Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics v1.5

Edward J. Maginn[1*,†§], **Richard A. Messerly**[2†††], **Daniel J. Carlson**[3†‡‡], **Daniel R. Roe**[4†¶], **J. Richard Elliott**[5†**]

[1]Department of Chemical and Biomolecular Engineering, The University of Notre Dame; [2]Thermodynamics Research Center, National Institute of Standards and Technology; [3]Chemical Engineering Department, Brigham Young University; [4]Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health; [5]Department of Chemical and Biomolecular Engineering, The University of Akron

This LiveCoMS document is maintained online on GitHub at https://github.com/ejmaginn/TransportCheckList; to provide feedback, suggestions, or help improve it, please visit the GitHub repository and participate via the issue tracker.

Contribution of the National Institute of Standards and Technology, not subject to US copyright.

This version dated April 26, 2018

Abstract

The ability to predict transport properties (i.e. diffusivity, viscosity, conductivity) is one of the primary benefits of molecular simulation. Although most studies focus on the accuracy of the simulation output compared to experimental data, such a comparison primarily tests the adequacy of the force field (i.e. the model). By contrast, the reliability of different simulation methodologies for predicting transport properties is the focus of this manuscript. Unfortunately, obtaining reproducible estimates of transport properties from molecular simulation is not as straightforward as static properties. Therefore, this manuscript discusses the best practices that should be followed to ensure that the simulation output is reliable, i.e. is a valid representation of the force field implemented. We also discuss procedures to use so that the results are reproducible (i.e. can be obtained by other researchers following the same methods and procedures).

There are two classes by which transport properties are predicted: equilibrium molecular dynamics (EMD) and non-equilibrium molecular dynamics (NEMD). This manuscript presents the best practices for EMD, leaving NEMD for a future publication. As self-diffusivity and shear viscosity are the most prevalent transport properties found in the literature, the discussion will also be limited to these properties with the expectation that future publications will discuss best practices for thermal conductivity, ionic conductivity, and transport diffusivity.

***For correspondence:**
ed@nd.edu (EM); Roe'semail (DR); Elliott'semail (JRE); richard.messerly@nist.gov (RAM); daniel.j.carlson@byu.edu (DC)

[†]These authors contributed equally to this work
[‡]These authors also contributed to this work

1 Introduction

Transport properties describe the rates at which mass, momentum, heat or charge move through a given substance. They involve mean squared displacements (MSDs) of molecules as the system evolves dynamically. In general, these properties can be computed by equilibrium molecular dynamics (EMD) or by non-equilibrium molecular dynamics (NEMD) methods. The EMD methods involve post-processing of a standard molecular dynamics (MD) trajectory while NEMD methods require modifications of the underlying equations of motion and/or boundary conditions of the system. Therefore, one advantage of EMD is that multiple transport properties can be obtained from a single simulation, whereas NEMD requires a separate simulation for each transport property of interest.

Many codes such as LAMMPS [21] and GROMACS [16] have analysis tools that automatically estimate transport properties from an EMD or NEMD simulation, but there are often insufficient checks as to whether the actual underlying simulations are adequate for making these estimates. For this reason, we strongly discourage using these analysis tools as a "black box." Following best practices is imperative to ensure that meaningful predictions are obtained. The purpose of this document is to improve the quality of published results and to reduce the time required for a novice in the field to obtain meaningful and reliable results.

In addition to the present manuscript, we highly recommend reviewing this list of existing resources:

1. Text books:
   (a) Ref. [5], pages 73-79, 274-281, and 292-296
   (b) Ref. [15], pages 87-90 and 509-523
   (c) Ref. [22], pages 374-382
2. Class notes
   (a) Ref. [1]
   (b) Ref. [2]
   (c) Ref. [3]
   (d) Ref. [4]
3. Published articles
   (a) Ref. [9]
   (b) Ref. [18]
   (c) Ref. [27], pages 13139-13140
   (d) Ref. [31]
4. Software manuals
   (a) Ref. [16]
   (b) Ref. [21]

Most text books and class notes provide a thorough discussion of EMD/NEMD theory with little discussion of practical considerations. Review articles tend to focus on the numerical advantages and disadvantages of different methods but assume that the reader already understands the subtleties of implementing each method. Furthermore, although software manuals describe some of the theory and implementation of these methods in their respective environments, the documentation is typically insufficient for someone not familiar with best practices for estimating transport properties. This document supplements the existing literature by providing a succinct checklist and discussing common pitfalls. We also provide some suggestions and recommendations based on our own experience, but ultimately it is up to the individual researcher to test and validate their methods.

2 Equilibrium Molecular Dynamics (EMD) for Estimating Transport Properties

It is most convenient to consider compiling the transport properties as an implicit part of any equilibrium MD simulation. The added computational overhead is relatively small, especially for the self-diffusivity. The main caveat is that longer simulations than normal may be required to achieve reasonable averages.

The general formula for computing a transport property via an EMD simulation is given as

$$\gamma = \int_0^\infty dt \langle \dot{\xi}(t)\dot{\xi}(0)\rangle \tag{1}$$

where γ is the transport coefficient (within a multiplicative constant), ξ is the perturbation in the Hamiltonian associated with the particular transport property under consideration and $\dot{\xi}$ signifies a time derivative. Integrals of the form given by Equation 1 are known as "Green-Kubo" integrals. It is easy to show that an integrated form of Equation 1 results in an equivalent expression for γ known as the "Einstein" formula

$$\gamma = \lim_{t\to\infty} \frac{\langle (\xi(t) - \xi(0))^2 \rangle}{2t} = \frac{1}{2} \lim_{t\to\infty} \frac{d}{dt} \langle (\xi(t) - \xi(0))^2 \rangle \tag{2}$$

where the derivative form is often preferred.

For self-diffusivity, ξ is the Cartesian atom position and the time correlation function, $\dot{\xi}$, in Equation 1 is of the molecular velocities. For the shear viscosity, the integral in Equation 1 is of the time correlation of the off-diagonal elements of the stress tensor. For the thermal conductivity the integral is over the energy current, and for the ionic conductivity the integral is over the ionic current. Table 1 provides the relevant equations for self-diffusivity (D) and shear viscosity (η), as these properties are the focus of this work.

Although both Equation 1 (Green-Kubo) and Equation 2 (Einstein) are theoretically rigorous, in practice one method is often preferred depending on the property being estimated.

Table 1. Equilibrium molecular dynamics equations.

Property	γ	ξ	Green-Kubo (Equation 1)	Einstein (Equation 2)
Self-diffusivity	D	r	$\frac{1}{3}\int_0^\infty dt \left\langle \frac{1}{N}\sum_{i=1}^{N} v_{\alpha,i}(t)v_{\alpha,i}(0)\right\rangle_{t_0}$	$\frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\left\langle \frac{1}{N}\sum_{i=1}^{N}\lvert r_i(t)-r_i(0)\rvert^2\right\rangle_{t_0}$
Shear viscosity	η	$r_\alpha v_\beta$	$\frac{V}{k_bT}\int_0^\infty dt \left\langle \tau_{\alpha,\beta}(t)\tau_{\alpha,\beta}(0)\right\rangle_{t_0}$	$\frac{V}{2k_bT}\lim_{t\to\infty}\frac{d}{dt}\left\langle \left(\int_0^t dt'\tau_{\alpha,\beta}(t')\right)^2\right\rangle_{t_0}$

$\tau_{\alpha,\beta}(t) = \frac{1}{V}\sum_{i=1}^{N}\left(mv_{\alpha,i}(t)v_{\beta,i}(t) + r_{\alpha,i}(t)f_{\beta,i}(t)\right), \alpha \neq \beta$
$\alpha, \beta = x, y,$ or z Cartesian coordinates of the atoms or molecule center of mass
$N =$ number of atoms or molecules (see Sec. 4.2.2)
$f_{\beta,i}$ is the force acting on particle i in direction β
$\langle\cdots\rangle_{t_0}$ denotes an average over time origins (see Sec. 4.2.1)

In the case of self-diffusivity, we recommend the Einstein (MSD) approach. In contrast, for shear viscosity we typically recommend Green-Kubo, although for some systems the Einstein approach may be preferable. As the simulation set-up and computational cost are essentially the same for the Green-Kubo and Einstein approaches, the primary difference is the post-simulation data analysis required. Precision and reproducibility of the estimated value are key factors for selecting between the Green-Kubo or Einstein methods. For this reason, we emphasize the importance of proper and clearly communicated data analysis and rigorous uncertainty quantification.

