JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

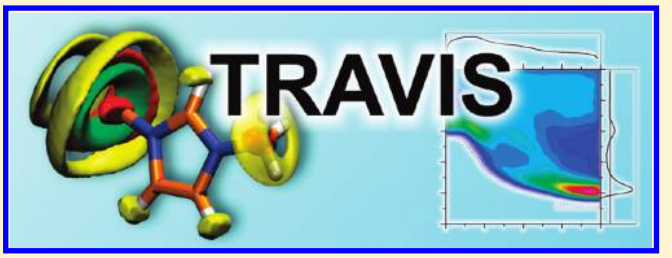
TRAVIS - A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories

Martin Brehm[†] and Barbara Kirchner*[,†]

[†]Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, Linnéstrasse 2, D-04103 Leipzig, Germany

Supporting Information

ABSTRACT: We present TRAVIS ("**TR**ajectory **A**nalyzer and **VIS**ualizer"), a free program package for analyzing and visualizing Monte Carlo and molecular dynamics trajectories. The aim of TRAVIS is to collect as many analyses as possible in one program, creating a powerful tool and making it unnecessary to use many different programs for evaluating simulations. This should greatly rationalize and simplify the workflow of analyzing trajectories. TRAVIS is written in C++, open-source freeware and licensed under the terms of the GNU General Public License (GPLv3). It is easy to install (platform independent, no external libraries) and easy to use. In this article, we present some of the algorithms that are implemented in TRAVIS - many of them widely known for a long time, but some of them also to appear in literature for the first time. All shown analyses only require a standard MD trajectory as input data.



INTRODUCTION

Due to the exponential increase of available computer resources, an increasing interest was dedicated to molecular dynamics (MD) and Monte Carlo (MC) simulations during the last decades, raising them to indispensable standard methods of nowadays computational chemistry. Complex and realistic chemical systems can be simulated, and insight can be gained even into properties that are not accessible by experiment.[1−8]

The direct result of any MC or MD simulation is a trajectory, containing the positions $\vec{r}(\delta t)$ and velocities $\vec{v}(\delta t)$ of the simulated particles at discrete time steps δt. If N particles have been simulated, this trajectory can be regarded as a path through the 6N-dimensional phase space of the system. Because of its high-dimensional nature, this path is not suitable for direct evaluation or interpretation. Algorithms for the reduction of dimensionality have to be applied, converting the input data into two- or three-dimensional data sets, which can be interpreted and visualized much easier.

Many of these dimension reduction algorithms have been known for a long time and are widely used, e.g. mentioning radial pair distribution functions (RDFs),[9−11] which provide insight into the system structure and can be found in almost any MC or MD article. There is an innumerable amount of programs available for calculating these functions. But typically, a certain analysis program contains only some of the desired algorithms, and different programs almost always have different requirements to the input data format. Circumstantial file format conversions are necessary - and of course a program containing the desired analysis has to be found. Due to the lack of programs for performing some analyses, many scientists write small scripts which can be only used on their particular system.

Our aim with developing TRAVIS was to create a unified program package which should contain all common kinds of analyses that can be applied to trajectories - and additionally some new ones that never appeared in literature before (e.g., the powerful combined distribution functions, see below). The availability of this kind of program package should greatly rationalize and simplify the process of evaluating a trajectories. Analyses performed with TRAVIS already appeared in the literature.[12−14]

A good way of introducing trajectory analyses is to apply them to a sample trajectory and show the results. The article is focused on the program and its features. We are not going to interpret the results related to the simulated system. This interpretation can be found elsewhere.[12,14] The aim of the system is just to serve as an example. Also out of the scope of this work is the extensive discussion of all possible applications of the analyses.

The sample system was chosen to be the result of an *ab initio* molecular dynamics simulation[1,8] of the room temperature ionic liquid 1-ethyl-3-methylimidazolium thiocyanate (referred to as $[C_2C_1Im]^+[SCN]^-$) in liquid state (at 350 K), like investigated before.[12,14] The cubic simulation cell contained 32 ion pairs, the cell vector was chosen to be 2027.7 pm, and periodic boundary conditions[15] were applied. The simulation was performed with the CP2k program package,[16] using the Quickstep method[17] and orbital transformation[18] for faster convergence. The B-LYP functional[19,20] of density functional theory (DFT)[21,22] with Grimmes dispersion correction[23] and DZVP-MOLOPT Gaussian basis sets[24] as well as GTH pseudopotentials[25,26] for all atoms were used. The simulation was performed in the NVT ensemble,

Received: May 16, 2011
Published: July 16, 2011

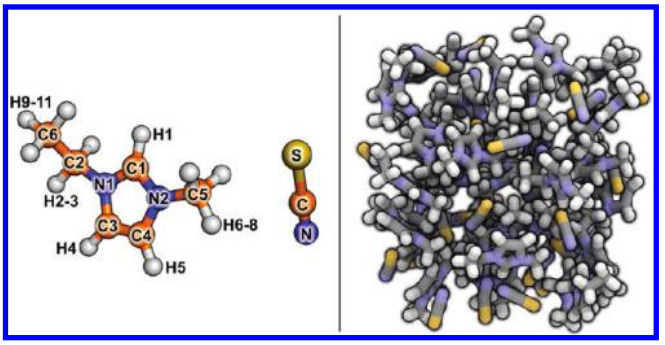


Figure 1. The analyzed system: (a) atom labels of the molecules according to the topological algorithm (left) and (b) snapshot of the simulation box (right).

applying a Nose−Hoover chain thermostat[27−29] for temperature coupling. Approximately 55 ps of physical time were simulated, using a time step of 0.5 fs.

The atom labels used in this article are shown on the left of Figure 1. The numbering scheme has been obtained from the topological algorithm implemented in TRAVIS (which is explained below). In the right part of Figure 1, a snapshot of the simulation box rendered with QuteMol[30] is illustrated.

This article is divided into six sections. After the introduction, some general features of TRAVIS are presented. In the following part, several structural analyses that can be performed are discussed. Section 4 introduces dynamical analyses implemented in TRAVIS. Following to that, some technical details of the implementation and a list of external programs that have been used for creating the figures are given. The last section concludes this article and gives a short outlook.

GENERAL FEATURES

Supported Trajectory Formats. Currently, TRAVIS is able to read trajectories in XYZ, PDB, and MOL2 file format. All three formats are text files and quite simple in their structure. The primary file format TRAVIS relies on is XYZ, and most of the coordinate input and output is performed in this format. If trajectories stored in different formats are to be analyzed, there are powerful tools available for converting almost any geometry format into XYZ (e.g., VMD[31]). TRAVIS only reads the atom coordinates from the trajectory file, disregarding any topological information (see section Molecule Recognition below).

The steps of the trajectory do not need to be successive in time (and a property like time does not even have to exist, like for most MC trajectories). The only requirement TRAVIS puts onto the trajectory is that the total number of atoms and the type of each atom stays constant during the simulation.

Most MC and MD simulations are performed with periodic boundary conditions (PBCs).[15] There exist complex and sophisticated schemes for applying PBCs.[32] TRAVIS supports periodic boundary conditions, but currently only orthorhombic simulation cells (i.e., the cell vectors are orthogonal to each other) can be used. Simulation cells which are only periodic in certain directions are also supported. There is support for changing cell vectors during the simulation (like, e.g., obtained when simulating in an NPT ensemble with a barostat), but in this case, only statical analyses can be performed.

Bond breaking and formation during the simulation is also supported by TRAVIS. However, this is normally not a point of major importance, as it can only happen within *ab initio* MC or MD simulations[1,8] or when applying a reactive force field in classical simulations.[33,34] Many analyses rely on a fixed molecular structure, and nonstatic bonding renders many of them useless to some degree. Only a small subset of analyses remains useful in that case.

Molecule Recognition and Atom Labeling. Most MD and MC simulations contain molecules, which themselves are composed of atoms. The input files TRAVIS typically reads do only contain the atomic positions but lack the connectivity between the atoms. TRAVIS therefore performs a molecule recognition to create this connectivity information. Bonds are recognized from the first time step (or any other specified step) according to a distance criterion, which is based on the covalent radii of the atoms, but scaled by a factor slightly above 1 to account for bond vibrations.

Another point which requires attention is the fact that most MC and MD simulations use periodic boundary conditions and that molecules may be ripped apart by atom-wise wrapping into the simulation box. TRAVIS recognizes molecules that are distorted in the described way and reconstructs them.

Also of importance is the labeling of atoms within molecules. The desired way would be that TRAVIS names the atoms according to IUPAC nomenclature. But as the IUPAC system contains innumerable rules and exceptions for all kinds of chemical systems, it would completely go beyond the scope of this project to implement a IUPAC compliant naming scheme into TRAVIS.

Instead of this, TRAVIS uses topological atom ordering to label the atoms, like described by Shelley et al.[35,36] Every atom gets a priority (a so-called *atom code*) assigned by an iterative algorithm. After this algorithm has converged, it can be guaranteed that all nonequivalent atoms possess different atom codes, and all equivalent atoms have the same atom code. This is very helpful when trying to identify equivalent atoms in complicated molecules. Atoms that are "more central" in the molecule get higher atom codes. The atoms of each element are sorted by their atom codes in descending order and given labels in this order (e.g., C1 is more central than C6, see Figure 1). This algorithm ensures that equivalent molecules are always labeled in the same way, no matter in which order the atoms are specified in the input trajectory. Furthermore, this enables TRAVIS to process trajectory files in which the atom order is nonuniform or mixed up.

If the simulated system contains several molecules, TRAVIS recognizes molecules of the same kind and puts them into groups. Molecules with the same sum formula but different connectivity (i.e., structural isomers) are distinguished, as the comparison algorithm is purely topological and relies on the atom codes and connectivity information.

Wrapping and Unwrapping. As already mentioned, most MC and MD simulations use periodic boundary conditions.[15] Some analyses require the molecules to be wrapped into the box (like e.g. RDFs and SDFs), other analyses (like mean square displacement) explicitly require the trajectory to be nonwrapped. Furthermore, some trajectory input files may be wrapped and others may not. According to this, TRAVIS needs to wrap/unwrap the trajectories by itself as required for the analysis specified.

The wrapping of molecules into a simulation box is a relatively easy procedure. There is the choice to wrap the molecules atom-wise, possibly breaking them into parts (useful for computation of RDFs between atoms), or to wrap them molecule-wise,

keeping molecules intact (useful for most other analyses). The box center for wrapping can be defined, it does not necessarily need to be the reference particle of the analysis. The choice of the box center offers additional flexibility, which is required for some tasks (e.g., analyzing trajectories of a big ion pair in a solvent).

It is in principle not possible to unwrap a trajectory which has been wrapped into a box before, as the information is lost (the wrapped trajectory is a kind of quotient space, and there is no unique mapping back to the original space). However, there is a possibility to get the information back and unwrap the trajectory, if some assumptions to the simulation can be made. If it can be guaranteed that no particle can move further than half of the box length during one time step (i.e., the time step is sufficient small), then a particle that nevertheless jumped further than half of the box length during one time step must have been wrapped. This wrapping can then be undone. TRAVIS uses this empirical algorithm to unwrap wrapped trajectories if required by the analysis.

Virtual Atoms. To perform most of the analyses presented here, some points within the simulated systems have to be chosen. Normally, these points will be chosen to be atoms of the system. With TRAVIS, it is also possible to define points in positions where no atoms are located. These auxiliary points are called *virtual atoms*. Virtual atoms can be used like normal atoms in all kinds of analyses. It is also possible to write out a trajectory containing the virtual atoms additional to the real atoms.

Virtual atoms can be defined in two ways. The first way is to specify a set of atoms and a weighting factor for each atom. The virtual atom is then defined as the weighted average of the atoms. Apart from entering weighting factors by hand, it is also possible to use the center of mass or the center of geometry of the specified atoms.

The other way of defining a virtual atom is to choose three atoms A, B, and C from one molecule. The virtual atom X is then defined by specifying the distance X-A, the angle X-A-B, and the dihedral angle X-A-B-C. This makes it possible, e.g., to locate virtual atoms above and below the $[C_2C_1Im]^+$ ring plane and to calculate functions like RDFs using these atoms.

Dipole Vectors and Dipole Moments. If the input trajectory contains electronic structure information (e.g., maximally localized wannier centers[37,38] from *ab initio* MD[1,8]), or if fixed partial charges for the atoms are given, TRAVIS can evaluate dipole moments for molecules, groups of molecules, or for the whole system. If the subsystem for which the dipole moment should be calculated has a total charge different from zero, there is no unique definition of a dipole moment. In this case, the value of the dipole moment depends on the choice of a reference point. Usually, the center of geometry of the subsystem is taken as reference point, but TRAVIS allows to choose any other point.

The obtained dipoles can be used for the calculation of IR spectra, like explained below, or to obtain histograms over dipole moments (see section Dipole Distribution Function). It is also possible to visualize the dependency of dipole moments from other quantities like distances or angles. Apart from that, dipole vectors can be used in all analyses that are defined using vectors (like e.g. angular distribution functions or vector reorientation dynamics). Higher order multipoles are currently not supported by TRAVIS.

