

***MN-GSM:***  
**A Module Incorporating the SM5.42, SM5.43, SM6, SM8, SM8AD,  
SM8T, VEM(d) and SM12 Solvation Models, the CM2, CM3, CM4, and  
CM4M Charge Models, Löwdin Population Analysis, and  
Redistributed Löwdin Population Analysis in the *Gaussian 09* Program**

**Users Manual**

Version 2012

date of finalization of this version of software: November 20, 2012

date of most recent change in this manual: November 20, 2012

Ryan M. Olson,<sup>a</sup> Aleksandr V. Marenich,<sup>a</sup> Adam C. Chamberlin,<sup>a</sup>  
Casey P. Kelly,<sup>a</sup> Jason D. Thompson,<sup>a</sup> James D. Xidos,<sup>a</sup>  
Jiabo Li,<sup>a</sup> Gregory D. Hawkins,<sup>a</sup> Paul D. Winget,<sup>a</sup> Tianhai Zhu,<sup>a</sup> Daniel Rinaldi,<sup>b</sup>  
Daniel A. Liotard,<sup>c</sup> Christopher J. Cramer,<sup>a</sup> Donald G. Truhlar,<sup>a</sup>  
and Michael J. Frisch<sup>d</sup>

<sup>a</sup>*Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431*

<sup>b</sup>*Laboratoire de Chimie Théorique, Université de Nancy I, Vandoeuvre-Nancy 54506, France*

<sup>c</sup>*Laboratoire de Physico-Chimie Théorique, Université de Bordeaux I, 351 Cours de la Libération, 33405 Talence Cedex, France*

<sup>d</sup>*Lorentzian, Inc., 140 Washington Avenue, New Haven, CT 06473*

## Contents:

MN-GSM Reference .....	3
Licensing and MN-GSM User Agreement .....	4
Executive Summary .....	5
Löwdin Population Analysis and Redistributed Löwdin Population Analysis .....	9
Charge Model 2, Charge Model 3, Charge Model 4/4M, Charge Model 5.....	9
SM5.42, SM5.43, SM6, SM8, SM8AD, SM8T, VEM(d), and SM12 Solvation Models.....	10
Allowed Combinations of Solvent Model, Electronic Structure Theory, .....	14
and Basis Set.....	14
Analytical Gradients and Geometry Optimization .....	16
Literature References.....	17
Usage, Required Iop Statements, and \$MNGSM Namelist .....	22
All Keywords.....	23
Extended Description of AtmProp .....	31
Extended Description of CDSRead.....	32
Extended Description of DeLEMin .....	34
Summary of Keywords and Their Default Values .....	34
MN-GSM Keywords Required for Running Standard SMx Calculations .....	35
Modified Gaussian links .....	36
Important Notes and Troubleshooting.....	37
Special Notes on Basis Sets.....	40
Input Examples .....	42
MN-GSM Test Suite.....	45
Density Functional Methods Available in Gaussian and MN-GFM .....	47
SCF Schemes and SCF Convergence .....	47
Installing MN-GSM .....	49
Running MN-GSM .....	50
Running MN-GSM for SM12 Calculations.....	51
Running MN-GSM for Solvatochromic Calculations .....	53
Tested Platforms .....	54
Revision History and Version Summaries.....	55
Further Information.....	65
Acknowledgments.....	65

## **MN-GSM Reference**

Publications including work performed with *MN-GSM* should cite the software package. The recommended forms of reference are as follows:

### ***ACS style***

Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM version 2012*, University of Minnesota, Minneapolis, MN 55455-0431, 2012.

### ***Elsevier style***

R. M. Olson, A. V. Marenich, A. C. Chamberlin, C. P. Kelly, J. D. Thompson, J. D. Xidos, J. Li, G. D. Hawkins, P. D. Winget, T. Zhu, D. Rinaldi, D. A. Liotard, C. J. Cramer, D. G. Truhlar, and M. J. Frisch, *MN-GSM version 2012*, University of Minnesota, Minneapolis, MN 55455-0431, 2012.

### ***Journal of Chemical Physics style or World Scientific style***

R. M. Olson, A. V. Marenich, A. C. Chamberlin, C. P. Kelly, J. D. Thompson, J. D. Xidos, J. Li, G. D. Hawkins, P. D. Winget, T. Zhu, D. Rinaldi, D. A. Liotard, C. J. Cramer, D. G. Truhlar, and M. J. Frisch, *MN-GSM version 2012*, University of Minnesota, Minneapolis, MN 55455-0431, 2012.

### ***Theoretical Chemistry Accounts style***

Olson RM, Marenich AV, Chamberlin AC, Kelly CP, Thompson JD, Xidos JD, Li J, Hawkins GD, Winget PD, Zhu T, Rinaldi, D, Liotard DA, Cramer CJ, Truhlar DG, Frisch MJ (2012) *MN-GSM version 2012*. University of Minnesota, Minneapolis, MN 55455-0431.

### ***Wiley style***

Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM version 2012* (University of Minnesota, Minneapolis, MN 55455-0431, 2012).

## Licensing and MN-GSM User Agreement

*Gaussian 09* is licensed by Gaussian, Inc. Modified code may not be redistributed except by Gaussian, Inc. Use of *MN-GSM* implies acceptance this user agreement.

1. No licensee or user will redistribute the source code or executable code to a third party in original or modified form without the written permission of one of the principal investigators (Christopher J. Cramer or Donald G. Truhlar) and Gaussian, Inc. *MN-GSM* is distributed only by Gaussian, Inc. and only to licensed users of *Gaussian*. A third party is a party not covered by the same *Gaussian* license as the party obtaining *MN-GSM* from Gaussian, Inc.
2. The licensee has no ownership rights in the *MN-GSM* software or in any copyrights of the *MN-GSM* software or documentation through this license. The copyright notice should not be removed from the program.
3. Publications resulting from using *MN-GSM* will cite the program. The recommended reference is given in the *MN-GSM* Users Manual and is printed by the program.
4. The message printed by the program that says that use implies acceptance of this agreement and that gives the recommended reference should not be removed.
5. No guarantee is made that this program is bug-free or suitable for specific applications, and no liability is accepted for any limitations in the mathematical methods and algorithms used within the program.
6. No consulting or maintenance services are guaranteed or implied.

## Executive Summary

*MN-GSM* (Minnesota Gaussian Solvation Module) is a module for incorporation of Löwdin population analysis, redistributed Löwdin population analysis (RLPA), the CM2, CM3, CM4, and CM4M charge models, and the SM5.42, SM5.43, SM6, SM8, SM8AD, and SM8T solvation models into revision C.01 of the *Gaussian 09* code and *MN-GFM* version 6.2 and later. The SM5.42, SM5.43, SM6, SM8, and SM8AD models can calculate free energies of solvation using either gas- or liquid-phase optimized geometries. The SM8T model calculates free energies of solvation in water at different temperatures. The current version of *MN-GSM* includes a tool to run a ground-state solvation energy calculation using CM5, CHELPG, and MKS charges and the SM12 solvation model. See the section entitled “Running *MN-GSM* for SM12 Calculations.” The current version of *MN-GSM* also includes a tool to run a vertical excitation energy calculations in solution using CM5, CHELPG, and MKS charges and the generalized Born VEM(d) model. See the section entitled “Running *MN-GSM* for Solvatochromic Calculations.”

*MN-GSM* adds the following capabilities to *Gaussian*:

- Gas-phase and liquid-phase Löwdin partial atomic charges can be obtained for all restricted and unrestricted Hartree-Fock (HF), density-functional theory (DFT), and hybrid-DFT wave functions. For calculations using the 6-31+G(d) and 6-31+G(d,p) basis sets, gas-phase and liquid-phase redistributed Löwdin population analysis (RLPA) partial atomic charges can be obtained for all restricted and unrestricted HF, DFT, and hybrid DFT methods.
- Gas-phase and liquid-phase CM2 class IV charges can be determined for the following combinations of electronic structure theory and basis set (using either a restricted or an unrestricted formalism):

HF/MIDI!	B3LYP/MIDI!
HF/MIDI!6D	BPW91/6-31G(d)
HF/6-31G(d)	HF/6-31+G(d)
BPW91/MIDI!	HF/cc-pVDZ
BPW91/MIDI!6D	BPW91/DZVP

- Gas-phase and liquid-phase CM3 class IV charges can be determined for the following combinations of electronic structure theory and basis set (using either a restricted or an unrestricted formalism):

HF/MIDI!6D	HF/6-31G(d)
MPWX/MIDI!	MPWX/MIDI!6D
MPWX/6-31G(d)	MPWX/6-31+G(d)
MPWX/6-31+G(d,p)	BLYP/6-31G(d)
B3LYP/MIDI!6D	B3LYP/6-31G(d)
B3LYP/6-31+G(d)	

MPWX is a method that uses the mPW exchange functional of Adamo and Barone (Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664), the PW91 correlation functional (Perdew, J. P. *Electronic Structure of Solids '91*; Ziesche, P., Eshrig, H., Eds.; Akademie: Berlin, 1991) and a percentage of HF exchange, X. Note that MPWX includes the following special cases:

$$\begin{aligned}\text{MPW0} &\equiv \text{mPWPW91} \\ \text{MPW6} &\equiv \text{MPW1S} \\ \text{MPW25} &\equiv \text{mPW1PW91} \\ \text{MPW42.8} &\equiv \text{MPW1K} \\ \text{MPW60.6} &\equiv \text{MPW1KK}\end{aligned}$$

For all of the MPWX methods listed above, CM3 and CM4 have been parameterized for five specific values of X, namely 0, 25, 42.8, 60.6, and 99.9, and these parameter sets are available in *MN-GSM*. Every CM3 and CM4 parameter is a linear or a quadratic function of the percentage of HF exchange used in the mPW exchange functional. So, in addition to the specific CM3 and CM4 parameter sets (i.e. when X in MPWX is 0, 25, 42.8, 60.6, and 99) the CM3 and CM4 Charge Models are available for any value of X in MPWX between 0.0 and 100.0. Note that the CM3 and CM4 parameters were optimized using a corrected version of the modified Perdew-Wang density functional as implemented in *Gaussian 03/09*. The details of this correction are described fully in “The Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory” by Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A*, **2003**, *107*, 1384.

The CM3 model for the BLYP and B3LYP methods uses a slightly modified mapping scheme for compounds that contain N and O. For more information, see “Parameterization of Charge Model 3 For AM1, PM3, BLYP, and B3LYP” by Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.*, **2003**, *24*, 1291. We have also developed a special CM3 model for assigning partial atomic charges to high-energy materials. This model is called CM3.1, and it uses the same mapping scheme as the CM3 model for BLYP and B3LYP. This model has been parameterized for use with HF/MIDI!, and is described in “Accurate Partial Atomic Charges for High-Energy Molecules with the MIDI! Basis Set” by Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.*, **2005**, *113*, 133.

- Gas-phase and liquid-phase CM4 class IV charges can be determined using any of the density functionals in *Gaussian 03/09* and *MN-GFM*, as well as any user-defined density functional with an arbitrary amount of Hartree-Fock exchange (see the section entitled “Density Functional Methods Available in *Gaussian* and *MN-GFM*” for more information). Thus, unlike CM2 and CM3, CM4 is independent of the choice of electronic structure method. However, like CM2 and CM3, CM4 is dependent on the choice of basis set. Currently, CM4 parameter sets exist for the following basis sets: MIDI!, MIDI!6D, 6-31G(d), 6-31+G(d), 6-31+G(d,p), 6-31G(d,p), cc-pVDZ, cc-pVDZ\_g03, DZVP, 6-31B(d), and 6-31B(d,p). Note that the basis sets here abbreviated to cc-pVDZ\_g03 are Dunning’s correlation consistent basis sets cc-pVDZ (Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007) that have had redundant functions removed and have been rotated in order to increase computational efficiency (Frisch, A.;

Frisch, M. J.; Trucks, G. W. *Manual Version 7.0, Gaussian 03 User's Reference*; Gaussian, Inc.: Carnegie, PA, 2003; p. 25).

- The CM4M charge model is an extension of the earlier CM4 model. The CM4M model was individually optimized for the M06 suite of density functionals (namely, M06-L, M06, M06-2X, and M06-HF) for eleven basis sets which are MIDI!, MIDI!6D, 6-31G(d), 6-31+G(d), 6-31+G(d,p), 6-31G(d,p), cc-pVDZ, cc-pVDZ\_g03, DZVP, 6-31B(d), and 6-31B(d,p). Note that the basis sets here abbreviated to cc-pVDZ\_g03 are Dunning's correlation consistent basis sets cc-pVDZ (Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007) that have had redundant functions removed and have been rotated in order to increase computational efficiency (Frisch, A.; Frisch, M. J.; Trucks, G. W. *Manual Version 7.0, Gaussian 03 User's Reference*; Gaussian, Inc.: Carnegie, PA, 2003; p. 25).
- Calculation of the solvent-accessible surface areas (SASAs) of the atoms of a given solute. The SASA is that defined by Lee and Richards (see Lee, B.; Richards, F. M. *Mol. Biol.* **1971**, *55*, 379.) and Hermann (see Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.). In this definition, the solvent is taken to be a sphere of radius  $r_S$  and the solute is represented by a set of atom-centered spheres of a given set of radii. By default, the van der Waals radii of Bondi are used when defined; in cases where the atomic radius is not given in Bondi's paper (Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441) a radius of 2.0 Å is used. The SASA is the area generated by rolling the spherical solvent molecule on the van der Waals surface of the molecule. The SASA is calculated with the Analytic Surface Area (ASA) algorithm (see Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422. By default, the solvent radius is set to 0.40 Å (see Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532 for a justification of this value for the solvent radius), but the user can specify a different value for the solvent radius (including zero, which yields the van der Waal's surface area) with the keyword "SolvRd". A solvent radius of 0.0 Å is recommended for predicting solvation free energies with SM5.42, while the default value of 0.40 Å is recommended for predicting solvation free energies with SM5.43 and SM6. See the section entitled *All Keywords* for more details.
- Liquid-phase calculations based on gas-phase geometries can be performed with SM5.42 for the following restricted and unrestricted Hartree-Fock, DFT, and adiabatic-connection-method wave functions (i.e. hybrid DFT wave functions) that employ spherical harmonic or Cartesian  $d$  functions:

HF/MIDI!	B3LYP/MIDI!
HF/MIDI!6D	BPW91/6-31G(d)
HF/6-31G(d)	HF/6-31+G(d)
BPW91/MIDI!	HF/cc-pVDZ
BPW91/MIDI!6D	BPW91/DZVP

- Liquid-phase calculations based on gas-phase geometries can be performed with SM5.43 for the following restricted and unrestricted Hartree-Fock, DFT, and adiabatic-

connection-method wave functions (i.e. hybrid DFT wave functions) that employ spherical harmonic or Cartesian *d* functions:

HF/6-31G(d)	B3LYP/6-31G(d)
MPWX/MIDI!	MPWX/MIDI!6D
MPWX/6-31G(d)	MPWX/6-31+G(d)
MPWX/6-31+G(d,p)	

- Liquid-phase calculations based on gas-phase geometries can be performed with SM6, SM8, SM8AD, and SM8T for any choice of electronic structure method and basis set combination for which CM4 charges can be calculated (see above). The SM6 and SM8T models are available only for aqueous solution (IAQU=1).
- For SM5.42, SM5.43, SM6, SM8, and SM8AD liquid-phase analytical gradients are only available for basis sets that use Cartesian *d* shells by default (for instance, 6-31G(d)) or can use Cartesian *d* shells if one specifies the option 6D (see the *Gaussian* manual).
- The availability of liquid-phase analytical gradients allows for efficient geometry optimization of liquid-phase minima and saddle points and for liquid-phase evaluation of numerical Hessians and vibrational frequencies based on liquid-phase analytical gradients.

**IMPORTANT: analytical second-order derivatives are not available in MN-GSM. Always use “freq=numer” for calculation of harmonic frequencies.**

- Numerical gradients are available for all SM5.42, SM5.43, SM6, SM8, and SM8AD wave functions mentioned above, including restricted and unrestricted Hartree-Fock, DFT, and adiabatic-connection-method wave functions (i.e. hybrid DFT wave functions) that employ spherical harmonic or Cartesian *d* functions. These numerical gradients can be used in conjunction with geometry optimization. Gradients are not yet available for SM8T.
- Reaction calculations (i.e., calculations using the IRC keyword) using the generalized Born model and the available SM5.42, SM5.43, SM6, SM8, and SM8AD solvation models can be carried out. Analytical gradients must be available to both the generalized Born model and the solvent model chosen to carry out the reaction path calculation. In addition, since numerical Hessians cannot be calculated with the calcfc keyword, the Hessian for the initial structure used in the calculation must be calculated and stored on the checkpoint file prior to running the reaction path calculation. For more details, see the test suite example test24 given later in this manual as well as the *Gaussian* manual.
- Generalized Born electrostatic solvation energy calculations using Löwdin partial atomic charges are available for all RHF, UHF, DFT, unrestricted DFT, and restricted and unrestricted hybrid-DFT wave functions that use any Cartesian or spherical harmonic basis functions that are supported by *Gaussian 03/09*. For methods using the 6-31+G(d) and 6-31+G(d,p) basis sets, generalized Born electrostatic solvation energy calculations

using redistributed Löwdin population analysis (RLPA) partial atomic charges using restricted and unrestricted HF, DFT, and hybrid DFT methods are also available.

- Analytical gradients of the generalized Born electrostatic solvation energy using Löwdin or RLPA partial atomic charges are available for the methods mentioned above. Where analytical gradients are available, so are geometry optimization of liquid-phase minima and saddle points, and the evaluation of liquid-phase numerical Hessian and vibrational frequencies evaluation based on liquid-phase analytical gradients for these wave functions.
- For methods in which analytical gradients are not supported, geometry optimizations based on numerical gradients are available.

### ***Löwdin Population Analysis and Redistributed Löwdin Population Analysis***

Löwdin population analysis, like Mulliken analysis, provides class II atomic partial charges, but the Löwdin method has certain advantages. We added it to the code because it is the first step in the CM2, CM3, and CM4 charge models, but, having added this capability, there may be some independent interest in Löwdin analysis since it can be used with any basis set (whereas CM2, CM3, and CM4 cannot), and it is expected often to yield more useful population analyses than Mulliken's method.

Partial atomic charges obtained from Löwdin population analysis can, however, be sensitive to basis set size, particularly for extended basis sets that include diffuse functions. We have developed and implemented a new method, called redistributed Löwdin population analysis (or RLPA), which alleviates some of this sensitivity to basis set size. RLPA charges are used as the starting point for CM3 and CM4 charges when the 6-31+G(d) or 6-31+G(d,p) basis sets are used.

### ***Charge Model 2, Charge Model 3, Charge Model 4/4M, Charge Model 5***

Charge Model 2 (CM2), Charge Model 3 (CM3), and Charge Model 4 (CM4) are our second-, third-, and fourth-generation models of class IV charges, respectively. Charge Model 3 has been parameterized with a larger training set than CM2 (398 data vs. 198 data), and it is available for different combinations of electronic structure theory and basis sets than CM2. Furthermore, it is parameterized for Li and for molecules that contain Si-O, Si-F, and Si-Cl bonds (CM2 is not). Charge Model 4 has been parametrized against the same training set that CM3 was, except that CM4 gives improved charges for aliphatic functional groups, which is important for modeling hydrophobic effects. Class IV charges are obtained by a semiempirical mapping of class II charges defined for a particular basis set. CM2, CM3, and CM4 use the same semiempirical mapping scheme. Class IV charges have the following advantages over class III charge models available in *Gaussian* (e.g., ChEIPG and Merz-Kollman algorithms):

- Class III charges are unreliable for buried charges (this problem is widely recognized, as discussed in work by Kollman and Francl and their respective coworkers). Class IV charge models provide a practical and stable way to obtain reasonable charges for buried atoms.

- Class III charges are at best as good as the basis set and wave function used, whereas class IV charges represent extrapolation to full CI with a complete basis.

Class IV charges are useful for any purpose for which ChElPG or Merz-Kollman charges are useful, but we believe that they are better.

