



A LETTERS JOURNAL EXPLORING
THE FRONTIERS OF PHYSICS

OFFPRINT

**Minimal parameter implicit solvent model for
ab initio electronic-structure calculations**

J. DZIEDZIC, H. H. HELAL, C.-K. SKYLARIS, A. A. MOSTOFI and
M. C. PAYNE

EPL, **95** (2011) 43001

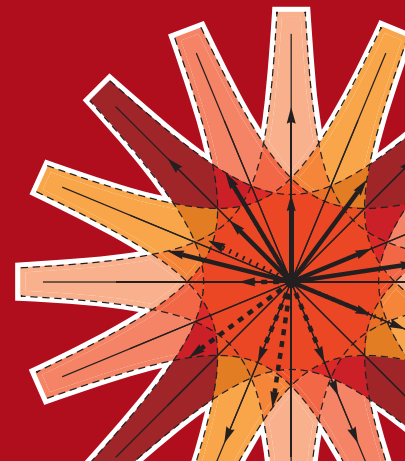
Please visit the new website
www.epljournal.org



A LETTERS JOURNAL
EXPLORING THE FRONTIERS
OF PHYSICS

The Editorial Board invites you
to submit your letters to EPL

www.epljournal.org



Six good reasons to publish with EPL

We want to work with you to help gain recognition for your high-quality work through worldwide visibility and high citations. As an EPL author, you will benefit from:

- 1 Quality** – The 40+ Co-Editors, who are experts in their fields, oversee the entire peer-review process, from selection of the referees to making all final acceptance decisions
- 2 Impact Factor** – The 2009 Impact Factor increased by 31% to 2.893; your work will be in the right place to be cited by your peers
- 3 Speed of processing** – We aim to provide you with a quick and efficient service; the median time from acceptance to online publication is 30 days
- 4 High visibility** – All articles are free to read for 30 days from online publication date
- 5 International reach** – Over 2,000 institutions have access to EPL, enabling your work to be read by your peers in 100 countries
- 6 Open Access** – Experimental and theoretical high-energy particle physics articles are currently open access at no charge to the author. All other articles are offered open access for a one-off author payment (€1,000)

Details on preparing, submitting and tracking the progress of your manuscript from submission to acceptance are available on the EPL submission website www.epletters.net

If you would like further information about our author service or EPL in general, please visit www.epljournal.org or e-mail us at info@epljournal.org



IOP Publishing

Image: Ornamental multiplication of space-time figures of temperature transformation rules
(adapted from T. S. Biró and P. Ván 2010 *EPL* **89** 30001; artistic impression by Frédérique Swist).

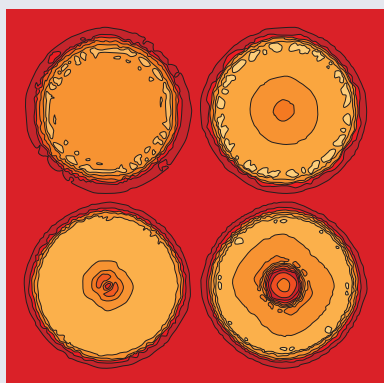
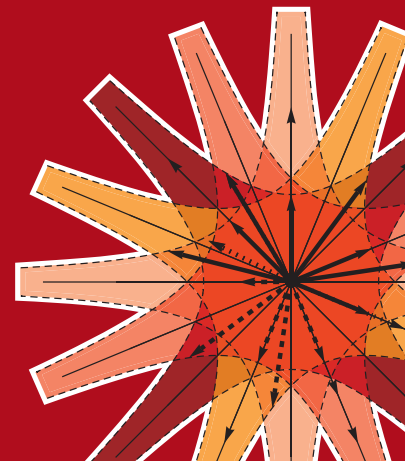


epl

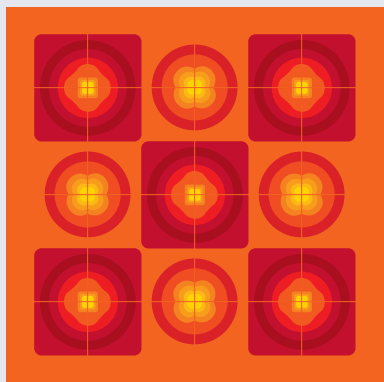
A LETTERS JOURNAL
EXPLORING THE FRONTIERS
OF PHYSICS

EPL Compilation Index

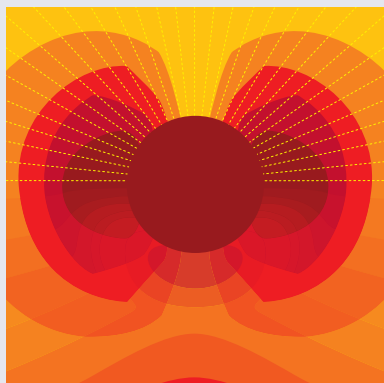
www.epljournal.org



Biaxial strain on lens-shaped quantum rings of different inner radii, adapted from **Zhang et al** 2008 *EPL* **83** 67004.



