The Gaussian Generalized Born model: application to small molecules

J. A. Grant,*[a] B. T. Pickup,*[b] M. J. Sykes,[b] C. A. Kitchen[b] and A. Nicholls[c]

Received 18th May 2007, Accepted 26th June 2007
First published as an Advance Article on the web 17th July 2007
DOI: 10.1039/b707574j

This work presents a Generalized Born model for the computation of the electrostatic component of solvation energies which is based on volume integration. An analytic masking function is introduced to remove Coulombic singularities. This approach leads to analytic formulae for the computation of Born radii, which are differentiable to arbitrary order, and computationally straightforward to implement.

1. Introduction

Modern drug design requires large scale modelling. The ultimate need is for many calculations in a short space of time. For example, the early phase of drug discovery requires the identification of molecules that exhibit even weak binding to a biological target. A common design strategy is to search for molecules that are similar in shape and electrostatic features to a known ligand.[1,2] Similar strategies are also adopted at a later phase in the drug discovery process to find iosteric replacements necessary to improve binding (or some other physical property). Crucial to the success of these approaches are large-scale multi-conformer databases. The modelling of large numbers of conformers requires reliable molecular mechanics force-fields. An essential component of modelling in this context is the ability to include solvation effects, because ligands bound to protein are more likely to resemble solvated conformers. Such high-throughput modelling obviates the use of explicit solvation models. Hence, there is a need for computationally inexpensive and reliable schemes for simulating the solvation properties of small drug-like molecules. Central to these requirements therefore is an understanding of the role of electrostatic interactions in aqueous solution. These computational requirements dictate that implicit (or continuum) solvent models, which describe the average behaviour of solvent, are preferred to atomistic models. Continuum models based on the Poisson–Boltzmann equation have given significant insight into the behaviour of biological systems.[3–8] In large scale modelling there is a need for methodologies with much greater computational efficiency. Atom-based 'semi-analytical' models have been developed which fulfil the role of greater efficiency with marginally less accuracy, and which can be effectively integrated into conventional molecular mechanics techniques. These methods have the advantage also that they enable more efficient implementation of the high order gradients, which are required to explore potential surfaces and facilitate the computation of normal modes.[9–11] Early atom-based approaches were based on treatments of how atomic surfaces, or volumes, were to a greater or lesser extent exposed to solvent.[12–14] Later, there were attempts to model the physics of the problem using different treatments for electrostatic and non-electrostatic contributions to the solvation energy.[15,16] In particular, the pioneering work of Still[17] introduced the Generalized Born (GB) approach to compute the electrostatic component of the solvation energy. The GB method divides the electrostatic term into atomic (self) and pair terms. Central to the model is the calculation of an effective atomic radius for each atom. These Born radii account for the degree to which each atom is screened from solvent by surrounding atoms. The original formulation of these ideas required a time consuming numerical treatment for the calculation of the Born radii. Subsequent improvements attempted to provide explicit parameterizations of the way in which an atom is screened from the solvent by its neighbours,[18,19] or hybrid treatments that used approximate analytic expressions for the self term combined with the GB approximation for the pair term.[20] These approaches led to explicit differentiable formulae for Born radii, with the drawback that they contain arbitrary parameters. Around the same time a development in Poisson–Boltzmann methodology[21–23] was the introduction of continuous smooth permittivity functions. These were introduced partly to address a range of algorithmic issues, but also as an attempt at a more realistic description of the dielectric boundary. Modifications of the GB model[24–29] have incorporated the smoothing concept to produce new expressions for Born radii. These methods use either numerical integration techniques or explicit formulae to compute Born radii. Two of these techniques which like the present work use Gaussians, namely the ACE method[20] and the AGBNP method,[28] are described in more detail in the Conclusions.

The present work introduces a simple analytical GB solvation model, which uses Gaussians to reproduce both the dielectric behaviour around a charged atom and the solvent screening effects in its locality. This produces a purely analytical expression for the Born radius which is differentiable to all orders, and dependent only on a single parameter. It also gives a method which is efficient, and is comparable in accuracy to the other GB approaches we have mentioned above.

[a] *AstraZeneca Pharmaceuticals, Mereside, Macclesfield, Cheshire, UK SK10 4TF. E-mail: Andrew.Grant@astrazeneca.com*
[b] *Department of Chemistry, The University of Sheffield, Sheffield, UK S3 7HF. E-mail: B.T.Pickup@sheffield.ac.uk*
[c] *OpenEye Scientific Software, Santa Fe, New Mexico 87507, USA*

2. Methods

The Generalized Born (GB) model decomposes the electrostatic contribution to solvation energy into atomic and pair contributions. The solvation energy can be written as

$$E^{\mathrm{solv}} = \sum_I E_I^{\mathrm{self}} + \sum_{I>J} E_{IJ}^{\mathrm{pair}} \tag{2.1}$$

where a self term (E_I^{self}) represents the interaction of a charged atom with the polarization it produces in the dielectric. The pair terms, on the other hand, represent the interaction between atoms and the polarization produced by other atoms. In the limit of a single ion of radius σ_I in a discontinuous dielectric, the expression for E_I^{self} reduces to the well known Born equation (in SI units)

$$E_I^{\mathrm{Born}} = -\frac{f_\varepsilon}{4\pi\varepsilon_0}\frac{Q_I^2}{2\sigma_I}, \tag{2.2}$$

where ε_0 is the vacuum permittivity. The factor

$$f_\varepsilon = \left(\frac{1}{\varepsilon_{\mathrm{in}}} - \frac{1}{\varepsilon_{\mathrm{out}}}\right), \tag{2.3}$$

where $\varepsilon_{\mathrm{in}}$ and $\varepsilon_{\mathrm{out}}$ are the relative permittivities (dielectric constants) in the solute and in the solvent, respectively. The essence of the GB model is to calculate the self terms using an effective radius (σ_I^B) to model the local atomic environment, hence

$$E_I^{\mathrm{self}} = -\frac{f_\varepsilon}{4\pi\varepsilon_0}\frac{Q_I^2}{2\sigma_I^B}. \tag{2.4}$$

"Born" radii are also used in an approximate expression for the pair terms. Although Poisson–Boltzmann approaches can be used to compute "perfect" Born radii,[30] the utility of GB models arises from approximate calculations of Born radii which are computationally efficient. In what follows we outline the basis of the approach of the Gaussian Generalized Born (GGB) model to compute Born radii.

The explicit expression in GB theory for a self term (derived in Appendix A) is

$$E_I^{\mathrm{self}} = -\frac{Q_I^2}{8\pi\varepsilon_0}\int \mathrm{d}\boldsymbol{r}\frac{1}{r_I^4}\frac{1}{4\pi}\left(\frac{1}{\varepsilon_{\mathrm{in}}} - \frac{1}{\varepsilon(\boldsymbol{r})}\right), \tag{2.5}$$

where our notation indicates the 'quantum chemistry' convention that unlabelled integrals are over all space, Q_I is the charge on atom I, r_I is the local radial co-ordinate around atom I, $\varepsilon_{\mathrm{in}}$ is the relative permittivity of atom I, and $\varepsilon(\boldsymbol{r})$ is an inhomogeneous scalar relative permittivity. This leads to an expression for the Born radius σ_I^B

$$\frac{1}{\sigma_I^B} = \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{r}\frac{1}{r_I^4}\left(\frac{1}{\varepsilon_{\mathrm{in}}} - \frac{1}{\varepsilon(\boldsymbol{r})}\right)\Big/\left(\frac{1}{\varepsilon_{\mathrm{in}}} - \frac{1}{\varepsilon_{\mathrm{out}}}\right), \tag{2.6}$$

which is an effective radius for an atom. The first step in achieving our goal is to produce a suitable expression for the relative permittivity function. The atomic nature of the expression for σ_I^B means that only the behaviour of the relative permittivity around centre I needs to be explored. Hence

$$\frac{1}{\varepsilon(\boldsymbol{r})} \rightarrow \frac{1}{\varepsilon_I(\boldsymbol{r})} = \frac{1}{\varepsilon_{\mathrm{in}}} + \left(\frac{1}{\varepsilon_{\mathrm{out}}} - \frac{1}{\varepsilon_{\mathrm{in}}}\right)\Theta_I^{\mathrm{mol}}(\boldsymbol{r}), \tag{2.7}$$

where the local exclusivity function, $\Theta_I^{\mathrm{mol}}(\boldsymbol{r})$, limits to 1 in the solvent region outside the molecule, and to 0 at the centre of atom I. The equation for the Born radius becomes

$$\frac{1}{\sigma_I^B} = \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{r}\Theta_I^{\mathrm{mol}}(\boldsymbol{r})\frac{1}{r_I^4}. \tag{2.8}$$

