

LIPFR

Philippe B. Wilson and Ian H. Williams

University of Bath

LIPFR was created by IHW (University of Bath). Its function is to calculate isotopic partition function ratios (hereafter referred to as IPFRs) from a Hessian obtained by means of an electronic-structure calculation for a full molecular system comprising N atoms and corresponding to a stationary point on a potential energy surface.

LIPFR is written in Fortran90, with some elements of earlier versions, but compiles fluidly under standard GNU compilers (gfortran...).

The aim of this user guide is to gently introduce the individual to the functionalities of LIPFR and provide explicit descriptions of the input and output one can expect from standard LIPFR runs. We will begin by discussing the format of the input, suggesting possible edits to the source code, which is available from the SULISO_source repository on Github: https://github.com/pbw20/SULISO_source

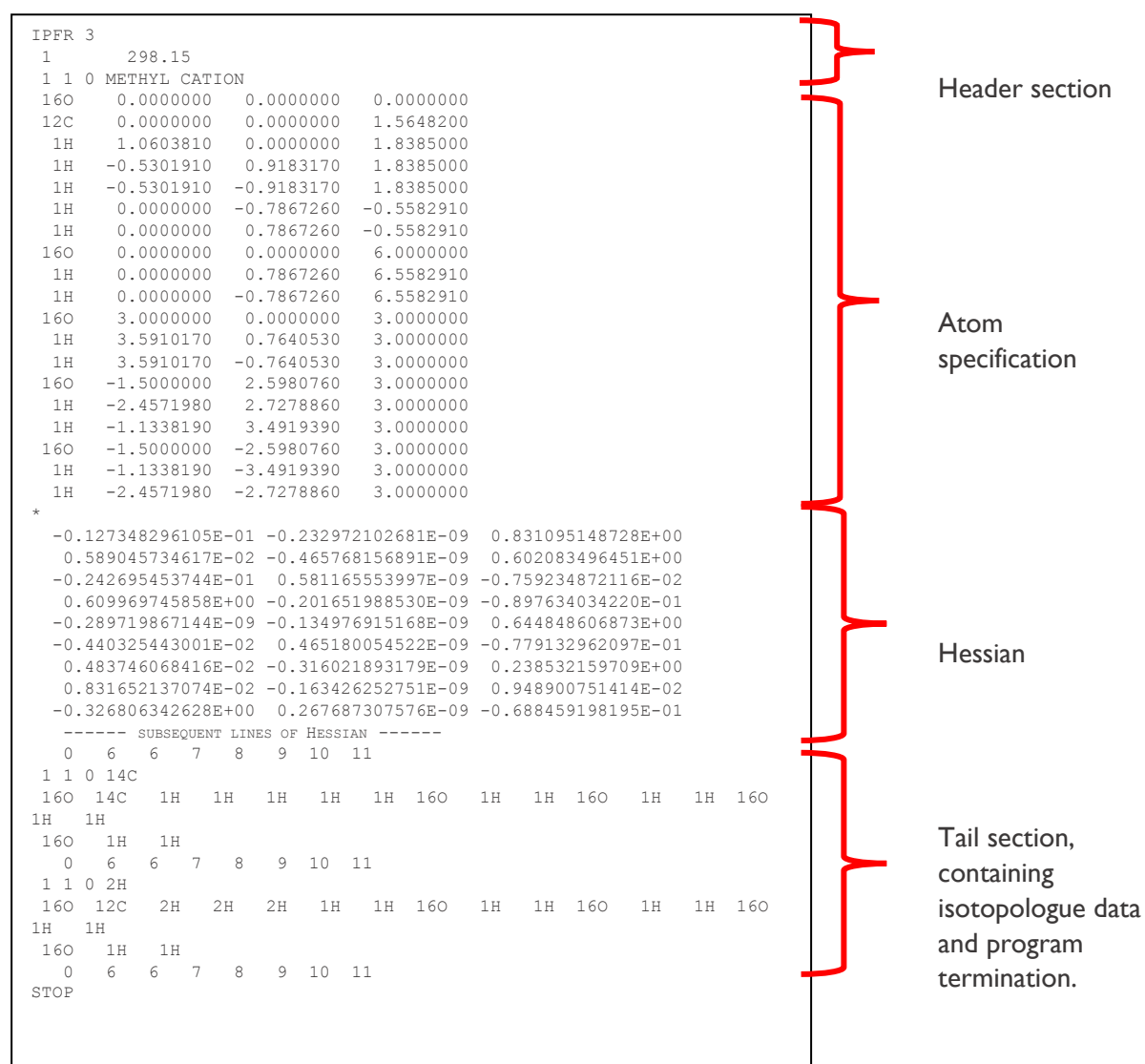


Figure 1. Sample input file for LIPFR. In the interest of space, only the first few lines of the Hessian are included, with the “----- SUBSEQUENT LINES OF HESSIAN -----” line indicating their original position.

Overview of LIPFR input

1. Header section, line 1: (FORMAT A4, I2, 1X, A4)

PROG, NISO, IFLIB

PROG takes the value 'IPFR' (A4) to start the program (or 'STOP' to finish).

NISO is the integer number (I2) of isotopologues **including** the parent.

IFLIB takes the value 'LIBR' (A4) if the 6 external degrees of freedom of the molecular species are to be treated as vibrations (or 'librations') instead of translations and rotations.

2. Header section, line 2: (FORMAT I2, 3X, 5F10.4)

NT, (T(I), I=1, NT)

NT is the number of temperatures (I2) and the values of

T are the absolute temperatures in Kelvin (5F10.4) to be used in the IPFR evaluations.

3. Header section, line 3: (FORMAT 3I3, 74A)

IS, IG, IFFREQ, TITLE

IS is a symmetry number (as appropriate, or set equal to 1 if symmetry is not required),

IG is the electronic degeneracy (usually = 1),

IFFREQ is the number of frequencies to be input (= 0 if a Hessian is supplied, but otherwise as appropriate).