Artistic impression of electrostatic particle-particle interactions in dielectrophoresis, adapted from **N Aubry and P Singh** 2006 *EPL* **74** 623.



Artistic impression of velocity and normal stress profiles around a sphere that moves through a polymer solution, adapted from **R Tuinier, J K G Dhont and T-H Fan** 2006 *EPL* **75** 929.

Visit the EPL website to read the latest articles published in cutting-edge fields of research from across the whole of physics.

Each compilation is led by its own Co-Editor, who is a leading scientist in that field, and who is responsible for overseeing the review process, selecting referees and making publication decisions for every manuscript.

- Graphene
- Liquid Crystals
- High Transition Temperature Superconductors
- Quantum Information Processing & Communication
- Biological & Soft Matter Physics
- Atomic, Molecular & Optical Physics
- Bose-Einstein Condensates & Ultracold Gases
- Metamaterials, Nanostructures & Magnetic Materials
- Mathematical Methods
- Physics of Gases, Plasmas & Electric Fields
- High Energy Nuclear Physics

If you are working on research in any of these areas, the Co-Editors would be delighted to receive your submission. Articles should be submitted via the automated manuscript system at **www.epletters.net**

If you would like further information about our author service or EPL in general, please visit **www.epljournal.org** or e-mail us at **info@epljournal.org**



IOP Publishing

Image: Ornamental multiplication of space-time figures of temperature transformation rules (adapted from T. S. Bíró and P. Ván 2010 *EPL* **89** 30001; artistic impression by Frédérique Swist).

Minimal parameter implicit solvent model for *ab initio* electronic-structure calculations

J. DZIEDZIC^{1(a)}, H. H. HELAL², C.-K. SKYLARIS¹, A. A. MOSTOFI³ and M. C. PAYNE²

¹ School of Chemistry, University of Southampton - Highfield, Southampton SO17 1BJ, UK, EU

² Theory of Condensed Matter group, Cavendish Laboratory, University of Cambridge
Cambridge CB3 0HE, UK, EU

³ The Thomas Young Centre for Theory and Simulation of Materials, Imperial College London
London SW7 2AZ, UK, EU

received 28 February 2011; accepted in final form 27 June 2011

published online 27 July 2011

PACS 31.70.Dk – Environmental and solvent effects

PACS 87.15.kr – Protein-solvent interactions

PACS 31.15.A– – *Ab initio* calculations

Abstract – We present an implicit solvent model for *ab initio* electronic-structure calculations which is fully self-consistent and is based on direct solution of the nonhomogeneous Poisson equation. The solute cavity is naturally defined in terms of an isosurface of the electronic density according to the formula of Fattebert and Gygi (*J. Comput. Chem.*, **23** (2002) 662). While this model depends on only two parameters, we demonstrate that by using appropriate boundary conditions and dispersion-repulsion contributions, solvation energies obtained for an extensive test set including neutral and charged molecules show dramatic improvement compared to existing models. Our approach is implemented in, but not restricted to, a linear-scaling density functional theory (DFT) framework, opening the path for self-consistent implicit solvent DFT calculations on systems of unprecedented size, which we demonstrate with calculations on a 2615-atom protein-ligand complex.

Copyright © EPLA, 2011

The role of solvent is critical to a multitude of chemical, biological and physical processes. The accurate simulation of such processes, therefore, requires careful treatment of solvation effects. However, explicit inclusion of the solvent with full atomic detail is very costly due to the significant increase in the number of simulated atoms and the need for extensive averaging over the solvent degrees of freedom [1]. Moreover, such explicit treatment may also be unnecessary, as it is often the long-range electrostatic effect of the solvent that is most significant, with only a small proportion of solvent molecules involved chemically. The implicit solvent approach addresses these issues by retaining only the atomic details of the solute, placed in a suitably defined cavity, and by representing the solvent environment by an unstructured dielectric continuum outside this cavity. The free energy of solvation is typically decomposed into two contributions —the

electrostatic energy of interaction of the solvent with the polarized dielectric, and a nonpolar term accounting for the work required to create a cavity in the solvent (cavitation energy), and, in more complex models, for dispersion-repulsion interactions between the solute and solvent.

A multitude of implicit solvent models of differing sophistication have been proposed to date [2] for use in *ab initio* calculations. Many of these models are derived from the self-consistent reaction field (SCRF) formalism, where the effect of the electric field due to the dielectric (polarized by the solute) is included in the Hamiltonian in a self-consistent fashion. Two widely used classes of SCRF-type models are the polarizable continuum model (PCM) of Tomasi *et al.* [3] and the conductor-like screening model (COSMO) of Klamt and Schüürmann [4].

The shape of the cavity containing the solute varies between models —early models used spherical or elliptical cavities; in more recent models the cavity is usually constructed from overlapping atomic spheres of varying radii, which necessitates using a number of parameters.

