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],
  – $\pi$-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

• 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

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

\[ \frac{1}{2} \int \rho(r) \phi(r) \, dr \]

in solvent and in vacuum.

only the total can be obtained experimentally

\[ \Delta G_{\text{sol}} = \Delta G_{\text{pol}} + \Delta G_{\text{npol}} \]
Obtaining $\phi(\mathbf{r})$

- **In vacuum** the Poisson equation is solved:
  \[ \varepsilon \nabla^2 \phi(\mathbf{r}) = -4\pi \rho(\mathbf{r}) \]
  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:
  \[ \nabla \cdot \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi \rho(\mathbf{r}) \]
  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.

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

Dependence of obtained $\Delta G_{\text{ele}}$ on cavity radius, for H$_2$O

- Gaussian PCM [10]
- Experiment (inferred) (refs 55-57 in [10])
- Scherlis et al. [10]
Density-dependent cavity

$\varepsilon = 80$

$\varepsilon = 1$
Density-dependent cavity of Fattebert and Gygi

\[ \varepsilon(\rho(r)) = 1 + \frac{\varepsilon_\infty - 1}{2} \left( 1 + \frac{1 - (\rho(r)/\rho_0)^{2\beta}}{1 + (\rho(r)/\rho_0)^{2\beta}} \right) \]
Density-dependent cavity

\[ \varepsilon = 5, q_{in} = 97.7\% \]
\[ \varepsilon = 40, q_{in} = 99.0\% \]
\[ \varepsilon = 70, q_{in} = 99.5\% \]
\[ \varepsilon = 79, q_{in} = 99.8\% \]
Implementing IS in ONETEP

- APBS is a multigrid solver that can calculate $\phi(r)$ given $\rho(r)$ and $\varepsilon(r)$. It also automatically calculates the polar term of the solvation energy as:

$$\Delta G_{\text{pol}} = \frac{1}{2} \int \rho(r) \phi_{\text{pol}}(r) dr - \frac{1}{2} \int \rho(r) \phi_{\varepsilon}(r) dr$$

- We are currently pursuing both the frozen density approach and the self-consistent one.
• Charge densities, potentials are represented on a grid commensurate with ONETEP’s *fine grid*.

• Currently using $d = 0.136 \, \text{Å}$ (KE cutoff $\approx 780 \, \text{eV}$).

• A small amount of padding ($\approx 1 \, \text{Å}$) is necessary.
Frozen density approach
Implementing IS in ONETEP

Frozen density approach:

1. Run usual ONETEP calculation in vacuum using `input.dat`.
2. Spill cores to grid.
3. Transfer core charge density to grid.
4. Add electronic charge density to get total charge density.
5. Run APBS with ε = 80.
6. Run APBS with ε = 1.
7. Compare, evaluate \( \Delta G_{pol} \).

Trivial conversion process.
First results

Frozen density approach

Mean unsigned error wrt experiment (kcal/mol):
- Scherlis et al.: 1.78 kcal/mol
- Gaussian PCM: 3.15 kcal/mol
- ONETEP+APBS, frozen density: 0.68 kcal/mol
First results

Frozen density approach
Correcting the effect of PBCs

Periodic images of the system interact with each other.

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


Frozen density approach
First results

Frozen density approach
First results

The graph shows the energy of solvation, $\Delta G$ [kcal/mol], for different ions: Cl$^-$, Ag$^+$, CN$^-$, and NO$_3^-$, with four different computational methods:

- **Experiment (refs in [1])**
- **Scherlis et al. [1]**
- **Gaussian PCM**
- **ONETEP+APBS, frozen density, MP-corrected**

The table below provides the mean unsigned error with respect to experiment (kcal/mol):

- **Scherlis et al.**: 7.63 kcal/mol
- **Gaussian PCM**: 5.58 kcal/mol
- **ONETEP+APBS, frozen density, MP-corrected**: 7.60 kcal/mol

Frozen density approach
Self-consistent approach
Preparation

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

Self-consistent approach
SCF iteration

part 1

run ONETEP iteration

- electronic charge density (+ bkg).

remove background charge

- electronic charge density.