TITLE is text string used as a description for the file or calculation.

4. Coordinate input, one line per atom, from line 4 to line 4 + N, where N is the number of atoms: (FORMAT A4, 1X, 3F12.7).

IA, (X(J, N), J=1, 3)

IA is the element symbol (right-justified in the A4 field) for an atom, and is followed by a space; X are the x, y, and z Cartesian coordinates in Ångström (3F12.7).

The following element symbols are currently accepted, although the UJISO source code may be edited to make additions or changes:

1H, 2H, 3H, 12C, 13C, 14C, 14N, 15N, 16O, 17O, 18O, F, 32S, 34S, 35CL, 37CL, 6LI, 7LI, 9BE, 10B, 11B, SI, P, 79BR, 81BR, NA, SP, MG, AL, CL, K, CA, ZN, BR, RB, I, CS, B, H, C, N, O, S, HE, NE, AR, KR, 33S, 36S, 11C

An element symbol preceded by an atomic mass specify a particular isotope (e.g. ¹H) whereas an element symbol without a mass specifies the chemical atomic weight (e.g. H = 1.00797). The special symbol 'SP' specifies a mass of 999.

5. Data separation line, line 5 + N: (FORMAT A4)

Consists of an asterisk '*' (left-justified in the A4 field) to separate the coordinate input from the Hessian. (This character is a special value for 'IA' in item 4.)

6. *Either* (if IFFREQ = 0):

Hessian input lines, from line 6 + N to line 6 + N + m: (FORMAT: 3E20.10)

(F(I), I=1, NNC)

This is the lower triangle of a square symmetric Hessian matrix, stored as a linear array. The number of unique elements is $NNC = N \times (N + 1) / 2$. The number of three-column lines of data in this section is therefore given by:

$$\text{Number of lines of Hessian, } m = \frac{3N}{2} \times \frac{3N+1}{3}$$

The number of lines of Hessian output from a calculation can be used to verify the reasonableness of the input, by comparing this with the number of atoms using the formula above.

Or (if IFFREQ = NFREQ):

Frequency input lines, line $6 + N$ to line $6 + N + m$: (FORMAT(6F12.4))

(FREQ(I), I=1, NFREQ)

FREQ are a selection vibrational frequencies (as wavenumbers / cm^{-1}) from a prior calculation on this molecular species. NFREQ does not need to be equal to either $3N$ or $3N - 6$ but may have any integer value.