The CM4M charge model is an extension of the earlier CM4 model. The CM4M model was individually optimized for the M06 suite of density functionals (see details in *OM07*).

Charge Model 5 is described in detail in *MJ12* and at <http://comp.chem.umn.edu/cm5pac>. The user does not need *MN-GSM* to calculate CM5 charges.

### ***SM5.42, SM5.43, SM6, SM8, SM8AD, SM8T, VEM(d), and SM12 Solvation Models***

SM5.42, our earliest solvation model implemented in *MN-GSM*, is a universal solvation model based on SM5 functional forms for atomic surface tensions (hence the first three characters in the name of the method are SM5), built on class IV point charges (hence .4 comes next) of the CM2 type (hence 2). A more recent model, called SM5.43, uses the same functional forms for atomic surface tensions as does SM5.42, but SM5.43 uses CM3 charges (hence the 3 in the name). The SM6 model is based on SM6 functional forms for atomic surface tensions and uses class IV CM4 point charges. The SM6 has only been parametrized for aqueous solvent.

The SM8 and SM8AD are the most recent universal continuum solvation models where "universal" denotes applicable to all solvents (see *MO07* for more details). With universal models, if desired, one can calculate solvation free energies for two different solvents (*e.g.*, water and 1-octanol) and use the results to calculate  $\log P$ , where  $P$  is the partition coefficient. SM8/SM8AD is applicable to any charged or uncharged solute composed of H, C, N, O, F, Si, P, S, Cl, and/or Br in any solvent or liquid medium for which a few key descriptors are known, in particular dielectric constant, refractive index, bulk surface tension, and acidity and basicity parameters. It may be used with any level of electronic structure theory as long as accurate partial charges can be computed for that level of theory; we recommend using it with self-consistently polarized Charge Model 4 or other self-consistently polarized class IV charges, in which case analytic gradients are available. The cavities for the bulk electrostatics calculation are defined by superpositions of nuclear-centered spheres whose sizes are determined by intrinsic atomic Coulomb radii. The difference between SM8 and SM8AD is that the SM8 model uses the formula of Still *et al.* for the Born radius used in the generalized Born approximation for bulk electrostatics while the SM8AD model utilizes the asymmetric descreening (AD) algorithm for the Born radius suggested by Grycuk. See *MC09* for more detail. The SM8 model is the default model.

The SM8T solvation model is an extension of SM8 to include the temperature dependence of the free energy of solvation relative to 298 K. The SM8T model models the temperature dependence of the solvation free energy using the same functional forms as those in SM8, but with additional terms added to account for temperature dependence (thus, a calculation carried out at 298 K with

the SM8T model will yield the same solvation free energy as the same calculation carried out with SM8). The SM8T model has only been parametrized for aqueous solution.

There was also the SM7 model. The SM7 model is an intermediate model between SM6 and SM8. Like in the case of SM5.42 and SM5.43, the non-bulk electrostatic part of the SM7 model was parametrized to predict solvation free energies in both aqueous and nonaqueous solutions. Unlike SM5.42 and SM5.43, the SM7 model is based on SM6 functional forms for atomic surface tensions and uses class IV CM4 point charges as well as the SM6 model. However, the electrostatic part of the SM7 model is based on the SM6 model's Coulomb radii which were optimized for aqueous solution only. In the new model called SM8, the radii depend on the nature of a solvent. This feature of the SM8 model makes it more accurate than SM7 when there is a need to calculate solvation energies in nonaqueous solutions. Thus, we skip the SM7 model hereafter.

There was also the SM6T model. The SM6T model is an extension of SM6 to include the temperature dependence of the free energy of solvation relative to 298 K. When the SM8 model came into existence, the old temperature-dependent terms from SM6T were augmented with a few new ones and the SM8T merged the SM6T. Since the SM8T model has some additional functionalities, we opt to skip the more inferior SM6T model hereafter.

The SMx solvation models provide a way to calculate electronic wave functions in liquid-phase solution and free energies of solvation. For solvation calculations based on gas-phase geometries, the standard-state free energy of solvation  $\Delta G_S^0(\mathbf{R})$  is given by two components:

$$\Delta G_S^0(\mathbf{R}) = \Delta G_{EP} + G_{CDS} \quad (1)$$

where

$$\Delta G_{EP} = \Delta E_E + G_P \quad (2)$$

In equations (1) and (2),  $\Delta G_{EP}$  is the bulk electrostatic component of the solvation free energy; it is the sum of the polarization energy  $G_P$  (representing favorable solute-solvent interactions and the associated solvent rearrangement cost) and the distortion energy  $\Delta E_E$  (the cost of distorting the solute electronic charge distribution to be self-consistent with the solvent electric polarization).  $G_{CDS}$  accounts for first-solvation-shell effects.

$\Delta G_{EP}$  is determined by a self-consistent reaction field (SCRF) calculation, which allows the solvent-induced change in the solute electronic wave function to be optimized variationally.  $G_{CDS}$  is not a self-consistent term; it has no effect on the solute electronic wave function. In its simplest form  $G_{CDS}$  is defined as:

$$G_{CDS} = \sum_k A_k \sigma_k \quad (3)$$

where  $A_k$  is the exposed surface area of atom  $k$  (this depends on the solute's 3-D geometry and is calculated by the Analytical Surface Area (ASA) algorithm as described in reference *LH95* and as included in recent versions of *AMSOL*, and  $\sigma_k$  is the atomic surface tension of atom  $k$ . The atomic surface tension  $\sigma_k$  is itself a function of the solute's 3-D geometry and a small set of solvent descriptors. References *LH98*, *ZL98*, and *LZ99* present a more expanded form of  $G_{\text{CDS}}$  than what appears in equation (3):

$$G_{\text{CDS}} = \sum_k \sigma_k^A A_k(\{r_k; r_S^{\text{CD}}\}) + \sigma^{\text{CS}} \sum_k A_k(\{r_k; r_S^{\text{CS}}\}) \quad (4)$$

where  $k$  represents atoms,  $\sigma_k^A$  is the partial atomic surface tension for atom  $k$ , and  $A_k(\{r_k; r_S^{\text{CX}}\})$ , with  $X = \text{C or D}$ , is the solvent-accessible surface area of atom  $k$  that is dependent upon a set of parameterized set or effective solute radii,  $\{r_k\}$ , and on one or another effective solvent radii,  $r_S^{\text{CD}}$  and  $r_S^{\text{CS}}$ . Although  $r_S^{\text{CD}}$  and  $r_S^{\text{CS}}$  have different nonzero values for some solvation models, they are both set to zero in the SM5.42 model. SM5.43 and SM6 use 0.40 Å for both of these values. Each  $\sigma_k^A$  is a function of one-, two-, and/or three-atom surface tension coefficients,  $\tilde{\sigma} \in \{\tilde{\sigma}_{z_k}, \tilde{\sigma}_{z_k z_{k'}}, \tilde{\sigma}_{z_k z_{k'} z_{k''}}\}$ , that depend on the atomic numbers  $z_k, z_{k'}$ , and  $z_{k''}$  of atoms  $k, k'$ , and  $k''$ , respectively, and the switching functions,  $f_{z_k z_{k'}}(R_{kk'})$  and  $f_{z_k z_{k'} z_{k''}}(R_{kk'}, R_{k'k''})$ , that are functions of interatomic distances,  $R_{kk'}$  and  $R_{k'k''}$ :

$$\sigma_k = \tilde{\sigma}_{z_k} + \sum_{k'} \tilde{\sigma}_{z_k z_{k'}} f_{z_k z_{k'}}(R_{kk'}) + \sum_{k', k''} \tilde{\sigma}_{z_k z_{k'} z_{k''}} f_{z_k z_{k'} z_{k''}}(R_{kk'}, R_{k'k''}) \quad (5)$$

For water solvent,  $\sigma^{\text{CS}}$  is absorbed into  $\sigma_k^A$ , and a set of  $\tilde{\sigma}_i \in \{\tilde{\sigma}_{z_k}, \tilde{\sigma}_{z_k z_{k'}}, \tilde{\sigma}_{z_k z_{k'} z_{k''}}\}$  is specifically determined constants (similar to  $\bar{\sigma}_k^A$ , which will be defined below).

For a particular organic solvent, each  $\tilde{\sigma}_i$  (with  $i = z_k, z_{k'}$ , or  $z_{k''}$ ) and  $\sigma^{\text{CS}}$  are defined to be linear functions of bulk solvent properties:

$$\tilde{\sigma}_i = \hat{\sigma}_i^{(n)} n + \hat{\sigma}_i^{(\alpha)} \alpha + \hat{\sigma}_i^{(\beta)} \beta \quad (6)$$

$$\sigma^{\text{CS}} = \hat{\sigma}^{(\gamma)} \gamma + \hat{\sigma}^{(\beta^2)} \beta^2 + \hat{\sigma}^{(\phi^2)} \phi^2 + \hat{\sigma}^{(\psi^2)} \psi^2 \quad (7)$$

where  $n, \alpha, \beta, \gamma, \phi$ , and  $\psi$  refer to the index of refraction, Abraham's hydrogen bond acidity and hydrogen bond basicity, macroscopic surface tension, electronegative halogenicity, and aromaticity, respectively. The constants  $\hat{\sigma}_i^{(n)}, \hat{\sigma}_i^{(\alpha)}, \hat{\sigma}_i^{(\beta)}, \hat{\sigma}^{(\gamma)}, \hat{\sigma}^{(\beta^2)}, \hat{\sigma}^{(\phi^2)}$ , and  $\hat{\sigma}^{(\psi^2)}$  have

been optimized using a large number of solvent/solute combinations for which experimental solvation energies are available.

The surface tension functional forms are the same in all SM5.42 and SM5.43 models. The models SM6-SM8 use a different set of functional forms. The SM6-SM8 functional forms are better for most purposes than those used in SM1–SM5.

The vertical excitation model VEM(d) is described in detail in *MC11*, and the SM12 solvation model is described in *MC12*. See also special sections in this document on how to run these models using *MN-GSM*.

### Incorporating temperature dependence into the SMx models: SM8T

To account for the variation of the free energy of solvation as a function of temperature, the temperature dependence of both the bulk electrostatics,  $\Delta G_{EP}$ , and the non-bulk electrostatic,  $\Delta G_{CDS}$ , contributions are included. The effect of temperature on the bulk-electrostatic contributions to the free energy of solvation is accounted for using a temperature dependent dielectric constant,  $\varepsilon(T)$  which was computed using the following equation

$$\varepsilon(T) = 249.21 - .79T + .00072T^2 \quad (8)$$

where  $T$  is the temperature of the aqueous solvent. This is an empirically derived equation found in the *CRC Handbook of Chemistry and Physics 76<sup>th</sup> edition*, ed. Lide, D. R., 1995, CRC Press, New York. The variation of the free energy of solvation due to bulk electrostatic contributions are quite small. The majority of the temperature dependence of aqueous free energies of solvation must be accounted for using  $\Delta G_{CDS}$ .

In SM8T the  $\Delta G_{CDS}$  term mimics the thermodynamic equation for the temperature dependence of free energies of solvation where the thermodynamic properties, the heat capacity and the entropy of solvation, have been replaced by parameterized atomic surface tensions:

$$G_{CDS}(T) = (T - 298) \sum_k A_k \sigma_k^B + \left( T - 298 - T \ln \frac{T}{298} \right) \sum_k A_k \sigma_k^C \quad (9)$$

where  $\sigma_k^B$  and  $\sigma_k^C$  are atomic surface tensions with identical functional forms to those of  $\sigma_k$ . Caution should be used in assigning any physical meaning to the atomic surface tensions shown above. Note that the model has only been developed for C, H, and O containing compounds in aqueous solutions for the temperature 273 to 373 K.

### A comment on using gas-phase geometries to calculate solvation free energies.

For SM1–4 and SM5.4, geometry optimization in solution was an essential part of the parameterization. SM5.42, SM5.43, SM6, and SM8/SM8T are parameterized in such a way that one fixes the geometry at a reasonable value (any reasonably accurate gas-phase geometry should be acceptable) and calculates the solvation energy without changing the geometry. Thus, geometry optimization in the presence of solvent is not required to obtain accurate solvation free energies. This method of obtaining solvation parameters based on gas-phase geometries was

adopted for several reasons. First, previous experience has shown that the difference one gets from re-optimizing the geometry in the presence of solvent in *almost* all cases is small—less than the average uncertainty in the method or in any competing method. Second, for many solutes, less expensive (e.g. semiempirical or molecular mechanics methods) can yield accurate gas-phase geometries. Third, for other solutes, such as transition states, solutes with low-barrier torsions, multiple low-energy conformations, weakly bound complexes, and in cases where one or more solvent molecules are treated explicitly, more expensive levels of theory might be needed to yield accurate geometries. Finally, solvation energies obtained using gas-phase geometries can be added conveniently to gas-phase energies for separable-equilibrium-solvation dynamics calculations.

In some cases, geometry optimization in the presence of solvent is important. In these cases, one can also apply the SM5.42, SM5.43, SM6, or SM8 models at a solute geometry  $\mathbf{R}$  that is not an approximation to an equilibrium gas-phase geometry. This type of calculation corresponds to the fixed- $\mathbf{R}$  solvation energy, which is still given by  $\Delta G_s^\circ(\mathbf{R})$  of equation (1). Evaluation of this quantity for geometries that do not correspond to an equilibrium structure is useful for dynamics calculations because the potential of mean force is given by

$$W(\mathbf{R}) = V(\mathbf{R}) + \Delta G_s^\circ(\mathbf{R}) \quad (10)$$

where  $V$  is the gas-phase potential energy surface (which is itself given by the sum of the gas-phase electronic energy and the gas-phase nuclear repulsion energy). If one applies the SM5.42, SM5.43, SM6, or SM8 models to a geometry optimized in solution and subtracts the gas-phase energy at a geometry optimized for the gas phase, one obtains the true solvation energy for the given method.

Note that both free energies of solvation and potentials of mean force depend on temperature; all of our solvation models except SM8T are currently only parameterized for 298.15 K.

$\Delta G_s^\circ(\mathbf{R})$  depends on standard state choices; the values given directly by the SM5, SM6, and SM8 models correspond to using the same molar density (e.g., one mole per liter) in the gas phase and in the liquid-phase solution. The liquid-solution standard state corresponds to an ideal dilute solution at that concentration. One may adjust the results to correspond to other choices of standard state by standard thermodynamic formulae. Changing the standard state corresponds to adding a constant to  $W$ ; thus the gradient of  $W$ , which is used for dynamics, is not affected.

### ***Allowed Combinations of Solvent Model, Electronic Structure Theory, and Basis Set***

When using MN-GSM, the SM5.42 models should only be applied with the ten combinations of basis sets and density functionals for which parameters are currently available in *Gaussian*:

<i>Standard notation</i>	<i>Gaussian keyword(s)</i>	<i>notes</i>
HF/MIDI!	HF/MIDIX	
HF/MIDI!6D	HF/MIDIX 6D	

HF/6-31G(d)	HF/6-31G(d)	
BPW91/MIDI!	BPW91/MIDIX	
BPW91/MIDI!6D	BPW91/MIDIX 6D	
B3LYP/MIDI!	B3LYP/MIDIX	
BPW91/6-31G(d)	BPW91/6-31G(d)	
HF/6-31+G(d)	HF/6-31+G(d)	
HF/cc-pVDZ	HF/GEN 5D	<i>a</i>
BPW91/DZVP	BPW91/GEN 6D	<i>b</i>

<sup>a</sup> See “Special notes on basis sets” section <sup>b</sup> Note that the DZVP basis set must be provided as a general basis

The SM6 model has been tested against several different density functionals, and has been shown to retain its accuracy when different density functionals besides MPWX (the method against which the CM4 and SM6 parameters were originally developed). Thus, the SM6 model is only basis-set-dependent. Currently, SM6 parameters exist for the following four basis sets:

MIDI!6D  
6-31G(d)  
6-31+G(d)  
6-31+G(d,p)

The CM4/CM4M parameters exist for additional basis sets:

MIDI!  
6-31G(d,p)  
cc-pVDZ  
DZPV  
6-31B(d)  
6-31B(d,p)

A list of density functionals that are available in *Gaussian* and that are recommended for use with CM4 and/or SM6 are listed in the section entitled “Density Functional Methods Available in *Gaussian* and *MN-GFM*”.

There is a single set of the SM8/SM8AD/SM12 parameters (radii and CDS terms) that can be used with any basis set as long as accurate partial charges can be computed for that basis set.

The SM8T model is applicable to the same combinations of theory and basis set as SM8, but it has been parametrized only for aqueous solution.

#### Why Use SMx?

- The semiempirical CDS terms make the above models more accurate than alternative models for absolute free energies of solvation of neutral solutes.
- SM5.42, SM5.43, SM8, SM8AD, and SM12 are universal models, *i.e.*, the semiempirical parameters are adjusted for water and for all solvents for which a small number of required solvent descriptors are known or can be estimated; this includes essentially any organic solvent.

- SM5.42, SM5.43, SM6, SM8, and SM8AD use class IV charges to calculate the bulk electrostatic contribution to the solvation free energy; this is typically more accurate than calculating the charge distribution directly from the approximate wave function. This has two consequences:
  - (1) The electrostatics are more accurate.
  - (2) CM2, CM3, and CM4/CM4M yield very accurate charges both in the gas phase and in liquid-phase solutions, and this is useful for a qualitative understanding of solvent-induced changes in the solute.
- SM5, SM6, SM8, SM8AD, and SM12 parameterizations included an extremely broad range of solute functional groups, including molecules containing phosphorus, which are very hard to treat.
- SM5, SM6, SM8, SM8AD, and SM12 do not need to be corrected for outlying charge error, which can be large in some other methods.

#### Why Use SM8T?

- SM8T is one of the few implicit models capable predicting aqueous free energies of solvation as a function of temperature.
- SM8T has been tested against a much broader range of compounds than prior models.
- SM8T is based on SM8, and thus shares its advantages (except the SM8 applicability to nonaqueous solutions).

### ***Analytical Gradients and Geometry Optimization***

No analytical gradients are available for SM8T, SM12 and VEM(d). Analytical gradients for other models are available.

Analytical gradients of SM5.42, SM5.43, and GB liquid-phase wave functions have been implemented in *MN-GSM* beginning with version 3.0. In particular, *MN-GSM* contains analytical gradients for restricted and unrestricted wave functions and for basis sets that employ Cartesian basis functions. However, analytical gradients are still not available for SCF scheme III, and for basis sets with spherical harmonic *d* and *f* functions.

The availability of gradients allows for efficient geometry optimization in liquid-phase solution. This is necessary in many cases. For example, the transition state geometry for the S<sub>N</sub>2 reaction of ammonia and chloromethane (the Menshutkin reaction) depends strongly on solvent. Other applications include the study of phase-dependent reaction mechanisms and molecular conformational preferences.

Equation (1) can be rewritten as:

$$\Delta G_s^\circ(\mathbf{R}) = G^\circ(l, R_e(g)) - G^\circ(g, R_e(g)) \quad (11)$$

where  $\Delta G_s^\circ(\mathbf{R})$  is the difference in the standard-state free energy of a molecular species fixed in its gas-phase equilibrium geometry,  $R_e(g)$ , in liquid-phase solution, *l*, and in the gas phase, *g*. Furthermore,

$$G^\circ(l, R_e(g)) = E(l, R_e(g)) + G_p(R_e(g)) + G_{\text{CDS}}(R_e(g)) \quad (12)$$

and

$$G^\circ(g, R_e(g)) = E(g, R_e(g)) \quad (13)$$

where  $E(l, R_e(g))$  is the SCRF energy in the presence of the solvent, while  $E(g, R_e(g))$  is the gas-phase SCF energy.

Geometry optimization in solution yields the solute equilibrium geometry,  $R_e(l)$ .