3 Checklist

This section provides an overview of the checklist items for each property (D and η) and method (Green-Kubo and Einstein). Detailed discussions for each checklist item are found in Sections 4-6.

4 General transport checklist items

4.1 General transport: Simulation set-up

4.1.1 Correct Ensemble

For a liquid solution, it is safest to run in the microcanonical (NVE, constant number of molecules, volume, energy) ensemble. This is because thermostats required to maintain constant temperature and barostats required to maintain constant pressure can interfere with the dynamics of the system, and thus the resulting transport properties can be skewed. However, it is most common to desire D and η at a specified temperature (T) and pressure (P). This requires performing a series of simulations in different ensembles:

1. NPT ensemble at desired T and P until equilibrium is well sampled
2. NVT ensemble where the volume is set such that the density is the average density computed from the NPT run
3. NVE ensemble where the final configuration of the NVT run is used as the initial configuration

The average pressure and temperature for the NVE production run are computed and should be close to (but not exactly the same) as the input P and T to the original NPT run. These average pressures and temperatures must be reported along with the self-diffusivity and viscosity.

Note that, although the best practice is to use the NVE ensemble (Steps 1-3), it is common to see values reported using the NPT (just Step 1) or NVT (Steps 1-2) ensemble. We strongly discourage the use of the NPT ensemble alone, because barostats (which alter positions through volume changes) greatly affect the dynamics of a system. In contrast, the NVT ensemble has been implemented successfully for transport property calculations and is quite common, especially for viscosity. For example, Fanourgakis et al. reported that the NVT and NVE ensembles provide nearly identical results for viscosity [13]. A study by Basconi and Shirts [6] reached a similar conclusion, and provides guidelines for how thermostats should be applied when computing transport properties. Therefore, we recommend using either the NVT or NVE ensemble, with NVE being preferred. If NVT simulations are used, we recommend that the papers above be consulted.

4.1.2 Replicate simulations

To smooth noise in the Green-Kubo integral or Einstein slope, we recommend generating independent replicate trajectories (i.e. different initial configurations or random seed to initialize velocities). The primary advantage of performing replicates as opposed to one longer simulation is the computational speed-up. Figure 1, borrowed from Ref. [28], demonstrates that an average of 10 short replicate simulations converges to the same value as a single long simulation. Since these replicates can be performed in parallel, the time required to obtain the result is reduced, although the CPU time may be the same or more.

CHECKLIST FOR COMPUTING SELF-DIFFUSIVITY WITH EINSTEIN EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 5.2.2.
    - ☐ Increase the output frequency. See Sec. 5.2.1.
  - ☐ Check for system size effects. See Sec. 5.1.2
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of D.
  - ☐ Improve precision by averaging over:
    - ☐ N molecules. See Sec. 4.2.1.
    - ☐ Three dimensions (xx, yy, zz). See Sec. 4.2.1.
    - ☐ Multiple replicate simulations. See Sec. 4.1.2.
  - ☐ Clearly communicate how D is obtained from Equation 2. See Secs. 4.2.2, 5.1.1, and 5.2.3.
  - ☐ Report the uncertainty in D:
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in D with respect to the time cut-off, etc. See Sec. 5.2.3.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Validation.** Compare your results with those from a reputable source. See Sec. 4.4.
- ☐ **Special topics.** Check if your system of interest requires some special considerations. See Sec. 5.4.

CHECKLIST FOR COMPUTING SELF-DIFFUSIVITY WITH GREEN-KUBO EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 5.3.2
    - ☐ Increase the output frequency. See Sec. 5.3.1.
  - ☐ Check for system size effects. See Sec. 5.1.2.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of D.
  - ☐ Improve precision by averaging over:
    - ☐ N molecules. See Sec. 4.2.1.
    - ☐ Three dimensions (xx, yy, zz). See Sec. 4.2.1.
    - ☐ Multiple replicate simulations. See Sec. 4.1.2.
  - ☐ Clearly communicate how D is obtained from Equation 1. See Secs. 4.2.2, 5.1.1, and 5.3.3.
  - ☐ Report the uncertainty in D:
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in D with respect to the time cut-off, etc. See Sec. 5.3.3.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Validation.** Compare your results with those from a reputable source. See Sec. 4.4.
- ☐ **Special topics.** Check if your system of interest requires unique considerations. See Sec. 5.4.

CHECKLIST FOR COMPUTING VISCOSITY WITH GREEN-KUBO EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 6.1.1.
    - ☐ Increase the output frequency. See Sec. 6.1.2.
  - ☐ Check for system size effects. See Sec. 6.1.3.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of η.
  - ☐ Improve precision by averaging over multiple:
    - ☐ Stress tensor elements (three off-diagonal or all six). See Sec. 6.1.4.
    - ☐ Replicate simulations. See Sec. 4.1.2 and 6.1.4.
  - ☐ Clearly communicate how η is obtained from Equation 1. See Secs. 4.2.2 and 6.2.1.
  - ☐ Report the uncertainty in η:
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in η with respect to the time cut-off, fitting model, etc. See Sec. 6.2.1.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Validation.** Compare your results with those from a reputable source. See Sec. 4.4.
- ☐ **Special topics.** Check if your system of interest requires some special considerations. See Sec. 6.4.

CHECKLIST FOR COMPUTING VISCOSITY WITH EINSTEIN EQUILIBRIUM APPROACH

- ☐ **Simulation set-up.** No amount of data analysis can compensate for a poorly designed experiment. It is imperative that the simulation sufficiently samples the relevant region of phase space.
  - ☐ Sample from the correct ensemble. See Sec. 4.1.1.
  - ☐ Increase the information extracted from simulation results.
    - ☐ Perform multiple replicate simulations. See Sec. 4.1.2.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 6.1.1.
    - ☐ Increase the output frequency. See Sec. 6.1.2.
  - ☐ Check for system size effects. See Sec. 6.1.3.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of η.
  - ☐ Improve precision by averaging over multiple:
    - ☐ Stress tensor elements (three off-diagonal or all six). See Sec. 6.1.4.
    - ☐ Replicate simulations. See Sec. 4.1.2 and 6.1.4.
  - ☐ Clearly communicate how η is obtained from Equation 2. See Secs. 4.2.2 and 6.3.1.
  - ☐ Report the uncertainty in η:
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis, i.e. variation in η with respect to the time cut-off, fitting model, etc. See Sec. 6.3.1.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Validation.** Compare your results with those from a reputable source. See Sec. 4.4.
- ☐ **Special topics.** Check if your system of interest requires some special considerations. See Sec. 6.4.

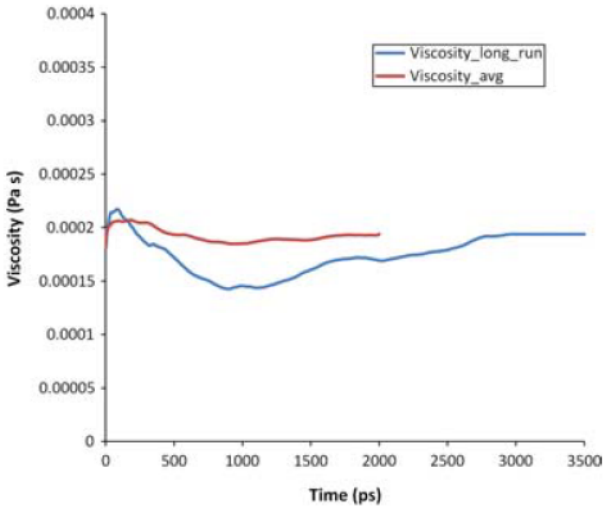

Figure 1. Green-Kubo viscosity plot. Copied from Figure 2 of Ref. [28]. Red curve represents the average viscosity over 10 independent 2 ns trajectories whereas the blue curve is obtained from a single 4 ns simulation. For further details, see Ref. [28]. **Note: When I went to Taylor and Francis to request re-use of this figure, they asked for 315 dollars. I think we need a substitute.**

The uncertainty is inversely proportional to the square root of the number of replicates (see Figure 7 of Ref. [34] and Figure 8 of Ref. [24]), so increasing the number of replicates is a simple, fast, and direct way to reduce the uncertainty. For example, note in Figure 1 the fluctuations in η are much smaller for the average of 10 replicates compared to that of a single longer simulation. As fluctuations in η are typically much larger than D, more replicate simulations are required for estimating viscosity (see Sec. 6.1.4).