The spirit of this model is to view the solvation shell arising from atoms in the vicinity of I as being a perturbation to a naked ion embedded in solvent. We write the exclusion function as a sum of two terms

$$\Theta_I^{\mathrm{mol}}(\boldsymbol{r}) = \Theta_{m,I}(\boldsymbol{r}) + \Theta_{s,I}(\boldsymbol{r}), \tag{2.9}$$

where the exclusion function is divided into a 'mask', $\Theta_{m,I}(\boldsymbol{r})$, and a 'screen', $\Theta_{s,I}(\boldsymbol{r})$ part. The 'mask' part represents the naked ion in solvent. The screening term describes the effects of neighbouring atoms. The definition of the molecular exclusion function given in eqn (2.9) leads to a division of eqn (2.8) into two terms

$$\frac{1}{\sigma_I^B} = \frac{1}{\sigma_{m,I}^B} + \frac{1}{\sigma_{s,I}^B}, \tag{2.10}$$

where

$$\frac{1}{\sigma_{m,I}^B} = \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{r}\Theta_{m,I}(\boldsymbol{r})\frac{1}{r_I^4}, \tag{2.11}$$

and

$$\frac{1}{\sigma_{s,I}^B} = \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{r}\Theta_{s,I}(\boldsymbol{r})\frac{1}{r_I^4}. \tag{2.12}$$

A central requirement of the model is that the mask exclusion function in eqn (2.11) should be defined so that $\sigma_{m,I}^B = \sigma_I$, the inherent radius of atom I. This ensures that we reproduce eqn (2.2) when other atoms are absent, and that the model is reduced to the need to calculate only eqn (2.12).

A. The Gaussian Generalized Born approximation

The computation of Born radii using eqn (2.8) presents the following challenges. Firstly, we require an analytic expression for the exclusion function which produces potentials and solvation energies which are as close as possible to the better founded Poisson–Boltzmann approach. Secondly, we want to be able to solve the volume integral in eqn (2.8) analytically. The motivation of the present model arose from our successful work on molecular volumes using Gaussian representations of molecular exclusion functions. Gaussians have well-known advantages with regard to their ability to produce analytic formulae for spatial integrals.[31,32] The mask is defined in terms of Gaussians as

$$\Theta_{m,I} = p_m(1 - e^{-\alpha r_I^2}), \tag{2.13}$$

where the exponent

$$\alpha = \kappa_m/\sigma_I^2, \tag{2.14}$$

is defined in terms of the radius, σ_I of the bare atom, and a dimensionless parameter, κ_m. The masking function defined in (2.13) is just strong enough to remove the r_I^{-2} singularity implied by the form of eqn (2.5) for the self energy. Our formalism is defined in such a way that the masking term, when applied to a single atom, should reproduce the Born eqn (2.2). In eqn (2.3) of Appendix B it is shown that

$$p_m = 1/\sqrt{\pi \kappa_m}. \tag{2.15}$$

This fixes the height factor of the masking function p_m, leaving κ_m as the single free variable in the method. This is similar to the approach used in Gaussian volume theory,[33–35] where the Gaussian height and scale factor are related by the need to conserve atomic volumes.

The role of the screening function is to account for modifications to local polarization arising from other atoms which have displaced solvent. This is primarily a volume effect, which is represented using the 'Gaussian shape-model',[33,34] hence we write

$$\Theta_{s,I} = (1 - e^{-\alpha r_I^2})\left(\sum_{J \neq I} g_J - \sum_{J > K \neq I} g_J g_K + \cdots\right), \tag{2.16}$$

where the volume Gaussians

$$g_J = p e^{-\kappa r_J^2/\sigma_J^2}. \tag{2.17}$$

We adopt the standard values $p = 2.70$ and $\kappa = 2.344$ for volume Gaussians.[33,34] In the present work we report only results using single atom terms in eqn (2.16). The use of higher order volume overlaps is straightforward as a result of the properties of the Gaussian product theorem.[31] This allows any high order intersection term to be represented using a single Gaussian. Thus every contribution to the Born radius arising from eqn (2.16) can be computed using a single equation given in Appendix B (2.16). This approach of simplifying the treatment of volume intersections is identical to that used in the computation of Gaussian molecular volumes[33,34] and similar to that used to scale screening contributions in the AGBNP model.[28] The screening mask function in eqn (2.16) contains a common r_I-dependent masking factor to eqn (2.13), but without the factor p_m. The r_I-dependent factor is required to remove the singularity at $r_I \to 0$, but because there is no formal constraint to apply to the screening terms, unlike for the self term (where there is a need to reproduce the Born energy), p_m is not included. The essence of our new method is to produce an expression for the Born radius derived from eqn (2.8) which is essentially analytic, in the sense that no quadrature or other approximate formulae need to be used. This is obtained using the Gaussian mask and screen functions described above, and detailed formulae described in Appendix B. The expression for the pair term within GB theory is

$$E_{IJ}^{\text{pair}} = -\frac{Q_I Q_J}{4\pi\varepsilon_0} \int \mathrm{d}\boldsymbol{r} \frac{\hat{\boldsymbol{r}}_I}{r_I^2} \cdot \frac{\hat{\boldsymbol{r}}_J}{r_J^2} \frac{1}{4\pi}\left(\frac{1}{\varepsilon_{\text{in}}} - \frac{1}{\varepsilon(\boldsymbol{r})}\right). \tag{2.18}$$

In Generalized Born models the analytic pair term is normally replaced by a simple *ansatz* of the form

$$E_{IJ}^{\text{Born,pair}} = -\frac{f_\varepsilon}{4\pi\varepsilon_0} \frac{Q_I Q_J}{\sqrt{\sigma_I^B \sigma_J^B e^{-R_{IJ}^2/4\sigma_I^B \sigma_J^B} + R_{IJ}^2}}. \tag{2.19}$$

This expression interpolates between the long and short range behaviour of the exact pair term, and depends only on distances between atoms, and their Born radii σ_I^B, and σ_J^B.

3. Results

The aim of this section is to demonstrate the performance of the GGB method by a statistical analysis comparing small molecule solvation energies with those obtained from well established GB and PB approaches. The analysis is based upon a data set of molecules obtained from the MDL Drug Database Report (MDDR).[36] For each of these molecules consistent tautomer and ionization states were assigned. Hydrogen positions and three-dimensional coordinates were computed using Corina.[37] These molecules were assigned partial charges and atomic radii based on the MMFF force-field parameters using MacroModel.[38] The electrostatic component of the GB solvation energy was calculated, for purposes of comparison, using a standard implementation in Macromodel. Care was taken to ensure that identical partial charges and radii were used for all subsequent calculations, ensuring comparability of the electrostatic component of the solvation energies between methods. This produced 64 000 neutral molecules (64 K set) and 387 charged molecules, with an average heavy atom count of 26.3. To provide a uniform standard for comparison high resolution (grid spacing 0.1667 Å) Poisson–Boltzmann (PB) calculations were carried out for all molecules, using a molecular surface (MS) dielectric boundary. This is the model often used to guide the parametrization of GB models. The mean solvation energy of the neutral set is −25.4 kcal mol^{-1}, and that of the charged set −70.7 kcal mol^{-1}.

