

SCIENTIFIC WORKFLOW 1.

CALCULATION OF ANHARMONIC X-H STRETCHING FREQUENCIES OF INTRAMOLECULAR VIBRATIONAL CHROMOPHORES RELEVANT TO ATMOSPHERIC CHEMISTRY

Any serious attempt to understand the fundamental physical basis of processes responsible for climate changes must rely on an in-depth understanding of the physics and chemistry of molecular species which constitute Earth's atmosphere. It is therefore self-understood that new milestones in climate science should aim at the molecular-level understanding of the aforementioned phenomena. In other words, climate science has to undergo a substantial paradigmatic shift, from mesoscale-level models to molecular models. In this context, the establishment of chemistry-climate models should be emphasized, aiming to address at a fundamental molecular level the essential issues such as long-term ozone trends as well as methane lifetime etc. In relation to these issues, organic acids have been thoroughly investigated as possible candidates for precursors in the process of formation of aerosols, which play one of the crucial roles in atmospheric, as well as health sciences. Among the organic acids, formic acid – the simplest representative of this class of compounds, is at the same time the most abundant, and also ubiquitous organic acid in the atmosphere.

In the present scientific workflow, through the example of this simplest organic acid we'll illustrate the approach to compute the anharmonic vibrational frequencies of characteristic X-H intramolecular chromophores which are treatable within 1D approximation (*i.e.* as one-dimensional anharmonic oscillators). The approach is straightforwardly extendable to an arbitrary X-Y oscillator.

The reported concentrations of this molecule are of the order of 10 parts per billion, with slightly higher values in urban areas. Besides in aerosols and in gas phase, formic acid has also been identified in atmospheric precipitates (acid rain). Both anthropogenic and biogenic emissions have been pointed out as main sources of this compound in the atmosphere. In a much wider context, formic acid is also important in the human metabolic processes. Also, notable interest for this simple molecular system has been shown in relation to its astrophysical relevance. Formic acid is the simplest organic acid that has been first identified in the interstellar medium. The chemistry and physics of its formation in interstellar

and cometary ices has been addressed as well. Aside from all arguments that have previously been outlined, one should bear in mind that the simplest molecule belonging to the class of organic acids is at the same time a prototypical model system for the whole class of compounds. A number of phenomena related to the physics and chemistry of organic acids may therefore be modeled and understood in a much easier way taking the simplest member as a representative example. Formic acid is also a sort of prototype for systems exhibiting intramolecular rotational (*i.e.* conformational) isomerism, as it can exist in two planar geometries that are mutually interconnected by rotation of the OH group (*i.e.* variation of the HCOH torsional angle – Fig. 1). The two conformations are known as *cis*- and *trans*-isomers, or by the systematic nomenclature, *Z* and *E* forms (Fig. 1).

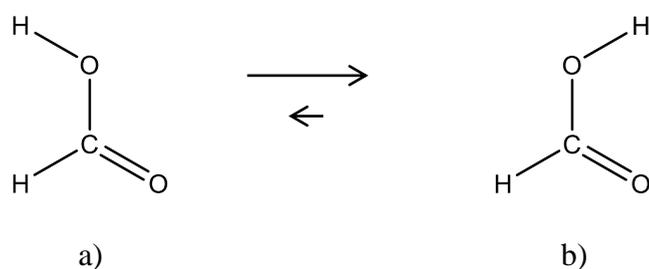


Fig. 1. Schematic presentation of the geometry of *cis*- (*i.e.* *E*) (a) and *trans*- (*i.e.* *Z*) (b) formic acid conformers.

The structural changes induced upon intramolecular torsional motion reflect correspondingly in the values of vibrational frequencies of intramolecular modes. In the context of the present study, we have focused on the values of O-H stretching vibrational frequencies, as this mode is of crucial importance for detection of formic acid, its rotamers, as well as its noncovalently bonded complexes under atmospheric (and even astrochemical) conditions. The O-H stretching mode, in this context, can be treated as a sort of “vibrational chromophore”, sensitive to both intra- and intermolecular environment. This chromophore can therefore be applied as a probe to sense both the different intra- and intermolecular environments. For these purposes, establishing a computational methodology for accurate computation of vibrational frequency of this particular vibrational mode seem crucial if one wants to gain deeper insights into the experimental spectroscopic data employing computational chemical physics as a support.

