

CHEMSOL 2.1

Jan Florián and Arieh Warshel
Department of Chemistry
University of Southern California
Los Angeles, CA 90089-1062

November 3, 1999

1. Introduction

Program ChemSol is designed for the calculations of solvation free energies using Langevin Dipoles (LD) solvation model, in which the solvent is approximated by polarizable dipoles fixed on a cubic grid. The implementation and parametrization for aqueous solution were described in the paper "*Langevin Dipoles Model for Ab Initio Calculations of Chemical Processes in Solution: Parametrization and Application to Hydration Free Energies of Neutral and Ionic Solutes, and Conformational Analysis in Aqueous Solution*".¹ The ChemSol 1.0 and 1.1 programs were used in studies of the chemical reactivity²⁻⁴, binding⁵, and conformational flexibility⁶ in aqueous solution. The extension of the predictive capabilities of the LD model to hydration entropies has been implemented in the 2.0 and 2.1 versions of the program.⁷

Copies of the program ChemSol 2.1 can be downloaded free of charge from anonymous ftp at: ftp.usc.edu, directory pub/warshel/cs2 or from a link in <http://apriori.usc.edu/florian>.

Specific comments and questions about running ChemSol should be addressed to Jan Florian (e-mail: florian@usc.edu, phone:(213) 740-7671).

- (1) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
- (2) Florián, J.; Warshel, A. *J. Am. Chem. Soc.* **1997**, *119*, 5473 .
- (3) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719 .
- (4) Florián, J.; Åqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 11524 .
- (5) Florián, J.; Sponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884 .
- (6) Florián, J.; Strajbl, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 7959.
- (7) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, in press.

2. List of ChemSol files

| | |
|-----------------|---|
| cs20_manual.rtf | this file |
| cs21.f | source file (Fortran 77) |
| cs | simple script to run cs |
| vdw.par | standard parameter file |
| cs.arc | archive file |
| test1.cs | input file with one sets of charges and default vdW radii (protonated cytosine) |
| test2.cs | input file with two sets of charges and user-defined vdW radii (methanol) |
| test*.log | corresponding output files |


```

4      atom, zan, q, x, y, z (format(1x,A8,F8.1,2F10.4,3F9.4))
      |      |      |      |      |
      |      |      |      |      | Cartesian coordinates
      |      |      |      |      |
      |      |      |      |      | Atomic charge
      |      |      |      |      |
      |      |      |      |      | Nuclear charge
      |
      |_____ Atom name (First two characters in the name
                  determine vdW radius on the atom -see part 2.)

```

Line 4 is repeated for each atom (i.e. natom-times)

```

5      keyword ('rp', 'pcm' or empty line (see below))

```

Keyword '*pcm*' (format a3, lower-case) used if corrections for solute polarization are to be evaluated explicitly (see example). This keyword must be followed by the atomic charge information in the format described above.

Keyword '*rp*' (format a2, lower-case) indicates that van der Waals (vdW) radii of some atoms will be changed from their standard values. If no parameter change is required, 'rp' keyword is skipped and atom specification is followed by an empty line.

```

6      n_new _____ number of atoms for which new vdW radii
                        will be specified
7      i_new  vdw_new (free format)
      |      |_____ value of the vdW radius
      |_____ sequential number of the atom for which
                        new vdW radius is defined.
8      empty line. (always required)

```

Atomic charges (*q*)

We suggest to use atomic charges that were obtained by fitting to the electrostatic potential (ESP) of the solute calculated from the PCM B3-LYP/6-31G*//HF/6-31G* wavefunction. However, another DFT or ab initio method/basis set/geometry may be used, possibly with user-defined van der Waals (vdW) atomic radii. Pauling atomic radii scaled by 1.2 are recommended for the polarized continuum model (PCM) calculation.

A possible command line for the calculation of atomic charges using the Gaussian94 program is

```
#B3LYP 6-31G* scrf=tomasi iop(1/11=200) population=(mk,dipole)
```

....

80. 400

A possible command line for the calculation of atomic charges using the Gaussian98 program is

```
#B3LYP 6-31G* scrf=oldpcm iop(1/11=200) population=(mk,dipole)
```

....

80. 400

or

```
#B3LYP 6-31G* scrf=(pcm,read) population=(mk,dipole)
```

....

```
radii=pauling alphah=1.20 ret=10 tnum=100
```

If one is not interested in the explicit evaluation of the solute polarization contribution (ΔG_{relax}) to the total hydration free energy (ΔG_{solv}), a single set of PCM charges calculated as described above should be given in the ChemSol input file. The solute polarization term will be implicitly accounted for in the magnitude of ΔG_{solv} for such calculations

ΔG_{relax} term can be explicitly calculated if both the gas-phase and PCM charges are specified on input (see test2.cs input for an example of such a calculation). First, gas-phase charges are given, followed by the keyword 'pcm' and the corresponding PCM charges. Both sets of charges can be generated in a single Gaussian94 calculation using the command line given above.

Choice of atom types

The default atomic vdW radii are determined from the value of the nuclear charge specified in the input file. However, for more accurate prediction of solvation energies, atom names also need to carry information about atomic hybridization in a given solute. This is done by using 2-character symbols defined in Table 1 (see below). If the first 2 characters of the atom name in the input file do not correspond to any atom type listed in Table 1, default van der Waals radii for the given nuclear charge will be used for the LD calculation.

Procedure to generate ChemSol input data from the Gaussian 94 /98 programs.

