Supporting Information: Observable-targeting global cluster structure optimization

Johannes M. Dieterich and B. Hartke

1 Methodic details

All the results reported in the manuscript have been obtained using our Scala implementation of the TTM3-F force field by Xantheas *et al.* (J. Phys. Chem. 128 (2008) 074506).

To obtain the dipole-dipole autocorrelation function, a NVE molecular dynamics propagation was carried out for 20,000 steps of 0.1 fs. Snapshots of the dipole moment were taken at every step and used to generate the autocorrelation function without cutting off correlation contributions.

As operators for the global cluster geometry optimization, a mixture of different crossover and mutation strategies was employed. The crossover operator contained contributions from our N-species phenotype operators employing planar and spherical cuts. The mutation operator contained contributions from our graph-based directed mutation operator and a straightforward Monte-Carlo-style mutation.

A full reference input specification can be found below.

2 Comparison of IR-spectra simulation techniques

As outlined in the manuscript, we chose to simulate the IR spectra by means of a NVE molecular dynamics propagation of cluster structures, dipole-dipole autocorrelation and Fourier transformation. Another access to IR spectra is possible through a diagonalization of the mass-weighted Hessian. However, diagonalization of the Hessian will only provide harmonic, uncoupled vibrations.

As can be seen from Fig. 1, our implementation of MD-based spectra is correct, since there is reasonable agreement with the results from Hessian diagonalization. However, the enlargement of the fingerprint region demonstrates that neglect of anharmonicity would lead to significant peak shifts in this area, which cannot be captured by uniform scaling. Therefore, the MD-based approach provides a spectrum that is more directly comparable to experiment.
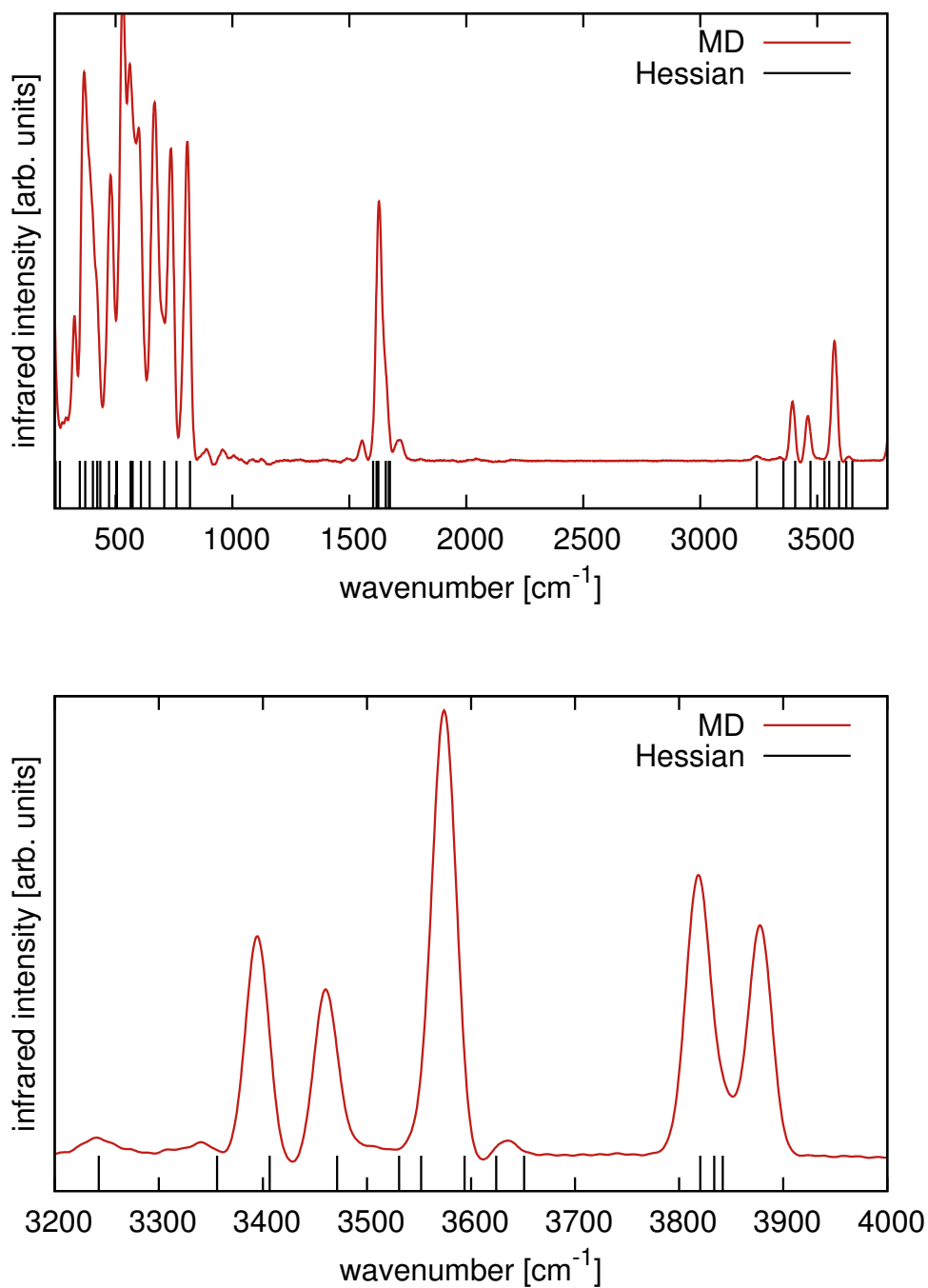
Figure 1: Comparison of IR absorption spectra as obtained from diagonalization of the Hessian (black) and through NVE propagation of the dipole moment (red) for the cage minimum of $(H_2O)_6$ and the TTM3-F potential. Upper panel: absorption range 240 to 3000 cm^{-1}, lower panel: fingerprint region, absorption range 3200 to 4000 cm^{-1}.

