tutorial 9: DFT+ $U$  on strongly correlated magnetic materials: A case study on antiferromagnetic Hematite
- Tutorial 10: Simulation cell relaxation
- Tutorial 11: Electrified electrode-electrolyte interfaces under potentiostatic control
- Tutorial 12: Quantum embedding with (time-dependent) embedded mean-field theory: hydrogenation and excitations of pentacene
- Tutorial 13: Electron Energy Loss Spectroscopy in ONETEP

## Tutorial 8: Implicit solvation, visualisation and properties: Protein-ligand free energy of binding for the T4 lysozyme

**Author:** Lennart Gundelach, Jacek Dziedzic

**Date:** June 2021 (revised June 2023)

### Introduction

#### Protein-Ligand Free Energies of Binding

The binding free energy is a measure of the affinity of the process by which two molecules form a complex by non-covalent association. An example of this, of central importance in biology, is the binding of a ligand to a protein. Many methods to computationally approximate the binding free energies of protein-ligand interactions have been proposed with the ultimate goal of computationally predicting small molecule drug candidates which bind strongly to the protein of interest.

#### Quantum Mechanics in Binding Free Energies

A key limitation common to most computational methods of estimating binding free energies is the assumption of the validity of classical mechanics. The atoms and electrons that constitute biological molecules, like proteins, are, however, governed by the laws of quantum mechanics. Charge transfer, polarization and non-local interactions are not captured by traditional classical mechanical force-fields. Thus, a true description of protein-ligand binding requires a quantum mechanical (QM) treatment of the problem. In theory, a full, *ab-initio* QM approach would be system-independent, parameter-free and would describe the full spectrum of physical phenomena at work.

Unfortunately, high-level QM methods like coupled-cluster (CC) are prohibitively expensive and often have cubic or worse scaling with system size. Thus, even the ligands alone are often too large for routine calculations with these methods.

# Results for small molecules

- ▶ After we devised and implemented several corrections to the FGS model, we obtain very good accuracy with our model (**MPSM**).

- ▶ Mean-square error:

**AMBER** - 3.3 kcal/mol

**PCM** - 4.9 kcal/mol

**FGS** - 5.0 kcal/mol

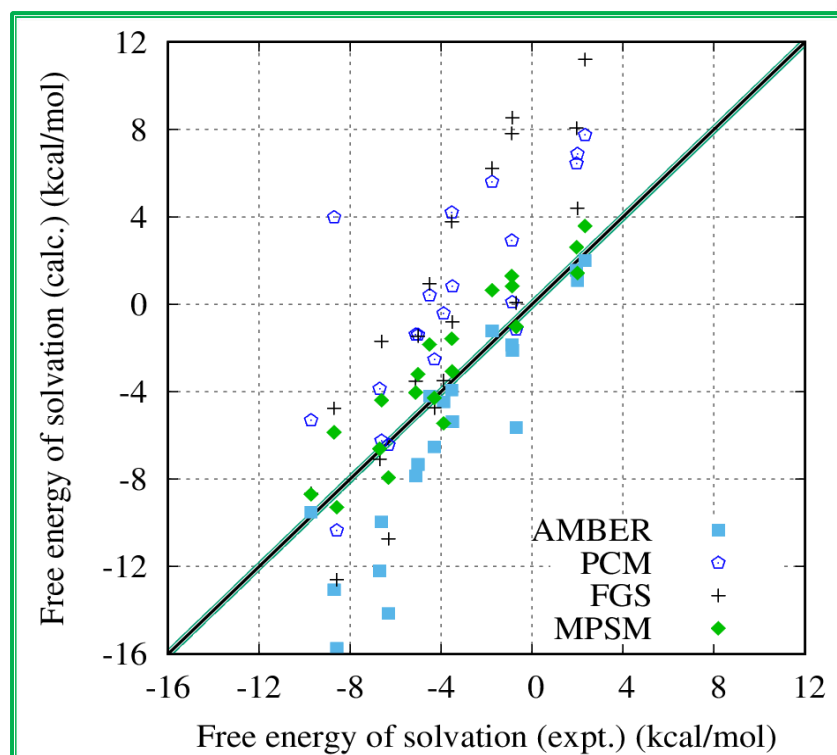
**MPSM** - **1.6** kcal/mol

**AMBER** - classical force field

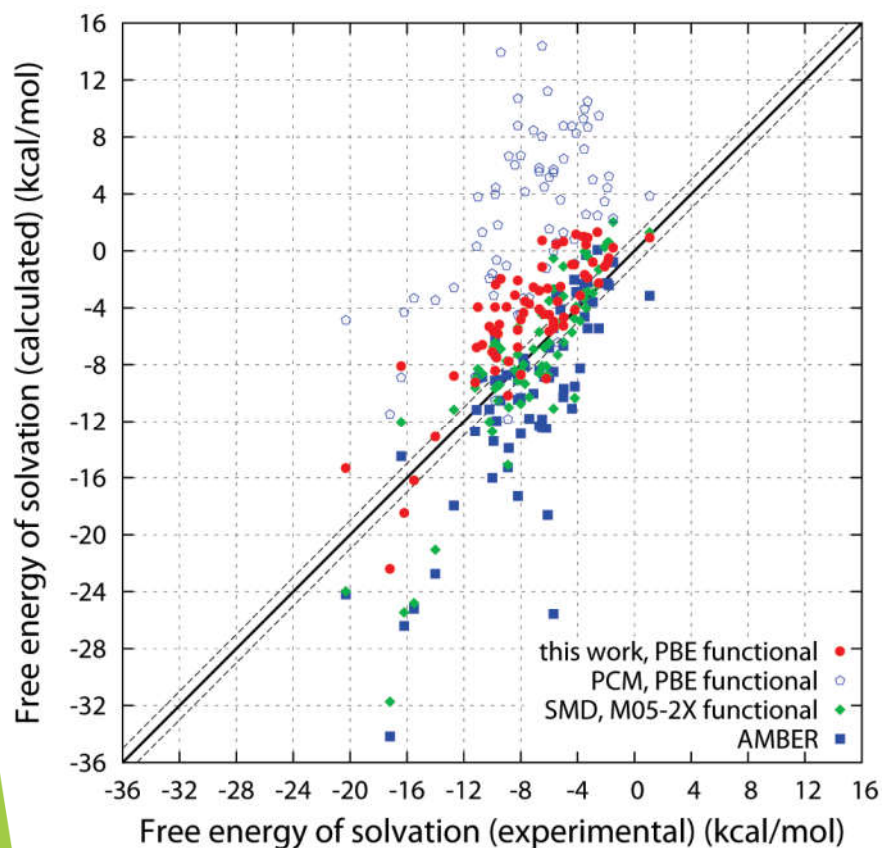
**PCM** - widely-used quantum approach

**FGS** - model before our corrections

**MPSM** - our model



# Results for industrially-relevant molecules



Approach	XC functional	rms error	max error	$r$
● this work <sup>a</sup>	PBE	3.8	8.3	0.83
● this work <sup>b</sup>	PBE	4.1	9.1	0.83
○ PCM	PBE	10.9	23.3	0.53
◆ SMD	M05-2X	3.4	14.5	0.87
■ AMBER	(classical)	5.1	19.9	0.77

Validation on 71 medium-size neutral molecules from

- Nicholls, Mobley, Guthrie, Chodera, Bayly, Cooper and Pande, "Predicting Small-Molecule Solvation Free Energies: An Informal **Blind Test** for Computational Chemistry", *J. Med. Chem.* **51** (2008).
- Guthrie, "A **Blind Challenge** for Computational Solvation Free Energies: Introduction and Overview", *J. Phys. Chem. B* **113** (2009).



## Implicit solvation with thousands of atoms

- ▶ We applied our approach to the full (untruncated) human T4 lysozyme protein to study its free energies of binding to small ligands.
- ▶ Such system sizes (~2600 atoms) are out of reach of conventional DFT.