- core charge density.

run APBS, \( \varepsilon = 80 \)

- total potential in solvent.

\( V_+ + V_- + V_{\text{die}} \)

run APBS, \( \varepsilon = 1 \)

- total charge density.

- total potential in vacuum.

\( V_+ + V_- \)

Self-consistent approach
Self-consistent approach

**SCF iteration part 2**

- Total potential in vacuum
  - $V_{\text{total}} = V_{\text{vacuum}} + V_{\text{die}}$

- Electronic potential in solvent
  - $V_{\text{electronic}} = V_{\text{solvent}} + V_{\text{die}}$

- Core potential
  - $V_{\text{core}}$

- Potential due to dielectric
  - $V_{\text{die}}$

- Potential due to dielectric
  - $V_{\text{die}}$

- Total potential in vacuum
  - $V_{\text{total}} = V_{\text{vacuum}} + V_{\text{die}}$

**HOW??**
How to inform ONETEP of the changes?

\[ z(r) = \sum_A Z_A \delta(r - R_A) \]
charge density of spilled cores

\[ n(r) \]
charge density of electrons

\[ q^{\text{die}}(r) \]
induced charge density (not directly accessible from APBS)

\[ V_+(r) \]
Coulombic potential due to the cores

\[ V_-(r) \]
Coulombic potential due to the electrons

\[ V^{\text{die}}(r) \]
potential due induced charge (dielectric’s response)
How to inform ONETEP of the changes?

\[ z(r) = \sum_A Z_A \delta(r - R_A) \]
charge density of spilled cores

\[ n(r) \]
charge density of electrons

\[ q^{\text{die}}(r) \]
induced charge density
(not directly accessible from APBS)

1. \[ E_{e-e} = \frac{1}{2} \int n(r)V_-(r)dr \]
How to inform ONETEP of the changes?

\[
z(r) = \sum_A Z_A \delta(r - R_A)
\]

charge density of spilled cores

\[n(r)\]

charge density of electrons

\[q^{\text{die}}(r)\]

induced charge density (not directly accessible from APBS)

1. \[E_{e-e} = \frac{1}{2} \int n(r) V_-(r) \, dr\]

2. \[E_{e-e} \neq \int n(r) V_+(r) \, dr\]

„usual“ pseudopot. energy

\[V_+(r)\]

Coulombic potential due to the cores

\[V_-(r)\]

Coulombic potential due to the electrons

\[V^{\text{die}}(r)\]

potential due induced charge (dielectric’s response)

Self-consistent approach
How to inform ONETEP of the changes?

\[ z(r) = \sum_A Z_A \delta(r - R_A) \]

charge density of spilled cores

\[ n(r) \]

charge density of electrons

\[ q^{\text{die}}(r) \]

induced charge density (not directly accessible from APBS)

1. \[ E_{e-e} = \frac{1}{2} \int n(r)V_-(r)\,dr \]

2. \[ E_{e-e} \neq \int n(r)V_+(r)\,dr \]

"usual" pseudopot. energy

3. \[ E_{e-die} = \int n(r)V^{\text{die}}(r)\,dr \]

could be bundled together with 1.

\[ V_+(r) \]

Coulombic potential due to the cores

\[ V_-(r) \]

Coulombic potential due to the electrons

\[ V^{\text{die}}(r) \]

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

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 \( \varepsilon = 1 \) 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)
How to inform ONETEP of the changes?

\[ z(r) = \sum_A Z_A \delta(r - R_A) \]
charge density of spilled cores

\[ n(r) \]
charge density of electrons

\[ q^{\text{die}}(r) \]
induced charge density
(not directly accessible from APBS)