In the SM5.42, SM5.43, SM6, and SM8 models, the definition of  $\Delta G_s^\circ$  is analogous to the definition of  $\Delta G_s^\circ(\mathbf{R})$  in equation (11):

$$\begin{aligned} \Delta G_s^\circ &= G^\circ(l, R_e(l)) - G^\circ(g, R_e(g)) \\ &= E(l, R_e(l)) + G_p(R_e(l)) + G_{\text{CDS}}(R_e(l)) - E(g, R_e(g)) \end{aligned} \quad (14)$$

A full derivation of the analytical gradient is presented in the paper by T. Zhu *et al.* entitled “Analytical Gradients of a Self-Consistent Reaction-Field Solvation Model Based on CM2 Atomic Charges” (*J. Chem. Phys.* **1999**, *110*, 5503-5513).

## ***Literature References***

### ASA algorithm

*LH95* Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. “Improved Methods for Semiempirical Solvation Models,” *J. Comp. Chem.* **1995**, *16*, 422.

### MIDI! Basis set

*EG96* Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. “The MIDI! Basis Set for Quantum Mechanical Calculations of Molecular Geometries and Partial Charges,” *Theor. Chim. Acta* **1996**, *93*, 281.

*LC98* Li, J.; Cramer, C. J.; Truhlar, D. G. “MIDI! Basis Set for Silicon, Bromine, and Iodine,” *Theor. Chem. Acc.* **1998**, *99*, 192.

*TW01* Thompson, J. D.; Winget, P.; Truhlar, D. G. “MIDIX Basis Set for the Lithium Atom: Accurate Geometries and Partial Atomic Charges with Minimal Computational Cost,” *PhysChemComm* **2001**, *4*, 72.

### Löwdin population analysis

*LD50* Löwdin, P. –O. *J. Chem. Phys.* **1950**, *18*, 365. (b) Del Re, G.; In *Quantum Science*; Calais, J. –L., Goscinski, O., Linderberg, J., Öhrn, Y., Eds.; Plenum Press: New York, 1976; p 53.

### Redistributed Löwdin population analysis

TX02 Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. "More Reliable Partial Atomic Charges when using Diffuse Basis Sets," *PhysChemComm* **2002**, *5*, 117.

### Original parameterizations of the CM2 Charge Model

LZ98 Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. "A New Class IV Charge Model for Extracting Accurate Partial Charges from Wave Functions," *J. Phys. Chem. A* **1998**, *102*, 1820.

### Parameterization of the CM2 model for HF/cc-pVDZ

LX99 Li, J.; Xing, J.; Cramer, C. J.; Truhlar, D. G. "Accurate Dipole Moments from Hartree-Fock Calculations by Means of Class IV Charges," *J. Chem. Phys.* **1999**, *111*, 885.

### Original parameterizations of the CM3 Charge Model

WT02 Winget, P.; Thompson J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. "Charge Model 3: A Class IV Charge Model based on Hybrid Density Functional Theory with Variable Exchange," *J. Phys. Chem. A*, **2002**, *106*, 10707.

### Extension of CM3 to AM1, PM3, BLYP, and B3LYP

TC03 Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. "Parameterization of Charge Model 3 For AM1, PM3, BLYP, and B3LYP," *J. Comput. Chem.*, **2003**, *24*, 1291.

### Extension of CM3 and CM3.1 to HF/MIDI!

KC05a Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. "Accurate Partial Atomic Charges for High-Energy Molecules with the MIDI! Basis Set," *Theor. Chem. Acc.* **2005**, *113*, 133.

### Parameterizations of the CM4 Charge Model

KC05b Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. "SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters," *J. Chem. Theory Comput.* **2005**, *1*, 1133.

### Parameterizations of the CM4M Charge Model

OM07 Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. "Charge Model 4 and intramolecular charge polarization," *J. Chem. Theory Comput.* **2007**, *3*, 2046.

### SM5.42/DFT method and ISCRF=1 option

ZL98 Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. "Density Functional Solvation Model Based on CM2 Atomic Charges," *J. Chem. Phys.* **1998**, *109*, 9117.

SM5.42/MIDI!6D method and more efficient version of ISCRF=1 option

LH98 Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. "Universal Reaction Field Model Based on *Ab Initio* Hartree-Fock Theory," *Chem. Phys. Lett.* **1998**, 288, 293.

Additional SM5.42 parameterizations and ISCRF=2 option

LZ99 Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. "Extension of the Platform of Applicability of the SM5.42R Universal Solvation Model," *Theor. Chem. Acc.* **1999**, *103*, 9.

SM5.42 parameterization for Si

WT02 Winget, P.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. "Parameterization of a Universal Solvation Model for Molecules Containing Silicon," *J. Phys. Chem. B*, **2002**, *106*, 5160.

SM5.43 parameterizations for HF/6-31G(d), B3LYP/6-31G(d), mPW1PW91/6-31G(d), and mPW1PW91/6-31+G(d)

TC04 Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. "New Universal Solvation Model and Comparison of the Accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM Continuum Solvation Models for Aqueous and Organic Solvation Free Energies and Vapor Pressures," *J. Phys. Chem. A* **2004**, *108*, 6532.

SM5.43 parameterizations for MPWX/MIDIX, MPWX/MIDI!6D, MPWX/6-31G(d), MPWX/6-31+G(d), and MPWX/6-31+G(d,p)

TC05 Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. "Density-Functional Theory and Hybrid Density-Functional Theory Continuum Solvation Models for Aqueous and Organic Solvents: Universal SM5.43 and SM5.43R Solvation Models for Any Fraction of Hartree-Fock Exchange," *Theor. Chem. Acc.* **2005**, *113*, 107.

SM6 aqueous model

KC05b Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. "SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters," *J. Chem. Theory Comput.* **2005**, *1*, 1133.

SM8 universal model

MO07 Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. "Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges," *J. Chem. Theory Comput.* **2007**, *3*, 2011.

#### SM8AD universal model

MC09 Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. "Universal Solvation Model Based on the Generalized Born Approximation with Asymmetric Descreening," *J. Chem. Theory Comput.* **2009**, *5*, 2447.

#### Temperature-dependent aqueous models

AC06 Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. "Predicting Free Energies of Solvation as Functions of Temperature," *J. Phys. Chem. B* **2006**, *110*, 5665.

AC08 Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. "Extension of a Temperature-Dependent Aqueous Solvation Model to Compounds Containing Nitrogen, Fluorine, Chlorine, Bromine, and Sulfur," *J. Phys. Chem. B* **2008**, *112*, 3024.

#### Analytical gradients of the SM5.42, SM5.43, SM6, and SM8 models

ZL99 Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. "Analytical Gradients of a Self-Consistent Reaction-Field Solvation Model Based on CM2 Atomic Charges," *J. Chem. Phys.* **1999**, *110*, 5503.

#### VEM(d) vertical excitation model

MC11 Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Guido, G. A.; Mennucci, B.; Scalmani, G.; Frisch, M. J. "Practical Computation of Electronic Excitation in Solution: Vertical Excitation Model," *Chem. Sci.* **2011**, *2*, 2143.

#### SM12 model

MC12 Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. "Generalized Born Solvation Model SM12," *J. Chem. Theory Comput.* **2012**, in press.

#### CM5 charge model

MJ12 Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. "Charge Model 5: an Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases," *J. Chem. Theory Comput.* **2012**, *8*, 527.

#### Bondi-type intrinsic atomic radii

MV10 Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. In *CRC Handbook of Chemistry and Physics*, 91st ed. (2010–2011); Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, **2010**; p 9-49.

## Quick index to literature

The references for the methods supported by *MN-GSM* are as follows:

<b>Method</b>	<b>Reference</b>
Löwdin charges	<i>LD50</i>
RLPA charges	<i>TX02</i>
CM2/HF/MIDI!	<i>LZ98</i>
CM2/HF/MIDI!6D	<i>LZ98</i>
CM2/HF/6-31G(d)	<i>LZ98</i>
CM2/BPW91/MIDI!	<i>LZ98</i>
CM2/BPW91/MIDI!6D	<i>LZ98</i>
CM2/B3LYP/MIDI!	<i>LZ98</i>
CM2/BPW91/6-31G(d)	<i>LZ98</i>
CM2/HF/6-31+G(d)	<i>LZ98</i>
CM2/HF/cc-pVDZ	<i>LX99</i>
CM2/BPW91/DZVP	<i>LZ98</i>
CM3/HF/MIDI!	<i>KC05a</i>
CM3/HF/MIDI!6D	<i>WT02</i>
CM3/HF/6-31G(d)	<i>WT02</i>
CM3/MPWX/MIDI!	<i>WT02</i>
CM3/MPWX/MIDI!6D	<i>WT02</i>
CM3/MPWX/6-31G(d)	<i>WT02</i>
CM3/MPWX/6-31+G(d)	<i>WT02</i>
CM3/MPWX/6-31+G(d,p)	<i>WT02</i>
CM3/BLYP/6-31G(d)	<i>TC03</i>
CM3/B3LYP/MIDI!6D	<i>TC03</i>
CM3/B3LYP/6-31G(d)	<i>TC03</i>
CM3/B3LYP/6-31+G(d)	<i>TC03</i>
CM3.1/HF/MIDI!	<i>KC05a</i>
CM4/DFT/MIDI!6D	<i>KC05b</i>
CM4/DFT/6-31G(d)	<i>KC05b</i>
CM4/DFT/6-31+G(d)	<i>KC05b</i>
CM4/DFT/6-31+G(d,p)	<i>KC05b</i>
CM4/DFT/MIDI!	<i>OM07</i>
CM4/DFT/6-31G(d,p)	<i>OM07</i>
CM4/DFT/cc-pVDZ	<i>OM07</i>
CM4/DFT/DZVP	<i>OM07</i>
CM4/DFT/6-31B(d)	<i>OM07</i>
CM4/DFT/6-31B(d,p)	<i>OM07</i>
CM4M/M06/MIDI!	<i>OM07</i>
CM4M/M06/MIDI!6D	<i>OM07</i>
CM4M/M06/6-31G(d)	<i>OM07</i>
CM4M/M06/6-31+G(d)	<i>OM07</i>

CM4M/M06/6-31+G(d,p)	OM07
CM4M/M06/6-31G(d,p)	OM07
CM4M/M06/cc-pVDZ	OM07
CM4M/M06/DZVP	OM07
CM4M/M06/6-31B(d)	OM07
CM4M/M06/6-31B(d,p)	OM07
SM5.42/HF/MIDI!	LZ99
SM5.42/HF/MIDI!6D	LH98
SM5.42/HF/6-31G(d)	LZ99
SM5.42/BPW91/MIDI!	LZ99
SM5.42/BPW91/MIDI!6D	ZL98
SM5.42/B3LYP/MIDI!	LZ99
SM5.42/BPW91/6-31G(d)	ZL98
SM5.42/HF/6-31+G(d)	LZ99
SM5.42/HF/cc-pVDZ	LZ99
SM5.42/BPW91/DZVP	ZL98
SM5.43/HF/6-31G(d)	TC04
SM5.43/B3LYP/6-31G(d)	TC04
SM5.43/MPWX/MIDI!	TC05
SM5.43/MPWX/MIDI!6D	TC05
SM5.43/MPWX/6-31G(d)	TC05
SM5.43/MPWX/6-31+G(d)	TC05
SM5.43/MPWX/6-31+G(d,p)	TC05
SM6/DFT/MIDI!6D	KC05b
SM6/DFT/6-31G(d)	KC05b
SM6/DFT/6-31+G(d)	KC05b
SM6/DFT/6-31+G(d,p)	KC05b
SM8	MO07
SM8AD	MC09
SM8T	AC06, AC08
SMx liquid-phase optimizations	ZL99 and CR99
VEM(d)	MC11
SM12	MC12

### ***Usage, Required Iop Statements, and \$MNGSM Namelist***

*MN-GSM* allows users to perform various types of calculations. The mechanics of setting up these calculations are described here.

All *MN-GSM* calculations require the inclusion of “Iop(3/70=3000)” (the “Iop” in Iop statements is not case sensitive in *Gaussian*) in the route section of the *Gaussian* input file. To request an analytic energy gradient calculation and/or a numerical Hessian and frequencies calculation, the “Iop(7/88=3000)” keyword is also required. These input-output parameter flags turn on the solvation calculation in the appropriate links in *Gaussian*.

The specifics of a particular calculation in *MN-GSM* are determined with the \$MNGSM namelist, which is identified by a line beginning with “\$MNGSM” (note that there is a required blank space before the “\$” symbol). Namelist input may span one or more lines, and is closed by an “\$END” (which, if placed on a line by itself, also requires a blank space before the “\$” symbol) statement. These namelists must be placed after a blank line following the last line of the regular *Gaussian* input file.

The remainder of this *Usage* section is subdivided into nine subsections as described below:

The \$MNGSM namelist input can accept the following 30 keywords:

AtmProp	HFE	JobName	SolC
Border	IAQU	LCoulR	SolG
CDSRead	ICDS	MeanK	SolH
ChgRead	ICMD	Mulliken	SolK
CMxSave	IGAS	NOPOL	SolN
DelEMin	IGRD	ReadK	SolvRd
Dielec	IRADII	SolA	
ETGAS	ISCRF	SolB	

The functions of all 30 keywords are presented in alphabetical order in the subsections entitled *All keywords*.

The use of the four keywords AtmProp, CDSRead, DelEMin, and LCoulR require more in-depth descriptions. Therefore, the subsections entitled *Extended description of AtmProp*, *Extended description of CDSRead*, *Extended description of DelEMin*, and *Extended description of LcoulR* discuss the function of these keywords in more detail.

The subsection entitled *Summary of keywords and their default values*, where the 30 keywords and their default values are tabulated, follows the extended description subsections.

Next, the subsection entitled *Substituting Gaussian links* describes how the *Gaussian* links that have been modified to implement *MN-GSM* are used to run *MN-GSM* calculations.

The final subsection is entitled *Important notes*, and presents a number of points concerning the running of *MN-GSM* calculations that have not been addressed elsewhere in this manual.

## ***All Keywords***

**IMPORTANT: MN-GSM uses logical variables. Some compilers will not correctly read in input if the logicals in the input do not have periods on either side of the value, i. e. ReadK=TRUE may cause problems whereas ReadK=.True. or ReadK=.T. should not.**

**AtmProp** If AtmProp=True, then atomic properties from the external file *JobName.atmp* are read in (default: AtmProp=False; see below for a description of the JobName keyword). Allowed atomic properties include Coulomb and van der Waals radii

and the number of frozen-core electrons. The format of the external file input is described below in the subsection *Extended description of AtmProp*.

- Border** If Border=True, then the gas-phase and/or liquid-phase Mayer bond order matrices are written to the output file (default: Border=False).
- CDSRead** If CDSRead=True, then the surface tension coefficients are read in from the external file *JobName.cds* (default: CDSRead=False; see below for a description of the JobName keyword). The format of the external file input is described below in the subsection *Extended description of CDSRead*.
- ChgRead** If ChgRead=True, then the partial atomic charges are read in (in free-format style) from the external file *JobName.Chg* (default: ChgRead=False; see below for a description of the JobName keyword). Also, ISCRF is automatically set to 3, and these charges are used in the evaluation of  $\Delta G_{EP}$ .
- CMxSave** If CMxSave=True, then the CM2, CM3, or CM4 partial atomic charges are written (in free-format style) to the external file *JobName.chg* (default: CMxSave=False; see below for the JobName keyword).
- DeLEMin** Manages the transition between the gas-phase SCF and the liquid-phase SCRF (default is 0.000001 hartrees). A detailed explanation of this transition and the utility of the DeLEMin keyword are presented below in the subsection *Extended description of DeLEMin*.
- Dielec** Dielectric constant,  $\epsilon$ , of solvent. When IAQU is set equal to 1, the default for Dielec is 78.3 for water. However, Dielec can be set to an arbitrary value. When IAQU is set equal to 0, there is not default for Dielec, and it must be provided in this case.
- ETGAS** Accepts a value provided as input for the gas-phase energy (in hartrees). The ETGAS keyword must be defined for IGAS = 2 or 3; the calculation will stop otherwise. Note that if IGAS = 0 or 1, and a value for ETGAS is provided, it will be ignored. Theoretically, the gas-phase SCF energy used in the evaluation of the free energy of solvation should correspond to the gas-phase equilibrium geometry. Supplying an energy for a non-equilibrium structure or using the IGAS = 1 option will lead to non-standard results.
- HFE** Defines the fraction of Hartree-Fock (HF) exchange (0.000 – 0.999) to be used when ICMD = 315 to ICMD = 319 or ICMD = 415 to ICMD = 510. For example, if a calculation of CM3 charges by the mPW1PW91/MIDI! method were to be carried out, then ICMD would be set to 315, and HFE would be set to 0.25. The HFE keyword only defines the fraction of HF exchange for a given calculation in *MN-GSM*, not in the entire *Gaussian* program. The fraction of HF exchange is defined in the entire program using the Iop(3/76) flag (which is described fully in the *Gaussian* manual). For example, if a calculation of CM3 charges were to be carried out by the MPW42.8/6-31+G(d,p) method, then ICMD would be set to 319 and HFE would be set to 0.428 the \$MNGSM namelist,

and appropriate definitions of the Iop(3/76) flag would have to be included in the route section of the input. This keyword is only active when ICMD = 315 – 319 or ICMD = 415 – 510, and in these instances HFE must be explicitly specified by the user in the \$MNGSM namelist. For use with ICMD  $\geq$  415 and the Hartree-Fock method, the value of HFE should be set to 0.999. **Note that, starting with MN-GSM version 2011, this option is no longer required because the program will try to determine the HFE value automatically based on Gaussian's input.**

- IAQU** Determines the solvent type:
- IAQU = 1 invokes the solvent model optimized specifically for aqueous solvent. This is the default if none of the solvent descriptors (SolN, SolA, SolB, SolG, SolC, and SolH) have been defined in the namelist. See below for a complete description of the solvent descriptor keywords.
  - IAQU = 0 invokes the general organic solvent model. This is the default if any of SolN, SolA, SolB, SolG, SolC, or SolH have been provided as input.

For a desired solvent, these values can be derived from experiment or from interpolation or extrapolation of data available for other solvents. Solvent parameters for common organic solvents are tabulated in the *Minnesota Solvent Descriptor Database*. The latest version of this database is available at: <http://comp.chem.umn.edu/solvation>.

- ICDS** Selects the set of coefficients that will be used to evaluate the atomic surface tensions. The table above matches the value of ICDS with the wave function for which the corresponding CDS parameter set has been optimized. ICDS = 800 is the default.

Method	ICDS
SM5.42/HF/MIDI!	1
SM5.42/HF/MIDI!6D	2
SM5.42/HF/6-31G(d)	3
SM5.42/HF/6-31+G(d)	8
SM5.42/HF/cc-pVDZ	9
SM5.42/BPW91/MIDI!	4
SM5.42/BPW91/MIDI!6D	5
SM5.42/BPW91/6-31G(d)	7
SM5.42/BPW91/DZVP	10
SM5.42/B3LYP/MIDI!	6
SM5.43/HF/6-31G(d)	303
SM5.43/B3LYP/6-31G(d)	313
SM5.43/MPWX/MIDI!	315
SM5.43/MPWX/MIDI!6D	316
SM5.43/MPWX/6-31G(d)	317
SM5.43/MPWX/6-31+G(d)	318
SM5.43/MPWX/6-31+G(d,p)	319
SM6/DFT/MIDI!6D	416
SM6/DFT/6-31G(d)	417

SM6/DFT/6-31+G(d)	418
SM6/DFT/6-31+G(d,p)	419
SM8 (default)	800
SM8AD	801

**ICMD** Selects the set of CM2, CM3, or CM4/CM4M coefficients used for the evaluation of the CM2 or CM3 or CM4/CM4M charges. Always required.