Equitable Binning. When creating a histogram by using "classical" binning, each piece of data is assigned to exactly one bin, and the value of this bin is then increased. If the value of the piece of data is very close to the border between two bins, a *bifurcation* exists: An infinitesimal change of the input value will cause that the original bin is left unchanged and the other bin is increased instead - a "macroscopic" effect relative to the infinitesimal change. This introduces unnecessary noise into the histogram.

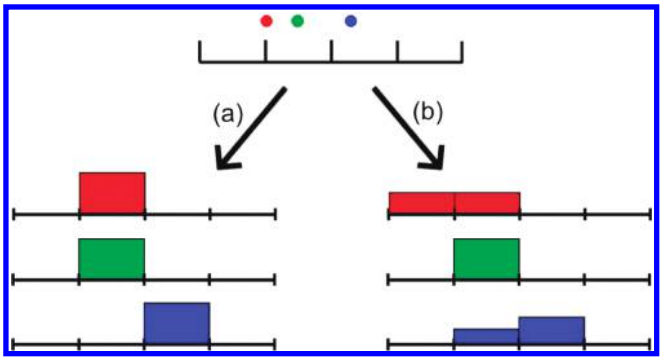


Figure 2. Comparison of binning methods: (a) standard binning (left) and (b) equitable binning (right).

TRAVIS uses a method that we call *equitable binning* to create histograms. To our knowledge, this method was not published before related to trajectory analysis. A piece of data that should be inserted into the histogram is shared in an equitable way between all adjacent bins. Adjacent bins are the ones that are closer than half the bin width to the exact position of the piece of data. For a visual description of this method, see Figure 2. Equitable binning can also be extended to higher-dimensional histograms. It is used, e.g., for the computation of combined distribution functions (2D binning) and spatial distribution functions (3D binning) by TRAVIS.

This algorithm should *not* be regarded as a smoothing procedure, as smoothing algorithms are applied to the histogram after the analysis. In contrast to that, equitable binning directly works on the raw input data, using additional information that is left out by classical binning. It does, e.g., not lead to line broadening, like expected when applying smoothing methods. As almost all analyses in TRAVIS use histograms in some stage, the use of equitable binning increases the visual quality of almost all plots that can be created.

Smoothing and Optimized Algorithms. Apart from the equitable binning method, a smoothing procedure can be applied to the output data (see, e.g., Figure 7). The question if smoothing improves the plot appearance is highly subjective and dependent on the system; therefore, the user may choose the strength of smoothing that is applied or leave out the smoothing at all.

As the correlation of large data sets is a very time-consuming step, TRAVIS performs all correlations by using fast Fourier transform according to the Wiener-Khinchin theorem.[39−41] This greatly reduces the required computer time if autocorrelations are involved by the selected analysis.

When calculating spectra, which are defined as the Fourier transform of some autocorrelation functions[42] (see below), TRAVIS allows the user to utilize sophisticated methods for obtaining neat results. Examples are *zero padding*, which increases the spectral resolution by adding a block of zeros at the end of the input data, or the appliance of a window function,[43,44] e.g. of Blackman-Harris type (cosine series) or Hanning type ($\cos^2$), which greatly improves the singal-noise ratio by suppressing the spurious Fourier spectrum of the input data envelope curve.

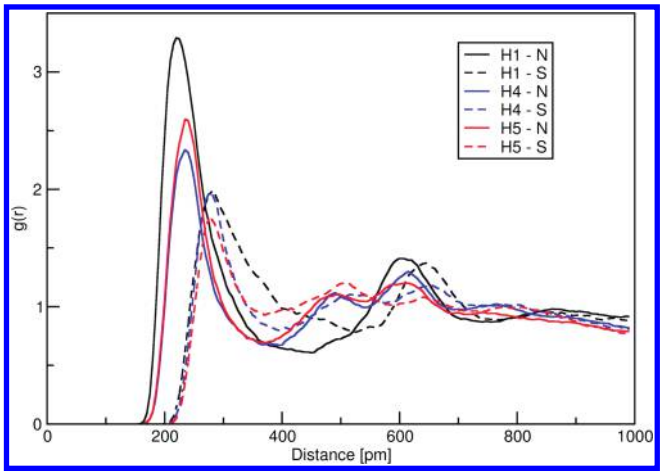


Figure 3. RDFs between $[C_2C_1Im]^+$ ring H atoms and $[SCN]^-$ S, N atoms.

STRUCTURAL ANALYSES

The term "Structural Analyses" in this article refers to those analyses that do not explicitly require a time axis. The algorithms presented in this section can be applied to any list of coordinate frames, including any MC or MD trajectory.

Please note that there are different meanings of terms like "two-dimensional" in common use when talking about functions. In this article, the dimensionality of functions refers to the sum of dimensionalities of preimage and image of the function. In simple words, a function mapping from a one-dimensional space to another one-dimensional space, like e.g. $f(x) = y$, is called a two-dimensional function here, as its plot is two-dimensional.

Two-Dimensional Distribution Functions. The functions shown in this subsection are histograms over the development of scalar quantities (like distance or angle) measured in the system during the simulation. According to the definition stated above, they are two-dimensional functions.

Radial Distribution Functions. A widely used way of analyzing trajectories and obtaining information about the system structure is the calculation of radial pair distribution functions (RDFs),[9,10] sometimes also referred to as pair correlation functions. By making available averaged distance information between two particles in the system, RDFs are a simple, flexible tool that can be universally applied. Another point contributing to the importance of RDFs is the fact that they can not only be computed from simulations but also be calculated from the results of neutron and X-ray diffraction experiments[45] for some kinds of samples. This enables validation of performed simulations by comparing the calculated RDFs to the ones obtained from the experiments. As RDFs involve different particles at the same time, we will use the terms *reference particle* and *observed particle* in the following.

A RDF gives the probability of finding an observed particle in a certain distance to a reference particle - relative to the uniform density of the observed particle. The uniform density is the density one would find if all the observed particles would be uniformly distributed in the simulation cell. RDF values larger than 1 depict that finding an observed particle in this distance to the reference particle is more probable than it should be on average. For large distances, the value of the function should tend toward 1. The definition of a RDF is given in eq 1, where r_i and r_j denote the position vectors of the i-th and the j-th particle, and δ takes the value 1 in the interval $[-w,w)$ (with w being the bin width), otherwise 0. RDFs calculated with TRAVIS already were published[12−14]

$$g_{ab}(r) = \frac{V}{N_a \cdot N_b} \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_b} \langle \delta(r - |\vec{r}_i(t) - \vec{r}_j(t)|) \rangle_t \qquad (1)$$

Some RDFs calculated for the analyzed model system are shown in Figure 3. Please note that for these RDFs the reference particle and the observed particle are chosen from different molecules - hence the information gained from the plot is of *intermolecular* nature. It is also possible to choose the reference particle and the observed particle from the same molecule to obtain *intramolecular* information, e.g. histograms over bond length developments.

Figure 4 shows the temporal development of the H1−N distances for one selected $[C_2C_1Im]^+$ and all 32 $[SCN]^-$ ions. This is a part of the raw data for the RDF represented by the black solid line in Figure 3 - the latter represents the distance histogram over this data. The right part of Figure 4 conveys how the distance developments forms the RDF. For five of the $[SCN]^-$ ions, the distance development is shown in color, while the remaining 31 curves are drawn in shades of gray to avoid confusion. The RDF is calculated from the distance information between all 32 $[C_2C_1Im]^+$ and all 32 $[SCN]^-$ ions - not only the 32 distance developments shown here. For our example cation, we observe four different anions in close coordination during the simulation (black, green, pink, and red curve). Plots of this kind are useful to analyze the exchange behavior of coordinated molecules. This is, e.g., useful for investigating anion−cation coordination or the solvation behavior of molecules or metal complexes (which was already performed by using TRAVIS and will be published soon).

Angular Distribution Functions. Like RDFs are histograms over a distance development, angular distribution functions (ADFs) are histograms over the development of a certain angle in the system, see eq 2. The term containing sin (α) corrects the uniform angular distribution (also referred to as *cone correction*[46]) and can be omitted by TRAVIS if desired. In addition to the angular distribution function itself, combined temporal development/distribution plots (like shown in Figure 4 for the distance) can be created.

$$ADF_{abc}(\alpha) = \frac{1}{\sin(\alpha)} \cdot \frac{1}{N_a \cdot N_b \cdot N_c} \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_b} \sum_{k=j+1}^{N_c} \langle \delta(\alpha - \angle(\vec{r}_i(t), \vec{r}_j(t), \vec{r}_k(t))) \rangle_t \qquad (2)$$

The angle of interest can be defined by specifying two vectors. Both vectors are specified by two points. The base points of the defined vectors may be different, which offers additional flexibility. The points defining the vectors do not necessarily have to be part of the same molecule. Apart from position vectors, it is also possible to use velocity or acceleration vectors of atoms or molecules, dipole vectors of molecules (see section Dipole Vectors and Dipole Moments), or any combination of these vectors. This enables powerful analyses like, e.g., the evaluation of the angle between a bond and the molecular dipole vector.

Dihedral Distribution Functions. Similar to RDFs and ADFs, dihedral distribution functions (DDF) are histograms over dihedral angles. There are two possible ways of defining a dihedral angle in TRAVIS. The standard definition, like used in

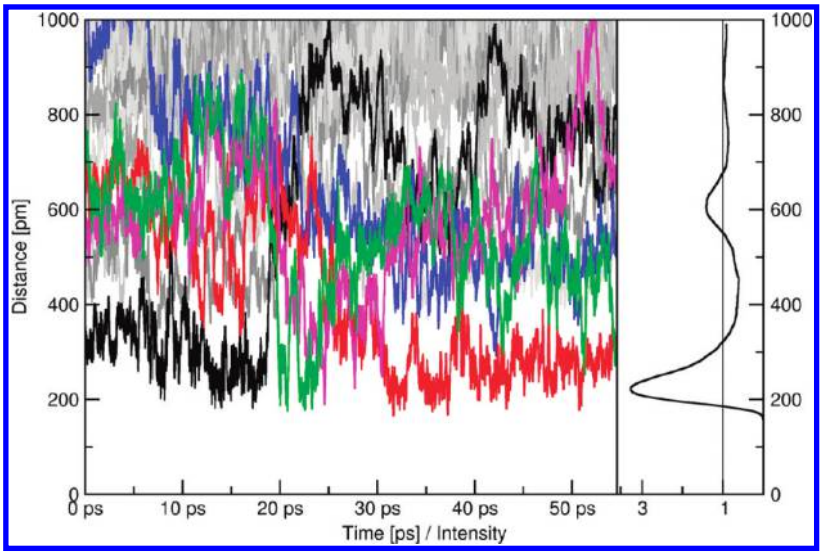


Figure 4. Temporal development of distance between one selected $[C_2C_1Im]^+$ H1 and all $[SCN]^-$ N atoms (left) and corresponding RDF (right).

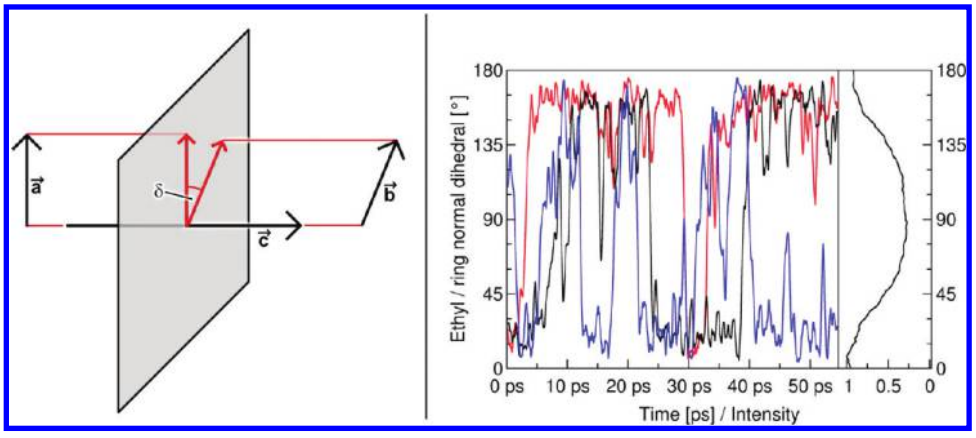


Figure 5. Generalized dihedral angle δ defined by three vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ (left) and temporal development of dihedral angle for three selected cations and corresponding DDF for all cations (right).

literature and many other programs, is to specify four points (e.g., atoms) which form the dihedral angle.

Additionaly, there is a generalized and more flexible method implemented in TRAVIS. The dihedral angle is specified by three vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$, which corresponds to six points instead of four. $\vec{a}$ and $\vec{b}$ are projected onto the normal plane of $\vec{c}$, and the angle between the projected vectors is the generalized dihedral angle, see the left side of Figure 5 and eq 3. The standard definition is identical to the generalized definition with the base point of vector $\vec{a}$ being the base point of $\vec{c}$ and the tip of $\vec{c}$ being the base point of $\vec{b}$

$$\delta_{\vec{a},\vec{b},\vec{c}} = \arccos\left(\frac{(\vec{a}\times\vec{c})\cdot(\vec{b}\times\vec{c})}{|\vec{a}\times\vec{c}|\cdot|\vec{b}\times\vec{c}|}\right) \quad (3)$$

As already explained when discussing ADFs, not only position vectors but also velocity, force, and dipole vectors can be used for defining dihedral angles.

