



UNIVERSITY OF
Southampton
School of Chemistry

Implicit Solvation Models in ONETEP

Jacek Dziedzic

School of Chemistry
University of Southampton

ONETEP Summer School, April 2010

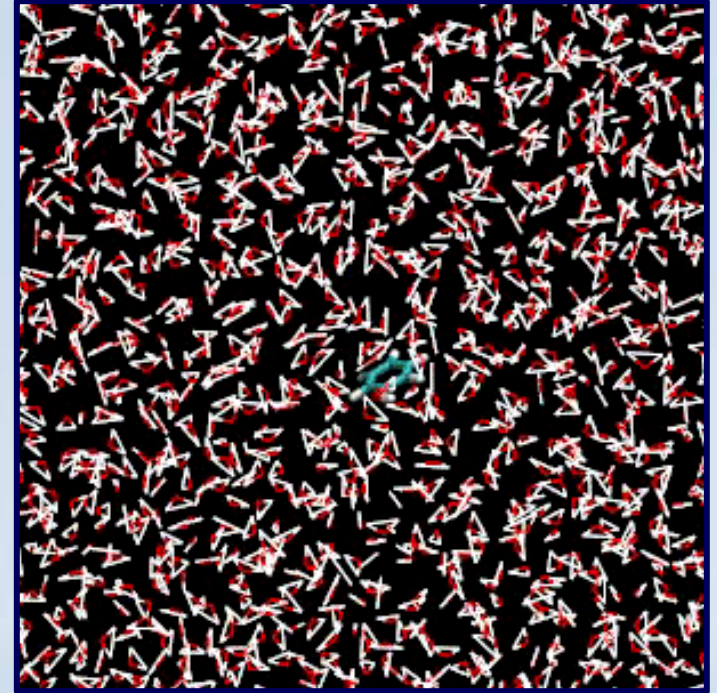
Importance of solvation

- 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],
 - π -facial sensitivity [2],
 - molecular (esp. protein-protein) associations [1,3],
 - **protein structures** [4],
 - **ligand binding free energies** [5].
- Thus proper description of the solvent environment is crucial in simulations of biological molecules.

Explicit solvent approach

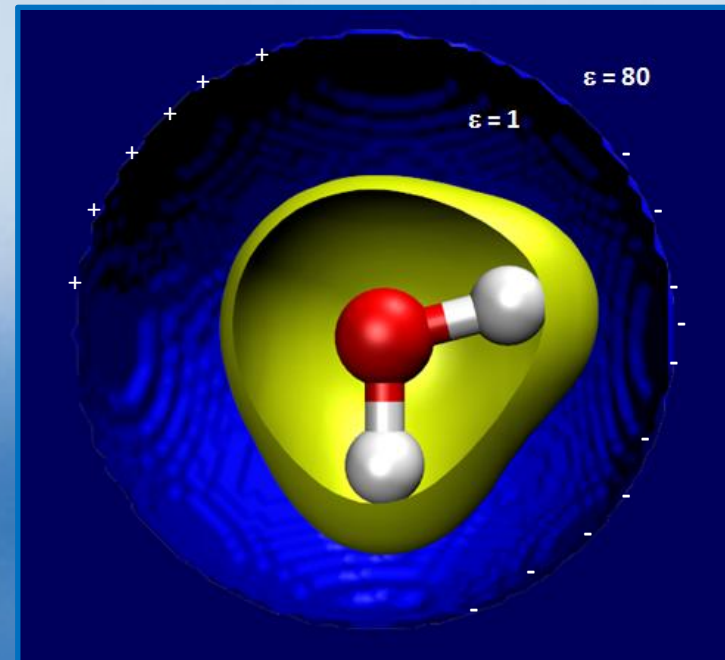
Animation by C. Pittock

- Introduce the solvent in molecular detail.
- Pros:
 - provides accurate treatment of solute-solvent interactions,
- Cons:
 - leads to an increase in system size, possibly by an order of magnitude [6],
 - must average out instantaneous interactions before the results become meaningful, ("integrate out" the degrees of freedom of the solvent) [4],
 - to avoid surface effects, PBCs must be introduced. This in turn requires large cells to avoid artificial interactions between replicated solutes.



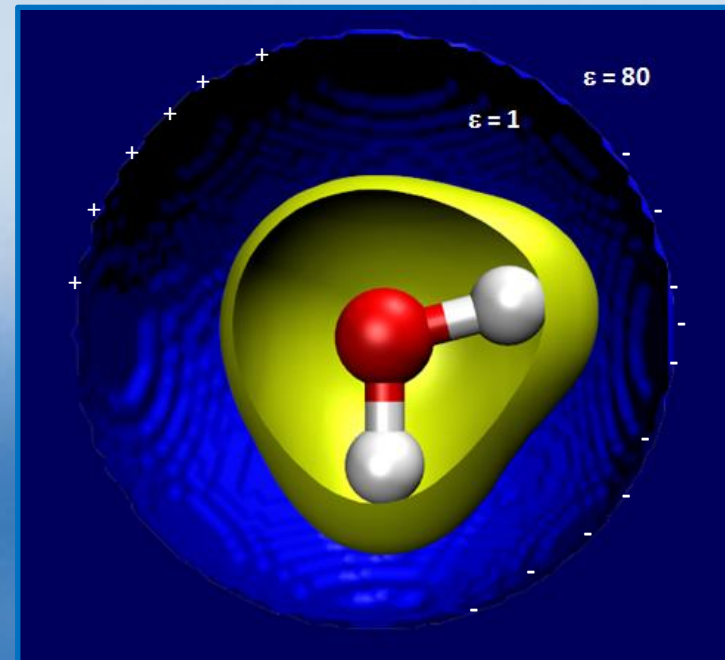
Implicit solvent approach

- Treat the solute explicitly and embed it within a suitably defined cavity, the inside of which is inaccessible to the solvent.
- Replace the solvent with an *unstructured dielectric continuum*, only retaining its **average** effect on the solute [7].
- The charge distribution of the solute polarizes the dielectric, creating a reflection charge, which then interacts with the solute [2].

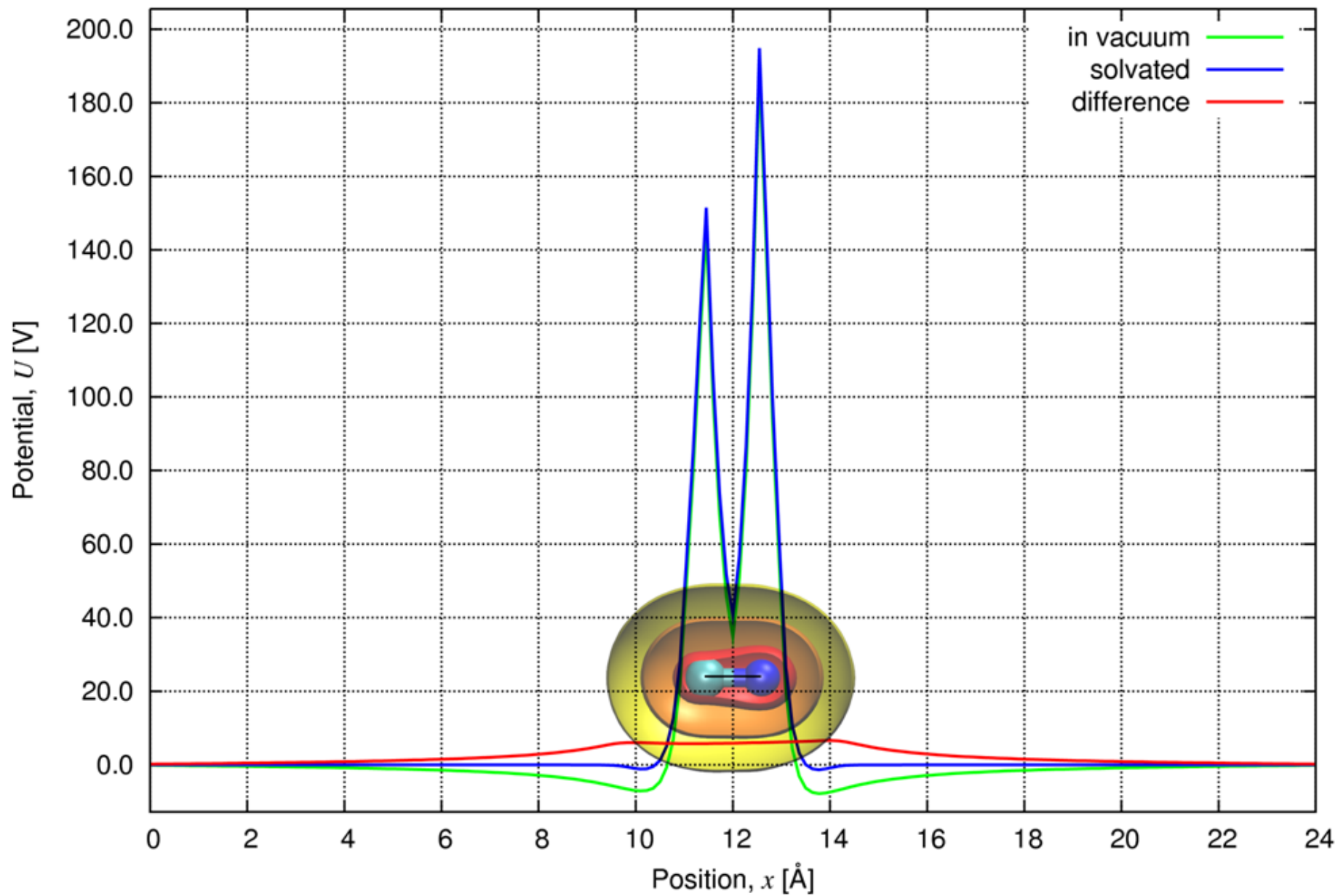


Implicit solvent approach

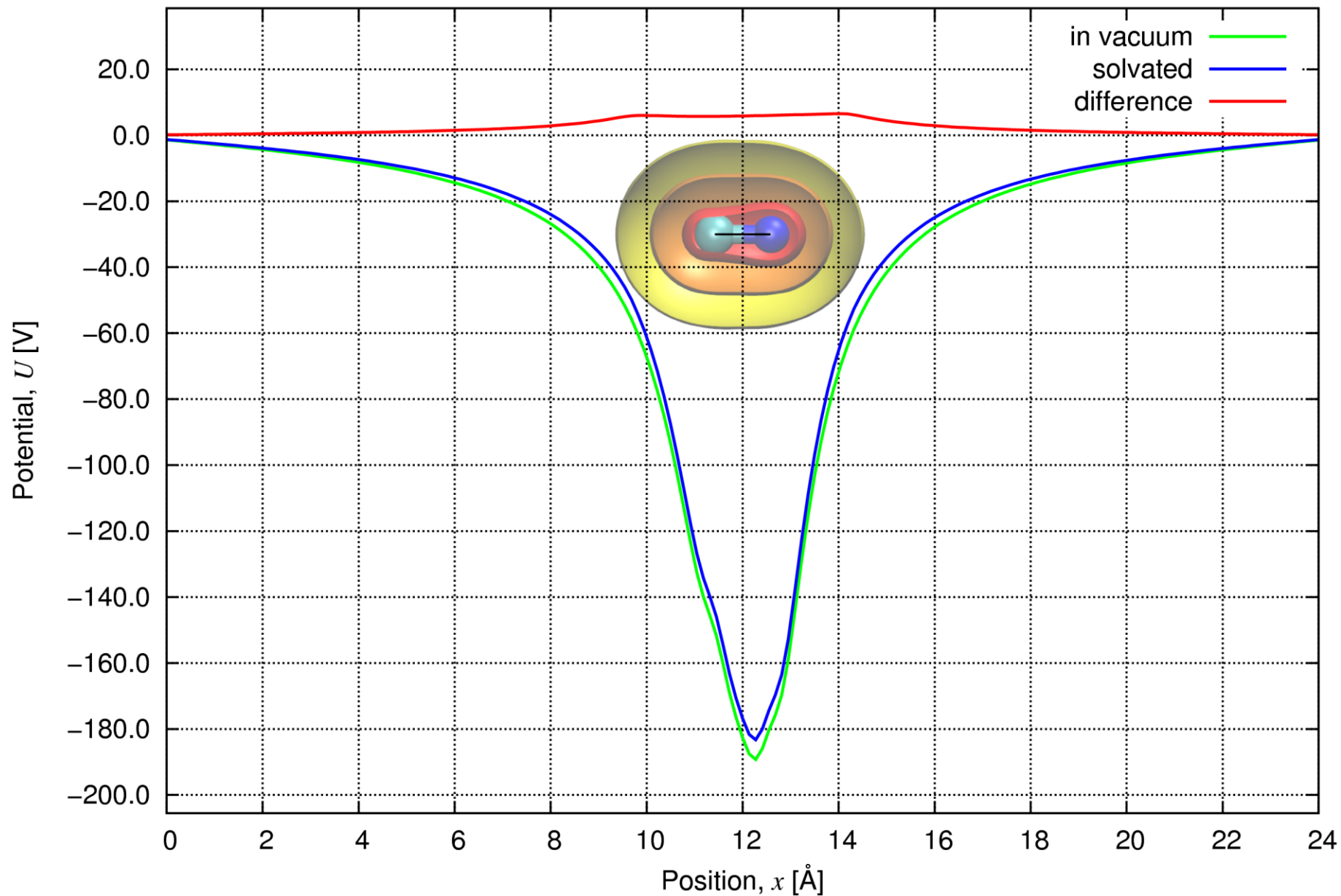
- In other words: one assumes that the macroscopic description of the solvent as continuous dielectric medium can be used as an approximation on the microscopic scale [8].
- This works because the largest part of molecular interactions in solution is characterized by low specificity and low directionality [9].
- Pros:
 - no solvent atoms necessary,
 - eliminates the costly sampling of the solvent degrees of freedom [6],
 - faster sampling of solute conformations owing to absence of solvent friction [5].



Effect of solvent on the potential, for CN^-



Effect of solvent on the potential, for CN^-



Two terms in free energy of solvation

only the total can

be obtained
experimentally

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

- 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.

Obtaining $\phi(\mathbf{r})$

- **In vacuum** the Poisson equation is solved:

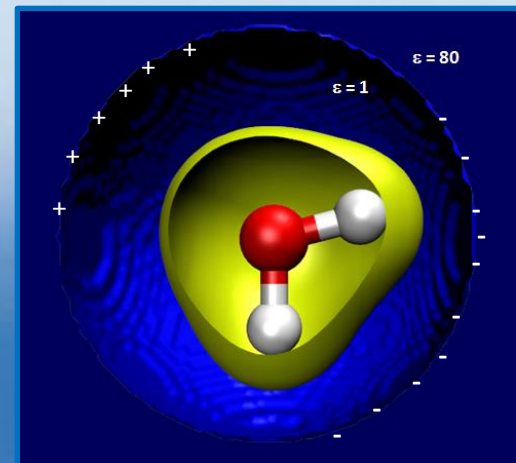
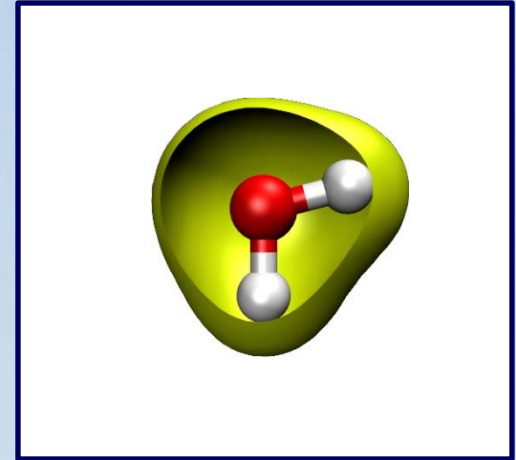
$$\epsilon \nabla^2 \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r})$$

This is easiest to compute in reciprocal space and that's the way it's done in ONETEP.

- **With an inhomogeneous dielectric** the *generalized* Poisson equation needs to be solved:

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

This must be solved in real space with a suitable solver.



