**Tartaric acid analysis**

**“In vino VERITAS”**

Table of Contents

[Theoretical background 4](#_Toc106117690)

[Classical mechanics 4](#_Toc106117691)

[Newton's law of motion 4](#_Toc106117692)

[Potential energy 4](#_Toc106117693)

[Molecular mechanics and dynamics 4](#_Toc106117694)

[**Quantum mechanics** 4](#_Toc106117695)

[Postulates 4](#_Toc106117696)

[Born–Oppenheimer approximation 4](#_Toc106117697)

[Hartree–Fock method (HF) 5](#_Toc106117698)

[Density functional theorem 5](#_Toc106117699)

[**Atomic orbital basis sets** 6](#_Toc106117700)

[**Methods** 7](#_Toc106117701)

[**Tinker - Simulation in water box** 7](#_Toc106117702)

[Simulation part 7](#_Toc106117703)

[Set Configuration 7](#_Toc106117704)

[Data processing – torsion angles 8](#_Toc106117705)

[Charts – torsion angles 9](#_Toc106117706)

[**Gaussian – Simulation in solvent of water continuum solvent** 10](#_Toc106117707)

[Simulation part 10](#_Toc106117708)

[Set configurations (general input) 10](#_Toc106117709)

[Data processing – vibrational spectra analysis 10](#_Toc106117710)

[Charts – Raman and ROA spectra 11](#_Toc106117711)

[**Amber** 12](#_Toc106117712)

[Optimization and molecular dynamics 12](#_Toc106117713)

[Amber instruction files 12](#_Toc106117714)

[Set configuration and data processing 12](#_Toc106117715)

[Charts – torsion angles, free MD, Amber 14](#_Toc106117716)

[**WHAM** 15](#_Toc106117717)

[Torsion angle WHAM analysis, 2D maps 15](#_Toc106117718)

[**Clusters** 18](#_Toc106117719)

[Option files 18](#_Toc106117720)

[Input files for Gaussian 18](#_Toc106117721)

[Amber MD – set of clusters 18](#_Toc106117722)

[Wham (torsion angle tau) – set of clusters 19](#_Toc106117723)

[**Notes on Analysis** 20](#_Toc106117724)

[**Geometry** 20](#_Toc106117725)

[Molecule chirality 20](#_Toc106117726)

|  |  |  |
| --- | --- | --- |
| Thanks to Petr Bour‘s research group for collaboration and professional assessment |  | Filip Zikeš |
|  | David Kocúrek |
|  | Petr Bouř |
|  | Jiří Kessler |
|  | Josef Kapitán |

# Theoretical background

## Classical mechanics

### Newton's law of motion

### Potential energy

### Molecular mechanics and dynamics

Following formulae describe computations performed by program Tinker.

|  |  |
| --- | --- |
| Minimized potential | , |
| Molecular dynamics (Newton's laws of motion) | , |

## Quantum mechanics

### Postulates

|  |  |
| --- | --- |
| Schrödinger equation defined by wave function |  |

Modification of SE to a time independent form:

### Born–Oppenheimer approximation

The approximation is based on separated functions for electrons and atomic nuclei justified by the great difference between mass of electrons and nucleus. Therefore, the motion of nuclei could be in this situation neglected.

R … atomic nuclei position vector

r … electron position vector

### Hartree–Fock method (HF)

The approximation uses simplified form of electron wave functions which is represented as Slater determinant consisting of one-particle functions and corresponds well with real behavior of particles. This results in set of equations with appearance of certain dependences on other electron functions (spinorbitals) comprising the self-consistent field (SCF) and exchange (correlation) energy which is a quantum effect occurred as physical characteristics for electrons.

#### Slater determinant

Because the electrons are described by half-integer spin and their wave function is antisymmetric, it is necessary to use convenient mathematical apparatus. The simplest mathematical form is therefore the Slater determinant consisting of one-particle wave functions.

#### Equation for HF energy

### Density functional theorem

The density functional theorem is using electrons’ density function, which is easier to represent by experimental value, instead of classical wave function. Mathematically it is defined as integration of N-particle function through all coordinates, except the one variable.

|  |  |
| --- | --- |
| The elementary condition |  |

#### Hellmann-Feynman theorem

Hellmann-Feynman theorem says that electron density function could be used for calculation of energy gradient even without requirement of wave function.

#### Hohenberg-Kohn theorem

According to Hohenberg-Kohn theorem, it is possible to use electron density function for definition of molecular characteristics instead of wave function, since the density comprises information about atomic nuclei as local extreme points, nuclear potential is stored in Hamiltonian and wave function occurs in Schrödinger equation.

#### Kohn-Sham equations

Kohn-Sham theorem made it possible to use DFT with convenient results. The innovation was done by addition of certain molecular orbitals into the density formula and thus the total energy.

## Atomic orbital basis sets

++diffusion \*\*polarizing

# Methods

## Tinker - Simulation in water box

### Simulation part

The target of the process was to simulate a molecule of tartaric acid in a water environment. Tinker is relatively simple open-source program for elementary chemical calculations with possibility to change its parameters, processes, and behavior in quite easy way.

The starting point of the research was creation of the tartaric acid molecule in program mcm95[[1]](#footnote-1). Using tinker, we minimized the energy state of the molecule (module minimize).