On the right side of Figure 5, an example of a temporal dihedral development and the corresponding histogram are shown. The dihedral specified using the general (three vectors) definition. Vector $\vec{a}$ was selected as the normal vector of the $[C_2C_1Im]^+$ ring plane. Vector $\vec{b}$ was chosen to be $\overrightarrow{C2-C6}$ and vector $\vec{c}$ was $\overrightarrow{N1-C2}$. The temporal development is shown for three selected $[C_2C_1Im]^+$ cations. The histogram on the right includes all cations. Plots like shown on the right side of Figure 5 are useful for investigating flexible groups (e.g., alkyl chains).

Dipole Moment Distribution Functions. If dipole moments for the simulated system are available (see section Dipole Vectors and Dipole Moments), TRAVIS can compute dipole moment distribution functions (DipDF). It is also possible to visualize the temporal development of the dipole moments similar to Figure 4.

Spatial Distribution Functions. Like RDFs show the probability of finding a particle in a certain distance to another particle, spatial distribution functions (SDFs)[47] depict the probability of finding a particle at a certain position in space around a fixed reference system of other particles. SDFs are a kind of "three-dimensional enhancement" to RDFs. A reference molecule has to be selected. For setting up a reference frame, three atoms (which must not be located in one line) within this

molecule have to be chosen: The first one is placed in the origin, the second atom is located on the positive X axis, and the third atom is put into the $X-Y$ plane (with positive Y value) of the reference frame. Every instance of the reference molecule in every time step is transformed into this unique orientation, which enables averaging of probabilities in the space around the reference molecule. This is only one possibility of setting up a reference frame from 3 points, but as the first reference atom is placed in the center of the frame, this enables direct control over the SDF center.

As explained in the Introduction, a probability in three-dimensional space is a four-dimensional function, which is obviously hard to visualize as it is. It is necessary to reduce the dimensionality to three (or even lower). A common and reasonable method for doing so is to visualize iso-surfaces on the data, which can be easily drawn in three-dimensional space. Iso-surfaces are surfaces that pass through all areas with the same probability, just like countour lines in topographic maps (but there in one dimension less). The values of iso-surfaces are typically given in the unit of particle density (nm^{-3} or pm^{-3}), specifying the particle density along the iso-surface. SDFs computed with TRAVIS already appeared in the literature.[12−14]

In Figure 6, an intramolecular SDF of the $[C_2C_1Im]^+$ cation is shown. This means that here one $[C_2C_1Im]^+$ at a time is the reference molecule, and only atoms within the reference molecule are observed. The "rotation" of the methyl and ethyl chains is shown, resulting in toroidal shapes. Please note that *rotation* is definitely a misleading term here, as these pictures do not predicate if the groups are actually rotating, as there is no dynamical information included. When considering Figure 6, it is also possible that the ethyl groups of each $[C_2C_1Im]^+$ stay in place all of the time, but in different positions for each $[C_2C_1Im]^+$.

When analyzing trajectories of medium size (e.g., 32 molecules over 55 ps, like performed in this work), the sampling of the data is usually too poor for direct SDF creation. The resulting SDFs look like the one shown in Figure 7 on the left. The appearance of this SDF can be improved by smoothing the underlying data. TRAVIS contains a smoothing algorithm for SDFs, the level of smoothing that is applied can be selected (including the choice not to smooth at all). As an example, two SDFs smoothened to different levels are shown in the middle and on the right of Figure 7. The SDFs shown there are intermolecular - one $[C_2C_1Im]^+$ cation at a time is fixed and the distribution of all $[SCN]^-$ ions (in this case by the $[SCN]^-$ center of mass) around the cations is observed.

One problem with iso-surfaces is that they often occlude important parts by their bulkiness. A good way for encountering this problem is to introduce a clipping plane into the system, as shown in Figure 8 on the left side (this is not a feature of TRAVIS but one of VMD[31]). Again, the purple color depicts the particle density of the $[SCN]^-$ anion around the $[C_2C_1Im]^+$ reference molecule. Without this clipping plane, only a purple colored spherical body would be visible, refusing any insight into the interior.

On the left side of Figure 8, it can be seen that there is an area close to the $[C_2C_1Im]^+$ cation where never any $[SCN]^-$ anion can be found - just because of the repulsive forces between close

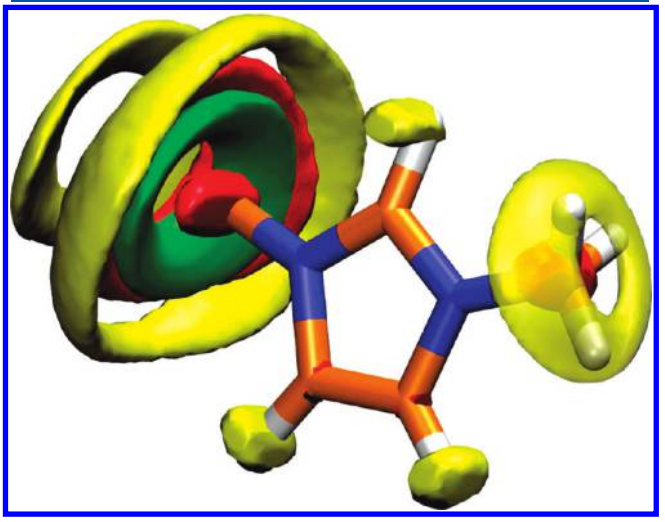


Figure 6. Intramolecular SDF of the $[C_2C_1Im]^+$ atoms (red: carbon; green: H2, H3; yellow: other hydrogen atoms).

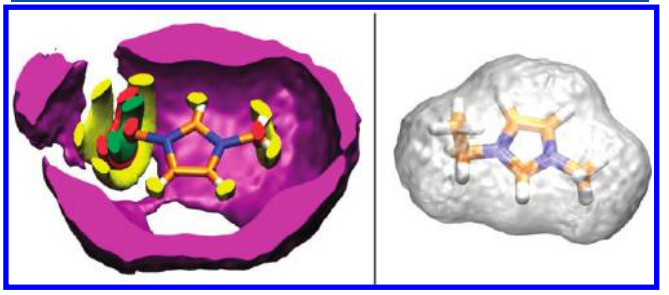


Figure 8. SDF of $[SCN]^-$ around $[C_2C_1Im]^+$ with clipping plane (left) and inverted SDF shows the excluded volume of $[C_2C_1Im]^+$ (right).

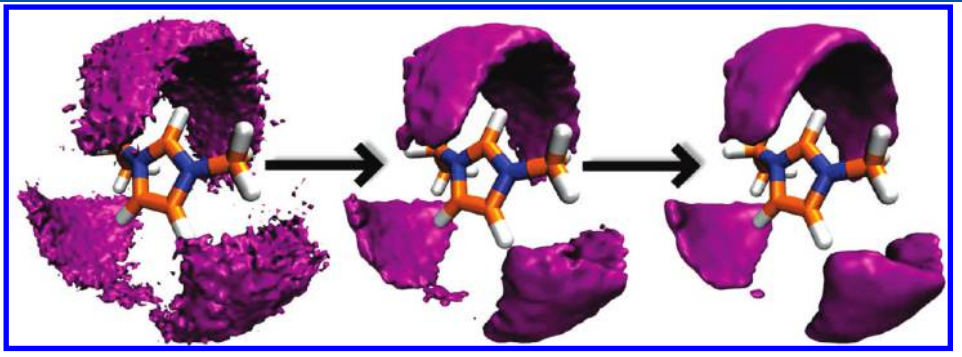


Figure 7. Spatial distribution function of the $[SCN]^-$ center of mass around $[C_2C_1Im]^+$ (left) and different smoothing levels of the spatial data (middle and right).

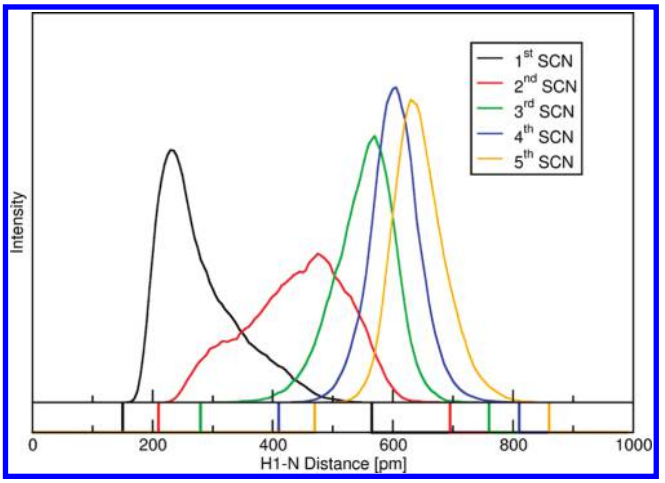


Figure 9. Distance distribution of the first 5 neighboring $[SCN]^-$ N atoms from $[C_2C_1Im]^+$ H1.

atoms. TRAVIS can invert the SDF to show only this inner region, where never any other particle enters, see Figure 8 on the right. The shown volume can be understood as the excluded volume of the $[C_2C_1Im]^+$ cation, which can be numerically integrated to give the molecular volume and the molecular surface area out of a MD or MC simulation.

Neighborhood Analysis. Sometimes it is of interest, at which distance to a reference particle the next neighboring particle of another kind can be found. TRAVIS can perform a neighborhood analysis for a given reference particle and a given observed particle. This is similar to creating a RDF (like shown in Figure 3), except for the fact that for every reference particle and every time step only the closest observed particle is taken into account. The resulting distribution shows the minimal and maximal distance of the first neighbor and in which distance the first neighbor can be found most of the time. This analysis can also be performed for further distant neighbors.

Such a neighborhood analysis is shown in Figure 9. The reference particle is the H1 atom of $[C_2C_1Im]^+$; the observed particle is the N atom from $[SCN]^-$. The black curve corresponds to the first neighbor, as explained above. The red curve is created by the same principle but only containing the second neighbor in every configuration. The green curve shows only the third neighbors and so on. The vertical lines in the lower part of the plot depict the minimal and maximal values where this kind of neighbor was ever found (technically, a strong magnification of the plot in the *Y* direction).

The first neighboring $[SCN]^-$ anion (black curve) can be found in the range between 200 and 300 pm most of the time, as expected. In some cases it approaches H1 as close as 150 pm. Also interesting is the fact that there are cases, in which the next neighboring $[SCN]^-$ nitrogen atom is as far as 550 pm away. This means that there are configurations in which no $[SCN]^-$ nitrogen atoms can be found in a 500 pm sphere around H1. These configurations were found with a probability of 78.3 ppm within the trajectory.

On the other hand, the fifth next neighboring $[SCN]^-$ nitrogen atom (orange curve) is sometimes nearer than 500 pm to H1, which means that there exist configurations in which five different $[SCN]^-$ nitrogen atoms can be found in a 500 pm sphere around H1 at the same time. These configurations were observed with a probability of 4.32 ppm.

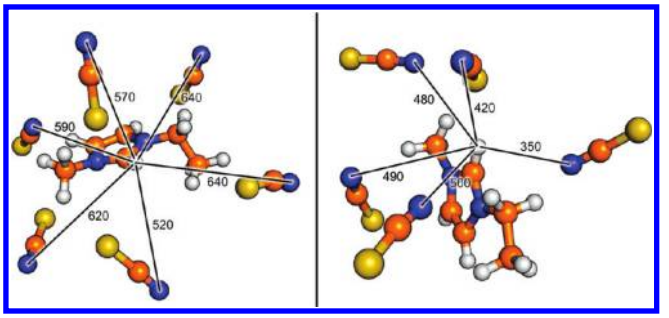


Figure 10. (a) Rare conformation with no $[SCN]^-$ N atom in a 500 pm sphere around H1 (left) and (b) another rare conformation with 5 $[SCN]^-$ N atoms in a 500 pm sphere around H1 (right). Distances in pm.

TRAVIS is able to write out snapshots of such rare events. As an example, in Figure 10 one selected configuration with no $[SCN]^-$ nitrogen atom in the 500 pm sphere around H1 is presented at the left (with all distances to the nitrogen atoms labeled). On the right side, a configuration with five different $[SCN]^-$ nitrogen atoms in the 500 pm sphere is shown. Without this feature, it would have been very hard to filter out these rare situations by hand.