Two terms in solvation energy

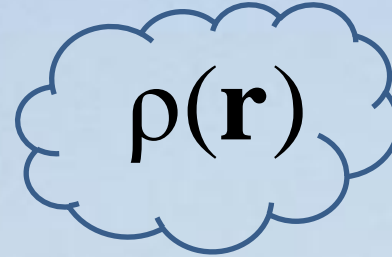
$$\Delta G_{\text{sol}} = \Delta G_{\text{pol}} + \Delta G_{\text{npol}}$$


- The **nonpolar term** accounts for the entropic cost of forming a cavity within the solvent and for the van der Waals interaction of the solute with the solvent [4].
- Difficult to describe rigorously, heuristic approaches are used.
- Most widely used approach is to represent it as a linear function of the molecular surface area [7]:

$$\Delta G_{\text{npol}} = \gamma A_{\text{SA}}$$

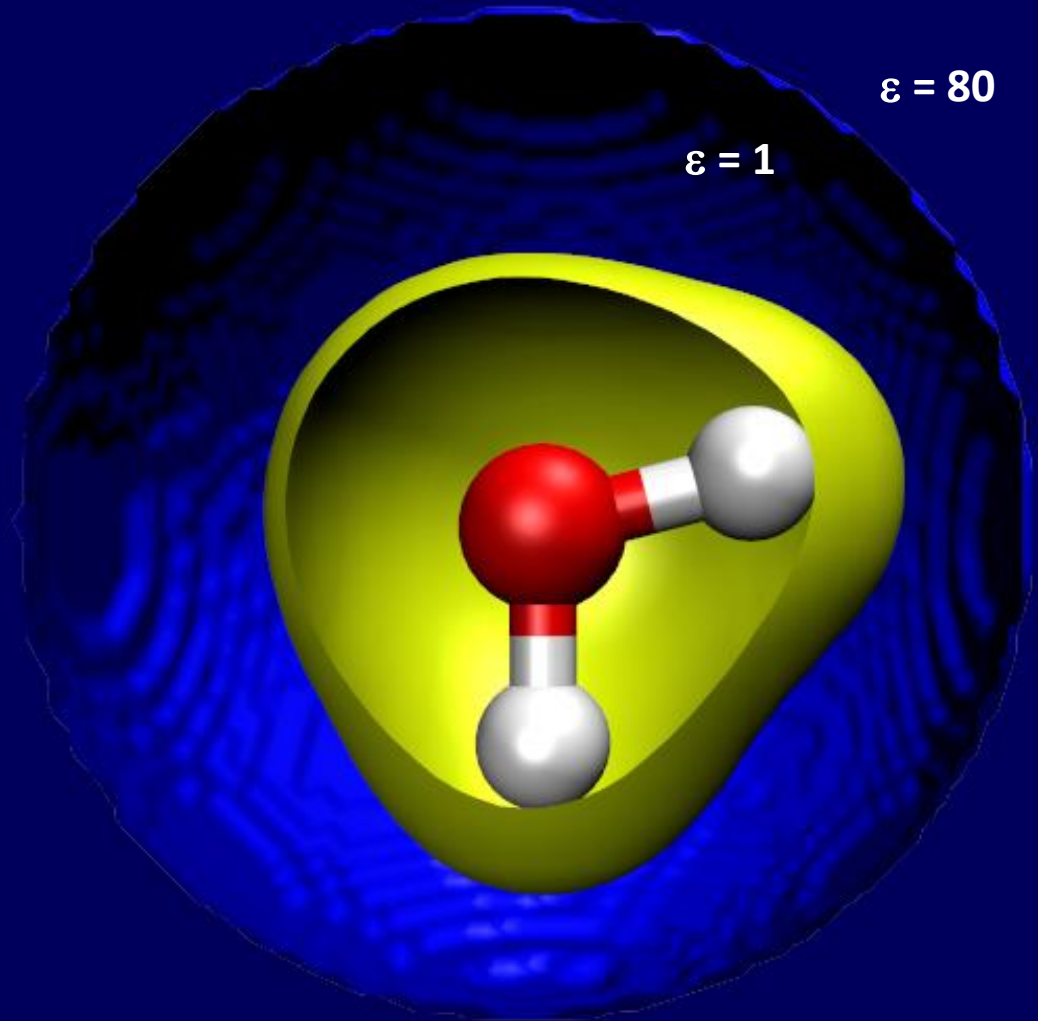
Important choices

- How the charges are represented:
 - as classical point charges,
 - vs.
 - distributed (charge densities).
- How solvation is treated:
 - for a frozen system: the energy of solvation is determined once, by adding the implicit solvent to the system and observing the resulting change in energy,
 - self-consistently: by re-calculating the solvent's influence during every step of energy minimization. This allows the electronic charge density to „feel“ the polarizing effect of the solvent and to adapt its shape accordingly.
- Cavity shape.



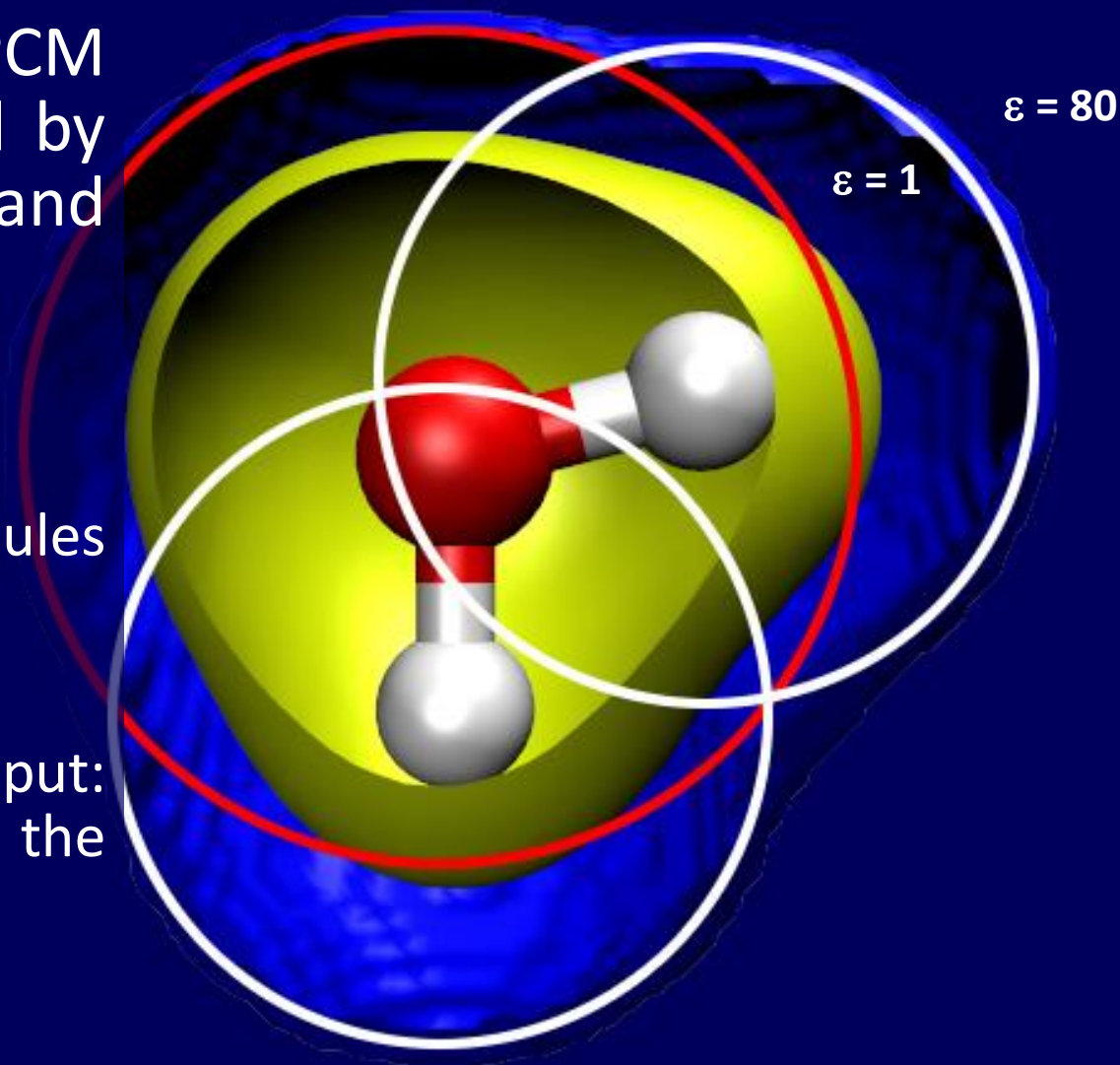
Spherical cavity

- Pros:
 - simple (to implement),
 - simple (runs fast).
- Cons:
 - simple (not especially realistic).
 - molecules that are not compact cannot be easily fit into a spherical cavity.
- Has enjoyed considerable success nevertheless ([2] & refs therein).

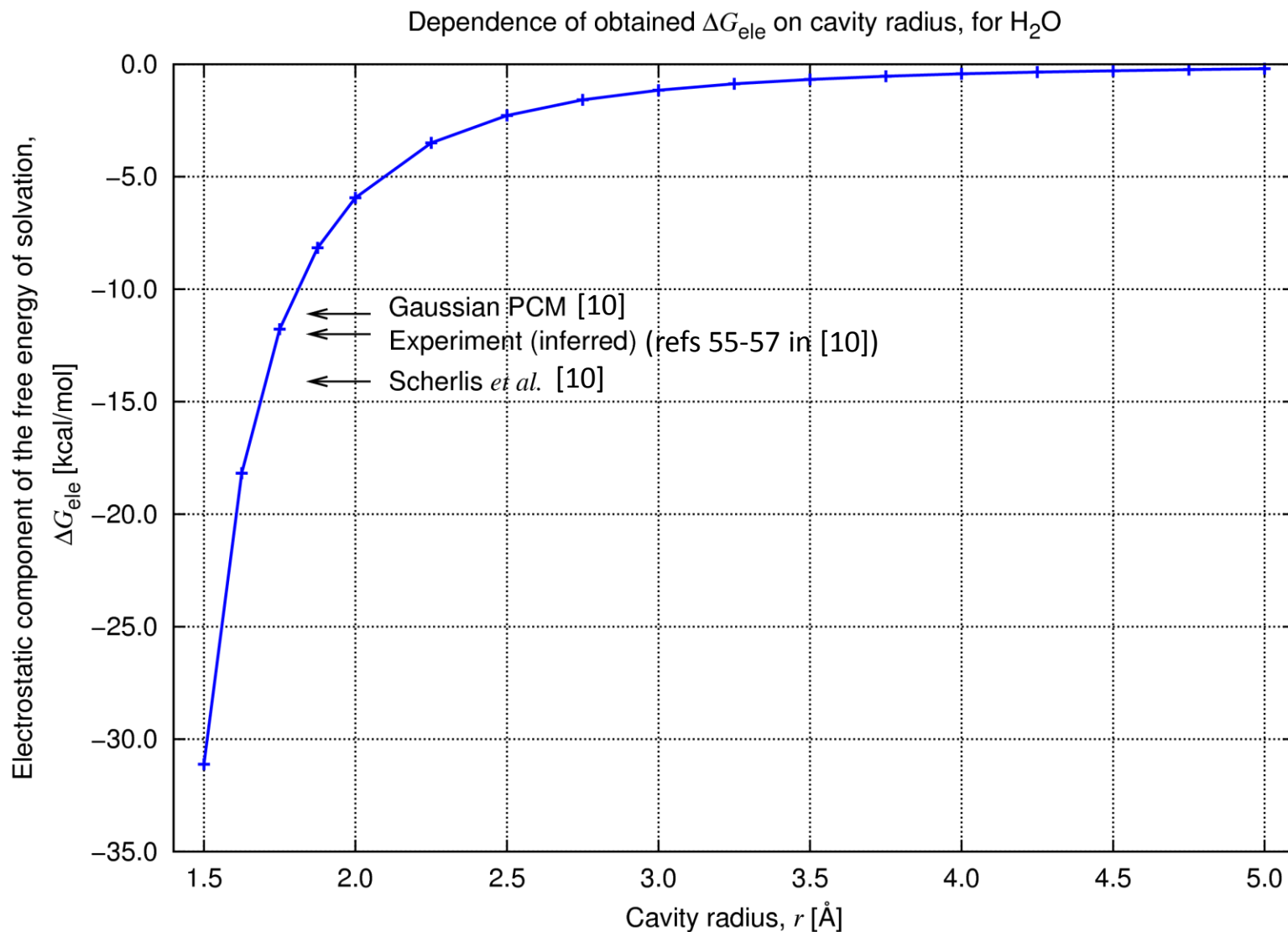


Interlocking spheres

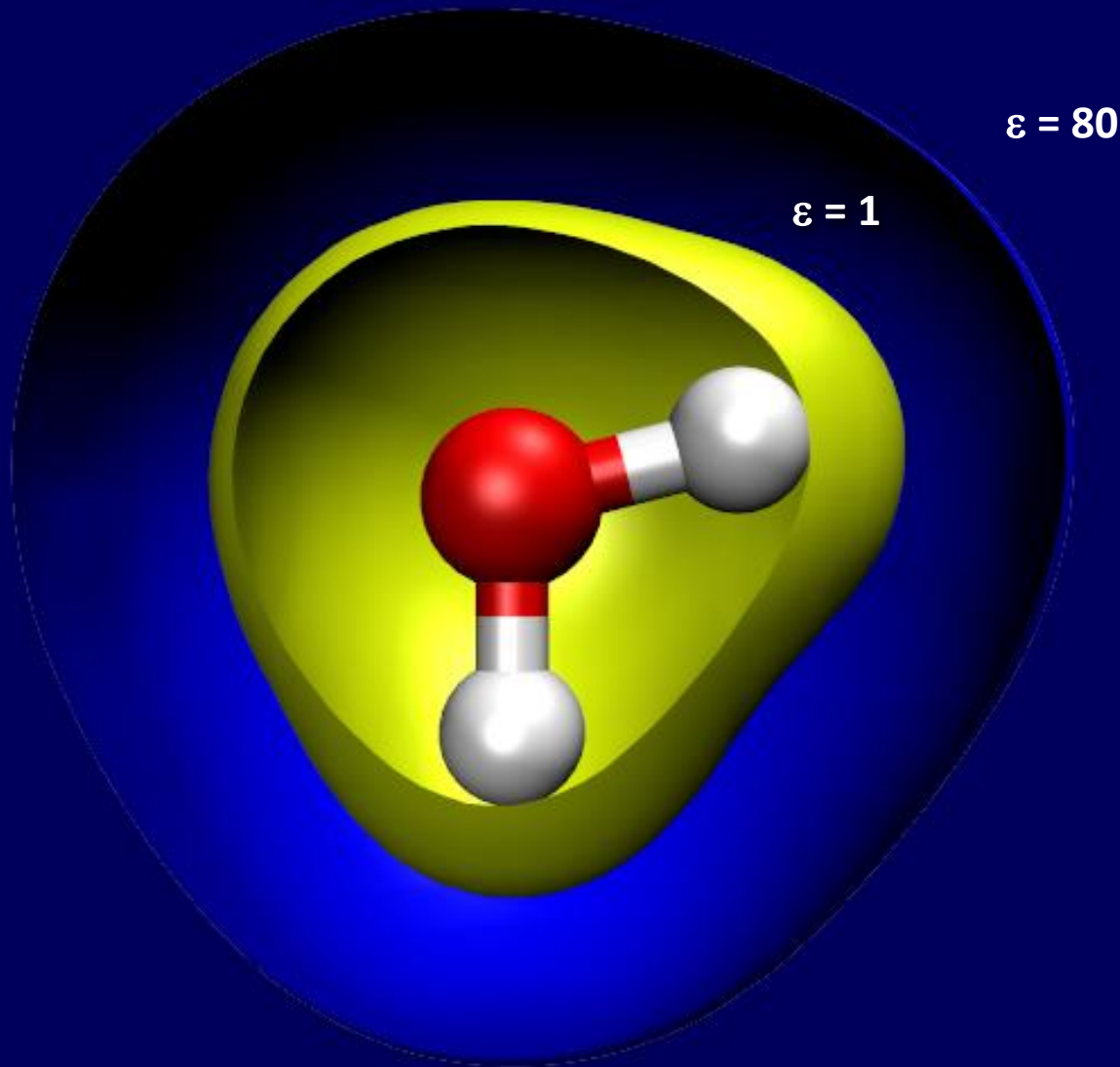
- Employed in the PCM approach developed by Tomasi, Scrocco and Miertus [9].
- Pros:
 - More realistic cavity.
 - Applicable to molecules of any shape.
- Cons:
 - Requires extra input: parametrization for the sphere radii.