Number of lines of frequencies, $m = \text{ROUNDUP}(\text{NFREQ}/6, 0)$

[The total number of lines of data up to this point is q .]

7. Frequency control line for parent species, line $q + 1$: (FORMAT 20I4)

ITS, NX, (IX(L), L=1, NX)

ITS = 0 if the species is **not** to be treated as a transition structure. Normally this is because the species is an equilibrium structure corresponding to a minimum on the potential energy surface.

ITS = n (a positive integer) specifies the ordinal number of the reaction coordinate frequency (RCF) in the list of all $3N$ vibrational frequencies for a species which **is** to be treated as a transition structure. Usually a transition structure corresponds to a first-order saddle point and possesses a single imaginary frequency which appears first in the ordered list of vibrational frequencies; in this case ITS = 1. Sometimes a species may correspond to a higher-order saddle point, but it may be desired to treat other imaginary frequencies as if they were real; for example, if the RCF appears as third in the list, then ITS = 3, and this frequency alone will be excluded from the partition function but used instead to determine the tunnelling correction.

NX quantifies the number of frequencies to be excluded from the IPFR evaluation.

If the species corresponds to a genuine zero-gradient stationary point, then 6 eigenvalues of the full $3N$ -dimensional Hessian should be zero. In practice, these values may not be exactly equal to zero for a number of reasons, including numerical rounding errors, and it is helpful to identify them explicitly. Usually, NX = 6, and the corresponding degrees of freedom are considered as translations and rotations which are excluded from evaluations of vibrational IPFRs and zero-point energies.

The third and subsequent integers IX (format I4) on this line refer to the ordinal numbers of the frequencies to be excluded from the calculations. If the molecular species is an equilibrium structure with $3N - 6$ real vibrational frequencies, this line would read as:

___0___6___1___2___3___4___5___6

8. Isotopologue 1 title specification, line $q + 2$: (FORMAT 3I3,74A)

IS, IG, IFFREQ, TITLE

This is the same as for item 3, but with appropriate data to describe the first isotopically substituted species.

9. Isotopologue 1 atom specification lines, from line $q + 2$ to line $(N/16)$ where $N/16$ refers to the 16 slots of atoms on one line of the specification, and is therefore equal to the nearest integer,

rounded up. For example, an 18-atom species would be specified by one line of 16 atoms and another of 2 atoms, i.e. 2 lines (FORMAT (16(A4,1X))).

(IAT(I), I=1, NAT)

Each of the n atoms specified in the coordinate input lines (from line 5 to line $5 + n$) is included in the format of 16 slots of format A4, separated by one space between each atom (1X). This is where the isotopic substitution can take place. Therefore, for the following atoms in the coordinate input:

12C, 1H, 1H, 1H

The isotopologue 1 atom specification line for a triply deuterated isotopologue would be:

_12C__2H__2H__2H

where again the underscores are used to represent the formatting.

(if IFFREQ = NF)):

This line is followed by m lines of frequencies for this isotopologue (also applies to subsequent isotopologues). Sample input files for a frequency-based IPFR calculation, as well as traditional Hessian-based calculations are included in the following GitHub repository:

https://github.com/pbw20/SULISO_sample_files

10. Frequency control line for isotopologue 1: (FORMAT 20I4)

As for item 7 for the parent structure, this line specifies whether or not the species is to be treated as a transition structure, and whether any frequencies are to be excluded from the IPFR calculation. The guidelines for item 7 can be followed here.

11. Subsequent isotopologue lines.

A repeat of items 8, 9 and 10 for subsequent isotopologues, therefore a minimum of 3 new lines per additional isotopologue. Remember that the number of isotopologues for IPFRs to be calculated here, including the parent structure, must match that specified in the Header section line 1.

12. Final line: (FORMAT A4)

The final line of the program consists of the word STOP, followed by a line break, in order to inform the program that no more input remains to be read.