In addition, replicate simulations are useful if a single simulation does not adequately sample phase space, i.e. is trapped in a local minimum or has slow dynamics. Furthermore, replicates can provide rigorous estimates of uncertainty (see Sec. 4.2.3).

Note that, although the best practice is to start each independent replicate at the NPT step, it is common to use the same density (NVT step) for each replicate. This approach is acceptable assuming that the authors provide the corresponding uncertainty in P (see Sec. 4.1).

4.2 General transport: Post-simulation analysis

4.2.1 Improved precision

In practice, several tricks-of-the-trade are employed to reduce fluctuations and, thereby, the standard deviation (σ). For self-diffusivity, it is a standard practice to average the mean-square-displacement or velocity autocorrelation function over all N molecules (see Table 1). For shear viscosity, it is not possible to average over the number of molecules because viscosity is a collective property that depends on the pressure/stress tensor of the system. For this reason, it is much easier to get precise diffusivity estimates than it is to get precise viscosity estimates; additional tactics are typically employed to improve the viscosity precision, namely, large amounts of replicate simulations.

The self-diffusivity is a tensor, and it is common practice in homogeneous systems to average the diagonal components, such that $D = \frac{1}{3}(D_{xx} + D_{yy} + D_{zz})$ where for example $D_{xx} = \frac{1}{2}\lim_{t\to\infty}\frac{d}{dt}\left\langle \frac{1}{N}\sum_{i=1}^{N}|x_i(t) - x_i(0)|^2 \right\rangle$. Since formally $D_{xx} = D_{yy} = D_{zz}$ for homogeneous systems, one can test the equivalence of the three terms as a check on a simulation and even to make a crude estimate of the uncertainty in D. In inhomogeneous systems, the diagonal terms will not necessarily be equivalent. Off diagonal terms should be zero, and we encourage the user to verify this.

For viscosity, the recommended practice is to use multiple components from the pressure/stress tensor. For example, although early studies only implemented a single off-diagonal component (typically xy), the common practice in recent studies is to use all three off-diagonal (xy, yz, zx) and sometimes three additional modified diagonal terms of the pressure/stress tensor (see Sec. 6.1.4).

Finally, for both self-diffusivity and shear viscosity it is common to average over multiple time origins (t_0). It is important that the difference between subsequent t_0 values (δt_0) be longer than the correlation time so that the different time intervals are independent.

4.2.2 Clear communication

Transport properties are estimated by integration of Equation 1 or calculating the slope of Equation 2 with respect to time. Both methods involve some judgment on the part of the user and results can vary depending on where the slope is taken (Einstein approach) and for how long the integral is carried out (Green-Kubo approach). Some recent work has suggested some guidelines for how to compute an objective estimate of the viscosity using the Green-Kubo approach [34]. Similar methods for estimating other transport properties from Equations 1 or 2 should be possible to develop.

As no single best practice can be recommended for the region over which the slope or integral is calculated, it is important to justify how this decision was made and then clearly communicate the approach used in any publication. Furthermore, it is critical to quantify the degree of variability in the estimated property that arises from assumptions in the data analysis, e.g. the time interval over which the Einstein slope is computed, etc. As post-simulation analysis is an essential step for estimating transport properties, we recommend

providing data analysis scripts as supporting information to improve future reproducibility.

4.2.3 Uncertainty quantification

Replicates can provide a rigorous uncertainty assessment. We recommend bootstrapping the uncertainties by randomly sampling which replicates are included in the data analysis procedure:

1. Randomly select (with replacement) a set of replicate simulations
2. Calculate the relevant average quantity from this random set, i.e. $\langle\dot{\xi}(t)\dot{\xi}(0)\rangle$ for Green-Kubo or $\langle(\xi(t)-\xi(0))^2\rangle$ for Einstein
3. Compute transport property (γ) from Equations 1 or 2
4. Repeat steps 1-3 thousands of times
5. Generate distribution of the estimated values for D or η
6. Determine lower and upper uncertainty bounds of D or η at desired confidence level, $1-\alpha$

The final step requires the probability density function (PDF, or alternatively the cumulative distribution function, CDF) for D or η. The bootstrapped distribution of D or η obtained in Step 5 is used to approximate the PDF, which is typically expressed as either a histogram or by fitting to a normal distribution. Solving for the lower and upper bounds of D or η can be performed in several different ways, but the two-sided tail approach is most common. With this approach, the lower and upper bounds correspond to the values that yield $\alpha/2\times100\%$ of the integrated PDF in the lower and upper tails. We recommend using $\alpha=0.05$, corresponding to a 95% confidence interval.

4.3 General transport: Common pitfalls

When simulating in the NVE ensemble, it is imperative that the integrator conserve energy. The most common method to check for energy conservation is to systematically adjust the time step and plot the energy versus time. The energy should show little to no drift over the timescale of the simulation. Haile [17] provides a detailed discussion of energy conservation and time step size (see Chapter 4.4 of his book). If constraints on bond lengths or angles are used, we also recommend checking to make sure that these constraints are maintained.

An important implicit assumption in Equations 1 and 2 is that the time over which these expressions are evaluated is much larger than the correlation time of the variable ξ. This assumption is often satisfied easily for simple liquids, where relaxation times are fast, but becomes problematical for systems with sluggish dynamics. Therefore, insufficient simulation time is a common pitfall in estimating transport properties. To avoid this pitfall, we recommend performing a series of progressively longer simulations to determine if the estimated values deviate significantly with increasing simulation time. Another way to test whether a simulation is long enough is to determine whether the molecules in the system explore a sufficiently diverse region of configuration space. This can be done by calculating the MSD of the molecules in the system and comparing this to either the radius of gyration of the largest molecule in the system (r_G) or the box length (L). If the square root of the MSD is larger than r_G (or better yet, is comparable to or larger than L), then the molecules have moved over long enough distances to sample a significant amount of configuration space.

4.4 General transport: Validation

Validation is an important step to demonstrate that the simulation set-up and post-simulation analysis are performed properly. One tool that can serve this purpose is the Standard Reference Simulation Website provided by the National Institute of Standards and Technology (NIST) [30]. "Benchmark Simulation Results" for static and transport properties are reported for both "toy" problems, such as the Lennard-Jones fluid, and more sophisticated systems, such as various water models, small *n*-alkanes, and light gases. We recommend that novice users attempt to replicate the transport properties reported for some of these simple systems. Subsequently, we recommend attempting to replicate literature values reported for a more similar system to the one of interest. In general, validation should be performed prior to simulating new systems for which a comparison is not possible.

5 Self-Diffusivity

We recommend the Einstein approach for computing self-diffusivity as it is robust and the most commonly used method. However, we also recommend validating that the Green-Kubo method provides similar estimates. Although systematic deviations are often observed between the two methods, if the analysis is done properly the values should agree within their statistical uncertainties [20, 23, 25]. Section 5.1 discusses self-diffusivity checklist items that apply to both the Einstein and Green-Kubo approaches. Sections 5.2 and 5.3 discuss checklist items that are specific to either the Einstein or Green-Kubo approaches, respectively, for estimating the self-diffusivity constant. Section 5.4 provides a brief discussion of some topics that are relevant in certain applications.

5.1 Self-Diffusivity: General

5.1.1 Data analysis

The equations for computing D listed in Table 1 require the use of "unwrapped coordinates". That is, periodic boundary

conditions should not be applied to the coordinates, or else the self-diffusivity will be underestimated. It is possible to use the coordinates / velocities of each atom or the center of mass of each molecule in the self-diffusivity expressions. In the long-time limit, the results should be the same (see Figures 1-2 of Ref. [25]). Nevertheless, we recommend using the molecular center of mass and not the individual atomic coordinates. The reason is that short-time vibrational displacements of individual atoms, that do not contribute to the self-diffusivity, are tracked when atomic coordinates are used while the center of mass displacements are much better behaved (see Figure 3 of Ref. [25]). In either case, it is imperative to use the correct value of N (number of atoms or number of molecules) and to clearly state which approach was used.

