

Description of the MULFIT program

February 1998

by G. G. Ferenczy, C. A. Reynolds and P. J. Winn

modified by A. J. Stone, 2006–2009

1 Introduction

This program is the realisation of a method for calculating electrostatic potential-fitted charges (Ferenczy 1991, Chipot *et al.* 1993, Winn *et al.* 1997, Ferenczy *et al.* 1997). In addition to the fit of charges to potentials, higher rank moments and the derivatives of the potential can also be involved in the calculation. In contrast to the usual way of fitting charges to the electrostatic potential from the wave-function, the present method uses a set of distributed multipole series to represent the wave-function and charges are fitted to the electrostatic potential of these multipole series (reference moments). Moreover, instead of directly calculating the electrostatic potential, the difference between the potential of the reference moments and fitted charges (or in general fitted moments) is evaluated analytically. Thus the distributed multipole series (typically obtained from the wave-function by a distributed multipole analysis (Stone 1981, Stone and Alderton 1985)) are input to the program and another distributed multipole series is calculated, using lower rank multipoles (e.g. charges), and possibly with different sites. The calculation involves a separate fitting procedure for each reference multipole series. The fitted multipoles are the best ones (in least square sense) reproducing the electrostatic potential of the reference multipoles (and optionally its derivatives). The final fitted multipole moments are obtained as the sum of multipoles at the same site fitted to different reference series.

2 Input

The program takes data from standard input and writes the results to standard output. The data file is keyword-driven. In the following, the keywords are all shown in upper case, but they can be given in upper, lower or mixed case in the data file. Items shown enclosed in square brackets are optional. If they are omitted, sensible default values will be used. The square brackets should not appear in the data.

TITLE

The following lines (up to 10, terminated by a blank line) are taken as title for the calculation.

UNITS { ANGSTROM | BOHR } Lengths and positions that follow in the data (including the DMA file) will be in the units specified. Default is Angstrom. This command also specifies the units to be used for lengths and positions printed by the program. It may be repeated to specify different units for different sections of the data.

DMA [source] [filename]

Read the reference multipole data. If a filename is provided, the data is read from the specified file. Otherwise the data follows. The *source* may be CADPAC, GDMA (default) or MULPOL; in the first two cases the format expected is that produced by the punch output option of CADPAC or GDMA respectively, while the third expects the output from the MULPOL program.

The GDMA output, by default, names each atom by the chemical symbol corresponding to its atomic number, but it is possible to define other names. (See the GDMA documentation.) MULFIT takes the name (or its first 4 characters, if longer) to specify the atom type. Alternatively the GDMA output may be edited to add a type definition. For example, if most H atoms are to be fitted with charges only, but the H atoms of an NH₂ group are to be fitted with charge+dipole, perhaps because they are involved in

hydrogen bonding, the GDMA output for these atoms might be edited to read

```
H1      -0.6423983213  -1.5273188918  -0.0000000000  Rank 6  Type Hb
... multipoles ...
H2      -1.9172826281  -0.3507197384  -0.0000000000  Rank 6  Type Hb
... multipoles ...
```

Types corresponding to the first 18 element names are predefined. The RANKS command (see below) should specify the rank of multipoles to be assigned to each type of atom.

TOTAL [RANK *L*] [ORIGIN *x y z*]

Calculate the total molecular multipole moments for the fitted and reference distributed multipoles. The moments are calculated up to the rank specified (default 3) and are referred to the specified origin (default (0, 0, 0)).

RADII

This keyword is followed by lines assigning radii to the atoms. Values should be given in Angstrom, unless changed by the UNITS command. The default is the van der Waals radius for the atom, if known by the program, or otherwise 2 Å. These values may be modified by lines following the RADII keyword, as follows:

t r

Set the radius for atom type *t* to be *r*.

The RADII section is terminated by a blank line.

ADD *incr*

Normally, only those fitting centres occurring within the specified radius from a particular atom are included in the fitting procedure for its reference moments. The inclusion radius is increased by *incr* (for all atom types) if this option is given.

The value of a radius is typically the van der Waals radius for the centre. (Reference centres are usually atoms.) The use of radii which involve the nearest (and possibly the next nearest) neighbours is recommended. In this way the reference multipole moments are reproduced by fitted moments on the nearby centres and the fitted moments reflect the chemical environment. The use of larger radii improves the fit, but results in loosely determined and chemically meaningless fitted moments.

QUIET

VERBOSE *n*

These commands control the amount of output. In the VERBOSE command, *n* = 0 specifies the minimum output (equivalent to QUIET), while higher values, up to 3, given increasing amounts of output. The default is *n* = 0.