Arbitrariness of the cavity

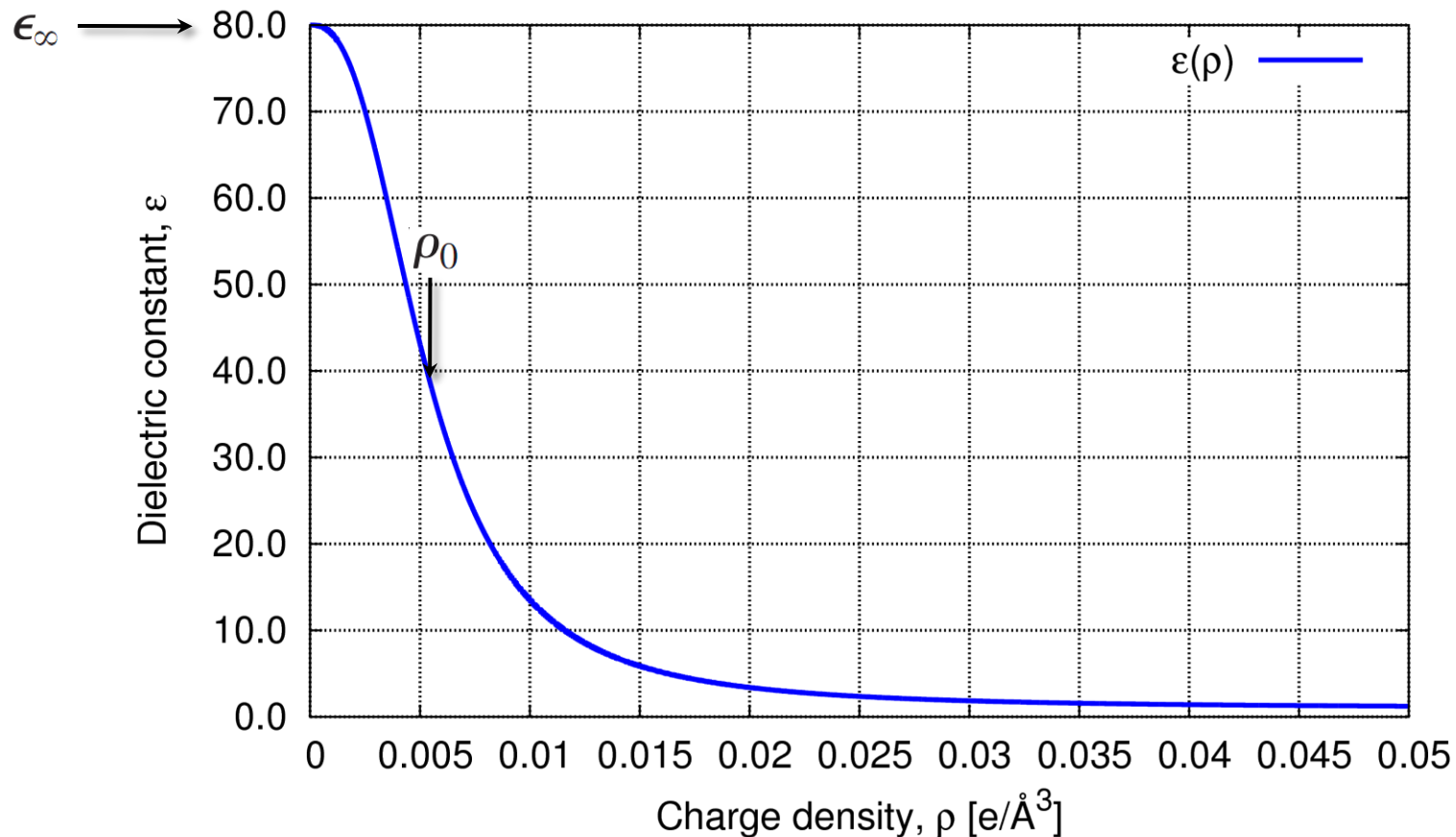


Density-dependent cavity

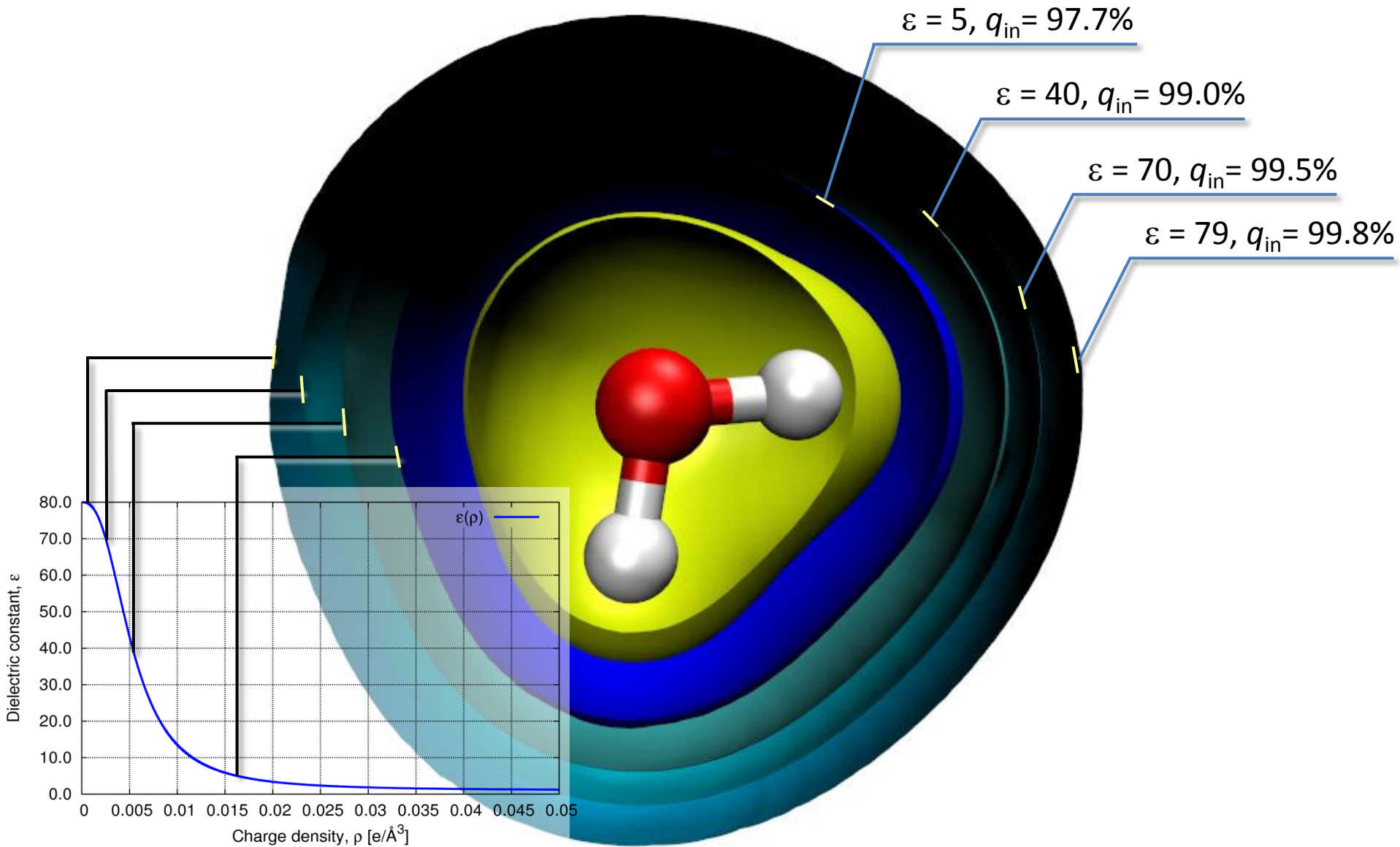


Density-dependent cavity of Fattebert and Gygi

$$\epsilon(\rho(\mathbf{r})) = 1 + \frac{\epsilon_{\infty} - 1}{2} \left(1 + \frac{1 - (\rho(\mathbf{r})/\rho_0)^{2\beta}}{1 + (\rho(\mathbf{r})/\rho_0)^{2\beta}} \right)$$



Density-dependent cavity



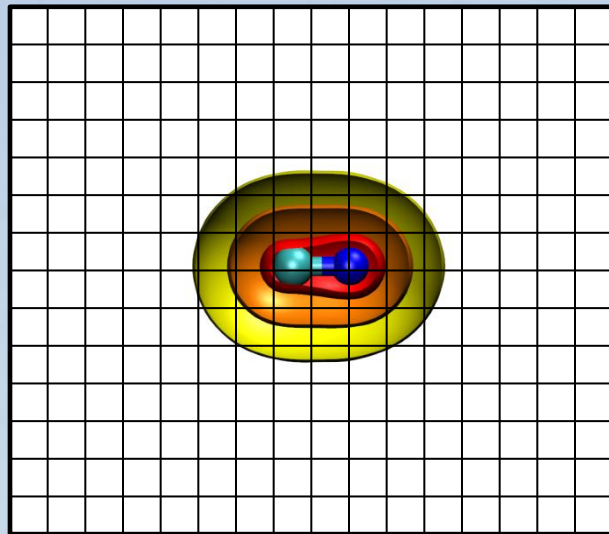
Implementing IS in ONETEP

- Interface with APBS [10] (Adaptive Poisson-Boltzmann Solver) code.
- APBS is a multigrid solver that can calculate $\phi(\mathbf{r})$ given $\rho(\mathbf{r})$ and $\epsilon(\mathbf{r})$. It also automatically calculates the polar term of the solvation energy as:

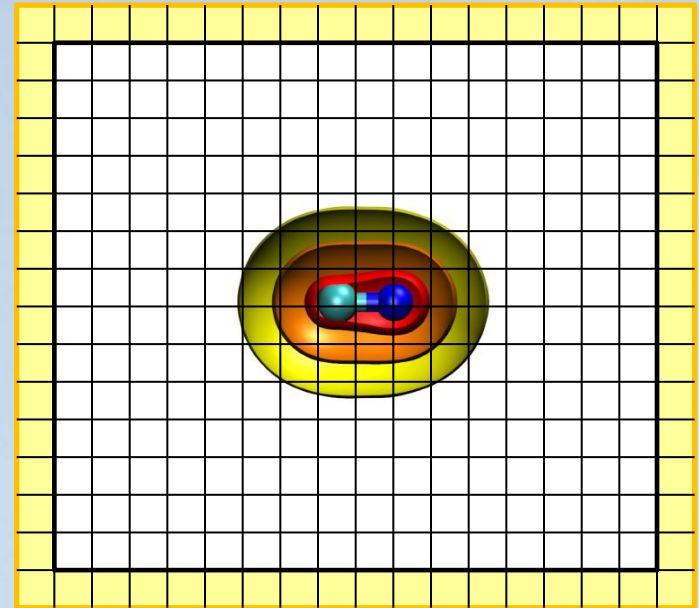
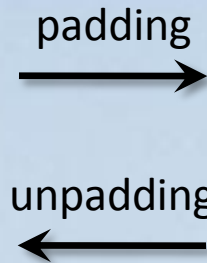
$$\Delta G_{\text{pol}} = \frac{1}{2} \int \rho(\mathbf{r}) \phi_{\epsilon=80}(\mathbf{r}) d\mathbf{r} - \frac{1}{2} \int \rho(\mathbf{r}) \phi_{\epsilon=1}(\mathbf{r}) d\mathbf{r}$$

- We are currently pursuing both the frozen density approach and the self-consistent one.

