

# INPUT DESCRIPTION FOR SQM version 2.0

## INTRODUCTION

SQM is an add-on module for the PQS program which scales force constants to produce a Scaled Quantum Mechanical (SQM) Force Field. This can correct for deficiencies in the calculated (harmonic) force constants, giving a better fit to experimentally observed vibrational fundamentals and infrared (IR) intensities. The method has considerable predictive power and is often of great help in understanding and assigning experimental vibrational spectra.

The first *ab initio* calculations of harmonic force constants were carried out at the Hartree-Fock level of theory. Hartree-Fock theory overestimates harmonic force constants significantly and empirical scaling is needed to produce force constants which can reproduce experimental vibrational frequencies. The scaling compensates for basis set deficiencies, anharmonicity and mostly for the lack of electron correlation. The need for empirical correction diminishes but is not completely eliminated if the quality of the wavefunction improves by adding electron correlation and increasing the size of the basis.

The first scaling methods applied to *ab initio* force constants used several different scale factors to correct for systematic errors in different types of molecular deformations, e.g., stretches, bends and torsions. This procedure requires the transformation of the molecular force field (the Hessian matrix) to chemically meaningful internal coordinates and cannot be applied directly to the calculated frequencies. It is thus less convenient than global scaling using a single scaling factor. Global scaling can be applied directly to the frequencies, the scale factor for frequencies being near 0.9, corresponding to a scale factor of 0.81 for force constants. Because of its simplicity, global scaling became popular, but using multiple scale factors yields much better results as was convincingly demonstrated by Blom and Altona in a series of papers starting in the mid 1970s [1]. Their method forms the basis of the SQM procedure which has been in widespread use for over 20 years [2].

In the original SQM procedure, the molecular geometry was expressed in terms of a full set of nonredundant natural internal coordinates [3]. Natural internals use individual bond displacements as stretching coordinates and localized linear combinations of bond angles and torsions as deformational coordinates. (They are the precursors to the delocalized internal coordinates used in PQS, which are linear combinations of *all* stretches, bends and torsions in the molecule [4].) On the basis of chemical intuition, the natural internal coordinates of all molecules under consideration are sorted into groups sharing a common scaling factor, and factors for each group are determined by a least-squares fit to experimental vibrational frequencies. Force constants, originally calculated in Cartesian coordinates, are transformed into an internal coordinate representation, and scaling is applied to the elements of the internal force constant matrix (*not* to the individual vibrational frequencies) according to

$$\mathbf{F}_{ij}(\text{scaled}) = (s_i s_j)^{1/2} \mathbf{F}_{ij}$$

where  $s_i$  and  $s_j$  are scaling factors for internal coordinates  $i$  and  $j$ , respectively.

The accuracy obtained by selective scaling in this way is naturally greater than if just a single overall scaling factor were used. Additionally, scaling the force constant matrix also affects the resultant normal modes, and hence the calculated intensities (which are unaffected if only the frequencies are scaled), leading to better agreement with experimental intensities.

The SQM procedure has been widely used in the interpretation of vibrational spectra. A further important role is the development of *transferable* scale factors which can be used to modify calculated force constants and so predict the vibrational spectrum a priori.

The SQM module uses a modified scaling procedure involving the scaling of *individual* valence coordinates [5] (*not* the linear combinations present in natural internal coordinates). This has immediate advantages in terms of ease of use, as no natural internals need to be generated (a procedure which may fail for complicated molecular topologies), and it simplifies the classification and presorting of the coordinates. In addition, the extra flexibility involved in the scaling of individual primitive internals generally leads to an increase in accuracy and to more transferable scale factors.

The user is encouraged to view the references provided, especially ref. 5.

## PROGRAM CAPABILITY AND INPUT

SQM capabilities include

1. Scaling a force constant matrix using a set of one or more precalculated scale factors. Eleven optimized scale factors are available for standard organics containing H, C, N, O and Cl for force constants calculated at B3LYP/6-31G\*. This is one of the most cost-effective and reliable theoretical methods currently available. The recommended scale factors are [5]:

stretch	X-X	0.9207
stretch	C-Cl	1.0438
stretch	C-H	0.9164
stretch	N-H	0.9242
stretch	O-H	0.9527
bend	X-X-X	1.0144
bend	X-X-H	0.9431
bend	H-C-H	0.9016
bend	H-N-H	0.8753
torsion	all	0.9523
linear bends	all	0.8847

where X denotes a non-hydrogen atom (C, N or O).

