MD-TRACKS: A Productive Solution for the Advanced Analysis of Molecular Dynamics and Monte Carlo simulations

Toon Verstraelen, Marc Van Houteghem, Veronique Van Speybroeck, and Michel Waroquier*

Center for Molecular Modeling, Ghent University, Proeftuinstraat 86, 9000 Ghent, Belgium

In this paper, we present MD-TRACKS, an advanced statistical analysis toolkit for Molecular Dynamics and Monte Carlo simulations. The program is compatible with different molecular simulation codes, and the analysis results can be loaded into spreadsheet software and plotting tools. The analysis is performed with commands that operate on a binary trajectory database. These commands process not only plain trajectory data but also the output of other MD-TRACKS commands, which enables complex analysis work flows that are easily programmed in shell scripts. The applicability, capabilities, and ease of use of MD-TRACKS are illustrated by means of examples, that is, the construction of vibrational spectra and radial distribution functions from a molecular dynamics run is discussed in the case of tetrahydrofuran. These properties are compared with the experimental data available in the literature. MD-TRACKS is open-source software distributed at http://molmod.ugent.be/code/.

1. INTRODUCTION

Molecular dynamics (MD) and Monte Carlo (MC) simulations are well-established modeling techniques in diverse research fields, ranging from catalysis over biochemistry to solid-state physics.[1−5] Historically, most MD/MC simulation software relied on molecular mechanics models to cope with the cost of computing the potential energy and the interatomic forces.[6] However, during the past decade, the development of specialized algorithms and continuously progressing computer technology have made ab initio molecular dynamics[7−11] a viable and attractive alternative to the conventional molecular mechanics methods. Despite the ever-increasing computing power, molecular mechanics remains relevant for its scalability toward large systems, longer time scales, and the applications in hybrid QM/MM methods.[12] Today, a myriad of MD/MC simulation software is available (CPMD,[7,13] CP2K,[8,14] LAMMPS,[15,16] DL_POLY,[17] CHARMM,[18] NAMD,[19] GULP,[20] GROMACS,[21] GROMOS,[22] CERIUS2,[23]...) that enables one to generate a vast amount of trajectory data by integrating the equations of motion of the system under study. The trajectory data does not only contain the time-dependent atomic coordinates, but also many other quantities as a function of time, for example, atomic velocities, forces, cell parameters, pressure, temperature, partial charges, dipole moments, polarizability, different types of energies, orbitals, and so on. To translate the raw trajectory data into relevant scientific results, a proper statistical analysis is indispensable.

Several trajectory analysis solutions are readily available, for example, Visual Molecular Dynamics[24] (VMD), GROMACS,[21] PTRAJ,[25] SIRIUS,[26] and SMILYS.[27] They are certainly of high quality, and some of them offer impressive visualization functions for molecular dynamics simulations. Despite the presence of these valuable analysis tools, we felt

the need for a more generic software program that enables complex analysis that goes beyond the standard functions present in the current tools. Ideally, a trajectory analysis program should also be flexible enough to be adapted for many different applications. In this paper, we present MD-TRACKS, a versatile, user-friendly, and freely available toolkit that addresses this challenge. We have tested the existing analysis software and prototypes of MD-TRACKS in foregoing studies[28−30] which has strongly influenced the software design. MD-TRACKS has some distinctive characteristics that make it suitable for a wide range of applications:

(1) MD-TRACKS is compatible with multiple MD/MC simulation codes. Prior to the actual analysis, the trajectory data is converted into a simple, manageable, fast, and cross-platform binary database. The current version of MD-TRACKS (0.003) has an interface to CP2K,[14] CPMD,[13] LAMMPS,[15,16] DL_POLY,[17] and CERIUS2,[23] and our software is easily extended to process trajectory data from other simulation codes.

(2) An analysis task is solved with a series of consecutive MD-TRACKS commands in a solution work flow. Each command loads only the actively used parts of the database in memory to analyze huge amounts of data without memory limitations.

(3) MD-TRACKS commands are orthogonal, that is, the output of each command is written to the database and can be used as input for any other command. MD-TRACKS has a built-in plotting function, but it is also possible to convert the analysis results from the binary database into plain text format that is supported by most spreadsheet applications and plotting tools.

(4) An MD-TRACKS programming library is provided to create custom Python scripts that can access the binary database. Once parts of the database are loaded in memory, efficient numerical operations are possible through NumPy.[31] If one can perform a very specific analysis task only partially

* To whom correspondence should be addressed. E-mail: michel.waroquier@ugent.be.

Table 1. Overview of the Most Important Modules in the MD-TRACKS Programming Library

module	functionality
tracks.core	low-level classes and auxiliary functions that efficiently read from or write to the binary MD-TRACKS database
tracks.convert	routines that convert the output of molecular dynamics and Monte Carlo programs into the binary format
tracks.parse	a set of functions that facilitate the interpretation of command line arguments
tracks.vector tracks.cell	tools for the manipulation of a collection of track files that represent time-dependent three-dimensional vectors or matrices

with the current MD-TRACKS commands, one can easily implement the remainder of the analysis in a specialized script using the MD-TRACKS programming library.

This paper describes both the implementation and the usage of MD-TRACKS. It is assumed that the reader has a basic knowledge of UNIX systems. The following section discusses the database format, the structure of a typical MD-TRACKS command, and an overview of the commands in the current MD-TRACKS version (0.003). In the third section, we give an impression of the capabilities of MD-TRACKS by showing how different types of vibrational spectra and radial distribution functions are easily computed from conventional trajectory data. In section four, we describe how MD-TRACKS can be obtained.

2. IMPLEMENTATION

2.1. MD-TRACKS Database Format. The MD-TRACKS toolkit stores all trajectory data and analysis output in a subdirectory `tracks` of the working directory where the molecular dynamics or Monte Carlo simulation software has written its output files. The commands that start with `tr-from-` convert the simulation output into binary files in the tracks directory. For example, `tr-from-cpmd-traj` converts a CPMD[13] trajectory file into binary data.

Each file in the directory `tracks` corresponds to a single time-dependent scalar, for example, `tracks/atom.pos.0000000.y` contains the values of the *Y* coordinate of the first atom for all iterations of the MD/MC simulation. This approach has the advantage that the database can be managed with simple UNIX commands like `ls`, `cp`, `rm`, and `mv`.

A single file in the database is called *a track*. Each track file consists of a fixed length header, followed by a binary data stream. The header contains a unique fingerprint and a format description of the binary data stream. The information in the header makes it possible to interpret the binary stream correctly on all computer architectures. The binary format has two major advantages, compared to conventional text files: (i) binary data is more efficient, both in terms of disk space and input/output performance, and (ii) binary data support random access, that is, one can read and write in track files at arbitrary positions without any overhead, while text files have to be read or written line by line.

MD-TRACKS follows these database-related conventions that facilitate the analysis process:

• All values in the MD-TRACKS database are stored in atomic units. The output files from MD/MC simulations are converted into atomic units by the `tr-from-*` scripts. Each MD-TRACKS command that produces human readable output, will present numerical data in the units specified by the user. All unit conversions are based on the CODATA basic constants and conversion factors.[32]

• Filenames in the MD-TRACKS database are reserved for a specific purpose. The most important reserved names are given below.

○ `tracks/time`: This track contains the time axis, that is, the time in the simulated molecular system at each iteration step. In most cases, this is the integration time step multiplied by the step counter.

○ `tracks/potential_energy`: This track stores the potential energies felt by the nuclei. It includes all interatomic interactions, that is, it is not limited to the Coulomb repulsion between the positively charged nuclei. In a molecular mechanics simulation, this also includes the valence interactions, Van Der Waals interactions, and so on. In the case of an ab initio simulation, this file contains the sum of the nuclei—nuclei, nuclei—electron and electron—electron interactions.