Grid and padding



ONETEP's box

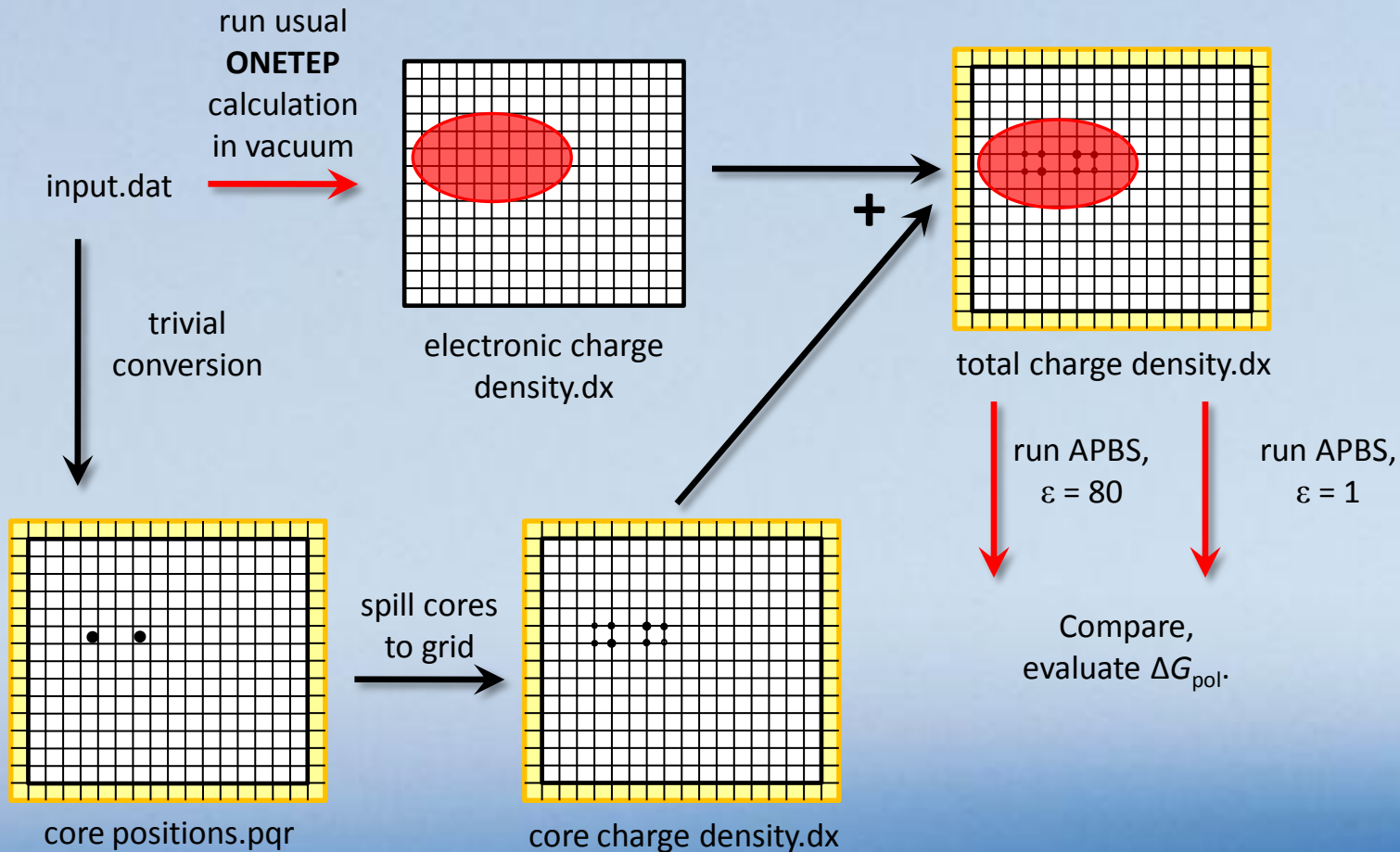


APBS's box

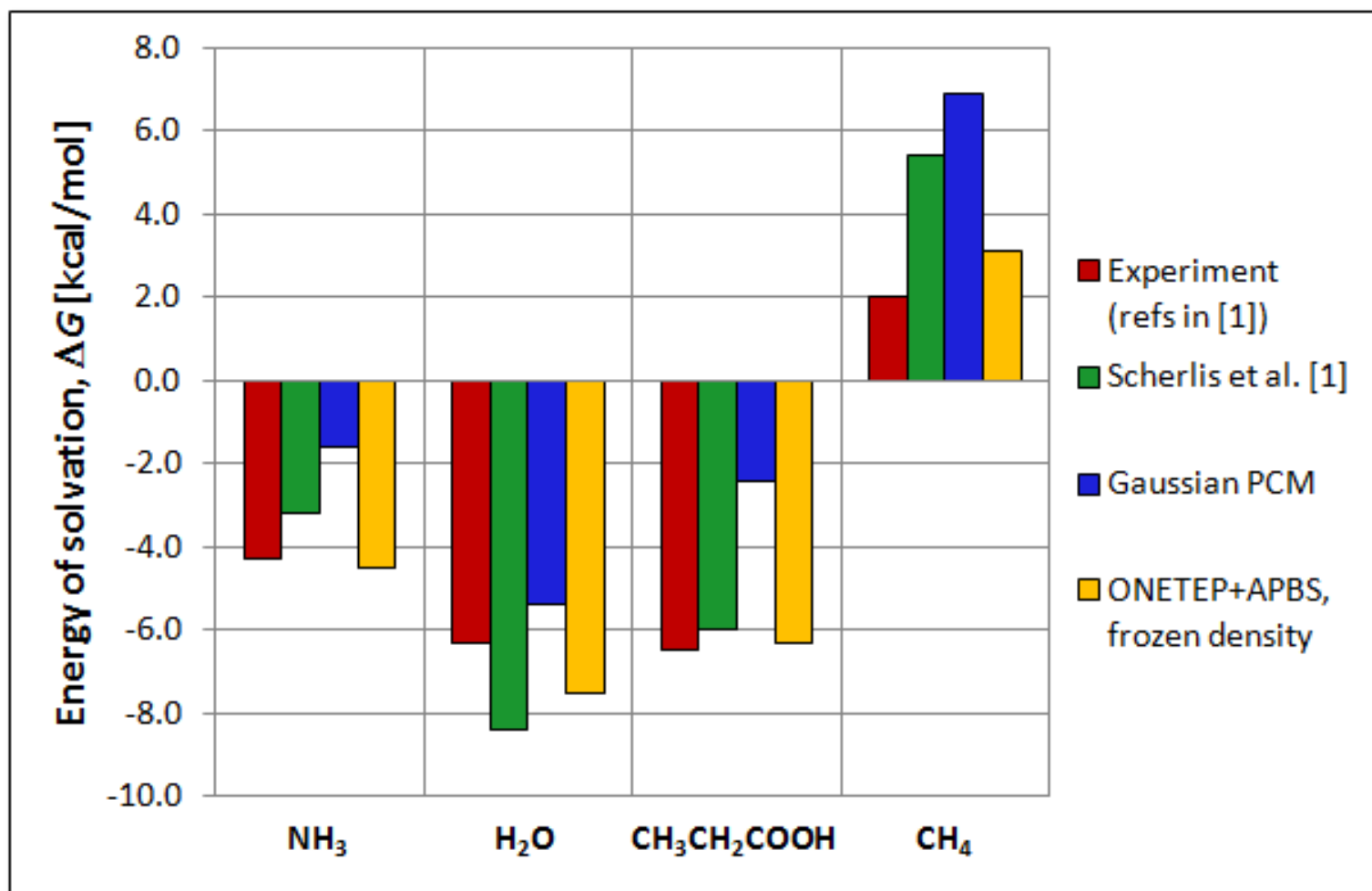
- Charge densities, potentials are represented on a grid commensurate with ONETEP's *fine grid*.
- Currently using $d = 0.136 \text{ \AA}$ (KE cutoff $\approx 780 \text{ eV}$).
- A small amount of padding ($\approx 1 \text{ \AA}$) is necessary.

Frozen density approach

Implementing IS in ONETEP



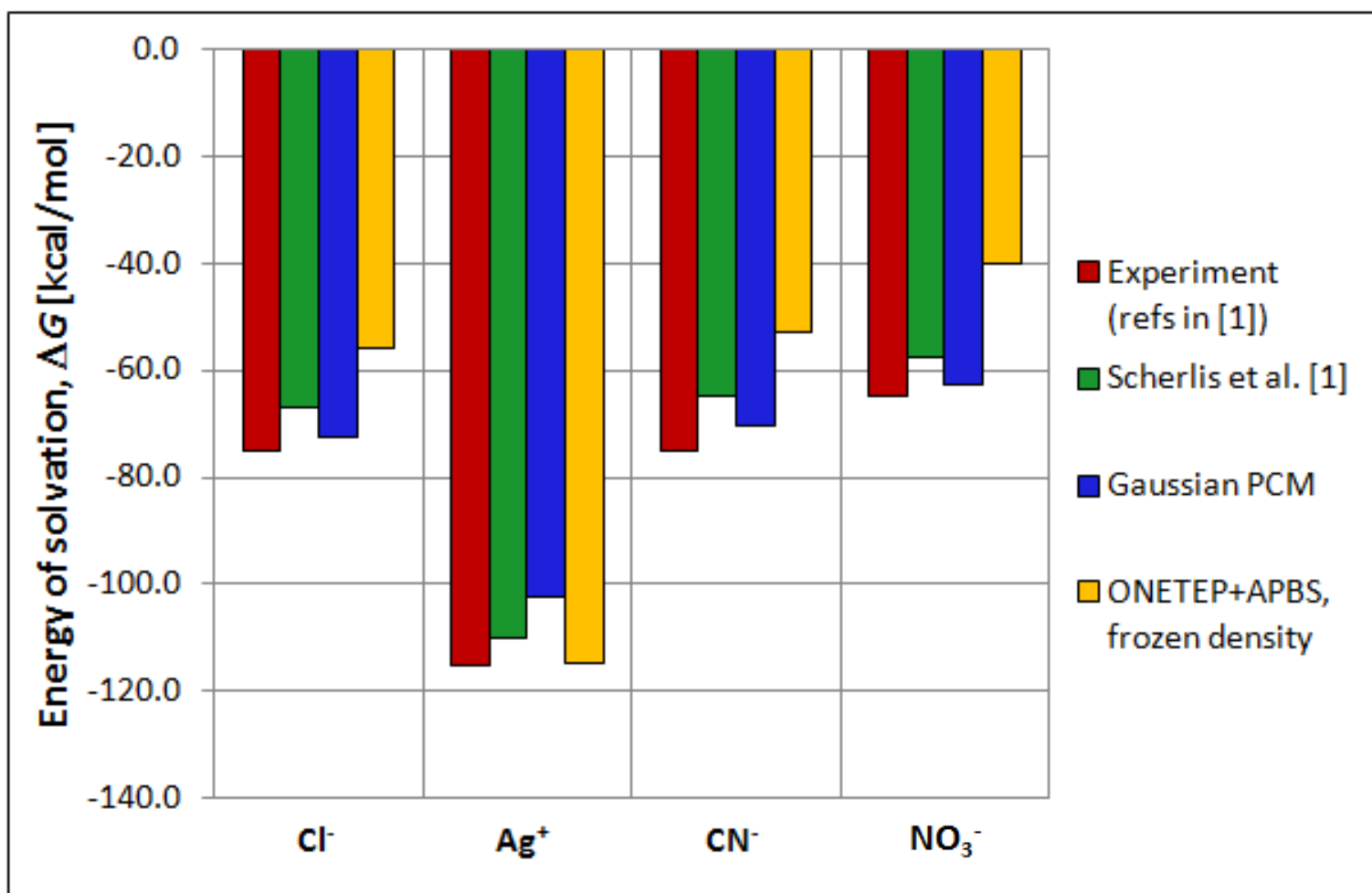
First results



Method	Mean unsigned error wrt experiment (kcal/mol)
Scherlis <i>et al.</i>	1.78
Gaussian PCM	3.15
ONETEP+APBS, frozen density	0.68

Frozen density approach

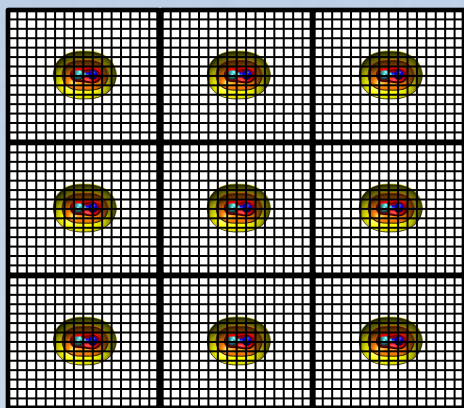
First results



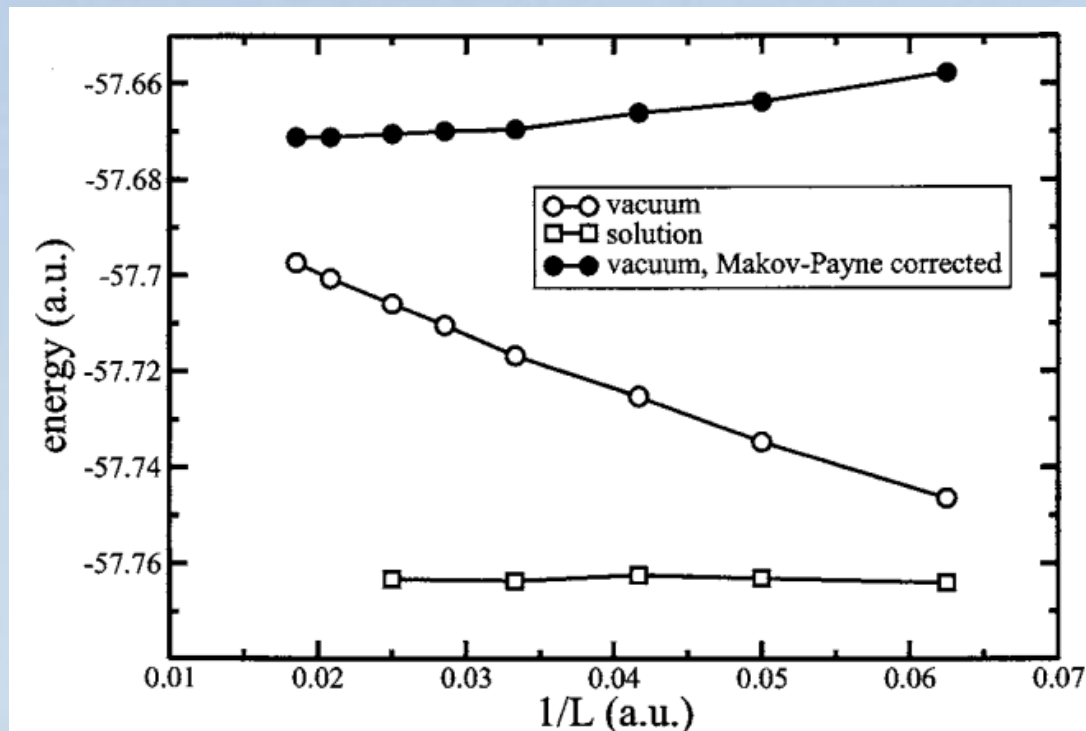
Method	Mean unsigned error wrt experiment (kcal/mol)
Scherlis <i>et al.</i>	7.63
Gaussian PCM	5.58
ONETEP+APBS, frozen density	16.55

Frozen density approach

Correcting the effect of PBCs



Periodic images of the system interact with each other.

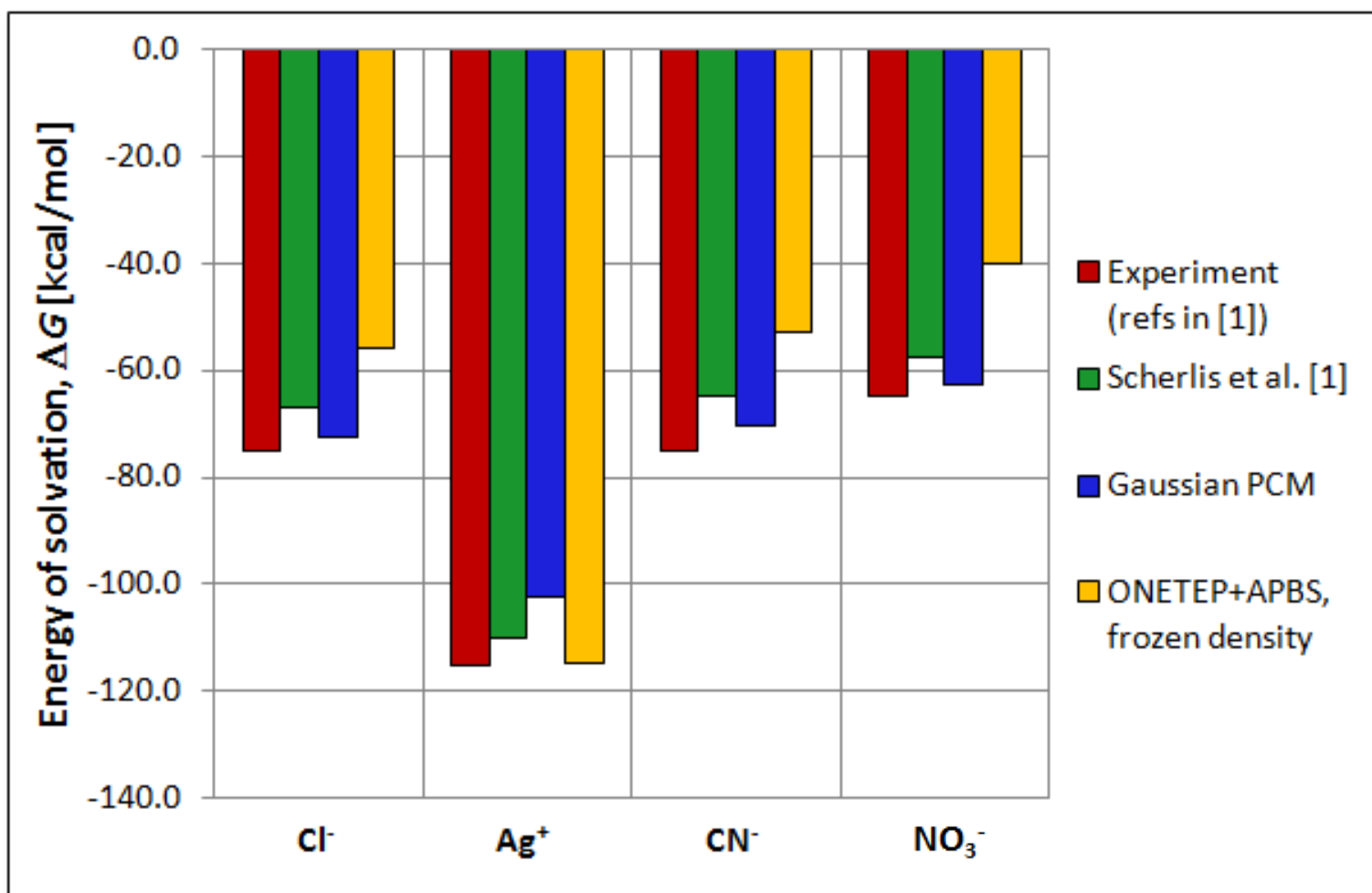


[11] Scherlis, Fattebert, Gygi, Cococcioni and Marzari, *J. Chem. Phys.* **124** (2006).

$$E_{\text{gas}} = E_{\text{PBC}} + c_1 \frac{q^2}{L} + c_2 \frac{qQ}{L^3} + O(L^{-5})$$

[12] Makov and Payne, *Phys. Rev. B* **51** (1995).