Combined Distribution Functions. In the last section, we presented four kinds of analyses (RDF, ADF, DDF, DipDF) which are histograms over scalar data (there is another analysis also fitting into this scheme, the velocity distribution function, but as a dynamical analysis, it is not in the scope of this section). With TRAVIS, it is also possible to combine several of those functions in order to create histograms of higher dimensionality, which are called combined distribution functions (CDFs). Each different analysis delivering scalar raw data to the CDF is called a *channel* of the CDF. Because every channel can be chosen to be one out of five two-dimensional analyses, and every of these analyses can be defined in a variety of ways, some of the possible CDFs never appeared in literature before. Just to mention one unusual example, it would be possible to plot the angle between the molecular dipole vector and the center of mass velocity vector against the total velocity of the molecule. Because high-dimensional histograms are hard to handle and to visualize, we will only discuss two-channel CDFs here, resulting in two-dimensional histograms.

In order to demonstrate the usefulness of CDFs, we will consider three examples in the following. We start with combining a RDF between the H1 atom of $[C_2C_1Im]^+$ and the N atom of $[SCN]^-$ and an ADF between the H1−C2 vector in $[C_2C_1Im]^+$ and the intermolecular H1−N vector between one $[C_2C_1Im]^+$ and one $[SCN]^-$. The underlying distance for the RDF is visualized in Figure 11 on the right as r and the angle for the ADF is labeled as α. This could be performed e.g. in order to observe possible hydrogen bonding, as a hydrogen bond involving three atoms is often defined by a distance and an angle.[48,49] For each time step of the input trajectory and for each pair of $[C_2C_1Im]^+$ and $[SCN]^-$ (all combinations), the RDF and the ADF both create one scalar value. Normally, we would have put the RDF value into the linear RDF histogram and the ADF value into the linear ADF histogram. But in this case, we regard the two values as 2-tuple, they are now connected and related to each other, because they stem from the same configuration. The obtained 2-tuple is put into a two-dimensional histogram (which can be regarded as a matrix).

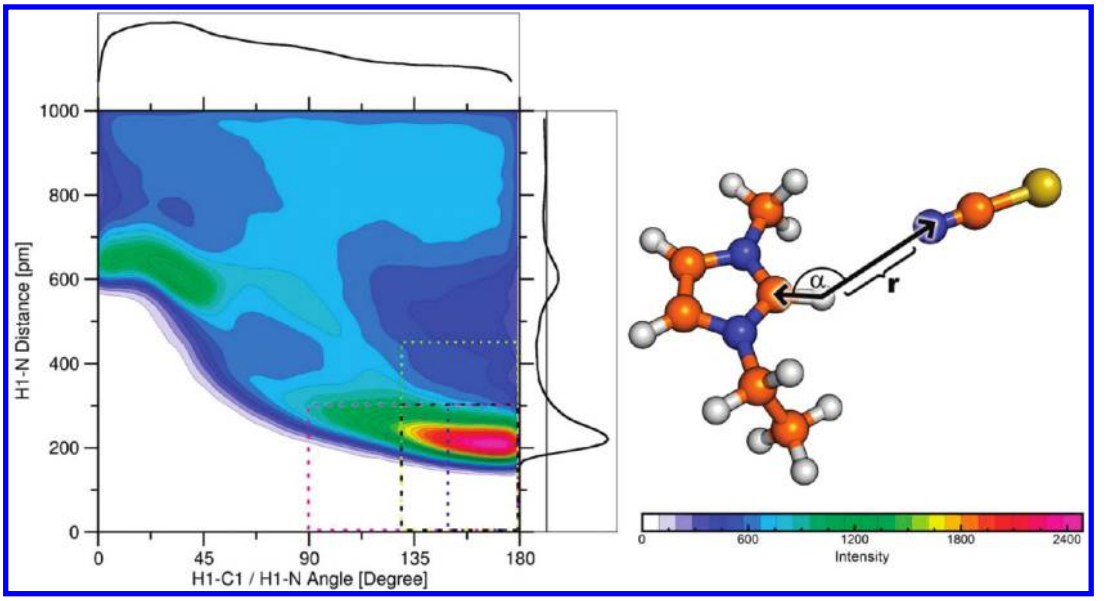


Figure 11. Combined radial/angular distribution function; four possible H bond criteria are shown by the dashed lines.

The resulting histogram over all time steps and all anion cation pairs is shown in Figure 11 on the left. If the distance and the angle would be independent of each other, the resulting plot would contain the same information as the RDF and the ADF on their own (the resulting plot would then be the product of the two distribution functions). But as the two values are dependent on each other, a CDF contains more information than just a RDF and an ADF together.

One can see that for angles $\alpha > 90°$ the preferred H1−N distance is around 200 pm, which corresponds to a configuration where the $[SCN]^-$ nitrogen atom is close in front of the $[C_2C_1Im]^+$ H1 atom. In this configuration, typically a hydrogen bond between H1 and N is formed. An angle of $\alpha = 180°$ would indicate a linear hydrogen bond. For small angles $\alpha < 45°$, there is another peak at fairly high distances around $r = 600$ pm. This peak corresponds to a configuration in which an $[SCN]^-$ anion is coordinated to the other two $[C_2C_1Im]^+$ ring hydrogen atoms H3 and H4. On the top and right axis of Figure 11, the two-dimensional functions, which act as data sources for the CDF, are plotted. Although not being used frequently, functions of similar kind already appeared in the literature.[47,50]

The dashed lines drawn in the plot in Figure 11 depict four different criteria that could be chosen to define hydrogen bonding between $[C_2C_1Im]^+$ and $[SCN]^-$. The standard approach when defining a hydrogen bond criterion is often to take the first minimum of the RDF as a distance cutoff ($0 \le r \le 450$ pm) and an angle interval of e.g. $130° \le \alpha \le 180°$. This is depicted by the dashed yellow line. But according to the visual impression from Figure 11, this seems to be a suboptimal criterion here, as it crops parts of the peak but takes into account contributions from quite large distances instead. A better choice could be a criterion with $0 \le r \le 300$ pm and $90° \le \alpha \le 180°$, which is visualized by the dashed pink line. A CDF is a powerful tool for defining criteria, as it offers much more information than just the RDF. Two other possible choices of criteria ($0 \le r \le 300$ pm/$130° \le \alpha \le 180°$, dashed black line, and $0 \le r \le 300$ pm/$150° \le \alpha \le 180°$, dashed blue line) are also shown. The dynamics according to these four criteria are analyzed in the section concerning Dimer Existence Autocorrelation Functions. It will be shown that these are very diverse.

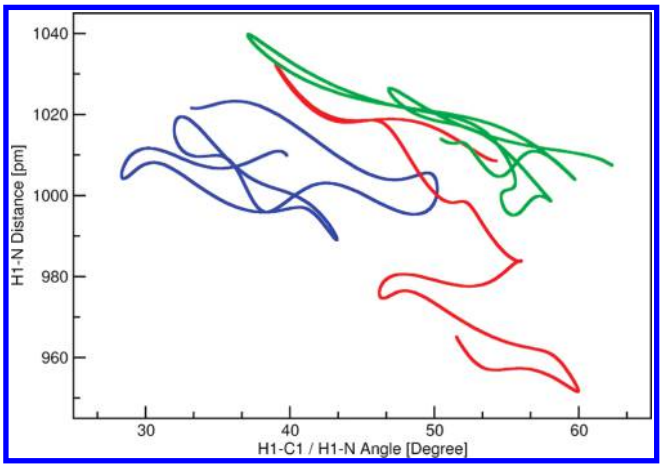


Figure 12. Temporal development of distange and angle between one $[C_2C_1Im]^+$ and three selected $[SCN]^-$ (red, green, blue). Raw data for the CDF shown in Figure 11.

When presenting the RDF and DDF, the temporal developments of the distance and dihedral angle, which are the raw data for the histograms, were also shown. The raw data forming the just explained CDF can be shown in a similar way, see Figure 12.

The two axes of the plot correspond to distance and angle - the two values which are the input data for the CDF. As the time proceeds, the changing values draw curves into this plot. One curve corresponds to one pair of $[C_2C_1Im]^+$ $[SCN]^-$. As there are 32 ion pairs simulated, there should be $32 \cdot 32 = 1024$ curves. To avoid confusion, only three of them are shown in Figure 11, and the shown time interval is very short (around 0.5 ps). TRAVIS can create this output as a movie as well. Such a movie can be found in the Supporting Information.

As stated before, all kinds of two-dimensional analyses can be combined to give three-dimensional CDFs. In Figure 13, we present an example, in which two RDFs are combined. The first

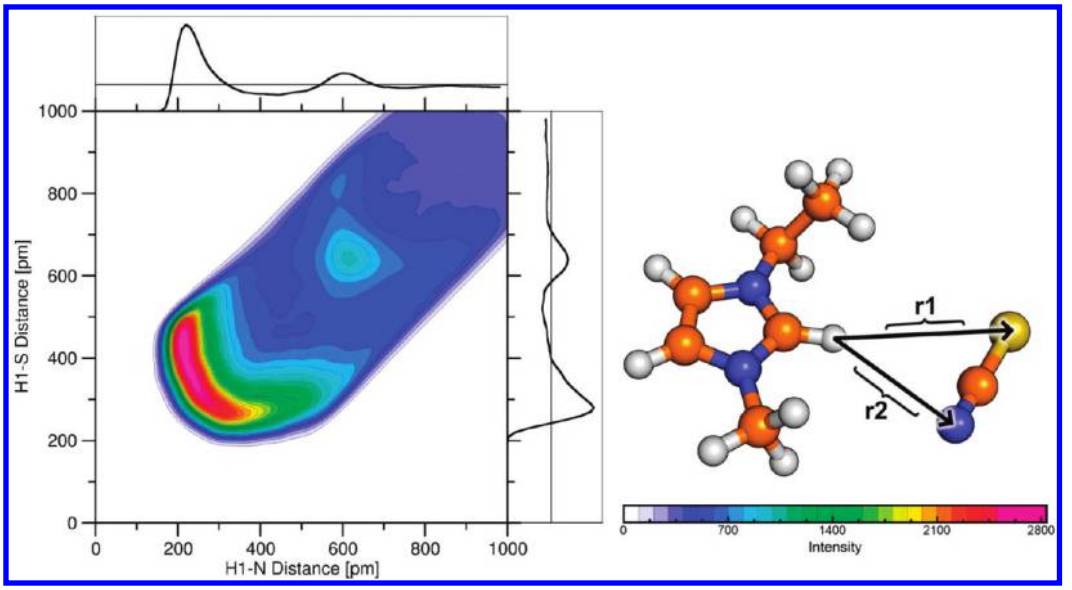


Figure 13. Combined radial/radial distribution function between $[C_2C_1Im]^+$ and $[SCN]^-$.

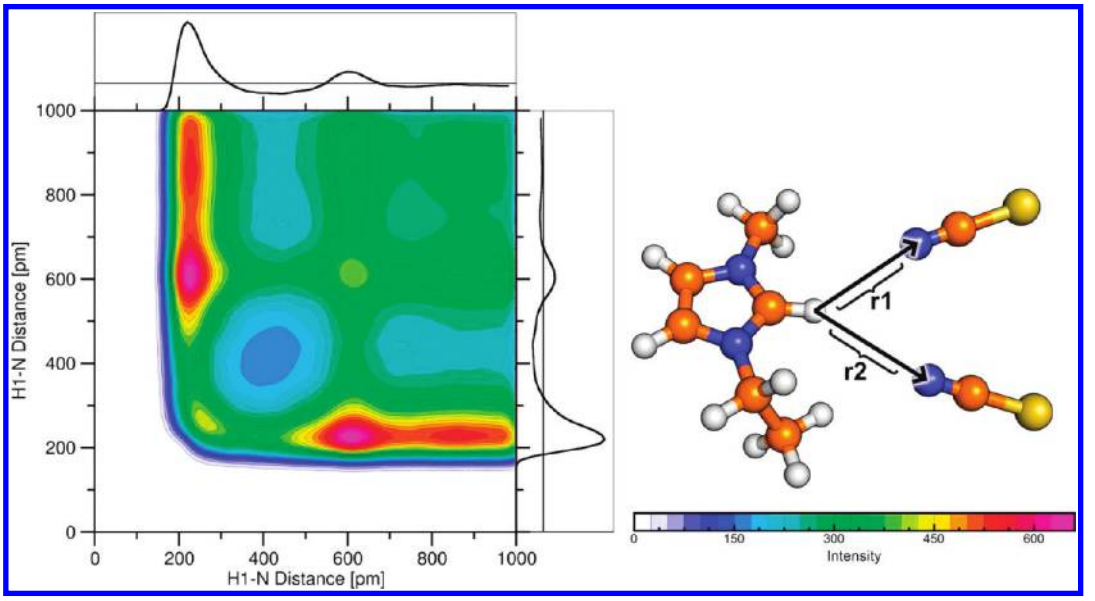


Figure 14. Combined radial/radial distribution function between $[C_2C_1Im]^+$ and two different $[SCN]^-$.

RDF observes the distance between the H1 atom of $[C_2C_1Im]^+$ and the N atom of $[SCN]^-$, and the second RDF observes the distance between H1 and the S atom of the same $[SCN]^-$ ion, as illustrated in Figure 13 on the right. The $[SCN]^-$ is mainly coordinated via the N atom, which is indicated by the peak at H1−N = 250 pm/H1−S = 450 pm. Configurations in which both N and S atoms have a similar distance of around 300 pm to H1 are also frequently found. However, cases with only the S atom coordinating to H1 are relatively rare, as the peak at H1−N = 450 pm/H1−S = 250 pm is weak. The asymmetry of the plot indicates the different coordination behavior of the N and S atom from $[SCN]^-$.