5.1.2 Finite size effects

Finite size effects are significant for self-diffusivity calculations and must be accounted for to obtain meaningful estimates. Self-diffusivities increase with increasing system size, as can be seen in Figure 2 from Ref. [26] where the self-diffusivity of high pressure CO_2 differs by approximately 10% depending on the size of the system. We therefore stress the importance of reporting the self-diffusivity in the “infinite" box limit. This can be determined in one of two ways.

First, simulations are run with progressively larger system sizes, and the computed self-diffusivities are plotted as a function of $1/N^{1/3}$, where N is the number of molecules. As shown in Figure 2, such a plot is approximately linear, and extrapolating to when $1/N^{1/3} = 0$ gives an estimate of the self-diffusivity (although note that some studies, such as Ref. [11], extrapolate D with respect to $1/N$). The downside of this approach is that it requires multiple simulations and the large system simulations are computationally intensive.

The second approach is to estimate the infinite system self-diffusivity from a single simulation using an analytic correction factor proposed by Yeh and Hummer [33]. The correction is given by

$$D_\infty = D(L) + \frac{k_B T \xi}{6\pi\eta L} \tag{3}$$

where D_∞ is the infinite system size self-diffusivity, $D(L)$ is the computed self-diffusivity for a cubic box with edge length L, k_B is the Boltzmann constant, T is the absolute temperature, η is the shear viscosity, and $\xi = 2.837298$ is a dimensionless constant. The shear viscosity must be computed separately but fortunately, it is not typically a strong function of system size (see Section 6.1.3). As can be seen in Figure 2, both methods give similar results (compare the blue and red dashed lines). The advantage of the Yeh-Hummer correction is that a good estimate of the self-diffusivity can be obtained from a single simulation. Note that a different correction is required for non-cubic simulation boxes.[19]

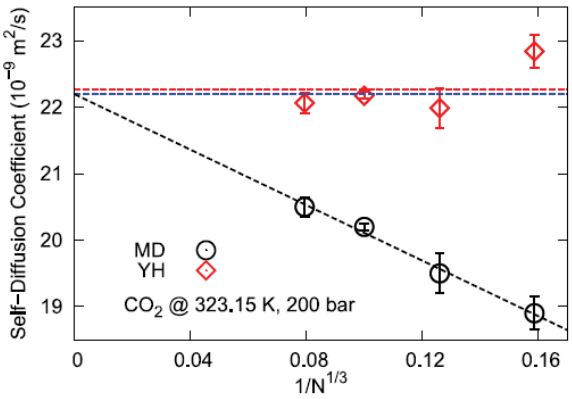

Figure 2. System size dependence of self-diffusivity obtained with Einstein approach. Reproduced with permission from J. Chem. Phys. 145, 074109 (2016). Copyright 2016 AIP Publishing [26]. Blue dashed lines are obtained by extrapolating the MD results to the infinite system size, i.e. $N^{-1/3} \to 0$. Red diamonds are the values of D after correcting for finite size effects, i.e. Equation 3. The red dashed line is an average of these corrected values of D. For further details, see Ref. [26].

5.2 Self-Diffusivity: Einstein

5.2.1 Output frequency

Self-diffusivities are computed by post-processing a trajectory. For the Einstein self-diffusivity, this means the positions of the atoms (or molecule centers of mass) should be stored as a function of time so that the MSD can be computed. How often should one save positions and at what frequency? There will always be a trade-off between accuracy (which argues for more configurations saved more frequently) and file size or runtime performance (both of which argue for fewer configurations saved less frequently). Since the long-time slope in MSD is required in the Einstein approach, configurations do not need to be saved at a high frequency. As a general guideline, to balance file size and accuracy, we recommend that approximately 1000 independent configurations be saved at uniform time intervals over the length of a production run.

5.2.2 Simulation length

Simulation length needed depends on number of molecules for which transport properties are desired. Fewer molecules requires more simulation time and vice versa. Regardless, the simulation must be long enough so that the molecules are in the diffusive regime. We recommend computing the slope from a log-log plot of MSD with respect to time, which should be approximately 1 in the diffusive regime (see Figure 3). As mentioned in Section 4.3, another heuristic is whether the MSD is sufficiently large, i.e. larger than the square of the radius of gyration of the molecule at the low end and larger than the square of half the box length at the high end. If

these criteria are met, then one can have confidence that the diffusive regime has been sampled.

5.2.3 Data analysis

In order to obtain reliable estimates of D, it is important to consider how the linear regression is performed for the MSD with respect to time (Equation 2). Specifically, the time interval that is included in the regression can have a significant impact on the predicted value of D. We recommend that only the "middle" of the MSD be used in the fit. Short time must be excluded as it follows a ballistic trajectory, while very long time is excluded due to the increased noise. Currently, we are unaware of an objective approach for defining the "middle" region. Until such an approach exists, we recommend that the author reports how the region was selected and how much variability in D can be attributed to the choice of this region. In addition, the uncertainty in the fit of the slope should be reported. A typical plot, borrowed from Ref. [20], is provided in Figure 3, where the linear regressions at long time are included.

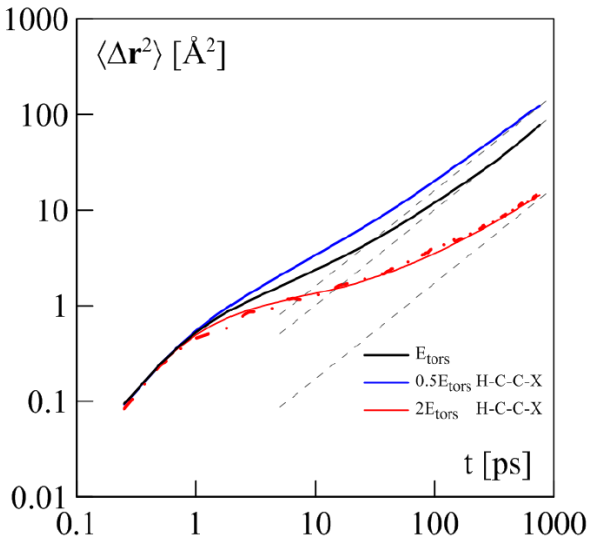

Figure 3. Log-log plot of MSD with respect to time. Copied from Figure 2 of Ref. [20]. The gray dashed lines are the long-time asymptotes of the MSD, as determined by the authors. For further details, see Ref. [20].

need to get permissions for this figure.

5.3 Self-Diffusivity: Green-Kubo

5.3.1 Output frequency

If the self-diffusivity is computed using a Green-Kubo approach, the velocities are needed as a function of time. We recommend that positions and velocities be saved as a function of time so that one can compute the self-diffusivity using both the Einstein and Green-Kubo methods; the file size increase is not that significant and other properties require knowledge of both positions and velocities. The frequency with which one must store velocities for the Green-Kubo diffusivity is much higher than that needed for applying the Einstein method. This is because the velocity autocorrelation function (VACF) that must be integrated decays very rapidly and fine time resolution is needed to perform an accurate numerical integration. Unless you have a good idea of how fast the VACF decays, we recommend saving velocities every 5 fs. This can quickly lead to unwieldy files, so one common strategy is to save velocities for a short length of time (say 50 ps) every 5 fs and then increase the time in between writes (say every 10-20 fs) for the duration.

5.3.2 Simulation length

Simulations should be long enough that the Green-Kubo integral has reached a plateau. Note that the plateau time is not the same as the required simulation time, since multiple time origins (t_0) are used to compute the Green-Kubo integral.

5.3.3 Data analysis

The most common method for computing the self-diffusivity from the VACF is to do a direct numerical integration of the VACF. If this is done, the author should provide details on how the integration was carried out (numerical procedure, algorithm, cutoffs, etc.). The running integral versus time is calculated and the self-diffusivity is estimated from the plateau value. The data are best at short time while noise dominates at long times. Like with the MSD, a cut-off needs to be determined when deciding when the integral has converged. It is important to report how sensitive the estimate is to this cut-off time.

5.4 Self-Diffusivity: Special topics

For systems that require anisotropic pressure control (e.g. membranes, etc.), use of a barostat/thermostat that maintains the correct isothermal/isobaric ensemble (e.g. extended system, Langevin piston) is required.