<b>ICMD</b>	<b>Wave function which corresponds to set of coefficients</b>
0	All coefficients equal zero, calculate Löwdin charges
1	CM2/HF/MIDI! (default for ICMD)
2	CM2/HF/MIDI!6D
3	CM2/HF/6-31G(d)
4	CM2/BPW91/MIDI!
5	CM2/BPW91/MIDI!6D
6	CM2/B3LYP/MIDI!
7	CM2/BPW91/6-31G(d)
8	CM2/HF/6-31+G(d)
9	CM2/HF/cc-pVDZ
10	CM2/HF/DZVP
300	Calculate RLPA charges
301	CM3/HF/MIDI!
302	CM3/HF/MIDI!6D
303	CM3/HF/6-31G(d)
313	CM3/B3LYP/6-31G(d)
314	CM3/B3LYP/6-31+G(d)
315	CM3/MPW $X$ /MIDI!
316	CM3/MPW $X$ /MIDI!6D
317	CM3/MPW $X$ /6-31G(d)
318	CM3/MPW $X$ /6-31+G(d), uses RLPA charges
319	CM3/MPW $X$ /6-31+G(d,p), uses RLPA charges
320	CM3/B3LYP/MIDI!6D
321	CM3/BLYP/6-31G(d)
322	CM3.1/HF/MIDI!
415	CM4/DFT/MIDI!
416	CM4/DFT/MIDI!6D
417	CM4/DFT/6-31G(d)
418	CM4/DFT/6-31+G(d), uses RLPA charges
419	CM4/DFT/6-31+G(d,p), uses RLPA charges
420	CM4/DFT/6-31G(d,p)
421	n/a
422	CM4/DFT/cc-pVDZ_g03
423	CM4/DFT/DZVP
424	CM4/DFT/6-31B(d)
425	CM4/DFT/6-31B(d,p)

500	CM4M/M06/MIDI!
501	CM4M/M06/MIDI!6D
502	CM4M/M06/6-31G(d)
503	CM4M/M06/6-31+G(d)
504	CM4M/M06/6-31+G(d,p)
505	CM4M/M06/6-31G(d,p)
506	CM4M/M06/cc-pVDZ
507	CM4M/M06/cc-pVDZ_g03
508	CM4M/M06/DZVP
509	CM4M/M06/6-31B(d)
510	CM4M/M06/6-31B(d,p)

In the above tables, *X* corresponds to the percentage of Hartree-Fock exchange (see the HFE keyword for more details). For SM6, DFT refers to any good density functional. A list of density functionals that are available in *Gaussian* and *MN-GFM* is given in the section entitled “Density Functional Methods Available in *Gaussian* and *MN-GFM*”. In cases where the default value of ICDS = 0, a solvation model does not exist for that particular set of charges. For ICMD = 300, all restricted and unrestricted HF, DFT, and hybrid DFT methods are available, but only the 6-31+G(d) or the 6-31+G(d,p) basis set should be used. Note that CM3 for the BLYP and B3LYP methods and CM3.1 for the HF/MIDI! method in the above table uses a different mapping function for compounds that contain N and O than CM3 for MPWX and HF does. For more information, see “Parameterization of Charge Model 3 For AM1, PM3, BLYP, and B3LYP” by Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.*, **2003**, *24*, 1291 and “Accurate Partial Atomic Charges for High-Energy Molecules with the MIDI! Basis Set” by Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133.

#### IGAS

Determines which gas-phase energy is used in the evaluation of  $\Delta G_S^0$ .

- For IGAS = 0, the gas-phase energy of the geometry provided in the input is used throughout the calculation (default).
- For IGAS = 1 the gas-phase energy is updated for every new geometry.
- For IGAS = 2 the gas-phase energy is read in using the ETGAS keyword (this keyword is described below).
- For IGAS = 3 the gas-phase energy is read in using the ETGAS keyword, and gas phase SCF is skipped, and the liquid phase SCRF begins at first SCF cycle.

#### IGRD

Controls gradient evaluation.

- For IGRD = 0, there is no analytical gradient evaluation (default).
- For IGRD = 1, gradients for SM5.42, SM5.43, SM6, or SM8 defined by ICMD and ICDS, or a generalized Born model, defined by setting ICMD > 0 and ICDS = 0 are evaluated [with Iop(7/88=3000)].

#### IRADII

Determines the set of atomic-number-dependent radii that will be used to build up the molecular cavity during the SCRF calculation (note that a different set of atomic radii are used to define the SASA; see the AtmProp keyword for more details). Different sets of atomic-number-dependent radii have been optimized for

SM5.42, SM5.43, SM6, and SM8. If used, this option overwrites the default setting for Coulomb's radii specified for each solvation model. This option should be skipped if there is no need to overwrite the default radii defined by the value of ICDS.

**IRADII = 1:** Use Bondi's values for the van der Waal's radii to build the molecular cavity, (2.0 Å is used in cases where the atomic radii has not been defined by Bondi). For evaluating the SASA, this is always the default set of radii. For other values of IRADII, these radii are used in cases where the radius of a given atom is not defined for a particular model (see the table below). This is also the default if ICDS = 0.

**IRADII = 2:** Use the SM5.42 optimized atomic radii. For any unparameterized element the radius is set to the default value of 2.0 Å. This option is the default for ICDS = 1 - 10.

**IRADII = 3:** Use the SM5.43 optimized atomic radii. For any unparameterized element the radius is set to the default value of 2.0 Å. This option is the default for ICDS = 303 - 319.

**IRADII = 4:** Use the SM6 optimized atomic radii. For any unparameterized element except for iodine the radius is set to the default value of 2.0 Å. For iodine the radius is set to 2.6 Å. This is the default for ICDS = 416 - 419.

**IRADII = 5:** Use the SM8 optimized atomic radii. For any unparameterized element except for iodine the radius is set to the value of the corresponding van der Waals radius from *MV10*. For iodine the radius is set to 2.6 Å. This is the default for ICDS = 800.

**IRADII = 6:** Use the SM8AD optimized atomic radii. For any unparameterized element except for iodine the radius is set to the value of the corresponding van der Waals radius from *MV10*. For iodine the radius is set to 2.6 Å. This is the default for ICDS = 801.

Note that the *AtmProp* keyword (which is discussed in detail below) can be used to define a set of atomic-number-dependent radii different from those listed in the table below. If the user wishes to define an atomic-number-independent set of radii (i.e. each atom in a molecule is assigned a specific radius, regardless of its atomic number), the *LCoulR* keyword should be set to "True" and the atomic radius of each atom in the molecule can be read from the external file named *JobName.coul*

## ISCRF

Controls the type of SCF to be performed:

- ISCRF = 0 invokes the gas-phase calculation of Löwdin and/or RLPA charges and CM2, CM3, or CM4 charges (default)
- ISCRF = 1 invokes the gas- and liquid-phase solution calculation of Löwdin and/or RLPA charges and CM2, CM3, or CM4 charges and the calculation of

solvation free energy using SCF Scheme I (see SCF Schemes). SCF Scheme I is no longer available for basis sets that contain diffuse basis functions.

- ISCRF = 2 invokes the gas- and liquid-phase solution calculation of Löwdin and/or RLPA charges and CM2, CM3, or CM4 charges and the calculation of solvation free energy using SCF Scheme II (see SCF Schemes). For liquid-phase calculations that employ diffuse basis functions, this is the default option. NOTE: ISCRF = 2 is not compatible with IGAS = 3 ( in this case, IGAS will be reset to 0).
- ISCRF = 3 uses user-supplied partial atomic charges (see ChgRead) in the evaluation of solvation free energy using SCF Scheme III (see SCF Schemes)

When RLPA charges are calculated and printed, so are Löwdin charges. However, for liquid-phase calculations that use RLPA charges, the Löwdin charges correspond to the converged SCRF achieved with RLPA charges, not with Löwdin charges. Also note that RLPA charges should only be calculated by methods that use the 6-31+G(d) or 6-31+G(d,p) basis sets, while Löwdin charges should be used for other methods.

**JobName** Accepts as input a file name prefix (default is ‘MNGSM’). The JobName keyword is a character string that can be up to 31 characters long, and it must be placed in single or double quotations (*e.g.*, JobName=‘prefix’ or JobName=”prefix”). Note that the value of the JobName keyword is case sensitive. For example, if the AtmPrp keyword is set to true, and JobName is set equal to ‘Test’, the program will expect to find a file called Test.AtmP in the same directory that the input file is in. In addition, the value of the JobName keyword should not contain blank spaces, tabs, and/or special characters such as \*, !, &, @, {, \$, {, etc.

The keywords AtmProp, CDSRead, ChgRead, MeanK, and ReadK control the writing or reading of information to or from external files. The names of these external files have the form *prefix.suffix*, where *prefix* = JobName, and where *suffix* has a specific value for each of the above keywords (see the AtmProp, CDSRead, and ChgRead keyword descriptions for the explicit form of *suffix*). External files will be referred to as *JobName.suffix* in this manual, where *suffix* will be the specific *suffix* value in normal print.

**LCoulR** If LCoulR = True, then an array of Coulomb radii for different atoms in the same molecule is read from the external file *JobName.coul*.

**MeanK** This option is similar to ReadK however instead of computing the electronic energy of the solute in solution at each temperature; it computes the electronic energy of the solute at the average temperature and scales the electronic energy of the solute using the following factor

$$\frac{\left(1 - \frac{1}{\epsilon_T}\right)}{\left(1 - \frac{1}{\epsilon_{Avg}}\right)} \quad (15)$$

where  $\epsilon_{Avg}$  is the dielectric constant of water at the average temperature and  $\epsilon_T$  is the dielectric constant of water at a given temperature,  $T$ . This corresponds assuming that the charge on the solute does not change significantly as a function of temperature. This approach can compute a large number of temperatures with nearly the same computation time as that of a single temperature. This approach does deviate mildly from the individual computations performed at each temperature; however, the deviation is minor.

- Mulliken** If Mulliken = True, then Mulliken atomic charges and the corresponding dipole moment are evaluated (default: Mulliken = False).
- NOPOL** If NOPOL = True, then  $G_p$  corresponding to the gas-phase wave function is evaluated (default: NOPOL = False).
- ReadK** If ReadK = True, this variable causes *MN-GSM* to read in a list of temperatures at which to compute free energies of solvation from a file labeled *JobName.kel*. The first line in the input file must list the number of temperatures at which the free energy of solvation is to be computed. Then each succeeding line should list only one temperature.

This method uses the converged wavefunction at each temperature as a starting guess for the succeeding temperature. This approach saves some time by avoiding the need to recompute the gas phase structure of the solute, however if the temperatures are closely spaced, the perturbation introduced by the new temperature may not be greater than the convergence criteria. This is not particularly problematic save that electronic energies computed in this fashion will differ slightly from those obtained by computing each temperature individually using SolK. This problem can be avoided by using very tight convergence criteria; however the increase in time due to the tighter criteria counterbalances the time savings from computing the gas-phase electronic structure of the solute only once. Tight convergence criteria are recommended in cases where two or more temperatures are less than 20 K apart.

- SolA** Abraham's hydrogen bond acidity,  $\Sigma\alpha_2^H$  (default is 0.0).
- SolB** Abraham's hydrogen bond basicity,  $\Sigma\beta_2^H$  (default is 0.0).
- SolC** Aromaticity,  $\phi$ : the fraction of non-hydrogenic solvent atoms that are aromatic carbon atoms (default is 0.0).
- SolG**  $\gamma = \gamma_m / \gamma^o$  (default is 0.0), where  $\gamma_m$  is the macroscopic surface tension at air/solvent interface at 298.15 K, and  $\gamma^o$  is  $1 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  (note that  $1 \text{ dyne/cm} = 1.43932 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ ).

- SolH** Electronegative halogenicity,  $\psi$  : the fraction of non-hydrogenic solvent atoms that are F, Cl, or Br (default is 0.0).
- SolK** The temperature at which the solute free energy of solvation is computed in aqueous solution. The current model is only applicable to H, C and O containing compounds for the range of temperatures in which water is liquid (273 to 373 K). Note the model uses a temperature-dependent dielectric constant, the equation for it was obtained from the *CRC Handbook of Chemistry and Physics 76<sup>th</sup> edition*, ed. Lide, D. R., 1995, CRC Press, New York.
- SolN** Index of refraction at optical frequencies at 293 K,  $n_{20}^D$  (default is 0.0).
- SolvRd** Value of the solvent radius (in angstroms) for the calculation of the solvent-accessible surface areas of the atoms of the solute (default 0.40 Å). For predicting solvation free energies with SM5.42, SolvRd should be set to 0 Å. For predicting solvation free energies with SM5.43 or later model, the default value of 0.40 Å should be used.

### ***Extended Description of AtmProp***

The atomic properties used in *MN-GSM* can be supplemented or substituted by the user by setting `AtmProp = True` and entering the data in the external file `JobName.AtmP`. Each atomic property has a property number (*Iprop*):

- Iprop* = 1: Atomic-number-dependent Coulomb radius (angstroms)  
*Iprop* = 2: Atomic-number-dependent van der Waals radius (angstroms)  
*Iprop* = 3: number of frozen-core electrons (default is zero)

The first line of `JobName.atmp` contains the total number of atomic property values provided in the file. Subsequent lines have the form:

$$Iprop \quad Atomic\_Number \quad Property\_Value$$

where *Iprop* and *Atomic\_Number* are integers, and *Property\_Value* is a floating point number (even for *Iprop* = 3).

#### Example:

The following input sets the number of frozen-core electrons for lutetium to 54, sets both the Coulomb and the van der Waals radius of lutetium to 2.45 angstroms, and sets the Coulomb radius of boron to 1.40 angstroms:

```
4
1 71 2.45
2 71 2.45
```

3	71	54.0
1	5	1.40

The use of the `AtmProp` keyword will override the user's definition of Coulomb radii defined via `LCoulR = True`. However, the user can still use `AtmProp` to define a set of atomic-number dependent radii for calculating the SASA.

### ***Extended Description of CDSRead***

Equations (4) through (7) in the *SM5.42*, *SM5.43*, and *SM6 Solvation Models* subsection of the *Executive summary* section introduce the naming conventions for the surface tension coefficients used in this section.

Surface tension coefficients can be provided by the user either as a set of solvent descriptor-independent coefficients  $\sigma_k^A$  or as a set of solvent descriptor-dependent coefficients  $\{\hat{\sigma}_i^{(n)}, \hat{\sigma}_i^{(\alpha)}, \hat{\sigma}_i^{(\beta)}, \hat{\sigma}_i^{(\gamma)}, \hat{\sigma}_i^{(\beta^2)}, \hat{\sigma}_i^{(\phi^2)}, \text{ and } \hat{\sigma}_i^{(\psi^2)}\}$ . It should be noted that the coefficients for any solvent or mixed solvent can be provided as input as a set of combined surface tensions,  $\bar{\sigma}_k^A = \sigma_k^A + \sigma^{\text{CS}}$ :

$$\bar{\sigma}_k^A = \sum_i \left( \hat{\sigma}_i^{(n)} n + \hat{\sigma}_i^{(\alpha)} \alpha + \hat{\sigma}_i^{(\beta)} \beta \right) + \hat{\sigma}_i^{(\gamma)} \gamma + \hat{\sigma}_i^{(\beta^2)} \beta^2 + \hat{\sigma}_i^{(\phi^2)} \phi^2 + \hat{\sigma}_i^{(\psi^2)} \psi^2 \quad (16)$$

where the sum is over the possible subscripts  $I = z_k, z_{k'}, \text{ and } z_{k''}$ .

User-provided surface tension coefficients are entered into the external file *JobName.cds* in the following fashion: the first line of file contains the total number of surface tension coefficients that the user has provided in the file, followed by the value of *Icoeff*, where:

- Icoeff* = 1: use the surface tension coefficients provided in the code for any surface tension coefficients that are not provided by the user in this input file
- Icoeff* = 2: set all surface tension coefficients that have not been provided to zero

All coefficient types have been assigned a four digit coefficient label (see table below); the user provides a coefficient label followed by its corresponding value (in cal/Å<sup>2</sup>) on each line following the first line until all data are entered.

Only one parameter convention can be entered: either a set of  $\sigma_k^A$  or  $\bar{\sigma}_k^A$  coefficients or a set of solvent descriptor-dependent coefficients ( $\hat{\sigma}_i^{(n)}, \hat{\sigma}_i^{(\alpha)}, \hat{\sigma}_i^{(\beta)}, \hat{\sigma}_i^{(\gamma)}, \hat{\sigma}_i^{(\beta^2)}, \hat{\sigma}_i^{(\phi^2)}, \text{ and } \hat{\sigma}_i^{(\psi^2)}$ ) can be entered, but not a mix of both (*i.e.*, coefficient labels must be either all less than 2000 or all greater than 2000).

One-, two-, and three-atom surface tension coefficients:

$\sigma_k^A, \bar{\sigma}_k^A$	$\hat{\sigma}_i^{(n)}$	$\hat{\sigma}_i^{(\alpha)}$	$\hat{\sigma}_i^{(\beta)}$	<b>coefficient type</b>
1000 + I, I = 1 - 100	2000 + I, I = 1 - 100	3000 + I, I = 1 - 100	4000 + I, I = 1 - 100	atomic coefficient for atoms with nuclear charge <i>i</i>
1100 + j, j = 1 - 100	2100 + j, j = 1 - 100	3100 + j, j = 1 - 100	4100 + j, j = 1 - 100	H-X, where <i>j</i> is the nuclear charge on atom X
1201	2201	3201	4201	C-C(1)
1202	2202	3202	4202	C-C(2)
1203	2203	3203	4203	O-C
1204	2204	3204	4204	O-O
1205	2205	3205	4205	N-C(1)
1206	2206	3206	4206	O-N
1207	2207	3207	4207	S-S
1210	2210	3210	4210	C-N
1211	2211	3211	4211	N-C(2)
1212	2212	3212	4212	H-N(2)
1213	2213	3213	4213	H-O(2)
1214	2214	3214	4214	O-P
1215	2215	3215	4215	S-P
1216	2216	3216	4216	N-C(3)
1217	2217	3217	4217	O-Si
1218	2218	3218	4218	O-S

molecular surface tension coefficients:

5001:	$\hat{\sigma}^{(\gamma)}$
5002:	$\hat{\sigma}^{(\beta^2)}$
5003:	$\hat{\sigma}^{(\phi^2)}$
5004:	$\hat{\sigma}^{(\psi^2)}$

Example:

The following input defines the values of the following four coefficients:  $\hat{\sigma}_H^{(n)} = 40.80$ ,  $\hat{\sigma}_C^{(\alpha)} = 24.51$ ,  $\hat{\sigma}_{H,O}^{(\beta)} = -346.42$ , and  $\hat{\sigma}^{(\gamma)} = 0.3136$ . All other coefficients are set to zero (*Icoeff* = 2)

```

4 2
2001    40.80D0
3006    24.51D0
4108   -346.42D0
5001    0.3136D0

```

## ***Extended Description of DeLEMin***

For IGAS = 0, 1, and 2, the liquid-phase SCRF is preceded by a gas-phase SCF. In a conventional *Gaussian* energy calculation, when the gas-phase SCF has converged, the program stops the SCF procedure. If instead it is required that the SCF procedure continue with the liquid-phase SCRF step, then the *Gaussian* code must be told that, in fact, the SCF procedure should not be stopped. Thus, to continue with the liquid-phase SCRF calculation, *MN-GSM* notifies *Gaussian* that the SCF is actually not finished, and the SCF procedure is continued.

*Gaussian* determines that an SCF or SCRF calculation has converged if the difference in the density matrix and/or the difference in electronic energy between two consecutive steps is below defined thresholds. Often, in the first few steps of the liquid-phase SCRF, the changes in the density matrix and the electronic energy between consecutive steps are still below these thresholds, and the liquid-phase SCRF procedure may stop prematurely. Sometimes many SCRF steps are required to get past this slow-change phase. The quantity DeLEMin is the minimum change in electronic energy between the gas-phase electronic energy and the current liquid-phase electronic energy that must be achieved in order for *MN-GSM* to stop forcing *Gaussian* to continue the SCRF procedure. Once this threshold has been overcome, it is expected that the SCRF will continue normally until the minimum energy of the liquid-phase wave function is found. DeLEMin accepts as input floating-point values in hartrees.

The default value of 0.000001 hartrees works for every test case in the test suite that uses SCF=Tight. A greater value of DeLEMin would be required if  $\Delta E_E$  for a molecular system is less than 0.000001 hartrees. Such a result can be verified by using IGAS = 3 (where DeLEMin is not required).

