Theor Chem Acc (1999) 101:396–408
DOI 10.1007/s00214990m184

Theoretical
Chemistry Accounts

Regular article

An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package

Cory C. Pye, Tom Ziegler

Department of Chemistry, The University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada

Received: 13 November 1998 / Accepted: 16 December 1998 / Published online: 16 March 1999

Abstract. The conductor-like screening model (COSMO) of solvation has been implemented in the Amsterdam density functional program with maximum flexibility in mind. Four cavity definitions have been incorporated. Several iterative schemes have been tested for solving the COSMO equations. The biconjugate gradient method proves to be both robust and memory-conserving. The interaction between the surface charges and the electron density may be calculated by integrating over either the fitted or exact density, or by calculating the molecular potential. A disk-smearing algorithm is applied in the former case to avoid singularities. Several self-consistent field/COSMO coupling schemes were examined in an attempt to reduce computational effort. A gradient-preserving algorithm for removing outlying charge has been implemented. Preliminary optimized radii are given. Applications to the benzene oxide-oxepin valence tautomerization and to glycine conformation are presented.

Key words: Solvation – Conductor-like screening model – Amsterdam density functional – Oxepin – Glycine

1 Introduction

The modeling of solvent effects in complex systems of technological, geochemical and biological significance is an important field of study. In the past several years, the conductor-like screening model (COSMO) of Klamt and Schüürmann [1] has emerged as a promising tool for efficiently treating these effects. The COSMO idea, initially implemented in the MOPAC semiempirical package, has subsequently been incorporated into DMol [2], Gaussian 92 [3], GAMESS [4], Gaussian 94 [5], MNDO/d [6], and PAW. We present here our implementation [7] into the Amsterdam density functional package (ADF) [8]. This COSMO implementation will be included in future releases of the ADF package (Scientific Computing and Modeling, Chemistry Department, Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands).

Correspondence to: C.C. Pye

2 Theory and implementation

2.1 Dielectric continuum model of solvation

A molecule in solution may be regarded as being situated in a cavity inside a dielectric continuum. The charge distribution of the molecule, which consists of the distribution due to the nuclear charges Z_A at R_A, $\sum_A Z_A \delta(r - R_A)$, as well as the electronic density $-\rho(r)$, will induce a charge density on the surface of the cavity, $\rho_s(r_s)$, where r_s is a point on the surface. The total charge induced will be opposite to the charge inside the cavity, which should be approximately equal to the total solute charge.

The electrostatic energy, formed from the interaction between the solute and the surface charge and the self-interaction of the surface charge, is given by

$$E^S = \int_S \sum_A \frac{Z_A \rho_s(r_s)}{|R_A - r_s|} dr_s + \int_S \int_S \frac{\rho_s(r_s)\rho_s(r'_s)}{|r_s - r'_s|} dr_s dr'_s + \int_V \int_S \frac{\rho(r)\rho_s(r_s)}{|r - r_s|} \quad . \tag{1}$$

The three terms of Eq. (1) represent the interaction of the surface charges with the nuclear charge, the surface charges (self-interaction), and the electron density, respectively. In practice, the surface charge distribution is approximated by a set of discrete point charges.

In addition to the electrostatic contribution, there is also a nonelectrostatic contribution arising from cavitation, dispersion and repulsion. These are usually modeled as functions of the surface area of the cavity.

2.2 Discretization of the cavity surface

In our implementation, the cavity may be constructed as either van der Waals (vdW), solvent-accessible (SA), solvent-excluding (SE) or Klamt [1] surfaces. The

GEPOL 93 program [9] is used to generate the first three surfaces. We have slightly modified the GEPOL program in order to generate the Klamt surface by a double application of the SA algorithm, first adding the solvent radius, which annihilates points near the cusp region, and then subtracting the solvent radius, which projects the surviving points on the original spheres.

The surface of each atom is discretized into 60 triangles. After application of the GEPOL algorithm, the surviving triangles will possess a center r_μ^{B}, surface area S_μ^{B} and charge q_μ^{B}, and will be associated with a parent atom B. This yields, for the solvation energy,

$$E^{\mathrm{S}} = \sum_{\mathrm{A}}\sum_{\mu}\frac{Z_{\mathrm{A}}q_\mu}{|R_{\mathrm{A}}-r_\mu|} + \frac{1}{2}\sum_{\nu\neq\mu}\sum_{\mu}\frac{q_\nu q_\mu}{|r_\mu - r_\nu|} + \sum_{\mu} 1.07\sqrt{\frac{4\pi}{S_\mu}}q_u^2 + \sum_{\mu}\int_V q_\mu V_\mu(r)\rho(r)dr \ . \tag{2}$$

In Eq. 2 a point-charge approximation is made for the first two terms, representing the nuclear-surface and surface-surface interactions. The third term represents the electrostatic interaction within each segment [1]. The final term represents the interaction between the electronic density and the surface, where the surface charge has been distributed with a potential $q_\mu V_\mu(r)$. We have not specified the potential at this point for the the final term, representing the electron density-surface interactions. The electrostatic screening energy may then be written

$$E^{\mathrm{S}} = \frac{1}{2}\sum_{\mu}\sum_{\nu} q_\mu A_{\mu\nu} q_\nu + \sum_{\mathrm{A}}\sum_{\mu} q_\mu B_{\mathrm{A}\mu} Z_{\mathrm{A}} + \sum_{\mu} q_\mu C_\mu \tag{3}$$

where $A_{\mu\nu} = |r_\mu - r_\nu|^{-1} (\mu \neq \nu)$, $A_{\mu\mu} = 1.07\sqrt{\frac{4\pi}{S_\mu}}$, $B_{\mathrm{A}\mu} = |r_\mu - r_{\mathrm{A}}|^{-1}$ and $C_\mu = \int_V V_\mu(r)\rho(r)dr$.

2.3 Solution of the COSMO equations

In COSMO, the surface charges vary in such a way as to equalize the electrostatic potential over the cavity surface, (or alternatively, by minimizing E_{s}), thus

$$\frac{dE^{\mathrm{s}}}{dq_\mu} = \sum_{\mu} A_{\mu\nu} q_\nu + \sum_{\mathrm{A}} B_{\mathrm{A}\mu} Z_{\mathrm{A}} + C_\mu = 0 \ , \tag{4}$$

or in matrix form, as

$$Aq = -(BZ + C) \tag{5}$$

from which q can be obtained as

$$q = -A^{-1}(BZ + C) \ . \tag{6}$$

One can use Gauss–Jordan techniques [10] to solve the linear systems, but since this equation must be solved at each self-consistent-field (SCF) cycle (same A), it is more efficient to invert the matrix and store A^{-1} externally; however, for large systems, the storage of this matrix in memory poses a significant challenge during the inversion process and one must resort to other means. Our initial attempt used a Gauss–Seidel iterative scheme [10], but several cases were encountered in which this method did not converge. We also implemented the Jacobi technique, and modified both of these techniques to allow for relaxation (both under- and overrelaxation, to allow for robustness and alacrity, respectively) [10]; however, in some cases, the charges diverged to infinity, with the solvation energy decreasing without bound. The lack of convergence stems from the lack of positive definiteness of the A matrix in certain cases, which causes these techniques to fail. Our solution to this problem was to implement the preconditioned minimum-residual biconjugate gradient technique [11] which allows for an iterative solution for these problems. The biconjugate gradient method is related to the conjugate gradient method of minimizing a function $f(x) = \frac{1}{2}xAx - bx$ but has the advantage that a positive definite matrix is not required. Four methods are thus available for solving these equations, with the biconjugate gradient method as the default.

At this point, we note that, if q satisfies Eq. (6), then the energy expression reduces to

$$E^{\mathrm{S}} = -\frac{1}{2}\sum_{\mu}\sum_{\nu} q_\mu A_{\mu\nu} q_\nu = \frac{1}{2}\left(\sum_{\mathrm{A}}\sum_{\mu} q_\mu B_{\mathrm{A}\mu} Z_{\mathrm{A}} + \sum_{\mu} q_\mu C_\mu\right) \ . \tag{7}$$

The COSMO equation is true only for a conductor. For real solvents of dielectric constant ϵ, the charges (or alternatively, the energies and potentials) are scaled by

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + x} \ , \tag{8}$$

where x is an empirical parameter. In Klamt's implementations [1, 2, 4, 12, 13], the value of 0.5 is chosen, whereas other authors [3, 5] choose 0.0. For the high-dielectric region, this choice makes no difference. We let the user make this choice for maximum flexibility and ease of comparison with published work, but set the default to 0.0 to satisfy Gauss' law.