○ `tracks/kinetic_energy`: This track contains the kinetic energy of the nuclei.

○ `tracks/atom.pos.${index}.${c}`: These tracks hold the atomic positions, where `${index}` identifies the atom by an integer of seven characters and `${c}` is *x*, *y*, or *z*.

○ `tracks/atom.vel.${index}.${c}`: These tracks contain the atom velocities, using the same conventions as in the previous item.

The proposed filename conventions are not strictly imposed but are strongly recommended when working with MD-TRACKS.

2.2. MD-TRACKS Commands. An overview of the MD-TRACKS commands (in version 0.003) is given in the Appendix. To execute a command, one enters its name at the UNIX command line shell, followed by options and arguments. The names of the commands are chosen to interact well with the tab-completion function in most popular UNIX shells. When the command line option `--help` is given as an argument to an MD-TRACKS command, the documentation for that command is printed on screen.

The MD-TRACKS commands are not meant to be entered manually, unless one explores new types of simulations or analysis techniques. In all other cases, it is much more efficient to collect a series of MD-TRACKS commands in a shell script that can be reused for the analysis of many MD/MC trajectories. This script-based approach automates the analysis process, but one still has the possibility to tune all parameters in this automated procedure.

Each command in the MD-TRACKS toolkit is based on the MD-TRACKS programming library which is a Python package that can be reused to write new MD-TRACKS commands for specialized applications. The most relevant modules in the MD-TRACKS programming library are listed in Table 1.
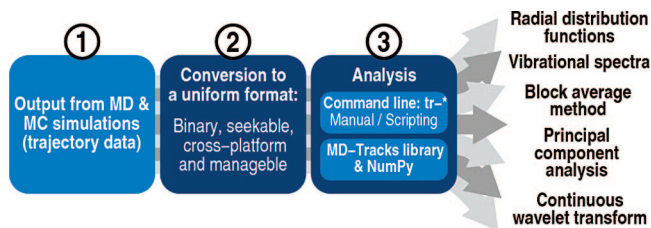
2416 *J. Chem. Inf. Model., Vol. 48, No. 12, 2008*

VERSTRAELEN ET AL.

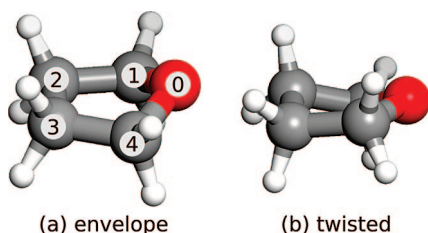**Figure 1.** Workflow of an MD-TRACKS application.

(a) envelope (b) twisted

Figure 2. Two stable conformers of the THF molecule. On the basis of the molecular mechanics model for linear and cyclic ethers of Vorobyov et al.,[36] the twisted conformer is 2.5 kJ mol^{-1} lower in energy than the envelope conformer. The pseudorotation phase (as defined by Cremer and Pople[51]), is $\pm 90°$ for the twisted and $0°$ or $180°$ for the envelope conformer.

3. EXAMPLE APPLICATION: VIBRATIONAL AND STRUCTURAL PROPERTIES OF TETRAHYDROFURAN

In this section, we present a selection of the various possibilities of MD-TRACKS by demonstrating how the vibrational and structural properties of tetrahydrofuran (THF) can be derived from a molecular dynamics simulation. This gives an idea of the usage pattern and the advantages of the MD-TRACKS software design. The typical MD-TRACKS workflow is illustrated in Figure 1. THF was chosen as a case study in view of its important role as structural unit in carbohydrates and biological systems and in the context of a general interest in the conformations and the ring-puckering of small-membered rings.[33,34]

A general consensus about the minimum energy structure of THF has not been achieved yet.[35,33] It is not the intention of this paper to resolve the question of the equilibrium geometry. Therefore, we will use a recently developed molecular mechanics force field for linear and cyclic ethers.[36] All conformational energies, vibrational frequencies and other analysis results mentioned below are obtained with this force field. The potential energy surface of THF in this molecular mechanics model has two local minima, corresponding to the twisted and envelope (or bended) conformer,[37,38] illustrated in Figure 2. The envelope conformer is 2.5 kJ mol^{-1} lower in energy. The vibrational frequencies of both conformers based on the harmonic oscillator approximation are listed in Table 2.

3.1. Vibrational Spectra of THF in the Gas Phase. In this example, we will apply MD-TRACKS to compute the infrared and inelastic neutron scattering (INS) spectrum of THF. In addition, we analyze the parts of the spectrum that are inherent to the pseudorotation motion, which is present in many puckered cyclic organic molecules.[39,40] The concept of pseudorotation was originally proposed by Kilpatrick et al. to explain the exceptionally high entropy of cyclopentane.[41] The IUPAC definition defines a psueodortation as a conformational change resulting in a structure that appears to have been produced by rotation of the entire initial

Table 2. Vibrational Frequencies of the two THF Conformers (in cm^{-1}) Based on the Harmonic Oscillator Approximation Applied to the Molecular Mechanics Model for Linear and Cyclic Ethers by Vorobyov et al[36]

envelope	twisted
262	252
573	558
574	585
756	790
827	846
856	860
936	936
950	950
982	971
1052	1049
1067	1080
1120	1135
1166	1163
1193	1183
1287	1286
1321	1319
1334	1331
1350	1354
1451	1450
1465	1472
1472	1489
1487	1514
1580	1571
1595	1606
2851	2850
2851	2851
2885	2884
2888	2887
2897	2897
2903	2905
2920	2923
2931	2932

molecule and is superimposable on the initial one, unless different positions are distinguished by substitution or isotopic labeling. No angular momentum is generated by this motion; this is the reason for the term. In the case of ring molecules, the conformational changes consist of puckering modes. The pseudorotation of tetrahydrofuran (THF) has been extensively investigated during the past decades, both in experimental and theoretical studies.[33,34,35,37,42,43]

We performed an NVE molecular dynamics simulation of a single THF molecule at an average temperature of 300 Kelvin in the gas phase. The simulation time is 1 ns, and the integration time step is 1 fs. The simulation has been carried out with CP2K.[14] In the following paragraphs, we give step-by-step instructions for the analysis of the molecular dynamics simulation. As the analysis progresses, we unravel the relation between the harmonic frequencies of the two conformers, the vibrational density of states from the molecular dynamics simulation and the experimentally measured infrared spectrum.

The remainder of the text contains a transcript from the command line terminal and can be used by a potential user to reproduce this specific example. The output of the commands is hidden or reduced for reasons of clarity. All commands are preceded by a ">" sign and are printed in bold. The screen output generated by the MD-TRACKS commands is printed using a normal font weight. Long

MD-TRACKS

J. Chem. Inf. Model., Vol. 48, No. 12, 2008 **2417**

commands are split over multiple lines because of the limited column width, but in practice, they should be entered without line breaks.