The GGB model contains a single parameter, κ_m, (*cf.* eqn (2.13) and (2.14)), which needs to be chosen to obtain the best possible result. This was done by using a least squared minimisation, in terms of the differences between PB and GGB calculations. The optimization code uses an analytical first derivative of the GGB solvation energy with respect to κ_m, taking advantage of the simple analytical form of the solvation energy. Convergence is rapid, because it has been found empirically that the minimization function has a single well-defined minimum. The minimisations were carried out using ten different subsets of 10 000 molecules randomly chosen from the 64 K set. The optimal value of κ_m was selected as the mean from the ten different minimization procedures. The average value of κ_m was 0.057768 with a very small standard deviation of 5.71×10^{-5}. This value of the κ_m is used in all subsequent GBB calculations referred to in this section. Fig. 1 shows the solvation energy computed using the hard-sphere based standard GB model,[38] compared to accurate Poisson–Boltzmann calculations carried out with an MS dielectric model at 6 grid points per Angstrom. The regression plot has a slope of 0.9611, an offset of −0.2422 kcal mol^{-1}, and an R^2 of 0.9568. The comparison of the GGB model to the standard PB model is demonstrated by the regression graph in Fig. 2. The GGB model captures much of the physics of the Poisson equation as reflected in the slope of 1.0014, and offset of 0.0870 kcal mol^{-1}, with an R^2 of 0.9717. The slopes for both graphs are very close to unity, indicating that they both describe the physics of solvation in a manner which is similar to that in the PB method.

We now introduce cumulative difference graphs to give a clearer representation of the expected accuracy for the GGB

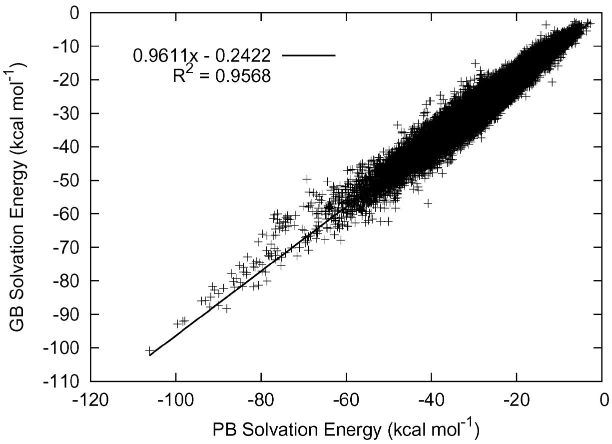

Fig. 1 Accurate GB *versus* Poisson–Boltzmann solvation energies.

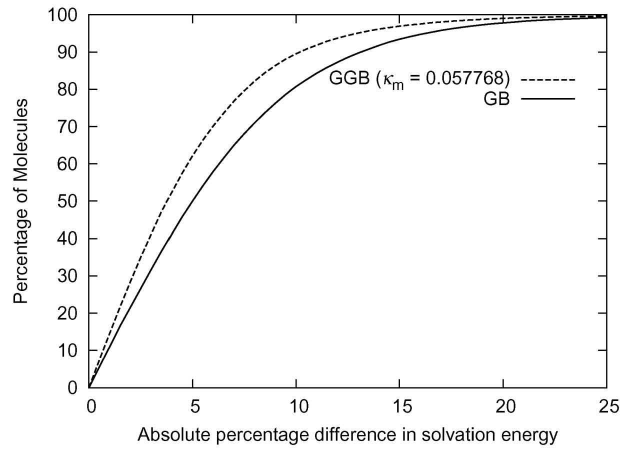

Fig. 3 Cumulative difference plots for GB and GGB models *versus* accurate Poisson–Boltzmann solvation energies.

method compared to the MS PB calculations. These graphs show what fractions of the data set have a certain percentage difference between solvation energies calculated with some given model and MS PB. Fig. 3 shows this for solvation energies computed using GGB, and for reference the GB model implemented in Macromodel. For the GGB model approximately 89% of the molecules have less than 10% difference in the solvation energy. This is similar to that of a conventional GB for which this figure is 81%. For the set of charged molecules, the GGB model has 96% of molecules that have solvation energies within 10% of the PB values (69% for the GB model).

Eqn (2.1) shows that the solvation energy in GB approaches can be decomposed into 'self' and 'pair' contributions in eqn (2.1). The individual contributions are easily computed using grid based Poisson–Boltzmann methods. In the current work, we have carried out this procedure using the high accuracy grids and molecular surface dielectric model, outlined earlier. The average values for the total self energy ($E^{\text{self}} = \Sigma_I E_I^{\text{self}}$), total pair energy ($E^{\text{pair}} = \Sigma_{IJ} E_{IJ}^{\text{pair}}$), and total solvation energy for our 64 K set of neutral molecules were -302.9, 277.2, and -25.4 kcal mol^{-1}, respectively. It follows that calculation of solvation energies may be very sensitive to errors because of the delicate balance of the large terms involved. Fig. 4 shows the cumulative difference graphs of the self and pair terms calculated using the GGB model. The differences in both terms are relatively small—over 96% of the molecules have less than 10% error in both self and pair terms. The errors in the self term are systematically smaller than those in the pair term. It is obvious that there must be a cancellation of differences between self and pair components in order to achieve results such as are shown in Fig. 3. A further insight into this balance of differences is provided by ensuring that the GGB model contains a self term that is identical to that obtained using PB theory. This is possible because the Poisson equation provides a simple (though expensive) route to the determination of Born radii. This is achieved by calculating atomic self energy contributions by repeated Poisson calculations where only single atoms are charged. The PB radii can then be obtained using eqn (2.4). Fig. 5 shows the cumulative difference in E^{pair} when PB radii are used compared to the pair term from the GGB model. However, the error in the GGB pair term is significantly reduced when PB radii are used. For example, 65% of molecules for GGB radii, compared to 98% using PB radii, have differences in pair solvation energies of less than 5%.

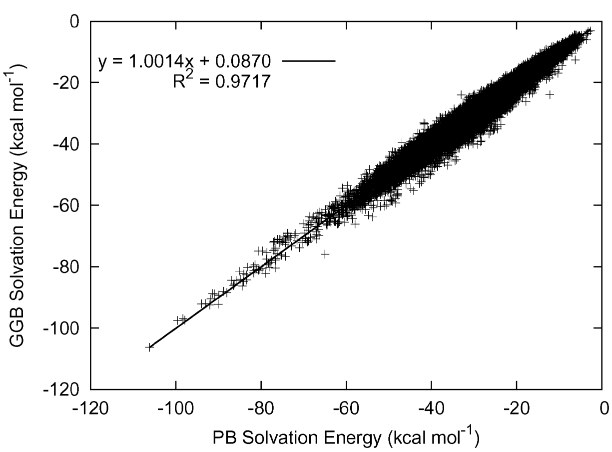

Fig. 2 Accurate GGB *versus* Poisson–Boltzmann solvation energies.