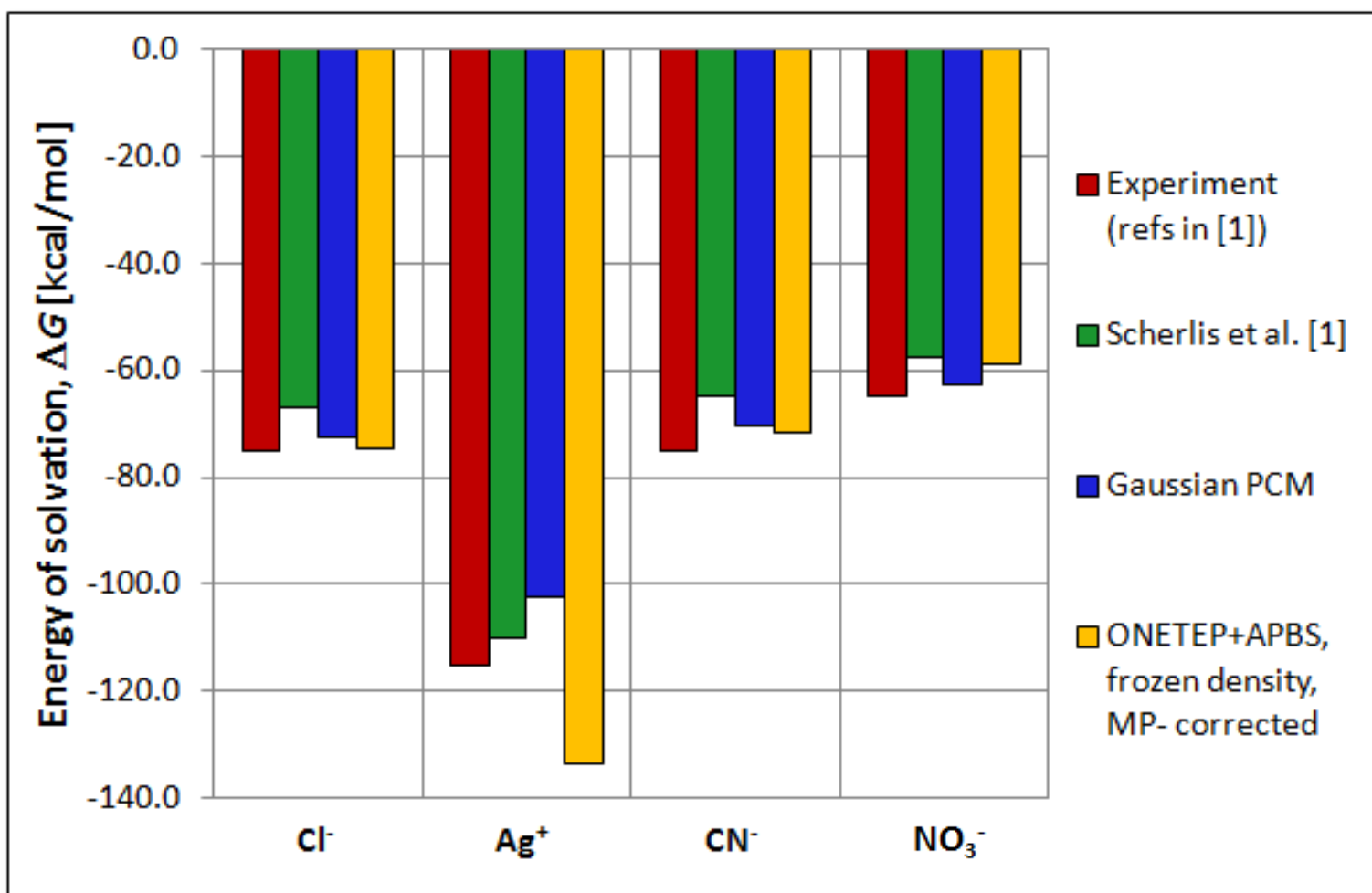
First results



Method	Mean unsigned error wrt experiment (kcal/mol)
Scherlis <i>et al.</i>	7.63
Gaussian PCM	5.58
ONETEP+APBS, frozen density	16.55

Frozen density approach

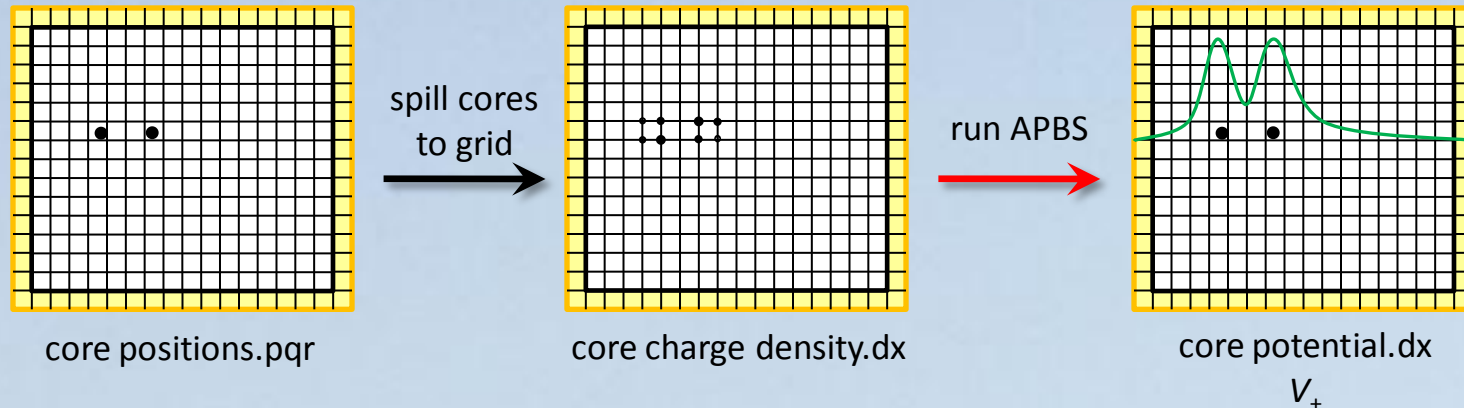
First results



Mean unsigned error wrt experiment	kcal/mol
Scherlis <i>et al.</i>	7.63
Gaussian PCM	5.58
ONETEP+APBS, frozen density, MP-corrected	7.60

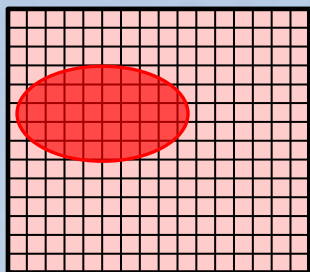
Self-consistent approach

Preparation



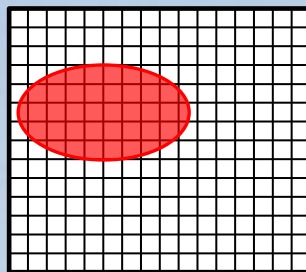
- An APBS calculation is performed in vacuum for the cores only, to obtain the **Coulombic potential due to cores**, $V_+(\mathbf{r})$.

run ONETEP
iteration

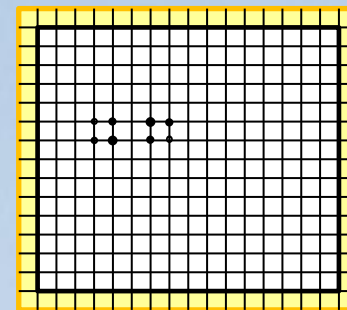


electronic charge
density (+ bkg).dx

remove
background
charge

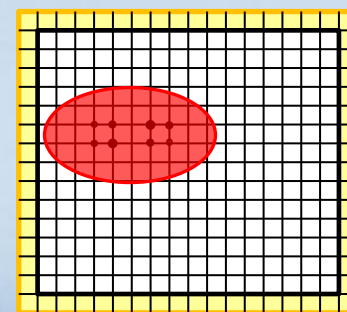


electronic charge
density.dx



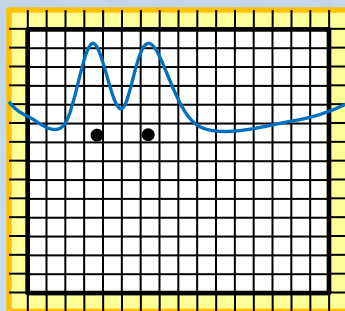
core charge density.dx

+



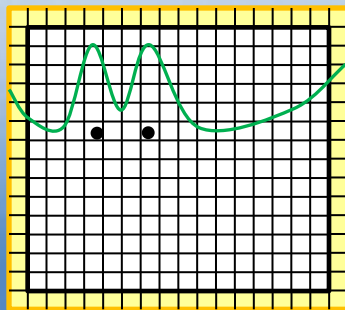
total charge density.dx

run APBS,
 $\epsilon = 80$



total potential
in solvent.dx
 $V_+ + V_- + V^{\text{die}}$

run APBS,
 $\epsilon = 1$

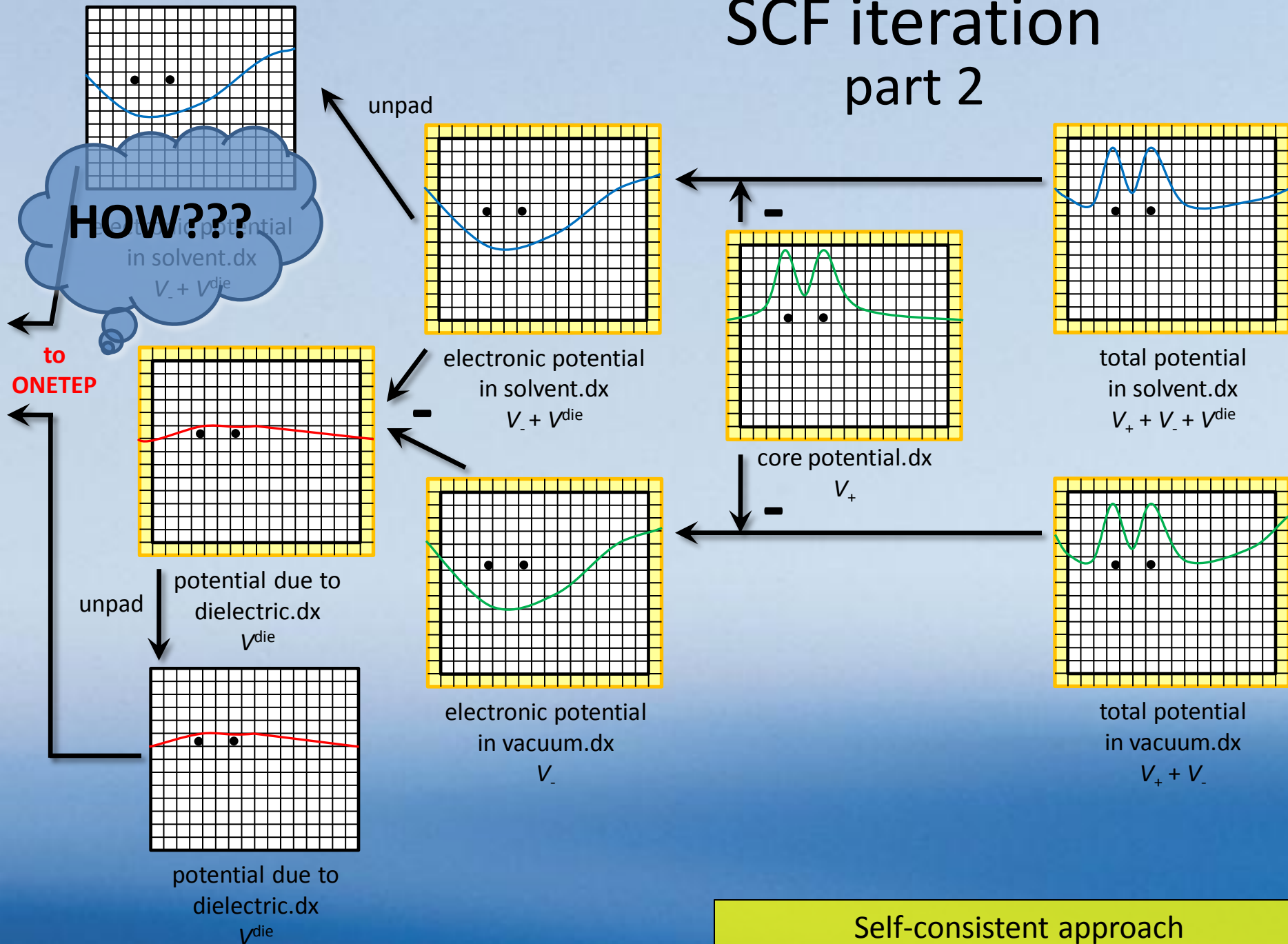


total potential
in vacuum.dx
 $V_+ + V_-$

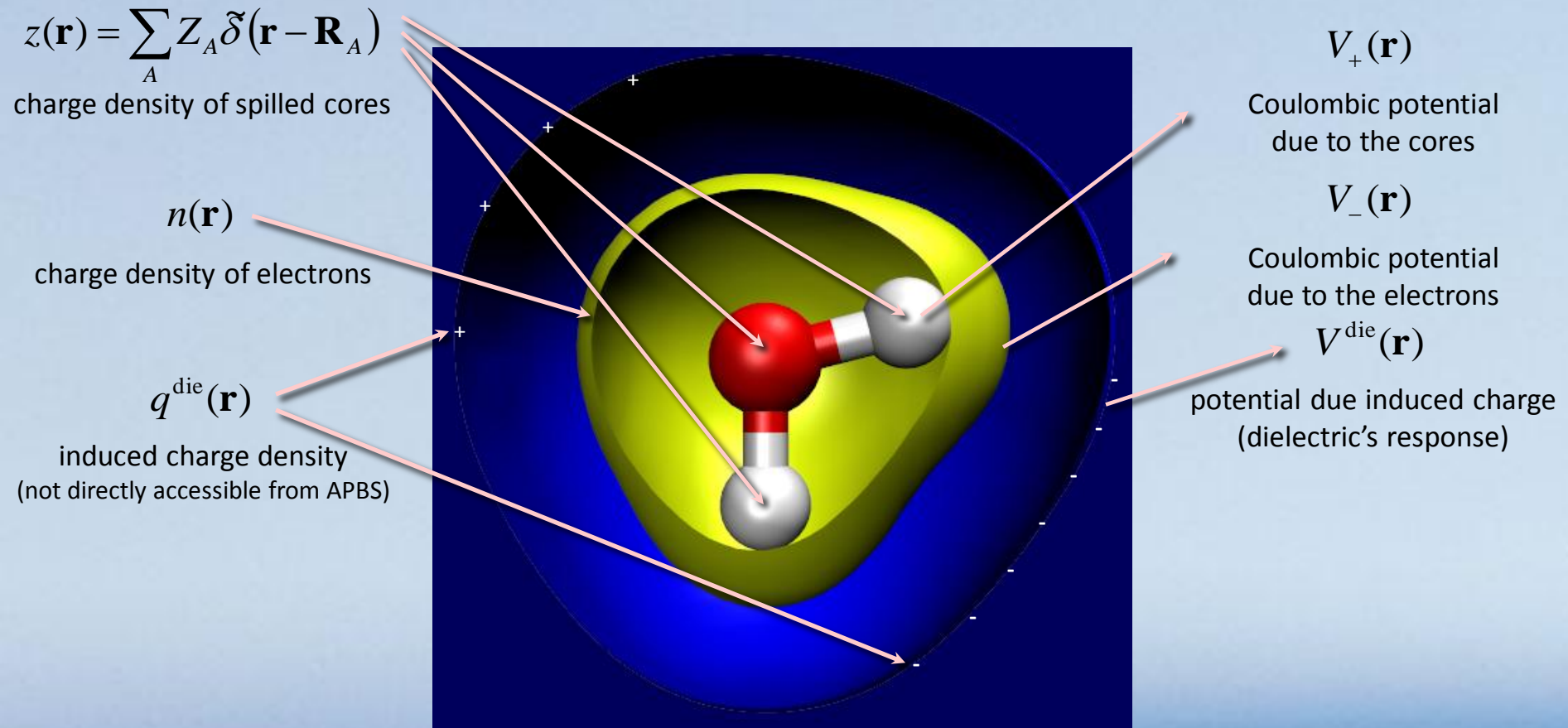
SCF iteration part 1

Self-consistent approach

SCF iteration part 2



How to inform ONETEP of the changes?