^(a)Also at: Faculty of Technical Physics and Applied Mathematics, Gdansk University of Technology - Narutowicza 11/12, 80-952 Gdańsk, Poland, EU; E-mail: jaca@kdm.task.gda.pl; jd12g09@soton.ac.uk

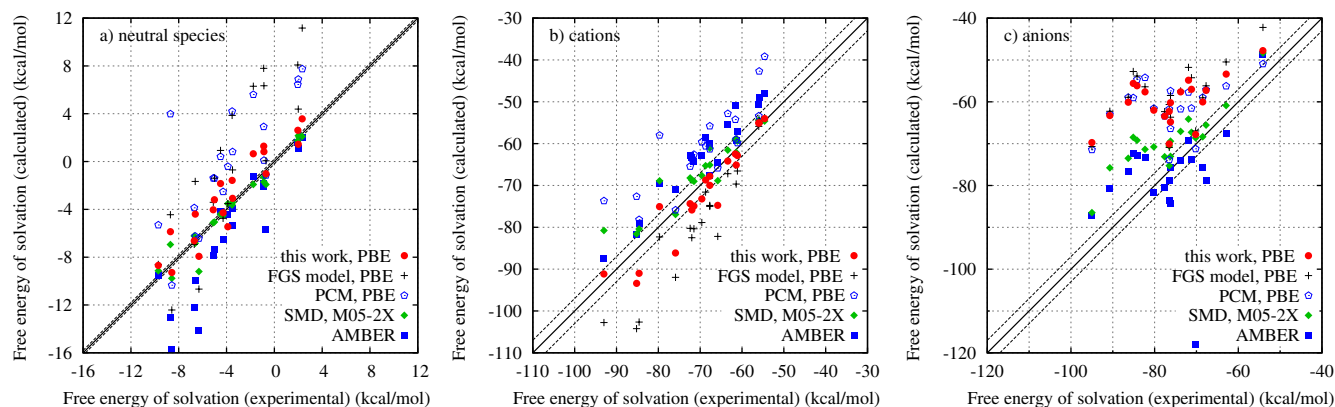


Fig. 1: (Colour on-line) Calculated free energies of solvation plotted against corresponding experimental values — a comparison between this work and other models. The solid diagonal line represents perfect agreement with experiment, the dashed black lines denote the estimated uncertainty of the experimental values.

In contrast, the recent model proposed by Fattbert and Gygi [5,6], and later developed by Scherlis *et al.* [7] (henceforth called the FGS model), utilizes a dielectric cavity constructed directly from the electronic density of the solute, which greatly reduces the number of parameters involved.

This “minimal-parameter” nature of the FGS model makes it attractive for *ab initio* calculations. However, this approach also has several shortcomings, which we address in this letter. First, the original model did not include dispersion-repulsion effects, which, as the authors themselves note, is likely to impact its accuracy for larger neutral molecules. Second, the model necessitates the use of an *a posteriori* correction to the energy in vacuum, obtained in periodic boundary conditions, to approximate open boundary conditions, whereas in the solvent the electrostatic energy is obtained subject to zero boundary conditions. Third, a severe numerical instability prevents this approach from being practical for large molecules.

This letter describes how we have addressed these limitations, by including dispersion-repulsion interactions, employing open (Coulombic) boundary conditions, and identifying and circumventing the root cause of the above-mentioned numerical instability. We then validate and evaluate the performance of the model on two sets of several tens of small molecules. Finally, by performing a calculation on a 2615-atom protein-ligand system, we demonstrate how the implemented model can be used to perform large-scale *ab initio* calculations in solution.

In contrast to other SCRF models where the solute cavity has a discontinuous boundary, the FGS model defines a smooth transition of the relative permittivity according to

$$\varepsilon(\mathbf{r}) = 1 + \frac{\varepsilon_\infty - 1}{2} \left(1 + \frac{1 - (\rho(\mathbf{r})/\rho_0)^{2\beta}}{1 + (\rho(\mathbf{r})/\rho_0)^{2\beta}} \right), \quad (1)$$

where $\rho(\mathbf{r})$ is the electronic density of the solute, ε_∞ is the bulk permittivity, the parameter β controls the

smoothness of the transition of $\varepsilon(\mathbf{r})$ from 1 to ε_∞ , and ρ_0 is the density value for which the permittivity drops to $\varepsilon_\infty/2$. The cavitation contribution to the free energy is assumed to be proportional to the surface area, S , of the cavity (calculated at $\rho = \rho_0$), that is $\Delta G_{\text{cav}} = \gamma S(\rho_0)$, where γ is the solvent surface tension. Values for β and ρ_0 are found by a least-squares fit to the hydration energies of ammonia, nitrate and methylammonium (representative of neutral, anionic and cationic molecules, respectively) [7]. The total potential of the solute in the presence of the dielectric, $\phi(\mathbf{r})$ is obtained by solving the nonhomogeneous Poisson equation

$$\nabla \cdot (\varepsilon[\rho] \nabla \phi) = -4\pi\rho_{\text{tot}} \quad (2)$$

directly in real space subject to zero Dirichlet boundary conditions. The total charge density $\rho_{\text{tot}}(\mathbf{r})$ is a sum of the electronic density $\rho(\mathbf{r})$ and a Gaussian-smeared density of the cores, as proposed in ref. [7].