The water environment was represented by water box with periodic boundary conditions, which has been made by multiplication of water molecule in tinker xyzedit utility with size of 18.56 Å. The insertion of the tartaric acid into the water box was done by tinker tool “Soak Current molecule in Box of Solvent” and saved as another xyz file. New file includes the whole system of molecules and can be used for energy minimization and converted into dynamic model.

For the dynamic simulation, we need all files defined for the compound (box xyz coordinates file, box key file and parameter file) and *the key file must be saved in the same directory as the coordinates file*. The utility dynamic requires determination of number of dynamics steps, time step length (femtoseconds), time between dumps (picoseconds) and type of mechanical ensemble.

#### Modification of parameter files

The parametric files were modified to handle bonds of tartaric acid between undefined atom combinations. In the amber forcefield file, the changes were mainly in atom definitions because of their specification for different compounds which was not significant for us. The amoeba forcefield needed deeper structural modifications because some atom combinations did not even exist and had to be derived from other bonding type, mostly their angle values.

Amoeba corrections consisted of defining these missing parameters based on acetic acid and similar compounds containing carboxyl groups and secondary carbons such as iso-propanol. The Amber parameter file modification was easier because of multiple occurrences of general parametric values for most of the straight chains.

### Set Configuration

#### Minimize calculation for tartaric acid

$tinker/minimize tartaric-acid.xyz 0.01

RMS Gradient per Atom Criterion: 0.01

#### Dynamic calculation for amber, amoeba

nohup $tinker/dynamic tartaric-acid-aq.xyz amber.prm 100000000 1 5 2 298 &

Number of Dynamic Steps: 1 000 000 000 (terminated: 226 500 000)

Time Step Length (Femtoseconds): 1

Time between Saves (Picoseconds): 5

Statistical Mechanical Ensembles: (2) Canonical (NVT)

Desired Temperature (Degrees K): 298

nohup $tinker/dynamic tartaric-acid-aq.xyz amoeba.prm 50000000 1 10 2 298 &

Number of Dynamic Steps:       50 000 000

Time Step Length (Femtoseconds): 1

Time between Saves (Picoseconds): 10

Statistical Mechanical Ensembles: (2) Canonical (NVT)

Desired Temperature (Degrees K): 298

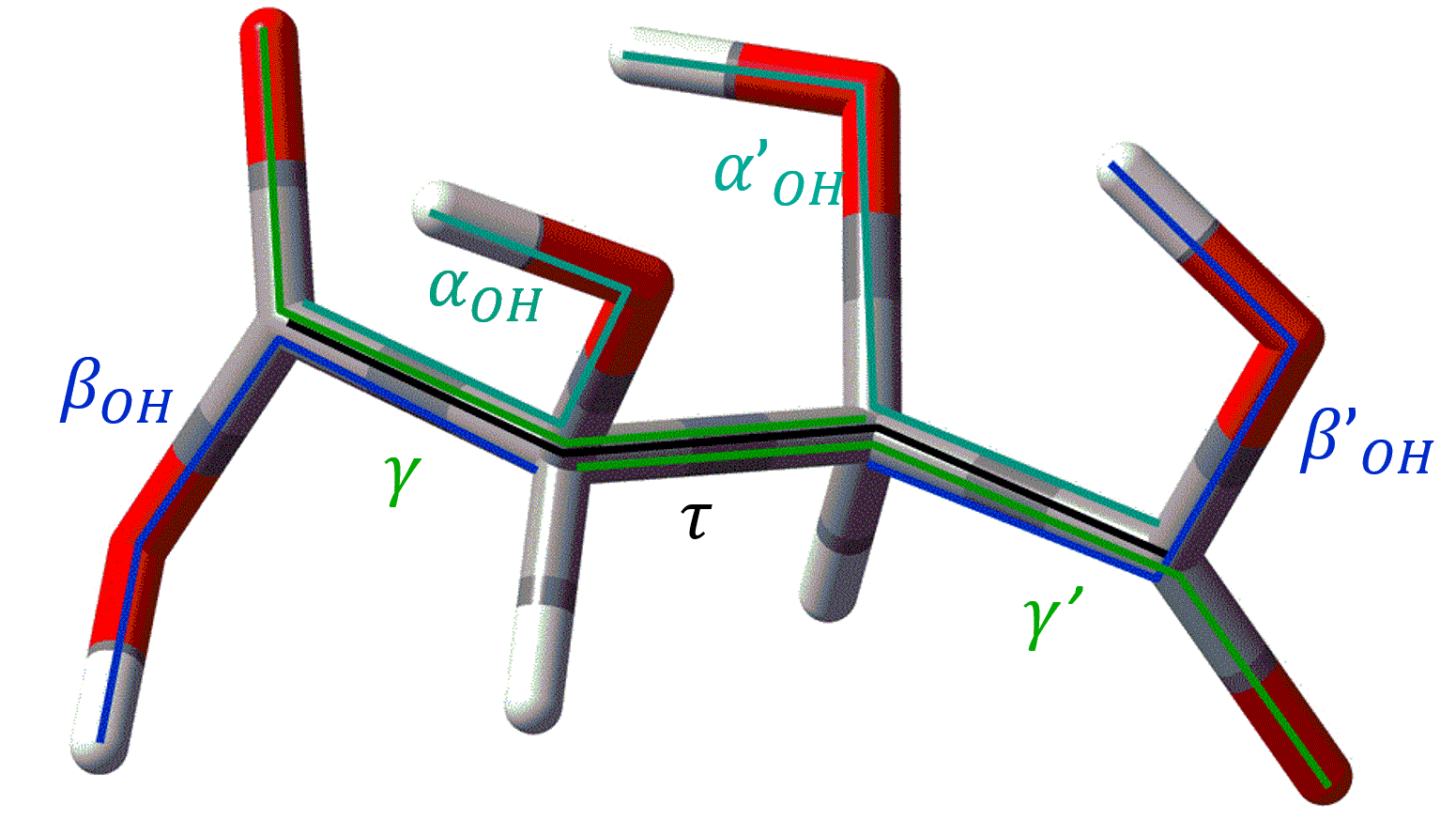
The computations were not completed till the end of operation but only for certain number of steps which was considered as satisfactory and converging. The amber simulation was terminated at 226 500 000 (equivalent of 45 300 files) steps and amoeba at \*\*\*.