3.1.1. Setup of the MD-TRACKS Database. The output of the molecular dynamics simulation, as generated by CP2K (development version of June 27, 2008), is distributed over several files. The file `md-1.ener` contains the elementary energy terms as a function of time. The files `md-pos-1.xyz` and `md-vel-1.xyz` contain the atomic coordinates and velocities at each integration time step. The file `md-MM_DIPOLE-1.data` contains the time derivative of the dipole moment. The general output file md.out will not be used in this example. We load the relevant files in the binary MD-TRACKS database with the commands `tr-from-cp2k-ener`, `tr-from-xyz`, and `tr-from-txt`. If an MD-TRACKS database does not yet exist, it is automatically created. The conversion is done as follows. (The CP2K input files are not shown.)

```
> ls
md-1.ener
md-MM_DIPOLE-1.data
md-pos-1.xyz
md-vel-1.xyz
md.out
> tr-from-cp2k-ener md-1.ener
> tr-from-xyz md-pos-1.xyz pos
> tr-from-xyz md-vel-1.xyz vel -u au
> ls
...
tracks/
> ls tracks/
atom.pos.0000000.x
atom.pos.0000000.y
atom.pos.0000000.z
atom.pos.0000001.x
...
atom.pos.0000012.z
atom.vel.0000000.x
...
atom.vel.0000012.z
conserved_quantity
kinetic_energy
potential_energy
time
```

The command `tr-from-xyz` has two mandatory arguments: the XYZ trajectory filename and a suffix (e.g., pos or vel) that is used to generate the filenames in the binary tracks database. By default `tr-from-xyz` assumes that the text based XYZ file contains atom coordinates in angstroms. When the velocity file is converted, the option `-u au` is used to indicate that the XYZ file contains data in atomic units.

In the remainder of this example, we also need the time derivative of the dipole moment generated by the THF molecule. The conversion of the file `md-MM_DIPOLE-1.data` demonstrates how the generic `tr-from-txt` command converts ASCII data to the binary format when a specific `tr-from-*` command is not available. The records of interest in the file `md-MM_DIPOLE-1.data` have the following format (one line):

```
MM DIPOLE [NON-PERIODIC] DERIVATIVE
(A.U.)| 0.000226 -0.000289 -0.00070
```

The three numbers correspond to the *x*-, *y*-, and *z*-components of the time derivative of the dipole moment in atomic units. The file contains such a record for each time step. We can use the ubiquitous UNIX tools `grep` and `cut` to filter out the data of interest:

```
> grep DERIVATIVE md-MM_DIPOLE-1.data |
  cut -c 50-
0.000226 -0.000289 -0.000070
0.000226 -0.000289 -0.000070
0.000220 -0.000262 -0.000066
0.000186 -0.000186 -0.000060
0.000135 -0.000079 -0.000053
...
```

The `grep` command prints only the records from the file `md-MM_DIPOLE-1.data` that contain the word `DERIVATIVE`. The pipe symbol, |, prevents that the output of grep is printed on screen. Instead, the filtered records are redirected as input to the cut command, which discards the first 50 characters from each line. The command `tr-from-txt` reads text data formatted in columns from the standard input and writes this information into the binary database. We can use a second pipe symbol to redirect the output of the cut command to `tr-from-txt`.

```
> grep DERIVATIVE md-MM_DIPOLE-1.data |
  cut -c 50- | tr-from-txt
tracks/dipole.derivt.x
tracks/dipole.derivt.y
tracks/dipole.derivt.z
> ls tracks
...
dipole.derivt.x
dipole.derivt.y
dipole.derivt.z
...
```

So far, all given MD-TRACKS commands were specific for CP2K. From now on, the analysis is completely generic. One can replace the setup of the tracks database by specific commands for another molecular dynamics program and then continue with the instructions below to perform a similar analysis.

3.1.2. Standard Spectral Analysis. The infrared adsorption spectrum can be derived from a molecular dynamics simulation based on linear response theory. The classical approximation of the infrared adsorption spectrum is given by[44,45]

$$\alpha(\nu) \approx \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{j=x,y,z} \left| \int_0^\tau dt \exp(-i\nu t) \frac{d\mu_j}{dt} \right|^2 \quad (1)$$

where ν is the frequency and μ_j are the Cartesian components of the dipole moment. This expression represents the power spectrum of the time derivative of the dipole moment. The command `tr-spectrum` is a generic tool to compute power spectra based on the numerical FFT algorithm.[46] When the time derivative of the dipole moment is given as input, the infrared spectrum is generated. When the atomic velocities are given as input, the inelastic neutron scattering (INS) spectrum or velocity power spectrum is computed. One could also compute the power spectrum of the time derivative of the polarizability, which leads to the Raman spectrum. The velocity power spectrum can be used as a classical approximation[47−49] of the vibrational density of states.[50]

Note that formula 1 is entirely equivalent to the Fourier transform of the autocorrelation function of the time-derivative of the dipole moment. The following MD-TRACKS commands compute the infrared adsorption and the INS spectrum:

```
> tr-spectrum tracks/dipole.derivt.*
  tracks/time tracks/spectrum.ir
  --blocks=250
> ls tracks/
...
spectrum.ir.frequencies
spectrum.ir.wavenumbers
spectrum.ir.amplitudes
...
> tr-spectrum tracks/atom.vel.*
  tracks/time tracks/spectrum.vib
  --blocks=50
> ls tracks/
...
spectrum.vib.frequencies
spectrum.vib.wavenumbers
spectrum.vib.amplitudes
...
```

In the first command line, the three components of the time derivative of the dipole moment are given as input. The time axis is used to create a proper frequency and wavenumber axis. The output is written to files that start with `tracks/spectrum.ir`.

The option `--blocks=250` instructs the `tr-spectrum` command to divide the input data in 250 blocks of the same size. The spectrum is computed for each block and finally the average over all spectra is computed. A higher number of blocks improves the statistical accuracy of the final spectrum, but reduces the resolution on the wavenumber axis. For the infrared spectrum, the resolution on the wavenumber scale (X-axis) is 8 cm^{-1}, and the relative statistical error on the infrared adsorption (Y-axis) is 8%. The INS spectrum is obtained in a similar way. The latter is obtained with a resolution of 1.5 cm^{-1} and a relative statistical error of 9%. The INS spectrum is less sensitive to statistical noise because it is based on more input data: for each atom, there are three Cartesian velocity components, while the dipole moment has only three components in total.

The output files `tracks/spectrum.ir.frequencies` or `tracks/spectrum.ir.wavenumbers` can be used as the X-axis when plotting the spectrum. The corresponding amplitudes of the spectrum are stored in `tracks/spectrum.ir.amplitudes`.

In Figure 3, the INS spectrum is compared with the frequencies obtained within the harmonic oscillator approximation. A strict one-to-one correspondence between harmonic frequencies and peaks in the INS spectrum cannot be made. The main reason is that the molecular dynamics simulation takes into account finite temperature effects, while the harmonic oscillator approximation is only valid close to zero Kelvin. In the remainder of the text, we will show that at room temperature the THF molecule does not oscillate very long in one of the local minima of the potential energy surface. Instead THF continuously alters from the twisted to the envelope state and visits all intermediate structures. It is therefore impossible to assign peaks in the vibrational spectra to one of the conformers. It is clear however that
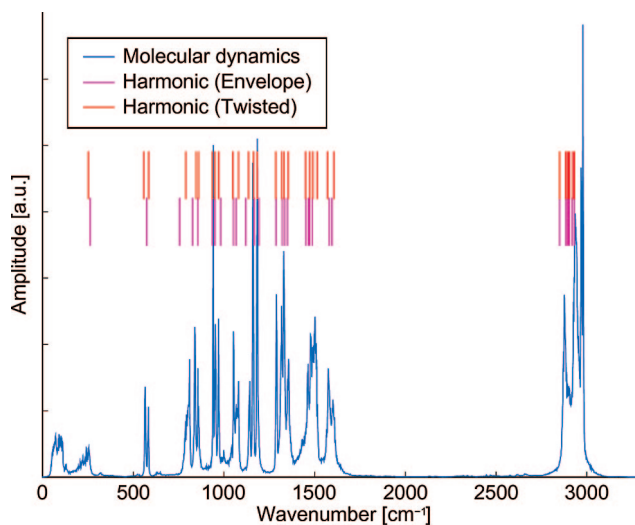