## ***Summary of Keywords and Their Default Values***

<b><i>MN-GSM</i> keyword</b>	<b>default value</b>
AtmProp	False
Border	False
CDSRead	False
CMxSave	False
ChgRead	False
LCoulR	False
DeLEMin	0.000001
DIELEC	78.3 if IAQU = 1, 0.0 if IAQU = 0; if IAQU = 0 and DIELEC = 0.0, 1/ $\epsilon$ is set to 0.0, <i>i.e.</i> , no polarization due to the solvent
ETGAS	N/A
HFE	-1.0, if ICMD is 315 to 319 or 416 to 419, the program will determine HFE automatically unless the value is defined in the \$MNGSM namelist
IAQU	0, if nonzero values for any of SolN, SolA, SolB, SolG, SolC, or SolH have been entered, 1, otherwise
ICDS	800
ICMD	must be defined
IGAS	0

IGRD	0
ISCRF	0; if ChgRead = True, ISCRF is set to 3
JobName	MNGSM
MeanK	False
Mulliken	False; if CM1Info = True, Mulliken is set to True
NOPOL	False
ReadK	False
SolA	0.0
SolB	0.0
SolC	0.0
SolG	0.0
SolH	0.0
SolK	298
SolN	0.0
SolvRd	0.40 Å

### ***MN-GSM Keywords Required for Running Standard SMx Calculations***

Shown in the below table are the *MN-GSM* keywords required for running standard SM5.42, SM5.43, and SM6 calculations. The keywords entered in the \$MNGSM namelist are used by the solvation and charge model portion of *MN-GSM*. Keywords used by the main *Gaussian* program, such as the level of theory, basis set, scf and geometry optimization options, etc., must also be included in the route section of the input file.

<b>Method <sup>a</sup></b>	<b>ICMD</b>	<b>ICDS</b>	<b>IRADII</b>	<b>SolvRd</b>	<b>HFE</b>
SM5.42/HF/MIDI!	1	1	2	0.00	N/A
SM5.42/HF/MIDI!6D	2	2	2	0.00	N/A
SM5.42/HF/6-31G(d)	3	3	2	0.00	N/A
SM5.42/HF/6-31+G(d)	8	8	2	0.00	N/A
SM5.42/HF/cc-pVDZ	9	9	2	0.00	N/A
SM5.42/BPW91/MIDI!	4	4	2	0.00	N/A
SM5.42/BPW91/MIDI!6D	5	5	2	0.00	N/A
SM5.42/BPW91/6-31G(d)	7	7	2	0.00	N/A
SM5.42/BPW91/DZVP	10	10	2	0.00	N/A
SM5.42/B3LYP/MIDI!	6	6	2	0.00	N/A
SM5.43/HF/6-31G(d)	303	303	3	0.40	N/A
SM5.43/B3LYP/6-31G(d)	313	313	3	0.40	N/A
SM5.43/MPWX/MIDI!	315	315	3	0.40	X
SM5.43/MPWX/MIDI!6D	316	316	3	0.40	X
SM5.43/MPWX/6-31G(d)	317	317	3	0.40	X
SM5.43/MPWX/6-31+G(d)	318	318	3	0.40	X
SM5.43/MPWX/6-31+G(d,p)	319	319	3	0.40	X
SM6/DFT/MIDI!6D	416	416	4	0.40	<i>b</i>
SM6/DFT/6-31G(d)	417	417	4	0.40	<i>b</i>
SM6/DFT/6-31+G(d)	418	418	4	0.40	<i>b</i>

SM6/DFT/6-31+G(d,p)	419	419	4	0.40	<i>b</i>
SM8/DFT/MIDI!	415	800	5	0.40	<i>b</i>
SM8/DFT/MIDI!6D	416	800	5	0.40	<i>b</i>
SM8/DFT/6-31G(d)	417	800	5	0.40	<i>b</i>
SM8/DFT/6-31+G(d)	418	800	5	0.40	<i>b</i>
SM8/DFT/6-31+G(d,p)	419	800	5	0.40	<i>b</i>
SM8/DFT/6-31G(d,p)	420	800	5	0.40	<i>b</i>
SM8/DFT/cc-pVDZ_g03	422	800	5	0.40	<i>b</i>
SM8/DFT/DZVP	423	800	5	0.40	<i>b</i>
SM8/DFT/6-31B(d)	424	800	5	0.40	<i>b</i>
SM8/DFT/6-31B(d,p)	425	800	5	0.40	<i>b</i>
SM8/M06/MIDI!	500	800	5	0.40	<i>b,c</i>
SM8/M06/MIDI!6D	501	800	5	0.40	<i>b,c</i>
SM8/M06/6-31G(d)	502	800	5	0.40	<i>b,c</i>
SM8/M06/6-31+G(d)	503	800	5	0.40	<i>b,c</i>
SM8/M06/6-31+G(d,p)	504	800	5	0.40	<i>b,c</i>
SM8/M06/6-31G(d,p)	505	800	5	0.40	<i>b,c</i>
SM8/M06/cc-pVDZ	506	800	5	0.40	<i>b,c</i>
SM8/M06/cc-pVDZ_g03	507	800	5	0.40	<i>b,c</i>
SM8/M06/DZVP	508	800	5	0.40	<i>b,c</i>
SM8/M06/6-31B(d)	509	800	5	0.40	<i>b,c</i>
SM8/M06/6-31B(d,p)	510	800	5	0.40	<i>b,c</i>

In the above table, *X* corresponds to the percentage of Hartree-Fock exchange used in the mPW hybrid density functional (see the HFE keyword for more details). For SM6/SM8, DFT refers to any good density functional. A list of recommended density functionals that are supported in *Gaussian* is given in the section entitled “Density Functional Methods Available in *Gaussian* and *MN-GFM*”.

<sup>a</sup> Analytic gradients are available only for basis sets with s, p, and Cartesian d, f functions.

<sup>b</sup> When using any of the CM4 or SM6/SM8 methods, the HFE keyword must always be specified (HFE = 0.00 should be entered when using pure DFT functionals). For more details, see the section entitled “Density Functional Methods Available in *Gaussian* and *MN-GFM*”.

<sup>c</sup> M06 stands for the M06 suite of density functionals (M06-L, M06, M06-2X, or M06-HF)

### ***Modified Gaussian links***

The *MN-GSM* module consists of two modified *Gaussian* links:

- 
- 1502 Contains additions and modifications that incorporate solvation energy and some of the components of a gradient evaluation. This link is required for any type of calculation performed by *MN-GSM* (*i.e.*, Löwdin, RLPA, CM2, CM3, or CM4 charge calculations, free energy of solvation calculations, gradient calculations, numerical frequency evaluation, and geometry optimizations).

L701 Contains additions and modifications for liquid-phase analytical gradient evaluation. This link is required for analytical gradient evaluation, numerical frequency evaluation and geometry optimization.

In order to use the modified links the %SUBST command is required. The syntax for this command is:

%SUBST *link\_number directory*

where *link\_number* is 1502 or 1701, and *directory* is the UNIX directory (absolute paths should be used) that contains the modified link executables (files 1502.exe, and 1701.exe). Any %SUBST commands must be placed before the '#' command line in a *Gaussian* input file, and a %SUBST command must be provided for each modified link required for a particular calculation. See the *Input examples* section below for explicit examples.

### ***Important Notes and Troubleshooting***

Below is a summary of important notes associated with the installation and proper execution of *MN-GSM*. Notes in this section are updated on a regular basis, often after particularly useful suggestions from *MN-GSM* users, or after a new version of *MN-GSM* is released. Thus, the notes below are continually updated (useful suggestions should be communicated to the *MN-GSM* developers; see the comp.chem.umn.edu website for information on how to contact *MN-GSM* developers).

#### **Notes on Installation**

1. On IBM platforms, the following environmental variable may have to be set in order to run *MN-GSM*:

```
setenv XLFRTOPTIONS namelist=old:xrf_messages=yes:err_recovery=yes
```

This line can be added to the users' ".cshrc" file, or any initialization file read during the login process.

#### **Notes on Input File Structure**

1. The Fortran namelist command looks for the first occurrence of the \$MNGSM keyword in the input file, where the "\$" is in column 2.
2. No namelist keywords are case sensitive. However, character arguments provided as input (*e.g.*, input for the JobName keyword) are case sensitive.
3. \$MNGSM input can span several lines, provided that the input is terminated by a single \$END.
4. The declarations of logical values on some platforms can cause errors while *MN-GSM* is reading in input from the \$MNGSM namelist. To avoid this problem it is highly

recommended that the user place periods on either side of the logical value, ie ReadK=True may cause problems whereas ReadK = .True. or ReadK=.T. will not.

## Notes on Basis Sets

1. The keyword for the MIDI! basis set in *Gaussian* is midix. The keyword for MIDI!6D is midix 6d.
2. When using effective core potentials, the number of frozen electrons must always be specified in the .AtmP file.
3. Analytical gradients are only available for basis sets that use Cartesian d and f functions. Note that when the keyword “gen” is specified in a *Gaussian* input file, spherical d and f functions (5D, 7F) are used by default. Thus, when performing liquid-phase geometry optimizations with either SM5.42, SM5.43, SM6, SM8, SM8AD, or the generalized Born model with the gen keyword, the 6d keyword must also be specified. If the 6d keyword is not specified and liquid-phase geometry optimization is invoked, the program will stop, and print out the following error:

**SYMSLV: T(1,1)=0**

## Notes on SCF Convergence

1. Accurate (i.e. SCF = Tight or SCF = (conver =  $N$ ), where  $N$  is an integer greater than 4) should be used for all single-point calculations. Note that specifying SCF = (conver = 8) invokes the same SCF convergence criterion as specifying SCF = Tight. For more information on SCF convergence options in *MN-GSM*, see the section “*SCF Convergence*”.
2. When using the ReadK keyword (used for performing calculations at temperatures other than 298 K) a tighter convergence criteria of SCF = (conver = 10) is recommended when reading in closely spaced temperatures (less than 20 K separation between successive temperatures).
3. If  $\Delta E_E$  is very small (e.g., less than 0.010 kcal/mol) then there is a possibility that the liquid-phase SCF calculation will not converge, leading to the following error message in the output:

```
Restarting incremental Fock formation.  
Restarting incremental Fock formation.  
Restarting incremental Fock formation.  
>>>>>>>> Convergence criterion not met.
```

A good example of a solute for which this occurs is the chloride anion, where carrying out a single point calculation in solution at the BPW91/6-31G(d) level of theory leads to the following error message:

```
>>>>>>>> Convergence criterion not met.  
SCF Done: E(RB-PW91) = -460.246528193 A.U. after 65 cycles
```

```
Convrg = .2387D-12          -V/T = 2.0034
S**2   = .0000
Convergence failure - run terminated.
```

The line “Convrg = .2387D-12” indicates that the difference between the density matrix obtained for the latest SCF iteration and the density matrix from the last iteration is small, even after 65 iterations.

To solve this problem, one should carry out a separate calculation of chloride in the gas phase. Then, with the gas phase energy in hand, the following line should be added to the \$MNGSM namelist for the calculation in solution:

```
IGAS=3 ETGAS=<Gas-phase_Energy>
```

where <Gas-phase\_Energy> is -460.246528193 for the example above. These keywords tell *MN-GSM* to skip the gas phase calculation and begin obtaining the wave function in solution immediately.

## Notes on Numerical Gradients

1. The *Gaussian* expression “Opt = CalcFC” causes the calculation of an analytical Hessian for HF and DFT methods. In order to use a liquid-phase numerical Hessian in a geometry optimization, the numerical Hessian must be calculated in a separate job, saved to the checkpoint file, and then read into the optimization job using “Opt = ReadFC”. This is illustrated in a test suite calculation discussed below. It has been determined that 10<sup>-5</sup> angstroms is often the optimum step size in the evaluation of numerical gradients.
2. The StepSize option of the Force keyword does not allow the user to input a step size smaller than 10<sup>-4</sup> angstroms. The user can bypass this problem by using Iop(1 / 39 = -1), which instructs the code to read in a user-supplied step size (format D20.13) from the line following a blank line after the molecule input (see Example 5 of the *Input Examples* section below).
3. Numerical gradients can be used for geometry optimizations in the same fashion already described in the *Gaussian* manual for any method not having analytical gradients available.

## Notes on Program Limits

The maximum number of atoms allowed in an *MN-GSM* calculation is currently set to **250** (instead of 128 in the previous versions of *MN-GSM*), and the maximum number of basis functions allowed is currently set to **2500** (instead of 1000). If the user wishes to increase these limits, then the parameters *MxDaWk*, *MxSolWk*, and *MxNU* in the *MN-GSM* specific subroutines (file MNI502.src) must be modified. (The memory used in file MNI701.src is dynamically allocated onto Gaussian’s main work array, therefore there are no extra issues with the size of a particular system).

If the user wishes to increase the maximum number of basis functions, he or she must set the value of  $MxNU$  to the new desired value. In addition, the user must adjust the size of  $MxSolWk$  by using the following equation:

$$MxSolwk = 4NBasis^2 + 3NBasis + (26 + 3NAtoms) \left( \frac{NAtoms}{2} (NAtoms + 1) \right) + NAtoms^2 + 303NAtoms + 1208 + 600 \quad (17)$$

where  $Natoms$  is the maximum number of atoms allowed by *MN-GSM* (which is 250 by default) and  $Nbasis$  is the maximum number of basis functions allowed by *MN-GSM* (which is 2500 by default). For a larger maximum value of the maximum number of atoms,  $MxSolWk$  must be adjusted according to the equation above, and the value of  $MxDaWk$  must be adjusted according to the following equation:

$$MxDaWk = 40NAtoms^2 + 50NAtoms + 23 \quad (18)$$

Note that  $MxDaWk$ ,  $MxSolWk$ , and  $MxNU$  appear in several places in `MN1502.src`, and they all must all be modified. Also, `link 502` must be recompiled after these modifications are made (see the section entitled *Installation and testing*).

## ***Special Notes on Basis Sets***

### ***MIDI! basis set***

MIDI! and MIDIX are two names for the same basis set; the basis set is written as MIDI! or MIDIX in the literature, but it is referred to as MIDIX in *Gaussian*. The MIDI! basis set is similar to the 3-21G basis set with the following key differences:

- With the exception of Si, Br, and I, the MIDI  $s$  and  $p$  orbitals have different exponents, whereas the 3-21G valence  $s$  orbitals are constrained to have the same exponents as the  $p$  orbitals (Si, Br, and I use the 3-21G basis set to which polarization functions were added).
- The bang (!) indicates that there is a  $d$  shell on N, O, Si, P, S, Cl, Br, and I, and a  $p$  shell on Li. There are neither  $p$ -polarization functions on hydrogen nor  $d$ -polarization functions on C. Thus bang is part-way between star and no-star. A simple description of ! is “heteroatom polarized”, whereas \* denotes “heavy atom polarized”. In the MIDI! basis set, the  $d$  shells (and the  $p$  shell for Li) have been carefully optimized to reproduce (as well as possible) high-quality geometries and partial atomic charges.

Please note the following convention regarding MIDI! basis sets: the original MIDI! basis set used 5 spherical harmonic components for each  $d$  shell; and MIDI! by itself denotes using 5-component  $d$  shells. However, we sometimes use 6 Cartesian components for  $d$  shells; this is denoted MIDI!6D. For emphasis, sometimes we use MIDI!5D to denote MIDI!. Thus MIDI!  $\equiv$  MIDIX  $\equiv$  MIDI!5D  $\equiv$  MIDIX5D and MIDI!6D  $\equiv$  MIDIX6D.

The MIDI! basis set is supported in *Gaussian* for H, C, N, O, F, Si, P, S, Cl, Br, I, but not for Li. To use MIDI! for Li, the user may supply it to *Gaussian* via the external file, *midi-bang.bas*, which can be found in the */support* directory in the *MN-GSM* distribution. Examples of how an external basis set is used in *Gaussian* are given in the test suite. As mentioned above, the keyword in *Gaussian* for MIDI! is “MIDIX”; the *Gaussian* keyword expression for MIDI!6D is “MIDIX 6D”. Note that the MIDIX and MIDIX 6D keywords are not case sensitive.

### ***cc-pVDZ basis set***

The cc-pVDZ basis set is expressed in *Gaussian* as follows:

```
-H 0
S 3 1.00
  0.1301000000D+02  0.1968500000D-01
  0.1962000000D+01  0.1379770000D+00
  0.4446000000D+00  0.4781480000D+00
S 1 1.00
  0.1220000000D+00  0.1000000000D+01
P 1 1.00
  0.7270000000D+00  0.1000000000D+01
```

These two ways of expressing the basis set give identical wave functions and energies, because the basis functions span the same space. Furthermore, they lead to identical Mulliken charges. However, the wavefunction coefficients and the Löwdin charges are different. Our CM2 charge model was parameterized using *Gaussian 94*, which uses a different expression than the one shown above. Hence, when a user wishes to specify cc-pVDZ in *Gaussian*, the keyword "GEN" should be specified, and the cc-pVDZ basis set should be provided using the original expression. To make this straightforward for the user, the cc-pVDZ basis set has been included in the *MN-GSM* package in the */support* directory.

Whenever the cc-pVDZ basis set is referred to in conjunction with CM2 and/or SM5.42, it is the one as explained in reference *LX99*. This extended basis set is identical to the cc-pVDZ basis set for all elements except Br and I; for which the MIDI! basis set is used (reference *LC98*). For the user's convenience, the cc-pVDZ basis set in the *MN-GSM* package is the cc-pVDZ basis set extended this way.

### ***6-31G(d) and 6-31+G(d) basis sets***

Whenever the 6-31G(d) and 6-31+G(d) basis sets are referred to in conjunction with CM<sub>x</sub> and/or SM<sub>x</sub>, they are those described in “A New Class IV Charge Model for Extracting Accurate Partial Charges from Wave Functions,” J. Li, T. Zhu, C. J. Cramer, and D. G. Truhlar, *Journal of Physical Chemistry A* **102**, 1820-1831 (1998). For CM2 and SM5.42, the 6-31G(d) basis set was not available for I, so the MIDI!6D basis set was used in its stead (reference *LC98*). The 6-31+G(d) basis set was also not available for I, so the MIDI!6D basis set augmented by diffuse

*s* and *p* shells with exponent 0.03 was used in its stead. For the user's convenience, the 6-31G(d) and 6-31+G(d) basis sets are included in the *MN-GSM* package.

### ***Other notes***

The following notation is recommended:

1. Geometry optimized at level X/Y in the gas phase, followed by a single-point SMx solvation calculation at level W/Z, where W/Z is one of the choices supported by ICMD:

SMx/W/Z//X/Y

2. If X/Y is the same as W/Z, then //X/Y may be substituted by //g, where g denotes gas-phase:

SMx/W/Z//g

*Previously, solvation calculations carried out using gas-phase geometries were denoted by including an "R" suffix after the name of the SMx model. Here, this older notation has been replaced with the notation above.*

3. For a liquid-phase geometry optimization the //X/Y is dropped, and this calculation is denoted as follows:

SMx/W/Z

*Previously, solvation calculations carried out using liquid-phase geometries were denoted by dropping the "R" suffix after the name of the SMx model. Here, we drop this suffix for all solvation calculations and use the notation described above.*

### ***Input Examples***

*MN-GSM* is transparent if one performs standard *Gaussian* calculations. However, if one would like to calculate Löwdin, RLPA, CM2, CM3, or CM4 atomic partial charges or perform a solvation calculation, both Iop(3/70=3000) and the \$MNGSM namelist input are required. In addition, Iop(7/88=3000) is required for analytical evaluation of the energy gradient, geometry optimizations with analytical gradients, and numerical evaluation of Hessians and frequencies. Below is a set of input examples for a variety of typical calculations. Note that SCF Scheme I and IGAS=0 are used where applicable, and ICMD = 0 is always selected; in practice, the value of ICMD should reflect the wave function employed in a given calculation. Finally, *directory* is the UNIX directory that contains the compiled *MN-GSM* links.