### Data processing – torsion angles

Then we used xtorsion*[[2]](#footnote-2)* script for torsion angles analysis for chosen atoms defined in file LIST.PAR. The operation results in set of files including values of the torsion angles in dependence on time during the dynamic simulation. Files called DISTR.\*.PRN summarize the occurrence frequency of individual torsion angles between atoms and LIST.PRN shows their states in all time moments of dynamic simulation.

Obsah obrázku text, bílá tabule

Popis byl vytvořen automatickyThese data could be processed with any software tool (Excel, SigmaPlot) with ability to work with spreadsheets and charts. The resulting charts are used to compare and review the final spectra. Program sc95 allows also easier retrieval of charts data and its numeric processing for more spectra.



### Charts – torsion angles

Charts of torsion angles for both amber  
and amoeba force field are shown below.



|  |  |
| --- | --- |
| Amber force field | Amoeba force field |

## Gaussian – Simulation in solvent of water continuum solvent

### Simulation part

The Gaussian is more advanced software for simulation of molecular behavior.

The basics of process represent the input file which configures the initial state of molecule, available computational resources with method of simulation based on theoretical background and other specifying parameters. For our purpose we used ***B3LYP*** method derived from density function theorem and orbital basis including polarizing functions of higher energy orbitals and diffusion functions for hydrogens and heavier elements.

### Set configurations (general input)

%chk=tartaric.chk

%mem=1GB

%nproc=2

#B3LYP/6-311++G\*\* opt scrf=(CPCM, solvent=water)

 Geom=(NoDistance, NoAngle)

... (name; molecule charge, multiplicity; coordinates data)

--link1--

...

#B3LYP/6-311++G\*\* scrf=(CPCM, solvent=water) geom=checkpoint guess=checkpoint

freq=roa

### Data processing – vibrational spectra analysis

The vibrational spectra could be recalculated from Gaussian output using scripts for extracting the optimized geometry with force field.

First of the scripts is gar9, which extracts the geometry and force field into separate files. New1 program starts the calculation. By new2 we can define atomic masses. New4 calculates harmonic vibrational frequencies and the cartesian – vibrational normal mode transformation matrix and saves them to F.INP. New5 reads derivatives of electric and magnetic dipoles (FILE.TEN) and vibrational frequencies with the S-matrix and calculates absorption and VCD intensities which are written to DOG.TAB. New6 reads derivatives of molecular polarizabilities (FILE.TTT) and vibrational frequencies with the S-matrix and calculates Raman and ROA intensities which are written to ROA.TAB. Tabprnf takes line intensities and makes smooth spectra from them.

### Charts – Raman and ROA spectra

Following spectra show the weighted arithmetic mean of vibrational spectra for 7 individual conformers of tartaric acid with the lowest level of energy and appropriate experimental spectra.

Weighted average mean RAM

Experimental RAM – D Tartaric Acid, 20 °C

Experimental ROA – D Tartaric Acid, 20 °C

Weighted average mean ROA

## Amber

**Location: s10**:/scratch/zikes/A01\_amber\_histograms/

### Optimization and molecular dynamics

Amber is vastly used package of programs in chemical simulations because of its power and adaptation to calculations processed on GPU. It is also designed for Weighted Histogram Analysis Method (WHAM). WHAM is a method which performs series of dynamic calculations for initial conditions of the substance and thus its energy levels in their local equilibrium state. This is being applied to depict more-dimensional charts of these energies and the most balanced states in accordance with initial position.

Firstly, we optimized tartaric acid molecule in Gaussian and generated structure file TAR.prmtop using scripts gen\_amber\_structure.sh, parmchk2, tleap and packmol. These procedures enable subsequently run dynamic simulation via pmemd.cuda program on GPU.

The result statement could be converted similarly to tinker analysis into torsion angle histograms.

### Amber instruction files

#### Min.in

Instruction file includes parameters for rough minimization and minimization according to Newton’s laws of motion. Also, definition of restraints. For the second minimization, ntr=0.

(File 1: min.in)

#### Eq.in

The file defines general parameters and restraints too. Moreover, there is a specified temperature, time between steps and a number of them. This simulation used 1 000 000 steps separated by 1 fs (free MD: 1 ns). The second equilibration used the parameter ntr=0 too.