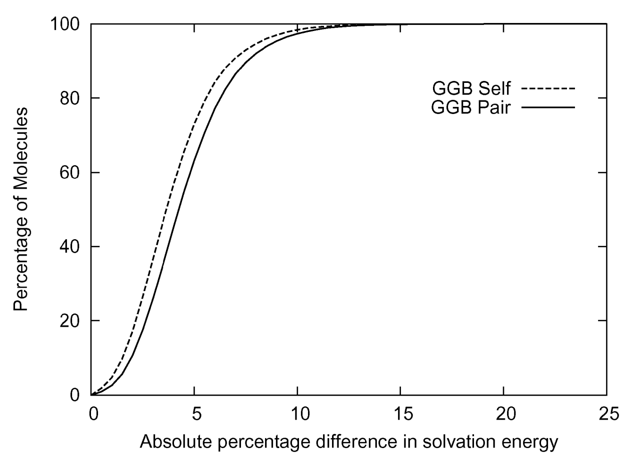

Fig. 4 Cumulative difference plots for GGB self and pair terms compared to accurate Poisson–Boltzmann calculations.

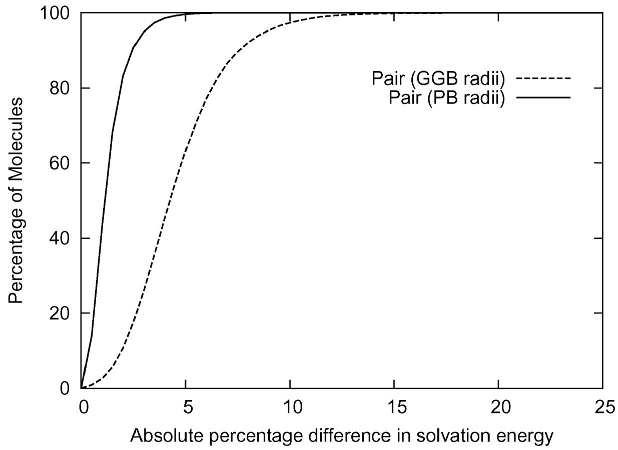

Fig. 5 Cumulative difference plots for total pair energies calculated with the Still formula using PB and GGB Born radii.

Fig. 6 shows the cumulative difference graph for the total solvation energy computed using PB radii in place of Born radii, which appears to perform much worse than the GGB model. The pair energy used in GB models with the Still *ansatz* (2.19) has limited accuracy because it lacks a full description of dielectric polarization. More specifically the GB model adopts the Coulomb-field approximation[39,40] which implies that the dielectric medium only responds to the field produced by the partial charges, and not to itself. However, an adjustment to the Still pair *ansatz* (2.19) can be made by writing

$$E_{IJ}^{\text{pair}} = -\frac{f_{\varepsilon_0}}{4\pi\varepsilon_0}\frac{Q_I Q_J}{\sqrt{\sigma_I^B \sigma_J^B e^{-cR_{IJ}^2/4\sigma_I^B\sigma_J^B} + R_{IJ}^2}}, \tag{3.20}$$

where the "switching" parameter c is allowed to vary from the factor 0.25 introduced by Still.[17] This parameter controls the interpolation between short range ($R_{IJ} \to 0$), and long range (Coulombic) behaviour in the pair term. Optimising the parameter c, by least squares minimization, to c = 0.19051 produces the final curve in Fig. 6. This shows that total solvation energies calculated with PB radii and an optimised Still pair *ansatz* exhibit a small improvement over the GGB

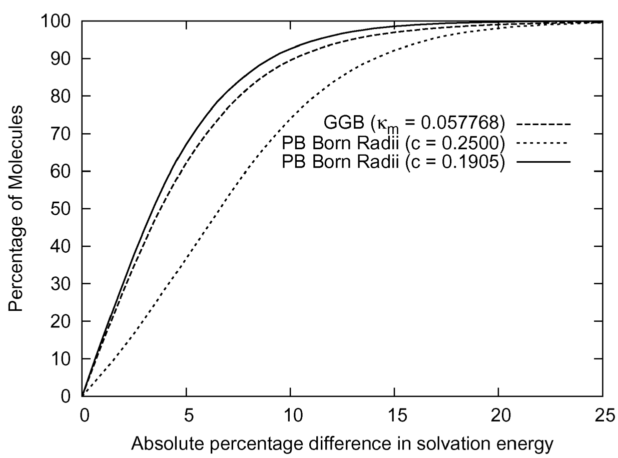

Fig. 6 Cumulative differences for total solvation energies calculated with PB Born radii using the Still formula using c = 0.25 and c = 0.19051. The GGB model is shown as a reference.

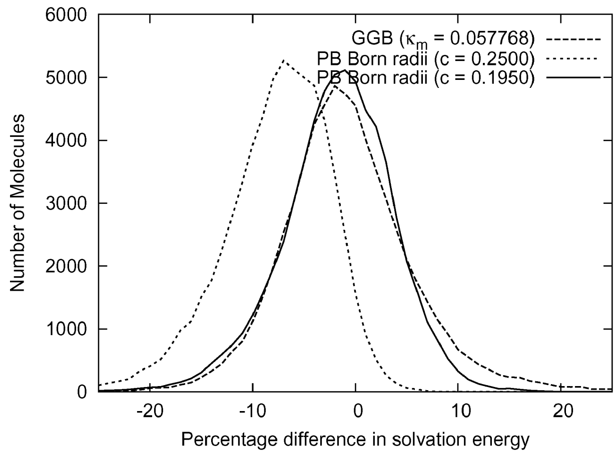

Fig. 7 Frequency distributions for differences in solvation energies with respect to accurate Poisson–Boltzmann calculations. Results are shown for the models in Fig. 6.

model. That is, being "perfect"[30] (*i.e.* using PB radii) for neutral molecules does not significantly improve the GB model. By contrast, for charged molecules for which the self term is more dominant, using PB radii produces measurably better answers than using Born radii. The importance of errors between self and pair contributions is further emphasised by Fig. 7, which shows the frequency distribution for the percentage errors using different models for the pair term. These errors are defined on a scale from −100% to 100%. The unoptimised version of the Still formula (using PB radii) has a distribution with a mean of −7.2%, and a standard deviation (SD) of 5.2%, where the optimised (c = 0.19051) version has a mean of −0.7%, and an SD of 5.5%. Hence, carrying out a least squares optimisation reduces the mean towards zero, although slightly broadening the distribution. It should be recalled that using PB radii implies that the E^{self} term has been fixed, so the optimisation of c only acts upon E^{pair}. The reference results for GGB exhibit a mean of −0.2%, and an SD of 6.5%. In the case of optimisation of κ_m in the GGB method, the least squares procedure acts upon both E^{self} and E^{pair} in an even-handed manner. The optimisation surface as a function of κ_m shows a well-defined narrow minimum such that errors between self and pair terms are cancelled. The calculation of the GGB solvation energy has been implemented using the formulation outlined in Appendix B, which requires the evaluation of the Dawson function. For this, we have used the double precision implementation in the GNU Scientific Library.[41] Our algorithm is relatively unoptimised, but takes 57.8 seconds for 10 000 evaluations of the GGB solvation energy. The timings could be greatly reduced by using a table look-up approach for the Dawson function, rather than evaluating it afresh for each energy point. The calculation of energy gradients is relatively inexpensive in our approach, because evaluations of exponential and Dawson functions for the energy point can be reused. The only extra operations are algebraic manipulations. It is important to understand that all GB methods have inherent complications in the evaluation of gradients. The coordinate dependence of the solvation energy arises *explicitly* in the usual way, through the dependence of the pair energies E_{IJ} on pair distances R_{IJ}. It also depends implicitly on the coordinates through the Born radii. In principle, each Born radius depends upon the coordinates of every other atom which screens it from the solvent. In the GGB case this dependence is truncated by the finite range of the masking function, and this can be used to build neighbourhood lists which are very effective in reducing the amount of work required to compute the gradient. The crucial contribution of the GGB method in obtaining solvent forces is that there is a simple analytical expression for the derivatives of Born radii. For example, there is no need for piecewise terms with ranges which need to be tested prior to evaluation of gradient expressions. It is hopefully clear also that the GGB method can easily be added to existing efficient GB implementations of solvation energies and forces with relatively few modifications.