A much larger set of examples of *MN-GSM* input and output is given in directory called "support" that is distributed with *MN-GSM*. These input examples are discussed in more detail in the next section.

**Example 1: Gas phase CM4 dipole moment calculation for water molecule**

```
%SUBST 1502 directory
# MPW1PW91/6-31G(d) SCF=Tight Iop(3/70=3000)

gas-phase CM4 dipole moment calculation for water

0 1
8      0.000000      0.000000      0.000000
1      0.967300      0.000000      0.000000
1      -0.210300      0.944200      0.000000

$MNGSM ISCRF=0 ICMD=417 HFE=0.25 $END
```

**Example 2: SM5.43 calculation of a water molecule in methanol solution**

```
%SUBST 1502 directory
# BLYP/6-31G(d) SCF=Tight Iop(3/70=3000)

solvation calculation of water in methanol

0 1
8      0.000000      0.000000      0.000000
1      0.967300      0.000000      0.000000
1      -0.210300      0.944200      0.000000

$MNGSM ISCRF=1 IGAS=0 ICMD=321 IAQU=0 SolN=1.3288 SolA=0.43
      SolB=0.47 SolG=31.77 Dielec=32.613 $END
```

**Example 3: SM6 geometry optimization of water in aqueous solution**

```
%SUBST 1502 directory
%SUBST 1701 directory
# MPW1PW91/MIDIX 6D Iop(3/70=3000) Iop(7/88=3000) Opt

geometry optimization of aqueous water molecule

0 1
8      0.000000      0.000000      0.000000
1      0.967300      0.000000      0.000000
1      -0.210300      0.944200      0.000000

$MNGSM ISCRF=1 IGAS=0 ICMD=416 IGRD=1 HFE=0.25 $END
```

**Example 4: Frequency evaluation of water in aqueous solution (SM6)**

```
%SUBST 1502 directory
%SUBST 1701 directory
# MPW1PW91/MIDIX 6D Iop(3/70=3000) Iop(7/88=3000) Freq=Numer

frequency evaluation of aqueous water molecule
```

```

0 1
8      0.000000      0.000000      0.000000
1      0.967300      0.000000      0.000000
1     -0.210300      0.944200      0.000000

```

```
$MNGSM ISCRF=1 IGAS=0 ICMD=416 IGRD=1 HFE=0.25 $END
```

**Example 5: Numerical gradient evaluation for water in aqueous solution (SM6)**

```
%SUBST 1502 directory
```

```
# MPW1PW91/6-31G(d) Iop(3/70=3000) Force=EnOnly Iop(1/39=-1)
```

```
numerical gradient evaluation of aqueous water
```

```

0 1
8      0.000000      0.000000      0.000000
1      0.967300      0.000000      0.000000
1     -0.210300      0.944200      0.000000

```

```
0.000010
```

```
$MNGSM ISCRF=1 IGAS=0 ICMD=417 IAQU=1 IGRD=2 HFE=0.25 $END
```

## MN-GSM Test Suite

Input and output files that test the various job types available in *MN-GSM* are included in the directory called *tests\_mngsm*. Given below are descriptions of the various input and output files contained in this folder:

Test	Description
<b>test1.inp</b>	Gas-phase CM2 charges for 1-( <i>P</i> -fluorophosphino)-1-silylmethanesulfonic acid are evaluated from a single-point HF/6-31G(d) calculation.
<b>test2.inp</b>	Gas-phase CM4 charges for 1-( <i>P</i> -fluorophosphino)-1-silylmethanesulfonic acid are evaluated from a single-point mPW1PW91/6-31G(d) calculation.
<b>test3.inp</b>	Aqueous CM4 charges for 1-( <i>P</i> -fluorophosphino)-1-silylmethanesulfonic acid are evaluated from a single-point mPW1PW91/6-31G(d) calculation using the SM8 solvation model.
<b>test4.inp</b>	Gas-phase CM4 charges for 2,2,4-trimethylpentane are evaluated from a single-point mPW1PW91/6-31G(d) calculation.
<b>test5.inp</b>	CM4 charges in a solvent with dielectric constant $\epsilon = 37$ for 2,2,4-trimethylpentane are evaluated from a single-point mPW1PW91/6-31G(d) calculation using the SM8 solvation model.
<b>test6.inp</b>	A single-point calculation is carried out for the acetate anion in water at the SM5.42/B3PW91/6-31G(d) level.
<b>test7.inp</b>	A geometry calculation is carried out for the acetate anion in water at the SM5.42/B3PW91/6-31G(d) level.
<b>test8.inp</b>	A geometry calculation is carried out for the acetate anion in acetonitrile at the SM5.42/B3PW91/6-31G(d) level.
<b>test9.inp</b>	A single-point calculation is carried out for the acetate anion in water at the SM5.43/B3LYP/6-31G(d) level.
<b>test10.inp</b>	A geometry calculation is carried out for the acetate anion in water at the SM5.43/mPW1PW91/6-31G(d) level.
<b>test11.inp</b>	A geometry calculation is carried out for the acetate anion in acetonitrile at the SM5.43/B3LYP/6-31G(d) level.
<b>test12.inp</b>	A single-point calculation is carried out for the acetate anion in water at the SM6/mPW1PW91/6-31G(d) level.
<b>test13.inp</b>	A geometry calculation is carried out for the acetate anion in water at the SM8/mPW1PW91/6-31+G(d) level.
<b>test14.inp</b>	A transition state optimization for the $S_N2$ reaction of $NH_3$ and $CH_3Cl$ in aqueous solution (the Menshutkin reaction) is performed at the SM8/mPW1PW91/MIDI!6D level. Two numerical Hessian calculations are carried out: one before the optimization to generate a good Hessian guess, and one after the optimization to ensure that the stationary-point is a first-order saddlepoint.
<b>test15.inp</b>	An analytical and a numerical gradient evaluation for hydrogen peroxide in water are performed at the SM8/mPW1PW91/MIDI!6D level.
<b>test16.inp</b>	A Generalized Born electrostatic solvation energy and gradient evaluation using Löwdin partial atomic charges and UHF/6-31G(d,f) for the peroxide radical in aqueous solution. The use of Cartesian $f$ functions and unrestricted wave functions is also tested.
<b>test17a.inp</b> <b>test17b.inp</b>	Test cases test17a and test17b test the JobName keyword, the CMxSave = True option (CM4 charges are written to file test17.chg), and the ChgRead = True option (CM4 charges are read from file test17.chg). Both test cases perform a single-point SM6/mPW1PW91/MIDI!6D calculation of hydrogen peroxide solute in water solvent, using ISCRF = 2, and both produce the same energetic results. Note that test17a needs to be run before test17b because test17b uses a file (test17.chg) generated by test17a.
<b>test18.inp</b>	Tests the AtmProp = True and the NOPOL = True options. $G_p$ for the gas-phase wave function is evaluated for HI using RHF with the 6-31G basis set for H and Stuttgart's RLC ECP basis and pseudopotential for I. $G_p$

	is evaluated using Löwdin atomic charges, and with Coulomb radii 1.00 and 2.50 angstroms for H and I, respectively.
<b>test19.inp</b>	A transition state optimization for the S <sub>N</sub> 2 reaction of NH <sub>3</sub> and CH <sub>3</sub> Cl in aqueous solution (the Menschutkin reaction) is performed at the SM6/BLYP/6-31+G(d) level. Test 19 does not use the density fitting approximation (via the Auto keyword). Test19 uses the ModRedundant keyword to avoid calculating initial force constants (Typically, <i>Gaussian</i> requires that initial force constants be calculated or read in when using Opt=TS. The ModRedundant keyword is specified with no options; note the <i>two</i> blank lines preceeding the \$MGNSM namelist).
<b>test20.inp</b>	Tests the Border keyword; the gas-phase and liquid-phase Mayer bond order matrices are printed for ethanol in water for mPW1PW91/MIDI!6D And for SM6/mPW1PW91/MIDI!6D.
<b>test21.inp</b>	<i>MN-GSM</i> versions up to and including version 99.6 evaluated the Mayer bond order matrix incorrectly for unrestricted wave functions (see <i>Revision History</i> for <i>MN-GSM</i> version 99.6.1 for more details). Test case test21 was added to illustrate the correct evaluation of Mayer bond orders, CM <sub>x</sub> charges, solvation energy, and gradient for a UHF wave function that yields results that are significantly different from those obtain when using <i>MN-GSM</i> versions up to 99.6. A single-point SM5.42/UHF/6-31G(d) calculation for CH <sub>4</sub> <sup>+</sup> in water solvent is performed.
<b>test22.inp</b>	Test the LCouLR keyword. A Generalized Born electrostatic solvation energy evaluation using Löwdin partial atomic charges obtained at the HF/6-31G(d) level of theory is performed for pyridinium in the aqueous phase, using user-defined atomic radii (from the user-provided file test22.coul). Such radii do not depend on the atomic number of the given atom.
<b>test23.inp</b>	Tests the calculation of the electrostatic contribution to the aqueous free energy of solvation of water using RLPA charges. An analytical gradient and a numerical gradient calculation are also carried out in test23.
<b>test24.inp</b>	A single-point energy calculation at SolK=273.15K is performed for furfural in water using the SM8T model.
<b>test25.inp</b>	A single-point energy calculation at SolK=373.15K is performed for DMSO in water using the SM8T model.
<b>test26.inp</b>	Tests the ReadK=True option, where the five temperatures are read in from a file labeled test26.kel also included in the directory. The solvation free energy of DMSO in water at 298.15 K is first calculated at the SM8/ mPW1PW91/6-31+G(d,p) level of theory. Then, the differences between the computed aqueous solvation free energy at 298.15 K and the solvation free energy at each of the temperatures are calculated using SM8T.
<b>test27.inp</b>	Tests the MeanK=True option, where the five temperatures are read in from the same file test26.kel as in the previous example.
<b>test28.inp</b>	Geometry optimization of water in the gas-phase, at the SM8/B3LYP/6-31+G(d) level of theory. Gas-phase Lowdin, RLPA, and CM4 charges are calculated during each step of the geometry optimization.
<b>test29.inp</b>	An SM8/B3LYP/6-31G(d) single-point calculation for acetonitrile in acetonitrile is performed using the SM8 CDS parameters from the file entitled test29.cds and the CDSread=True option.
<b>test30.inp</b>	An SM8/B3LYP/6-31G(d) single-point calculation for acetonitrile in acetonitrile is performed using the default SM8 CDS parameters. This example also tests IGAS=3.
<b>test31.inp</b>	An analytical gradient calculation for 2,2-dichloroethenyl dimethyl phosphate in water is performed at the SM8/HF/6-31G(d) level of theory (HFE=0.999, ICMD=417).
<b>test32.inp</b>	An analytical gradient calculation for 5-fluorouracil in fluorobenzene is performed at the SM8/HF/6-31+G(d,p) level of theory (HFE=0.999, ICMD=419).
<b>test33.inp</b>	An analytical gradient calculation for some molecule containing 47 atoms in water is performed at the SM8/HF/STO-3G level of theory (HFE=0.999, ICMD=417).
<b>test34.inp</b>	An analytical gradient calculation for some molecule containing 47 atoms in water is performed at the SM8AD/HF/STO-3G level of theory (HFE=0.999, ICMD=417).
<b>test35.inp</b>	SM8AD/M05-2X/6-31+G(d) electrostatics, analit./numer. gradients with RLPA - water in water

<b>test36.inp</b>	SM8AD/HF/MIDIX 6D analytical/numer. gradients - 5-fluorouracil in fluorobenzene
<b>test37.inp</b>	SM8AD/HF/STO-3G 6d analytical/numer. gradients - 1,1,1,3,3,3-hexafluoropropan-2-ol water cluster in water
<b>test38.inp</b>	SM8AD/HF/6-31G(d) analytical gradient - 1,1,1,3,3,3-hexafluoropropan-2-ol water cluster in water
<b>test39.inp</b>	SM8AD/HF/MIDIX 6D analytical optimization - 5-fluorouracil in fluorobenzene
<b>test40.inp</b>	SM8AD/HF/MIDIX 6D seminumerical frequencies - 5-fluorouracil in fluorobenzene
<b>test41.inp</b>	SM8AD/M05-2X/MIDIX(6D) analytical gradient calculation in some zwitterion in methylene chloride
<b>mngfm_test1.inp</b>	An SM8/M06-2X/6-31G(d) single-point energy and force constant calculation of methanol in water is performed.
<b>mngfm_test2.inp</b>	An SM8AD/M06-HF/6-31G(d,p) geometry optimization for methanol in water is performed.
<b>mngfm_test3.inp</b>	An SM8AD/M062X/6-31G(d,p) geometry optimization for methanol in acetonitrile is performed.
<b>mngfm_test4.inp</b>	A numerical frequency calculation for methanol in acetonitrile is performed at the SM8/M06/6-31+G(d,p) level.
<b>mngfm_test5.inp</b>	The SM8 model is tested for geometry optimization the acetate anion in a methanol-water mixture at the M06L/6-31B(d) 6D level of theory. The CM4M charge model for M06L/6-31B(d) is tested as well. The basis is read in from a file entitled 631Bd.gbs.

The user should use the script *compare.pl* in the directory *tests\_mngsm* to verify that all the test calculations were successful by executing the following command:

```
./compare.pl *.log
```

## Density Functional Methods Available in Gaussian and MN-GFM

See <http://comp.chem.umn.edu/mn-gfm> for the most recent information.

### SCF Schemes and SCF Convergence

If ISCRF = 0, only a gas-phase calculation is performed, and this is the default. If ISCRF equals either 1 or 2, then the code does a liquid-phase SCF calculation. Two different SCF schemes have been implemented. The standard scheme, called scheme I, uses the current solution-phase bond order matrix values in the Fock or Kohn-Sham operator at every step of the iteration. Scheme I is chosen by setting ISCRF = 1.

Scheme II uses the gas-phase bond order matrix values at the current geometry to calculate CM2, CM3, or CM4 corrections of Löwdin or RLPA charges in solution. Scheme II and its physical meaning may be “justified” as follows. The modification to the Löwdin or RLPA charges is parameterized to fit to experimental gas-phase dipole moments. So one can assume that the same modification applies to the solvated molecule with the same geometry in the liquid phase as in the gas phase. The change of atomic partial charges due to the solvation effects is reflected by the change of Löwdin or RLPA charges, and all  $q_k^M$  are constants (see the references ZL98, LH98, and LZ99 for more detail). This leads to a very simple expression for solvated Fock matrix elements:

$$F_{ij}^S = F_{ij}^g - \sum_n V_{k_n} (\mathbf{S}^{\frac{1}{2}})_{in} (\mathbf{S}^{\frac{1}{2}})_{nj} \quad (14)$$

where  $V_{k_n}$  is the reaction field acting on atom  $k$ , the indices  $i, j$ , and  $n$  refer to basis functions, and  $k_n$  is the atom on which basis function  $n$  is centered.

Usually scheme I and scheme II give very similar results. In principle (and in practice), scheme II converges to slightly different results, and thus it could be considered to be a different model. In our experience the difference is usually small, and scheme II sometimes has better convergence properties.

SCF Scheme III (ISCRF = 3) performs a liquid-phase SCRF based on user-provided atomic charges (ChgRead = True). These atomic charges remain constant during the SCRF calculation. Analytical gradients have not been implemented for ISCRF = 3.

SMx calculations cannot proceed if non-DIIS SCF convergence schemes (e.g., SCF=QC or SCF=DM) are chosen.

SMx calculations cannot use ROHF, but UHF may be used.

For single point direct SCF calculations, the *Gaussian* program uses, as default, a modest convergence criterion (SCF = Sleazy) in the interest of speed (when the opt keyword is specified in the route section of a *Gaussian* input file, SCF=Tight is the default option). For most gas-phase calculations, this convergence criterion is sufficient for 0.1 kcal mol<sup>-1</sup> accuracy in the SCF energy and 3 decimal places in the density matrix. However, for liquid-phase calculations and for calculations that use diffuse basis functions, higher accuracy is often required, and the default convergence criterion is usually not sufficient for calculating accurate wave-function-dependent properties (e.g. partial atomic charges). Thus, for *MN-GSM* calculations, it is recommended that the SCF energy be converged to at least 5 decimal places in the density matrix (for calculations that use diffuse basis functions, even tighter convergence should be used, if possible). The SCF convergence criterion can be changed by specifying either SCF = Tight, or SCF=(conver = N), which sets the convergence criterion to 10<sup>-N</sup>, in the route section of the *Gaussian* input file. Note that specifying SCF = (conver = 8) invokes the same SCF convergence criterion as specifying SCF = Tight. Also note that if the SCF convergence is too loose (less than 10<sup>-5</sup>), *MN-GSM* will print out a warning message in the output file and skip the liquid-phase part of the calculation.

When calculating temperature-dependent solvation free energies with SM8T at multiple temperatures (via the ReadK option) it is recommended that an even tighter convergence criterion of SCF = (conver = 10) be used. (For SM8T calculations at a single temperature, SCF = Tight is usually sufficient). Tighter convergence criteria is often required when using the ReadK option with SM8T because when this keyword is invoked, the converged liquid-phase wavefunction corresponding to the initial temperature is used as a starting guess for successive calculations at a different temperatures (e.g. if 310 K, 320 K, and 330 K were specified with the ReadK option, the converged liquid-phase wavefunction at 310 K would be used as a starting guess for the 320 K calculation). The necessity to use tighter convergence criteria is especially important when multiple temperatures within 20 K of one another are specified. In these cases,

using even more stringent SCF convergence criteria is often required to avoid complications in SCF convergence. In these cases, SCF = (conver = 12) is typically sufficient for obtaining a converged wavefunction within a reasonable number of iterations.

## Installing MN-GSM

The *MN-GSM* installation kit contains a single directory called *MN-GSM-v2012\_source* (that comes as a tarred file) with several files (*cm5pac.exe*, *cm5pac.f*, *gbsolg09*, *gbvemg09*, *gbvemg09.exe*, *gbvemg09.f*, *mngsm.part1*, *mngsm.part2*, *mngsm.part3*, *modgau*) and three subdirectories called *tests\_gbsolg09*, *tests\_gbvemg09*, and *tests\_mngsm*.

The user should first install the original (unmodified) version of *Gaussian* or *MN-GFM* and make sure that it works successfully. Then, the user should run the script *modgau* and follow the instructions printed by the script. This script will automatically modify two *Gaussian* files, namely, *l502.F* and *l701.F*. Then, the user should recompile *l502.exe* and *l701.exe* according to *Gaussian*'s installation instructions.

The following paragraph explains how to compile *l502.exe* and *l701.exe* in the case of *Gaussian 03* on Minnesota Supercomputing Institute's machines, provided the user could successfully modify the original *Gaussian* files *l502.F* and *l701.F* using the script *modgau*. The modified files *l502.F* and *l701.F* for the case of *Gaussian 09* (revision C.01) can be found in the *MN-GSM* developers' archive.

Compilation of *l502.exe* and *l701.exe* depends where they are located on the user's machine. The following steps need to be taken to properly install *MN-GSM*:

- 1) `module load g03/mn` or `module load g03/mngfm41`
- 2) Make sure that the `$g03root` environment variable is properly set. The `$g03root` environment variable is the location of the *Gaussian 03* or *MN-GFM* directory.
- 3) Make sure that `$g03root/g03/bsd` is defined in the environment variable `$PATH`. This can be checked by typing "echo \$PATH" at the UNIX prompt. If it is not in `$PATH`, it is placed in `$PATH` by typing:

```
setenv PATH $PATH:$g03root/g03/bsd
```

- 4) In the directory where the modified links reside, a soft link to the directory `/bsd` needs to be created by typing:

```
ln -s $g03root/g03/bsd
```

and a soft link to *util.a* needs to be created by typing

```
ln -s $g03root/g03/util.a
```

- 5) On platforms running AIX (e.g. the IBM Regatta) an additional soft link to *utila.o* needs to be created by typing:

```
ln -s $g03root/g03/utila.o
```

**IMPORTANT: On platforms running AIX, the following environmental must be set in order to run MN-GSM:**

```
setenv XLFRTLOPTS namelist=old:xrf_messages=yes:err_recovery=yes
```

This line can be added to the users' ".cshrc" file, or any initialization file read during the login process.