2.4. Total energy expression and the Kohn-Sham equations

The total Kohn–Sham (KS) energy, E^{T}, for the combined system of solute and induced surface charges is obtained by adding E^{s} (Eq. 3) to the gas-phase KS energy expression

$$\begin{aligned} E^0 = &-\frac{1}{2}\sum_{i}^{n}\int \psi_i^*(1)\nabla^2\psi_i(1)d\tau_1 \\ &+\frac{1}{2}\sum_{\mathrm{A}}^{N}\sum_{\mathrm{B}\neq\mathrm{A}}^{N}\frac{Z_{\mathrm{A}}Z_{\mathrm{B}}}{|R_{\mathrm{A}}-R_{\mathrm{B}}|} \\ &-\sum_{i}^{n}\int \psi_i^*\left[\sum_{\mathrm{A}}^{N}\frac{Z_{\mathrm{A}}}{|R_{\mathrm{A}}-r_1|}\right]\psi_i(1)d\tau_1 \\ &+\frac{1}{2}\sum_{i}^{n}\sum_{j}^{n}\int\int \psi_i^*(1)\psi_j^*(1)\frac{1}{|r_1-r_2|} \\ &\times\psi_i(2)\psi_j(2)d\tau_2 d\tau_1 + E_{xc} \ , \end{aligned} \tag{9}$$

where $\{\psi_i(1); i = 1, n\}$ are the one-electron spin-orbitals, and the electron density ρ is written as

$$\rho(1) = \sum_{j=1}^{n} \psi_j^*(1)\psi_j(1) \ . \tag{10}$$

The first term in Eq. (9) represents the kinetic energy of a noninteracting model system with the same density as the real interacting system. The second term represents the nuclear-nuclear repulsion term, while the third term gives the nuclear-electron attraction energy. The fourth term, also known as the Hartree term, describes the interaction of the electron density with itself. The final term, the exchange correlation energy, contains all additional terms, including the difference between the kinetic energies of the model and real system. The KS orbitals are determined to minimize the total energy, which now includes the solvation term. This condition is obeyed if the functions $\psi_i(1)$ satisfy the one-electron KS equations

$$h(1)\psi_i(1) = \varepsilon_i\psi_i(1) \tag{11}$$

under the orthonormality constraint

$$\int \psi_i^*(1)\psi_j(1)d\tau_1 = \delta_{ij} \ , \tag{12}$$

where the one-electron KS operator $h(1)$ is given by

$$h(1) = T + \sum_{\mathrm{A}}^{N} V_{\mathrm{N}}^{\mathrm{A}} + \sum_{\mu}^{M} q_\mu V_\mu(1) + V_{\mathrm{C}}(1) + V_{\mathrm{xc}} \ . \tag{13}$$

In Eq. (13),

$$T = -\frac{1}{2}\nabla^2 \tag{14}$$

is the kinetic energy operator,

$$V_{\mathrm{N}}^{\mathrm{A}}(1) = -\frac{Z_{\mathrm{A}}}{|R_{\mathrm{A}} - r_1|} \tag{15}$$

is the nuclear-electron attraction operator due to nucleus A,

$$V_{\mathrm{C}}(1) = \sum_{j}^{n} \int \psi_j^*(2)\psi_j(2)\frac{1}{|r_1 - r_2|}d\tau_2 \tag{16}$$

is the electrostatic potential from the total electronic density ρ, and

$$V_{\mathrm{xc}} = \frac{\partial E_{\mathrm{xc}}}{\partial \rho} \tag{17}$$

is the exchange-correlation potential.

To simplify the calculation of V_{C}, one introduces the fitted density

$$\tilde{\rho}(2) = \sum_{a} c_a f_a(r_2) \tag{18}$$

into Eq. (16) to yield the fitted potential. The f_a are single-center Slater functions and the c_a are determined by a least-squares fitting procedure.

The operator differs from the gas-phase operator only in the added external potential $\sum_\mu q_\mu V_\mu(1)$ due to the induced surface charges, where $V_\mu(1) = \frac{1}{|r_\mu - r|}$.

2.5 Implementation of the KS equations in the ADF program

The KS orbitals of Eq. (11) are expressed in the ADF program as linear combinations of primitive Slater-type spin-orbitals (STO) $\{\chi_\mu(1); \mu = 1, 2M_{\mathrm{T}}\}$ as

$$\psi_i(1) = \sum_{\mu=1}^{2M_{\mathrm{T}}} d_{\mu i}\lambda_\mu(1) \ , \tag{19}$$

where the primitive STOs are given as

$$\lambda_\mu(1) = (x_1 - X_{\mathrm{A}})^{l_x}(y_1 - Y_{\mathrm{A}})^{l_y}(z_1 - Z_{\mathrm{A}})^{l_z} \times (|r_1 - R_{\mathrm{A}}|)^{l_r}\exp[-\alpha_\mu|r_1 - R_{\mathrm{A}}|]\gamma(\sigma_1) \ . \tag{20}$$

With the expansion of Eq. (19), the set of one-electron KS equations takes the form

$$\sum_{\gamma=1}^{2M_v}\Big[F_{\tau\gamma} - \varepsilon_i(S_{(L)})_{\tau\gamma}\Big]C_{\gamma i} = 0, \tau = 1, 2M_v \tag{21}$$

with the orthonormality condition becoming

$$\sum_{\gamma=1}^{2M_v}\sum_{\tau=1}^{2M_v} C_{\tau i}^* C_{\gamma j}(S_{(L)})_{\tau\gamma} = \delta_{ij}, \tag{22}$$

where the KS integral, $F_{\tau\gamma}$, and the overlap integral, $(S_{(L)})_{\tau\gamma}$, are defined as

$$F_{\tau\gamma} = \int \lambda_\tau^*(1)h(1)\lambda_{\gamma(1)}d\tau_1 \tag{23}$$

$$(S_{(L)})_{\tau\gamma} = \int \lambda_\tau^*(1)\lambda_{\gamma(1)}d\tau_1 \tag{24}$$

The matrix elements $F_{\tau\gamma}$ are calculated in the ADF program by numerical integration as

$$F_{\tau\gamma} = \sum_{k=1}^{N_{\mathrm{S}}} W(r_k)\lambda_\tau(r_k)\big[h^0(r_k) + h^s(r_k)\big]\lambda_\gamma(r_k), \tag{25}$$

where $W(r_k)$ is a weight associated with each integration point r_k. The matrix $S_{(L)}$ is calculated by analytical integration. We note again the small modification of the KS operator by the surface charge potential.

The results for a suite of small molecules, taken primarily from Ref. [12], using the vdW, SE and Klamt surfaces are given in Tables 1–4. Because of the coarse discretization used, the sets of triangles corresponding to the vdW and SE surfaces are often identical, and would give exactly the same solvation energy. The vdW and SE surfaces begin to differ for larger systems, with the SE surface possessing a smaller solvation energy. The Klamt surface always has the smallest solvation energy. We prefer to use the SE surface because it possesses neither cusps nor holes. Klamt has proposed a modification to his original surface which also fills in these holes [13].

2.6 Numerical integration grid and surface charges

The potential V_μ due to the surface charges is needed at each numerical integration point in order to evaluate the

Table 1. Monatomic and diatomic test cases. Results obtained using van der Waals (*vdW*), solvent-excluding (*SE*) and klamt surfaces

	$\langle\rho\|H^g\rangle$ (a.u.)				Electrostatic			Destabilization			Total electronic		
	Gas	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt
Monatomics													
H^-	−0.035871	−0.031909	−0.031909	−0.031909	−89.47	−89.47	−89.47	0.59	0.59	0.59	−88.88	−88.88	−88.88
F^-	−0.125772	−0.125658	−0.125658	−0.125658	−112.31	−112.31	−112.31	0.07	0.07	0.07	−112.24	−112.24	−112.24
Cl^-	−0.139652	−0.138712	−0.138712	−0.138712	−98.38	−98.38	−98.38	2.49	2.49	2.49	−95.89	−95.89	−95.89
Diatomics													
CN^-	−0.639786	−0.639491	−0.639491	−0.639522	−76.53	−76.53	−75.13	0.18	0.18	0.17	−76.35	−76.35	−74.97
N_2	−0.638751	−0.638699	−0.638699	−0.638715	−0.96	−0.96	−0.76	0.03	0.03	0.02	−0.93	−0.93	−0.74
OH^-	−0.339214	−0.335016	−0.335016	−0.334678	−106.11	−106.11	−106.91	2.63	2.63	2.85	−103.48	−103.48	−104.06
CO	−0.572585	−0.572463	−0.572463	−0.572480	−1.54	−1.54	−1.23	0.08	0.08	0.07	−1.46	−1.46	−1.17
NO^+	−0.112181	−0.111886	−0.111886	−0.112053	−99.40	−99.40	−96.43	0.18	0.18	0.08	−99.22	−99.22	−96.35
$O_2(^1\Sigma)$	−0.341613	−0.341605	−0.341605	−0.341610	−0.97	−0.97	−0.49	0.00	0.00	0.00	−0.97	−0.97	−0.49
$O_2(^3\Pi)$	−0.380298	−0.380292	−0.380292	−0.380296	−0.94	−0.94	−0.48	0.00	0.00	0.00	−0.94	−0.94	−0.48
HF	−0.299772	−0.297973	−0.297973	−0.298114	−8.44	−8.44	−8.88	1.13	1.13	1.04	−7.31	−7.31	−7.84
F_2	−0.144116	−0.144088	−0.144081	−0.144075	−1.30	−1.27	−1.03	0.02	0.02	0.03	−1.28	−1.25	−1.01
SH^-	−0.299986	−0.297338	−0.297338	−0.297005	−79.99	−79.99	−80.21	1.66	1.66	1.87	−78.32	−78.33	−78.33
CS	−0.379939	−0.376848	−0.376854	−0.376869	−7.37	−7.27	−7.03	1.94	1.94	1.93	−5.43	−5.33	−5.10
HCl	−0.225486	−0.223887	−0.223887	−0.224145	−6.70	−6.70	−6.36	1.00	1.00	0.84	−5.70	−5.70	−5.52
OCl^-	−0.288527	−0.284082	−0.284147	−0.284896	−83.14	−82.98	−80.28	2.79	2.74	2.28	−80.35	−80.23	−78.01
FCl	−0.159222	−0.158613	−0.158574	−0.158549	−3.73	−3.77	−3.48	0.38	0.41	0.42	−3.35	−3.36	−3.05
Cl_2	−0.130156	−0.129949	−0.129891	−0.129872	−3.55	−3.64	−3.22	0.13	0.17	0.18	−3.42	−3.47	−3.05