(File 2: eq.in)

### Set configuration and data processing

#### Gaussian optimization parameters

Output: optim.out

#B3LYP/6-31G\*\* 5d Iop(6/33=2,6/41=10,6/42=14) opt scrf=(CPCM,solvent=water)

Geom=(NoDistance,NoAngle)

Output: wham\_opt.out

#pop=MK B3LYP/6-31G\*\* 5d Iop(6/33=2,6/41=10,6/42=14) opt scrf=(CPCM,solvent=water)

Geom=(NoDistance,NoAngle)

#### Generating structure file (commands)

gen\_amber\_structure.sh wham\_opt.out structure.prepc

parmchk2 -i structure.prepc -o TAR.frcmod -f prepc

tleap -f prep.in

packmol < box.inp

tleap -f box.in

#### Dynamic simulation (commands)

**General command:**

**pmemd.cuda** -O -i **instruction** -o **output** -p **FILE.prmtop** …

**First minimization:**

… -c TAR.crd -r TAR-min.crd *-ref TAR.crd*

**First equilibration:**

… -c TAR-min.crd -r TAR.rst *-x TAR.mdcrd -ref TAR-min.crd*

**Second minimization:**

… -c TAR.rst -r TAR-min.crd

**Second equilibration:**

… -c TAR-min.crd -r TAR.rst *-x TAR.mdcrd*

**Prod. run:**

… -c TAR.old.rst -r TAR.rst *-x TAR.mdcrd*

#### Torsion angles histograms (commands)

cpptraj TAR.prmtop < text.in

ln -s prod-text.mdcrd GEO

ln -s TAR.prmtop TXT

makea (requires: TXT, GEO)

Obsah obrázku text, bílá tabule

Popis byl vytvořen automatickyxtorsion (requires: FILE.X, LIST.PAR)

### Charts – torsion angles, free MD, Amber

|  |  |
| --- | --- |
| ***Amber MD*** | ***Tinker MD:*** **Amber** and **Amoeba** force field |