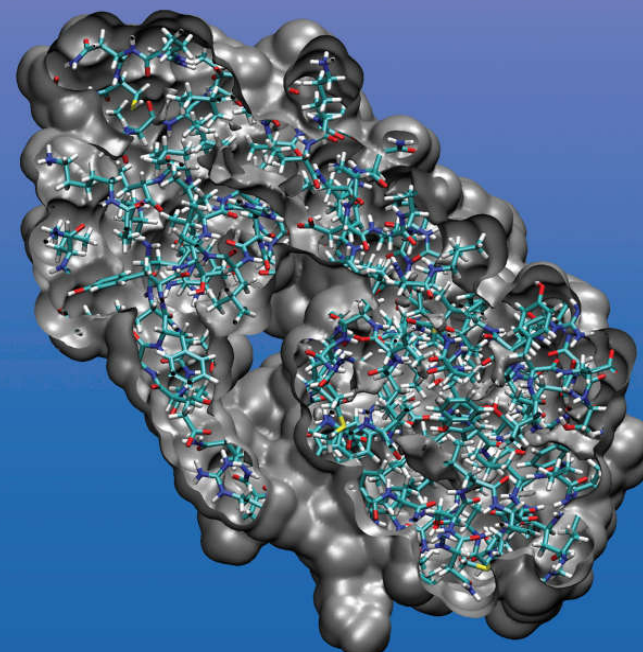
Molecule	$\Delta G_{\text{lig.solv}}^{\text{SMD}}$	$\Delta G_{\text{lig.solv}}^{\text{exp}}$	$\Delta G_{\text{lig.solv}}^{\text{MM}}$	$\Delta G_{\text{lig.solv}}^{\text{QM}}$
Catechol	-9.3	-9.4 [54]	-20.9	-8.1
3-Chlorophenol	-6.7	-	-9.9	-3.6
2-Fluoroaniline	-4.5	-	-5.4	-3.2
2-Aminophenol	-9.8	-	-13.9	-8.0
2-Methylphenol	-6.3	-5.9 [55]	-9.0	-2.9
1-Phenylsemicarbazide	-15.1	-	-16.2	-13.8
Toluene	-1.3	-0.9 [55]	-1.4	1.4
Phenol	-6.7	-6.6 [55]	-9.7	-3.7
Relative hydration energies				
Catechol	-2.5	-2.8	-11.1	-4.3
3-Chlorophenol	0.1	-	-0.1	0.1
2-Fluoroaniline	2.3	-	4.4	0.6
2-Aminophenol	-3.1	-	-4.1	-4.2
2-Methylphenol	0.4	0.7	0.8	0.9
1-Phenylsemicarbazide	-8.4	-	-6.4	-10.1
Toluene	5.4	5.7	8.3	5.1
Phenol (reference)	0.0	0.0	0.0	0.0
Max error <sup>a</sup>			8.6	1.8
RMS error <sup>a</sup>			3.4	1.2