THE WORKFLOW OF THE APPROACH

STEP 1

Construct the initial Z-matrix, choose a computational methodology and make an input file for QM calculation. Since Gaussian is one of the most widely used (and most widely available) quantum chemical codes, we'll illustrate the principles of the approach through worked examples with this particular code. In Figures 2 and 3 given below, two example input files are presented for geometry optimization and subsequent harmonic vibrational frequency computations for free formic acid starting from initial geometries corresponding to its trans- (Fig. 2) and cis- conformers (Fig. 3).

```
%chk=fa-trans-mp2
%mem=40000MB
%nprocshared=12
# mp2/6-311++g(3df,3pd) opt freq=numer

trans formic acid free, optimization and freq calculations

0 1
C1
H2,1,RH2C1
O3,1,RO3C1,2,AO3C1H2
O4,1,RO4C1,3,AO4C1O3,2,D0,0
H5,4,RH5O4,1,AH5O4C1,2,D,0
  Variables:
RH2C1=1.00
RO3C1=1.50
RO4C1=1.50
AO3C1H2=120.0
AO4C1O3=120.0
D0=180.0
RH5O4=1.00
AH5O4C1=118.0
D=180.0
```

Fig. 2. Example Gaussian input file for simple geometry optimization and subsequent analytical harmonic frequency calculation starting from trans-conformer of the free formic acid molecule.

Since formic acid is a rather small molecular system, of course that also much higher levels of theory can be employed, especially on HPC architectures, but we here focus on the

methodology of computation of anharmonic O-H stretching frequencies, and therefore choose an appropriate level of theory which is sufficient for these purposes.

```
%chk=fa-cis-mp2
%mem=40000MB
%nprocshared=12
# mp2/6-311++g(3df,3pd) opt freq=numer

cis formic acid free, optimization and freq calculations

0 1
C1
H2,1,RH2C1
O3,1,RO3C1,2,AO3C1H2
O4,1,RO4C1,3,AO4C1O3,2,D0,0
H5,4,RH5O4,1,AH5O4C1,2,D,0
  Variables:
RH2C1=1.00
RO3C1=1.50
RO4C1=1.50
AO3C1H2=120.0
AO4C1O3=120.0
D0=180.0
RH5O4=1.00
AH5O4C1=118.0
D=0.0
```

Fig. 2. Example Gaussian input file for simple geometry optimization and subsequent analytical harmonic frequency calculation starting from cis-conformer of the free formic acid molecule.

As we focus on quite another aspect in this workflow, our intention was to keep the structure of the Gaussian input file as simple as possible. For more details concerning the options and Gaussian keywords, the reader is referred to Gaussian documentation (www.gaussian.com). In the current example we adhere to numerical computation of harmonic vibrational frequencies (requested by the `freq=numer` G09 keyword), in order to avoid the necessary reference (HF) wavefunction stability checks that need to be done prior to analytic computation of harmonic vibrational frequencies at MP2 level. This keeps the computation simple, though computationally somewhat more expensive.

STEP 2

Run the Gaussian jobs. The Gaussian jobs are run on a HPC system in a standard way (depending on the particular workflow management system, architecture and user policies; in certain cases, modifications may be required in the input file especially regarding the amount of RAM allocated and the number of processors used through shared memory). Once the stationary points on the molecular particular potential energy surface (PES) are located (the default Gaussian criteria are implied in the former input files), the computation automatically subsequently continues to analytical evaluation of the mass-weighted Hessian, its diagonalization and computation of harmonic vibrational eigenvalues within the normal mode approximation. Aside from vibrational analysis, the computed vibrational frequencies are also an indication of the character of the located stationary point on the molecular PES (minimum or an n -th order saddle point). We hereby seek for true minima on the PES, so that the finally “optimized” structures should be characterized with absence of imaginary frequencies (*i.e.* negative eigenvalues of the Hessian matrices).