Figure 3. Vibrational frequencies in the THF molecule. The harmonic frequencies of the envelope (purple) and twisted (red) conformer are plotted as vertical lines. The vibrational density of states based on the molecular dynamics simulation is plotted in blue.
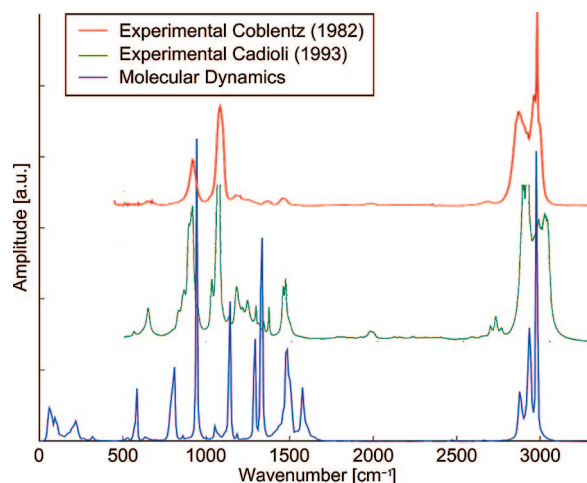

Figure 4. Infrared spectrum. The blue line represents the simulated spectrum. The green[58] and red[34] curves are experimental spectra from the literature.

clusters of harmonic frequencies correlate with clusters of peaks in the spectrum. The correlation fails visibly in two cases:

(1) In the region below 150 cm^{-1}, a band in the INS spectrum appears, which is completely absent in the harmonic oscillator approximation. This band represents a genuine vibration of the THF molecule and can not be attributed to coupling with rotational degrees of freedom. (The angular momentum is set to zero at the beginning of the molecular dynamics simulation and remains negligible.)

(2) The peaks around 3000 cm^{-1} are blue-shifted in the velocity power spectrum when compared to the harmonic frequencies. Figure 4 compares the simulated infrared spectrum with the experimental result. The molecular mechanics model approximates the experimental peak positions, but it fails to predict the infrared activity. In the remainder of this section, we will study the origin of adsorption band below 150 cm^{-1}.

3.1.3. Transitions between the THF Conformers. The distinct conformers of the THF molecule are well characterized by the ring puckering coordinates that can be computed
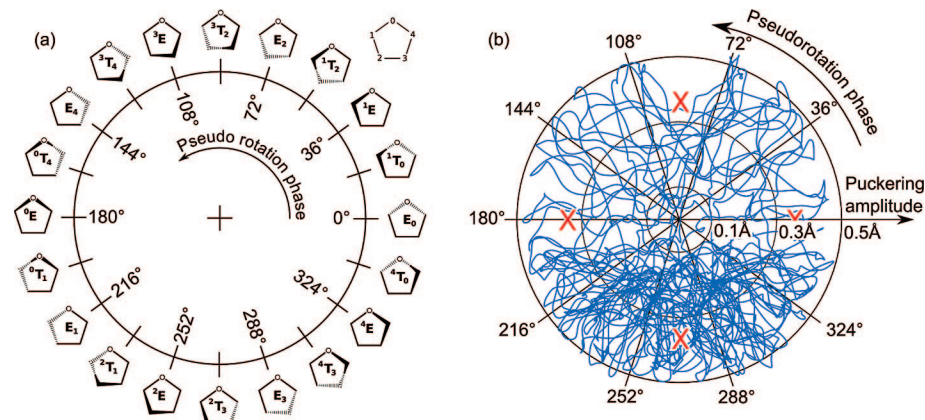
Figure 5. (a) Illustration of the pseudorotation phase, based on Figure 3 in the work of Altona and Sundaralingam.[59] (b) A polar plot of the puckering coordinates during the first 10 ps of the MD simulation. The radius is the puckering amplitude; the angle corresponds to the pseudorotation phase. The coordinates of the stable THF conformers are indicated with crosses.

with `tr-ic-puckering`. Our implementation relies on the general definition of puckering coordinates by Cremer and Pople.[51] In the case of a five-membered ring structure, there are two puckering coordinates: the puckering amplitude, q, which expresses the deviation from the planar ring structure and the pseudorotation phase, ϕ, which discriminates between all possible envelope and twisted geometries (see Figure 5a).

The time evolution of the ring puckering coordinates are computed with the following command:

```
> tr-ic-puckering 5
tracks/atom.pos.0000000
tracks/atom.pos.0000001
tracks/atom.pos.0000002
tracks/atom.pos.0000003
tracks/atom.pos.0000004 tracks/puck.pos
> ls tracks/
...
puck.pos.amplitude.0000002
puck.pos.phase.0000002
...
```

The first argument (5) stands for the number of atoms in the ring structure. The following five arguments are file prefixes that correspond to the five ring atoms (in consecutive order). The last argument is a filename prefix for the output files. The number of ring puckering coordinates is N-3 where N is the number of atoms in the ring. Conventionally, these coordinates are labeled with an integer index that starts from two.[51] For an eight-membered ring, one would obtain the following output files:

```
> ls tracks/
...
puck.pos.amplitude.0000002
puck.pos.amplitude.0000003
puck.pos.amplitude.0000004
puck.pos.phase.0000002
puck.pos.phase.0000003
...
```

During the molecular dynamics simulation, the THF molecule passes through all possible envelope and twisted conformers. This is demonstrated in Figure 5b where the time-dependent puckering coordinates during the first 10 ps are plotted as a solid line. The coordinates of the stable conformers are drawn as crosses. From this picture, it is clear

the ring puckering motion can not be considered as a harmonic oscillation around a minimum on the potential energy surface. This suggests that the vibrational spectra from the MD simulation will exhibit features that are not present in the frequencies from the harmonic oscillator approximation. In other words, the band below 150 cm^{-1} is most likely related with the pseudorotation of the THF molecule. In the next part of this example application, this correspondence is unambiguously demonstrated.

3.1.4. Peak Assignment in Simulated Vibrational Spectra. For a proper understanding of the relation between the pseudorotation phase and the vibrational spectra, we should compute the contribution of the pseudorotation to the spectrum. A straightforward power spectrum of the time derivative of the pseudorotation phase can be misleading. The unit of the amplitude of this spectrum differs from the original INS spectrum, which disturbs a strict comparison. Alternatively, one can project the Cartesian velocity vector on the tangent of the internal coordinate of interest at each time step. The spectrum of these projected velocities only includes contributions of the motion along the selected internal coordinate and has the same unit as the original INS spectrum. The projected velocities are always smaller than or equal to the original velocities. The total spectrum will therefore be an (approximate) upper limit for the spectrum of the original velocities. When a peak in the *projected* spectrum coincides with its counterpart in the total spectrum, one can assume that there is no motion along other (orthogonal) coordinates that contributes to this peak. Such a peak is then completely resolved.