Local density approximation (LDA) double-zeta polarized (DZP), $\int$ accuracy = 10^{-6}. All energies in kcal/mol unless otherwise specified. C matrix constructed with exact density. Solvent radius 1.4 Å. Diatomics —tetratomics surface testing with ADF 2.0 version. Remainder—ADF 2.3

Table 2. Triatomic test cases

	$\langle\rho\|H^g\rangle$ (a.u.)				Electrostatic			Destabilization			Total electronic		
	Gas	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt
HCC^-	−0.727414	−0.719854	−0.719854	−0.720478	−82.62	−82.62	−79.91	4.74	4.74	4.35	−77.88	−77.88	−75.56
NH_2^-	−0.522047	−0.516285	−0.516285	−0.515714	−95.80	−95.80	−94.44	3.62	3.62	3.97	−92.18	−92.18	−90.47
HCN	−0.755637	−0.750642	−0.750642	−0.751243	−12.27	−12.27	−11.16	3.13	3.13	2.76	−9.14	−9.14	−8.40
N_3^-	−0.941783	−0.941569	−0.941569	−0.941478	−70.80	−70.80	−69.56	0.13	0.13	0.19	−70.67	−70.67	−69.36
H_2O	−0.539929	−0.537299	−0.537299	−0.537475	−10.48	−10.48	−10.17	1.65	1.65	1.54	−8.83	−8.83	−8.63
OCN^-	−0.970435	−0.969598	−0.969598	−0.969386	−73.13	−73.13	−70.64	0.53	0.53	0.66	−72.60	−72.60	−69.98
N_2O	−0.840841	−0.840213	−0.840213	−0.840364	−4.41	−4.41	−2.98	0.39	0.39	0.30	−4.02	−4.02	−2.68
HOO^-	−0.533136	−0.527420	−0.527542	−0.528063	−96.97	−96.56	−92.87	3.59	3.51	3.18	−93.38	−93.05	−89.69
CO_2	−0.902048	−0.901035	−0.901035	−0.901173	−5.49	−5.49	−4.61	0.64	0.64	0.55	−4.85	−4.85	−4.06
NO_2^+	−0.362420	−0.360895	−0.360895	−0.361543	−96.56	−96.56	−91.51	0.96	0.96	0.55	−95.61	−95.60	−90.96
NO_2^-	−0.792041	−0.791746	−0.791746	−0.791833	−73.75	−73.75	−72.08	0.19	0.19	0.13	−73.56	−73.56	−71.95
O_3	−0.536474	−0.536053	−0.536078	−0.536218	−3.04	−2.87	−2.02	0.26	0.25	0.16	−2.77	−2.62	−1.86
FCN	−0.757497	−0.754144	−0.754144	−0.754693	−9.44	−9.44	−7.43	2.10	2.10	1.76	−7.34	−7.34	−5.67
FNO	−0.645275	−0.644231	−0.644226	−0.644407	−4.08	−4.05	−3.33	0.65	0.66	0.54	−3.43	−3.39	−2.78
PH_2^-	−0.441485	−0.439131	−0.439131	−0.438887	−67.91	−67.91	−66.76	1.48	1.48	1.63	−66.43	−66.43	−65.13
H_2S	−0.415328	−0.414066	−0.414066	−0.414236	−4.99	−4.99	−4.63	0.79	0.79	0.69	−4.20	−4.20	−3.95
COS	−0.759208	−0.758490	−0.758497	−0.758497	−4.11	−4.06	−3.75	0.45	0.45	0.45	−3.66	−3.61	−3.30
SO_2	−0.626525	−0.623970	−0.623979	−0.624084	−8.34	−8.27	−7.85	1.60	1.60	1.53	−6.74	−6.67	−6.32
CS_2	−0.618310	−0.617978	−0.617978	−0.618020	−3.68	−3.51	−2.75	0.21	0.21	0.18	−3.48	−3.30	−2.57
ClCN	−0.692846	−0.688559	−0.688616	−0.688912	−10.57	−10.30	−9.23	2.69	2.65	2.47	−7.88	−7.65	−6.76
ClNO	−0.583842	−0.582256	−0.582287	−0.582742	−4.35	−4.22	−3.20	1.00	0.98	0.69	−3.35	−3.24	−2.51
ClNS	−0.452863	−0.449652	−0.449587	−0.449709	−7.13	−7.08	−6.79	2.01	2.06	1.98	−5.12	−5.02	−4.82

Table 3. Tetratomic test cases

	$\langle\rho\|H^g\rangle$ (a.u.)				Electrostatic			Destabilization			Total electronic		
	Gas	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt
CH_3^+	−0.306404	−0.306012	−0.306012	−0.306123	−94.02	−94.02	−90.98	0.25	0.25	0.18	−93.77	−93.77	−90.80
CH_3^-	−0.678403	−0.673264	−0.673264	−0.672843	−84.29	−84.29	−82.04	3.22	3.22	3.49	−81.07	−81.07	−78.55
HCCH	−0.880788	−0.879576	−0.879576	−0.879901	−6.31	−6.31	−5.11	0.76	0.76	0.56	−5.55	−5.55	−4.55
NH_3	−0.740767	−0.738731	−0.738731	−0.738765	−7.91	−7.91	−7.81	1.28	1.28	1.26	−6.63	−6.63	−6.56
H_3O^+	−0.349492	−0.348448	−0.348448	−0.348818	−101.76	−101.76	−96.44	0.65	0.65	0.42	−101.11	−101.11	−96.01
H_2CO	−0.851644	−0.848318	−0.848318	−0.848665	−8.37	−8.37	−7.75	2.09	2.09	1.87	−6.28	−6.28	−5.88
HOCN	−1.018937	−1.012244	−1.012244	−1.013621	−18.69	−18.69	−16.12	4.20	4.20	3.34	−14.49	−14.49	−12.79
HNCO	−1.071485	−1.069303	−1.069303	−1.069588	−8.22	−8.22	−7.63	1.37	1.37	1.19	−6.85	−6.85	−6.44
HCNO	−0.972176	−0.968264	−0.968264	−0.968556	−9.25	−9.25	−8.63	2.45	2.45	2.27	−6.80	−6.80	−6.36
HOOH g	−0.702290	−0.699311	−0.699286	−0.699699	−12.20	−12.28	−11.34	1.87	1.89	1.63	−10.33	−10.39	−9.72
HOOH t	−0.699726	−0.689193	−0.689200	−0.689433	−12.62	−12.56	−11.65	6.61	6.61	6.46	−6.01	−5.95	−5.19
$HCOO^-$	−1.065903	−1.061760	−1.061760	−1.061996	−75.42	−75.42	−73.48	2.60	2.60	2.45	−72.82	−72.82	−71.03
CO_3^{2-}	−1.011490	−1.010056	−1.010056	−1.010481	−277.58	−277.58	−267.88	0.90	0.90	0.63	−276.68	−276.68	−267.25
NO_3^-	−1.068370	−1.068143	−1.068143	−1.068161	−68.58	−68.58	−66.71	0.14	0.14	0.13	−68.44	−68.44	−66.58
HCCF	−0.882943	−0.882391	−0.882391	−0.882538	−4.27	−4.27	−3.20	0.35	0.35	0.25	−3.92	−3.92	−2.94
HF–HF	−0.614969	−0.612257	−0.612387	−0.612873	−14.35	−13.90	−12.49	1.70	1.62	1.32	−12.65	−12.28	−11.18
COF_2	−0.953682	−0.952527	−0.952527	−0.952925	−5.64	−5.64	−4.00	0.73	0.73	0.48	−4.91	−4.91	−3.53
FOOF g	−0.564086	−0.562434	−0.562498	−0.562960	−6.53	−5.89	−4.49	1.04	1.00	0.71	−5.50	−4.89	−3.78
FOOF t	−0.505428	−0.505409	−0.505409	−0.505384	−1.62	−1.16	−0.77	0.01	0.01	0.03	−1.61	−1.15	−0.74
NF_3	−0.638232	−0.638220	−0.638230	−0.638235	−1.37	−0.99	−0.51	0.01	0.01	0.00	−1.36	−0.99	−0.51
PH_3	−0.579760	−0.579463	−0.579463	−0.579454	−1.49	−1.49	−1.44	0.19	0.19	0.19	−1.30	−1.30	−1.25
PF_3	−0.759905	−0.758659	−0.758677	−0.758848	−5.66	−5.40	−4.40	0.78	0.77	0.66	−4.88	−4.63	−3.74
H_3S^+	−0.236068	−0.235058	−0.235058	−0.235703	−89.07	−89.07	−84.27	0.63	0.63	0.23	−88.44	−88.44	−84.04
SOF_2	−0.708787	−0.706712	−0.706748	−0.707016	−7.36	−7.00	−5.90	1.30	1.28	1.11	−6.06	−5.72	−4.79
HCCCl	−0.820550	−0.819686	−0.819668	−0.819793	−5.75	−5.64	−4.66	0.54	0.55	0.47	−5.21	−5.09	−4.18
$COCl_2$	−0.780269	−0.779050	−0.779131	−0.779175	−5.13	−4.73	−3.88	0.76	0.71	0.69	−4.37	−4.02	−3.19
NCl_3	−0.458388	−0.457988	−0.457894	−0.457782	−3.88	−3.33	−3.23	0.25	0.31	0.38	−3.62	−3.02	−2.85
PCl_3	−0.514271	−0.513462	−0.513580	−0.513910	−4.53	−3.72	−2.68	0.51	0.43	0.23	−4.02	−3.29	−2.45