Self-consistent approach

How to inform ONETEP of the changes?

$$z(\mathbf{r}) = \sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A)$$

charge density of spilled cores

$$n(\mathbf{r})$$

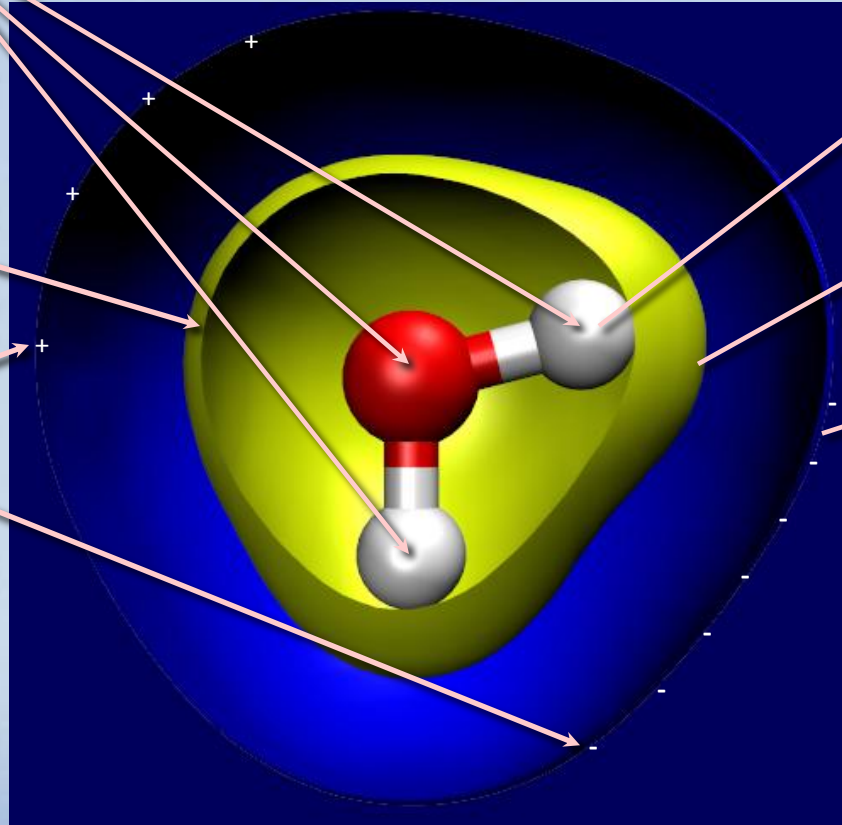
charge density of electrons

$$q^{\text{die}}(\mathbf{r})$$

induced charge density

(not directly accessible from APBS)

$$1. E_{e-e} = \frac{1}{2} \int n(\mathbf{r}) V_-(\mathbf{r}) d\mathbf{r}$$



$$V_+(\mathbf{r})$$

Coulombic potential
due to the cores

$$V_-(\mathbf{r})$$

Coulombic potential
due to the electrons

$$V^{\text{die}}(\mathbf{r})$$

potential due induced charge
(dielectric's response)

Self-consistent approach

How to inform ONETEP of the changes?

$$z(\mathbf{r}) = \sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A)$$

charge density of spilled cores

$$n(\mathbf{r})$$

charge density of electrons

$$q^{\text{die}}(\mathbf{r})$$

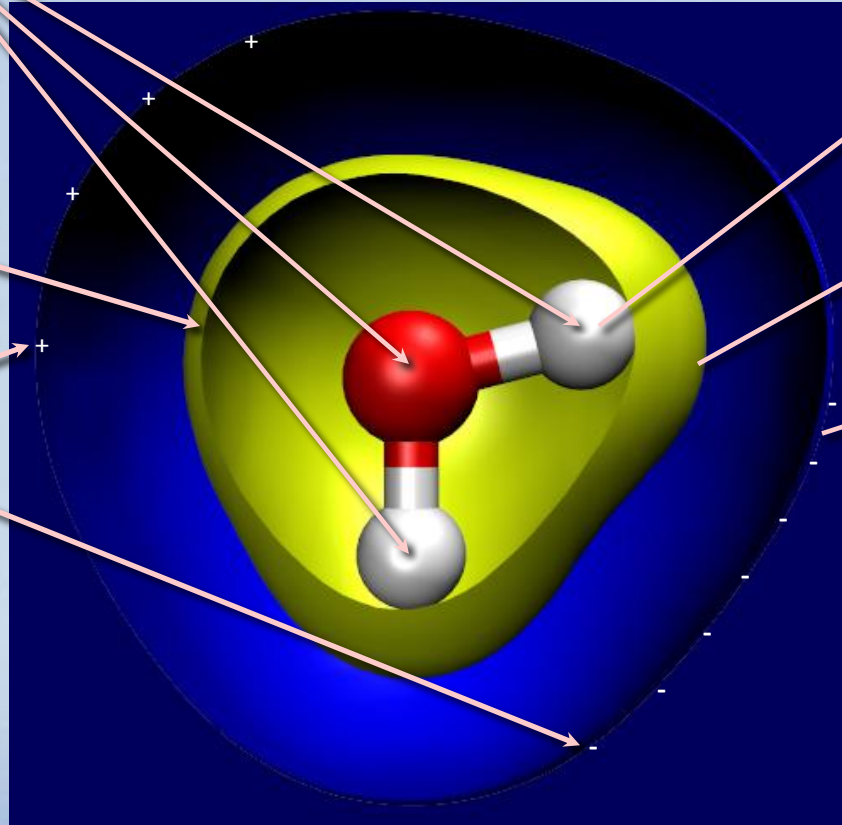
induced charge density

(not directly accessible from APBS)

$$1. E_{e-e} = \frac{1}{2} \int n(\mathbf{r}) V_-(\mathbf{r}) d\mathbf{r}$$

$$2. E_{e-c} \neq \int n(\mathbf{r}) V_+(\mathbf{r}) d\mathbf{r}$$

„usual” pseudopot. energy



$$V_+(\mathbf{r})$$

Coulombic potential
due to the cores

$$V_-(\mathbf{r})$$

Coulombic potential
due to the electrons

$$V^{\text{die}}(\mathbf{r})$$

potential due induced charge
(dielectric's response)

Self-consistent approach

How to inform ONETEP of the changes?

$$z(\mathbf{r}) = \sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A)$$

charge density of spilled cores

$$n(\mathbf{r})$$

charge density of electrons

$$q^{\text{die}}(\mathbf{r})$$

induced charge density

(not directly accessible from APBS)

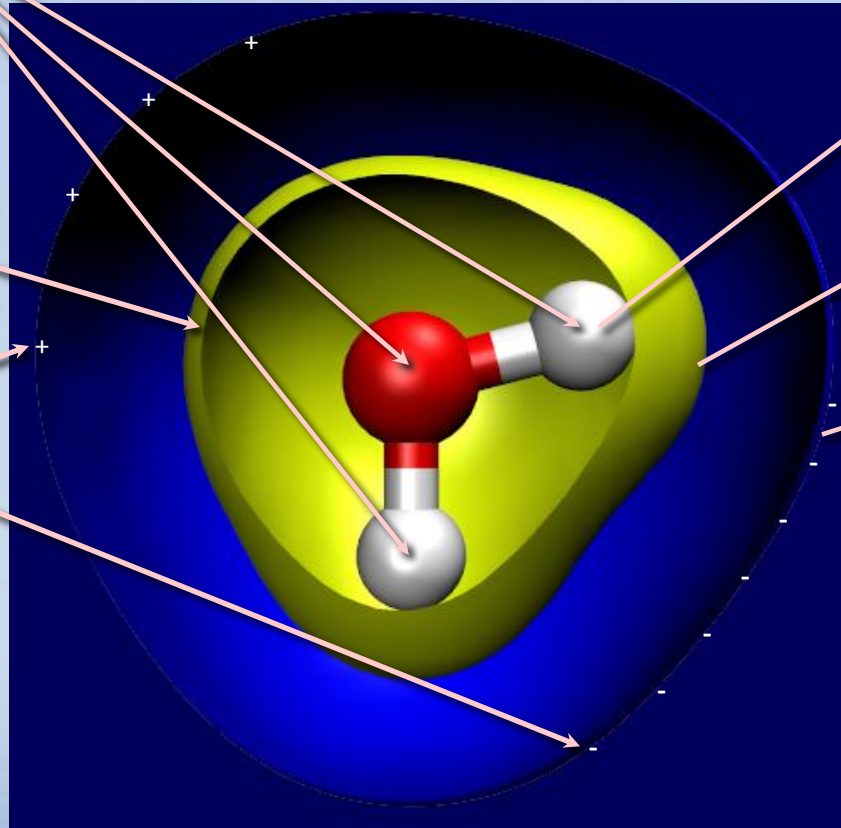
$$1. E_{e-e} = \frac{1}{2} \int n(\mathbf{r}) V_-(\mathbf{r}) d\mathbf{r}$$

$$2. E_{e-c} \neq \int n(\mathbf{r}) V_+(\mathbf{r}) d\mathbf{r}$$

„usual” pseudopot. energy

$$3. E_{e-\text{die}} = \int n(\mathbf{r}) V^{\text{die}}(\mathbf{r}) d\mathbf{r}$$

could be bundled together with 1.



$$V_+(\mathbf{r})$$

Coulombic potential
due to the cores

$$V_-(\mathbf{r})$$

Coulombic potential
due to the electrons

$$V^{\text{die}}(\mathbf{r})$$

potential due induced charge
(dielectric's response)

Self-consistent approach

How to inform ONETEP of the changes?

$$z(\mathbf{r}) = \sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A)$$

charge density of spilled cores

$$n(\mathbf{r})$$

charge density of electrons

$$q^{\text{die}}(\mathbf{r})$$

induced charge density

(not directly accessible from APBS)

$$1. E_{e-e} = \frac{1}{2} \int n(\mathbf{r}) V_-(\mathbf{r}) d\mathbf{r}$$

$$2. E_{e-c} \neq \int n(\mathbf{r}) V_+(\mathbf{r}) d\mathbf{r}$$

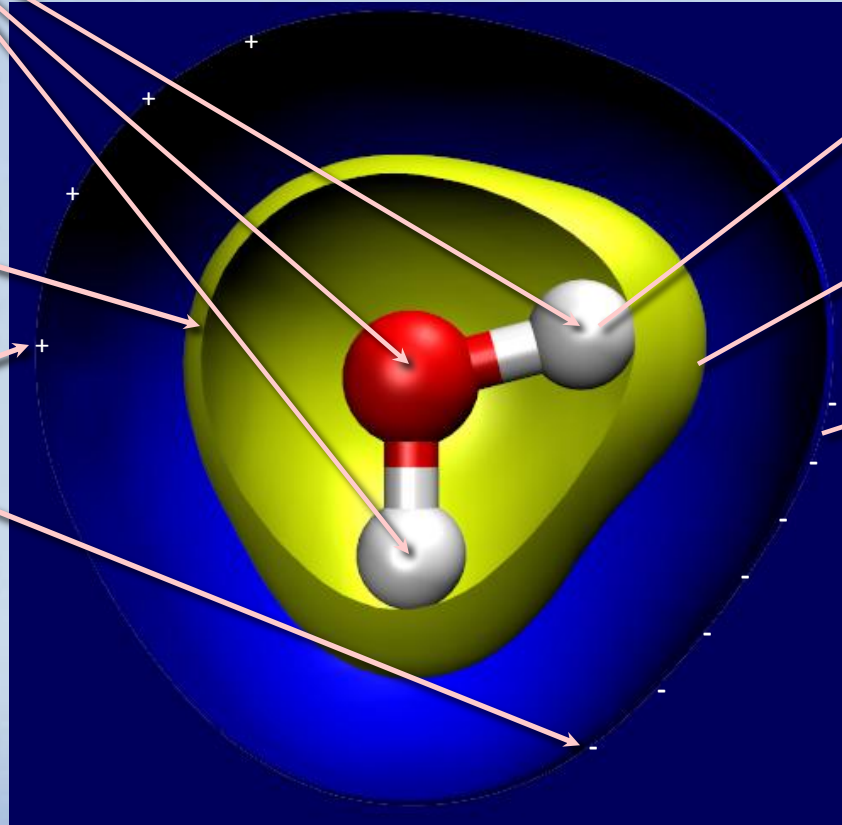
„usual” pseudopot. energy

$$3. E_{e-\text{die}} = \int n(\mathbf{r}) V^{\text{die}}(\mathbf{r}) d\mathbf{r}$$

could be bundled together with 1.

$$4. E_{c-c} = \frac{1}{2} \sum_{A,B} \frac{Z_A Z_B}{R_{AB}}$$

Assuming open BC and $\epsilon=1$
between the cores



$$V_+(\mathbf{r})$$

Coulombic potential
due to the cores

$$V_-(\mathbf{r})$$

Coulombic potential
due to the electrons

$$V^{\text{die}}(\mathbf{r})$$

potential due induced charge
(dielectric's response)

Self-consistent approach

How to inform ONETEP of the changes?

$$z(\mathbf{r}) = \sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A)$$

charge density of spilled cores

$$n(\mathbf{r})$$

charge density of electrons

$$q^{\text{die}}(\mathbf{r})$$

induced charge density

(not directly accessible from APBS)

$$1. E_{e-e} = \frac{1}{2} \int n(\mathbf{r}) V_-(\mathbf{r}) d\mathbf{r}$$

$$2. E_{e-c} \neq \int n(\mathbf{r}) V_+(\mathbf{r}) d\mathbf{r}$$

„usual” pseudopot. energy

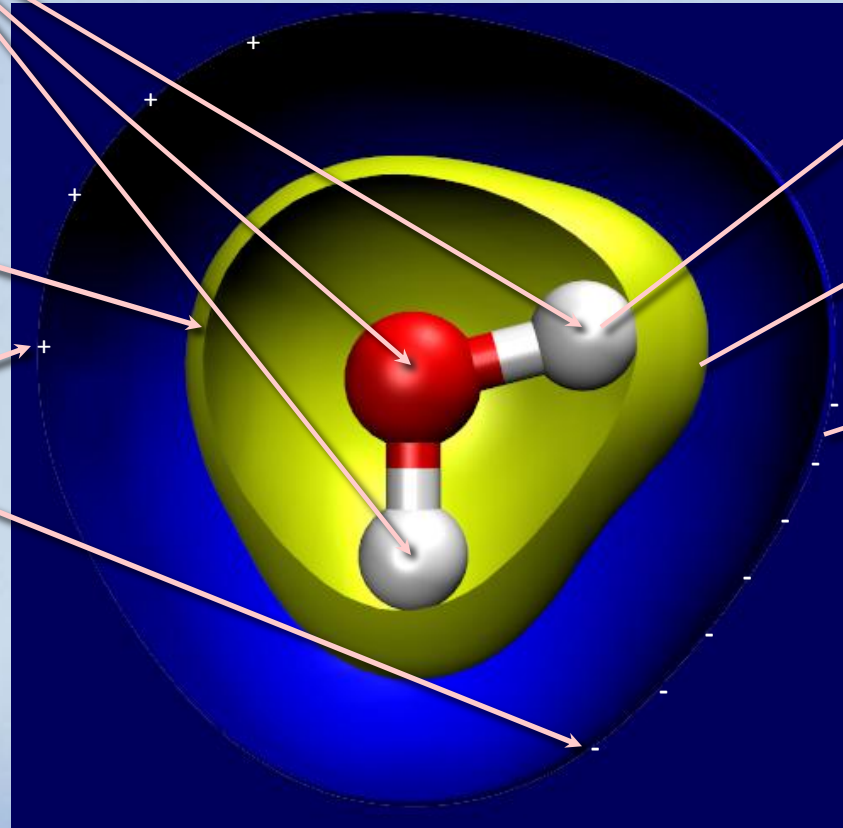
$$3. E_{e-die} = \int n(\mathbf{r}) V^{\text{die}}(\mathbf{r}) d\mathbf{r}$$

could be bundled together with 1.

$$4. E_{c-c} = \frac{1}{2} \sum_{A,B} \frac{Z_A Z_B}{R_{AB}}$$

Assuming open BC and $\epsilon=1$
between the cores

$$5. E_{c-die} = \int z(\mathbf{r}) V^{\text{die}}(\mathbf{r}) d\mathbf{r}$$



$$V_+(\mathbf{r})$$

Coulombic potential
due to the cores

$$V_-(\mathbf{r})$$

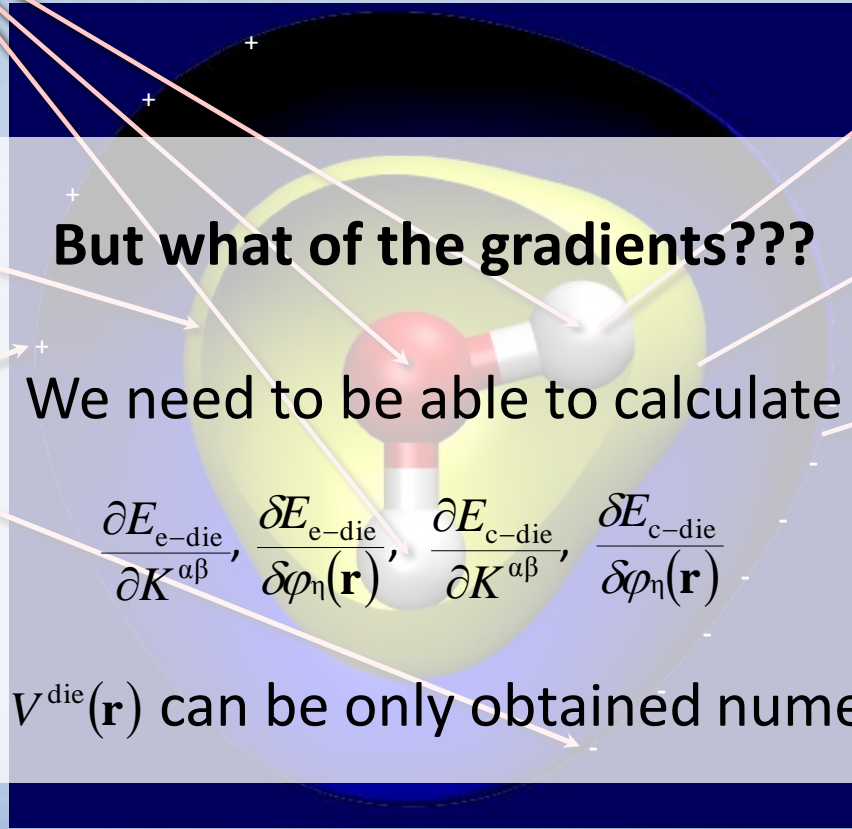
Coulombic potential
due to the electrons

$$V^{\text{die}}(\mathbf{r})$$

potential due induced charge
(dielectric's response)

Self-consistent approach

How to inform ONETEP of the changes?



But what of the gradients???