The following command will compute the ring puckering coordinates of the THF molecule, their time derivatives, and the projection of the Cartesian velocity vector on the tangent of each puckering coordinate:

```
> tr-ic-puckering 5
tracks/atom.pos.0000000
tracks/atom.pos.0000001
tracks/atom.pos.0000002
tracks/atom.pos.0000003
tracks/atom.pos.0000004
tracks/atom.vel.0000000
tracks/atom.vel.0000001
tracks/atom.vel.0000002
tracks/atom.vel.0000003
```
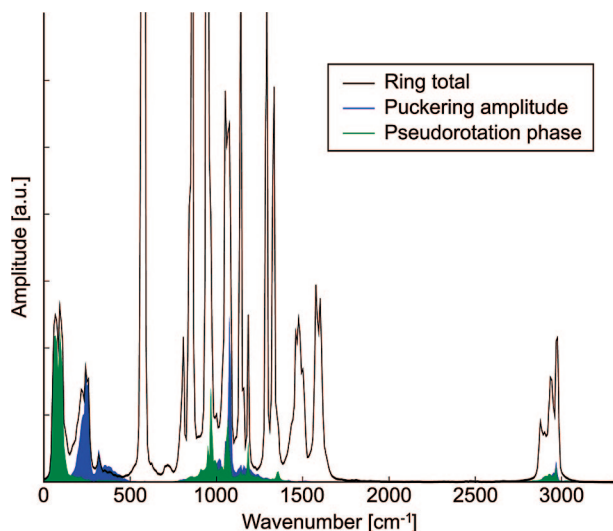
Figure 6. Analysis of the low-frequency bands in the velocity power spectrum of THF. The black line is the total velocity power spectrum of the five ring atoms. The blue curve represents the power spectrum of the velocity vector projected on the tangent of the puckering amplitude coordinate. The green curve is a similar spectrum that corresponds to the pseudorotation phase.

```
tracks/atom.vel.0000004
tracks/puck.pos tracks/puck.vel --project
> ls tracks/
...
atom.vel.0000000.x.proj.puck.vel.amplitude.0000002
atom.vel.0000000.x.proj.puck.vel.phase. 0000002
atom.vel.0000000.y.proj.puck.vel. amplitude.0000002
...
atom.vel.0000004.y.proj.puck.vel.amplitude.0000002
atom.vel.0000004.z.proj.puck.vel.amplitude.0000002
atom.vel.0000004.z.proj.puck.vel.phase.0000002
...
puck.pos.amplitude.0000002
puck.pos.phase.0000002
puck.vel.amplitude.0000002
puck.vel.phase.0000002...
```

Then we compute the power spectra of the projected velocities.

```
>tr-spectrum
tracks/atom.vel.*.proj.puck.vel.amplitude.0000002 tracks/time
tracks/spectrum.puck.amplitude
--blocks=250 > tr-spectrum
tracks/atom.vel.*.proj.puck.vel.phase.0000002
tracks/time tracks/spectrum.puck.phase --blocks=250
> ls tracks/
...spectrum.puck.amplitude.amplitudes
spectrum.puck.amplitude.wavenumbers
spectrum.puck.phase.amplitudes
spectrum.puck.phase.wavenumbers
...
```

Figure 6 gives an overview of the results. The black curve represents the total velocity power spectrum of the five ring atoms. The blue curve is the spectrum of the velocity vector projected on the tangent of the puckering amplitude coordinate, and the green curve corresponds to the pseudorotation phase. The contribution of the hydrogen atoms is not included in this figure because the ring puckering coordinates are not influenced by the hydrogen

positions. The plot clearly reveals the origin of the lowest frequency band (from 20 to 150 cm^{-1}). It is entirely the result of the pseudorotation motion. The second-lowest band (from 150 to 280 cm^{-1}) is mainly caused by the puckering amplitude vibration. Apparently the projected velocities also correlate with bond stretch and bending angle modes at higher frequencies, mainly, because the tangents of the bond stretch, bending angle, and puckering coordinates are not orthogonal.

3.1.5. Final Remarks. This example stresses an important technical advantage of MD-TRACKS. The orthogonal design of the MD-TRACKS commands does not impose a pre-defined work flow during the analysis. In this case, the output of the command `tr-ic-puckering` is used as input for the command `tr-spectrum`. The puckering coordinates could have been used as an input for other commands, such as `tr-hist` or `tr-blav`. As demonstrated above, `tr-spectrum` also processes the output of many commands, for example, `tr-from-xyz`, `tr-from-txt`, `tr-ic-*`, etc. More advanced analysis tasks are carried out by a whole series of MD-TRACKS commands, grouped in a shell script. Each line in such a script processes the results from the previous commands. The final analysis results are converted to ASCII format with `tr-to-txt`, or they can be directly plotted with `tr-plot`.

3.2. Structural Properties of Liquid THF. In this section, we study radial distribution functions (RDF's) or pair distribution functions[52] of THF in the liquid phase. We consider both the center-of-mass RDF and the atom−atom RDF, which can be compared to neutron diffraction experiments.[53] The comparison between an experimental and a simulated RDF is a stringent test for the validity of the nonbonding interactions in the molecular dynamics simulation.

All radial distribution functions below are derived from an NVT molecular dynamics simulations of 64 THF molecules in a cubic box of 20.5 Å, using periodic boundary conditions. A Nosé−Hoover thermostat[54] with a relaxation time of 0.1 ps was applied to control the temperature of the system. The integration time step is 1 fs, and the total simulation time is 1 ns. The simulation has been carried out with CP2K.[14] The following subsections are organized in the same style as in the previous example.

3.2.1. Setup of the MD-TRACKS Database. This part is very similar to the previous example. We extract the time dependent atom positions with the following commands:

```
> ls
init.psf
md-1.ener
md-MM_DIPOLE-1.data
md-pos-1.xyz
md-vel-1.xyz
md.out
> tr-from-xyz md-pos-1.xyz pos
--slice=::1000
> ls tracks/
atom.pos.0000000.x
atom.pos.0000000.y
atom.pos.0000000.z
atom.pos.0000001.x
...
atom.pos.0000831.z
```

The option `--slice=::1000` instructs the script `tr-from-xyz` to read a frame from the trajectory file `md-pos-1.xyz` every 1000 time steps. The file `init.psf` is a CHARMM[18] topology file that will be used below to determine which atoms belong to the same molecule or to identify the chemical environment of an atom.

3.2.2. Center-of-Mass Radial Distribution Function. A center-of-mass RDF expresses the probability of finding two THF molecules at a certain distance apart in the liquid, relative to the probability of finding a pair of molecules that are homogeneously distributed at the same density. We will first derive the centers of mass of each molecule:

```
> tr-split-com tracks/atoms.pos pos
init.psf
> ls tracks/
...
com.pos.0000000.x
com.pos.0000000.y
...
com.pos.0000063.z
...
```

The arguments of `tr-split-com` are interpreted as follows: (i) the prefix for the track files that contain the atom coordinates, (ii) a tag for the output files, and (iii) the CHARMM topology file to identify the individual molecules. Consequently, we run the command that computes the radial distribution function based on these centers of mass:

```
> all_mol_prefixes=$(tr-select init.psf mol
  True --prefix=tracks/com.pos)
> echo $all_mol_prefixes
tracks/com.pos.0000000
tracks/com.pos.0000001...
> tr-rdf $all_mol_prefixes 20.5*A, 15*A 100
tracks/com.rdf
> ls
...
tracks/com.rdf.bins
tracks/com.rdf.hist
...
```