Of course, to make the connection to experiment reliable, both ways of calculating vibrational spectra need a potential that correctly models this property. Although TTM3-F was conceived for exactly this purpose, we do not claim that is it the best possible choice. Instead, our investigations are meant as proof-of-concept only.

3 Input data

Based on the current development snapshot of OGOLEM (SVN-ID r1250), we provide a reference input for spectral fitting of a $(H_2O)_6$ cluster against a reference spectrum as defined in the file `ref.spectrum`.

```
###OGOLEM###
<MOLECULARDYNAMICS>
MDEnergyBackend=scattm3f:6,1,false
MDEnsemble=NVE
MDStepLength=0.1
StepsToSnapshot=1
DynamicsSteps=20000
MDPropagator=velocityverlet
MDCoordinatesFile=start.xyz
MDStartTemperature=298.0
VeloInitFromFile=false
MDVelocityFile=start.velo
MDRemoveRotTrans=true
MDRemoveMolInitTransRot=false
MDRemoveRotTrans=true
MDDissocMolMode=2
MDDissocMolFrames=10
MDDissocClusterMode=2
MDDissocClusterFrames=10
</MOLECULARDYNAMICS>
GlobOptAlgo=cluster{
xover(multiple:75%sweden:cutstyle=2|25%snaefellsjoekull:cutstyle=0,
doinflate=true)mutation(multiple:25%montecarlo:mode=some|65%montecarlo:mode=one|
10%advancedgdm:dolocopt,blowbonds=3.08,blowcoll=1.8,gdmprovider=waterspecific
gdmstrategy=euleropt)}
LocOptAlgo=lbfgs:backend=xyz:scattm3f:6,0.5
PoolSize=42
NumberOfGlobIterations=24242
GrowCell=false
CellSize=12;12;12
BlowFacDissoc=3.05
BlowInitialBonds=1.8
BlowBondDetect=1.8
DiversityCheck=percfitnessbased:0.1
```

```
ThreshLocOptGradient=1E-6
PostSanityCD=true
PostSanityDD=true
OptimizationTarget=spectral:corrframes=20000;corrframesperblock=20002;corrgap=20
000;
maxwavenumbers=5000;fullthresh=1e-1;refspecfile=ref.spectrum;intervalstart=3100;
intervalend=3700
<GEOMETRY>
NumberOfParticles=6
<MOLECULE>
MoleculeRepetitions=6
MoleculePath=water.xyz
</MOLECULE>
</GEOMETRY>
```

For a description of these input keywords, we refer to the OGOLEM manual as distributed from `https://www.ogolem.org`. It should be noted that not all of the options are required and some of the choices are system dependent (e.g., `PoolSize=`, `GlobOptAlgo=`, `NumberOfGlobIterations=`).

Additionally, a configuration for the Gaussian-based folder is required:

```
# specify the folder
gaussfolder:1e-10;-1.0
```

as `folder.info`.