As outlined in ref. [7], the fact that the dielectric cavity responds self-consistently to changes in the electronic density means that the functional derivative of the electrostatic energy, E_{es} , is no longer equal to the potential that is the solution of eq. (2), but rather

$$\frac{\delta E_{\text{es}}}{\delta \rho}(\mathbf{r}) = \phi(\mathbf{r}) - \frac{1}{8\pi} (\nabla \phi(\mathbf{r}))^2 \frac{\delta \varepsilon}{\delta \rho}(\mathbf{r}). \quad (3)$$

The original FGS model does not set out to address dispersion-repulsion effects. This makes the results obtained for larger molecules dubious, especially for those that are neutral, as then the electrostatic contribution to solvation would be dwarfed by the nonpolar terms. As the authors duly note, this deficiency already becomes evident for the case of benzene where this model predicts a ΔG of 7.9 kcal/mol [7] whereas the experimental value is -0.87 kcal/mol [8]. To appreciate the magnitude of the problem we refer to fig. 1 and the top of table 1, where results obtained with the FGS model are shown for a representative selection of 20 neutral, 20 cationic and 20 anionic molecules chosen from ref. [8]. The geometries of

Table 1: Error (root mean square (rms) and maximum, in kcal/mol), with respect to experiment [8], in the calculated free energies of solvation, and the corresponding coefficient of correlation, r , between the calculated and experimental values, for the 20 neutral, 20 cationic and 20 anionic species studied.

Approach	XC functional	Neutral species			Cations			Anions		
		rms err.	max err.	r	rms err.	max err.	r	rms err.	max err.	r
FGS	PBE	5.0	8.8	0.87	9.7	19.0	0.95	19.5	32.4	0.55
This work ^a	PBE	1.6	2.8	0.93	4.4	10.2	0.95	18.1	29.5	0.53
This work ^b	PBE	5.0	8.9	0.87	10.4	19.0	0.95	21.2	35.1	0.54
This work ^c	PBE	1.8	3.1	0.93	3.9	8.3	0.94	18.1	29.4	0.54
PCM	PBE	4.9	12.7	0.75	10.5	21.7	0.83	17.8	29.5	0.36
PCM	B3LYP	4.7	12.0	0.78	10.4	21.8	0.83	17.0	28.4	0.41
PCM	M05-2X	4.4	11.1	0.79	10.2	21.7	0.81	15.7	26.8	0.46
SMD	M05-2X	0.9	2.9	0.97	4.6	12.2	0.95	8.5	16.6	0.86
AMBER [9]	(classical)	3.3	7.84	0.64	6.9	10.8	0.96	12.8	47.8	0.32

^aWith the cavity responding self-consistently to changes in density.

^bWith the cavity responding self-consistently to changes in density, and without dispersion-repulsion effects.

^cWith the cavity fixed.

the molecules were not re-optimized in solvent, instead geometries optimized in the gas phase readily available from ref. [8] were used. The FGS model underestimates the solvation effect for anions and overestimates it for cations. The predictions for neutral species are in moderate agreement with experiment. By examining the coefficient of correlation between the calculated and experimental values, we demonstrate that the obtained values do not correlate well with experiment (with the notable exception of cations), which makes the calculation of *relative* free energies of solvation, $\Delta\Delta G$, unreliable.

The second shortcoming of this approach is related to the boundary conditions used for the solution of eq. (2). References [5,6] are concerned only with calculations in solution, where zero boundary conditions are used. Owing to the dielectric screening, this is a reasonable approximation, as long as the relative permittivity of the solvent is large. Reference [7] uses the same approach in solution, whereas for the reference vacuum calculation (needed to obtain free energies of solvation), where the Poisson equation becomes homogeneous, standard periodic plane-wave DFT calculations are performed. *In vacuo* energies thus obtained are subsequently corrected with the Makov-Payne formula [10] to mimic the effect of open boundary conditions. This too is an approximation, since the correction cannot fully capture polarization effects [10]. Furthermore, only the energy is corrected, while the shape of the electronic density, and, in turn, the cavity generated in solution corresponds to periodic boundary conditions. As we demonstrate later, this subtly affects the free energies of solvation obtained for charged species, leading to a degree of cancellation of errors.