The first command `tr-select` lists molecules or atoms based on a filter expression. The first argument is the CHARMM topology file for the system under study. The second argument can be `at` or `mol` to indicate which objects one would like to select (atoms or molecules). The last argument is a filter expression that must evaluate to true for the atoms or molecules of interest. When `True` is literally given as filter expression, all atoms, or in this case molecules, are listed. The option `--prefix=tracks/com.pos` determines the format in which the list is printed. The construction `all_mol_prefixes=$(...)` assigns the output of the command `tr-select` to the variable `all_mol_prefixes`. The command `echo $all_mol_prefixes` prints the contents of this variable on screen. The command `tr-rdf` computes the actual radial distribution function for a list of time-dependent Cartesian coordinates. In this example, the centers of mass are used to compute the radial distribution function. The second argument specifies the box dimension so that the periodicity is properly taken into account when computing pair distances. The third argument is the maximum distance for which the RDF is computed. The fourth argument defines the number of bins
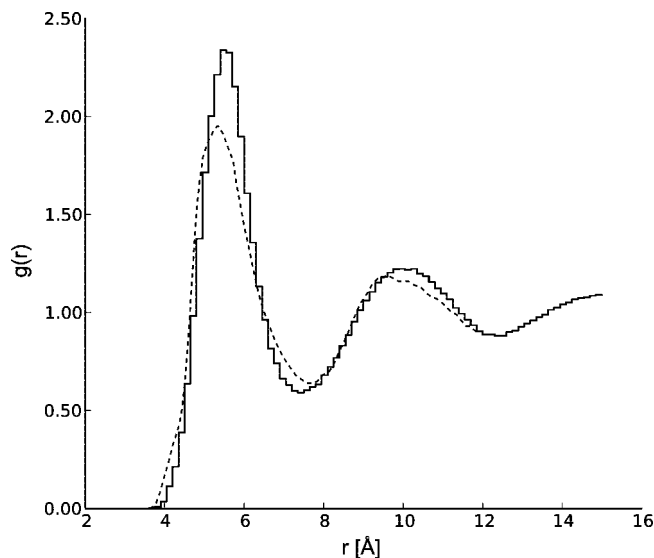

Figure 7. Center-of-mass radial distribution function of liquid THF at room temperature. The solid line represents the histogram derived from the molecular dynamics simulation. The dashed line is the experimentally observed radial distribution function by Bowron et al.[53]

in the histogram and the last argument is a prefix used for the output files. In this example, the file `tracks/com.rdf.hist` contains the *y*-values of the radial distribution function.

The results are depicted in Figure 7, together with the experimental center-of-mass RDF by Bowron et al.[53] The overall correspondence is satisfactory, except for the maximum of the first peak, which is slightly overestimated in the simulated distribution. The area under the first peak of the simulated pair distribution reveals that (on average) there are 12.9 ± 0.5 molecules in the first shell that surrounds a given THF molecule. This compares very well to the experimental value of 12.6 ± 0.3.[53]

3.2.3. Atom—Atom Radial Distribution Functions. A more fine-grained picture of the relative position and orientation of THF molecules in the liquid phase is given by the atom—atom RDF, which expresses the probability of finding an atom of type A and B at a certain distance in the liquid, relative to the probability of finding these atoms at the same distance when they are homogeneously distributed. Analogously to the center-of-mass RDF, we first select the atoms for which we want to compute the RDF. In a second step, the actual RDF's are computed:

```
> O_prefixes=$(tr-select init.psf at
'a.symbol=="O"' --prefix=tracks/atom.pos)
> C1_prefixes=$(tr-select init.psf at
'a.symbol=="C" and a.nsymbols="O,C,H_2"
--prefix=tracks/atom.pos)
> C2_prefixes=$(tr-select init.psf at
'a.symbol=="C" and a.nsymbols="C_2,H_2"
--prefix=tracks/atom.pos)
> echo
tracks/atom.pos.0000001
tracks/atom.pos.0000004...
> tr-rdf $O_prefixes 25*A, 15*A 100
tracks/O.rdf
> tr-rdf $O_prefixes - $C1_prefixes 25*A,
15*A 100 tracks/OC1.rdf
> tr-rdf $O_prefixes - $C2_prefixes 25*A,
```

```
15*A 100 tracks/OC2.rdf
> ls tracks/
...
O.rdf.bins
O.rdf.hist
OC1.rdf.bins
OC1.rdf.hist
OC2.rdf.bins
OC2.rdf.hist
...
```

In the first three lines in the transcript above, three groups of atoms are defined with the command `tr-select`: the oxygen atoms (O), the carbon atoms that are directly bonded to an oxygen atom (group C1), and the carbons atoms that are not directly bonded to an oxygen atom (group C2). For a detailed description of the filter expressions, we refer to the documentation of `tr-select`, which can by consulted with the command `tr-select --help`. Consequently, the RDF's are computed with `tr-rdf`. The first example is based on distances between atoms in a single set, in this example, the set of oxygen atoms. The latter two RDF's consider the distances between the atoms in set A and B but not the distances within each set A or B. In this case, A is the set of oxygen atoms, and B is the set of C1 or C2 atoms.

The three radial distribution functions are plotted in Figure 8. The simulated distributions do not match perfectly with the experimental data. Mainly at short distances, the experimental RDF's show sharp peaks that are not present in their simulated counterparts. The experimental data suggest that our simulations underestimate the liquid structure in terms of relative orientation of neighboring THF molecules.

4. PROGRAM AVAILABILITY

The MD-TRACKS toolkit is distributed as open source software under the conditions of the GNU General Public License, version 3. The software can be downloaded from the Code Web site of the Center for Molecular Modeling: http://molmod.ugent.be/code/. Documentation, installation instructions, and technical support are also available on this web site. In addition, there is a web-interface to the revision control systems that logs all changes in the source code. MD-TRACKS is released together with ZEOBUILDER,[55] which is a highly suitable GUI toolkit for the construction of initial molecular geometries for molecular dynamics simulations.

5. CONCLUSIONS

MD-TRACKS is a powerful and free molecular trajectory analysis toolkit. The MD-TRACKS toolkit consists of many commands that operate on an efficient and cross-platform binary trajectory database. There are three levels at which the MD-TRACKS toolkit can be used: one enters the individual MD-TRACKS commands at a UNIX command line shell, or one collects these commands in specialized shell scripts that automate the analysis work, or one creates new MD-TRACKS commands based on the MD-TRACKS programming library. Currently, the MD-TRACKS program has an interface CP2K,[14] CPMD[13] LAMMPS,[15,16] DL_POLY,[17] and Cerius2.[23]

The analysis of a molecular dynamics simulation of tetrahydrofuran with MD-TRACKS gives nontrivial insights in the vibrational and structural properties of this solvent.
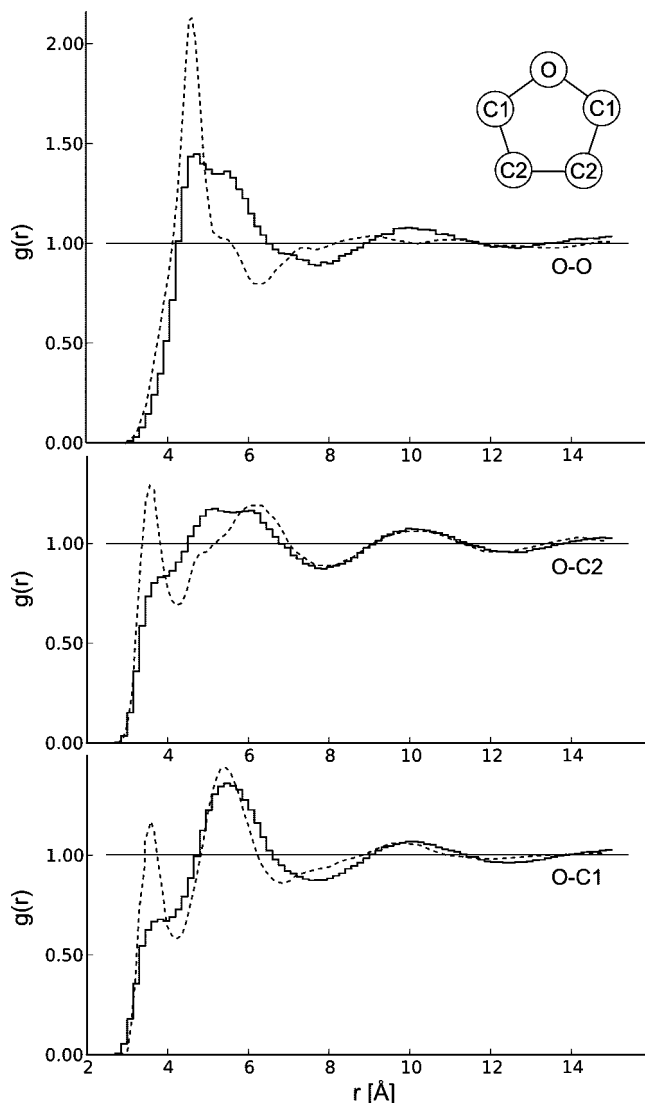

Figure 8. Atom−atom radial distribution functions for the pairs O−O, O−C1, and O−C2. C1 is the set of carbon atoms directly bonded to oxygen, and the C2 set contains to the remaining carbon atoms. The simulated RDF is plotted as a solid line, while the experimental curves are plotted as dashed lines.