## WHAM

**Location: s10**:/scratch/zikes/ *(A02 ~ A07)*

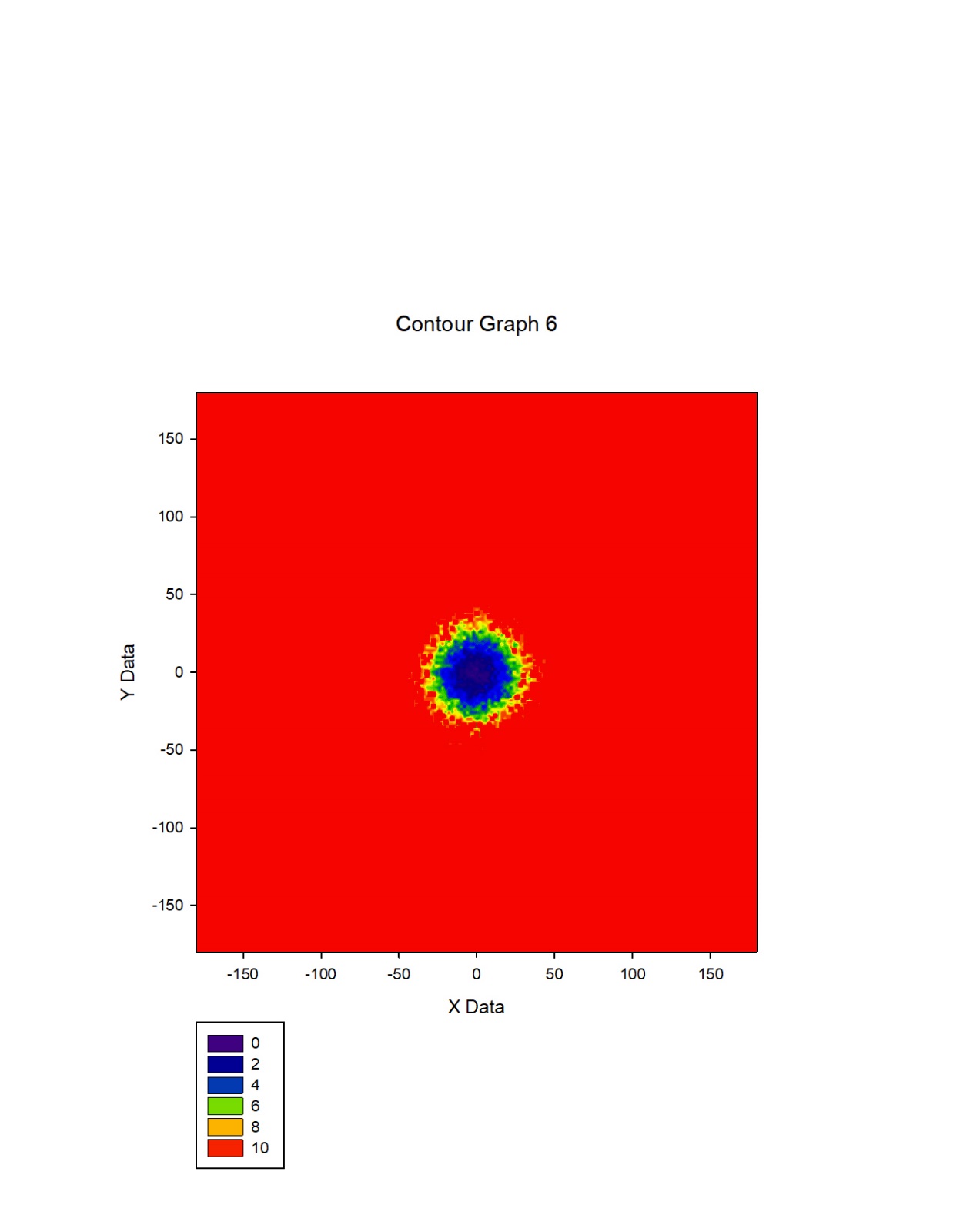
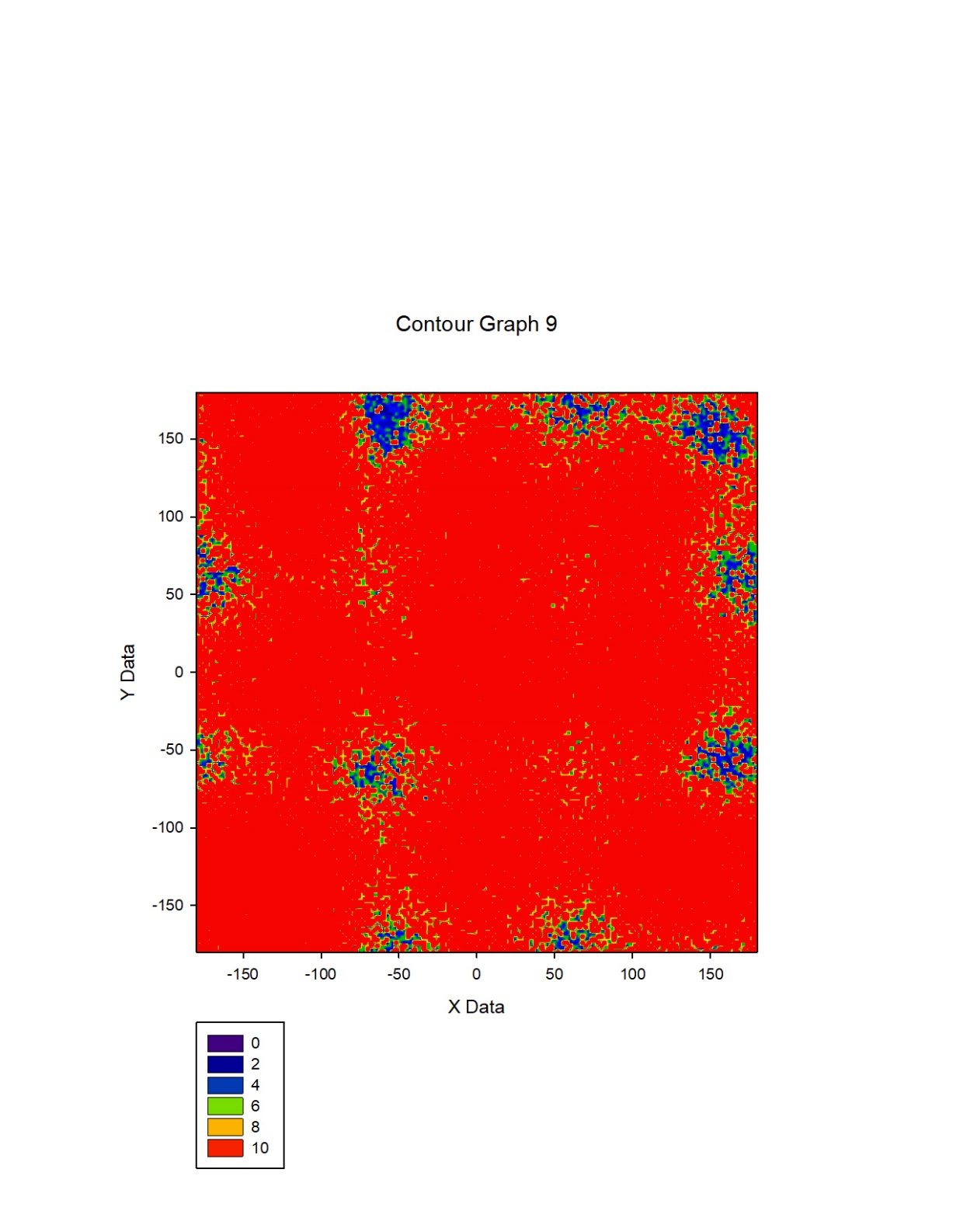
The Wham process was based on Amoeba optimized coordinate files. We ran MD at each angle step for 1 ns where the molecule was held in the position of particular torsion angle. We analyzed the gamma and tau angle by one-dimensional wham and the two-dimensional wham was used for the dependency of pair gamma-tau.