The following analysis, which is presented in Figure 14, shows again a CDF which is constituted from two RDFs. But a major difference to the preceding cases is that here the two RDFs observe different $[SCN]^-$ ions. Both RDFs are defined by the H1−N distance. The 2-tuples described above are therefore created not only for every possible pair of $[C_2C_1Im]^+$ and $[SCN]^-$ also but for every possible triple of $[C_2C_1Im]^+$, $[SCN]^-1$, and $[SCN]^-2$. Excluded are only cases where $[SCN]^-1$ equals $[SCN]^-2$, which would result in a strong diagonal contribution in the plot.

The two main peaks in Figure 14 at around 600 pm/200 and 200 pm/600 pm depict configurations with one $[SCN]^-$ N atom tightly coordinated to H1 and another $[SCN]^-$ N atom coordinated to the rear of $[C_2C_1Im]^+$ (H4 or H5). Cases in which two $[SCN]^-$ N atoms coordinate to H1 at the same time (200 pm/200 pm) or cases in which no $[SCN]^-$ N atom coordinates to H1 and both are in the rear (600 pm/600 pm) are found only infrequently.

dx.doi.org/10.1021/ci200217w |J. Chem. Inf. Model. 2011, 51, 2007–2023

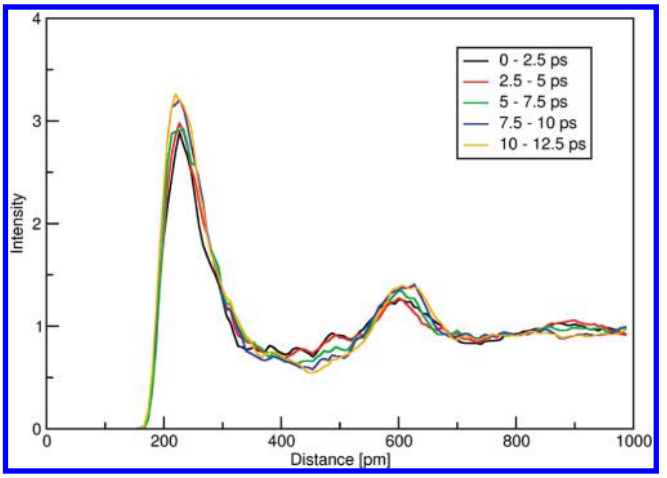


Figure 15. RDFs between $[C_2C_1Im]^+$ H1 and $[SCN]^-$ 1 N, computed for different intervals of the trajectory.

Multi-Interval Analysis. All analyses in TRAVIS (not only the ones presented up to this point) can be performed in a multi-interval manner. This means that the input trajectory is automatically separated into parts. The analyses are performed, and results are written out separately for each part. This enables to see if some properties change during the trajectory, e.g. if the system is well equilibrated from the beginning on or not.

As an example of a multi-interval analysis, the H1−N RDF, which was already presented in Figure 3 by the solid black line, is shown in Figure 15. The first 12.5 ps of the trajectory have been automatically dissected into 5 nonoverlapping intervals of the same length. Taking the average of these 5 curves would give the overall RDF over the first 12.5 ps. It can be seen that the black, red, and green curves (depicting the first 3 intervals) show some differences to the other 2 curves over the whole plot. The remaining 2 curves show good accordance to each other. This behavior usually means that the system is not well equilibrated from the beginning on, and some picoseconds should be left out from analyzing.

DYNAMICAL ANALYSES

In contrast to the structural analyses presented above, dynamical analyses take into account the simulation time and therefore can only be applied to time-dependent trajectories.

Temporal Development Overlay Plots. One way for direct visualization of the temporal motion of single molecules is the temporal development overlay (TDO) plot. To create a plot of this kind, one molecule from the trajectory and some points in time have to be chosen. If desired, a defined count of neighboring molecules can also be taken into account. TRAVIS then creates a file which contains the molecular conformation at the defined points in time. This file is a small script which executes PyMol[51] and automatically loads the coordinates. A bleaching in the atom colors visualizes consecutive geometries, and the initial structure is shown with full color saturation. TRAVIS gives the choice if the defined cluster should be centered (to remove translation) and if the cluster should be rotated to an uniform orientation (to remove rotational movement). TDO plots created with TRAVIS were already published.[13]

Figure 16 shows an example of a TDO plot. The temporal motion of one selected $[C_2C_1Im]^+$ ion, together with its five nearest $[SCN]^-$ neighbors around, is shown. Each snapshot was taken after 0.25 ps; therefore, the total plot with 5 snapshots comprises a time span of 1 ps. The ring center is fixed to the origin, which results in no translational movement of the $[C_2C_1Im]^+$ being visible. Rotation of the $[C_2C_1Im]^+$ is not removed but does not occur on this short time scale. However, the strong rotation of the methyl group is clearly visible.

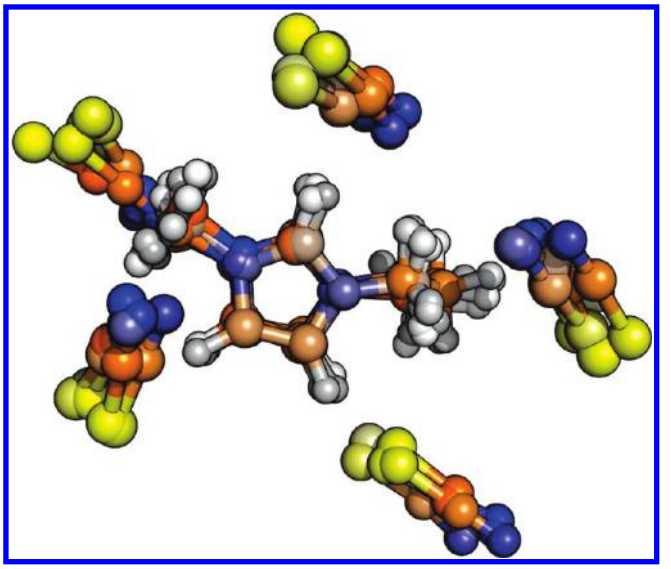


Figure 16. TDO plot of one $[C_2C_1Im]^+$ ion with five $[SCN]^-$ around.

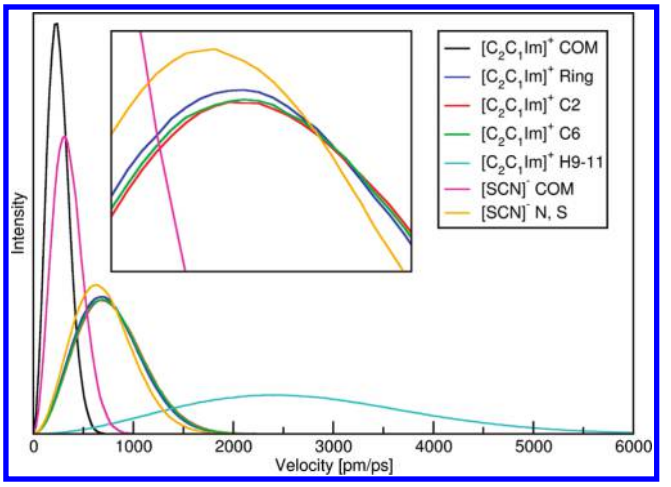


Figure 17. VDFs for different atoms and centers of mass (COM) in $[C_2C_1Im]^+$ and $[SCN]^-$.

Velocity Distribution Functions. When analyzing the dynamic properties of a trajectory, an obvious choice for an analysis is the velocity distribution, as the velocity is the central property of dynamics. TRAVIS can create velocity distribution functions (VDFs) for any defined center. This can be any atom or group of atoms and even points in space where no atom is located, e.g. center of mass or center of geometry.

As an example, Figure 17 illustrates the velocity distribution functions for selected atoms as well as the VDFs for the centers of mass (COM) for both ions and the $[C_2C_1Im]^+$ ring center. Like expected, each velocity shows a distribution which is qualitatively similar to the Maxwell−Boltzmann distribution from kinetic gas

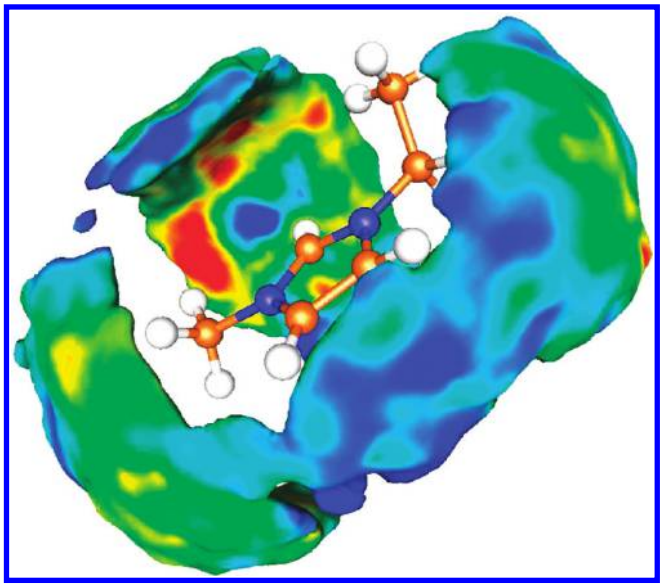


Figure 18. Spatial distribution function of the $[SCN]^-$ center of mass (COM) around $[C_2C_1Im]^+$; the coloring represents the averaged $[SCN]^-$ COM velocity (blue = slow, red = fast).

theory. Lighter atoms possess higher velocities on average, whereas heavier atoms or whole molecules (defined by their center of mass) move slower.

An interesting option is to add velocity information to SDFs like shown in Figure 7. The result is shown in Figure 18. The surface is defined in the same way as for the "classical" SDF: It is an isosurface of the particle density of $[SCN]^-$ around a fixed $[C_2C_1Im]^+$ molecule. Additional information is here given by the coloring of the surface. The color represents the averaged velocity of all $[SCN]^-$ at this point in space, relative to the fixed $[C_2C_1Im]^+$. Blue represents slow motion, whereas red depicts high average velocities.

Dimer Existence Autocorrelation Functions. In order to judge the lifetime of an aggregated species (called *dimer* here) defined by one or more criteria, a dimer existence autocorrelation function (DACF) can be computed by TRAVIS. The DACF for a given pair of particles *i,j* is defined as the autocorrelation of a simple function β_{ij}, which has the value 1 as long as the criteria are fulfilled, and switches to 0 as soon as the criteria fail for the first time (see eq 4). It stays at 0, no matter if the criteria are again fulfilled later on or not. The total DACF is the average over all different pairs *i,j*. According to this definition, DACFs start with the value 1 and then fall to 0 with increasing τ. The statement contained in the resulting function $DACF(\tau)$ is as follows: *"How large is the probability of the criteria being still fulfilled (without interruption) at a given time τ, when they were fulfilled at time 0?"*. DACFs calculated with TRAVIS already have been published[13]

$$DACF(\tau) = N \cdot \langle \sum_{t=0}^{T-\tau} \beta_{ij}(t+\tau) \cdot \beta_{ij}(t) \rangle_{i,j} \qquad (4)$$

TRAVIS supports two ways of defining dimers. One way is to specify a criterion based on angular and distance intervals. For each angle and distance that has been defined between two molecules, there may be one or more intervals given in which this criterion is fulfilled. Different angle and distance criteria may be combined with the logical operations "and" and "or", allowing complex combinations. Another way of defining a dimer criterion

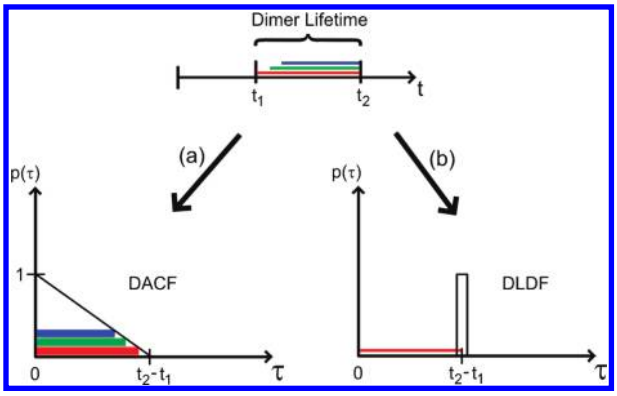


Figure 19. One dimer with specific lifetime: (a) corresponding DACF (left) and (b) corresponding DLDF (right).