g – gauche, t – trans

Table 4. Pentatomic test cases

	$\langle\rho\|H^q\rangle$ (a.u.)				Electrostatic			Destabilization			Total electronic		
	Gas	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt	vdW	SE	Klamt
CH_4	−0.907350	−0.907295	−0.907295	−0.907314	−0.84	−0.84	−0.75	0.03	0.03	0.02	−0.81	−0.81	−0.72
CF_4	−1.042217	−1.042130	−1.042130	−1.042234	−2.61	−2.61	−0.50	0.05	0.05	0.00	−2.56	−2.56	−0.51
CCl_4	−0.670707	−0.670514	−0.670344	−0.670315	−4.29	−3.23	−2.94	0.12	0.23	0.25	−4.17	−3.00	−2.70
NH_4^+	−0.616104	−0.615930	−0.615930	−0.615982	−91.57	−91.57	−87.65	0.11	0.11	0.08	−91.46	−91.46	−87.57
PH_4^+	−0.423614	−0.423115	−0.423115	−0.423323	−83.98	−83.98	−80.33	0.31	0.31	0.18	−83.67	−83.67	−80.14
ClO_4^-	−0.931560	−0.931457	−0.931452	−0.931495	−59.97	−59.91	−58.37	0.06	0.07	0.04	−59.90	−59.84	−58.33
CH_3F	−0.919152	−0.917945	−0.917936	−0.918024	−4.67	−4.70	−4.39	0.76	0.76	0.71	−3.91	−3.94	−3.68
CH_3Cl	−0.849116	−0.847882	−0.847885	−0.848004	−4.75	−4.69	−4.39	0.77	0.77	0.70	−3.97	−3.92	−3.69
CH_3Br	−0.825298	−0.824157	−0.824163	−0.824285	−4.55	−4.48	−4.19	0.72	0.71	0.64	−3.83	−3.77	−3.56
CH_3O^-	−0.993240	−0.984271	−0.984271	−0.985311	−83.10	−83.10	−79.31	5.63	5.63	4.98	−77.47	−77.47	−74.34
CH_3S^-	−0.921767	−0.914262	−0.914288	−0.914136	−77.45	−77.41	−75.58	4.71	4.69	4.79	−72.74	−72.72	−70.79
CHF_3	−1.001797	−1.000535	−1.000539	−1.000882	−5.72	−5.62	−4.07	0.79	0.79	0.57	−4.93	−4.83	−3.50
$CHCl_3$	−0.735327	−0.734255	−0.734257	−0.734777	−6.00	−5.46	−3.95	0.67	0.67	0.34	−5.33	−4.79	−3.61
$CFCl_3$	−0.759070	−0.758873	−0.758817	−0.758629	−3.85	−3.07	−2.41	0.12	0.16	0.28	−3.72	−2.91	−2.13
CF_3Cl	−0.945710	−0.945590	−0.945604	−0.945492	−2.97	−2.64	−1.08	0.07	0.07	0.14	−2.89	−2.57	−0.95
NOF_3	−0.880230	−0.880175	−0.880188	−0.880236	−2.64	−2.10	−0.56	0.03	0.03	0.00	−2.60	−2.07	−0.56
POF_3	−1.045715	−1.044181	−1.044268	−1.044827	−6.91	−6.51	−4.12	0.96	0.91	0.56	−5.94	−5.60	−3.57
$POCl_3$	−0.798186	−0.795274	−0.795553	−0.795597	−9.56	−8.09	−6.90	1.83	1.65	1.62	−7.74	−6.44	−5.28
PSF_3	−0.928201	−0.927812	−0.927885	−0.928020	−3.94	−3.49	−1.87	0.24	0.20	0.11	−3.69	−3.29	−1.76
$PSCl_3$	−0.676487	−0.675690	−0.675907	−0.675938	−4.89	−3.61	−2.80	0.50	0.36	0.34	−4.39	−3.25	−2.45
CH_2F_2	−0.954875	−0.953270	−0.953268	−0.953532	−6.26	−6.20	−5.29	1.01	1.01	0.84	−5.26	−5.19	−4.45
CH_2Cl_2	−0.793428	−0.791924	−0.791945	−0.792440	−6.35	−6.11	−4.97	0.94	0.93	0.62	−5.41	−5.18	−4.35
CH_2Br_2	−0.743256	−0.741684	−0.741729	−0.742299	−6.55	−6.14	−5.00	0.99	0.96	0.60	−5.56	−5.18	−4.40
CF_2Cl_2	−0.851039	−0.850883	−0.850877	−0.850659	−3.44	−2.91	−1.79	0.10	0.10	0.24	−3.35	−2.81	−1.56
SO_2F_2	−0.950396	−0.949589	−0.949676	−0.950090	−4.91	−4.46	−2.18	0.51	0.45	0.19	−4.40	−4.01	−1.99
SO_2Cl_2	−0.803701	−0.801830	−0.802039	−0.801957	−7.31	−6.12	−4.88	1.17	1.04	1.09	−6.14	−5.08	−3.78
CH_2CN^-	−1.272173	−1.268093	−1.268093	−1.268182	−69.86	−69.86	−68.06	2.56	2.56	2.50	−67.30	−67.30	−65.55
CH_2CO	−1.198349	−1.197165	−1.197165	−1.197261	−5.35	−5.35	−4.78	0.74	0.74	0.68	−4.61	−4.61	−4.10
CH_2NN	−1.199679	−1.197975	−1.197975	−1.198154	−5.62	−5.62	−4.96	1.07	1.07	0.96	−4.55	−4.55	−4.00
CH_2FCl	−0.871505	−0.869871	−0.869893	−0.870281	−6.43	−6.29	−5.19	1.03	1.01	0.77	−5.41	−5.28	−4.42
CHF_2Cl	−0.908594	−0.907334	−0.907372	−0.907760	−5.85	−5.63	−3.99	0.79	0.77	0.52	−5.06	−4.86	−3.47
$CHFCl_2$	−0.819456	−0.818265	−0.818304	−0.818752	−5.96	−5.52	−3.91	0.75	0.72	0.44	−5.21	−4.80	−3.47
H_2CNH	−1.048983	−1.046012	−1.046012	−1.046388	−8.65	−8.65	−8.00	1.86	1.86	1.63	−6.79	−6.79	−6.37
H_2COH^+	−0.669594	−0.668082	−0.668082	−0.668737	−91.44	−91.44	−85.45	0.95	0.95	0.54	−90.49	−90.49	−84.92
HPO_3 s	−1.137053	−1.128109	−1.128584	−1.131950	−30.32	−29.06	−21.11	5.61	5.31	3.20	−24.70	−23.75	−17.91
HPO_3 e	−1.150327	−1.142329	−1.143241	−1.145491	−27.20	−25.03	−19.74	5.02	4.45	3.03	−22.18	−20.58	−16.71
HCO_3^- s	−1.350894	−1.346682	−1.346681	−1.347231	−77.96	−77.96	−74.64	2.64	2.64	2.30	−75.32	−75.32	−72.34
HCO_3^- e	−1.367280	−1.363430	−1.363490	−1.363886	−75.30	−75.23	−72.98	2.42	2.38	2.13	−72.88	−72.85	−70.85
HCOOH s	−1.165307	−1.162287	−1.162367	−1.162823	−12.08	−11.83	−10.56	1.90	1.85	1.56	−10.19	−9.98	−9.01
HCOOH a	−1.158905	−1.151837	−1.152225	−1.153571	−18.02	−17.35	−15.22	4.44	4.19	3.35	−13.59	−13.16	−11.87

s – staggered, e – eclipsed (HCOOH s – syn, a – anti)

matrix element $F_{\tau\lambda}$ of Eq. (25). It is also required if C_μ of Eq. 3 is evaluated as

$$C_\mu = \sum_k V_\mu(r_k)\rho(r_k)W(r_k), \tag{26}$$

where r_k is an integration point and $W(r_k)$ is its associated weight.

In our implementation, the numerical integration grid is chosen by the ADF program, whereas the surface points are chosen independently by the GEPOL algorithm. It is possible that a surface point may nearly coincide with an integration point. In this case, representing the surface potential as a point charge would result in numerical instability. To avoid this possibility, we smear the charge uniformly on a disk of radius R_μ^{B}, centered on triangle μ. The disk radius was initially chosen such that it inscribes the triangle. The unit charge potential associated with this smearing is given [14] by

$$V_\mu^{\mathrm{B}} = \frac{2}{R_\mu^{\mathrm{B}}}\left[1 - \lambda_\mu|P_1(\cos\omega)| + \frac{1}{2}\lambda_\mu^2 P_2(\cos\omega) - \frac{1}{8}\lambda_\mu^4 P_4(\cos\omega) + \cdots\right], \lambda_\mu < 1 \tag{27}$$

and

$$V_\mu^{\mathrm{B}} = \frac{2}{R_\mu^{\mathrm{B}}}\left[\frac{1}{2}\lambda_\mu^{-1} - \frac{1}{8}\lambda_\mu^{-3}P_2(\cos\omega) + \frac{1}{16}\lambda_\mu^{-5}P_4(\cos\omega) + \cdots\right], \lambda_\mu > 1 \ , \tag{28}$$

where $\lambda = (R_\mu^{\mathrm{B}})^{-1}|r_\mu^{\mathrm{B}} - r|$, and ω is the angle between $r - r_\mu^{\mathrm{B}}$ and $R_{\mathrm{B}} - r_\mu^{\mathrm{B}}$. P_n is the nth order Legendre polynomial.