The transition of THF between the two (symmetric) twisted conformers is an anharmonic oscillation. It results in a broad band in the vibrational spectra between 20 and 150 cm^{-1}. The molecular mechanics model of Vorobyov et al.[36] leads in general to analysis results that correlate well with experimental observations. It should be clear that these examples merely cover a small part of the functionality of the MD-TRACKS toolkit. Even a full listing of the current MD-TRACKS commands does not reflect the continuous development of new features and improvements.

APPENDIX

The algorithms implemented in the commands below are discussed in standard text books on molecular dynamics and Monte Carlo simulations.[52,56]

- `tr-ac`: Computes the autocorrelation function.
- `tr-ac-error`: Computes the error on an autocorrelation function.
- `tr-angular-momentum`: Determines the angular momenta of one or more molecules.
- `tr-blav`: Applies the block-average method.
- `tr-calc`: Evaluates arbitrary mathematical functions on the data in track files
- `tr-corr`: Computes correlation coefficients
- `tr-cwt`: Computes the continuous wavelet transform.
- `tr-derive`: Numerically differentiates a function.
- `tr-fit-peaks`: Fits peaks in a spectrum.
- `tr-fluct`: Computes fluctuations.
- `tr-from-atrj`: Converts Cerius2 trajectory files (into the binary track files).
- `tr-from-cp2k-cell`: Converts CP2K unit cell data.
- `tr-from-cp2k-ener`: Converts CP2K energy files.
- `tr-from-cp2k-stress`: Converts CP2K stress tensor files.
- `tr-from-cpmd-ener`: Converts CPMD energy files.
- `tr-from-cpmd-traj`: Converts CPMD atom trajectory files.
- `tr-from-dlpoly hist`: Converts DL_POLY history files.
- `tr-from-dlpoly output`: Converts DL_POLY output files.
- `tr-from-lammps-dump`: Converts LAMMPS dump files.
- `tr-from-txt`: Reads data from a column-based plain text file and writes the data to binary track files.
- `tr-from-xyz`: Converts XYZ trajectory files.
- `tr-hist`: Computes histograms.
- `tr-ic-bend`: Computes a (time-dependent) bending angle.
- `tr-ic-dihed`: Computes a dihedral angle.
- `tr-ic-dist`: Computes an interatomic distance.
- `tr-ic-dtl`: Computes a distance to a line.
- `tr-ic-oop`: Computes an out-of-plane distance.
- `tr-ic-psf`: Enumerates and computes (a subset of) the internal coordinates based on the molecular topology.
- `tr-ic-puckering`: Computes the generalized puckering coordinates for an n-membered ring.
- `tr-integrate`: Numerically integrates a function.
- `tr-irfft`: Computes the inverse real Fourier transform.
- `tr-length`: Prints the length of a track file.
- `tr-mean-std`: Computes the (time-dependent) mean and the standard deviation.
- `tr-msd`: Derives the mean square displacement of a set of coordinates as a function of the time interval.
- `tr-msd-fit`: Derives the diffusion coefficient from the data obtained with tr-msd.
- `tr-norm`: Computes the time-dependent norm of a vector.
- `tr-pca`: Applies the principal component analysis method.
- `tr-plot`: Generates charts directly from data in the binary MD-TRACKS database.
- `tr-qh-entropy`: Computes the vibrational entropy, using the quasi-harmonic approximation.
- `tr-rdf`: Computes different types of radial distribution functions:
- `tr-reduce`: Reduces a data set with block averages.

- `tr-rfft`: Computes the forward real Fourier transform.
- `tr-select`: Prints atom or molecule indexes that fulfill a given filter expression.
- `tr-select-rings`: Prints atom indexes that belong to n-membered strong rings.[57]
- `tr-shortest-distance`: Computes the (time-dependent) shortest distance between two sets of atoms.
- `tr-slice`: Reduces a data set with subsampling.
- `tr-spectrum`: Computes various types vibrational spectra.
- `tr-split-com`: Computes the time dependent centers of mass of the molecules in the trajectory.
- `tr-to-txt`: Reads data from the binary database and convert it into plain text format.
- `tr-to-xyz`: Converts the trajectory data in the database to the XYZ format.

REFERENCES AND NOTES

(1) Rapaport, D.D.C. Introduction. In *The Art of Molecular Dynamics Simulationo*, 1st ed.; Cambridge University Press: Cambridge, U.K., 1997; pp 1–10.

(2) Tornroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P. Structural mechanism of plant aquaporin gating. *Nature.* **2006**, *439*, 688–694.

(3) Beckstein, O.; Tai, K.; Sansom, M. Not ions alone: Barriers to ion permeation in nanopores and channels. *J. Am. Chem. Soc.* **2004**, *126*, 14694–14695.

(4) Caleman, C.; van der Spoel, D. Picosecond melting of ice by an infrared laser pulse: A simulation study. *Angew. Chem. Int Ed.* **2008**, *47*, 1417–1420.

(5) Berendsen, H. J. C. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, 1st ed.; Cambridge University Press: Cambridge, U.K., 2007; p 624.

(6) Allen, M. P.; Tildesley, D. J. Introduction. In *Computer Simulation of Liquids*; Oxford University Press: New York, 1989; pp 1–32.

(7) Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55*, 2471.

(8) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(9) McGrath, M. J.; Siepmann, J. I.; Kuo, I. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. Isobaric–isothermal Monte Carlo simulations from first principles: Application to liquid water at ambient conditions. *Chem. Phys. Chem.* **2005**, *6*, 1894–1901.

(10) Kuhne, T. D.; Krack, M.; Mohamed, F. R.; Parrinello, M. Efficient and accurate Car–Parrinello-like approach to Born–Oppenheimer molecular dynamics. *Phys. Rev. Lett.* **2007**, *98*, 066401.

(11) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. Investigating biological systems using first principles Car–Parrinello molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149–156.

(12) Lin, H.; Truhlar, D. QM/MM: What have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(13) Parrinello, M.; Hutter, J.; Marx, D.; Focher, P.; Tuckerman, M.; Andreoni, W.; Curioni, A.; Fois, E.; Roetlisberger, U.; Gianozzi, P.; Deutsch, T.; Alavi, A.; Sebastiani, D.; Laio, A.; Vandevondele, J.; Seitsonen, A.; Billeter, S. *CPMD*, version 3.11.1; IBM Corp, MPI für Festkörperforschung Stuttgart: Stuttgart, Germany, 1990–2006, 1997–2007.

(14) CP2K. http://cp2k.berlios.de (accessed Mar 24, 2008).

(15) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(16) LAMMPS. http://lammps.sandia.gov/cite.html (accessed Jul 9, 2008).