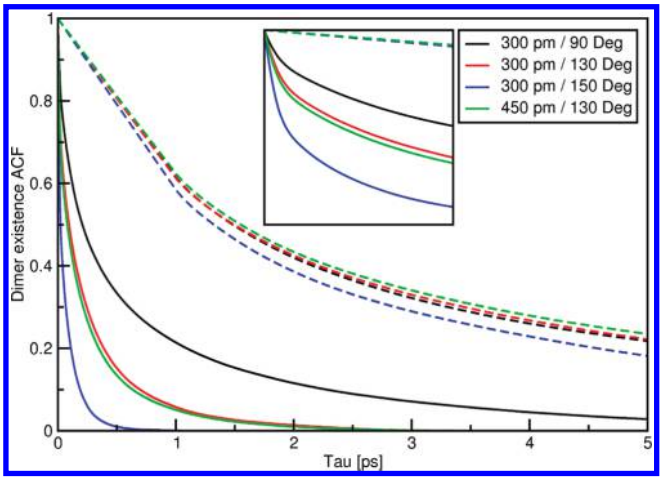


Figure 20. DACFs for the 4 criteria shown in Figure 11 (solid line = continuous; dashed line = intermittend 1 ps).

is to define a distance between molecules and take into account the nearest neighbors according to this distance definition. This makes it possible for one molecule to form a dimer with its next neighbor (or with its next *n* neighbors) at each time. Dimers defined in this way break if the neighborhood relations change, i.e. if the formerly next neighbor is no longer the next neighbor. The distance for determining the neighborhood may be based on several or even all atoms of the molecules. In the latter case, the atoms of the two molecules that are closest to each other are used to calculate the distance at each time, which corresponds to the smallest distance between the two molecules.

If there is only one pair of particles and this pair fulfills the conditions only during one single time interval, the resulting curve is a straight line, starting with value 1 at $\tau = 0$ and falling to 0, giving the shape of a triangle (which exactly corresponds to the statement in quotation marks given above). This behavior is visualized in Figure 19 on the left. In contrast to the DACF, the dimer lifetime distribution function (DLDF), which is shown there on the right, exhibits only one peak per dimer at the position of the total lifetime.

In complex systems, there are many dimer pairs, each of them fulfilling the conditions during many different time intervals. But the resulting DACF curves are still composed as sum of these triangles, one for every single event. In the case of poor sampling,

dx.doi.org/10.1021/ci200217w |J. Chem. Inf. Model. 2011, 51, 2007–2023

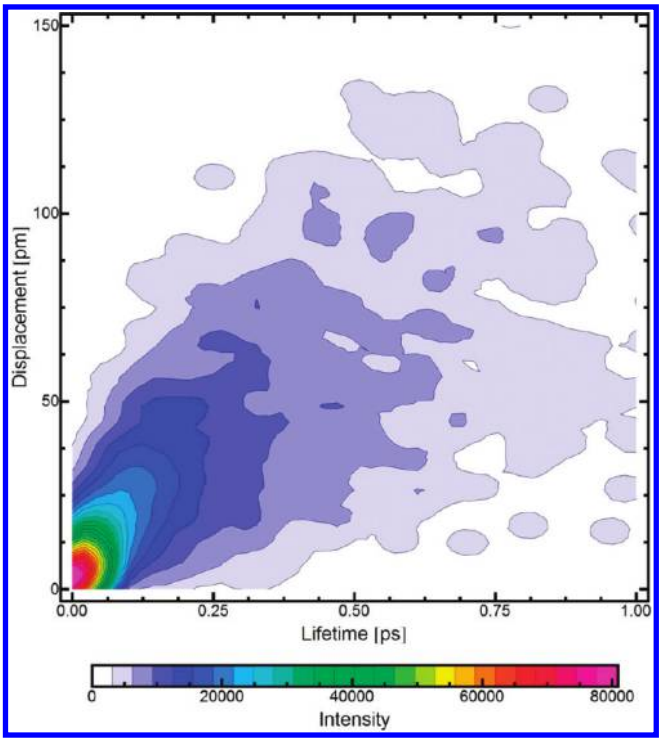


Figure 21. DLDisp function (continuous) for the 300 pm/130° criterion shown in Figure 11.

this can lead to visible edges in the curves, like, e.g., showing up in the dotted line of Figure 20 at $\tau \approx 1$ ps.

The description given above corresponds to "normal" DACFs, often denoted as *continuous* DACFs. In contrast to this, also *intermittend* DACFs can be created using TRAVIS. The definition for these intermittend functions is almost the same, with the difference that interruptions of the criteria fulfillment are now allowed up to a specified duration. In other words, dissociation and recombination of the aggregates is allowed, as long as the duration of this process does not exceed the given time criterion.

In Figure 20, DACFs between $[C_2C_1Im]^+$ and $[SCN]^-$ corresponding to the four different criteria, which are shown in Figure 11 and also described there, are shown. The coloring of the curves corresponds to the used criteria. Solid lines depict continuous DACFs, and dashed lines represent intermittend DACFs with a reformation threshold of 1 ps. DACFs are very sensitive in regard to the chosen dimer definition. Please note that, in spite of being based on strongly differing criteria, the red and the green curve show almost the same behavior. This seems to be just a case of coincidence.

Dimer Lifetime/Displacement Functions. If not only the lifetime of a dimer but also the distance it travels while fulfilling the criteria is of interest, a dimer lifetime/displacement function (DLDisp) can be of great use. It is a three-dimensional plot with one axis representing the lifetime of the dimer, as explained for the DLDF above (see Figure 19 on the right). The other axis depicts the distance that this dimer (e.g., the center of mass of the contributing molecules) did travel from its formation up to its cleavage. This kind of function was already published before.[52]

As an example, a DLDisp function between $[C_2C_1Im]^+$ and $[SCN]^-$ based on the 300 pm/130° criterion is shown in Figure 21 (only the continuous function, no reformation allowed here). It can be seen that some $[C_2C_1Im]^+$ $[SCN]^-$ dimers travel

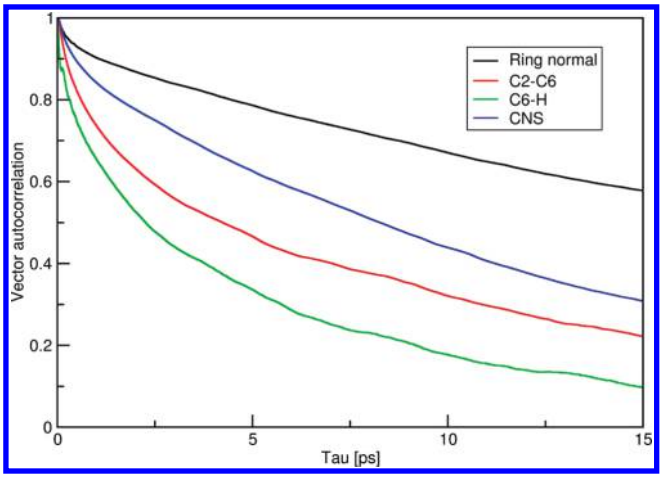


Figure 22. Vector reorientation dynamics for several vectors in $[C_2C_1Im]^+$ and $[SCN]^-$.

up to 150 pm during time intervals up to 1 ps while fulfilling the criterion. But the majority of the dimers stay only together for approximately 25 pm and 0.05 ps, which is, e.g., less than a bond length and less than a typical chain reorientation motion.

Vector Reorientation Dynamics. For any vector $\vec{a}$ in the trajectory (defined, e.g., by two points, velocity vector or dipole vector), vector reorientation dynamics analyses can be performed. The resulting functions are the autocorrelation of the vector, which is defined as the normalized sum over the dot product between the vector at some time t and the same vector at some later time $t + \tau$ for all starting times t. The definition of the vector reorientation dynamics function is given in eq 5, where T denotes the total simulation time. This kind of function is very useful for comparing MD simulations with experiments,[13] as reorientational correlation times can be obtained by NMR experiments[53]

$$VRD(\tau) = N \cdot \langle \sum_{t=0}^{T-\tau} \vec{a}_i(t) \cdot \vec{a}_i(t+\tau) \rangle_i \quad (5)$$

These functions are similar to the DACFs described above. They also start at value 1, tend to 0 for increasing t, and give information about how fast a vector in the system changes its direction. Some vector reorientation functions are shown in Figure 22. The black curve depicts the reorientation of the $[C_2C_1Im]^+$ ring normal vector. The blue curve corresponds to the reorientation of the $[SCN]^-$ S−N vector. The red and green curve give information of the reorientation of the bond vectors C2−C6 and C6−H in $[C_2C_1Im]^+$, indicating that the movement of the side chains is faster than that of the $[C_2C_1Im]^+$ ring.

Mean Square Displacement. When analyzing the dynamics of a simulated system, a popular and widely used method is to calculate the mean square displacement (MSD) of some kind of particles. It is a two-dimensional function and defined as the square of the average distance that a kind of particle has moved away from its starting point within the time interval τ (averaged over all particles of this kind and all starting times), see eq 6. These functions contain information on the *diffusivity* of the observed particles. The steeper they rise to higher values, the faster the observed particle is diffusing within the system. If the system is solid, the MSD saturates to a finite value, while if the system is liquid or gaseous, the MSD grows rather linearly with time. The slope of the MSD is the diffusion coefficient D of

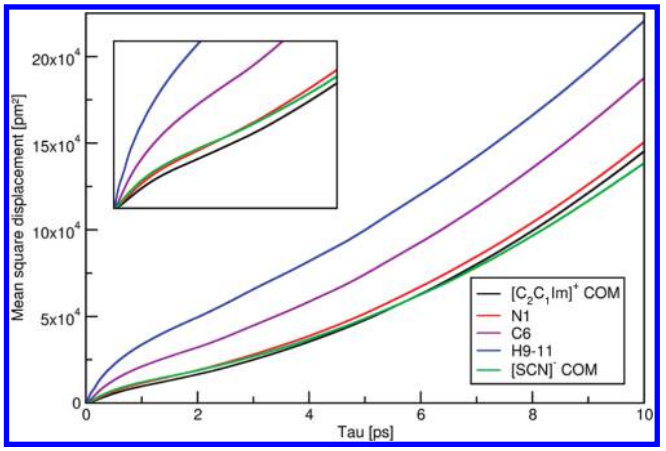


Figure 23. MSD plots for several atoms and centers of mass (COM) in $[C_2C_1Im]^+$ and $[SCN]^-$.

the observed particle,[11] see eq 7. TRAVIS also allows to calculate the components of the mean square displacement parallel and perpendicular to some vector $\vec{a}$ defined in the molecule, see eq 8 and eq 9. This is important, e.g., to make good estimates for diffusion coefficients of molecules from MD simulation.[54] Mean square displacements are an important quantity, as they can be used to validate simulations by means of measured values, e.g. from PFG NMR experiments.[55]

Figure 23 shows the MSDs for some particles. As expected, light atoms like hydrogen have a large mean square displacement, whereas heavier atoms and whole molecules move over shorter distances in the same time. Interesting is the fact that for small values of τ, $[SCN]^-$ (green curve) seems to move faster than $[C_2C_1Im]^+$ (red curve), but for larger τ, $[SCN]^-$ is moving slower (see inlay)

$$MSD(\tau) = \langle | \vec{r}_i(t+\tau) - \vec{r}_i(t)|^2 \rangle_{t,i} \tag{6}$$

$$D = \lim_{\tau \to \infty} \frac{1}{6\tau} \langle | \vec{r}_i(t+\tau) - \vec{r}_i(t)|^2 \rangle_{t,i} \tag{7}$$

$$MSD_{\parallel \vec{a}}(\tau) = \left\langle \left| \left(\vec{r}_i(t+\tau) - \vec{r}_i(t) \right) \cdot \frac{\vec{a}(t)}{|\vec{a}(t)|} \right|^2 \right\rangle_{t,i} \tag{8}$$

$$MSD_{\perp \vec{a}}(\tau) = \left\langle \left| \left(\vec{r}_i(t+\tau) - \vec{r}_i(t) \right) - \left(\left(\vec{r}_i(t+\tau) - \vec{r}_i(t) \right) \cdot \frac{\vec{a}(t)}{|\vec{a}(t)|} \right) \cdot \frac{\vec{a}(t)}{|\vec{a}(t)|} \right|^2 \right\rangle_{t,i} \tag{9}$$

Van Hove Correlation Functions. Van Hove correlation functions (VHCFs) are a time-dependent generalization of RDFs. An additional time dimension τ is introduced into the RDF. For $\tau = 0$, the function is identical to the "normal" RDF between $[C_2C_1Im]^+$ H1 and $[SCN]^-$ N. In cases with $\tau > 0$, the observation is performed between the position of H1 at time 0 and the position of N at time τ, see eq 10 (δ is explained in eq 1). In other words, H1 is allowed to diffuse away during the time interval τ, but this is not taken into account - the initial position of H1 is used, although H1 is no longer located there. This leads to the broadening of the RDF peaks with increasing τ. For very large values of τ, the function takes the value 1 for all distances r. An example of a VHCF between $[C_2C_1Im]^+$ H1 and $[SCN]^-$ N is shown in Figure 24.