Further, we point out the root cause of the numerical instability inherent in the FGS model. The second term in the RHS of eq. (3) is extremely difficult to evaluate accurately, because $\frac{\delta\epsilon}{\delta\rho}$ is very close to zero everywhere, except on the boundary of the cavity, where, in turn, $(\nabla\phi(\mathbf{r}))^2$ is almost zero and thus difficult to distinguish from

numerical noise. Because of this, the energy gradient calculated from eq. (3) is not numerically accurate and the method is found to converge only when high-order finite-differences and extremely fine grids (with a spacing of $0.15a_0$ or finer) are used, as only then the gradient of the potential can be evaluated to sufficient accuracy. The memory requirements necessitated by such fine grids quickly make the technique impractical for larger molecules.

By addressing each of these limitations, we obtain a highly accurate and usable approach which retains the conceptual elegance of the FGS model.

We solve eq. (2) by means of a second-order multigrid [11,12] approach, which is subsequently defect-corrected [13] in an iterative fashion using 10th order finite-difference stencils for the first and second derivatives. We find that with a grid spacing of $0.125a_0$ as few as 3–4 defect-correction iterations are sufficient to reduce the algebraic error in the obtained potential by four orders of magnitude with respect to the initial, uncorrected solution. The corresponding reduction in the magnitude of the residual is two orders of magnitude, due to the approximate nature of the calculated boundary conditions. With a grid spacing of $0.25a_0$ ten iterations, on average, were necessary.

We have recast the solvation problem into open boundary conditions by computing the core-core and the local pseudopotential terms in real space and by using open (Coulombic) boundary conditions when solving eq. (2)—that is, we set up Dirichlet boundary conditions by evaluating the Coulombic potential due to the charge distribution ρ_{tot} . Since with a spatially localized density the calculation of the Coulombic integral for all the points on the boundary scales as $\mathcal{O}(L^2N_{\text{at}})$ with the box length L and number of atoms N_{at} , charge coarse-graining and interpolation were used to reduce the prefactor in this calculation by about three orders of magnitude. In so doing, we obtain *in vacuo* energies and densities that need

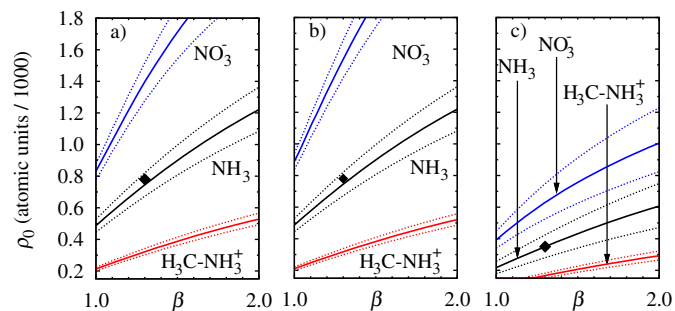


Fig. 2: (Colour on-line) Isolines of zero error with respect to experiment for the possible parametrizations of a) the FGS model, b) the FGS model with the boundary conditions we propose and c) the model proposed in this work, for three representative molecules. The dashed lines indicate an error of ± 1 kcal/mol. The black diamonds indicate the final parametrization of the respective models.

not be corrected. In the solvated case, where the nonhomogeneity in ϵ prevents such an approach, we calculate the boundary conditions by approximating the dielectric as homogeneous, with the permittivity of the bulk solvent. Figure 2, panels a) and b) and table 1, rows 1 and 3 illustrate that for the simulation cell we used (a cube with an edge length $L = 40.5 a_0$) this alone offers no improvement in accuracy. A difference in the free energy of solvation of 0.1 kcal/mol, 0.9 kcal/mol and -1.8 kcal/mol is observed on average for neutral, cationic and anionic species, respectively, compared to the predictions of the original FGS model (refer to fig. 3 for details). For charged species the application of consistent open boundary conditions leads to a slight increase in the error.

A parameter sweep for the three molecules used to parametrize the FGS model demonstrates (cf. fig. 2) that there exists no parametrization that would result in even moderate agreement with experiment for the three species simultaneously. The model would consistently either underestimate free energies of solvation for anions or overestimate them for cationic species.

We attribute this failure to a combination of factors—the poor performance of the Perdew-Burke-Ernzerhof (PBE) exchange-correlation (XC) functional; the fact that *any* isodensity formulation will use larger cavities for anions than for corresponding cations, whereas the charge asymmetry in solvation effects is in fact opposite [14]; and, finally, to the lack of inclusion of dispersion-repulsion effects, which leads to an overestimation of the nonpolar component of solvation. The middle rows of table 1 show, on the example of PCM, how using a hybrid functional such as B3LYP or a hybrid meta-GGA functional such as M05-2X [15] addresses the first problem, by reducing the self-interaction error, which otherwise leads to excessive delocalization of the electrons, but does not address the other two problems.