Numerical experiments suggested that this problem occurred rarely. Although the disk potential is accurate, it is much more costly than using a point-charge model if always used, so we have employed an adaptive strategy to choose which pairs of points require a disk potential. Our algorithm determines, given a pair of points, if the point-charge model (calculated initially) and an approximate disk potential differ by a user-specified tolerance (default 0.01). If so, the algorithm determines the number of terms in the expansion parameter λ required to guarantee the desired accuracy, up to a user-specified maximum order (default 4), and calculates the resulting disk potential. These pairs are then stored externally. This approach reduces the computational costs by an order of magnitude, and is only slightly more expensive than using the point-charge model, which has also been implemented. Another option for reducing expense is to scale the size of the disk. A comparison of the effects of changing the disk scaling and the potential tolerance is given in Table 5. There is little change in the solvation energy overall, but there is some variation with disk scaling. Once the scaling factor exceeds 0.1, the computational time increases dramatically. The solvation energy, with the default scale of 0.01, does not vary with extension of the potential tolerance, although the computational time starts growing large if the tolerance is smaller than 10^{-4}. The defaults of 0.01 for the scale factor and 0.01 for the potential tolerance are seen to give a good balance between accuracy and efficiency.

Table 5. Effect of disk scaling and potential tolerance on HCl calculation

Scale	−log(tol)	Number of Special pairs	t (s)	E^{s} (kcal/mol)	E^0 (a.u.)
0.00	0	0	15.16	−5.113597	−0.22420528
0.01	2	0	15.57	−5.113597	−0.22420528
0.1	2	58	16.35	−5.112736	−0.22420528
0.2	2	576	29.10	−5.103269	−0.22420671
0.3	2	2336	145.30	−5.078991	−0.22421076
0.01	3	0	15.58	−5.113597	−0.22420528
0.01	4	58	16.35	−5.113588	−0.22420528
0.01	5	2784	192.89	−5.113559	−0.22420529

In Eq. (26), ρ represents the electron density of the solute. In our implementation, the user can choose whether the exact density (Eq. 10) or the fitted density (Eq. 18) is used. The fitted density usually takes less time to evaluate. By integrating analytically over the electronic coordinate instead, the alternative expression

$$C_\mu = \int \frac{\rho(r)}{|r_\mu - r|}dr = V_{\mathrm{C}}(r_\mu) \tag{29}$$

is obtained, which is the molecular Coulomb potential at r_μ. This potential can be evaluated from the fitted density according to Eq. 18.

The solvation energies, evaluated using the three different methods for calculating the C matrix (exact density, fitted density and fitted potential), are given in Table 6. For ionic species, the various methods are within 2 kcal/mol of each other, whereas for neutral species the agreement is much better at 0.2 kcal/mol. Plots of processing time versus species for gas-phase calculations and the three methods for C matrix evaluation are shown in Fig. 1. It is clear that the fitted-potential method cuts the extra time required to do a solvation calculation by half compared to either the fitted or exact density. This has been made the default.

2.7 SCF procedure modifications

The computation of the C matrix and the calculation of the contribution of the surface charges to the Fock operator add significant computational effort to each cycle of the SCF. Several schemes were examined in an attempt to reduce the computational effort. Table 7 shows that the total solvation energy obtained from a post-SCF approach is a reasonable estimate of the full variational energy for neutral nonpolar compounds. For ions, the results are within 2 kcal/mol of each other. The deviation for neutral species increases with dipole moment. Attempts to improve variational calculations by turning on the COSMO charges later in the SCF via several strategies did not lead to improvement in processing time.

2.8 Outlying charge correction

In real quantum-mechanical systems, a significant portion of the electron density will lie outside the cavity

Table 6. Diatomic test cases: C matrix[a]

	$\langle\rho\|H^g\rangle$ (a.u.)			Electrostatic			Destabilization			Total electronic		
	Gas	E	F	P	E	F	P	E	F	P	E	F	P
CN^-	−0.639784	−0.639432	−0.639434	−0.639431	−74.49	−73.96	−73.73	0.22	0.22	0.22	−74.27	−73.74	−73.51
N_2	−0.638750	−0.638697	−0.638699	−0.638696	−0.75	−0.74	−0.69	0.03	0.03	0.03	−0.72	−0.71	−0.66
OH^-	−0.339213	−0.335101	−0.335036	−0.335053	−104.90	−103.91	−103.69	2.58	2.62	2.61	−102.32	−101.29	−101.08
CO^+	−0.572583	−0.572448	−0.572450	−0.572446	−1.35	−1.34	−1.23	0.08	0.08	0.09	−1.27	−1.26	−1.14
NO^+	−0.112180	−0.111874	−0.111875	−0.111871	−99.81	−100.12	−100.14	0.19	0.19	0.19	−99.62	−99.93	−99.95
$O_2(^1\Sigma)$	−0.341607	−0.341603	−0.341602	−0.341601	−0.50	−0.59	−0.45	0.00	0.00	0.00	−0.50	−0.59	−0.45
$O_2(^3\Pi)$	−0.380291	−0.380289	−0.380287	−0.380286	−0.49	−0.58	−0.43	0.00	0.00	0.00	−0.49	−0.58	−0.43
HF	−0.299771	−0.297974	−0.297956	−0.297965	−8.24	−8.38	−8.22	1.13	1.14	1.13	−7.11	−7.24	−7.09
F_2	−0.144112	−0.144088	−0.144087	−0.144090	−0.87	−0.92	−0.77	0.02	0.02	0.01	−0.85	−0.90	−0.76
SH^-	−0.299077	−0.297676	−0.297626	−0.297611	−76.35	−76.36	−76.04	1.44	1.48	1.48	−74.91	−74.88	−74.56
CS	−0.379940	−0.376679	−0.376569	−0.376574	−6.74	−6.92	−6.72	2.05	2.12	2.11	−4.69	−4.80	−4.61
HCl	−0.225486	−0.224205	−0.224197	−0.224209	−5.11	−5.22	−5.05	0.80	0.81	0.80	−4.31	−4.41	−4.25
OCl^-	−0.288517	−0.284201	−0.284403	−0.284458	−81.14	−79.89	−79.17	2.71	2.58	2.55	−78.43	−77.31	−76.62
FCl	−0.159215	−0.158616	−0.158549	−0.158558	−2.99	−3.21	−3.00	0.38	0.42	0.41	−2.61	−2.79	−2.59
Cl_2	−0.130153	−0.129944	−0.129893	−0.129917	−2.75	−2.94	−2.50	0.13	0.16	0.15	−2.62	−2.78	−2.35

[a] E: exact density, F: fitted density, P: fitted potential

boundary, leading to the well-known problem of outlying charge [12]. If we let Q_{tot} be the total charge on the molecule, then $\Sigma_{\mu=1}^{M} q_\mu + Q_{\text{tot}}$ is a measure of the error in Gauss' law due to both the discretization and to outlying charge. We may correct for both at once by adding the constraint to the energy expression in Eq. 3 to form the Lagrangian

$$\mathscr{L}(q,\lambda) = E(q) + \lambda\left(\sum_{\mu=1}^{M} q_\mu + Q_{\text{tot}}\right) \quad . \tag{30}$$

Setting the derivative of the Lagrangian with respect to charge to zero gives

$$q^{\text{corr}} = -A^{-1}(BZ + C + \lambda\vec{1}) \tag{31}$$

$$= -A^{-1}(BZ + C) - \lambda A^{-1}\vec{1} \tag{32}$$

$$= q^{\text{uncorr}} + \lambda q^{\lambda}, \tag{33}$$

where $\vec{1} = [11\ldots1]^T$. An explicit expression for λ may be garnished by ensuring that the constraint is satisfied, which gives

$$\lambda = \frac{Q_{\text{tot}} + \sum_{\mu=1}^{M} q^{\text{uncorr}}}{\sum_\mu \sum_\nu (A^{-1})_{\mu\nu}} \quad . \tag{34}$$

The outlying-charge-correction formula depicted here depends on the availability of A^{-1}, which depends on how the COSMO equations are solved. For the iterative schemes, the unavailability of A^{-1} may be remedied by solving simultaneously both the COSMO equation and the equation $Aq_\lambda = \vec{1}$. We note that $q_i^\lambda = \sum_\mu (A^{-1})_{i\mu}$. Alternatively, we extrapolate $\sum q$ linearly from the cases $\lambda = 0, 1$ to determine λ for which the error disappears. Our implementation solves the COSMO equations first for a conductor, including any correction for outlying charge, and then scales the corrected charges by $f(\epsilon)$. For a spherical cavity, this approach would be equivalent to adding a constant charge to each (identical) surface element. For a heteronuclear diatomic species in which the atomic radii of the two atoms are equal, upon separation to form ions, both atoms would possess the same correction, since this depends only on the geometry. We agree with Klamt and Jonas [12] that local errors in the charges could result, since the anion would be expected to have a greater outlying charge.

We note that Eq. 7 is no longer valid for the charge-corrected case. A simple substitution gives

$$E^{\text{s}} = \frac{1}{2}(BZ + C)q - \frac{1}{2}\lambda q^T\vec{1} \tag{35}$$

$$= \frac{1}{2}(BZ + C)q + \frac{1}{2}\lambda Q_{\text{tot}} \quad , \tag{36}$$

where q are the corrected charges. For neutral species, the second term disappears, and the expression is the same.

Some of the results for the outlying charge correction are presented in Table 7. Our scheme makes little difference to the solvation energy, but as expected, the magnitude of the energy correction decreases in the order anions > neutrals > cations. The default does not correct for outlying charge because of the small size of the correction and because of the concern expressed above.