One important point about VHCFs is that they form an intersection between simulations and experiments, as they can be determined experimentally.[56] The result of neutron scattering experiments[57] is a three-dimensional plot with frequency and reciprocal distance on its axes. This plot is the dual Fourier transform of the VHCF

$$VHCF_{ab}(\tau, r) = \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_b} \langle \delta(r - | \vec{r}_i(t+\tau) - \vec{r}_j(t)|) \rangle_t \tag{10}$$

With TRAVIS, it is also possible to create *auto*-VHCFs between a certain particle and itself. The distance between the initial position of the particle at time 0 and the new position of the particle at time τ is then observed. This is a kind of generalized mean square displacement, as it contains not only the average distance after time τ but actually the complete distance distribution. In Figure 25, a VHCF with $[SCN]^-$ center of mass observing itself is shown. Please note that the axes in the right part of the plot are swapped to strike out the similarity to the MSD (see Figure 23). The MSD plot shows the square of the displacement, whereas the auto-VHCF contains the nonsquared displacement. Obviously, it is not possible to draw a function of this kind for $\tau = 0$, as this function would possess a peak of infinite height at $r = 0$ per definition.

Power Spectra/Infrared Spectra. Another important part of trajectory analysis is to compute spectra of different kinds. Spectra are readily available by experiments and probably one of the best ways of validating performed simulations. As infrared (IR) spectra are of outstanding dispersiveness and available for almost every substance, an intuitive choice would be to calculate IR spectra from trajectories. But as an IR spectrum observes the change in the dipole vectors of the molecules, time-dependent dipole information is required to calculate one.[58] If this information is supplied (e.g., in the form of time-dependent partial charges or of maximally localized wannier centers[37,38]), TRAVIS can compute IR spectra from the trajectory. But as this information cannot be obtained from classical MD runs, the calculation of IR spectra is usually impractical (it is, however, possible to assign fixed partial charges to the atoms and calculate IR spectra from the resulting dipole moments, but these spectra will be of very poor quality).

A more general kind of spectrum that can be computed for every kind of time-dependent trajectory is the power spectrum. Power spectra are defined as the Fourier transform of the autocorrelation function of the particle's velocities.[42] Power spectra contain all peaks that can be found in the IR spectrum, and even more - they can be imagined as the sum of the IR spectrum, the Raman spectrum and all motions that are neither IR nor Raman active. In Figure 26, the power spectra for $[C_2C_1Im]^+$ and $[SCN]^-$ are shown. TRAVIS allows the user to utilize sophisticated methods for obtaining neat spectra, see section Smoothing and Optimized Algorithms.

One great advantage of simulated spectra is the possibility to assign peaks to specific compounds present in the sample - which is, of course, not possible by performing a single experiment. But not only can the peaks be assigned to $[C_2C_1Im]^+$ and $[SCN]^-$ (like shown in Figure 26), also can

dx.doi.org/10.1021/ci200217w |J. Chem. Inf. Model. 2011, 51, 2007–2023

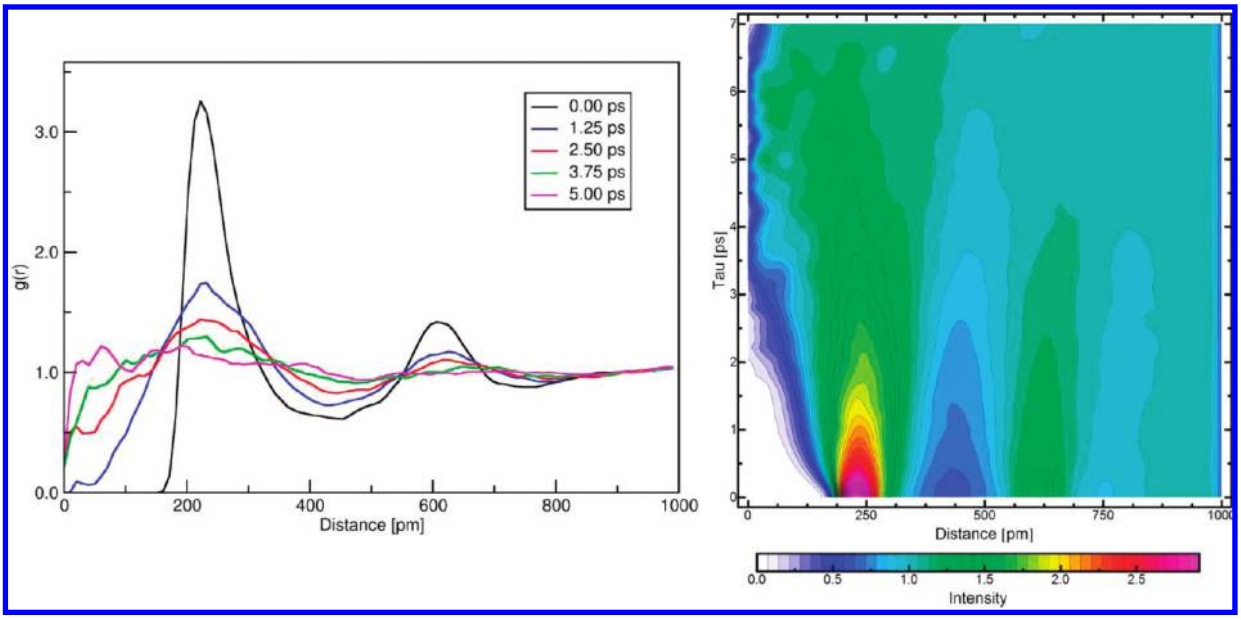


Figure 24. Van Hove correlation function of H1 observing N (left: 2D slices; right: 3D plot).

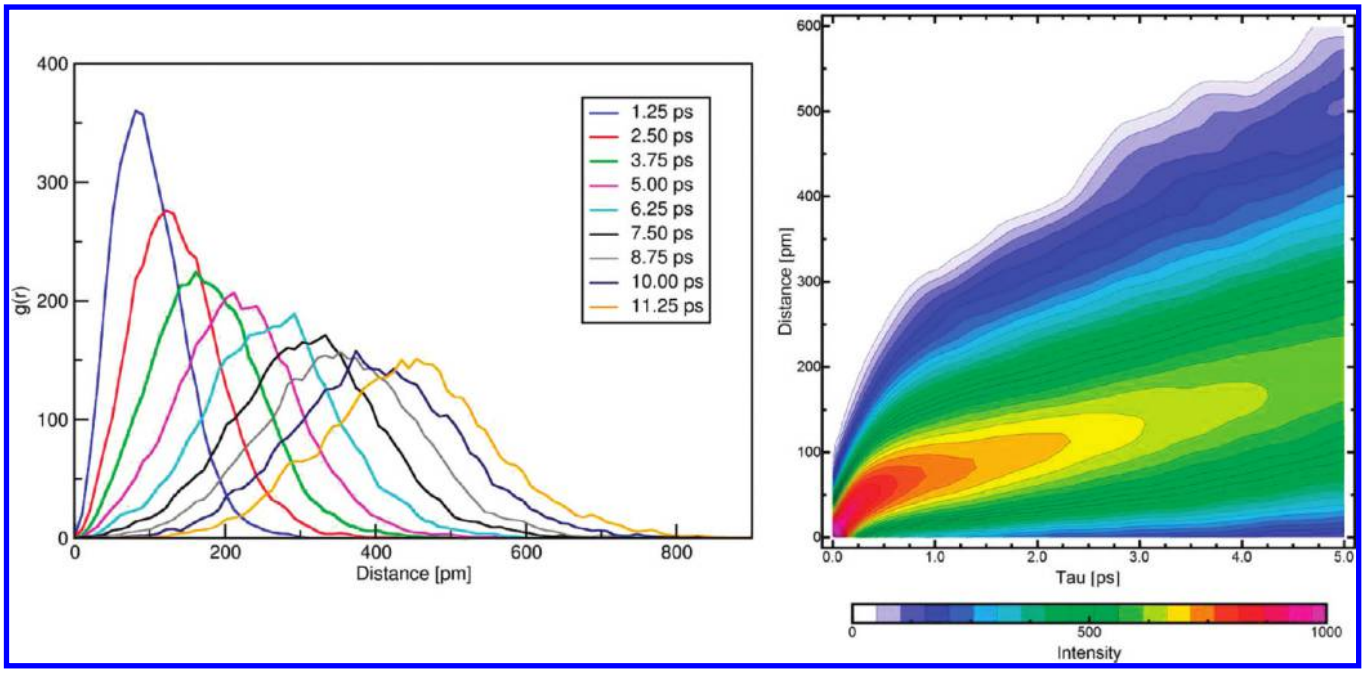


Figure 25. Van Hove correlation function of the $[SCN]^-$ center of mass observing itself (left = 2D slices; right = 3D plot - note the swapped axes).

the peaks be assigned to specific motions in the molecules. In Figure 27 on the left, the spectra of some C−H bonds in the $[C_2C_1Im]^+$ cation are shown. The plot on the right side of Figure 27 contains the spectra of some C−C and C−N bonds in $[C_2C_1Im]^+$. These spectra only take into account bond-stretching modes. Angular vibration modes are left out here.

TECHNICAL DETAILS

Source Code and Compiling Requirements. TRAVIS is written in C++, strongly relying on the powerful concept of object-oriented programming. The functionality is encapsulated into 80 different classes, collecting similar code segments and separating disjunct ones from each other. The program is distributed as open-source freeware, licensed under the GNU general public license (version 3).[59] The source code as well as some precompiled executable files and the user manual can be downloaded from http://www.uni-leipzig.de/~travis/.

TRAVIS is proud of not requiring any external library. The only demand for building TRAVIS from source code is a working C++ compiler. Resulting from this fact, there is in principle no limitation on hardware platforms and operating systems TRAVIS will run on. The development and testing has been performed

using different types of Microsoft Windows and Linux operating systems as well as several C++ compilers (e.g., g++, Intel C++ Compiler, Microsoft Visual C++, Bloodshed Dev-C++).

Required External Programs. Although TRAVIS bears the predicate "visualizer" in its title, it does not produce any graphics files itself. What it produces are close precursors to graphics files. Besides the data to be visualized, these precursors already contain design information and can be transformed into "publication grade" figures with almost no additional effort. This step is performed by some external programs, which are an integral part of the trajectory analysis pipeline of TRAVIS. Those programs will be presented in this subsection.

Please note, that in addition to the output of the specialized input files for the visualization programs, the raw data is written out in plain text format to enable further processing by different programs or other purposes.

- For the rendering of two-dimensional plots (like shown in Figure 3 to 5,9,12,20,22 to 25, and 27), TRAVIS writes AGR input files for the **Grace** plotting package,[60] already containing design information, e.g., the axis labels and the frame style.
- The spatial distributions (like the ones in Figure 6 to 8 and 18) are created with **VMD**.[31] To store spatial densities, TRAVIS writes files in gOpenMol PLT and Gaussian CUBE format, which can be, e.g., directly loaded into VMD. The high-quality rendering of these pictures was performed with the **POV-Ray** rendering package,[61] which can be employed by VMD.
- Three-dimensional distribution functions like shown in Figures 11, 13, 14, 21, 24, and 25 have been obtained using **Wolfram Mathematica**.[62] TRAVIS directly outputs Mathematica notebook files, which can be opened with one click, and rendered with a second click directly into "publication grade" images. The files written by TRAVIS contain a sophisticated framework for plotting 3D graphs, enabling to adjust many parameters as well as powerful features like zooming into the plot or direct smoothing of the data if desired.
- The Temporal Development Overlay plot shown in Figure 16 was rendered with **PyMOL**.[51] TRAVIS creates a custom PyMOL startup script, which automatically loads the different overlay geometries into PyMOL and sets the colors for all atoms as well as nice rendering options.
  The molecule drawings in Figures 1, 10, 11, 13, and 14 have also been created with PyMOL, but they do not stem directly from the output of the TRAVIS program package and required some additional processing with an image manipulation program.
- The snapshot of the simulation box in Figure 1 on the right was rendered using the freeware program **QuteMol**.[30]

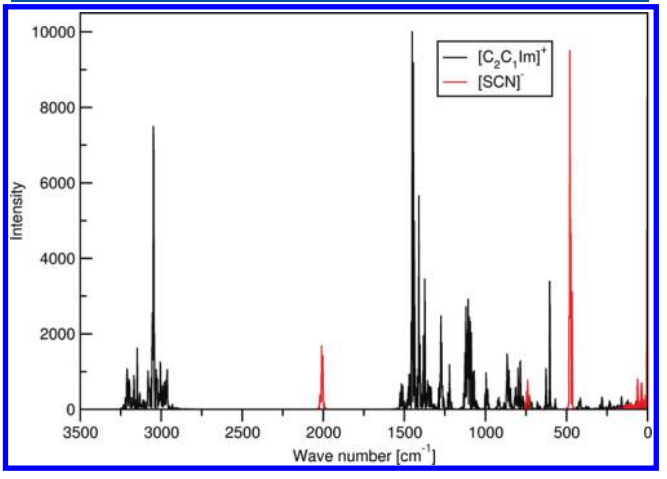


Figure 26. Power spectra (Fourier transform of velocity autocorrelation) for $[C_2C_1Im]^+$ and $[SCN]^-$.