Calculating diffusion in membrane systems with periodic boundary conditions require some additional consideration, e.g. Saffman-Delbruck model [8, 32].

The standard non-bonded long-range cut-off corrections are not straightforward when computing diffusivity in a heterogeneous system.

6 Viscosity

Although the popularity of NEMD methods for predicting viscosity has increased in recent years, Ref. [9] demonstrates that EMD methods can be of equal accuracy and reliability to NEMD as long as best practices are followed, i.e. proper

system set-up and thorough data analysis. That being said, EMD works best for fluids with relatively low viscosity, i.e. typically less than 20 cP although EMD has been successfully implemented for systems near 50 cP. Higher viscosity systems are extremely difficult to compute with EMD and so NEMD methods are often preferred in this case.

The recommended EMD approach for predicting viscosity is Green-Kubo. We should note that Hess claims that the Einstein relation is more convenient than Green-Kubo for viscosity because "inaccuracies in the long-time correlations can be ignored by only considering integral over shorter times." [18] However, several advances have been implemented with the Green-Kubo approach since 2002 (when Ref. [18] was published). The Green-Kubo approach now appears to be the most popular EMD method found in the literature. More importantly, less arbitrary data analysis methods exist that improve the reliability and reproducibility (see Sec. 6.2.1). Section 6.1 discusses shear viscosity checklist items that apply to both the Einstein and Green-Kubo approaches. Sections 6.2 and 6.3 discuss checklist items specific to the Green-Kubo and Einstein approaches, respectively, for estimating viscosity. Section 6.4 provides a brief discussion of some topics that are relevant in certain applications.

6.1 Viscosity: General

6.1.1 Simulation length

Overall you need about 10X more data to compute viscosity than diffusivity, since viscosity is a collective property. As with the self-diffusivity, the simulation time needs to be long enough so that all the relaxation processes are adequately sampled. We recommend applying similar heuristics as those described in Section 4.3 to determine the length of the simulation required.

Figure 4, borrowed from Ref. [34], demonstrates that if the length of each independent trajectory is too short the viscosity will not converge to the correct value, regardless of how many replicates are used. Specifically, the average viscosity obtained from 100 replicates of 500 ps appears to diverge from the 1, 2, and 4 ns simulation results, suggesting that 500 ps is not sufficiently long for this system. Since it is very hard to know how long an individual trajectory needs to be, we recommend performing an analysis similar to that shown in Figure 4 to ensure adequate sampling.

It is important not to confuse the Green-Kubo integration time (the abscissa for the top panel of Figure 4) with the simulation length (the different color lines in both panels of Figure 4). Recall that the Green-Kubo integral (plotted in the top panel) is evaluated using multiple time origins (t_0), so the Green-Kubo integral contains more independent trajectories for the 4 ns line than the 500 ps line. Therefore, the time at

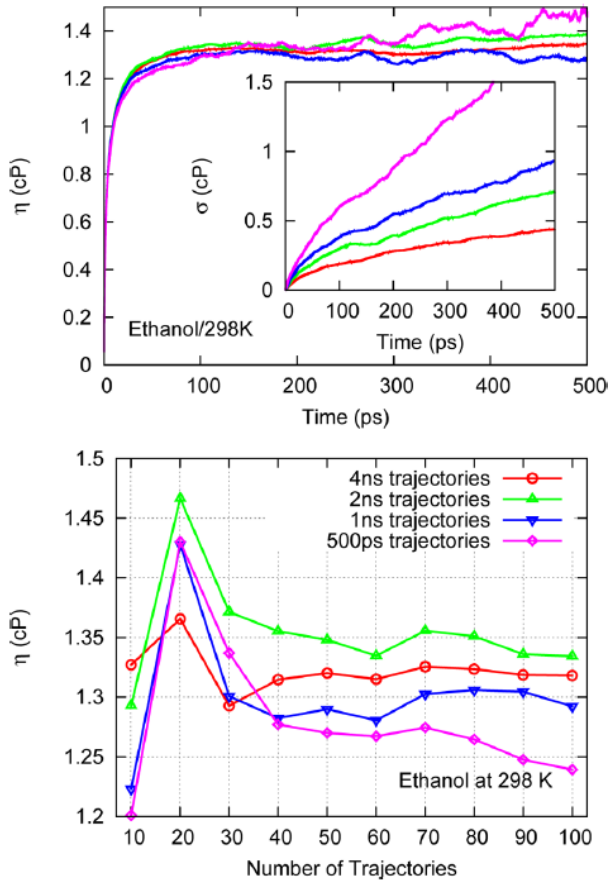

Figure 4. Viscosity dependence on simulation length. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society Ref. [34]. Different lines and symbols correspond to different simulation length, i.e. trajectory time. The inset in the top panel plots the standard deviation, σ. For further details, see Ref. [34].

which the Green-Kubo integral reaches a plateau (around 100 ps in the top panel of Figure 4) is not the same as the required simulation time. For sufficient independent trajectories, the required simulation time should typically be around an order of magnitude greater than the plateau time.

Figure 5, borrowed from Ref. [34], demonstrates that the plateau time increases with increasing viscosity, where an order of magnitude increase in viscosity corresponds to approximately an order of magnitude increase in the plateau time. In order to account for the increase in the plateau time, higher viscosity fluids require longer overall simulation times.

6.1.2 Output frequency

As with the self-diffusivity, shear viscosity is computed by post processing a data file. If the Green-Kubo procedure is used, stress tensor components need to be written out frequently enough so that an accurate estimate of the time integral can be made. Since the integral decays quickly with time, we rec-

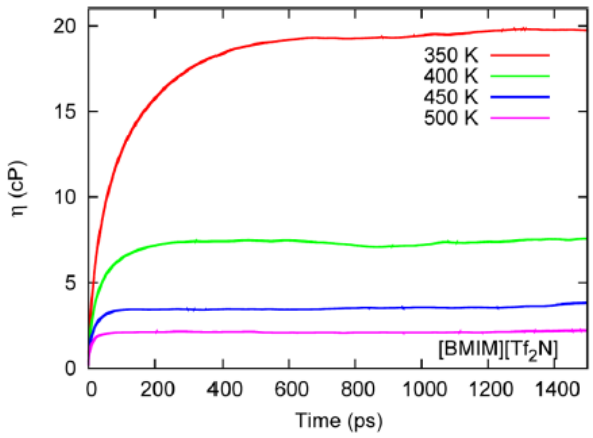

Figure 5. Plateau time dependence on viscosity. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society Ref. [34]. Different lines correspond to different temperatures and, thus, different viscosities. For further details, see Ref. [34].

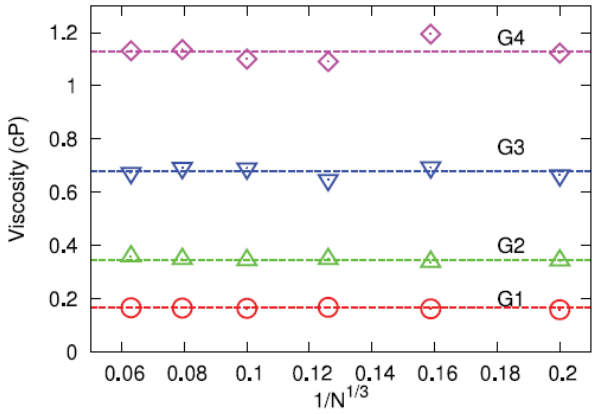

Figure 6. Finite size effects for viscosity obtained with Green-Kubo approach. Reproduced with permission from J. Chem. Phys. 145, 074109 (2016). Copyright 2016 AIP Publishing [26]. Different symbols correspond to different types of glymes (Gi). Dashed lines are average value for each glyme from various system sizes (N). For further details, see Ref. [26].

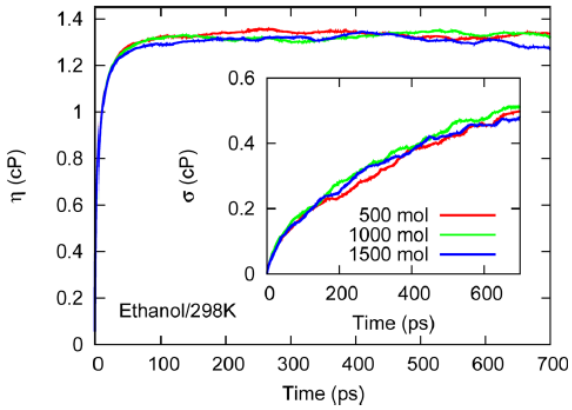

Figure 7. Finite size effects for viscosity obtained with Green-Kubo approach. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society Ref. [34]. Different colors correspond to different number of molecules. The inset plots the standard deviation, σ. For further details, see Ref. [34].

ommend writing the stress tensor every 5-10 fs. If the Einstein relationship is used, less frequent writes can be made over the length of the simulation. The user should perform some preliminary tests to ensure write frequencies are sufficient as well as to estimate file sizes for a given simulation.