The increase in the magnitude of the difference between the calculated and experimentally obtained free energies

of solvation with the size of the molecule, especially in the case of neutral molecules, demonstrated in fig. 3 indicates that the neglect of dispersion-repulsion effects is detrimental to the predictive quality of the FGS model. We propose including dispersion-repulsion effects in the free energy of solvation, $\Delta G_{\text{dis,rep}}$, using an approximate relation derived by Floris *et al.* [16]. Since this relation is linear, it amounts to a simple rescaling of the surface tension of the solvent, including the approximate $\Delta G_{\text{dis,rep}}$ in the cavitation term. From the slope of the linear relation plotted in fig. 1 of ref. [16] it follows that the surface tension should be rescaled by a factor of 0.281. Even this crude method for taking dispersion-repulsion into account dramatically improves the accuracy of the model, as evidenced by figs. 1 and 2 and table 1, from which it is apparent that the resulting approach is in much better agreement with experiment than both PCM and the force-field Poisson-Boltzmann (PB) approach of AMBER [9], offering comparable quality to the much more complex SMD¹ [17] model. The improvement offered by the inclusion of dispersion-repulsion effects is evidenced by fig. 3 and can be quantified by comparing rows denoted with *a* and *b* in table 1. The results corresponding to the row denoted with *b* were obtained by turning off the dispersion-repulsion contribution whilst using the parameters proposed in ref. [7], denoted with a diamond in fig. 2, panels a) and b).

The numerical instability caused by the second term in the RHS of eq. (3) can be circumvented without loss of accuracy. We first note that this term disappears when, instead of responding to changes in the electronic density, the dielectric cavity is fixed. We propose constructing the cavity by the application of eq. (1) to the converged electronic density of the solute obtained in the vacuum calculation and keeping the cavity fixed throughout the calculation in solvent. We show (cf. tables 1, 2) that the associated reduction in accuracy is insignificant, while both the wall time and the memory requirements of the computation are reduced by about an order of magnitude, as convergence is readily achieved with a more reasonable real-space grid spacing of $0.25 a_0$. We should point out that a similar attempt to fix the cavity in the FGS model would probably lead to larger errors due to the fact that the fixed cavity would come from the *periodic* density of the vacuum calculation—as the Makov-Payne correction [10] only corrects the energy. We note that this simplified approach is still suitable for geometry optimization in solution, provided the additional contribution to the forces due to the cavity variation with atomic positions is included. Sánchez *et al.* [18] also note the above-mentioned

¹SMD is a recently proposed model based on the integral-equation-formalism PCM (IEF-PCM), which yields excellent agreement with experiment. This requires, however, making use of an extensive set of parameters to describe the solute (intrinsic Coulomb radii, atomic surface tension parameters) and the solvent (refractive index and acidity and basicity parameters; in addition to the dielectric constant and bulk surface tension needed in the proposed model).

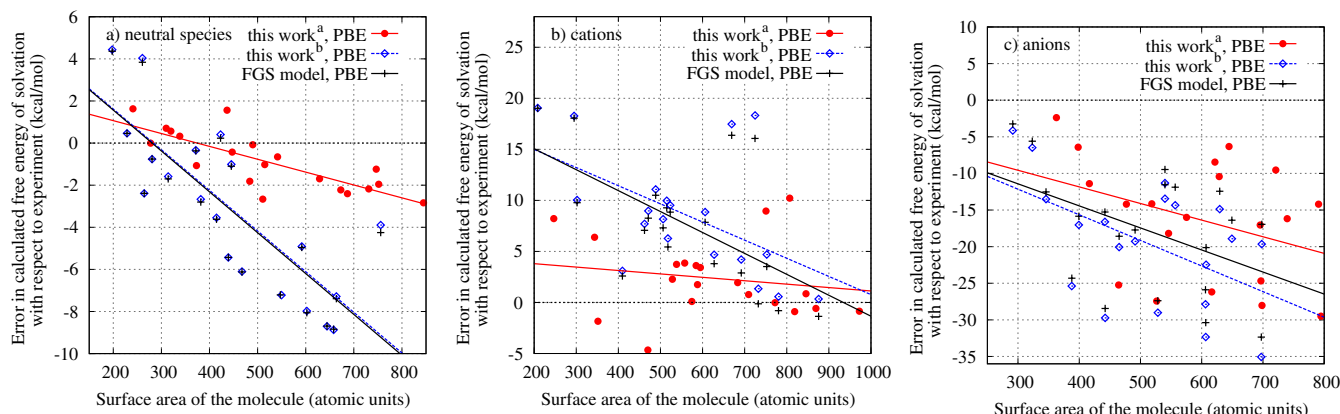


Fig. 3: (Colour on-line) Absolute error in the calculated free energies of solvation with respect to experimental values, plotted against the surface area of the molecule. The surface areas differ between models because the parametrization is different. The solid and dashed lines represent a linear fit. The horizontal dotted line corresponds to perfect agreement with experiment.

Table 2: Error (root mean square (rms) and maximum, in kcal/mol), with respect to experiment, in the calculated free energies of solvation, and the corresponding coefficient of correlation, r , between calculated and experimental values for the 71 molecules studied [19,20].