4. Conclusions

We have established a successful GB method with expressions for gradients which are smooth to all orders. This method is based upon the use of volume Gaussians to describe dielectric screening. It is competitive with other GB approaches in terms of numerical accuracy, but it has the advantange of requiring only a single parameter. The method uses a simple analytic representation of the Born radius in terms of easily computed functions, and leads to an analytic expression for the nuclear derivatives required in molecular modelling.

The breakdown of the solvation energy into "self" (atomic) and "pair" terms (2.1) is exact in PB and GB theories. An analysis of the cancellation of errors between pair and self terms demonstrates the extraordinary sensitivity of the total solvation energy to these errors. This confirms the general view that it is not possible to find better GB theories of solvation without radical change of the *ansatz* used for the pair function.

The success of the GGB model relies upon three separate factors. Firstly, the nature of the term used to cancel the $1/r_I^2$ singularity which is apparent in eqn (2.11) and (2.12) for the Born radius. This singularity does not occur in theories using discontinuous dielectrics, because regions around the nuclei are perfectly excluded, *i.e.* the integrations occur outside the solute cavity. In our case, where a smooth representation of the dielectric is used, it is necessary to use a function which cancels the singularity. Secondly, the smoothness of the dielectric switch ensures that the resulting expressions for Born radii are smooth, *i.e.* they are differentiable to all orders in coordinates with no discontinuities. Thirdly, the expressions which result from using a Gaussian mask, and Gaussian screening for solute atoms result in integrals which can be expressed in terms of known functions. The Gaussian masking function was introduced because of its mathematical advantages. It cannot be argued that it is a less physical model than one which treats an atom in a molecule as though it is a hard sphere. Continuum solvation models are practical models for describing the screening of electrostatic interactions in solvents. The switch between an atomistic representation of the solute and the continuum representation of the solvent is clearly a key issue in such theories. The replacement of a "sudden" discontinuous dielectric shift by a smooth one is a relatively small change in methodology compared to the inherent assumptions of the continuum solvation approach.

It is instructive to contrast our approach with other GB methods that have utilized Gaussian functions. The ACE procedure[42] uses Gaussians to represent the effective volumes (to allow for the effects of multiple volume intersections) of atoms in a manner similar to that used in eqn (2.17). However, ACE[42] replaces atom-centred point charges by Gaussian charge distributions which are constructed to reproduce the Born energy (2.2) for a Born ion. This adoption of non-zero size charges cleverly removes the Coulombic singularity. However, it leads to a complicated integral expression for the screening of the self energy (eqn (21) of the ACE reference) which cannot be solved analytically. As a consequence this expression needs to be replaced by an *ansatz* which correctly describes long and short range contributions to the screening of the self energy by a given solute atom. These terms contain parameters which add to the complexity of the model. In our case point charges are retained and the Born ion limit is achieved by using a masking Gaussian to remove the singularity.

The AGBNP[28] method is a complete solvation model in the sense that it includes electrostatic, cavitation and dispersion contributions. The electrostatic term is based upon GB theory and produces an inverse Born radius correction based upon pair screening of the self atom from solute by neighbouring solute atoms in a similar manner to our eqn (2.18). In the AGBNP case the calculation of the r^{-4} integral (our eqn (2.12)) is accomplished using hard-spheres to model the atoms, resulting in a piecewise expression (eqn (52)–(57) of ref. 26), with continuous but non-smooth first gradients and discontinuous second gradients. The resulting inverse Born radius is then corrected for overcounting of multiple intersections by using formulae derived from Gaussian volume intersections.[33] In contrast the GGB model uses a single equation to describe the screening of solvent by solute atoms. The resulting expressions can be calculated analytically (*i.e.* using only transcendental functions), and there are no discontinuities in derivatives at any order.

The role of the masking function $(1-e^{-\alpha r_I^2})$ in GGB theory is more subtle than just the cancellation of the Coulomb singularity. It acts also to scale the contributions to screening from the surrounding atoms. The value of the optimal masking parameter (κ_m = 0.057768) in the present work suggests that this scaling occurs over a width of approximately $4\sigma_I$. This is illustrated by looking at the schematic plot in Fig. 8. This shows the terms affecting the variation of the screen exclusion function ($\Theta_{s,I}$) in eqn (2.16) that contribute to the Born radius. The screen exclusion function is a product of the mask term with an exponent proportional to κ_m through eqn (2.14), and a set of volume Gaussians, g_J defined *via* eqn (2.17). For large values of κ_m the mask function rapidly approaches unity, and nearby volume Gaussians are not affected by the mask. Conversely, for small values of κ_m, the contribution of volume Gaussians to screening is reduced because the scaling function is less than unity. The GGB formulation in eqn (2.16) allows for volume intersections in screening atoms to be included to arbitrary order, but this gives only slightly improved overall results with a larger optimal κ_m (0.12). This larger value arises

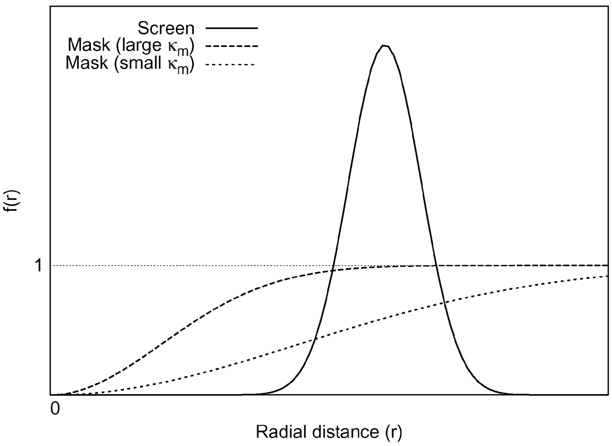

Fig. 8 Schematic diagram showing the components of the screening exclusion function in eqn (2.16) for a diatomic. The atom for which the Born radius is being computed is at the origin and a "screen atom" is at the peak of the Gaussian function. Masking functions with large and small values of κ_m are shown. The y-axis shows the relative values of the component functions.

because the rescaling effect now need occur to a smaller extent than before, because the higher intersections remove spurious volume intersections. The greater computational expense involved in including high order volume intersections leads us to suggest that for small molecules at least, they are not useful in the GGB model. It remains to be established if higher order intersections are significant for macromolecules, and this is something that we are currently investigating.

It is clear that one could design many other masking functions which would satisfy the mathematical requirements of cancelling the singularity inherent at the (Born) atomic centre I in eqn (2.11). It would be very hard, however, to propose a combination of masking and screening functions which lead to simple analytical expressions, such as those obtained in the GGB model.