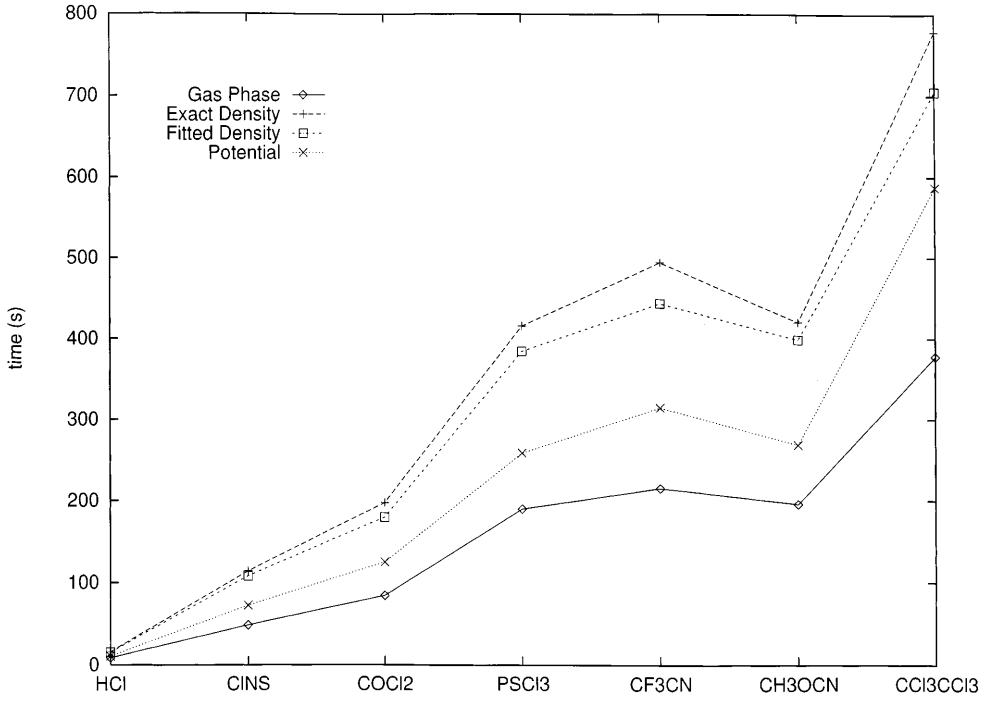


Fig. 1. Timings for various C-matrix calculation methods

Table 7. Diatomic test cases: outlying charge and perturbational

	Outlying charge corrected				Perturb.	Uncorrected
	$\langle\rho\|H^g\rangle$ (a.u.)	E^{es}	E^{destab}	E^{tot}	E^{tot}	E^{tot}
CN^-	−0.639374	−73.96	0.26	−73.70	−74.08	−74.27
N_2	−0.638698	−0.61	0.03	−0.58	−0.68	−0.72
OH^-	−0.335056	−104.57	2.61	−101.96	−100.19	−102.32
CO	−0.572450	−1.15	0.08	−1.07	−1.18	−1.27
NO^+	−0.111877	−99.76	0.19	−99.57	−99.42	−99.62
$O_2(^1\Sigma)$	−0.341603	−0.35	0.00	−0.35	−0.49	−0.50
$O_2(^3\Pi)$	−0.380289	−0.33	0.00	−0.33	−0.48	−0.49
HF	−0.297972	−8.16	1.13	−7.03	−6.14	−7.11
F_2	−0.144087	−0.78	0.02	−0.76	−0.84	−0.85
SH^-	−0.297553	−75.34	1.52	−73.82	−73.70	−74.91
CS	−0.376699	−6.36	2.03	−4.33	−3.34	−4.69
HCl	−0.224212	−4.86	0.80	−4.06	−3.71	−4.31
OCl^-	−0.284161	−80.17	2.73	−78.04	−76.70	−78.43
FCl	−0.158612	−2.80	0.38	−2.42	−2.27	−2.61
Cl_2	−0.129946	−2.46	0.13	−2.33	−2.46	−2.62

2.9 The energy gradient

If we consider the energy as being $E = E(X, q, C)$, where q and C are treated as implicit functions of X, then the total derivative of the energy with respect to the nuclear coordinate X_A can be written as

$$\frac{dE^T}{dX_A} = \left(\frac{\partial E^T}{\partial X_A}\right)_{q,C} + \sum_{i=1}^{n}\sum_{\mu=1}^{2M_v}\left(\frac{\partial E^T}{\partial C_{i\mu}}\right)_{X,q}\left(\frac{\partial C_{i\mu}}{\partial X_A}\right)_q + \sum_{\mu=1}\left(\frac{\partial E^T}{\partial q_\mu}\right)_{X,C}\left(\frac{\partial q_\mu}{\partial X_A}\right)_C . \quad (37)$$

The gas-phase terms of Eq. (37) require that we modify the gas-phase one-particle KS operator by adding the surface potential. The gas-phase energy E^0 does not depend on charge explicitly, therefore

$$\left(\frac{\partial E^T}{\partial q_\mu}\right)_{X,C} = \left(\frac{\partial E^s}{\partial q_\mu}\right)_{X,C} . \quad (38)$$

Since the COSMO equation ensures that this is zero (Eq. 4), the final term of Eq. (37) disappears. The first term is the sum of the explicit partial derivatives of the gas-phase and solvation energies. The latter can be written as

$$\frac{\partial E^s}{\partial X_A} = \frac{1}{2}\sum_\mu\sum_\nu q_\mu \frac{\partial A_{\mu\nu}}{\partial X_A} q_\nu + \sum_\mu q_\mu Z_A \frac{\partial B_{A\mu}}{\partial X_A} + 2\sum_\mu q_u \int_V \frac{\partial \psi_i(1)}{\partial X_A} V_\mu(1)\psi_i(1)d\tau + \sum_\mu \int_V \psi_i(1)\frac{\partial V_\mu^A(1)}{\partial X_A}\psi_i(1)d\tau . \quad (39)$$

The first and second terms of Eq. (39) are readily calculated from the analytical expressions for $A_{\mu v}$ and $B_{A\mu}$ (as given in Refs. [1] and [14]). The third term can be obtained from numerical integration in a way similar to C_μ itself. Because of the translational invariance condition,

$$\sum_B \frac{\partial}{\partial X_B} \int_V \psi_i(1) V_\mu^A(1) \psi_i(1) d\tau = 0, \tag{40}$$

the fourth term can be rewritten as

$$\sum_\mu \int_V \psi_i(1) \frac{\partial V_\mu^A(1)}{\partial X_A} \psi_i(1) d\tau = -2 \sum_B \sum_\mu q_\mu^A \int_V \frac{\partial \psi_i(1)}{\partial X_B} V_\mu^A(1) \psi_i(1) d\tau \ , \tag{41}$$

which can also be integrated numerically.

We would like to point out that taking the derivative of the Lagrangian corresponding to constrained charges (Eq. 30) affords exactly the same expressions for the derivative, provided the corrected charges are used.

2.10 Radii optimization

The construction of the various surfaces requires the radii of specific atoms or atom types to be specified. The expression of Truong and Stefanovich [3] was chosen for the nonelectrostatic term. The molecules of Table 8 provided the test set for optimization. The radii obtained are intended to be treated as preliminary values to be refined later. All radii were optimized by least-squares fitting on a grid with spacing 0.1 Å . The experimental values were taken from Table 2 of Ref. [16] and from Table 4.1 of Ref. [6].

First, the H and C radii were optimized for hydrocarbons. A uniform carbon radius of 2.4 Å was initially obtained. We found that the error was insensitive to the hydrogen radius, so this value was fixed at 1.16 Å. Subsequently, sets of compounds including O, N, S and P were examined separately to find radii for these elements and to possibly improve on the carbon radii. The radii obtained from this procedure were $C_{sp^3} = 2.3$, $C_{sp^2} = 2.2$, O = 1.3, N = 1.4, S = 1.7 and P = 2.4 (Å). Optimizations for other atoms, including the halogens, is in progress. A comparison of experimental and calculated solvation energies is given in Table 8. Approximately 70% of the values are within 1 kcal/mol of the experimental result. The secondary and tertiary amines are the largest outliers. The hydration energies of hydrocarbons are systematically underestimated with these radii.

Table 8. Hydration energies (kcal/mol) with optimized radii

	Expt	Calc
Compounds containing H,C (C = 2.4 Å)		
Methane	2.00	1.46
Ethane	1.83	1.54
Ethene	1.28	0.76
Ethyne	−0.01	−0.47
Propane	1.96	1.58
Propyne	−0.48	−0.43
Butane	2.08	1.56
1-Butyne	−0.16	−0.27
1-Butyn-3-ene	0.04	−0.49
Pentane	2.33	1.68
Benzene	−0.90	−0.59
Methylbenzene	−0.80	−0.68
Ethylbenzene	−0.80	−0.54
Compounds containing O		
Methanol	−5.1	−3.97
Methanal	−1.7	−2.48
Ethanol	−5.0	−4.15
Ethanal	−3.5	−3.71
Methyl methanoate	−2.8	−3.09
Ethanoic acid	−6.7	−6.01
Dimethyl ether	−1.9	−1.12
2-Propanol	−4.8	−4.53
Propanone	−3.9	−3.93
Methyl ethanoate	−3.3	−3.45
2-Pentanone	−3.5	−3.91
4-Methyl-2-pentanone	−3.1	−3.64
Phenol	−6.6	−6.95
Compounds containing N		
Nitromethane	−3.7	−4.43
Methylamine	−4.57	−3.56
Dimethylamine	−4.3	−2.13
Ethanamide	−9.7	−10.32
Ethylamine	−4.5	−3.67
cis-N-Methylethanamide	−10.1	−8.56
trans-N-Methylethanamide	−10.1	−8.58
Trimethylamine	−3.2	0.88
Propylamine	−4.4	−3.62
Diethylamine	−4.1	−2.04
Butylamine	−4.4	−3.84
4-Methylpyridine	−4.9	−4.82
3,5-Dimethylpyridine	−5.5	−4.75
Compounds containing S		
Methanethiol	−1.2	−1.57
Dimethyl sulfide	−1.5	−1.22
Ethanethiol	−1.2	−1.92
Methyl ethyl sulfide	−1.4	−1.32
Diethyl sulfide	−1.4	−1.38
Compounds containing P		
Phosphine	0.6	0.65

3 Applications

3.1 Benzene oxide – oxepin valence tautomerism

Considerable experimental and theoretical interest has been given to the benzene oxide – oxepin and related valence tautomerisms [17]. Four species are of interest here: benzene oxide, oxepin, the transition state for their interconversion (valence tautomerization) and the planar transition state for oxepin inversion. The structures and energies of these species can be found in Tables 9 and 10 respectively, and a gas-phase reaction profile is shown in Fig. 2 (without zero-point-energy effects). The non-local (NL) energy of tautomerization from oxepin to benzene oxide (0.2 kcal/mol) is reasonably close to the QCISD(T)/6-31G* value (−0.3 kcal/mol). Predicting the correct sign when the values are so close to zero is a very difficult task. The barriers to tautomerization and to inversion are lower than both the estimated experimental and the QCISD(T) values.