Approach	XC functional	rms error	max error	r
This work ^a	PBE	3.8	8.3	0.83
This work ^b	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

^aWith the cavity responding self-consistently to changes in density.

^bWith the cavity fixed.

instability and propose a somewhat different way of circumventing it.

We further validate our model on 71 neutral molecules taken from the blind tests of refs. [19,20], for which the experimental energies of solvation are reported in ref. [8]. Again, the geometries were not re-optimized in solution, but rather the gas-phase geometries from ref. [8] were used. The results, shown in fig. 4 and table 2, again show that our approach is consistently more accurate than both PCM and the force-field PB approach of AMBER [9] and that our model offers a level of agreement with experiment which is comparable to the SMD [17] model, even when the cavity is fixed.

Conventional *ab initio* calculations are typically limited to only a few hundred atoms at most. However, with recent advances in linear-scaling density functional theory (LS-DFT) approaches [21] a number of codes [22–24] have been developed which are capable of performing calculations with many thousands of atoms. The combination of LS-DFT with implicit solvent models would enable highly realistic simulations of important phenomena such as

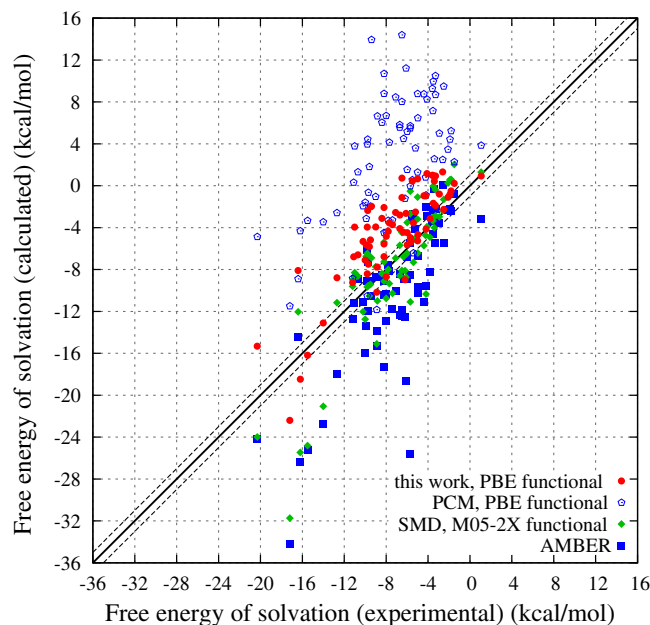


Fig. 4: (Colour on-line) Calculated free energies of solvation plotted against corresponding experimental values for the 71 neutral molecules studied [19,20] —a comparison between models. The solid diagonal line represents perfect agreement with experiment, the dashed black lines denote the estimated uncertainty in the experimental values.

biomolecular processes or the chemical modification and self-assembly of nanostructures.

As a demonstration of the potential applications of this approach in large-scale DFT calculations, we have implemented our solvent model in the LS-DFT code ONETEP [22] and used it to calculate the free energy of solvation for a 2615-atom system composed of the catechol ligand bound to a L99A/M102Q mutant of the T4 lysozyme. The results of this calculation are shown in table 3 and the overview of the system in question, along with an outline of the dielectric cavity (at $\rho = \rho_0$) is

Table 3: Free energies of solvation (in kcal/mol) of L99A/M102Q T4 lysozyme (ΔG_{host}), its complex with catechol (ΔG_{cplx}), catechol (ΔG_{lig}), desolvation energy of the ligand ($\Delta G_{\text{d}} = \Delta G_{\text{cplx}} - \Delta G_{\text{host}} - \Delta G_{\text{lig}}$), binding energy *in vacuo* (ΔE_{gas}) and in solvent (ΔE_{sol}), as predicted by our model (with PBE and a fixed cavity) and AMBER.

Approach	ΔG_{cplx}	ΔG_{host}	ΔG_{lig}	ΔG_{d}	ΔE_{gas}	ΔE_{sol}
This work	-2423.0	-2421.3	-7.5	5.8	-28.6	-22.8
AMBER	-2428.3	-2433.0	-17.6	22.4	-27.7	-5.3
expt. [25]	—	—	-9.3	—	—	—

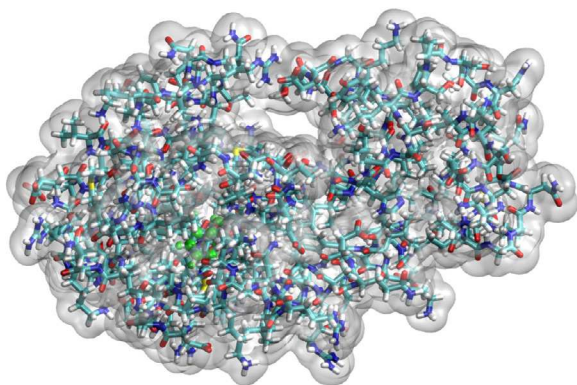


Fig. 5: (Colour on-line) An overview of the lysozyme-catechol complex, with the dielectric cavity indicated in grey and catechol in green.

shown in fig. 5. AMBER greatly overestimates the solvation energy of catechol, and consequently the binding energy in solvent differs significantly between the two models. The need for an *ab initio* model, where the density is polarized by the solvent is demonstrated by the different behavior of ΔG_{host} (as compared to ΔG_{cplx}) between the proposed model and AMBER.