Overview of LIPFR output

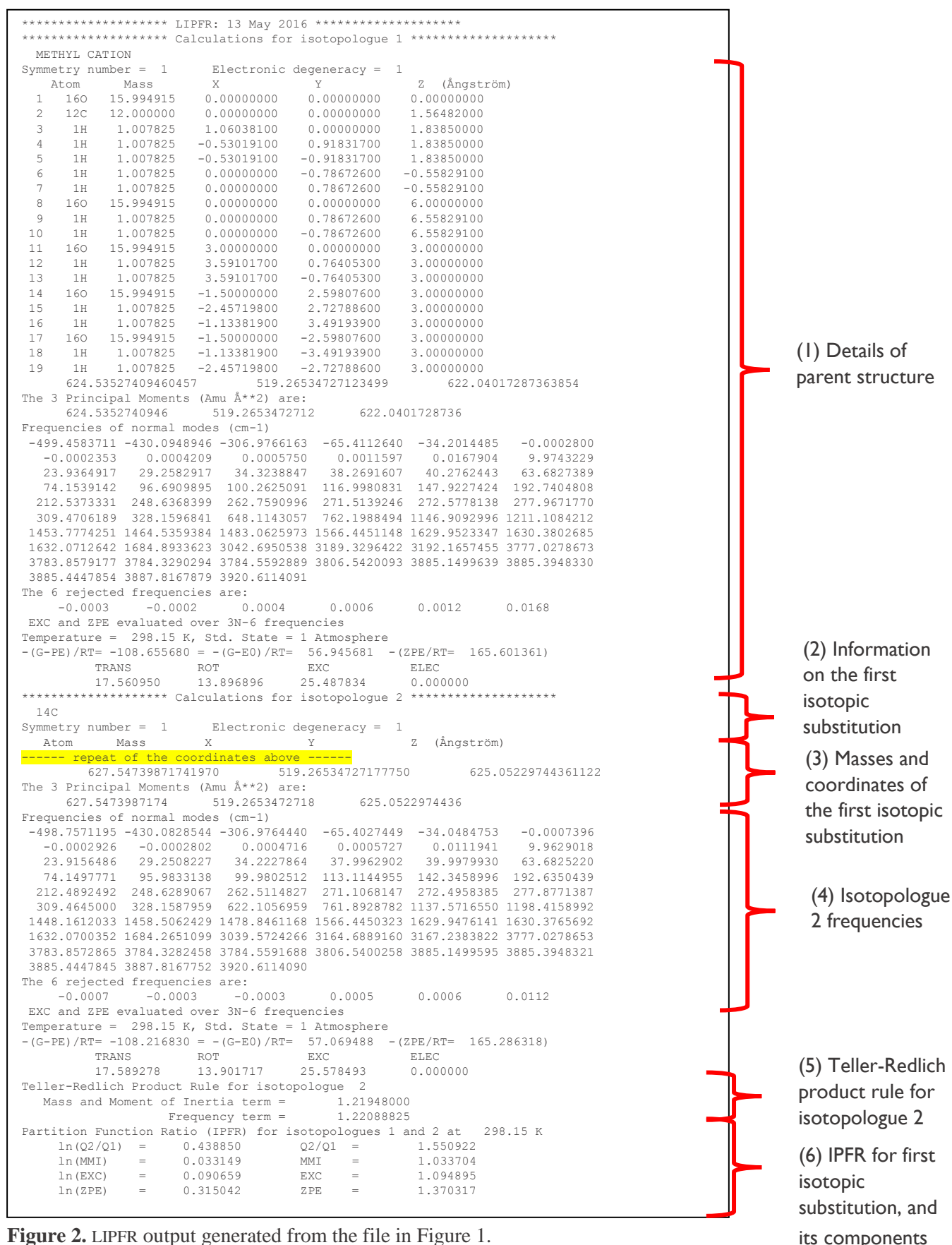


Figure 2. LIPFR output generated from the file in Figure 1.