An experimental value of the difference in the free energy of tautomerization ($\Delta\Delta G$) for two different

Table 9. Benzene oxide–oxepin valence tautomerism: Z matrices

INTEGRATION 5.0 5.0 5.0

1.	XX	0	0	0	0.0	0.0	0.0	0.00
2.	XX	1	0	0	1.000000	0.0	0.0	0.00
3.	C	1	2	0	CD	90.0000	0.0	1.63
4.	C	1	2	3	CD	90.0000	180.0000	1.63
5.	H	3	1	2	H1C1	H1C1D	DH1	1.18
6.	H	4	1	2	H1C1	H1C1D	−DH1	1.18
7.	C	3	4	2	C2C1	C2C1C6	0.0000	1.63
8.	C	4	3	2	C2C1	C2C1C6	0.0000	1.63
9.	H	7	3	4	H2C2	H2C2C1	DH3	1.18
10.	H	8	4	3	H2C2	H2C2C1	−DH3	1.18
11.	C	7	3	4	C3C2	C3C2C1	DH4	1.63
12.	C	8	4	3	C3C2	C3C2C1	−DH4	1.63
13.	H	11	7	3	H3C3	H3C3C2	DH5	1.18
14.	H	12	8	4	H3C3	H3C3C2	−DH5	1.18
15.	O	2	1	3	OD2	OD2D1	90.0000	1.46

		Benzene oxide		Transition state		Oxepin		Planar oxepin[a]	
		LDA	NL	LDA	NL	LDA	NL	LDA	NL
CD	=	0.7129	0.7213	0.6943	0.7018	0.6765	0.6797	0.6666	0.6713
H1C1	=	1.0951	1.0916	1.0955	1.0918	1.0962	1.0926	1.0949	1.0913
C2C1	=	1.3494	1.3568	1.3851	1.3925	1.4272	1.4451	1.4525	1.4673
H2C2	=	1.0950	1.0915	1.0959	1.0924	1.0967	1.0932	1.0956	1.0923
C3C2	=	1.4416	1.4608	1.3779	1.3948	1.3366	1.3433	1.3297	1.3376
H3C3	=	1.0991	1.0947	1.0981	1.0934	1.0973	1.0924	1.0938	1.0894
OD2	=	2.0755	2.1249	1.9966	2.0388	1.9696	2.0123	2.1205	2.1504
H1C1D	=	118.26	118.24	117.63	117.70	118.14	118.28	118.37	118.38
C2C1C6	=	121.41	121.46	122.24	122.25	123.99	124.33	125.72	125.99
H2C2C1	=	120.82	120.89	119.73	119.68	118.58	118.14	116.23	116.13
C3C2C1	=	120.37	120.54	121.78	122.02	124.05	124.80	130.77	130.56
H3C3C2	=	119.22	118.76	123.61	123.37	123.90	123.90	119.90	120.41
OD2D1	=	134.86	136.77	130.16	131.45	135.00	137.52	N/A	N/A
DH1	=	171.01	172.43	165.92	166.36	171.91	172.90	180.0f	180.0f
DH3	=	165.30	167.53	148.99	150.62	146.67	148.01	180.0f	180.0f
DH4	=	−13.06	−10.80	−25.25	−23.32	−31.43	−30.22	0.0f	0.0f
DH5	=	164.93	163.50	168.13	167.73	174.74	175.50	180.0f	180.0f

[a] Different Z matrix

5.	XX	2	1	3	1.000000	90.0000	90.0000	0.00
16.	O	2	5	1	OD2	90.0000	180.0000	1.46

atom numbers between [6–15] incremented by one. f = fixed

Table 10. Benzene oxide – Oxepin valence tautomerism: energies

	$\langle\rho\|H^0\rangle$									
	LDA	NL	$\epsilon = 1.95$	$\epsilon = 72$	E^{s}	E^{s}	E^{dest}	E^{dest}	E^{tot}	E^{tot}
Benzene Oxide	−3.15003235	−2.90570874	−2.90509293	−2.90245196	−3.86	−9.90	0.39	2.04	−3.47	−7.86
Transition State	−3.14323472	−2.89843559	−2.89785323	−2.89558836	−3.46	−8.70	0.37	1.79	−3.09	−6.91
Oxepin	−3.14798464	−2.90606211	−2.90553833	−2.90359695	−3.30	−8.11	0.33	1.22	−2.97	−6.89
Planar	−3.13891435	−2.89856774	−2.89825249	−2.89699446	−3.20	−7.47	0.20	0.79	−3.00	−6.68

solvent systems (isooctane, $\epsilon = 1.95$; 15% aqueous methanol, $\epsilon = 72$) can be estimated as 1.8 kcal/mol. Our previous theoretical estimation [17] was 0.4 kcal/mol. COSMO calculations were performed at the gas-phase geometries to give a value of 1.0 kcal/mol, in much better agreement with experiment.

3.2 Glycine structure in aqueous media

The structure of glycine in the gas phase [18–22] is very different from the structure in aqueous solution or in the solid state [23]. Glycine, the simplest amino acid, possesses both an amino and a carboxylic acid group in the gas phase, whereas in aqueous solution, the favored structure is the result of a proton transfer from the acid to the amine parts of the molecule.

Initially, microwave spectroscopy was only able to identify the gas-phase conformer of glycine which had the largest dipole moment (4.6 D), and thus the most intense spectra [18, 19]. This conformer (**1**, shown in Fig. 3) possesses an internal hydrogen bond between the hydroxyl hydrogen and the nitrogen lone pair. Later, a second conformer (**2**) was identified, possessing a dipole

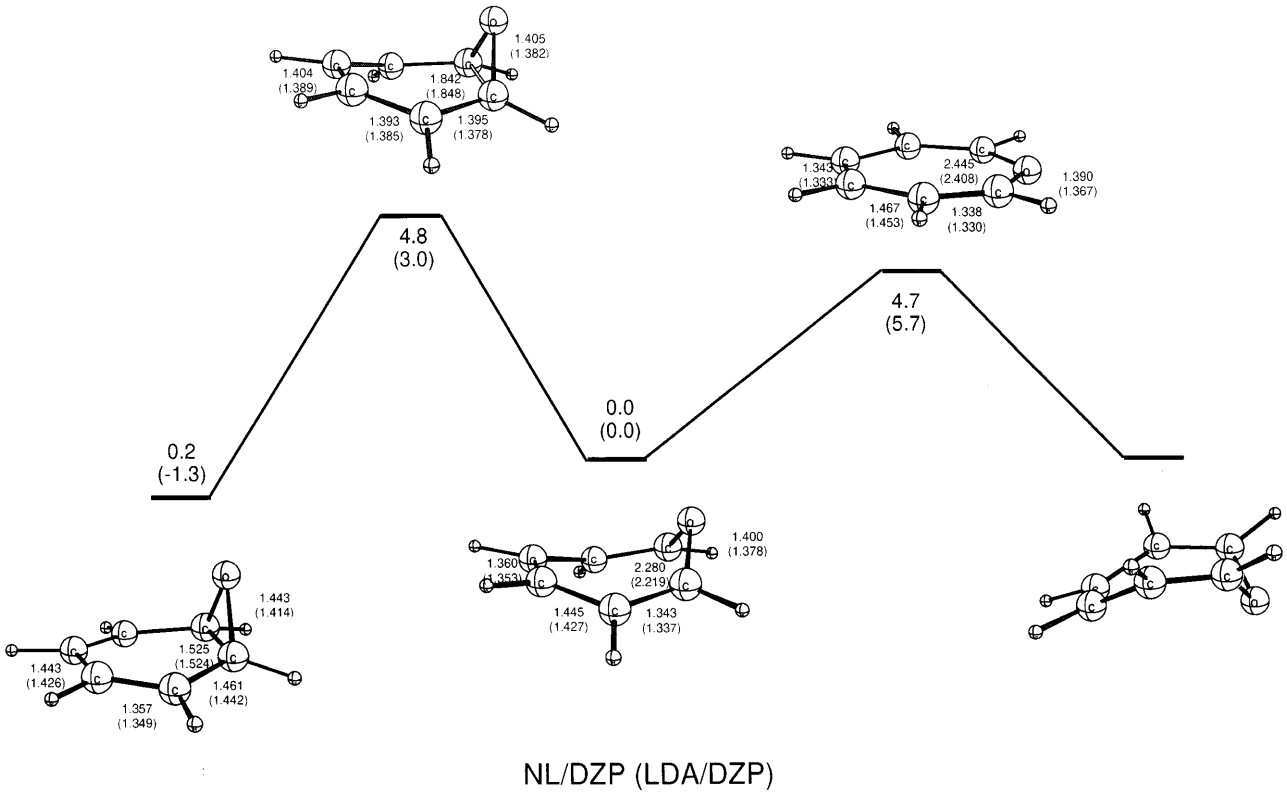


Fig. 2. Energy profile for benzene oxide—oxepin valence tautomerism

moment of 1.0 ± 0.15 D, and which is 1.4 ± 0.4 kcal/mol lower in energy [20, 21]. This conformer posseses hydrogen bonds between the hydrogens of the amine with the carbonyl oxygen. An electron diffraction study is consistent with the identification of the lowest energy conformer above [22], but identifies a third conformer (**3**) as being next highest in energy, with hydrogen bonds between the hydrogens of the amine and the hydroxyl oxygen. The difference in enthalpy between the neutral and zwitterionic forms (**4**) of glycine in aqueous solution has been estimated as 9.9 kcal/mol [24].