6.1.3 Finite size effects

Figures 6-7 from Refs. [26] and [34], respectively, suggest that finite size effects are not significant for systems with as few as 125 and 500 molecules, respectively. Other authors, including Daivis and Evans [11], have also reported that shear viscosity has a weak dependence on system size (see Figure 4 of Ref. [11]). It is thus reasonable to neglect a system size correction, although if possible we recommend that users carry out some additional calculations to justify this assumption.

To test for system size dependence, one can run a series of simulations over a range of N molecules, where N is varied at least by a factor of two and ideally an order of magnitude. By plotting the computed shear viscosity versus $N^{-1/3}$, it is possible to ascertain if there are system size effects. We encourage authors to report these findings to help further verify system size effect trends on viscosity. If a linear trend is observed with respect to $N^{-1/3}$, the infinite system size viscosity can be extrapolated as the intercept from a linear regression. The author should report the uncertainty associated with this linear fit and extrapolation.

6.1.4 Improved precision

To improve statistical averaging, it is common to include multiple terms from the stress tensor. For example, Figure 8, borrowed from Ref. [18], demonstrates the improvement of averaging the three off-diagonal elements of the pressure tensor, compared to a single off-diagonal element. To maximize simulation efficiency for an isotropic system, we recommend that users employ a generalized form of the Green-Kubo integral [10, 12], which uses all six independent components of the stress tensor. Details are given in the Appendix of Ref. [10]. This generalized integral is given by

$$\eta = \frac{V}{10k_BT}\int_0^\infty \left\langle \tau_{ij}^{os}(0)\tau_{ij}^{os}(t)\right\rangle_{t_0} dt \qquad (4)$$

where the components τ_{ij}^{os} of the traceless, symmetric part of the stress tensor are given by

$$\tau_{ij}^{os} = \frac{\tau_{ij}+\tau_{ji}}{2} - \delta_{ij}\left(\frac{1}{3}\sum_k \tau_{kk}\right) \qquad (5)$$

where δ is the unit tensor. Note that the factor of 10 in the denominator of Eq. 4 results from assigning weighting factors of 3/3 and 4/3 for each of the six off-diagonal terms and the three diagonal terms, respectively [7, 23, 25] (although some authors have argued for an equal weighting [9], which would modify the normalization factor in the denominator of Eq. 4). The equivalent generalization of the Einstein relation is

$$\eta = \lim_{t\to\infty} \frac{V}{20k_BT}\frac{d}{dt}\sum_i\sum_j \left\langle \int_0^t \tau_{ij}^{os}(t')dt' \right\rangle_{t_0} \quad (6)$$

We are not aware of any studies that rigorously quantify the improvement in precision obtained by using all six terms. Figure 9, borrowed from Ref. [9], demonstrates that the average viscosity is nearly identical when using the three off-diagonal terms or when using six terms.

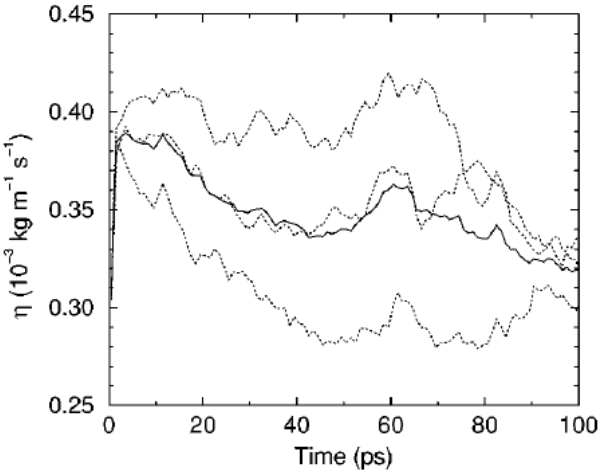

Figure 8. Green-Kubo viscosity plot. Copied from Figure 5 of Ref. [18]. Dashed lines represent a single off-diagonal element of the pressure tensor while solid line is the average of the three off-diagonal elements. For further details, see Ref. [18].**need to get permissions for this figure**.

Although fluctuations in η are significantly reduced by including multiple terms from the stress tensor, the key to improved precision of viscosity estimates is to perform several replicate simulations. For example, Figure 10 demonstrates that averaging three stress tensors is not sufficient to obtain a reliable Einstein slope as $t \to \infty$. By contrast, averaging a large number of replicates results in a near linear trend at high time.

The number of replicates used in the literature varies widely. In their study of the shear viscosity of alkanes, Payal and co-workers [28] used 10 replicates, whereas Zhang et al. [34] performed a systematic investigation of the minimal number of replicates required for convergence. They observed that a value of 30-40 replicates was statistically equivalent to 100 replicates for their system. However, the necessary number of replicates depends on the system. Specifically,

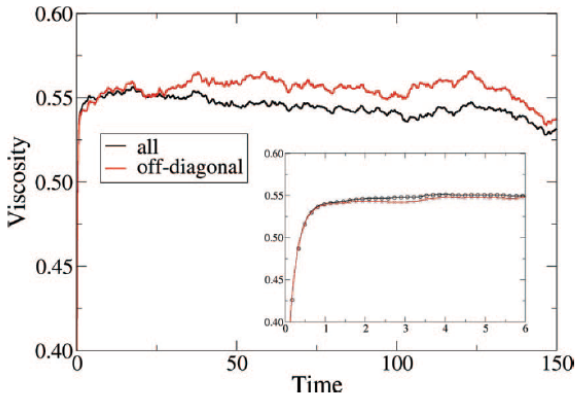

Figure 9. Green-Kubo viscosity plot. Copied from Figure 1 of Ref. [9]. Red line is obtained by averaging the three off-diagonal elements while the black line is obtained from all six pressure tensor elements. For further details, see Ref. [9].**need to get permissions for this figure**.

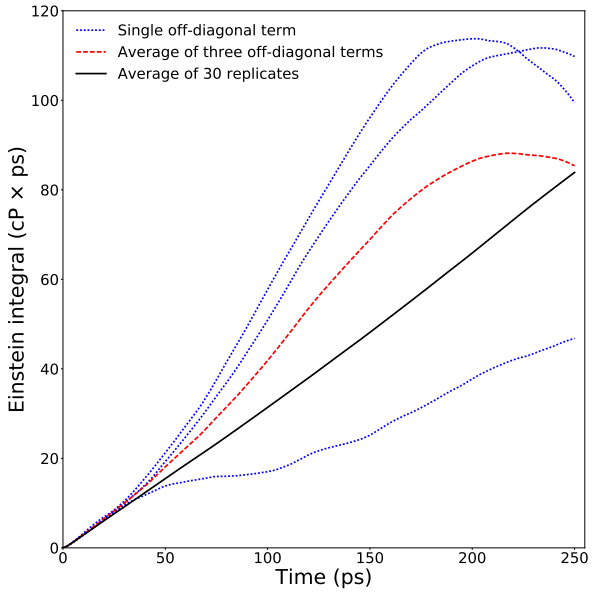

Figure 10. Improvement with three stress tensor terms and 30 replicate simulations. Einstein integral is defined as $\frac{V}{2k_bT}\left\langle \left(\int_0^t dt' \tau_{\alpha,\beta}(t') \right)^2 \right\rangle_{t_0}$. Simulation results were obtained for saturated liquid ethane at 137 K using the TraPPE-UA model in Gromacs. [16]

the compound, the temperature, the number of molecules, and the simulation time all influence the optimal number of replicates. We recommend that researchers plot how η varies with respect to the number of replicates for a range of 10–30 replicates to determine if additional simulations are needed.