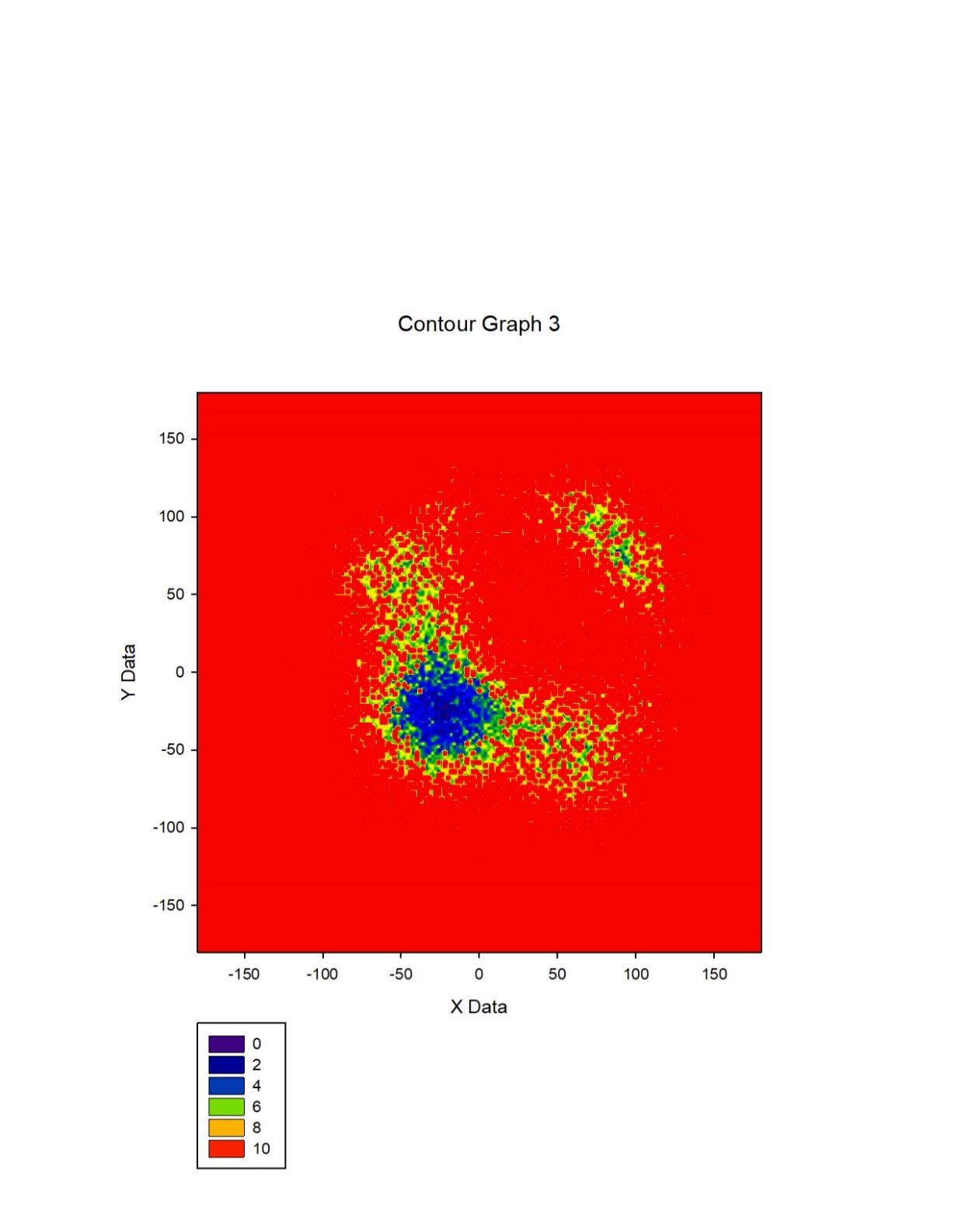
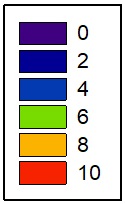
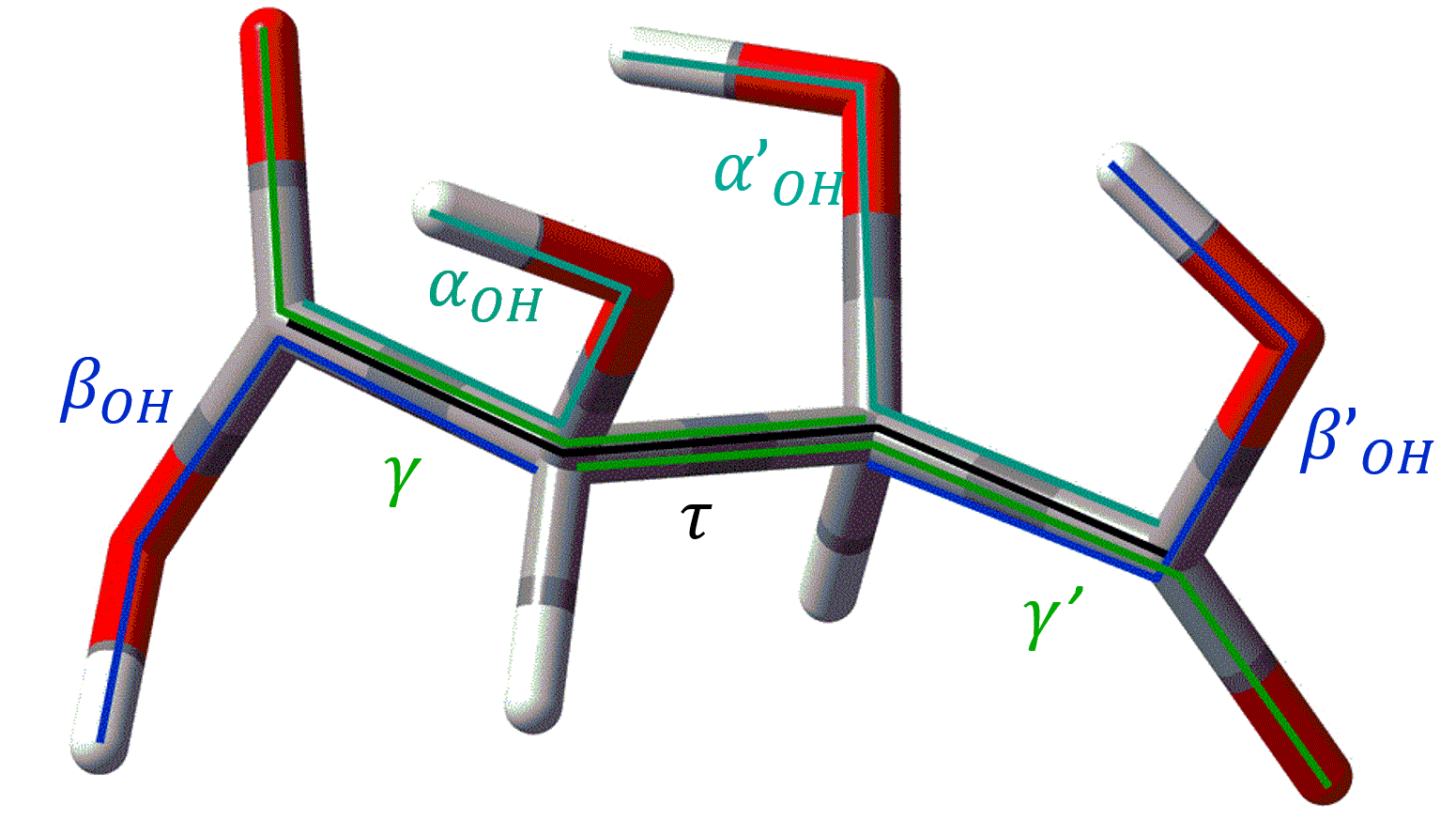
### Torsion angle WHAM analysis, 2D maps