There have been several COSMO investigations on this interesting system [1, 2, 4, 15]. The energy differences between the two lowest gas-phase conformers are calculated to be 3.3 and 1.1 kcal/mol, at the LDA/DZP and BP/DZP levels, respectively, but the order is reversed. The same trend was noted by other workers [2], who obtained values of 2.5 and 0.3 kcal/mol, respectively.

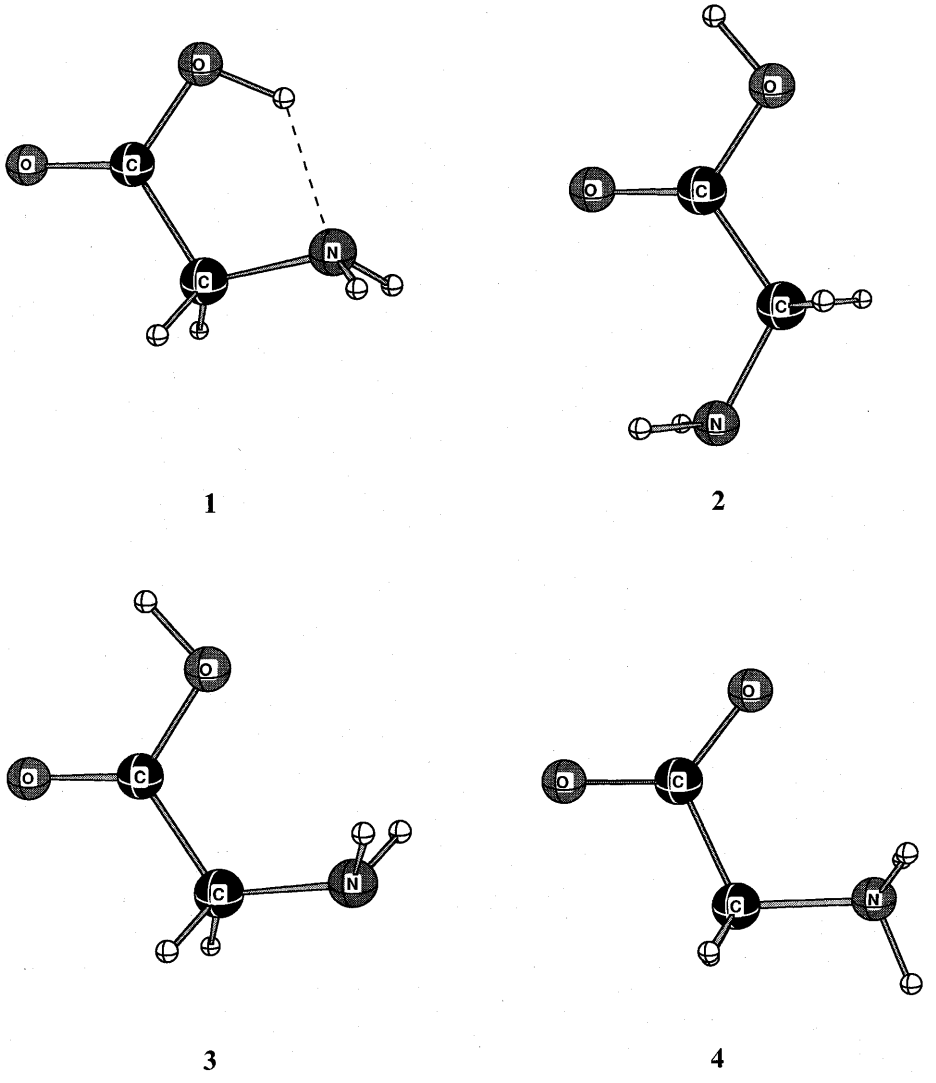
For an aqueous environment, we have calculated the difference in the neutral and zwitterionic forms to be 9.4 kcal/mol (BP/DZP), in excellent agreement with the experimental estimate. Other workers have obtained similar values of 10.7 (LDA) and 8.7 (BP) [2], 7.0 (HF/6-31G(2d,p)) [4] and 8.3 kcal/mol (AM1) [1]. The importance of this example is that geometry optimization in solution is required for the zwitterionic form of glycine, because this structure does not exist in the gas phase.

4 Conclusions

The COSMO model of solvation has been implemented in the ADF program. Several schemes for defining the surface, solving the COSMO equations, C matrix construction, and SCF convergence improvement have been examined. A gradient-preserving algorithm for removing outlying charge has been implemented. Preliminary energy-optimized radii are given as H = 1.16, $C_{sp^3} = 2.3$, $C_{sp^2} = 2.2$, O = 1.3, N = 1.4, S = 1.7 and P = 2.4. The majority of solvation energies calculated with these radii are within 1 kcal/mol of experiment. The difference in the enthalpy of tautomerization of benzene oxide to oxepin between two solvent systems was calculated to be 1.0 kcal/mol, in good agreement with the experimental result of 1.7 kcal/mol. For the enthalpy difference of neutral and zwitterionic forms of glycine in aqueous solution, we obtained 9.4 kcal/mol, in excellent agreement with the experimental estimate of 9.9 kcal/mol.

Acknowledgements. The assistance of a fellowship (PDF) granted by NSERC Canada to C. C. P. is gratefully acknowledged. T. Z. acknowledges support from NSERC and the donors of the Petroleum Research Fund, administered by the American Chemical Society (grant ACS-PRF no. 31205-AC3), as well as by Novacor Research and Technology Corporation (NRTC) of Calgary. The Izaak Walton Killam Memorial Foundation is highly appreciated by T. Z. The assistance of Kumar Vanka and Eva Zurek during the radii optimization process, and of Yunzhu Han in the initial stages of implementation, is highly appreciated. The authors also acknowledge fruitful discussions with Andreas Klamt during his November 1998 visit to the University of Calgary.

Fig. 3. Glycine conformers examined in this study



References

1. Klamt A, Schüürmann G (1993) J Chem Soc Perkin Trans 2: 799–805
2. Andzelm J, Kölmel C, Klamt A (1995) J Chem Phys 103: 9312–9320
3. Truong TN, Stefanovich EV (1995) Chem Phys Lett 240: 253–260
4. Baldridge K, Klamt A (1997) J Chem Phys 106: 6622–6633
5. Barone V, Cossi M (1998) J Phys Chem A 102: 1995–2001
6. Geleßus A (1997) Ph. D. thesis. University of Zurich, Zurich, Switzerland
7. Pye CC, Ziegler T ADF 2.3 version of COSMO
8. (a) Baerends EJ, Ellis DE, Pos P (1973) Chem Phys 2: 41–51; (b) Ravenek W (1987) In: te Riele HJJ, Dekker TJ, van der Vorst HA (eds) Algorithms and applications on vector and parallel computers. Elsevier, Amsterdam; (c) te Velde G (1994) ADF user manual, Amsterdam; (d) Boerrigter PM, te Velde G, Baerends EJ (1988) Int J Quantum Chem 33: 87–113; (e) te Velde G, Baerends EJ (1992) J Comput Phys 99: 84–98; (f) Fonseca C, Visser O, Snijders JG, te Velde G, Baerends EJ (1995) In: Clementi E, Corongiu G (eds) Methods and techniques in computational chemistry (METECC-95). STEF Cagliari, pp 305–395
9. Pascual-Ahuir JL, Silla E, Tuñon I (1994) J Comput Chem 15: 1127–1138
10. Burden RL, Faires JD (1989) Numerical analysis, 4th edn. PWS-Kent, Boston, Chs. 6,7 pp 311–424
11. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in fortran, 2nd edn. Cambridge University Press, Cambridge
12. Klamt A, Jonas V (1996) J Chem Phys 105: 9972–9981
13. Klamt A, Jonas V, Bürger T, Lohrenz JCW (1998) J Phys Chem A 102: 5074–5085
14. Morse PM, Feshback H (1953) Methods of theoretical physics. McGraw-Hill, New York pp 1266–1267
15. Truong TN, Stefanovich EV (1995) J Chem Phys 103: 3709–3717
16. Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, Ringnalda M, Goddard WA III, Honig B (1994) J Am Chem Soc 116: 11875–11882
17. Pye CC, Xidos JD, Poirier RA, Burnell DJ (1997) J Phys Chem A 101: 3371–3376
18. Brown RD, Godfrey PD, Storey JWV, Bassez M-P (1978) J Chem Soc Chem Commun 547–548
19. Suenram RD, Lovas FJ (1978) J Mol Spectrosc 72: 372–382
20. Suenram RD, Lovas FJ (1980) J Am Chem Soc 102: 7180–7184
21. Godfrey PD, Brown RD (1995) J Am Chem Soc 117: 2019–2023
22. Iijima K, Tanaka K, Onuma S (1991) J Mol Struct 246: 257–266
23. Jönsson P-G, Kvick Å(1972) Acta Crystallogr Sect B 28: 1827–1833
24. Haberfield P (1980) J Chem Ed 57: 346–347