Appendix A. Electrostatics in the presence of a varying permittivity

The aim of this Appendix is to derive the explicit expressions for the self and pair solvation energies, in the context of a Generalized Born (GB) model. The solute charge distribution is represented as a set of discrete charges Q_I at positions $\boldsymbol{R}_I$ thus:

$$\rho(\boldsymbol{r}_1) = \sum_I Q_I \delta(\boldsymbol{r}_{1I}), \qquad (1.1)$$

where the local relative coordinate $\boldsymbol{r}_{1I} = \boldsymbol{r}_1 = \boldsymbol{R}_I$. A key quantity in the theory of electric fields in materials is the displacement vector $\boldsymbol{D}(\boldsymbol{r}_1)$ which satisfies the Maxwell equation (in SI units)

$$\nabla_1 \cdot \boldsymbol{D}(\boldsymbol{r}_1) = \rho(\boldsymbol{r}_1) \qquad (1.2)$$

which is proportional to the electric field through[43,44]

$$\boldsymbol{D}(\boldsymbol{r}_1) = \varepsilon_0 \varepsilon(\boldsymbol{r}_1) \boldsymbol{E}(\boldsymbol{r}_1), \qquad (1.3)$$

where ε_0 is the vacuum permittivity. In atom-based solvation models $\varepsilon(\boldsymbol{r})$ is an isotropic relative permittivity function that takes values ε_{in} in regions near the atoms (*i.e.* in the solute), and values ε_{out} in the solvent. Substitution of (1.3) into (1.2) gives the familiar Poisson equation

$$\nabla_1 \cdot \varepsilon_0 \varepsilon(\boldsymbol{r}_1) \nabla_1 \phi(\boldsymbol{r}_1) = -\rho(\boldsymbol{r}_1) \qquad (1.4)$$

The polarization P produced in solvent is defined by the relation

$$\boldsymbol{D} = \varepsilon_{\text{in}} \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}. \qquad (1.5)$$

The equation for the polarization P is obtained from (1.3, 1.5) as

$$\begin{aligned} \boldsymbol{P}(\boldsymbol{r}_1) &= \varepsilon_0(\varepsilon(\boldsymbol{r}_1) - \varepsilon_{\text{in}}) \boldsymbol{E}(\boldsymbol{r}_1) \\ &= -\varepsilon_0(\varepsilon(\boldsymbol{r}_1) - \varepsilon_{\text{in}}) \nabla \phi(\boldsymbol{r}_1) \end{aligned} \qquad (1.6)$$

where $\phi(\boldsymbol{r}_1)$ is the "total" potential obtained from solving (1.4). This potential includes contributions from the source charge and that arising from dielectric polarization. The polarization produced at a point $\boldsymbol{r}_1$ gives rise to a dipole

$$\mathrm{d}\boldsymbol{p}_1 = \boldsymbol{P}(\boldsymbol{r}_1)\mathrm{d}\boldsymbol{r}_1. \qquad (1.7)$$

The potential at point $\boldsymbol{r}_2$ arising from this infinitesimal dipole is

$$\mathrm{d}\phi^{\text{pol}}(\boldsymbol{r}_2) = \frac{1}{4\pi\varepsilon_0\varepsilon_{\text{in}}} \mathrm{d}\boldsymbol{p}_1 \cdot \nabla_1 \frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}. \qquad (1.8)$$

Integrating this over all space gives

$$\begin{aligned} \phi^{\text{pol}}(\boldsymbol{r}_2) &= \frac{1}{4\pi\varepsilon_0\varepsilon_{\text{in}}} \int \mathrm{d}\boldsymbol{r}_1 \boldsymbol{P}(\boldsymbol{r}_1) \cdot \nabla_1 \left(\frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\right) \\ &= -\frac{1}{4\pi} \int \mathrm{d}\boldsymbol{r}_1 \left(\frac{\varepsilon(\boldsymbol{r}_1)}{\varepsilon_{\text{in}}} - 1\right) \nabla_1 \phi(\boldsymbol{r}_1) \cdot \frac{\hat{\boldsymbol{r}}_{12}}{r_{12}^2}. \end{aligned} \qquad (1.9)$$

The nature of (1.9) is that it would need to be solved iteratively, because the total potential ϕ on the right hand side contains a contribution from ϕ^{pol}. However in GB theory an approximation that avoids the requirement to obtain an iterative solution is to assume that the form of the displacement is, for a single charge Q_I located at $\boldsymbol{R}_I$:

$$\boldsymbol{D}(\boldsymbol{r}_1) = \frac{Q_I \hat{\boldsymbol{r}}_{1I}}{4\pi r_{1I}^2}, \qquad (1.10)$$

which implies that the term $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$ in (1.9) is replaced by

$$\boldsymbol{E}(\boldsymbol{r}_1) = \frac{Q_I}{4\pi\varepsilon_0\varepsilon(\boldsymbol{r}_1)} \frac{\hat{\boldsymbol{r}}_{1I}}{r_{1I}^2}. \qquad (1.11)$$

This approximation is known as the Coulomb-field approximation.[39,40] It is only exact for a narrow range of situations, the mathematics of which is elegantly explored in Kharkats *et al.*[39] Hence, labelling the polarization potential for the charge Q_I,

$$\phi^{\text{pol}}_{\text{GB},I}(\boldsymbol{r}_2) = -\frac{Q_I}{4\pi\varepsilon_0} \int \mathrm{d}\boldsymbol{r}_1 \frac{1}{4\pi} \Delta\varepsilon(\boldsymbol{r}_1) \frac{\hat{\boldsymbol{r}}_{1I}}{r_{1I}^2} \cdot \frac{\hat{\boldsymbol{r}}_{12}}{r_{12}^2}, \qquad (1.12)$$

where

$$\Delta\varepsilon(\boldsymbol{r}_1) = \left(\frac{1}{\varepsilon_{\text{in}}} - \frac{1}{\varepsilon(\boldsymbol{r}_1)}\right). \qquad (1.13)$$

The potential arising from the polarization coming from *all* of the charges is just the sum

$$\phi^{\rm pol}_{\rm GB}(\boldsymbol{r}_2) = \sum_I \phi^{\rm pol}_{{\rm GB},I}(\boldsymbol{r}_2). \tag{1.14}$$

The solvation energy $E^{\rm solv}$ in the GGB model is the interaction between the fixed charges Q_I and the polarization potential, and is given by

$$\begin{aligned} E^{\rm solv} &= \frac{1}{2}\int {\rm d}\boldsymbol{r}_2 \phi^{\rm pol}_{\rm GB}(\boldsymbol{r}_2)\rho(\boldsymbol{r}_2) \\ &= \frac{1}{2}\sum_J Q_J \phi^{\rm pol}_{\rm GB}(\boldsymbol{R}_J) \\ &= \frac{1}{2}\sum_{IJ} Q_J \phi^{\rm pol}_{{\rm GB},I}(\boldsymbol{R}_J). \end{aligned} \tag{1.15}$$

The solvation energy can thus be split into 'self' and 'pair' terms as in eqn (2.1). The self terms are

$$\begin{aligned} E^{\rm self}_I &= \frac{1}{2} Q_I \phi^{\rm pol}_{{\rm GB},I}(\boldsymbol{R}_I) \\ &= -\frac{Q_I^2}{8\pi\varepsilon_0}\int {\rm d}\boldsymbol{r} \tfrac{1}{r_I^4}\tfrac{1}{4\pi}\Delta\varepsilon(\boldsymbol{r}), \end{aligned} \tag{1.16}$$

and the pair terms are

$$\begin{aligned} E^{\rm pair}_{IJ} &= \frac{1}{2}\left(Q_I \phi^{\rm pol}_{{\rm GB},J}(\boldsymbol{R}_I) + Q_J \phi^{pol}_{{\rm GB},J}(\boldsymbol{R}_J)\right) \\ &= -\frac{Q_I Q_J}{4\pi\varepsilon_0}\int {\rm d}\boldsymbol{r} \tfrac{\hat{\boldsymbol{r}}_I}{r_I^2}\cdot\tfrac{\hat{\boldsymbol{r}}_J}{r_J^2}\tfrac{1}{4\pi}\Delta\varepsilon(\boldsymbol{r}), \end{aligned} \tag{1.17}$$

where we have used the reciprocity

$$Q_J \phi^{\rm pol}_{{\rm GB},I}(\boldsymbol{R}_J) = Q_I \phi^{\rm pol}_{{\rm GB},J}(\boldsymbol{R}_I), \tag{1.18}$$

to simplify the relationship in eqn (1.17). Eqn (1.16) and (1.17) are exact within the approximation given in (1.10). The assumption of linearity is what implies the existence of the pairwise summation formula for the solvation energy (2.1). It is also responsible for the reciprocity relation (1.18) noted above.