1. LIPFR output begins by printing the version of the code currently being executed, in this case, the 13th May 2016 release. Then follows the data for the parent structure, in this case a methyl cation structure within a cage of water molecules.
2. Information on the first isotopic substitution (isotopologue 2) is output, including the ‘Title’ line, which was read from the input, as well as the electronic degeneracy and symmetry of the molecule.
3. The masses, atoms, and coordinates are then printed out for isotopologue 2, the coordinates should be the same as for the parent structure (isotopologue 1). In the interest of space, these have been omitted in this output in order to show more of the calculation results.
4. Frequencies calculated from the input un-mass-weighted Hessian and the specified isotopic substitutions are calculated. For heavier isotopologues the frequencies should be lower, and *vice versa*. This is another option to check the reasonableness of the calculations.
5. The product rule is printed out to test the frequency/ mass term equality which should be present in all calculations. If the two terms are equal, then there is little error in the calculated IPFR and input structures, however the holding of the Teller-Redlich product rule is a good measure of the quality of input structures and it is always a good idea to inspect it in order to check that both components are equal to all or the majority of decimal places. For more information on the Teller-Redlich product rule, see:
 E. Teller, cf. Angus, W. R.; Bailey, C. R.; Hale, J. B.; Ingold, C. K.; Leckie, A. H.; Raisin, C. G.; Thompson, J. W.; Wilson, C. L. *J. Chem. Soc.* **1936**, 971-987.
 Redlich, O. Z. *Physik. Chem. B* **1935**, 28, 371-382.
 Williams, I. H., *Journal of Chemical Theory and Computation* **2012**, 8 (2), 542-553.
6. The IPFR calculated as a quotient of partition functions of the heavy isotopologue over the light, is printed out as a component of the different factors: mass (MMI), excitational (EXC), and zero-point (ZPE).

Kinetic isotope effects may be obtained from IPFRs evaluated for light (L) and heavy (H) isotopologues of RS and TS species as follows:

$$\text{KIE} = \frac{k_{\text{light}}}{k_{\text{heavy}}} = \frac{\left(\frac{Q_L^{\text{TS}}}{Q_L^{\text{RS}}}\right)}{\left(\frac{Q_H^{\text{TS}}}{Q_H^{\text{RS}}}\right)} = \left(\frac{Q_H^{\text{RS}}}{Q_L^{\text{RS}}}\right) / \left(\frac{Q_H^{\text{TS}}}{Q_L^{\text{TS}}}\right) = \frac{IPFR^{\text{RS}}}{IPFR^{\text{TS}}}$$

(where Q denotes a molecular partition function). This expression explains the convention of quoting IPFRs as “heavy/light”.

Equilibrium isotope effects may be obtained from IPFRs evaluated for light (L) and heavy (H) isotopologues of reactant and product species, by replacing TS by PS in the equation above.

To obtain a KIE averaged over multiple thermally-accessible conformations of a flexible system, within which the atomic subset considered by UJISO is embedded, it is convenient and appropriate to perform independent averaging of IPFRs over individual RS species for the reactant state and individual TS species for the transition state. This may be performed easily by collecting IPFRs from individual UJISO calculations into a spreadsheet.

$$\langle \text{KIE} \rangle = \langle IPFR^{\text{RS}} \rangle / \langle IPFR^{\text{TS}} \rangle$$

see: Ruiz-Pernía, J. J.; Williams, I. H. *Chem. Eur. J.* **2012**, 18, 9405-9414.

LIPFR Troubleshooting guide

There are a number of errors that users can come across, which will hopefully decrease in number as they familiarise themselves with the code. The most common of these are input errors. Luckily, the program will inform the user within the UNIX environment of the last read format and line number in the fortran source code, and therefore a likely area linking to the error. Throughout this manual, the formats of input have been specified to allow the user easy identification of the areas which require specific types. This should be the first reference point in the case of incomplete termination.

The advantage of the SULISO suite is that when programs terminate successfully, a line similar to that below will terminate the output:

```
***** THE END *****
```

This will ensure that users are certain of normal termination before considering results.

One error that has been noted is the format of certain coordinate/Hessian outputs from electronic structure codes. Some codes simply print out the coordinates, followed directly by the Hessian. However, sometimes the gradient is printed following the coordinates and preceding the Hessian. This consists of n lines of a similar format to the Hessian, where N is the number of atoms in the structure. This is often noticed upon processing, where erroneous frequencies are printed from the program and the input needs to be re-inspected. The simplest solution is to use the equation in Part 6 of the *Overview of LIPFR input* to deduce the expected number of lines of Hessian, and a program such as Notepad++ with a line count to identify if lines of gradient are present. When located, these can simply be deleted from the input and the programs rerun, which should fix the errors.