1. \[ E_{c-c} = \frac{1}{2} \int n(r)V_-(r)d\mathbf{r} \]
2. \[ E_{c-c} \neq \int n(r)V_+(r)d\mathbf{r} \]
   „usual“ pseudopot. energy

3. \[ E_{c-\text{die}} = \int n(r)V^{\text{die}}(r)d\mathbf{r} \]
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. \[ E_{c-\text{die}} = \int z(r)V^{\text{die}}(r)d\mathbf{r} \]

\[ V_+(\mathbf{r}) \]
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)
How to inform ONETEP of the changes?

\[ z(r) = \sum_A Z_A \delta(r - R_A) \]

charge density of spilled cores

\[ n(r) \]

charge density of electrons

\[ q_{\text{die}}^\text{c}(r) \]

induced charge density (not directly accessible from APBS)

1. \( E_{\text{e-e}} = \frac{1}{2} \int n(r)V_-(r)dr \)

... but \( V_{\text{die}}(r) \) can be only obtained numerically.

2. \( E_{\text{e-c}} \neq \int n(r)V_+(r)dr \)

“usual” pseudopot. energy

3. \( E_{\text{e-die}} = \int n(r)V_{\text{die}}(r)dr \)

could be bundled together with 1.

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

Assuming open BC and \( \varepsilon=1 \) between the cores

5. \( E_{\text{c-die}} = \int z(r)V_{\text{die}}(r)dr \)

\[ V_+(r) \]

coulombic potential due to the cores

\[ V_-(r) \]

coulombic potential due to the electrons

\[ V_{\text{die}}(r) \]

potential due induced charge (dielectric’s response)

But what of the gradients???

We need to be able to calculate

\[ \frac{\partial E_{\text{e-die}}}{\partial \varphi} \quad \frac{\delta E_{\text{e-die}}}{\delta \varphi(r)} \quad \frac{\partial E_{\text{c-die}}}{\partial \varphi} \quad \frac{\delta E_{\text{c-die}}}{\delta \varphi(r)} \]
Current state of affairs

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

![Effect of solvent on the potential, for CN⁻](image.png)
Convergence -- standard ONETEP

Self-consistent approach
Convergence -- standard ONETEP vs ONETEP+APBS

-2.0
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0
-5.5
-6.0
-6.5

Self-consistent approach
Convergence -- standard ONETEP vs ONETEP+APBS vs ONETEP+APBS (nonuniform dielectric)

Self-consistent approach
Self-consistent approach
Self-consistent approach
Further difficulties

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

$$\frac{\delta E_{es}}{\delta \rho(r)}(r) = \phi(r) + c(\nabla \phi(r))^2 \frac{\delta \epsilon(r)}{\delta \rho(r)}$$
Density-dependent cavity of Fattebert and Gygi

\[ \varepsilon(\rho(r)) = 1 + \frac{\varepsilon_\infty - 1}{2} \left( 1 + \frac{1 - \left(\frac{\rho(r)}{\rho_0}\right)^{2\beta}}{1 + \left(\frac{\rho(r)}{\rho_0}\right)^{2\beta}} \right) \]

\[ \approx 0.5 \text{ Å} \]
Further difficulties

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

\[
\frac{\delta E_{es}}{\delta \rho(r)}(r) = \phi(r) + c(\nabla \phi(r))^2 \frac{\delta \epsilon(r)}{\delta \rho(r)}
\]

- 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.

<table>
<thead>
<tr>
<th>$L$ [Å]</th>
<th>RAM req'd</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>80 MB</td>
</tr>
<tr>
<td>20</td>
<td>640 MB</td>
</tr>
<tr>
<td>50</td>
<td>10 GB</td>
</tr>
<tr>
<td>100</td>
<td>80 GB</td>
</tr>
<tr>
<td>200</td>
<td>640 GB</td>
</tr>
</tbody>
</table>

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

- 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