STEP 3

Generate input files for computation of the O-H (X-H in general case) stretching potential. Sequentially to the optimization followed by frequency calculation (harmonic vibrational analysis), and the confirmation that one has located a true minimum on the explored PES, we “extract” the Cartesian coordinates of the optimized structures (of the cis- and trans- conformers in the present case), we proceed to the generation of input files to compute the O-H stretching vibrational potentials for the two considered rotamers. We will illustrate this on the example of cis- conformer. First, one needs to identify the vibrating O and H atoms within the HCOOH molecule, which actually move upon excitation of the O-H stretching vibrational mode. This is done rather straightforwardly either by identifying the numeration of atoms within the molecule, or by using a visualization program (such as, e.g. Gaussview). Once the vibrating fragment has been identified, we “cut” the corresponding segment from the geometry optimization + frequency output file, *i.e.* from the part containing Cartesian coordinates of the optimized structure. Out of these data, we create an input file for generation of geometries on which single-point energy calculations will be further performed. Such input file contains two rows, each containing the “initial” (for the next phase), *i.e.*

optimized coordinates of the vibrating O and H atoms (“8” and “1” are the atomic numbers thereof Fig. 3):

```
8 -1.1817589598 0. -0.7725539693
1 -1.898412676 0. -0.1296337467
```

Fig. 3. Example file for the xh-stretch code

This input file (named, e.g. as fa-cis-mp2.txt) is further subjected to processing by our locally developed code “xh-stretch”, which actually reads the input and generates the required configurations. The code is based on the equations of motions for both atoms constituting the vibrating fragment; bond length stretch is carried out by fixing the position of the center-of-mass of the O-H bond. The user can adjust the increment of the change in O-H distance during O-H stretching motion (the “grid density”), as well as the initial distance and the number of points to be generated by the code. Of course, appropriate choice of these parameters needs to be done carefully, doing certain preliminary explorations of the region around the minimum of the O-H stretching potential, approximate location of the first excited vibrational energy level, etc.

Subsequently to generating the O-H stretching configurations, one needs to generate Gaussian input files to perform single-point energy calculations for the series of geometries. Of course, the code “xh-stretch” can in principle be appended to a code that would generate such input file from the geometries corresponding to the minima on certain PES, or geometries extracted from MD snapshots. We have actually developed a series of such code, but these are very “problem-specific” and not useful for a general purpose. They are, however, available from our group upon request. Below is a fragment from such multi-step Gaussian job file (Fig. 4):

```
%mem=40000MB
%nprocshared=12
#mp2/6-311++g(3df,3pd) scf=tight density=current

cis formic acid free, optimization and freq calculations

0 1
```

```
6 -0.0033939966 0. -0.1142886418
1 -0.111683472 0. 0.9793522509
8 1.0481629025 0. -0.6829566288
8 -1.186735 .000000 -.768090
1 -1.819442 .000000 -.200479
```

```
--link1--
```

```
%mem=40000MB
```

```
%nprocshared=12
```

```
# mp2/6-311++g(3df,3pd) scf=tight density=current
```

```
cis formic acid free, optimization and freq calculations
```

```
0 1
```

```
6 -0.0033939966 0. -0.1142886418
1 -0.111683472 0. 0.9793522509
8 1.0481629025 0. -0.6829566288
8 -1.184529 .000000 -.770069
1 -1.854454 .000000 -.169070
```

```
...
```

```
--link1--
```

```
%mem=40000MB
```

```
%nprocshared=12
```

```
# mp2/6-311++g(3df,3pd) scf=tight density=current
```

```
cis formic acid free, optimization and freq calculations
```

```
0 1
```

```
6 -0.0033939966 0. -0.1142886418
1 -0.111683472 0. 0.9793522509
8 1.0481629025 0. -0.6829566288
8 -1.155850 .000000 -.795798
1 -2.309610 .000000 .239257
```

Fig. 4. Example Gaussian input file for computation of the O-H stretching vibrational potential in the case of cis-conformer of the free formic acid molecule.

STEP 4

Extract the computed energies from the Gaussian output files. Once the series of single-point energy calculations has finished, one further needs to extract the computed energies. This can be done either by the awk language, which is a part of almost every

standard LINUX distribution, or, alternatively, by making a template in Mathematica or R package. Of course, one can also write relatively simple home-made purpose-specific codes for this particular task. The series of pair values for the O-H distance and the corresponding energy are used to generate the vibrational potential. Representative potentials for cis- and trans- formic acid monomers in gas phase are shown in figure 1, shown below.