2. Adjusting atomic masses to give normal modes, frequencies and IR and potentially Raman and VCD intensities for isotopomers.
3. Optimizing scale factors to give the best least-squares fit to a set of experimental vibrational frequencies.
4. Carrying out a total energy distribution analysis [6] to determine how much a given primitive (stretch, bend or torsion) contributes to a particular normal mode.
5. Determining invariant diagonal force constants for all the stretches, bends and torsions in a molecule
6. Carrying out a full vibrational and thermodynamic analysis based on the scaled force constant matrix. The analysis is the same as that in PQS.

## INPUT FILE

A sample input file for formamide (HCONH<sub>2</sub>) with all keywords shown is given below:

```
$molecules
formamide
$scaling
stre  X    X          1    0.9202    fixed
stre  C    H          2    0.9163    fixed
stre  N    H          3    0.9239
bend  X    X    X      4    1.0108    fixed
bend  X    X    H      5    0.9438    fixed
bend  H    N    H      6    0.8765
tors  H    X    X    X  7    0.9525    fixed
tors  H    X    X    H  7    0.9525    fixed
$print_ted  3.0
$print_level  4
$max_atoms 10
$end
```

**\$molecules:** file prefix names for <hess>, <deriv>, <aat> and <evib> files

SQM needs molecular geometries, force constant (Hessian) matrices, dipole (and possibly quadrupole) derivatives and atomic axial tensors as input. The latter are available following a PQS frequency run in the files `jobname.hess`, `jobname.deriv` and `jobname.aat` - in this example these files are called `formamide.hess`, `formamide.deriv` and `formamide.aat`. **These files must exist.** (SQM will still function if the <deriv> or <aat> files are missing, but no IR, Raman or VCD intensities will be available.)

Various molecules can be grouped together (for example to determine the best scale factors for the whole set). In this case all the file prefix names should be given, one per line with no blank lines, following the **\$molecules** keyword.

**\$scaling:**            scaling parameters

<scale type>   <atom types>   <scale group>   <value>   <action>

<scale type>:        one of `stre`, `bend`, `tors`, `linc`, `linp`  
(corresponding to stretches, planar bends, proper torsions, and colinear and coplanar bend, respectively. Out-of-plane bends (`outp`) are currently not accessible)

<atom types>:        up to four atomic symbols (X for any non-hydrogen atom)  
depending on the <scale type>

<scale group>:        scale factors with the same (integer) scale group number will be  
grouped together during any optimization of the scale factors.  
The scale group number should start at 1 and must be listed  
consecutively as shown in the example input

<value>:            initial scale factor value

<action>:            either `fixed` (for a fixed scale factor) or `optimize`  
The default if this field is blank is to optimize the scale factor

In the formamide example, there are seven scale factors, with the two different torsions in the molecule grouped together. The scale factors for the N-H stretch and the H-N-H bend will be optimized to give the best least-squares fit to a set of experimental frequencies, keeping all the other scale factors fixed at their initial input values.

**\*\*IMPORTANT\*\*** When applying the scale factors, SQM takes each scale factor in turn *in the order they appear in the input file* and scales any Hessian element that fits. Care must be taken that scale factors involving the use of "X" - for a general non-hydrogen atom - appear *first* in the list of each scale type (`stre`, `bend`, `tors`, `linc`, `linp`) as, if these appear *subsequent* to a specific atom type (say a C-C stretch), then the specific atoms will be taken as general non-hydrogen atoms and the wrong scale factor will be applied.

Note also that "X" *cannot* be used in place of a hydrogen atom. Thus if all torsions, say, in a given molecule/set of molecules are to be scaled with the same scale factor, then any torsions involving hydrogen must be specifically provided. Thus `tors X X X X` alone will *not* scale any torsions involving hydrogen; you also need to specify both `tors H X X X` and `tors H X X H` to scale these torsions.

The other input options are:

**\$print\_ted=<real>**: print threshold for total energy distribution analysis. This analysis gives the percentage contribution of each primitive stretch, bend and torsion in the molecule to each normal mode. The default if no value is given is 5.0, i.e., only primitives which contribute 5% or more to a given normal mode will be printed for that mode.

**\$print\_level=<integer>**: controls the amount of printout (larger integer - more printout). In particular, a value of 4 will print the normal modes. Values higher than 4 will progressively output more and more intermediate quantities constructed during the SQM procedure and are essentially for debug printout.