6.2 Viscosity: Green-Kubo

6.2.1 Data analysis

It is imperative to report how the viscosity was estimated from Equation 1. There are three common methods: average over a specified time interval, fit the autocorrelation function to a model and analytically integrate the model fit, or fit the "running integral" to a model and extrapolate the model to infinite time. We recommend the latter methodology but discuss each approach below.

Average over time interval

A slightly ambiguous but common practice is to report an average shear viscosity that is obtained over a specified time interval. Due to large fluctuations at long times, the initial plateau of the integral at short times (around 10–100 ps) is typically the region of choice, see Refs. [9, 13]. However, it is important to explain how this time interval was selected (i.e. visual inspection, test of convergence, magnitude of fluctuations, etc.) and to quantify how much the estimated viscosity changes if the time interval were modified.

Model fit to autocorrelation function

An alternative method is to fit a model to the autocorrelation function before calculating the "running integral." The integral of the model fit can then be evaluated in the limit as $t \to \infty$. This helps to overcome large fluctuations at long times and, thereby, reduces uncertainties. The primary difficulty is finding a model that can adequately match the autocorrelation function without introducing bias into the estimate of viscosity. A common function found in the literature is

$$\frac{S^{\mathrm{f}}_{\mathrm{ACF}}(t)}{S^{\mathrm{f}}_{\mathrm{ACF}}(0)} = (1-C)cos(\omega t)\exp\left(-t/\tau_f\right)^{\beta_f} + C\exp\left(-t/\tau_s\right)^{\beta_s} \tag{7}$$

where $C, \omega, \tau_f, \tau_s, \beta_f, \beta_s$ (and sometimes $S^{\mathrm{f}}_{\mathrm{ACF}}(0)$) are fitting parameters. ω is the frequency of rapid pressure oscillations, τ_f and β_f are the time constant and exponent of fast relaxation in a stretched-exponential approximation, τ_s and β_s are constants for slow relaxation, C is the pre-factor that determines the weight between fast and slow relaxation, $S^{\mathrm{f}}_{\mathrm{ACF}}(t)$ is the stress autocorrelation function at time t, and $S^{\mathrm{f}}_{\mathrm{ACF}}(0)$ is the initial (time-zero) autocorrelation function [16].

Figure 11, from Ref. [13], demonstrates that Equation 7 has the correct shape to fit the stress autocorrelation function for this system. However, notice the significant deviation between the model fit $(S^{\mathrm{f}}_{\mathrm{ACF}})$ and the raw simulation output (S_{ACF}) for time less than 0.02 ps and the relatively small deviations in the first two peaks around 0.1 ps. These systematic deviations in the model fit can lead to significant bias in the estimated viscosity. One method to overcome this issue is to place a larger weight on short-time data or to include a cut-off time beyond which S_{ACF} data are not included in the model fit. Alternatively, it is sometimes preferable to integrate the raw S_{ACF} simulation output for short time and then integrate the model fit, $S^{\mathrm{f}}_{\mathrm{ACF}}$, to infinite time.

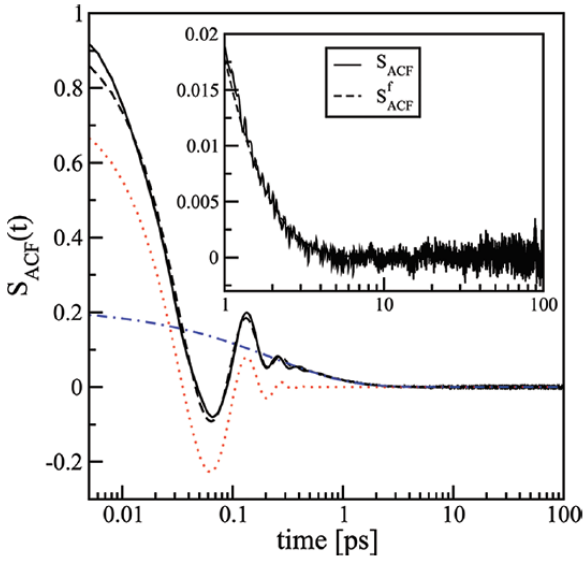

Figure 11. Fit of autocorrelation function to Equation 7. Reproduced with permission from J. Phys. Chem. A. 116 (10), pp 2564-2570. Copyright 2012 ACS Publications [13]. S_{ACF} and $S^{\mathrm{f}}_{\mathrm{ACF}}$ correspond to the raw autocorrelation function and the fit to Equation 7, respectively. The red dotted line and blue dashed–dotted line correspond to the fast and slow autocorrelation components, respectively, i.e. the first and second terms of Equation 7. For further details, see Ref. [13].

The advantage of this hybrid integration approach is that the raw data are used in the time region where small deviations in the model fit can lead to large biases in η, whereas the model fit is utilized in the time region where integration of the raw data does not converge. The time where the Green-Kubo integration switches from using S_{ACF} to $S^{\mathrm{f}}_{\mathrm{ACF}}$, referred to as the switch-time (t_{s}), should be after the "fast" autocorrelation component has dissipated (the first term in Equation 7 and the red dotted line in Figure 11).

The hybrid integration approach is especially preferred when S_{ACF} is highly oscillatory, such as that shown in Figure 12, from Ref. [14]. The dashed line in Figure 12 is obtained by fitting the S_{ACF} data (denoted $\Phi_\eta(t)$) for $t > 0.015$ ps to the model $S^{\mathrm{f}}_{\mathrm{ACF}} = a\exp(-t/b)$, where a and b are fitting parameters. Note that a simpler exponential decay function can be used with the hybrid integration approach because the model does not need to fit the autocorrelation function over the entire time range, just for $t > t_{\mathrm{s}}$.

Similar to the methods discussed previously, it is important to quantify the variability in viscosity that arises from the model fit. For example, we recommend bootstrapping the uncertainties by repeating the model fit for hundreds of randomly selected subsets of S_{ACF}. If the hybrid integration approach is utilized, it is important to investigate and report

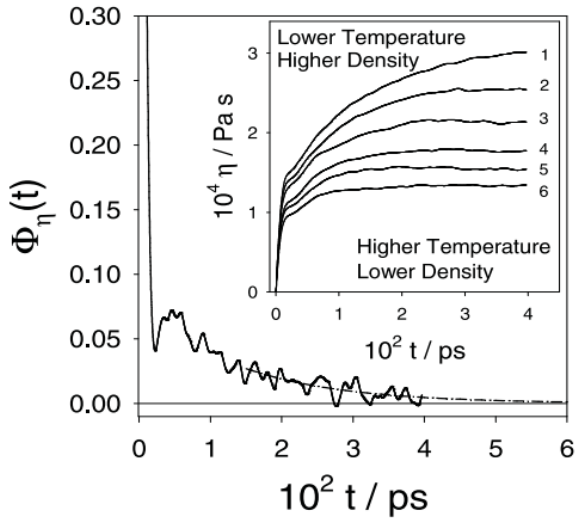

Figure 12. Fit of autocorrelation function ($\Phi_\eta(t)$) to $a\exp(-t/b)$. Copied from Figure 7 of Ref. [14]. Solid line is the raw autocorrelation function while the dashed line is the model fit for $t > 0.015$ ps. For further details, see Ref. [14].**Note: When I went to Taylor and Francis to request re-use of this figure, they asked for 315 dollars. I think we need a substitute.**

how sensitive the final viscosity value is to the switch-time and/or to discuss how t_s is chosen. Furthermore, if a weighting function or cut-off time is implemented when fitting S^f_{ACF}, the impact of these parameters should be discussed.

Model fit to running integral

The method we recommend for obtaining viscosity from EMD is to fit an analytic function directly to the "running integral". The primary advantage of fitting a model to the "running integral" over the previous approach of fitting a model to the autocorrelation function (i.e. Equation 7) is that uncertainties in the model fit do not propagate through the integration.

For example, Refs. [29] and [34] recommend fitting the "running integral" to a double-exponential function

$$\eta(t) = A\alpha\tau_1\left(1-\exp\left(-t/\tau_1\right)\right) + A(1-\alpha)\tau_2\left(1-\exp\left(-t/\tau_2\right)\right) \quad (8)$$

where A, α, τ_1, and τ_2 are fitting parameters. Note that the "true" estimate of η is obtained as $t \to \infty$, i.e. $\eta_\infty = A\alpha\tau_1 + A(1-\alpha)\tau_2$.