Appendix B. Expressions for Gaussian Born radii

The contribution to the Born radius from the mask term from eqn (2.11) is

$$\frac{1}{\sigma^B_{m,I}} = \frac{p_m}{4\pi}\int {\rm d}\boldsymbol{r} \tfrac{1}{r_I^4}\left(1 - e^{-\alpha r_I^2}\right), \tag{2.1}$$

where we have used eqn (2.13) to define the mask part of the exclusion function. It is convenient to use an identity

$$\int_0^\alpha {\rm d}a e^{-ar_I^2} = \frac{1}{r_I^2}\left(1 - e^{-\alpha r_I^2}\right), \tag{2.2}$$

in subsequent equations. We can derive an analytic expression from (2.1) by noting that the integrand is spherically symmetric about the local radial coordinate r_I. Integrating over angles gives

$$\begin{aligned} \frac{1}{\sigma^B_{m,I}} &= p_m \int_0^\alpha {\rm d}a \int_0^\infty {\rm d}r_I e^{-ar_I^2} = \frac{p_m}{2}\int_0^\alpha {\rm d}a\left(\frac{\pi}{a}\right)^{\frac{1}{2}} \\ &= p_m\sqrt{\pi\alpha} \\ &= \frac{p_m}{\sigma_I}\sqrt{\pi\kappa_m}. \end{aligned} \tag{2.3}$$

In order to achieve a "zeroth order" model which agrees with the Born atom result of eqn (2.2), we require the mask height p_m to be set as in eqn (2.15). Considering a single screen atom J eqn (2.16) becomes

$$\Theta_{s,I} = p(1 - e^{-\alpha r_I^2})e^{-\beta r_J^2}, \tag{2.4}$$

where the exponent

$$\beta = \frac{\kappa}{\sigma_J^2}, \tag{2.5}$$

as in eqn (2.17). This leads to

$$\begin{aligned} \frac{1}{\sigma^B_{s,I}} &= -\frac{p}{4\pi}\int {\rm d}\boldsymbol{r}\frac{1}{r_I^4}(1 - e^{-\alpha r_I^2})e^{-\beta r_J^2} \\ &= -\frac{p}{4\pi}\int_0^\alpha {\rm d}a \int {\rm d}\boldsymbol{r} \tfrac{1}{r_I^2} e^{-ar_I^2} e^{-\beta r_J^2}, \end{aligned} \tag{2.6}$$

where we have used the relationship in eqn (2.2). We can use the cosine rule to expand the radial factor $r_J{}^2$ in the screening Gaussian, to give

$$r_J^2 = r_I^2 + R_{IJ}^2 - 2r_I R_{IJ} c, \tag{2.7}$$

where we have substituted $c = \cos\theta$ for the polar angle between the vectors $\boldsymbol{r}_I$ and $\boldsymbol{R}_{IJ}$. We can now integrate over spherical polar angles, since the expressions arising from eqn (2.6) are independent of the axial angle ϕ, and the integral over $c = \cos\theta$ is trivial. We obtain

$$\frac{1}{\sigma^B_{s,I}} = -\frac{pe^{\beta R_{IJ}}}{2\beta R_{IJ}}\int_0^\alpha {\rm d}a \int_0^\infty {\rm d}re^{-(a+\beta)r^2}\sinh(2\beta R_{IJ} r). \tag{2.8}$$

This can be integrated[45] to give

$$\frac{1}{\sigma^B_{s,I}} = -\left(\frac{\pi p}{4\beta R_{IJ}}\right)e^{-\beta R_{IJ}^2}\int_0^\alpha {\rm d}a\,{\rm erfi}\frac{\beta R_{IJ}}{\sqrt{a+\beta}}, \tag{2.9}$$

where the imaginary error function

$$ {\rm erfi}\,(z) = i^{-1}{\rm erf}\,(iz). \tag{2.10}$$

We use the change of variable below to simplify the argument of the error function within the integral sign:

$$\begin{aligned} t = \frac{\beta R_{IJ}}{\sqrt{\alpha+\beta}} &\Rightarrow {\rm d}t = -\frac{1}{2}\frac{\beta R_{IJ}}{(\alpha+\beta)^{\frac{3}{2}}}{\rm d}\alpha \\ &\Rightarrow {\rm d}\alpha = -2\,(\beta R_{IJ})^2\frac{{\rm d}t}{t^3}. \end{aligned} \tag{2.11}$$

The relation in (2.11) can be substituted into (2.9) to give us the new expression

$$\frac{1}{\sigma^B_{s,I}} = \frac{\pi\beta R_{IJ} p}{2}e^{-\beta R_{IJ}^2}\int_{\sqrt{\beta R_{IJ}}}^{\frac{\beta R_{IJ}}{\sqrt{\alpha+\beta}}}\frac{{\rm erfi}\,(t)}{t^3}{\rm d}t. \tag{2.12}$$

The integral appearing in (2.12) can be simplified using partial integration to obtain

$$\begin{aligned}\int \frac{\mathrm{erfi}(t)}{t^3}\mathrm{d}t &= \left(1-\frac{1}{2t^2}\right)\mathrm{erfi}\,(t) - \frac{e^{t^2}}{\sqrt{\pi}t}\\ &= 2\frac{e^{t^2}}{\sqrt{\pi}t}\left[\left(t-\frac{1}{2t}\right)\mathrm{Daw}\,(t)-\frac{1}{2}\right],\end{aligned} \qquad (2.13)$$

where the Dawson integral is defined as

$$\mathrm{Daw}\,(t) = e^{-t^2}\int_0^t e^{x^2}\mathrm{d}x = \frac{i\sqrt{\pi}}{2}e^{-t^2}\mathrm{erfc}\,(-it). \qquad (2.14)$$

It is useful to define a function

$$\mathrm{N}_0(t) = \left(t-\frac{1}{2t}\right)\mathrm{Daw}\,(t) - \frac{1}{2}, \qquad (2.15)$$

which allows us to write eqn (2.12) in its final form:

$$\frac{1}{\sigma^B_{s,I}} = p\left[K_{IJ}\sqrt{\pi(\alpha+\beta)}\mathrm{N}_0\left(\frac{\beta R_{IJ}}{\sqrt{\alpha+\beta}}\right)\sqrt{\pi\beta}N_0\left(\sqrt{\beta}R_{IJ}\right)\right], \qquad (2.16)$$

where

$$K_{IJ} = e^{-\alpha\beta R_{IJ}^2/(\alpha+\beta)}. \qquad (2.17)$$

Eqn (2.16) is used to sum over screening contributions arising from each Gaussian in the summation coming from eqn (2.16). It is straightforward to apply this equation to Gaussians representing any intersection order. Eqn (2.16) contains explicit functions of the co-ordinates of atoms, through their dependence upon the internuclear distance, R_{IJ}. It follows that the Born radii, σ^B_I, depend explicitly upon the co-ordinates of all the screening atoms. These expressions may be directly, and simply, differentiated with respect to nuclear co-ordinates, to give 'analytic' expressions for the dependence of the GGB solvation energy upon nuclear co-ordinates. This can be accomplished by defining the derivatives of the function $N_0(x)$ using the definition of the Dawson function. Hence,

$$\begin{aligned}N_1(x) &= \frac{\mathrm{d}N_0(x)}{\mathrm{d}x}\\ &= \left(\frac{1}{2x^2}-2x^2+2\right)\mathrm{Daw}\,(x) - \frac{1}{2x} + x,\end{aligned} \qquad (2.18)$$

and similarly,

$$\begin{aligned}N_2(x) &= \frac{\mathrm{d}N_1(x)}{\mathrm{d}x}\\ &= \left(-\frac{1}{x^3}+4x^3-\frac{1}{x}-8x\right)\mathrm{Daw}\,(x)\\ &\quad + 3 + \frac{1}{x^2} - 2x^2.\end{aligned} \qquad (2.19)$$