OVERALL

CUMULATIVE

These commands have no effect when only fitted charges are calculated. When different rank moments are calculated then the fit can be an overall or a cumulative one. In the first case, different rank moments are fitted to a particular reference series in one fitting procedure. In the second case, the lowest rank (usually charge) moments are fitted first. Then the next lowest rank (usually dipole) moments are fitted to reproduce that part of the potential which is not reproduced by the lowest rank moments; and so on with the higher moments. The advantage of such a procedure is that charges from a charge + dipole fit will be the same as charges obtained from a charge fitting procedure. When charges and higher rank moments are fitted together in an overall fitting procedure then a better fit can be achieved but the resulting multipoles have less chemical significance. The default is to use a cumulative fit.

SHELL [INNER *x*] [OUTER *y*]

The parameters x and y define the region of space where the reproduction of the potential (and its derivatives) is required. The potential around each atom is fitted in a spherical shell between $r + x$ and $r + y$, where r is the radius given for that type of atom on the VDW card. Both x and y must be non-negative and y has to be larger than x . Default values are $x = 1.2 \text{ \AA}$, $y = 3.0 \text{ \AA}$.

LIMIT L

The distributed multipole data used as input may include high-rank multipoles. This option specifies the highest rank that is to be included in the fit. Any higher moments are ignored. If a reference multipole site has moments up to a lower rank than L , zero is assumed for the missing moments. Default value is $L = 6$.

TENSORS $n w_0 [w_1 [w_2]]$

n is the highest derivative of the potential to be used in the fit. That is, if $n = 0$, only the potential is used; if $n = 1$, the electric field is fitted too; and if $n = 2$, the potential, the field and the field gradient are all used. w_0 , w_1 and w_2 are the weights to be associated with potential, field and gradient, and must all be between 0 and 1. The higher the weight the better the reproduction of the corresponding quantity. Default values: $n = 0$, $w_0 = 1.0$.

ORIENT *file* Print the results to the specified file in a form suitable for input to the Orient program.

RANKS

This data section specifies the ranks of fitted multipole that may be used for each type of atom in the fitting procedure. Each line following the RANKS keyword is of the form

type n1 n2 ...

Here *type* is an atom type, and the numbers that follow are the ranks of multipole that may be fitted to atoms with that atomic number. For example,

C 0 2

Hb 0 1

H 0

specifies that charges (rank 0) and quadrupole (rank 2) may be used on carbon atoms, charges and dipoles on atoms of type Hb, and only charges on other hydrogen atoms. The last line could be omitted, since the default is to fit only charges to each type of atom. The maximum rank that may be specified is 4.

This section is terminated by a blank line. The fitting is carried out immediately, so any options must be specified before the RANKS section. It is an error if no DMA specification has been provided. There may be further RANKS sections, preceded if necessary by different options.

3 Output

The goodness of fit parameter is designed to indicate the quality of the result. Experience suggests that it should be viewed cautiously. The comparison of parameters obtained in an overall and in a cumulative fit make no sense. The calculation of this parameter includes a division by the number of multipole moments to be reproduced, which is higher in a cumulative fit since it includes separate fits for each rank of fitted moments.

The agreement between reference and fitted total moments is a good indicator of the quality of the result.

4 Author

All notes and inquiries about this program are to be addressed to

Gyorgy Ferenczy
Chinoin, Sanofi-aventis
1045 Budapest,
To utca 1-5
Hungary
e-mail: gyorgy.ferenczy@sanofi-aventis.com

References

- Chipot, C., Ángyán, J. G., Ferenczy, G. G. and Scheraga, H. A. (1993) 'Transferable net atomic charges from a distributed multipole analysis for the description of electrostatic properties. a case study of saturated hydrocarbons.' *J. Phys. Chem.* **97**, 6628–6636.
- Ferenczy, G. G. (1991) 'Charges derived from distributed multipole series,' *J. Comp. Chem.* **12**, 913–917.
- Ferenczy, G. G., Winn, P. J. and Reynolds, C. A. (1997) 'Toward improved force fields. 2. Effective distributed multipoles,' *J. Phys. Chem. A* **101**, 5446–5455.
- Stone, A. J. (1981) 'Distributed multipole analysis; or how to describe a molecular charge distribution,' *Chem. Phys. Lett.* **83**, 233–239.
- Stone, A. J. and Alderton, M. (1985) 'Distributed multipole analysis—methods and applications,' *Molec. Phys.* **56**, 1047–1064.
- Winn, P. J., Ferenczy, G. G. and Reynolds, C. A. (1997) 'Toward improved force fields. 1. Multipole-derived atomic charges,' *J. Phys. Chem. A* **101**, 5437–5445.