In summary, we have outlined and validated an implicit solvent model for *ab initio* calculations, which, despite using only two parameters, offers a substantial improvement over existing models, as measured by the agreement of absolute and relative free energies of solvation with experiment (compared to PCM and AMBER) or the number of parameters needed to achieve similar agreement (compared to SMD). We have shown how the implementation of the proposed model in the LS-DFT code ONETEP paves the way for first-principles implicit-solvent calculations for molecules with thousands of atoms.

The calculations were carried out on the Iridis3 super-computer of the University of Southampton and on the Darwin Supercomputer of the University of Cambridge High Performance Computing Service provided by Dell

Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England. JD, C-KS, AAM and MCP acknowledge funding from EPSRC Grant EP/G055904/1. HHH acknowledges the support of the Yousef Jameel PhD scholarship administered by the University of Cambridge. C-KS would like to thank the Royal Society for a University Research Fellowship. AAM acknowledges the support of an RCUK academic fellowship. MCP acknowledges funding from EPSRC Grant EP/F0327731/1.

REFERENCES

- [1] HENCHMAN R. H. and ESSEX J. W., *J. Comput. Chem.*, **20** (1999) 499.
- [2] TOMASI J., MENNUCCI B. and CAMMI R., *Chem. Rev.*, **105** (2005) 2999.
- [3] TOMASI J. and PERSICO M., *Chem. Rev.*, **94** (1994) 2027.
- [4] KLAMT A. and SCHÜÜRMAN G., *J. Chem. Soc. Perkin Trans. 2*, issue No. 5 (1993) 799.
- [5] FATTEBERT J.-L. and GYGI F., *J. Comput. Chem.*, **23** (2002) 662.
- [6] FATTEBERT J.-L. and GYGI F., *Int. J. Quantum Chem.*, **93** (2003) 139.
- [7] SCHERLIS D., FATTEBERT J., GYGI F., COCCIONI M. and MARZARI N., *J. Chem. Phys.*, **124** (2006) 074103.
- [8] MARENICH A. *et al.*, *Minnesota solvation database, version 2009* (University of Minnesota Minneapolis) 2009.
- [9] CASE D. A. *et al.*, *AMBER 10* (University of California, San Francisco) 2010.
- [10] MAKOV G. and PAYNE M. C., *Phys. Rev. B*, **51** (1995) 4014.
- [11] BECK T. L., *Rev. Mod. Phys.*, **72** (2000) 1041.
- [12] HOLST M. and SAIED F., *J. Comput. Chem.*, **14** (1993) 105.
- [13] TROTTERBERG U., OOSTERLE C. W. and SCHULLER A., *Multigrid* (Academic Press) 2000.
- [14] MOBLEY D. L., BARBER A. E., FENNELL C. J. and DILL K. A., *J. Phys. Chem. B*, **112** (2008) 2405.
- [15] ZHAO Y., SCHULTZ N. E. and TRUHLAR D. G., *J. Chem. Theory Comput.*, **2** (2006) 364.
- [16] FLORIS F. M., TOMASI J. and AHUIR J. L. P., *J. Comput. Chem.*, **12** (1991) 784.
- [17] MARENICH A. V., CRAMER C. J. and TRUHLAR D. G., *J. Chem. Theory Comput.*, **5** (2009) 2447.
- [18] SÁNCHEZ V., SUED M. and SCHERLIS D., *J. Chem. Phys.*, **131** (2009) 174108.
- [19] NICHOLLS A. *et al.*, *J. Med. Chem.*, **51** (2008) 769.
- [20] GUTHRIE J. P., *J. Phys. Chem. B*, **113** (2009) 4501.
- [21] GOEDECKER S., *Rev. Mod. Phys.*, **71** (1999) 1085.
- [22] SKYLARIS C.-K., HAYNES P. D., MOSTOFI A. A. and PAYNE M. C., *J. Chem. Phys.*, **122** (2005) 084119.
- [23] BOWLER D. R., CHOUDHURY R., GILLAN M. J. and MIYAZAKI T., *Phys. Status Solidi (b)*, **243** (2006) 989.
- [24] ARTACHO E. *et al.*, *J. Phys.: Condens. Matter*, **20** (2008) 064208.
- [25] MORDASINI T. Z. and MCCAMMON J. A., *J. Phys. Chem. B*, **104** (2000) 360.