our protein,  
with the  
dielectric  
cavity shown



International Journal of  
**QUANTUM**  
**CHEMISTRY**

www.q-chem.org



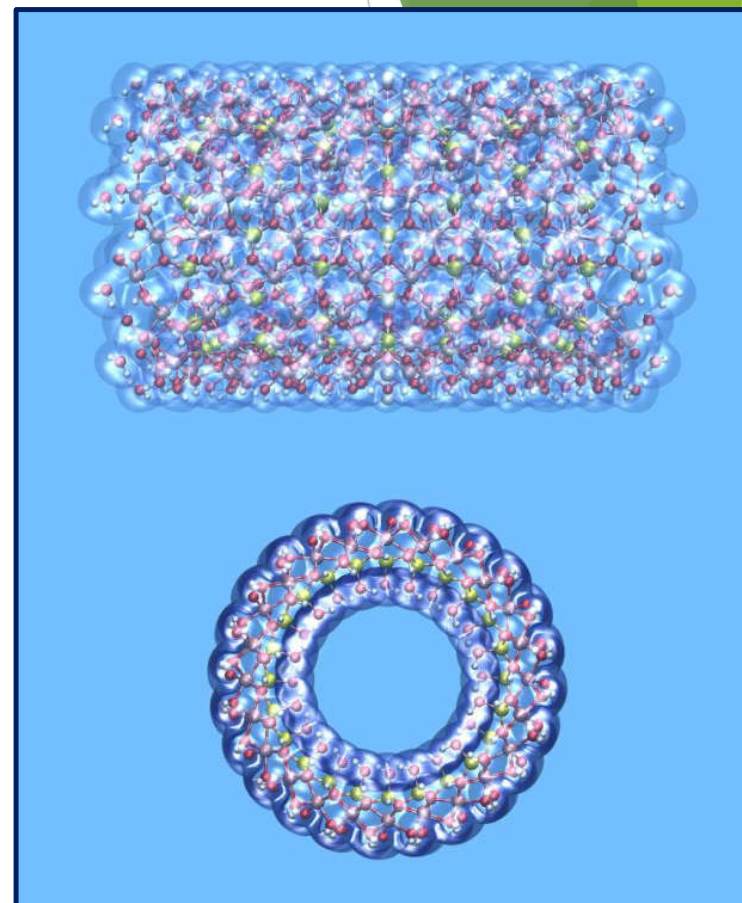
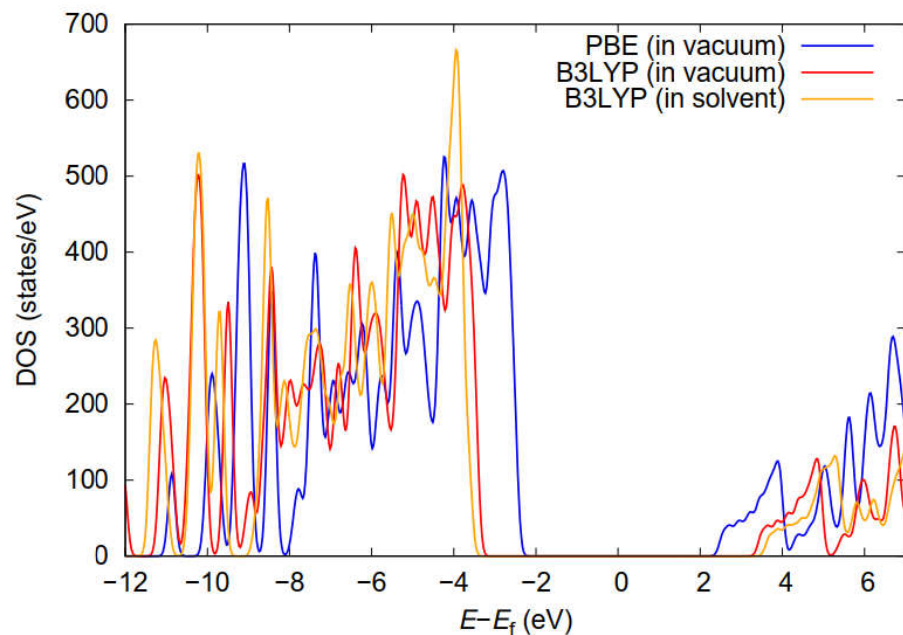
Special Issue: 14th International Density Functional Theory Conference  
Guest Editors: Nikitas I. Gidopoulos and Andreas K. Theophilou

WILEY

Volume 113 | Issues 5-6 | 2013  
Included in this print edition:  
Issue 5 (March 5, 2013)  
Issue 6 (March 15, 2013)

## Implicit solvation with thousands of atoms

- Hybrid functional (B3LYP) calculations on an **aluminosilicate imogolite nanotube** with 1416 atoms and implicit solvent, with near-complete basis set accuracy.



## Other features not discussed here

- ▶ **Full support for PBCs and OBCs.** OBCs are used by default.
- ▶ **Forces from all solvation terms:**
  - ▶ geometry optimisation in solvent is possible,
  - ▶ MD in solvent is possible.
- ▶ **Alternative solvation models available in ONETEP:**
  - ▶ Fisicaro's soft-sphere model.
  - ▶ Andreussi's Self-Consistent Continuum-Solvation model.
- ▶ **Solvent exclusion regions.**
- ▶ **Electrolyte.**

## Conclusions

- ▶ Building on the isodensity model of Fattebert and Gygi, we have developed a solvation model which:
  - ▶ is based on first principles,
  - ▶ has predictive power that is superior to classical models and to PCM,
  - ▶ has only two parameters ( $\beta$ ,  $n_0$ ), whose values have been optimized and which appear to be universal (do not depend on the solute),
  - ▶ uses only two fundamental quantities ( $\epsilon$ ,  $\gamma$ ), with clear physical interpretation, to describe the solvent.
- ▶ With our current implementation we can do implicit solvent calculations on systems with  $\sim 10^4$  atoms within a day.
- ▶ (!) **ONETEP is free for academics.** More info including tutorials and case studies: [www.onetep.org](http://www.onetep.org).