#### Dependency of same torsion angles (1d wham)

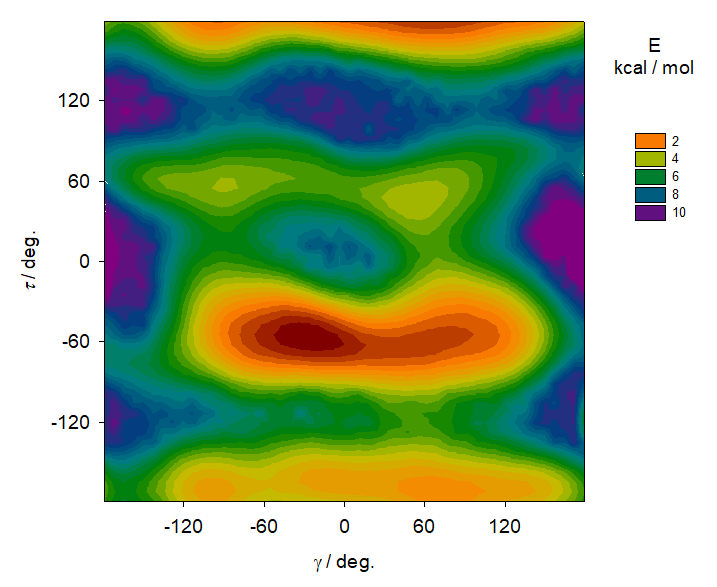
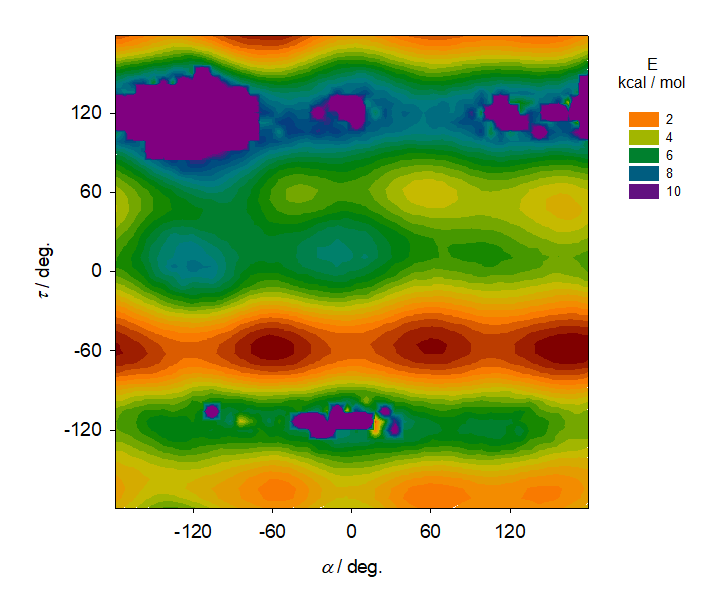
The histograms show us that torsion angle alpha reaches in most cases 3 values, where both of them cannot simultaneously achieve the angle 50° and its values with different sign.

Torsion angle beta is the most stable at angle 0°. The gamma has its own stable angle too, but sometimes, it could be deviated from its basic position.