- 6) Finally, the links can be compiled by typing:

```
make -f bsd/g03.make l502.exe  
make -f bsd/g03.make l701.exe
```

In the case of *Gaussian 09*, the variable *\$g03root* is replaced with *\$g09root*, *g03* with *g09*, and the file *g03.make* is replaced with *g09.make*. The user should verify that the file *g09.make* exists in the directory *bsd*. If not, the user should create *g09.make* from *bsd/i386.make* by modifying the latter accordingly. **Note that the *module* command should not be used if the user sets the *\$g03root* or *\$g09root* environment manually using the *export* or *setenv* commands (depending on a particular shell).** In this case, the user should verify that the *Gaussian* environment is really defined by running the command *which g03* or *which g09*.

## Running MN-GSM

The input file should contain the *%SUBST* commands as follow

```
%SUBST link_number directory
```

where *link\_number* is 1502 or 1701, and *directory* is the UNIX directory (absolute paths should be used) that contains the modified link executables (files *l502.exe*, and *l701.exe*). Any *%SUBST* commands must be placed before the '#' command line in a *Gaussian* input file, and a *%SUBST* command must be provided for each modified link required for a particular calculation. See the *Input examples* section below for explicit examples. Note that the *%SUBST* commands are not needed if the *Gaussian* directory defined by *\$g03root/g03* or *\$g09root/g09* already contains the *MN-GSM* (i.e., modified) versions of *l502.exe* and *l701.exe*.

To run *MN-GSM*, execute the following commands:

*g03* input  
or  
*g09* input

## ***Running MN-GSM for SM12 Calculations***

The current version of *MN-GSM* contains a tool (*gbsolg09*) that allows for running iterative SM12 solvation energy calculations in the ground state using CM5, CHELPG, or Merz–Kollman partial atomic charges. This model is described in more detail in reference *MC12*.

To run such calculations, the user should make sure that the *gbsolg09* tool is properly installed. There are three files (*gbsolg09*, *cm5pac.exe*, *cm5pac.f*) in */MN-GSM-v2012\_source* that should be placed in the directory defined by the variable *\$g09root/g09*.

The command ``which gbsolg09`` should yield the path to the respective executable. To run an SM12 calculation, execute the following command:

*gbsolg09 filename.inp*

Below please find an example for *filename.inp* (where *filename.inp* is *H3C1I1\_octn\_m062x\_mks.inp* given in the directory called *tests\_gbsolg09*).

```
%chk=H3C1I1_octn_m062x_mks.chk
#T m062x/GEN pseudo=read scf=(tight,MaxCycle=200) pop=(MK,ReadRadii) iop(3/70=3000)

H3C1I1 iodomethane in octanol

0 1
  6   1.8081600000   0.0000200000  -0.0000080000
  1   2.1349630000   1.0296710000  -0.0619960000
  1   2.1349400000  -0.5686110000  -0.8604870000
  1   2.1349310000  -0.4611750000   0.9225270000
 53  -0.3255430000   0.0000000000   0.0000000000

I 0
MWB46
****
H C 0
midix
****

I 0
MWB46

$MNGSM ChgRead=.True. ICMD=0 ICDS=820 JobName='H3C1I1_octn_m062x_mks'
IAQU=0 DIELEC=9.8629 SOLN=1.4295 SOLA=0.37 SOLB=0.48 SOLG=39.0057 SOLC=0 SOLH=0
$END
53 1.98
```

The *gbsolg09* script will create the *filename.log* output file. After a successful execution of *gbsolg09*, the file *filename.log* for the example above should contain the following output in the end:

```
FINAL MNGSM RESULTS:
Summary of solvation calculation by
Minnesota-Gaussian Solvation Module MN-GSM v2012 at 298.15 K
-----
Options: ISCRF = 3, IGAS = 0, ICMD = 0, ICDS = 820, IAQU = 0, DIELEC = 9.86,
        SolK = 298.15

(0) E-EN(g) gas-phase RM062X elect-nuc energy          -51.013420113 a.u.
(1) E-EN(liq) elect-nuc RM062X energy of solute        -51.012905625 a.u.
(2) G-P(liq) polarization free energy of solvation      -1.874 kcal/mol
(3) G-ENP(liq) elect-nuc-pol free energy of system     -51.015892024 a.u.
(4) G-CDS(liq) cavity-dispersion-solvent structure
    free energy                                         -2.009 kcal/mol
(5) G-P-CDS(liq) = G-P(liq) + G-CDS(liq) = (2) + (4)  -3.883 kcal/mol
(6) G-S(liq) free energy of system = (1) + (5)        -51.019094004 a.u.
(7) DeltaE-EN(liq) elect-nuc reorganization
    energy of solute molecule (7) = (1) - (0)          0.323 kcal/mol
(8) DeltaG-ENP(liq) elect-nuc-pol free energy
    of solvation (8) = (3) - (0)                       -1.551 kcal/mol
(9) DeltaG-S(liq) free energy of solvation
    (9) = (6) - (0)                                     -3.560 kcal/mol
-----

SCF Done: E(RM062X) = -51.0129056249      A.U. after 24 cycles
          Convgt = 0.2299D-10              -V/T = 2.2022

CONVERGENCE ACHIEVED for H3C1I1_octn_m062x_mks.inp
```

The '(9) = (6) - (0)' line contains the resulting SM12 free energy of solvation calculated self-consistently at the fixed geometry.

For the example above, the SM12 calculation was performed using the MK atomic charges specified by the keyword *pop=MK* and the SM12 CDS parameters optimized for the use with ESP charges (either CHELPG or Merz-Kollman) as specified by the keyword *ICDS=820*. The calculation of MK charges in this case employs the value of van der Waal radius on iodine (1.98 Å) provided by the user through the keyword *pop=readradii*. If one would like to use CHELPG charges instead of MK charges, the keyword *pop=chelpg* should be used along with the keyword *ICDS=820*. See *Gaussian 09's* manual about CHELPG and MK charges.

If the user would like to run an SM12 calculation based on CM5 charges, the keywords *pop=hirshfeld* and *ICDS=810* should be specified. The keyword *pop=hirshfeld* invokes a ground-state calculation of Hirshfeld charges which will be converted into CM5 charges by the auxiliary program *cm5pac.exe* automatically.

The keyword *ChgRead=True* invokes reading the file *filename.chg* containing an initial set of partial atomic charges for the SM12 solvation free energy calculation (where "filename" is the name of your input file without an extension). If there is no file *filename.chg* given then the initial charges are set to the charges evaluated in the gas phase. The keyword *ICMD=0* should not be changed.

The user should use the keyword *IAQU=0* for a nonaqueous solvent along with the correct values of the solvent's descriptors, namely, *DIELEC*, *SOLN*, *SOLA*, *SOLB*, *SOLG*, *SOLC*, and *SOLH*. If the solvent is water, then the keyword *IAQU=-1* should be specified. In the latter case, the solvent descriptors are set to their default values for water unless they are specified in the input file. In the example above, the solvent is octanol.

See the directory called *tests\_gbsolg09* for more examples.

## ***Running MN-GSM for Solvatochromic Calculations***

The current version of *MN-GSM* contains a tool (*gbvemg09*) that allows for running iterative vertical excitation energy calculations within the GB/VEM(d) model. This model is described in more detail in reference *MC11*.

To run such calculations, the user should make sure that the *gbvemg09* tool is properly installed. There are five files (*gbvemg09*, *gbvemg09.exe*, *gbvemg09.f*, *cm5pac.exe*, and *cm5pac.f*) in */MN-GSM-v2012\_source* that should be placed in the directory defined by the variable *\$g09root/g09*. The user is advised to recompile *gbvemg09.exe*, for example, by using the GFortran compiler as

```
gfortran -o gbvemg09.exe, gbvemg09.f
```

The commands ``which gbvemg09`` and ``which gbvemg09.exe`` should yield the path to the corresponding executables. To run the VEM calculation, execute the following command:

```
gbvemg09 filename.inp 'tddft/cis options'
```

where *'tddft/cis options'* could be, for example, *'td=(nstates=6,root=1,conver=6)'* (see *Gaussian 09's* manual for more detail on the *td/cis* keyword).

Below please find an example for *filename.inp* (where *filename.inp* is *fo\_wat\_m06\_sm8\_chg.inp* given in the directory called *tests\_gbvemg09*).

```
# m06/6-311+G(2df,2p) scf=tight integral=ultrafine pop=chelpg Iop(3/70=3000)
Formaldehyde in water ground state using sm8, chelpg, and geometry in water
0 1
8 0.000000 0.000000 0.674199
6 0.000000 0.000000 -0.531116
1 0.000000 0.938483 -1.103447
1 0.000000 -0.938483 -1.103447

$MNGSM ICMD=0 ISCRF=3 IAQU=0 ICDS=800 DIELEC=78.4 SOLN=1.3416 SOLA=0.82
JobName='fo_wat_m06_sm8_chg' $END
```

The *gbvemg09* script will create the *filename.log* output file. After a successful execution of *gbvemg09*, the file *filename.log* for the example above should contain the following output in the end:

```
CURRENT MNGSM RESULTS FOR THE EXCITED STATE:  
GSRF Excitation Energy (cm-1), w0(4) = 32606.65  
VEM(d) Excitation Energy (cm-1), w(4) = 32273.72
```

```
EXCITED-STATE CONVERGENCE ACHIEVED for fo_wat_m06_sm8_chg.inp  
... Ending MNGSM/G09 excited-state nonequilibrium calculation
```

Note that a line containing the final (converged) value of the VEM(d) excitation energy is the only result that is needed from this calculation. The remaining output can be used for debugging purposes.

For the example above, the VEM(d) calculation was performed using the SM8 Coulomb radii and the SM8 generalized Born formalism specified by the keyword *ICDS=800*. If one would like to invoke the SM8AD generalized Born formalism along with the SM8AD Coulomb radii then the keyword *ICDS=801* should be used. The keywords *ICDS=800* and *ICDS=801* are the only options available for such calculations.

The keyword *pop=chelpg* invokes the use of CHELPG charges. If one would like to use Merz-Kollman charges then the keyword *pop=mk* should be used. See *Gaussian 09's* manual about CHELPG and MK charges.

If one wants to use CM5 charges in this calculation then the keyword *pop=hirshfeld* should be specified. It will invoke ground- and excited-state calculations of Hirshfeld charges which will be converted into CM5 charges by the auxiliary program *cm5pac.exe* automatically.

The keywords "*ICMD=0 ISCRF=3 IAQU=0*" should always be used without modification. The keywords *DIELEC*, *SOLN*, and *SOLA* should be assigned the values corresponding to a given solvent. In the example above, the solvent is water.

Note also that the *gbvemg09* script creates files with current-iteration ground-state charges (*filename.chg*) and excited-state charges (*filename.neq*) (where "filename" is the name of your input file without an extension). These files can be used to restart the calculation.

See the directory called *tests\_gbvemg09* for more examples.

## Tested Platforms

The current version of *MN-GSM* was tested in combination with *Gaussian 09* (revision A.02) using the following specifications:

- HP Linux cluster with 1,091 HP ProLiant BL280c G6 blade servers, each with two-socket, quad-core 2.8 GHz Intel "Nehalem" processors, with pgfortran 9.0-4 64-bit target on x86-64 Linux -tp nehalem-64
- HP Linux cluster with 1,091 HP ProLiant BL280c G6 blade servers, each with two-socket, quad-core 2.8 GHz Intel "Nehalem" processors, with Intel(R) Fortran Intel(R) 64 Compiler Professional for applications running on Intel(R) 64, Version 11.1 Build 20100414 Package ID: 1\_cprof\_p\_11.1.072

- SGI Altix XE 1300 Linux cluster, with 256 SGI Altix XE 300 compute nodes, each containing two quad-core 2.66 GHz Intel Xeon "Clovertown"-class processors, with pgfortran 9.0-3 64-bit target on x86-64 Linux -tp core2-64
- Sun Fire X4600 Linux cluster, with 192 AMD Opteron (Model 8356) compute cores with a clock speed of 2.3GHz and 8 AMD Opteron (Model 8222) cores with a clock speed of 3.0 GHz, with pgfortran 11.0-0 64-bit target on x86-64 Linux -tp k8e

## Revision History and Version Summaries

### ***MN-GSM-v.98.1 through MN-GSM-v.98.3 (Feb., 1998–Mar., 1999)***

Developers' versions of *MN-GSM-v.98* for *Gaussian 94* revision D.4 and *Gaussian 98* revision A.6.

### ***MN-GSM-v.98.4 (June, 1999)***

by Jiabo Li, Gregory D. Hawkins, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

First standard version of *MN-GSM-v.98*. This version is for *Gaussian 94* revision D.4 and *Gaussian 98* revision A.6.

### ***MN-GSM-v.98.5 (June, 2000)***

by Jiabo Li, Gregory D. Hawkins, James D. Xidos, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch

This version is for *Gaussian 94* revision D.4 and *Gaussian 98* revision A.6.

- Several corrections have been made to the parameter sets employed in the evaluation of the CM2 charges and the SM5.42R surface tensions. All parameters that are corrected are ones that were entered in the code incorrectly in version 98.1.
- An extraneous line of code in *MN-GSM* versions 2.1–2.5 and 3.1 for *Gaussian 94* that affected UHF calculations was removed.
- A new comprehensive test suite has been added.

Some errors exist in both the *Gaussian 94* and *98* versions of *MN-GSM-v.98.5* that are rectified in *MN-GSM-v.99.0*.

### ***MN-GSM-v.99.0 (June, 2000)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version is for *Gaussian 98* revision A.9. *MN-GSM-v.99.0* is an extensive revision of *MN-GSM-v.98.x*, with the following enhancements:

- Analytical gradients for SCF Scheme I employed with wavefunctions that use Cartesian  $d$  functions (*i.e.*, for HF/MIDI!6D, HF/6-31G(d), BPW91/MIDI!6D, BPW91/6-31G(d), HF/6-31+G(d) and BPW91/DZVP).
- Numerical gradients for both SCF Schemes and all wavefunctions.
- Numerical second derivatives where analytical gradients are available.
- Efficient geometry optimization where analytical gradients are available.

- Dynamically allocated arrays: *MN-GSM-v.99.0* can efficiently handle the same size molecules that *Gaussian* can.
- The IGAS and ETGAS options were added. These allow the user to:
  - input a gas-phase energy for the evaluation of  $\Delta G_S^0$
  - control which gas-phase energy is used for the evaluation of  $\Delta G_S^0$  during a liquid-phase geometry optimization
  - skip the gas-phase SCF and begin the liquid-phase SCRF right away.
- New, more informative output.
- New test cases were added to the test suite.

### ***MN-GSM-v.99.0.1 (July, 2000)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch

This version is for *Gaussian 98* revision A.9. Scalars *NCarbon*, *NOxygen*, *NSulfur*, and *SUB2* are initialized to zero, and *L2C*, *L2O*, and *L2S* are initialized to **.false.** in subroutine *DoSolv*. These variables were not initialized in *MN-GSM-v.99.0*, resulting in some errors in the test suite.

### ***MN-GSM-v.99.1 (July, 2000)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version is for *Gaussian 98* revision A.9. If *ICSAVE* = 2, information required to determine new charge model parameterizations is saved in a file called *CM2\_info*. The output consists of (1) the number of atoms, (2) the nuclear charge and the *x*-, *y*-, and *z*-Cartesian coordinate (in angstroms) for each atom in the structure provided as input, (3) the Löwdin partial atomic charges, and (4) the Mayer bond order matrix.

### ***MN-GSM-v.99.2 (July, 2000)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

*MN-GSM-v.99.2* consists of two versions of *MN-GSM*: one compatible with *Gaussian 98* revision A.7 and one compatible with *Gaussian 98* revision A.9. Both versions are equivalent in performance and in capability to *MN-GSM-v.99.1*.

### ***MN-GSM-v.99.3 (December, 2000)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- SCRF evaluations can now be performed using Löwdin charges. This option is available for restricted or unrestricted, HF, DFT, or hybrid-DFT wavefunctions that employ Cartesian or spherical harmonic basis functions. Analytical gradients of SCRF energies with respect to the nuclear coordinates are available for restricted or unrestricted, HF, DFT, or hybrid-DFT wavefunctions that employ Cartesian basis functions up to *f* shell (i.e., 6D/10F).
- The new \$CM2 keyword *CDSOnly* has been added. When *CDSOnly* is set to true, only  $G_{CDS}$  is evaluated (i.e., the SCRF calculation is skipped).

- The new \$SCM2 keyword Mulliken has been added. When Mulliken is set to true, Mulliken atomic charges and dipole moment are evaluated and printed.
- The new \$SCM2 keywords CDSInfo, CM1Info, and CM2Info have been added. When CDSInfo, CM1Info, or CM2Info is set equal to a file name, information required for CDS, CM1, or CM2 parameterization, respectively, will be printed to that file.
- The new \$SCM2 keyword CM2Save has been added. When CM2Save is set equal to a file name, the CM2 charges will be printed to that file. This keyword replaces the ICSAVE keyword that existed in previous versions of *MN-GSM* but has been eliminated in this version.
- The new \$SCM2 keyword AtmProp has been added. When AtmProp is set equal to a file name, user-specified Coulomb radii, van der Waals radii, and/or number of frozen-core electrons will be read in from that file.
- The new \$SCM2 keyword ChgRead has been added. When ChgRead is set equal to a file name, user-specified atomic charges will be read in from that file.
- The new \$SCM2 keyword CDSRead has been added. When CDSRead is set equal to a file name, user-specified surface tension coefficients will be read in from that file.
- The “Single-subscript surface-tension-coefficient contributions to G\_CDS” section of the *MN-GSM* output now prints out information for any atom. Previous implementations of *MN-GSM* would only print information for H, C, N, O, F, P, S, Cl, Br, and I.
- Test cases test13, test14a, test14b, test14c, test14d, test15, and test16 have been added to the test suite.

### ***MN-GSM-v.99.3.1 (February, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- The parameter ZERO in subroutine *RijUij* is now initialized in a data statement. The parameter ZERO was not initialized in previous versions of *MN-GSM* starting with *MN-GSM-v.99.0*, but this oversight did not affect any results in the previous versions. Minor discrepancies in  $E(\text{liq})$  and  $G_P$  were observed in the development phase of *MN-GSM-v.99.4* on IBM SP computers (but not on the SGI Origin computers).

### ***MN-GSM-v.99.4 (February, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- The \$SCM2 keyword NOPOL option has been added: if NOPOL = True, then  $G_P$  based on the gas-phase wave function will be evaluated and printed.
- The \$SCM2 keyword JobName has been added. JobName is set equal to a character string that can be up to 31 characters long that is placed inside single or double quotation marks. JobName is the prefix of the file name for all external files read from or written to by *MN-GSM*, with the exception of the standard *Gaussian* input and output files.
- The \$SCM2 keywords AtmProp, ChgRead, CM2Save, CM1Info, CM2Info, CDSInfo, and CDSRead introduced in *MN-GSM-v.99.3* have been modified to accept logical arguments instead of character arguments. The external files accessed by *MN-GSM* if any of these

keywords are set to equal True are named *JobName.suffix*, where *JobName* is set by the JobName keyword, and there is a specific value of *suffix* for each of the above keywords: *suffix* = AtmP, Chg, CM2, CM1I, CM2I, CDSI, and CDS if AtmProp, ChgRead, CM2Save, CM1Info, CM2Info, CDSInfo, or CDSRead equal True, respectively.