To avoid cumbersome manual input of the atomic charges and coordinates in the format required by the ChemSol, it is recommended that ChemSol users modify the ESPFit subroutine in the l602 link and recompile the l602.exe as described below:

i/ Insert the lines

```
Dimension IEL(0:104)
1030 Format(//,10x,'ESP atom charges and Cartesian coordinates')
1035 Format(i4)
1040 Format(1x,A2,6x,F8.1,2F10.4,3F9.4)
```

in an appropriate (upper) part of the ESPFit subroutine.

ii/ Insert the following instructions in the end of the ESPFit subroutine (starting from the CHF comment line):

```
else
  Call AClear(NAtoms,Q)
  Do 305 I = 1, NATFit
305   Q(IAtUse(I)) = Scr(I)
  Call AtomPr(NAtoms,IAn,Q,NAtoms,1,1)
CHF  Print atom coordinates with charges
     b2a=0.52918
     Call Fillel(0,104,IEL)
```

```

write(Iout,1030)
write(Iout,1035) NatFit
do i=1,NatFit
ii=ian(i)
write(Iout,1040) IE1(ii),float(ian(i)),
$ q(iatuse(i)),(b2a*c(j,iatuse(i)),j=1,3)
end do
endIf
Return
End

```

iii/ Recompile the link 1602 (see G94 User's Reference, page 206).

iv/ The modified link 602 prints the atomic charges and coordinates in the ChemSol format directly into the standard G94 output file (*.log), provided the 'population' keyword was specified on input.

Alternatively, one can use a conversion program to extract the necessary information for the generation of the ChemSol input file from the g94 or g98 outputs. This conversion code is currently not included in the ChemSol 2.0 package, but it can be emailed to interested users upon request.

4. The parameter input file (vdw.par)

This file contains the values of selected parameters of the LD-CS2 model.

```

1 25 0.20 0.90 0.48 1.4 ! iterld, ndxp, dxp0(1), clgvn, slgvn, tds0 (
i2,i3,4f6.3)
0.88 2.65 3.00 2.65 2.65 ! srp(H), rp(C0), rp(C1), rp(N0), rp(N1) (5f6.3)
2.32 2.65 2.80 3.20 3.20 ! rp(O0), rp(O1), rp(O2), rp(P), rp(S0) (5f6.3)
2.46 3.16 ! rp(F), rp(Cl) (5f6.3)
1.0 1.0 0.5 0.5 1.0 ! vdwc6: H, C0, N, O, P
1.0 1.0 ! vdwc6: S, Cl
0.120 0.138 0.002 0.038 1.5 ! vdwsl, 10*phobsl, ephil1, ephil2, rzcut (5f6.3)

```

iterld = 0 -----> noniterative LD calculation (iterld = 0 is not supported in cs2)

1 -----> iterative LD calculation

ndxp -----> dG is calculated as an average over ndxp grids that differ
in their positions wrt the solute (use ndxp 20 - 50 for
production calculations, and ndxp < 10 for test purposes).

dxp0(1) -----> grid shift

slgvn -----> scale factor for iterative lgvn energy

slgvn*clgvn-----> scale factor for noniterative lgvn energy

tds0-----> a constant part of the hydration entropy (kcal/mol)

rp -----> Van der Waals radii for selected atoms

vdwc6 -----> London coefficients

Note that *rp* values (vdw radii) listed in the *vdw.par* correspond to the LD-CS2 parametrization. The changes of these parameters can be done in the solute input file (using the keyword 'rp'). Alternatively, global parameter changes can be done for selected atoms in the *vdw.par* file. The values of vdW radii present in the input file overwrite their values defined in the *vdw.par* file and the source code.

Table 1: Definition of atom types

C0 = C.....sp3 carbon (default)
 C1.....sp2 and sp carbon
 N0 = N.....sp3 nitrogen (default)
 N1.....sp2 and sp nitrogen
 O0 = O.....sp3 oxygen (default)
 O1.....sp2 oxygen
 O2....."inorganic" oxygen
 S0 = S.....sulphur
 H0 = H.....hydrogen (default), $rp(H0) = srp(H) * rp$ of the closest nonhydrogen atom

5. Output data

Results of the calculations are written in two files. The detailed information about ChemSol run is written in the **.log* file. In addition, if the ChemSol run was successfully completed, a single output line is appended in the end of *cs.arc* file. This short output information will be sufficient for most purposes.

Description of the results written in the *cs.arc* file

| Molecule | lgvn | VdW | -TdS | Relax | Born | dHsolv | dGsolv |
|---------------|-------|------|------|-------|-------|--------|--------|
| CytN3H+, 631g | -56.1 | -5.6 | 5.3 | -1.6 | -10.1 | -71.8 | -66.5 |

lgvn Electrostatic part of the solvation free energy in kcal/mol.
VdW London (dispersion) part of the free energy of solvation (kcal/mol).
-TdS Hydration entropy (kcal/mol).
Relax Contribution of solute polarization to the solvation free energy (kcal/mol). This term is always included in the 'lgvn' energy given in the first column. Relax is reported only if gas-phase charges were specified in the ChemSol input file.
Born Continuum correction for a finite size of the sphere of point dipoles (kcal/mol)
dHsolv Hydration enthalpy (kcal/mol)
dGsolv Hydration free energy (kcal/mol).

The total solvation free energy:

$$\Delta G_{\text{solv}} = \Delta H_{\text{solv}} - T\Delta S = \text{lgvn} + \text{VdW} + \text{Born} - T\Delta S$$