STEP 5

Solve the vibrational Schrödinger equation. The computed vibrational potentials are further fitted to (usually) a fifth-order polynomial in $r(\text{O-H})$, or $\Delta r(\text{O-H}) = r(\text{O-H}) - r_e(\text{O-H})$, where the last quantity denotes the equilibrium $r(\text{O-H})$ value, corresponding to the minimum energy. The fitting polynomial is subsequently cut after fourth order, and a coordinate transformation into Simons-Parr-Finlan (SPF) coordinate type is performed. The SPF coordinates are defined as:

$$\rho = (r_{\text{OH}} - r_{\text{OH},e}) / r_{\text{OH}}$$

Alternatively, one may also work with the “ordinary” bond-stretch coordinates. However, SPF coordinates have been shown to allow for certain advantages over the bond stretch ones. The vibrational Schrödinger equation is solved variationally using often a basis set composed by harmonic oscillator eigenfunctions. If one works with SPF coordinates, usually much lower number of basis functions is required to achieve sufficient convergence. Aside from the variational method, one could also apply a variant of DVR (discrete variable approximation) methods to solve the vibrational Schrödinger equation. When one obtains the vibrational energy levels, the fundamental O-H stretching frequency is computed from the energy difference between the ground ($|0\rangle$) and the first excited ($|1\rangle$) vibrational energy levels.

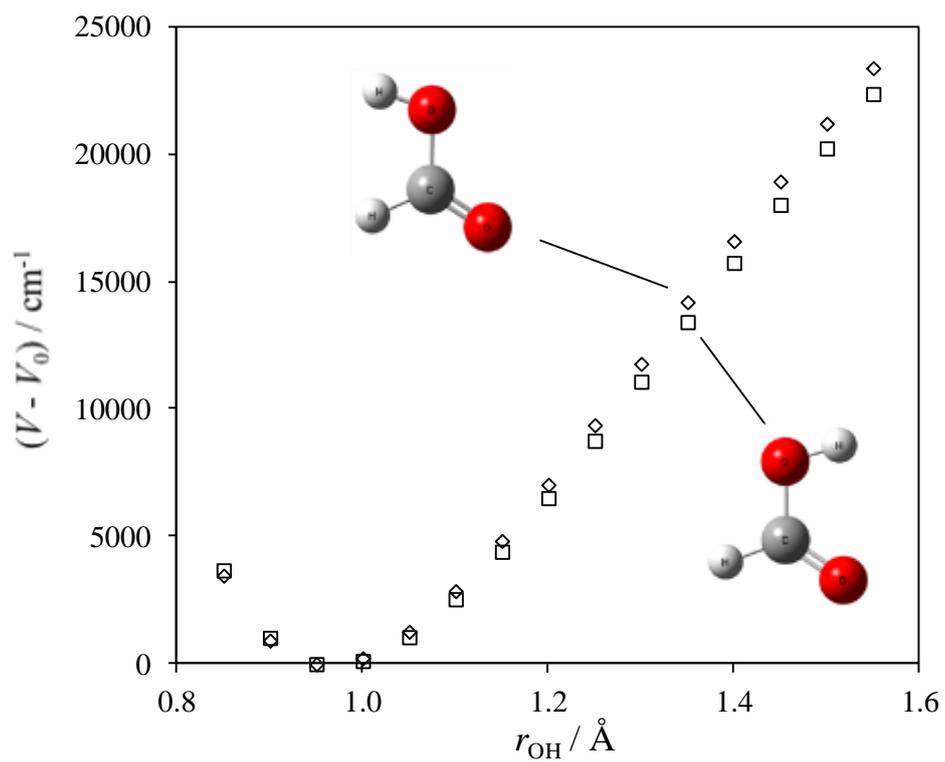


Fig. 1. The O-H stretching potentials of the cis- and trans- rotamers of formic acid computed at MP2/6-311++G(3df,3pd) level of theory.