We need to be able to calculate

$$\frac{\partial E_{e\text{-die}}}{\partial K^{\alpha\beta}}, \frac{\delta E_{e\text{-die}}}{\delta \varphi_{\eta}(\mathbf{r})}, \frac{\partial E_{c\text{-die}}}{\partial K^{\alpha\beta}}, \frac{\delta E_{c\text{-die}}}{\delta \varphi_{\eta}(\mathbf{r})}$$

$V^{\text{die}}(\mathbf{r})$ can be only obtained numerically.

$$z(\mathbf{r}) = \sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A)$$

charge density of spilled cores

$$n(\mathbf{r})$$

charge density of electrons

$$q^{\text{die}}(\mathbf{r})$$

induced charge density

(not directly accessible from APBS)

$$1. E_{e-e} = \frac{1}{2} \int n(\mathbf{r}) V_-(\mathbf{r}) d\mathbf{r}$$

... but

$$2. E_{e-c} \neq \int n(\mathbf{r}) V_+(\mathbf{r}) d\mathbf{r}$$

„usual” pseudopot. energy

$$3. E_{e\text{-die}} = \int n(\mathbf{r}) V^{\text{die}}(\mathbf{r}) d\mathbf{r}$$

could be bundled together with 1.

$$4. E_{c-c} = \frac{1}{2} \sum_{A,B} \frac{Z_A Z_B}{R_{AB}}$$

Assuming open BC and $\epsilon=1$
between the cores

$$5. E_{c\text{-die}} = \int z(\mathbf{r}) V^{\text{die}}(\mathbf{r}) d\mathbf{r}$$

$$V_+(\mathbf{r})$$

Coulombic potential
due to the cores

$$V_-(\mathbf{r})$$

Coulombic potential
due to the electrons

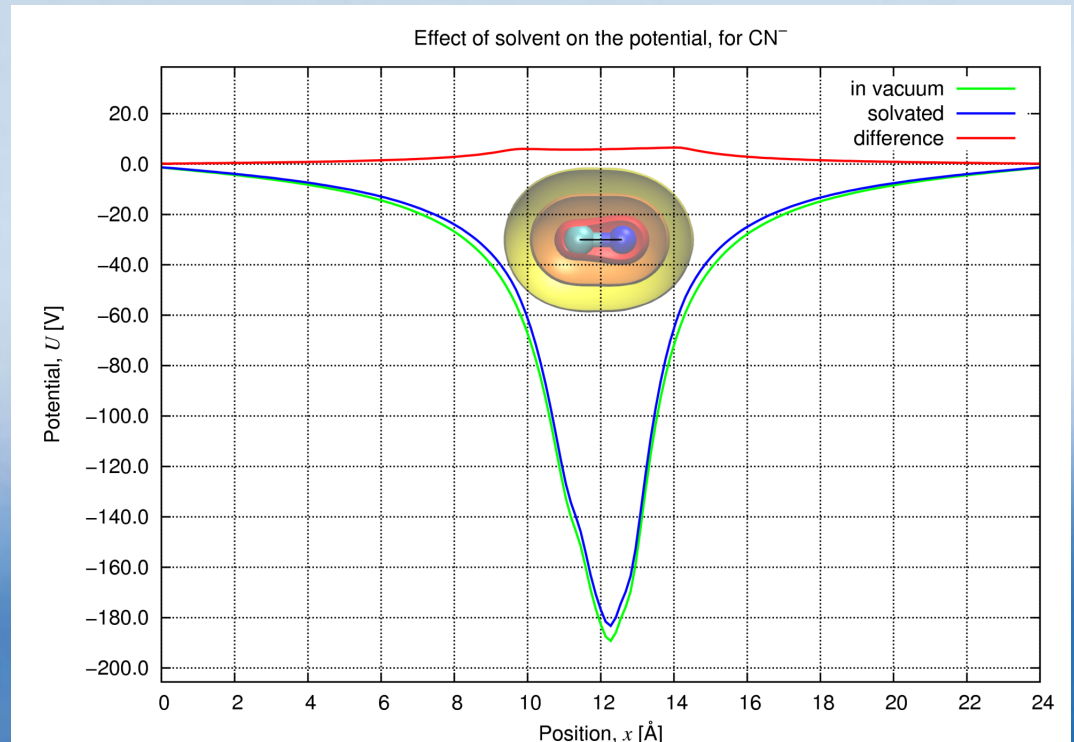
$$V^{\text{die}}(\mathbf{r})$$

potential due induced charge
(dielectric's response)

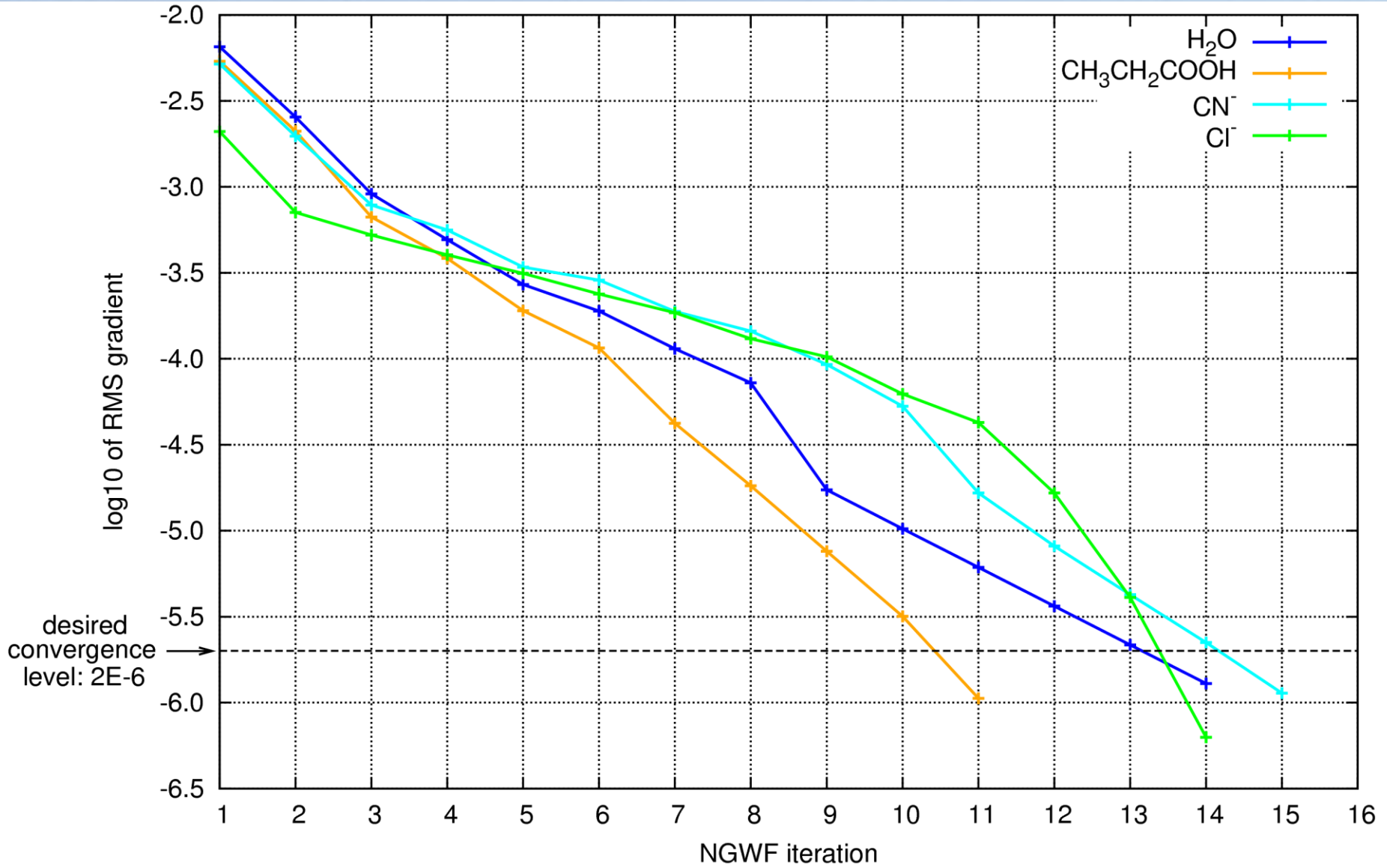
Self-consistent approach

Current state of affairs

- Ignore the gradient discrepancy until we think of something better.
- Backed up by the fact that V^{die} is small compared to V_- and V_+ .