Adjusting this for other systems than $(H_2O)_6$ can be accomplished by changing the geometry definition, reference spectrum and, if necessary, choosing a different backend than TTM3-F. As mentioned above, further tuning of input choices may be benefical for optimal performance of the global optimization.

4 Output data

The Cartesian coordinate set for the $(H_2O)_{25}$ structure optimized by our approach and depicted in the manuscript as an inset in Fig. 4 is:

```
75

O            -0.3160008             1.5249116             0.8435426
H            -0.1399346             2.0210169            -0.0064158
H             0.5092785             1.5575762             1.3754891
O             0.5846956             4.5890034             2.2918239
H            -0.3287684             4.3099587             2.4974519
H             1.1408780             3.8128410             2.5046905
O            -0.0276937            -2.7745718            -2.3650767
H             0.0929129            -3.4864959            -1.7000241
H             0.6847788            -2.1177986            -2.1847078
```

```
O        -1.4743813        -0.8341435         4.7736205
H        -2.4499600        -0.7506771         4.8072971
H        -1.3354570        -1.5278487         4.0964418
O         3.3648388         1.6933332         0.2631477
H         3.3783504         0.7192437         0.4706738
H         2.8116114         2.0891992         0.9677499
O         1.0840815         5.3800930        -0.3217976
H         0.7765973         6.2783220        -0.4584160
H         0.9203763         5.1861748         0.6346661
O         0.2867227         2.9350319        -1.4083934
H         0.5595948         3.8459666        -1.1454549
H         1.0261962         2.5660303        -1.9287495
O        -0.1459516         1.5370720         4.6657338
H         0.0592123         1.7834228         5.5750586
H        -0.5765478         0.6357589         4.7122769
O         0.7055124        -1.8328644         0.6321496
H         0.4993437        -2.7822616         0.5188092
H        -0.1171452        -1.3635571         0.4148399
O        -2.0389094        -0.5458979         0.6077308
H        -1.4145518         0.2174926         0.7331459
H        -2.0244268        -1.1025461         1.4228482
O        -3.8975539         1.0738684         2.2256743
H        -3.4701806         0.6402547         1.4634248
H        -3.3235775         1.8375870         2.4541086
O         2.2297587        -1.0537927        -2.8238401
H         2.3115591        -0.0754113        -2.7427959
H         1.8834748        -1.2465732        -3.7254326
O         3.3018320        -0.9885941         0.9738231
H         2.3491080        -1.2664401         0.9362947
H         3.7546065        -1.5315021         0.2822899
O        -2.3058137        -2.4972147         2.6412485
H        -2.5317646        -3.2447413         2.0470480
H        -3.1371486        -2.2022043         3.0741200
O         1.8011368         2.0124515         2.6986373
H         1.1508995         1.8213795         3.4183052
H         2.6277512         1.5206385         2.9623515
O         4.0607714         0.5295516         3.2039038
H         3.8908235        -0.1915570         2.5548944
H         4.8376731         1.0180670         2.8557766
O         4.1075472        -2.3911853        -1.2740743
H         3.4715720        -1.9603455        -1.8960964
H         4.8944143        -2.5775574        -1.7899828
O        -2.7651815        -4.3664318         0.5302106
H        -3.4855840        -4.9924335         0.3988217
H        -2.8446118        -3.6850835        -0.2007042
```

```
O        2.6770386         1.7196648        -2.4074957
H        3.3153239         2.2352508        -2.9120353
H        2.9638911         1.7770946        -1.4559697
O       -0.0208704        -4.5001930        -0.0840290
H        0.4563651        -5.2855774         0.2057054
H       -0.9610743        -4.6163890         0.1947162
O        5.8425883         2.0954687         1.6351510
H        6.0749383         3.0202263         1.7411890
H        5.0885723         2.0820936         1.0091433
O       -2.6546007        -2.4152164        -1.3594780
H       -1.8175594        -2.4803186        -1.8701544
H       -2.5460200        -1.6219871        -0.7847649
O       -4.2353922        -0.9439534         4.0388162
H       -5.1113795        -0.9437977         4.4350745
H       -4.2085879        -0.1680083         3.4124834
O        0.7292492        -2.0607845        -4.9794837
H        0.0117956        -1.6541017        -5.4709871
H        0.3055948        -2.4952051        -4.2057523
O       -1.7621716         2.9231832         2.7819334
H       -1.3312673         2.5189210         3.5666691
H       -1.3109461         2.5107644         2.0080043
```