The functions defined above are depicted in Fig. 9. They are well behaved, even or odd functions, and contain no singularities as $x \to 0$. The behaviour of $N_0(x)$ at the origin is

$$N_0(x) = -1 + \frac{4}{3}x^2 - \frac{4}{5}x^4 + O(x^5).$$

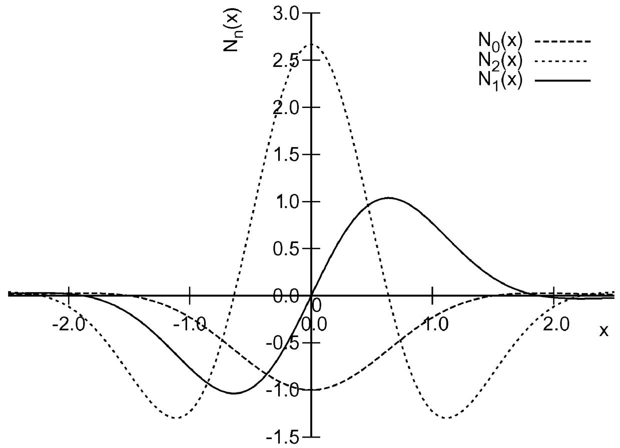

Fig. 9 The function $N_0(x)$ and its first and second derivatives $N_1(x)$ and $N_2(x)$, plotted as a function of x.

With these definitions, the equations for nuclear gradients can be derived using standard differentiation formulae.

References

1 T. S. Rush, J. A. Grant, L. Mosyak and A. Nicholls, *J. Med. Chem.*, 2005, **48**, 1489–1495.
2 S. W. Muchmore, A. J. Souers and I. Akritopoulou-Zanze, *Chem. Biol. Drug Design*, 2006, **67**, 174–176.
3 J. Warwicker and H. C. Watson, *J. Mol. Biol.*, 1982, **157**, 671–679.
4 B. Honig and A. Nicholls, *Science*, 1995, **268**, 1144–1149.
5 A. Warshel and A. Papazyan, *Curr. Opin. Struct. Biol.*, 1998, **8**, 211–217.
6 N. A. Baker and J. A. McCammon, *Methods Biochem. Anal.*, 2003, **44**, 427–440.
7 N. A. Baker, *Curr. Opin. Struct. Biol.*, 2005, **15**, 137–143.
8 P. Koehl, *Curr. Opin. Struct. Biol.*, 2006, **16**, 1–10.
9 K. Moritsugu and A. Kidera, *J. Phys. Chem. B*, 2004, **108**, 3890–3898.
10 P. Carnevali, G. Tóth, G. Toubassi and S. Meshkat, *J. Am. Chem. Soc.*, 2003, **125**, 14244–14245.
11 V. Alexandrov, U. Lehnert, N. Echols, D. Milburn, D. Engelman and M. Gerstein, *Protein Sci.*, 2005, **14**, 633–643.
12 Z. I. Hodes, G. Némethy and H. A. Scheraga, *Biopolymers*, 1979, **18**, 1565–1610.
13 D. Eisenberg and A. D. McLachlan, *Nature*, 1986, **319**, 199–203.
14 J. Vila, R. L. Williams, M. Vásquez and H. A. Scheraga, *Protein Struct. Funct. Genet.*, 1991, **10**, 199–218.
15 Y. Kang, G. Némethy and H. A. Scheraga, *J. Phys. Chem.*, 1987, **91**, 4105–4109.
16 Y. Kang, G. Nemethy and H. A. Scheraga, *J. Phys. Chem.*, 1987, **91**, 4109–4117.
17 W. C. Still, A. Tempczyk, R. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6128.
18 D. Qiu, P. S. Shenkin, F. P. Hollinger and W. C. Still, *J. Phys. Chem. A*, 1997, **101**, 3005–3014.
19 A. Ghosh, C. S. Rapp and R. A. Friesner, *J. Phys. Chem. B*, 1998, **102**, 10983–10990.
20 M. Schaefer and M. Karplus, *J. Phys. Chem.*, 1996, **100**, 1578–1599.
21 W. Im, D. Beglov and B. Roux, *Comput. Phys. Commun.*, 1998, **11**, 59–75.
22 J. Grant, B. Pickup and A. Nicholls, *J. Comput. Chem.*, 2001, **22**, 608–640.
23 M. E. Davis and J. A. McCammon, *J. Comput. Chem.*, 1991, **12**, 909–912.
24 M. S. Lee, F. R. Salsbury and C. L. Brooks, *J. Chem. Phys.*, 2002, **116**, 10606–10614.
25 W. Im, M. S. Lee and C. L. Brooks, *J. Comput. Chem.*, 2003, **24**, 1691–1702.

26 M. S. Lee, M. Feig, F. R. Salsbury Jr. and C. L. Brooks, *J. Comput. Chem.*, 2000, **24**, 1348–1356.
27 M. Feig, A. Onufriev, M. Lee, W. Im, D. Case and C. L. Brooks, *J. Comput. Chem.*, 2003, **25**, 265–284.
28 E. Gallicchio and R. M. Levy, *J. Comput. Chem.*, 2003, **25**, 479–499.
29 Z. Yu, M. P. Jacobson and R. A. Friesner, *J. Comput. Chem.*, 2005, **27**, 72–89.
30 A. Onufriev, D. Case and D. Bashford, *J. Comput. Chem.*, 2002, **23**, 1297–1304.
31 I. Shavitt, *Methods Comput. Phys.*, 1962, **2**, 1–45.
32 S. F. Boys, *Proc. R. Soc. London, Ser. A*, 1950, **200**, 542–554.
33 J. A. Grant and B. T. Pickup, *J. Phys. Chem.*, 1995, **99**, 3503–3510.
34 J. A. Grant, M. A. Gallardo and B. T. Pickup, *J. Comput. Chem.*, 1996, **17**, 1653–1666.
35 J. A. Grant and B. T. Pickup, *KLUWER/ESCOM*, Kluwer Scientific Publishers, Dordrecht, 1998, vol. 3, ch. Gaussian Shape Methods, pp. 150–176.
36 MDDR, *MDL Information System*, San Leandro, CA, USA, 2004.
37 J. Gasteiger, C. Rudolph and J. Sadowski, *Tetrahedron Comput. Methodol.*, 1992, **3**, 537–547.
38 F. Mohamadi, N. Richards, W. Guida, R. Liscamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson and W. Still, *J. Comput. Chem.*, 1990, **11**, 440–666.
39 Y. I. Kharkats, A. A. Kornyshev and M. A. Vorotyntsev, *J. Chem. Soc., Faraday Trans. 2*, 1976, **72**, 361–371.
40 T. Grycuk, *J. Chem. Phys.*, 2003, **119**, 4817–4826.
41 M. Galassi, GNU Scientific Library reference manual, Technical report, GNU Software Foundation, 2006.
42 M. Schaefer and M. Karplus, *J. Phys. Chem.*, 1996, **100**, 1578–1599.
43 J. D. Jackson, *Classical Electrodynamics*, John Wiley and Sons, New York, 1975.
44 J. R. Reitz and F. J. Milford, *Foundations of Electromagnetic Theory*, Addison Wesley, New York, 1967.
45 I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, Academic Press, New York, 6th edn, 2000.