## Recommended reading

- ▶ J. Dziejczak, H. H. Helal, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, *Minimal parameter implicit solvent model for ab initio electronic structure calculations*, Europhysics Letters 95, 43001 (2011).
- ▶ J. Dziejczak, S. J. Fox, T. Fox, C. S. Tautermann, and C.-K. Skylaris, *Large-scale DFT calculations in implicit solvent - a case study on the T4 lysozyme L99A/M102Q protein*, Int. J. Quantum Chem 113, 771 (2013).
- ▶ S. Fox, J. Dziejczak, T. Fox, C.S. Tautermann, and C.-K. Skylaris, *DFT calculations on entire proteins for free energies of binding: application to a model polar binding site*, Proteins: Structure, Function and Bioinformatics 82, iss. 12 (2014).

# References

1. Fattebert and Gygi, *J. Comp. Chem.* **23** (2002).
2. Foresman, Keith, Wiberg, Snoonian and Frisch, *J. Phys. Chem.* **100** (1996).
3. Moreira, Fernandes and Ramos, *J. Mol. Struct.: THEOCHEM* **729** (2005).
4. Feig and Brooks III, *Curr. Opinion in Struct. Biol.* **14** (2004).
5. Chen, Brooks III and Khandogin, *Curr. Opinion in Struct. Biol.* **18** (2008).
6. Dong, Wagoner and Baker, *Phys. Chem. Chem. Phys.* **10** (2008).
7. Baker, *Curr. Opinion in Struct. Biol.* **15** (2005).
8. Feig, Onufriev, Lee, Wonpil, Case and Brooks III, *J. Comput. Chem* **25** (2004).
9. Miertus, Scrocco and Tomasi, *Chem. Phys.* **55** (1981).
10. Fattebert and Gygi, *International Journal of Quantum Chemistry* **93** (2003).
11. Scherlis, Fattebert, Gygi, Cococcioni and Marzari, *J. Chem. Phys.* **124** (2006).
12. Holst, "Multilevel Methods for the Poisson-Boltzmann Equation", PhD thesis (1993).
13. Trottenberg, Oosterles and Schuller, "Multigrid" (2001).
14. Schaffer, *Math. Comp.* **43** (1984).
15. Clark, Segall, Pickard, Hasnip, Probert, Refson and Payne, *Zeitschrift für Kristallographie* **220** (2005).
16. Skylaris, Haynes, Mostofi and Payne, *J. Chem. Phys.* **122** (2005).
17. Floris, Tomasi and Ahuir, *J. Comp. Chem.* **12** (1991).
18. Marenich, Kelly, Thompson, Hawkins, Chambers, Giesen, Winget, Cramer, and Truhlar, *Minnesota Solvation Database - version 2009*, University of Minnesota, Minneapolis, 2009.
19. Marenich, Cramer, and Truhlar, *J. Phys. Chem. B* **113** (2009).
20. Cramer and Truhlar in "Free Energy Calculations in Rational Drug Design" (2001).
21. Shestakov, Milovich and Noy, *Journal of Colloid and Interface Science* **247** (2002).
22. Lu, Zhang and McCammon, *J. Chem. Phys.* **122** (2005).
23. Fogolari, Brigo and Molinari, *Journal of Molecular Recognition* **15** (2002).
24. Tomasi and Persico, *Chem. Rev.* **94** (1994).
25. Swanson, Henchman and McCammon, *Biophysical Journal* **86** (2004).
26. Rocchia, Sridharan, Nicholls, Alexov, Chiaberra and Honig, *J. Comput. Chem.* **23** (2002).
27. Tan, Yang and Luo, *J. Phys. Chem. B* **110** (2006).
28. Tomasi, Mennucci and Cammi, *Chem. Rev.* **105** (2005).
29. Sanchez, Sued and Scherlis, *J. Chem. Phys.* **131** (2009).
30. Kaukonen, Soederhlejm, Heimdal and Ryde, *J. Phys. Chem. B* **112** (2008).
31. Baker, Sept, Joseph, Holst and McCammon, *Proc. Natl. Acad. Sci. USA* **98** (2001).