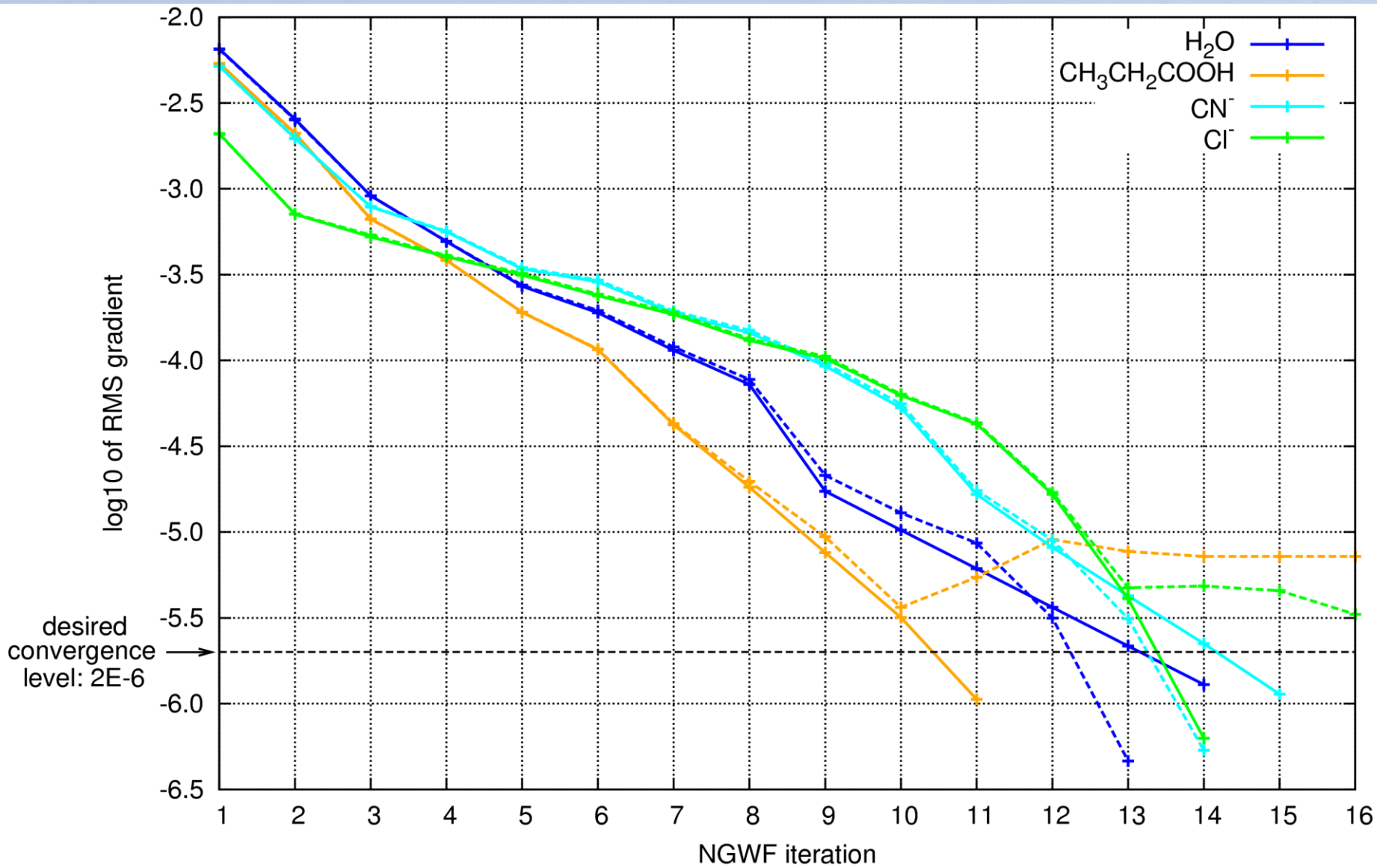


Convergence -- standard ONETEP



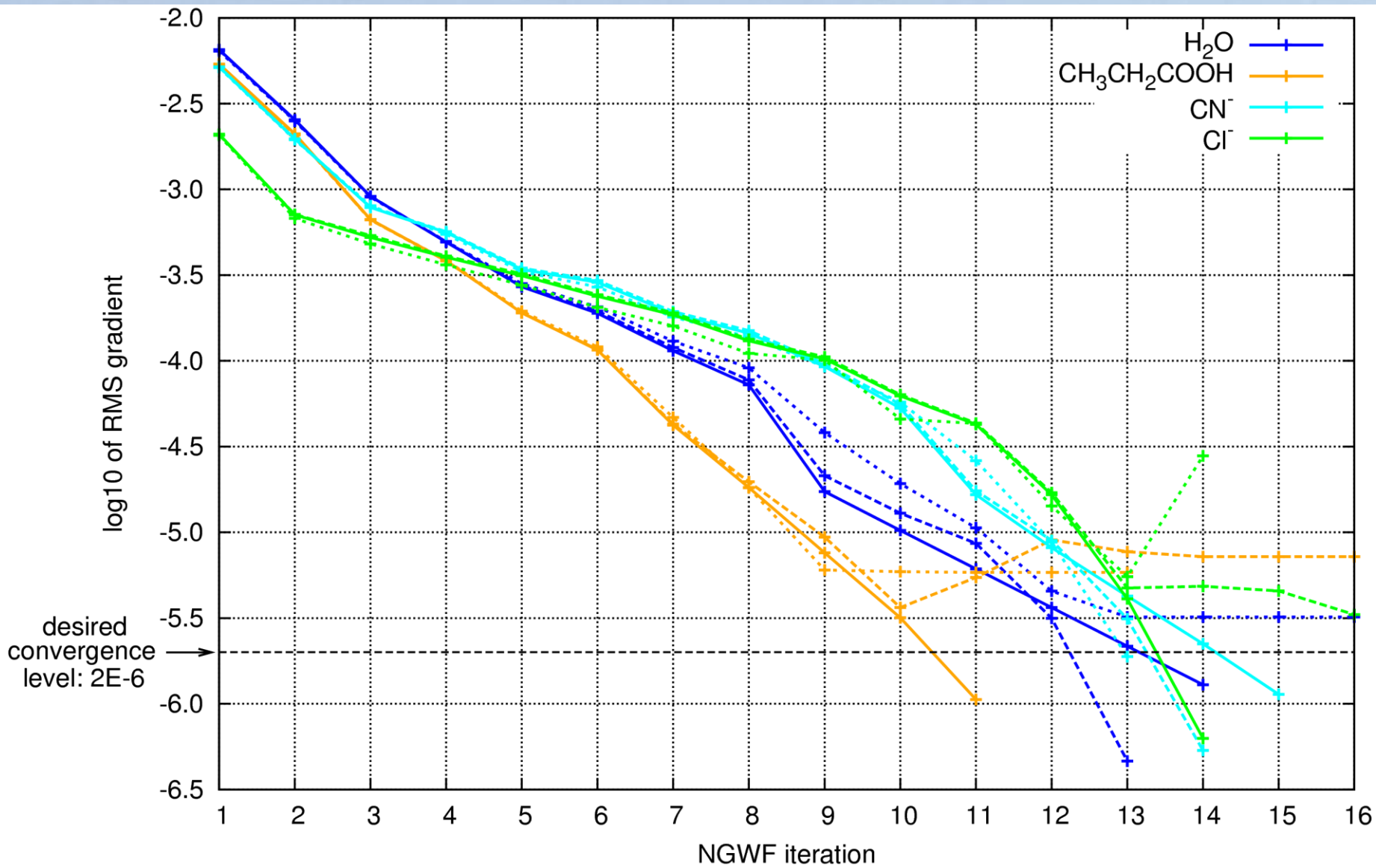
Self-consistent approach

Convergence -- standard ONETEP vs ONETEP+APBS



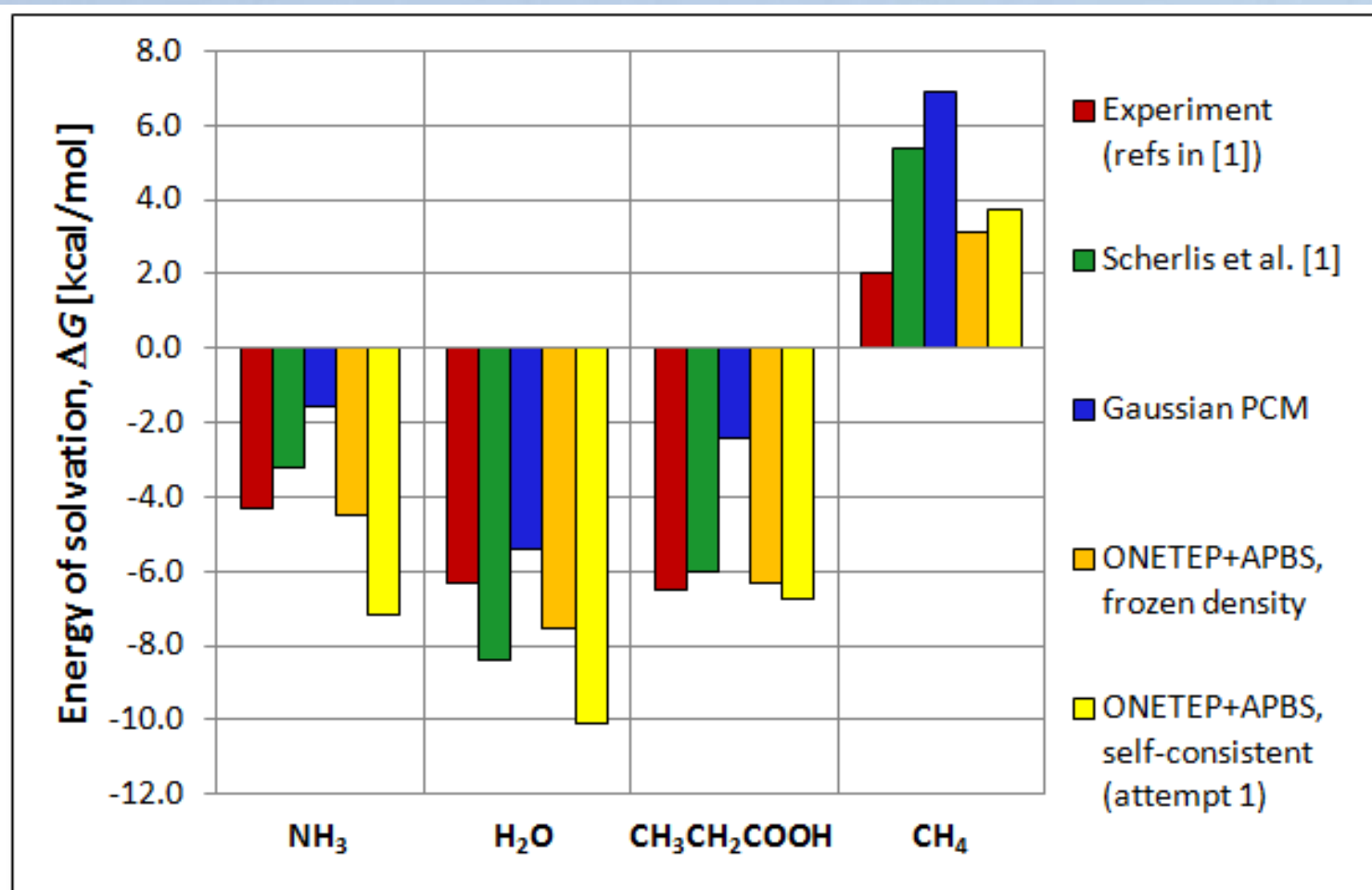
Self-consistent approach

Convergence -- standard ONETEP vs ONETEP+APBS vs ONETEP+APBS (nonuniform dielectric)



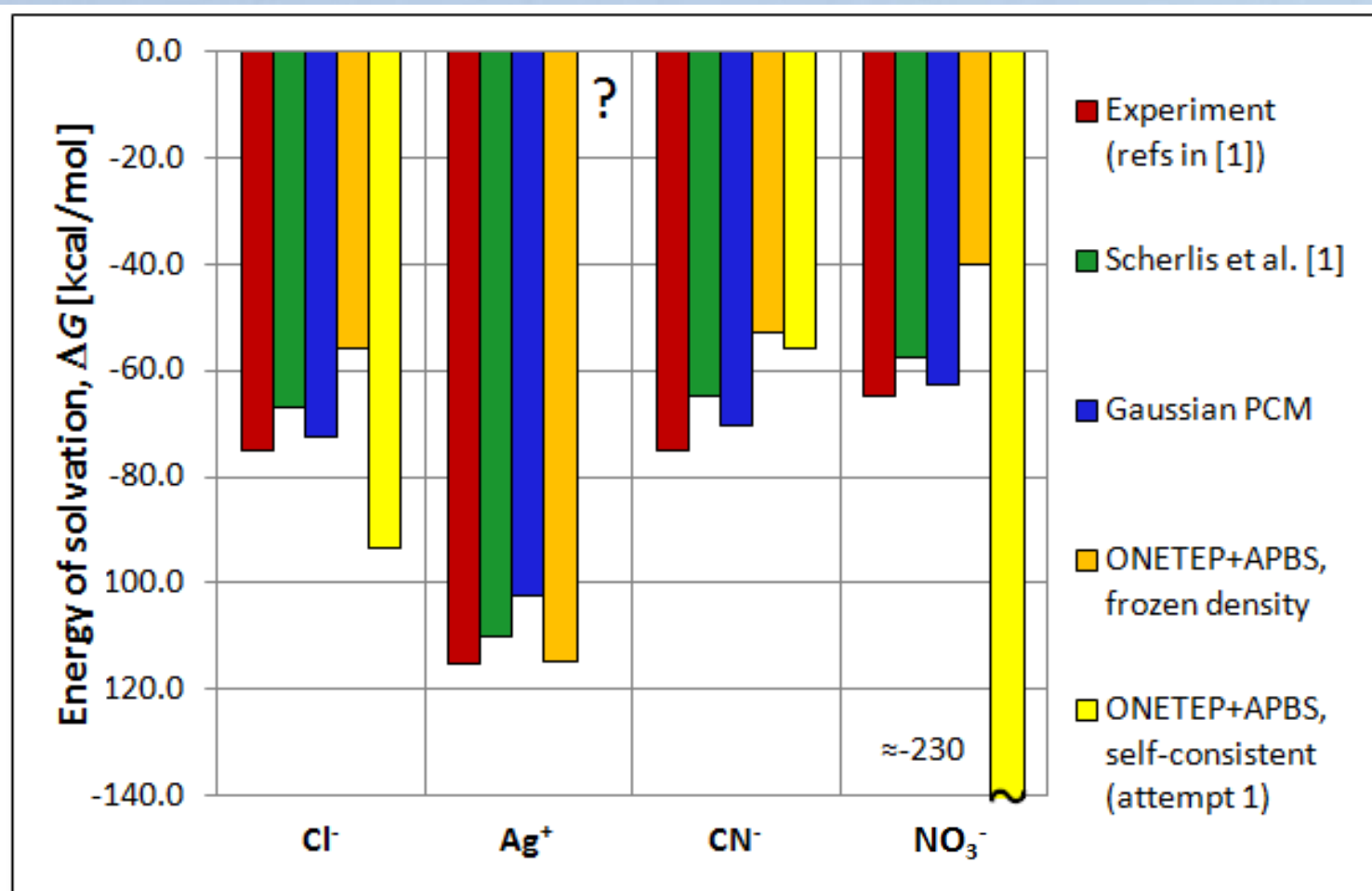
Self-consistent approach

Self-consistent approach



Method	Mean unsigned error wrt experiment	kcal/mol
Scherlis <i>et al.</i>		1.78
Gaussian PCM		3.15
ONETEP+APBS, frozen density		0.68
ONETEP+APBS, self-consistent (attempt 1)		2.16

Self-consistent approach



Method	Mean unsigned error wrt experiment	kcal/mol
Scherlis <i>et al.</i>		7.63
Gaussian PCM		5.58
ONETEP+APBS, frozen density, MP-corrected		7.60
ONETEP+APBS, self-consistent (attempt 1)		67.96

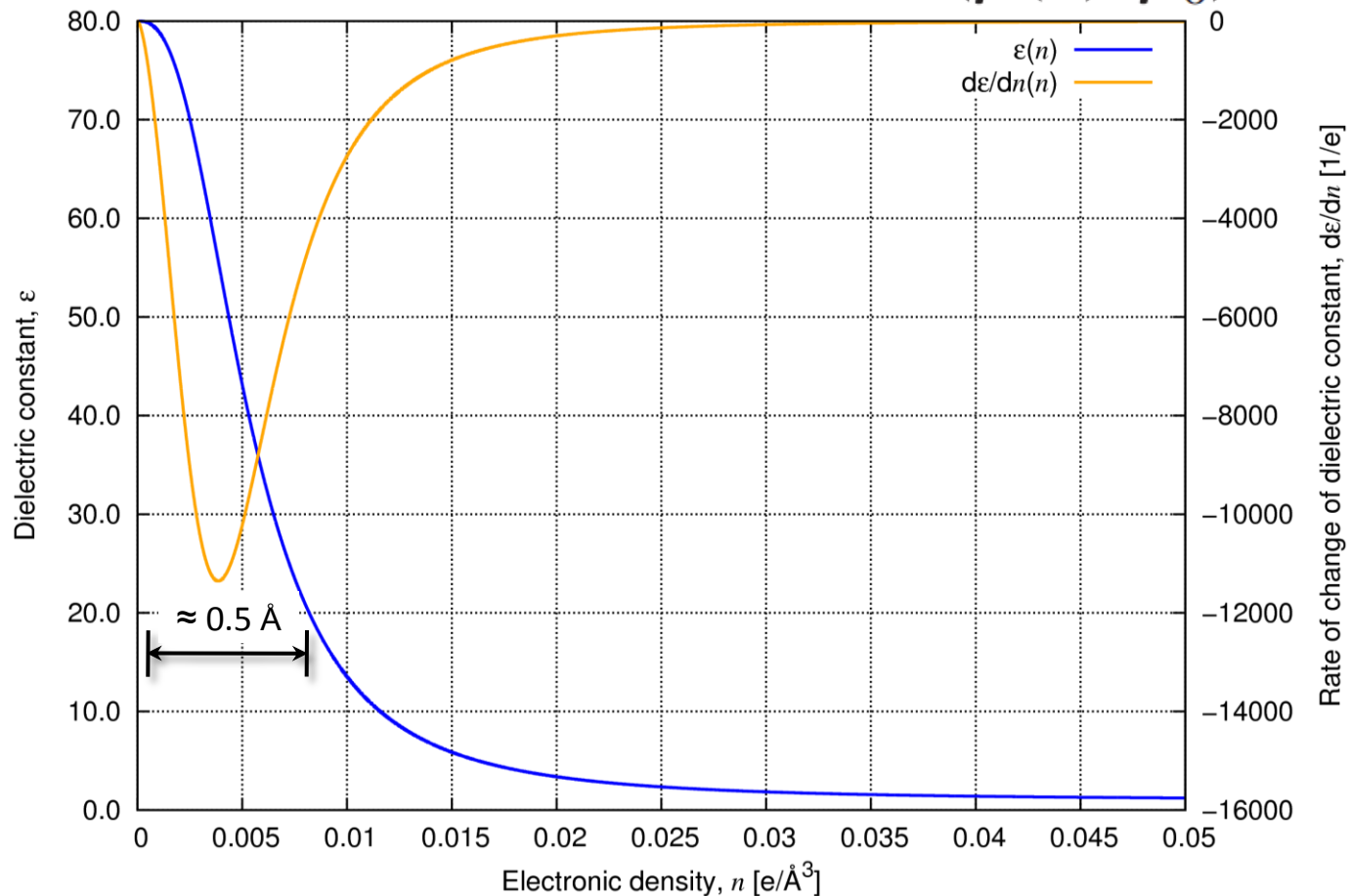
Further difficulties

- The density-dependent cavity changes shape between iterations, because the density itself changes.

$$\frac{\delta E_{\text{es}}}{\delta \rho(\mathbf{r})}(\mathbf{r}) = \phi(\mathbf{r}) + c(\nabla \phi(\mathbf{r}))^2 \frac{\delta \varepsilon(\mathbf{r})}{\delta \rho(\mathbf{r})}$$

Density-dependent cavity of Fattebert and Gygi

$$\epsilon(\rho(\mathbf{r})) = 1 + \frac{\epsilon_{\infty} - 1}{2} \left(1 + \frac{1 - (\rho(\mathbf{r})/\rho_0)^{2\beta}}{1 + (\rho(\mathbf{r})/\rho_0)^{2\beta}} \right)$$



Further difficulties

- The density-dependent cavity changes shape between iterations, because the density itself changes.

$$\frac{\delta E_{\text{es}}}{\delta \rho(\mathbf{r})}(\mathbf{r}) = \phi(\mathbf{r}) + c(\nabla \phi(\mathbf{r}))^2 \frac{\delta \varepsilon(\mathbf{r})}{\delta \rho(\mathbf{r})}$$

- For now, to define the cavity, we are using the fixed density obtained from a converged calculation in vacuum.

Further difficulties

- APBS is memory-hungry. Treating large systems (proteins of several thousand atoms) at fine grid spacings requires 40-80 GB of memory.

L [Å]	RAM req'd
10	80 MB
20	640 MB
50	10 GB
100	80 GB
200	640 GB

Assuming
fine grid
spacing of
 $d=0.136$ Å
(KE cutoff ≈ 780 eV)

- APBS lacks a "proper" parallel implementation.
- There is a kludgy parallel mode for APBS, but it's Bad and Wrong.

Conclusions

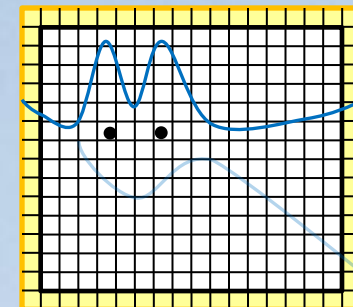
- Capturing solvation effects is important, particularly for biochemical applications.
- Implicit solvation models are computationally cheaper than explicitly modelling the solvent, but not necessarily easy to develop or to implement.
- Simple approach of correcting the *in vacuo* calculation with a postprocessing step of solving the PB equation for the frozen density appears to work surprisingly well.
- ... for the simplest molecules we've studied so far.
- Some clever maths will be required to further the advances towards a self-consistent approach.

Acknowledgments

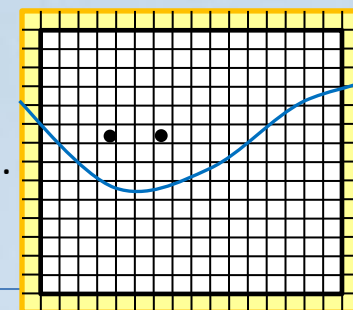
- Chris-Kriton Skylaris, head of the group.
 - Quintin Hill
 - Stephen Fox
 - Álvaro Ruiz Serrano
 - Nicholas Zonias
 - Chris Pittock
- } members of the group
- Funding from EPSRC grant "*Development of wide-ranging functionality in ONETEP*".
 - iSolutions, the guys who manage our PCs.
 - Iridis3 and HECToR, the parallel machines.

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. Baker, Sept, Joseph, Holst and McCammon, *Proc. Natl. Acad. Sci. USA* **98** (2001).
11. Scherlis, Fattebert, Gygi, Cococcioni and Marzari, *J. Chem. Phys.* **124** (2006).
12. Makov and Payne, *Phys. Rev. B* **51** (1995).
13. Fattebert and Gygi, *International Journal of Quantum Chemistry* **93** (2003).
14. Shestakov, Milovich and Noy, *Journal of Colloid and Interface Science* **247** (2002).
15. Lu, Zhang and McCammon, *J. Chem. Phys.* **122** (2005).
16. Fogolari, Brigo and Molinari, *Journal of Molecular Recognition* **15** (2002).
17. Tomasi and Persico, *Chem. Rev.* **94** (1994).
18. Swanson, Henschman and McCammon, *Biophysical Journal* **86** (2004).
19. Rocchia, Sridharan, Nicholls, Alexov, Chiaberra and Honig, *J. Comput. Chem.* **23** (2002).
20. Tan, Yang and Luo, *J. Phys. Chem. B* **110** (2006).
21. Tomasi, Mennucci and Cammi, *Chem. Rev.* **105** (2005).
22. Sanchez, Sued and Scherlis, *J. Chem. Phys.* **131** (2009).
23. Kaukonen, Soederhlejm, Heimdal and Ryde, *J. Phys. Chem. B* **112** (2008).



a Cheshire cat
without the smile



just the smile,
without the cat