(17) Smith, W.; Forester, T. R. DL_POLY_2.0: A general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph.* **1996**, *14*, 136–141.

(18) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable

molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(20) Gale, J. GULP—A computer program for the symmetry adapted simulation of solids. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 629.

(21) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.

(22) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Gunsteren, W. F. V. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.

(23) *Cerius2 Modelling Environment*; Accelrys Software Inc.: San Diego, CA, 2006.

(24) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(25) Case, D.; Darden, T.; T. E. Cheatham, I. I. I.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Pearlman, D.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D.; Schafmeister, C.; Ross, W.; Kollman, P.; *AMBER 9*; University of California: San Francisco, CA, 2006.

(26) Sirius. http://sirius.sdsc.edu/ (accessed Jul 10, 2008).

(27) Krüger, P.; Szameit, A. SIMLYS version 2.0. *Comput. Phys. Commun.* **1992**, *72*, 265–268.

(28) Pauwels, E.; Verstraelen, T.; Waroquier, M. Effect of temperature on the EPR properties of a rhamnose alkoxy radical: A DFT molecular dynamics study. *Spectrochim. Acta, Part A* **2008**, *69*, 1388–1394.

(29) Pauwels, E.; Verstraelen, T.; De Cooman, H.; Van Speybroeck, V.; Waroquier, M. Temperature study of a glycine radical in the solid state adopting a DFT periodic approach: Vibrational analysis and comparison with EPR experiments. *J. Phys. Chem. B* **2008**, *112*, 7618–7630.

(30) Lesthaeghe, D.; Vansteenkiste, P.; Verstraelen, T.; Ghysels, A.; Kirschhock, C. E. A.; Martens, J. A.; Speybroeck, V. V.; Waroquier, M. MFI fingerprint: How pentasil-induced IR bands shift during zeolite nanogrowth. *J. Phys. Chem. C* **2008**, *112*, 9186–9191.

(31) Oliphant, T. Numpy. http://www.numpy.org/ (accessed Jun 15, 2008).

(32) Mohr, P. J.; Taylor, B. N. CODATA recommended values of the fundamental physical constants: 1998. *Rev. Mod. Phys.* **2000**, *72*, 351.

(33) Legon, A. C. Equilibrium conformations of four- and five-membered cyclic molecules in the gas phase: determination and classification. *Chem. Rev.* **1980**, *80*, 231–262.

(34) Cadioli, B.; Gallinella, E.; Coulombeau, C.; Jobic, H.; Berthier, G. Geometric structure and vibrational spectrum of tetrahydrofuran. *J. Phys. Chem.* **1993**, *97*, 7844–7856.

(35) Duffy, P.; Sordo, J. A.; Wang, F. Valence orbital response to pseudorotation of tetrahydrofuran: A snapshot using dual space analysis. *J. Chem. Phys.* **2008**, *128*, 125102.

(36) Vorobyov, I.; Anisimov, V.; Greene, S.; Venable, R.; Moser, A.; Pastor, R.; Mackerell, A. Additive and classical drude polarizable force fields for linear and cyclic ethers. *J. Chem. Theory Comput.* **2007**, *3*, 1120–1133.

(37) Rayon, V. M.; Sordo, J. A. Pseudorotation motion in tetrahydrofuran: An ab initio study. *J. Chem. Phys.* **2005**, *122*, 204303.

(38) Harris, D. O.; Engerholm, G. G.; Tolman, C. A.; Luntz, A. C.; Keller, R. A.; Kim, H.; Gwinn, W. D. Ring Puckering in five-membered rings. I. General theory. *J. Chem. Phys.* **1969**, *50*, 2438–2445.

(39) Vansteenkiste, P.; VanSpeybroeck, V.; Verniest, G.; DeKimpe, N.; Waroquier, M. Applicability of the hindered rotor scheme to the puckering mode in four-membered rings. *J. Phys. Chem. A* **2006**, *110*, 3838–3844.

(40) Vansteenkiste, P.; VanSpeybroeck, V.; Verniest, G.; DeKimpe, N.; Waroquier, M. Four-membered heterocycles with a carbon–heteroatom exocyclic double bond at the 3-position: puckering potential and thermodynamic properties. *J. Phys. Chem. A* **2007**, *111*, 2797–2803.

(41) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. The thermodynamics and molecular structure of cyclopentane. *J. Am. Chem. Soc.* **1947**, *69*, 2483–2488.

(42) Coulombeau, C.; Jobic, H. Contribution to the vibrational normal modes analysis of d8-THF by neutron inelastic scattering. *THEOCHEM.* **1995**, *330*, 127–130.

(43) Yang, T.; Su, G.; Ning, C.; Deng, J.; Wang, F.; Zhang, S.; Ren, X.; Huang, Y. New diagnostic of the most populated conformer of tetrahydrofuran in the gas phase. *J. Phys. Chem. A* **2007**, *111*, 4927–4933.

(44) Berens, P.; Wilson, K. Molecular dynamics and spectra. I. Diatomic rotation and vibration. *J. Chem. Phys.* **1981**, *74*, 4872–4882.

(45) McQuarrie, D. A. The time-correlation function formalism, I. In *Statistical Mechanics*; Rice, S., Ed.; Harper & Row: New York, 1976; pp 467–542.

(46) Cooley, J. W.; Tukey, J. W. An algorithm for the machine calculation of complex fourier series. *Math. Comput.* **1965**, *19*, 297–301.

(47) Dickey, J. M.; Paskin, A. Computer simulation of the lattice dynamics of solids. *Phys. Rev.* **1969**, *188*, 1407.

(48) Loong, C.; Vashishta, P.; Kalia, R. K.; Jin, W.; Degani, M. H.; Hinks, D. G.; Price, D. L.; Jorgensen, J. D.; Dabrowski, B.; Mitchell, A. W.; Richards, D. R.; Zheng, Y. Phonon density of states and oxygen-isotope effect in Ba1-xKxBiO$_3$. *Phys. Rev. B.* **1992**, *45*, 8052.

(49) Rahman, A.; Mandell, M. J.; McTague, J. P. Molecular dynamics study of an amorphous Lennard-Jones system at low temperature. *J. Chem. Phys.* **1976**, *64*, 1564–1568.

(50) McQuarrie, D. A. Crystals. In *Statistical Mechanics*; Rice, S., Ed.; Harper & Row: New York, 1976; pp 194–221.

(51) Cremer, D.; Pople, J. A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

(52) Allen, M. P.; Tildesley, D. J. Statistical Mechanics. In *Computer Simulation of Liquids*; Oxford University Press: New York, 1989; pp 33–70.

(53) Bowron, D.; Finney, J.; Soper, A. The structure of liquid tetrahydrofuran. *J. Am. Chem. Soc.* **2006**, *128*, 5119–5126.

(54) Nose, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(55) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. ZEO-BUILDER: A GUI toolkit for the construction of complex molecular structures on the nanoscale with building blocks. *J. Chem. Inf. Model.* **2008**, *48*, 1530–1541.

(56) Frenkel, D.; Smit, B. "Monte Carlo Simulations" and "Molecular Dynamics Simulations". In *Understanding Molecular Simulation*, 2nd ed.; Academic Press: San Diego, CA, 2002; pp 23–108.

(57) O'Keeffe, M.; Hyde, B. Nets and infinite polyhedra. In *Crystal Structures I. Patterns and Symmetry*; Mineralogical Society of America: Washington, DC, 1996; pp 289–375.

(58) Smith, A. *The Coblentz Society Desk Book of Infrared Spectra*, 2nd ed.; Carver, C., Ed.; The Coblentz Society: Kirkwood, MO, 1982; pp 1–24.

(59) Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.