**\$max\_atoms=<integer>**: for allocating memory. Should be set to at least the number of atoms in the largest molecule under consideration (the default is 50).

**\$end** terminates input (**must** be present)

## The <evib> File

The <evib> file contains both the molecular geometry and, if scale factors are being optimized, the experimental vibrational frequencies. The *formamide.evib* file is given below:

```
$coordinates      angstrom
C      0.4121292508   0.0816374866   0.0000000000
O      0.4653658528  -1.1331720094   0.0000000000
H      1.3139166617   0.7266945306   0.0000000000
N     -0.7368390692   0.8133281236   0.0000000000
H     -1.6267160783   0.3334309776   0.0000000000
H     -0.7250294253   1.8221287816   0.0000000000
$point_group     cs
$nsymop          2
$frequencies      J.Raman Spectros.  25 (1994) 183
0.0
608.
646.
841.    0.0d0
1090.
1309.
1391.
1602.
1692.
2882.
3190.
0.0
$end
```

The **\$coordinates** section contains the geometry in Cartesian coordinates. The format is: atomic symbol X Y Z atomic mass (A8,2X,4F20.14).

This is essentially the same format as the PQS <coord> file *except that the atomic charge is missing*. Coordinates can be given either in bohr or angstrom; the units are specified following the **\$coordinates** string as shown. If no units are specified, the default is bohr.

### Atomic Masses

There are two ways of specifying isotopomers. The atomic mass can be given following the coordinates (as per the above format) or the isotope can be specified as a part of the atomic symbol, e.g., H-2 will use deuterium instead of hydrogen for that atom, C-13 will give carbon 13. If *neither* of these options is specified, then the isotopically averaged atomic mass will be used.

The SQM program has built-in isotopically averaged atomic masses for all elements up to and including Xenon (N=54). There are also up to four individual isotope masses for each of these elements (if an element has more than four isotopes, then the four with the highest percentage abundance are available). The atomic masses for any atoms not included in the above description **must** be specified in the **\$coordinates** section.

### Use of Symmetry

The inclusion of full point group symmetry is the major difference between SQM version 2.0 and previous versions. Knowing the molecular point group allows appropriate symmetry labels to be given to each vibrational mode and also allows for a proper thermodynamic analysis. The data required are the point group symbol and the number of symmetry operations.

**\$point\_group**=<string>: character string (maximum four characters) giving the point group symbol, e.g., cs, c2v, td etc....

**\$nsymop**=<integer>: the number of symmetry operations in the point group, e.g., 2, 4, 24 etc.....

Symmetry information can either be provided in the <evib> file following the geometry (as shown above) or it can be read from a PQS <sym> file, if one exists. Note that if a <sym> file does exist then it will be read and any symmetry information given in the <evib> file will be overwritten. If no symmetry data is given, C<sub>1</sub> symmetry (i.e., no symmetry) will be assumed. If there *is* symmetry and it is not used, then certain quantities calculated in the thermodynamic analysis will be incorrect.

SQM version 2.0 makes use of the same symmetry routines as PQS and we have made a conscious effort to make the output from SQM as close as possible to the corresponding output in PQS. Indeed if all you want to do is an SQM scaling of the force constant matrix using known scale factors, then this can now be done within PQS itself using the newly introduced **SQM** keyword (see the PQS manual).

The **\$frequencies** section contains the experimental (or other) frequencies that will be used in the least-squares fit. The format is <frequency> (in  $\text{cm}^{-1}$ ) <weight> with each frequency on a separate line.

**weight**=<real>: gives the weight each vibrational frequency is given in the least-squares fit. Frequencies that are not known accurately can be given a lower weight in the fitting; conversely frequencies that are regarded as being reliable or for which a good agreement is particularly desired can be given a higher weighting. The default if no weight is given is  $1000 \times$  inverse experimental frequency (in  $\text{cm}^{-1}$ ).

If the experimental frequency for a particular vibration is very suspect or if it is not known at all, it should be given a zero weight. The number of fundamentals a molecule has is given - for a non-linear system it is  $3N-6$ , where  $N$  is the number of atoms. Quite often there are *less* than  $3N-6$  vibrational fundamentals that are reliably known. In this case, the **\$frequencies** section should have one or more lines containing zeroes (which correspond to an unknown frequency with a zero weight in the fit). It may be necessary to vary the position of these zero lines in the input depending on the accuracy of the fit; note also that although frequencies should generally be given in increasing order, it may be that two fairly close values with different symmetries may need to be switched. This can often be easily detected if one mode is strongly IR active whilst the other is only weakly active or IR inactive; the theoretical mode with the large IR intensity should be fit to the experimentally IR active mode.