#### Dependency of two different torsion angles (2d wham)



## Clusters

**Location: electra**:/home/zikes/cluster/ **lenka**:/home/zikes/cluster/

From free MD (10 ns) we made 1000 snapshot geometries (separated by 10 ps). Water further than 3.1 A from the acid was deleted in case of keeping the water environment. The source clusters from Amber MD were used for three individual spectra calculations with different starting methods, bases and bordered space of surrounding water molecules. The next three clusters originate from wham MD snapshots at and differ only in number of water molecules in the nearest space. All the clusters use corresponding increasing naming from C01 to C06.

### Option files

XSHELL.OPT, Q.OPT, EA.LST, ROA.OPT

### Input files for Gaussian[[3]](#footnote-3)

File: G.TXT (→ freq.inp)

%chk=tartaric.chk

%mem=4GB

%nproc=4

(#method/base) 5d freq=noraman iop(2/11=1) nosymm scrf=(CPCM,solvent=water)

File: Gmode.TXT (→ G98.INP)

(#method/base) 5d force iop(2/11=1) nosymm scrf=(CPCM,solvent=water)

File: Groa.TXT, roa.inp (→ roa.inp)

(#method/base) 5d freq=(roa,vcd) iop(2/11=1) nosymm scrf=(CPCM,solvent=water)

...

532 nm

### Amber MD – set of clusters

#### Cluster C01

**Source:** **Amber MD** (min/eqv/min/eqv/prod)

**Location:** **electra**:/home/zikes/cluster/C01

**Used parameters**

**xshell** 3.1

**G.TXT, Gmode.TXT:** HF/6-31G

**G98.INP:** b3pw91/6-31G\*\*

This cluster uses traditional method of vibrational modes calculation starting at rough optimization using HF/6-31G after removal of water beyond interval of 3. 1 A, the process continues at b3pw91/6-31G\*\* with polarizing functions, calculations of gradient and ends with generating harmonic frequencies in Gaussian resulting in spectra histograms, specifically the creation of Raman spectrum and vibrational circular dichroism.

#### Cluster C02

**Source:** **Amber MD** (min/eqv/min/eqv/prod)

**Location:** **lenka**:/home/zikes/cluster/C02

**Used parameters**

**xshell** 0

**G.TXT, Gmode.TXT, Groa.TXT:** b3pw91/6-311++G\*\*

The second cluster was based on elimination of the water environment and whole optimization process was running at the optimal and more complex method and base.

#### Cluster C03

**Source:** **Amber MD** (min/eqv/min/eqv/prod)

**Location:** **lenka**:/home/zikes/cluster/C02

**Used parameters**

**xshell** 0

**G.TXT, Gmode.TXT, Groa.TXT:** b3pw91/6-31G\*\*

The water molecules were also deleted in the last cluster of the set with simpler base.

### Wham (torsion angle tau) – set of clusters

#### Cluster C04

**Source:** **Wham Amber MD** ()

**Location:** **electra**:/home/zikes/cluster/C04

**Used parameters**

**xshell** 3.1 (*HYD:* **t**)

**G.TXT:** b3pw91/6-311++G\*\*

#### Cluster C05

**Source:** **Wham Amber MD** ()

**Location:** **lenka**:/home/zikes/cluster/C05

**Used parameters**

**xshell** 0

**G.TXT:** b3pw91/6-311++G\*\*

#### Cluster C06

**Source:** **Wham Amber MD** ()

**Location:** **-----**:/home/zikes/cluster/C05

**Used parameters**

**xshell** 3.1 (*HYD:* **f**)

**G.TXT:** b3pw91/6-311++G\*\*

### Charts of clusters spectra

The average spectra were taken from the spectra generated by clusters method for each snapshot. The evolution of the average spectra with respect to increasing number of calculated snapshots was drawn into following charts. The ROA spectra is variable in comparison to Raman spectra which acts as an check of correct geometry.

##### Raman evolution chart – cluster C05 (calculated on lenka)

##### 

##### ROA evolution chart – cluster C05 (calculated on lenka)

## Further analysis methods

### Program Kuk

### Program Specomp

Xtorsion used (viewed)

# Notes on Analysis

## Geometry

### Molecule chirality

|  |  |
| --- | --- |
| R | **L** (C01, C02, C03) |
| S | **D** |

## Online Resources

### Petr Bour’s Page

[Link to Homepage](http://hanicka.uochb.cas.cz/~bour/)

[Link to Programs](http://hanicka.uochb.cas.cz/~bour/programs/list.html)

[Link to Scripta](http://hanicka.uochb.cas.cz/~bour/prednaska/prednaska.htm)

### Public GitHub Repository

[filipzikes/IOCB-roa](https://github.com/filipzikes/IOCB-roa)

## Content Comments

### Torsion angles

* A01: long
* A02: tau (f=3)
* A03: tau (f=30)
* A04: gamma (f=3)
* tau: -47.329 (file 5000)
* tau: -169.87 (file 5300)

[Link to charts](#_Charts_–_torsion)

### Analysis scripts

* Suball (sublist) – runs optimization and generates frequency for cluster input in Gaussian
* Group\_sort – sorts the clusters (snapshots) according to its torsion angle into groups
* Folave – calculates average spectrum, including subaverage (SPLIT.AVE)

1. Mcm95 – Molecular graphics program (Pascal, by Petr Bour and Petr Malon). [↑](#footnote-ref-1)
2. Xtorsion – tool calculating internal coordinates for a set of geometries (Fortran, by Petr Bour). [↑](#footnote-ref-2)
3. Generated by script xinp. [↑](#footnote-ref-3)