CONCLUSION

We presented the TRAVIS program package, which is able to perform a variety of analyses with trajectories obtained from MC and MD simulations; many of them were discussed in this article. Some of the analyses, like e.g. radial distribution functions, spatial distribution functions, and mean square displacements (all three discussed above), are very common and can be regarded as ubiquitous standard trajectory analyses today. Others were enhanced by new features to offer additional flexibility, like the three-vector definition of the dihedral angle (see Figure 5 on the left). To our knowledge, some of the analyses even appear in the literature for the first time. This accounts, e.g., for some of the combined distribution functions (CDFs), which offer a very powerful tool for investigating the structure of simulated systems by combination of scalar quantities like distances, angles, or dipole moments. As these scalar quantities can be chosen in many different ways, there is a manifold of different CDFs arising from these combinations, that can be applied to every system. A CDF bears more information than just the distribution functions

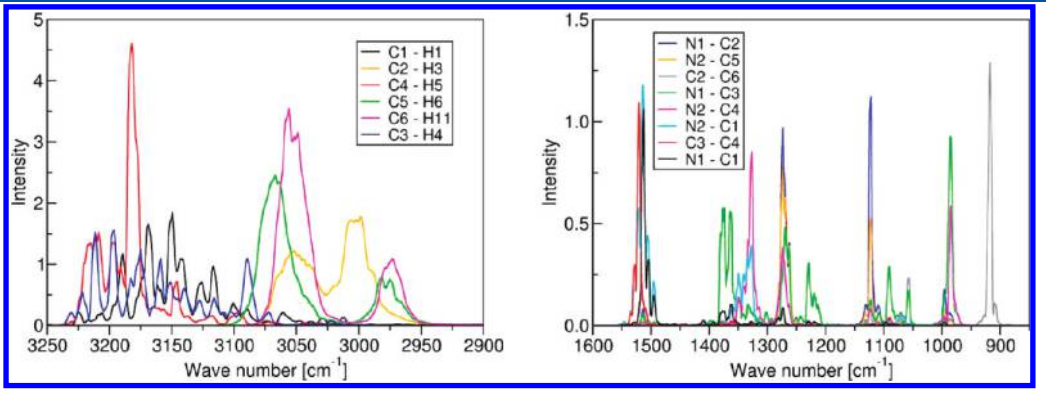


Figure 27. Power spectra of different C−H bond stretching motions in $[C_2C_1Im]^+$ (left) and power spectra of different C−C and C−N bond stretching motions in $[C_2C_1Im]^+$ (right).

dx.doi.org/10.1021/ci200217w |J. Chem. Inf. Model. 2011, 51, 2007–2023

of the scalar base quantities. Because of this, the creation of CDFs is only possible with a program package which unites all these scalar analyses in its scope.

Apart from the variety of analyses, TRAVIS offers some general features that are very helpful for analyzing trajectories (see general features section). Such are, e.g., automatic molecule recognition, topological atom labeling, support for periodic boundary conditions with changing cell vectors, definition of new virtual atoms at any position in space, and calculation of dipole vectors and dipole moments. The equitable binning method used in TRAVIS reduces the noise and enhances the visual quality of the created one- and multidimensional histograms. To our knowledge, equitable binning was not applied to problems of trajectory analysis and visualization before.

TRAVIS is open-source freeware. The source code as well as some precompiled executables and the user manual can be downloaded from http://www.uni-leipzig.de/~travis. The development of TRAVIS is an ongoing process. New features will be implemented from time to time.

■ ASSOCIATED CONTENT

Supporting Information. A movie and text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: bkirchner@uni-leipzig.de.

■ ACKNOWLEDGMENT

This work was supported by the DFG, in particular by the projects SPP 1191, KI-768/4-1, KI-768/4-2, KI-768/5-2, and KI-768/5-3. Support from the DFG within the Graduate School of Excellence Building with Molecules and Nano-objects (BuildMoNa), the ESF, and the IRTG "Diffusion in Porous Materials" are gratefully acknowledged.

■ REFERENCES

(1) Huber, H.; Dyson, A. J.; Kirchner, B. Calculation of bulk properties of liquids and supercritical fluids from pure theory. *Chem. Soc. Rev.* **1999**, *28*, 121–133.

(2) Maginn, E. J.; Elliott, J. R. Historical Perspective and Current Outlook for Molecular Dynamics As a Chemical Engineering Tool. *Ind. Eng. Chem. Res.* **2010**, *49*, 3059–3078.

(3) Maginn, E. J. Molecular simulation of ionic liquids: current status and future opportunities. *J. Phys.: Condens. Matter* **2009**, *37*, 373101.

(4) Brüssel, M.; Zahn, S.; Hey-Hawkins, E.; Kirchner, B. Theoretical Investigation of Solvent Effects and Complex Systems: Toward the calculations of bioinorganic systems from ab initio molecular dynamics simulations and static quantum chemistry. *Adv. Inorg. Chem.* **2010**, *62*, 111–142.

(5) Senn, H. M.; Thiel, W. QM/MM methods for biological systems. *Top. Curr. Chem.* **2007**, *268*, 173–290.

(6) Yockel, S.; Schatz, G. C. Dynamic QM/MM: A Hybrid Approach to Simulating Gas-Liquid Interactions. *Top. Curr. Chem.* **2011**ahead of print.

(7) Delle Site, L.; Holm, C.; van derVegt, N. F. A. Multiscale Approaches and Perspectives to Modeling Aqueous Electrolytes and Polyelectrolytes. *Top. Curr. Chem.* **2011**ahead of print.

(8) Jaramillo-Botero, A.; Nielsen, R.; Abrol, R.; Su, J.; Pascal, T.; Mueller, J.; Goddard, W. A., III First-Principles-Based Multiscale, Multiparadigm Molecular Mechanics and Dynamics Methods for Describing Complex Chemical Processes. *Top. Curr. Chem.* **2011**ahead of print.

(9) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.

(10) Hansen, J. P.; McDonalds, I. R. *Theory of simple liquids*, 2nd ed.; Academic Press: New York, 1986.

(11) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1989.

(12) Thar, J.; Brehm, M.; Seitsonen, A. P.; Kirchner, B. Unexpected Hydrogen Bond Dynamics in Imidazolium-Based Ionic Liquids. *J. Phys. Chem. B* **2009**, *113*, 15129–15132.

(13) Kohagen, M.; Brehm, M.; Thar, J.; Zhao, W.; Müller-Plathe, F.; Kirchner, B. Performance of Quantum Chemically Derived Charges and Persistence of Ion Cages in Ionic Liquids. A Molecular Dynamics Simulations Study of 1-n-Butyl-3-methylimidazolium Bromide. *J. Phys. Chem. B* **2011**, *115*, 693–702.

(14) Brüssel, M.; Brehm, M.; Kirchner, B. Ab initio molecular dynamics simulations of a binary system of ionic liquids. *Phys. Chem. Chem. Phys.* **2011**, DOI:10.1039/C1CP21550G

(15) Born, M.; Von Karman, T. Über Schwingungen in Raumgittern. *Physik. Z.* **1912**, *13*, 297–309.

(16) CP2k. 2011. http://cp2k.berlios.de/index.html (accessed Mar 21, 2011).

(17) 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.

(18) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365–4369.

(19) Becke, A. Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(20) Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **1988**, *37*, 785–789.

(21) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* **1964**, *136*, 864.

(22) Kohn, W.; Sham, L. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, 1133.

(23) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(24) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.

(25) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(26) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.

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

(28) Nose, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(29) Martyna, G.; Klein, M.; Tuckerman, M. Nose-Hoover chains - The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(30) Tarini, M.; Cignoni, P.; Montani, C. Ambient Occlusion and Edge Cueing for Enhancing Real Time Molecular Visualization. *IEEE T. Vis. Comput. Gr.* **2006**, *12*, 1237–1244.

(31) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(32) Frascoli, F.; Todd, B. D.; Searles, D. J. Boundary condition independence of molecular dynamics simulations of planar elongational flow. *Phys. Rev. E* **2007**, *75*, 066702.

(33) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

(34) Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W.-Q.; Goddard, W. A., III Development of the ReaxFF Reactive Force Field for Describing Transition Metal Catalyzed Reactions, with Application to the Initial Stages of the Catalytic Formation of Carbon Nanotubes. *J. Phys. Chem. A* **2005**, *109*, 493–499.
(35) Shelley, C. A.; Munk, M. E. Computer Perception of Topological Symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 110–113.
(36) Shelley, C. A. Heuristic Approach for Displaying Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 61–65.
(37) Wannier, G. H. The Structure of Electronic Excitation Levels in Insulating Crystals. *Phys. Rev.* **1937**, *52*, 191–197.
(38) Marzari, N.; Vanderbilt, D. Maximally localized generalized-Wannier functions for composite energy bands. *Phys. Rev. B* **1997**, *56*, 12847–12865.
(39) Futrelle, R. P.; McGinty, D. J. Calculation of spectra and correlation functions from molecular dynamics data using the fast fourier transform. *Chem. Phys. Lett.* **1971**, *12*, 285–287.
(40) Kestemont, E.; Van Craen, J. On the computation of correlation functions in molecular dynamics experiments. *J. Comput. Phys.* **1976**, *22*, 451–458.
(41) Smith, W. An introduction to the discrete Fourier transform. *CCP5 Quarterly* **1982**, *5*, 34–41.
(42) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. In *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: 1993; Chapter Power Spectra Estimation Using the FFT, pp 542−551.
(43) Harris, F. J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **1978**, *66*, 51–83.
(44) Nuttall, A. Some windows with very good sidelobe behavior. *IEEE Trans. Acoust., Speech, Signal Process.* **1981**, *29*, 84–91.
(45) Head-Gordon, T.; Hura, G. Water Structure from Scattering Experiments and Simulation. *Chem. Rev.* **2002**, *102*, 2651–2670.
(46) Kroon, J.; Kanters, J. A. Non-linearity of hydrogen bonds in molecular crystals. *Nature* **1974**, *248*, 667–669.
(47) Svishchev, I. M.; Kusalik, P. G. Structure in liquid water - A study of spatial-distribution functions. *J. Chem. Phys.* **1993**, *99*, 3049–3058.
(48) Wendler, K.; Thar, J.; Zahn, S.; Kirchner, B. Estimating the Hydrogen Bond Energy. *J. Phys. Chem. A* **2010**, *114*, 9529–9536.
(49) Lehmann, S. B. C.; Roatsch, M.; Schöppke, M.; Kirchner, B. On the physical origin of the cation-anion intermediate bond in ionic liquids Part I. Placing a (weak) hydrogen bond between two charges. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7473–7486.
(50) Kumar, R.; Schmidt, J. R.; Skinner, J. L. Hydrogen bonding definitions and dynamics in liquid water. *J. Chem. Phys.* **2007**, *126*, 204107.
(51) The PyMOL Molecular Graphics System, Version 1.3. 2010. http://www.pymol.org/ (accessed Mar 21, 2011).
(52) Zhao, W.; Leroy, F.; Heggen, B.; Zahn, S.; Kirchner, B.; Balasubramanian, S.; Müller-Plathe, F. Are There Stable Ion-Pairs in Room-Temperature Ionic Liquids? Molecular Dynamics Simulations of 1-n-Butyl-3-methylimidazolium Hexafluorophosphate. *J. Am. Chem. Soc.* **2009**, *131*, 15825–15833.
(53) Antony, J. H.; Mertens, D.; Dölle, A.; Wasserscheid, R.; Carper, W. R. Molecular reorientational dynamics of the neat ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate by measurement of C-13 nuclear magnetic relaxation data. *ChemPhysChem* **2003**, *4*, 588–594.
(54) Tildesley, D. J.; Madden, P. A. Time correlation functions for a model of liquid carbon disulphide. *Mol. Phys.* **1983**, *48*, 129–152.
(55) Vasenkov, S.; Bohlmann, W.; Galvosas, P.; Geier, O.; Liu, H.; Kärger, J. PFG NMR study of diffusion in MFI-type zeolites: Evidence of the existence of intracrystalline transport barriers. *J. Phys. Chem. B* **2001**, *105*, 5922–5927.
(56) Dahlborgt, U.; Gudowskit, W.; Davidovic, M. Van Hove correlation functions from coherent neutron inelastic scattering. *J. Phys.: Condens. Matter* **1989**, *1*, 6173–6179.
(57) Embs, J. P.; Juranyi, F.; Hempelmann, R. Introduction to Quasielastic Neutron Scattering. *Z. Phys. Chem.* **2010**, *224*, 5–32.
(58) Iftimie, R.; Tuckerman, M. E. Decomposing total IR spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized Wannier orbitals. *J. Chem. Phys.* **2005**, *122*, 214508.
(59) GNU General Public License. http://www.gnu.org/licenses/ (accessed Mar 21, 2011).
(60) Grace. 1996−2008. http://plasma-gate.weizmann.ac.il/Grace (accessed Mar 21, 2011).
(61) Persistence of Vision Raytracer (Version 3.6). 2004. http://www.povray.org (accessed Mar 21, 2011).
(62) Mathematica, Version 8.0. 2010. www.wolfram.com/mathematica/ (accessed Mar 21, 2011).