It is not so uncommon to find experimentally assigned bands that you simply cannot fit at all because they have, in fact, been misassigned. The SQM procedure, when used correctly with a good theoretical method (such as B3LYP/6-31G\*), usually gives average errors in band positions of around  $8 \text{ cm}^{-1}$ , and maximum errors of the order of  $20-30 \text{ cm}^{-1}$ . If you find maximum errors significantly outside this range, there is a good chance that the experimental assignment is wrong.

**\$end** terminates input (**must** be present)

In the formamide example, both the lowest and the highest frequency fundamentals are not known experimentally (hence the two zero lines) and the fundamental assigned experimentally at  $841 \text{ cm}^{-1}$  is considered to be unreliable and has been given zero weight in the fit. The source of the experimental data (J.Raman Spectros. **25** (1994) 183) has been given after the **\$frequencies** keyword as a reminder to the user.

## Invariant (“relaxed”) Force Constants

Force constants in internal coordinates can be obtained by a suitable transformation of the Cartesian force constant matrix. Diagonal elements of the internal coordinate force constant matrix give individual stretching, bending and torsional force constants. Unfortunately, the values of these internal coordinate force constants depend on which primitive internals have been chosen to describe the molecular geometry. For example, if a given stretch is present in two sets of internal coordinates (which contain different bends and/or torsions), both sets of which can be used to describe a molecular geometry, then the stretching force constant calculated using one set of coordinates will *not* have the same value in the other set. This fact is perhaps not as widely known as it ought to be.

However, if the force constants are defined *not* as the diagonal elements of the Hessian matrix in internal coordinates, but instead as the inverse of the diagonal elements of the *inverse* Hessian, then the two stretching force constants in the example above *will* be the same. In this way force constants in internal coordinates can be defined in an *invariant* way, independent of the precise choice of coordinates. The inverse Hessian matrix is known as the compliance matrix. For a fairly recent discussion on relaxed force constants see [7].

Invariant force constants defined in this manner can be output from the SQM program by setting **\$sprint\_level** to 4 in the SQM input file.

**\*\*IMPORTANT\*\*** Both invariant force constants and a total energy distribution of the normal modes will only be performed for *single* molecules, i.e., there must be only one molecule filename specified in the **\$molecules** section.

## Program Usage

```
sqm.x filename
```

where the input file should be `filename.inp` and only the file prefix (i.e., no `.inp`) should be given. Output will be in `filename.out` and files associated with the molecule(s) specified in the **\$molecules** section in `filename.inp`, namely `molecule.evib`, `molecule.hess` and (optionally) `molecule.deriv` and `molecule.aat` must exist.

The SQM program produces four temporary files: `sqm-temp1`, `sqm-temp2`, `sqm-temp3` and `sqm-temp4`. As these do not have unique names associated with the input only *one* SQM job should be run in a given directory at any one time.



## REFERENCES

---

- [1] (a) C. E. Blom and C. Altona, *Mol. Phys.* **31** (1976) 1377  
(b) C. E. Blom, L. P. Otto and C. Altona, *Mol. Phys.* **32** (1976) 1137  
(c) C. E. Blom and C. Altona, *Mol. Phys.* **33** (1977) 875; *ibid.* **34** (1977) 1137
- [2] P. Pulay, G. Fogarasi, G. Pongor, J. E. Boggs and A. Vargha,  
*J. Am. Chem. Soc.* **105** (1983) 7037
- [3] (a) P. Pulay, G. Fogarasi, F. Pang, J. Boggs, *J. Am. Chem. Soc.* **101** (1979) 2550;  
(b) G. Fogarasi, X. Zhou, P. W. Taylor, P. Pulay, *J. Am. Chem. Soc.* **114** (1992) 8191
- [4] J. Baker, A. Kessi and B. Delley, *J. Chem. Phys.* **105** (1996) 192
- [5] J. Baker, A. A. Jarzecki and P. Pulay, *J. Phys. Chem. A* **102** (1998) 1412
- [6] P. Pulay and F. Török, *Acta. Chim. Acad. Sci. Hung.* **47** (1965) 273
- [7] J. Baker, *J. Chem. Phys.*, **125** (2006) 14103