- The algorithm for handling the program flow from the gas-phase SCF calculation to the liquid-phase SCRF calculation has been changed. In the current implementation, the liquid-phase SCRF is forced to continue during its first few iterations until the difference between the gas-phase energy minimum and the current liquid-phase energy is at least the value of the new keyword DeEMin (default is 0.000001 hartrees). Once this threshold has been overcome, it is expected that the SCRF will continue normally until the minimum energy of the liquid-phase wave function is found.
- The default for IAQU has been modified. In previous implementations of *MN-GSM*, the default value of IAQU was 1. In this version, the default value of IAQU is 0 if any of SolN, SolA, SolB, SolG, SolC, or SolH have been provided as input by the user, and the default value of IAQU is 1 if none of those six solvent descriptors are provided as input by the user.
- The default value of SolN has been changed from 1.00 to 0.00.
- In previous implementations of *MN-GSM*, if an error was made in the \$CM2 namelist input, the job would continue as a conventional *Gaussian 98* job. Now, if an error is made in the \$CM2 namelist input, the job will crash and print out an error message that lists all of the possible namelist keywords and their possible arguments.
- New printing has been added to the beginning of every *MN-GSM* calculation. A table containing all of the keywords, their values, and a brief description is printed, followed by a more informative summary of the calculation to be performed.
- Test case test17 has been added to the test suite.

### ***MN-GSM-v.99.5 (March, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- The \$CM2 keyword BOrder has been added; when BOrder is set to True, the gas-phase and/or liquid-phase Mayer bond order matrices are printed to the standard output.
- Tests cases test18a, test18b, and test19 have been added to the test suite.

### ***MN-GSM-v.99.6 (March, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- The author list has been expanded to include Tianhai Zhu and Daniel Rinaldi in recognition of their contributions to code that has been imported into *MN-GSM* in earlier versions.
- The citation for *MN-GSM* is now printed at the beginning of every run of *MN-GSM*.
- The *MN-GSM User Agreement* has been added to all *MN-GSM* documentation, and to an ASCII file called README that is now part of the *MN-GSM* distribution package.

### ***MN-GSM-v.99.6.1 (April 6, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

Corrected the implementation of the Mayer bond order for unrestricted wave functions:

In Mayer's well known paper on bond order (Mayer, I. *Chem. Phys. Lett.* **1983**, 97, 270), the bond order  $\mathbf{B}_{AB}$  between atoms A and B is defined as:

$$\mathbf{B}_{AB} = \sum_{\lambda \in A} \sum_{\omega \in B} (\mathbf{P}\mathbf{S})_{\omega\lambda} (\mathbf{P}\mathbf{S})_{\lambda\omega} \quad (1)$$

where  $\mathbf{P}$  is the density matrix and  $\mathbf{S}$  is the overlap matrix, and where basis functions  $\lambda$  and  $\omega$  reside on atoms A and B, respectively. This formulation is only correct for restricted wave functions where all orbitals are doubly occupied. Later, in an addendum and in a full paper (Mayer, I. *Chem. Phys. Lett.* **1985**, 117, 396; Mayer, I. *Int. J. Quant. Chem.* **1986**, 29, 73), Mayer provided a formula for UHF wave functions:

$$\mathbf{B}_{AB} = \sum_{\lambda \in A} \sum_{\omega \in B} \left[ (\mathbf{P}^{\alpha}\mathbf{S})_{\omega\lambda} (\mathbf{P}^{\alpha}\mathbf{S})_{\lambda\omega} + (\mathbf{P}^{\beta}\mathbf{S})_{\omega\lambda} (\mathbf{P}^{\beta}\mathbf{S})_{\lambda\omega} \right] \quad (2)$$

$$= 2 \sum_{\lambda \in A} \sum_{\omega \in B} \left[ (\mathbf{P}^{\alpha}\mathbf{S})_{\omega\lambda} (\mathbf{P}^{\alpha}\mathbf{S})_{\lambda\omega} + (\mathbf{P}^{\beta}\mathbf{S})_{\omega\lambda} (\mathbf{P}^{\beta}\mathbf{S})_{\lambda\omega} \right] \quad (3)$$

where  $\mathbf{P}^{\alpha}$  and  $\mathbf{P}^{\beta}$  are the alpha and beta spin density matrices, respectively,  $\mathbf{P} = \mathbf{P}^{\alpha} + \mathbf{P}^{\beta}$  is the spinless density matrix, and  $\mathbf{P}^{\mathbf{S}} = \mathbf{P}^{\alpha} - \mathbf{P}^{\beta}$  is the spin density matrix. Equations (2) and (3) reduce to equation (1) for RHF wave functions.

The implementation of the Mayer bond order matrix for UHF wave functions in previous versions of *MN-GSM* used the form:

$$B_{AB} = \sum_{\lambda \in A} \sum_{\omega \in B} \left[ (\mathbf{P}^{\alpha} + \mathbf{P}^{\beta})\mathbf{S} \right]_{\omega\lambda} \left[ (\mathbf{P}^{\alpha} + \mathbf{P}^{\beta})\mathbf{S} \right]_{\lambda\omega} \quad (4)$$

Equation (4) is correct for restricted wave functions but is incorrect for unrestricted wave functions. In version 99.6.1, equation (3) has been implemented for the evaluation of bond order for both restricted and unrestricted wave functions.

- The form of the  $\partial\mathbf{B}/\partial\mathbf{P}$  part of  $\partial G_{\mathbf{P}}/\partial\mathbf{P}$  (the solvent reaction field component of the Fock matrix) and the form of the  $\partial\mathbf{B}/\partial\mathbf{R}$  part of  $\partial G_{\mathbf{P}}/\partial\mathbf{R}$  (the energy gradient) both change due to the implementation of equation (3) for the Mayer bond order. The necessary modifications for these two terms have been made in this version.

### ***MN-GSM-v.99.7 (April 6, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- Subroutine *SolGrd* was extensively rewritten to make it more efficient, yielding a reduction of 50% in the memory required by subroutine *SolGrd* for scratch arrays, a net savings of  $6 \times NBasis \times NBasis$  double precision words, where *NBasis* is the number of contracted basis functions.

- A new test run, test20, was added to illustrate the calculation of Mayer bond orders using equation (3), CM2 charges, solvation energy, and gradient for a UHF wave function that yields results that are significantly different from those obtain when using *MN-GSM* versions up to 99.6. The calculation is for  $\text{CH}_4^+$  in water using SM5.42/UHF/6-31G(d).

### ***MN-GSM-v.99.8 (June 5, 2001)***

by James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version supports both revision A.7 and revision A.9 of *Gaussian 98*.

- In previous versions of *MN-GSM*, the value of the total energy that is stored in the code and in the checkpoint file for a liquid-phase calculation is  $E(l)$ , the electronic-nuclear energy of the solute in the liquid phase. When the modified link 1716 was employed (during geometry optimization), the value of the total energy used during geometry optimization was set to be  $G_S^0$ , the absolute free energy of the solute in the liquid phase ( $G_S^0 = E(g) + \Delta G_S^0$ ). In this version, the total energy that is stored in the code and in the checkpoint file is set to  $G_S^0$  after every SCRF procedure. Consequently, *MN-GSM* no longer requires any modification of 1716, and the modified 1716 has been dropped from the *MN-GSM* distribution. Moreover, the addition of  $G_S^0$  to the checkpoint file allows for the use of *MN-GSM* to be used with codes that used *Gaussian* as a back-end (e.g., *GAUSSRATE*).
- The *Gaussian* keyword *FCheck* has been added to test run test19 to test the update of the value of the total energy in the checkpoint file (see above). The substitution of 1716 in test cases test10 and test11 has been removed.

### ***MN-GSM-v.1.6 (August, 2001)***

by James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version of *MN-GSM* only supports revision A.10 of *Gaussian 98*. The appropriate additions and modifications of links 502 and 701 were made to prepare *MN-GSM* for its implementation into revision A.10 of *Gaussian 98*. From the user's perspective, the only notable difference between this version of the code and the previous version is the addition of the `Iop(7/88=3000)` flag, which is described in the *usage* and *input and output examples* sections of this manual.

### ***MN-GSM-v.1.7 (December, 2001)***

by James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The `CoulR`, `LCoulR`, and `PrtCoulR` keywords were added, and two new test runs were added to illustrate their use.

### ***MN-GSM-v.1.8 (January, 2002)***

by James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The SM5.42 parameters for Si have been implemented in this version of *MN-GSM*. For more details, refer to “Parameterization of a Universal Solvation Model for Molecules Containing Silicon”, Winget, P.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B*, **2002**, *106*, 5160. The option SCF=(ConVer=5) was added to the input of test cases test18a and test18b. These two test cases illustrate the utility of the keyword DeLEMin. Before the change, the SCF for these two test cases prematurely converged in revisions A.10 and A.11 on the basis on energy, not the density. The addition of the above option corrects this problem.

### ***MN-GSM-v.2.1 (July, 2002)***

by James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The CM3 parameter sets, which allow for evaluation of CM3 charges, were added to this version of *MN-GSM*. This provides the capability to evaluate gas-phase CM3 charges, liquid-phase CM3 charges, and the electrostatic contribution to the free energy of solvation using the generalized Born model. The corresponding CM3/GB gradient may also be evaluated and used for geometry optimizations. The redistributed Löwdin population analysis (RLPA) method was also added to this version of *MN-GSM*. This new method can be used to evaluate gas-phase RLPA charges and liquid-phase RLPA charges. The RLPA charges can further be used in a calculation of the electrostatic contribution to the free energy of solvation using the generalized Born model and of the corresponding free energy gradient, which can also be used for geometry optimizations. The namelist has been broadened to support the new options as well as the previous ones. The namelist is now called \$MNGSM (although the old name, \$CM2, may also be used if desired). To accommodate the new methods, the allowed values of the ICMD keyword have been extended to include 302, 303, 315 – 319, and 400. Also, the HFE keyword, which specifies the percentage of HF exchange used in the mPW exchange functional, has been added.

### ***MN-GSM-v.2.2 (September, 2002)***

by James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The CM3 parameter sets for BLYP/6-31G(d), B3LYP/MIDI!6D, B3LYP/6-31G(d), and B3LYP/6-31+G(d) were implemented. This provides the capability to evaluate gas-phase CM3 charges, liquid-phase CM3 charges, and the electrostatic contribution to the free energy of solvation using the generalized Born model using these wave functions. The corresponding CM3/GB gradient may also be evaluated and used for geometry optimizations. The keyword ICMD was extended to allow values of 313, 314, 320, and 321 to accommodate the new methods.

### ***MN-GSM-v.3.1 (November, 2003)***

by James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The keyword SolvRd was added. This keyword allows the user to specify a value for the solvent radius, which is used for the calculation of the solvent-accessible surface areas of the atoms of the solute.

### ***MN-GSM-v.5.1 (June, 2005)***

by Casey P. Kelly, James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The keywords IRADII and ISTS was added. The IRADII keyword specifies the set of atomic-number-dependent radii that are used to build the molecular cavity. The ISTS keyword determines whether SM5- or SM6-type functionals are used. The data array “Coulomb”, which held the values of the intrinsic Coulomb radii used by SM5.42, was renamed Coulom2. The data arrays Coulom3, and Coulom4 were added. These data arrays contain the Coulomb radii used by the SM5.43, and SM6 models, respectively. In previous versions of the code, a default value of 0.0 Å was assigned to elements for which an intrinsic Coulomb radii had not been optimized. In version 5.1, the default atomic radius for elements for which an intrinsic Coulomb radius has not been optimized is Bondi’s value for the van der Waals radius when available, and 2.0 Å for all other atoms. The keyword LPrtCoulRd was removed. The atomic radii used in the SCRF calculation (intrinsic Coulomb radii) are now printed out by default. The keywords CM2Save and CM2Info were changed to CMxSave and CMxInfo. The extensions of the files resulting from the use of these keywords are renamed CMx and CMxI, respectively. In previous versions of *MN-GSM*, the LCoulRd keyword could be used to assign different atomic radii to different hydrogen atoms in a molecule. This keyword has been modified so that all atoms (not just hydrogen) in a given molecule can be assigned atomic-number-independent radii. The default value for the solvent radius (SolvRd keyword) was changed from 0 Å to 0.40 Å. The CM3 and CM3.1 parameter sets for HF/MIDI! were added (ICMD = 301 and ICMD = 322, respectively). The CM4 parameter sets for DFT/MIDI!6D (ICMD = 416), DFT/6-31G(d) (ICMD = 417), DFT/6-31+G(d) (ICMD = 418), and DFT/6-31+G(d,p) (ICMD = 419) were added. The SM5.43 parameter sets for aqueous and organic solvents for HF/6-31G(d) (ICDS = 303), B3LYP/6-31G(d) (ICDS = 313), MPWX/MIDI! (ICDS = 315), MPWX/MIDI!6D (ICDS = 316), MPWX/6-31G(d) (ICDS = 317), MPWX/6-31+G(d) (ICDS = 318), and MPWX/6-31+G(d,p) (ICDS = 319) were added. The SM6 parameter sets for aqueous solution for DFT/MIDI!6D (ICDS = 416), DFT/6-31G(d) (ICDS = 417), DFT/6-31+G(d) (ICDS = 418), and DFT/6-31+G(d,p) (ICDS = 419) were added.

### ***MN-GSM-v.5.2 (August, 2005)***

by Casey P. Kelly, James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

SCF Scheme I is no longer available for liquid-phase calculations that involve the use of diffuse basis functions, due to convergence problems. Four new test jobs have been added to the test suite (13a-13d) that demonstrate this. A warning message has been added in the subroutine ConOpn. It warns the user that if appropriate SCF convergence options are not specified, the liquid-phase part of the calculation will be skipped.

### ***MN-GSM-v.5.3 (Feb, 2006)***

by Adam C. Chamberlin, Casey P. Kelly, James D. Xidos, Jiabo Li, Jason D. Thompson, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

Minor corrections to the code were implemented. The atomic surface tensions used to compute the temperature dependence of the free energy of solvation have been added. A new keyword (SolK) to read in the temperature of liquid aqueous solution and to compute free energies of solvation at the given has been added. The keyword only accepts values within the range of 273 to 373 K. A new keyword (ReadK) to read in temperatures from a file named *JobName.kel* and to compute free energies of solvation for a particular molecule in aqueous solution has been added. Note that this approach used the solution phase electronic structure as an initial guess for successive temperatures, so one may obtain identical electronic energies if the temperatures are very near one another. To avoid this either use very strict convergence criteria, or do not order similar temperatures very near one another. The differences in electronic energies are very small anyway. Another keyword (MeanK) which is similar to ReadK computes the electrostatics by computing the electronic energy of the solute in solution at the average of all the temperatures in the file *JobName.kel* and then uses a scaling factor to compute the electronic energy at a given temperature. Note this option is considerably faster than the ReadK option. An additional test suite for SM6T has been added.

#### ***MN-GSM-v. 6.0 (April, 2006)***

by Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version of *MN-GSM* is compatible with *Gaussian 03* Revision D.01 and *MN-GFM* version 1.0. This version of *MN-GSM* is no longer compatible with *Gaussian 98*. The following features were deemed obsolescent, and have been removed from this version of *MN-GSM*:

- The \$CM2 namelist has been removed. All *MN-GSM* calculations must now use the \$MNGSM namelist.
- The keywords CDSOnly, CM1Info, CMxInfo, and CDSInfo have been removed.

#### ***MN-GSM-v. 6.1 (October, 2006)***

by Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version of *MN-GSM* is compatible with *Gaussian 03* Revision D.01 and *MN-GFM* version 2.0.1. In previous versions of *MN-GSM*, specifying ISCRF=0 in the \$MNGSM namelist and opt in the *Gaussian* route section resulted in a program error and subsequent exit. In this version, these two keywords can be specified simultaneously, which allows for Löwdin, RLPA, and/or CMx charges to be calculated at each step of a gas-phase geometry optimization. A test job (test28) has been added to the test suite to demonstrate this new capability.

#### ***MN-GSM-v. 6.2 (October, 2007)***

by Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version of *MN-GSM* is compatible with *Gaussian 03* Revision D.01 and *MN-GFM* version 2.0.1. Corrected a bug in prior versions of *MN-GSM* beginning with version 5.1. The bug caused the gradients to be calculated incorrectly.

### ***MN-GSM-v2008 (July, 2008)***

by Ryan M. Olson, Aleksandr V. Marenich, Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version of *MN-GSM* is compatible with *Gaussian 03* Revision D.01/E.01 and *MN-GFM* version 3.0. All major modifications of the code are listed below:

- The SM8 model was added.
- The CM4M model was added.
- The SM8T model was added.
- The storage of CDS parameters for each solvation model was organized through individual subroutines.
- The MN-GSM output was made to be more user-friendly.
- The CoulR option was removed. If LCoulR=True, the user must provide a file \*.coul that should contain user-defined Coulomb radii for all of the atoms in a molecule.
- If CMxSave=True, MN-GSM will print CMx charges into a file \*.chg (instead of \*.cmx as it was before). The same file can be used if ChgRead=True.
- The ISTS option is deprecated.
- The IRADII option can be omitted unless there is a need to overwrite the default value of IRADII for a particular solvation model.
- The maximum number of atoms has been extended to 250 instead of 128.
- The maximum number of basis functions has been extended to 2500 instead of 1000.
- A bug in subroutine SetSol in all of the prior versions of *MN-GSM* was corrected. The bug was related to an undeclared array and it caused a stop in running the SCRF calculation in the code compiled by PGI version 6.2 on IBM Blade.
- The previous version of MNGSM stopped running at the first SCF cycle without a solvation calculation if the user used accurate molecular orbitals as a guess from a checkpoint file. This bug has been fixed.
- The SM5.43 sigma parameters in the array called SDEPTA were ordered in the wrong way that has been fixed now (see the string DATA SDEPTA). This bug caused wrong CDS contributions for any solute in any nonaqueous solvent for ICDS = 303 – 319.
- The O-O CDS expression for IST=6 [STS(I)=STS(I)+RTKK\*SIGMAX(104)] was corrected to [STS(I)=STS(I)+COT(IJ0(I,J))\*SIGMAX(104)]. This bug was found by comparing the gradients corresponding to the bound oxygen atoms (like in peroxides) calculated numerically and analytically.
- Bondi's van der Waals radius of Br in Data VANWAL was incorrect. The wrong value (1.80 Å) was replaced with the correct one (1.85 Å).
- Equation (17) in the manual was corrected by the term of NAtoms + 600 that corresponds to the addition of the temperature-dependent model's parameters.

### ***MN-GSM-v2010 (September, 2010)***

by Ryan M. Olson, Aleksandr V. Marenich, Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

This version is compatible with *Gaussian 03* Revision E.01 and *MN-GFM* Version 4.1. The only major modifications of the code in this version is that the SM8AD model (energies and gradients) was added.

### ***MN-GSM-v2011 (June, 2011)***

by Ryan M. Olson, Aleksandr V. Marenich, Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The *MN-GSM* module is now compatible with both *Gaussian 03* (revision E.01) and *Gaussian 09* (revision A.02). A new installation kit has been created. The *MN-GSM* package does no longer include the *MN-GSM* modified versions of 1502.F and 1701.F. Instead, there is a special script that creates these modified files automatically from their original *Gaussian 03/09* versions. The user just needs to re-compile the modified versions of 1502.F and 1701.F as usual. An additional feature is a special tool to calculate solvatochromic shifts using the generalized Born version of the GSRF, IBSF, IESRF, and VEM(d,RD) models.

### ***MN-GSM-v2012 (November, 2012)***

by Ryan M. Olson, Aleksandr V. Marenich, Adam C. Chamberlin, Casey P. Kelly, Jason D. Thompson, James D. Xidos, Jiabo Li, Gregory D. Hawkins, Paul D. Winget, Tianhai Zhu, Daniel Rinaldi, Daniel A. Liotard, Christopher J. Cramer, Donald G. Truhlar, and Michael J. Frisch.

The SM12 parameters have been added, and a special tool called *gbsolg09* to run an iterative SM12 calculation using CM5, CHELPG, or MK charges has been created.

## **Further Information**

If and when needed, updated information will be made available at the University of Minnesota Computational Chemistry Web site: <http://comp.chem.umn.edu/mn-gsm>.

## **Acknowledgments**

The authors are grateful to Shuxia Zhang, Yao-Yuan (John) Chuang, Patton Fast, and Laura Masgrau for assistance. The authors are also grateful to Benjamin J. Lynch for the correction to the mPW hybrid functional, which was used for the CM3 and CM4 parameterizations and the development of the RLPA method.