Ref. [7] proposes an alternative model by integrating the slow stretched-exponential function (second term in Equation 7) which results in the expression

$$\eta(t) = \eta_\infty(1-\exp(-(t/\tau_s)^{\beta_s})) \quad (9)$$

where η_∞, τ_s, and β_s are fitting parameters that relate to the infinite-time viscosity, decay time, and the exponent of slow relaxation.

We recommend the use of Equation 8 as we have found it to be a more flexible fitting model, i.e. the optimized sum-squared-error is typically lower than that of Equation 9. That being said, the η_∞ estimates obtained with Equations 8 and 9 are quite similar. Deviations in η_∞ between the two equations are generally less than 1% for both low (gas phase) and high (compressed liquid phase) viscosities. Regardless of whether Equation 8 or 9 is implemented, it is important to include a description of how the fit is performed, i.e. the objective function, range of data included, etc.

Ref. [34] recommends that the data be weighted by the inverse of the standard deviation (σ) with respect to time. They fit σ to a model At^b, where t is time and A and b are fitting parameters. This fit is used to develop a weighting model of the form $w \propto t^{-b}$, where w is the weight and b is the weighting exponent obtained from the σ model fit. If such a model is utilized, the resulting estimate of η may depend strongly on b, the weighting exponent. For example, Figure 13, borrowed from Ref. [34], compares η for two different values of b in the weighting model, namely, when b is a predetermined value of 0.5 and when b is fit to σ in the replicate averages. Note that Ref. [29] recommended a value of $b = 2$. Ref. [34] demonstrated that for $b = 2$ the estimated value of η for an ionic liquid ([BMIM][Tf_2N]) at 350 K is approximately 11 cP (compared to ≈ 19 cP in the bottom panel of Figure 13). For these reasons, we recommend that the author quantifies the uncertainty in the estimated viscosity due to the value of b. Propagating the uncertainty in η from b can be accomplished by implementing a two-step bootstrap method. First, a distribution of b values are obtained by bootstrapping the σ model fit. Second, a distribution of η values are computed by fitting Equation 8 with each value of b from the distribution generated in the previous step.

Ref. [34] also suggests that a cut-off time be implemented to improve the fit. They provide a heurestic that the cut-off time correspond to when the standard deviation is 40% of the plateau value. Regardless of how the cut-off is determined, it is important to quantify the degree to which the estimated viscosity depends on this parameter. For example, Zhang et al. reported that the viscosity decreased by 0.8% and 6.1% when using a cut-off time corresponding to a standard deviation of 30% or 20% the plateau value, respectively. However, the magnitude of variability depends strongly on the system. We recommend that the author quantify the cut-off time dependence.

Furthermore, Ref. [34] recommends excluding short-time data from the fitting procedure. In Figure 14, borrowed from Ref. [34], we observe large oscillations at very short times, ca. $t < 2$ ps. A weighting function with a t^{-b} form assigns an inappropriately large weight to these short-time data points. Therefore, it is important to exclude data in this short-time

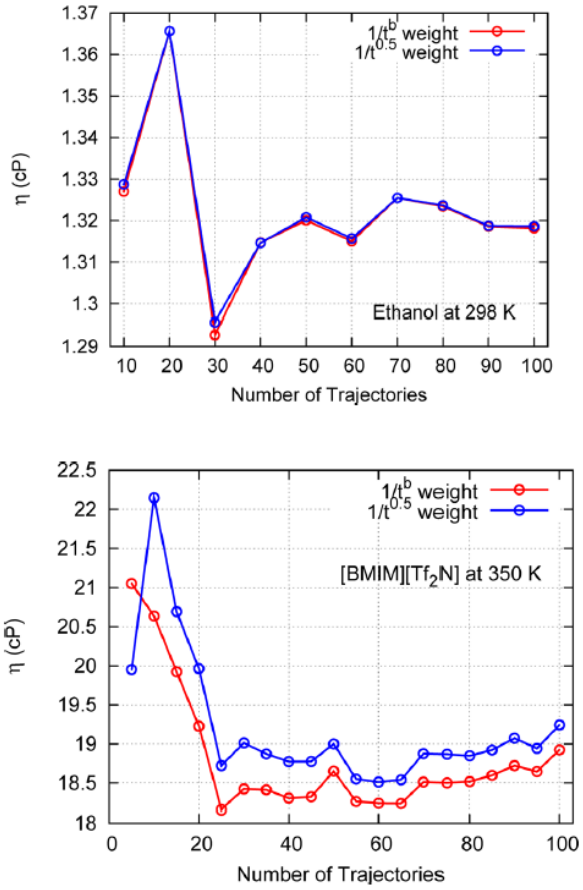

Figure 13. Viscosity dependence on the exponent of the weighting model, b. $b = 0.52$ for Ethanol at 298K, top panel, while b is between 0.60–0.73 for [BMIM][Tf_2N] at 350 K, bottom panel. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society. For further details, see Ref. [34].

region from the model fitting.

6.3 Viscosity: Einstein

6.3.1 Data analysis

Since the Einstein relation is valid in the limit of infinite time, in theory the slope should only be computed at long time. However, by contrast with self-diffusivity, the long-time slope may behave poorly (for example, recall Figure 10). For this reason, it is common to fit the slope over an intermediate time interval, e.g. 10–50 ps. In either case, we recommend that the author explain why the slope was calculated using a given time interval and how much variability is introduced if a different region is selected. For example, similar to the bootstrap method discussed in Sec. 4.2.3, a useful measure of uncertainty is the range of η values obtained by analyzing numerous (order of 100) different time regions.

Since the viscosity is estimated from the slope of the Einstein integral, the average of replicates can be performed in one of two ways. The first option is to calculate the viscosity (i.e. the slope) with respect to time for each replicate and then average the replicate viscosities. However, this approach results in large fluctuations and, therefore, large uncertainties.

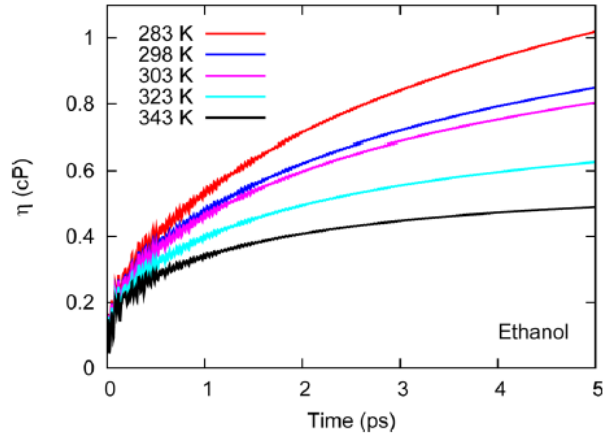

Figure 14. Large fluctuations at very short time. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society. For further details, see Ref. [34].

The second, and recommended, method when implementing the Einstein approach is to average the Einstein integral of the multiple replicates. The resulting Einstein integral is often linear over a large time interval if sufficient replicates are used. Subsequently, the slope is determined from this average Einstein integral. Fortunately, with sufficient replicate simulations the slope tends to be fairly constant over intermediate and long-time intervals.

The number of replicates needed has not been rigorously investigated as it has for the Green-Kubo approach. For this reason, we recommend creating a plot of viscosity with respect to number of replicates (see Figure 4) to determine when sufficient replicates have been simulated. It is our experience that the necessary number of replicates is similar to that for Green-Kubo. As recommended for Green-Kubo, we also recommend bootstrapping the uncertainty. This is done by randomly sampling which replicates are included in the average Einstein integral, calculating the viscosity from the slope, and producing a distribution of these viscosity values from thousands of different random sets of replicates.

6.4 Viscosity: Special topics

The GROMACS manual reports that viscosity "is very dependent on the treatment of the electrostatics. Using a (short) cut-off results in large noise on the off-diagonal pressure elements, which can increase the calculated viscosity by an order of magnitude." [16, 18]

7 Conclusions

Molecular simulation is commonly used to predict transport properties, however, without careful simulation design and acute analysis, results may not be meaningful. For example, an EMD simulation with an insufficient number of molecules can severely underestimate self-diffusivity. This work outlines the best practices in the design and analysis of MD simulations for transport properties.

For self-diffusivity, it is suggested that the Einstein method is employed. In liquid systems, the NVE ensemble is suggested over the NVT ensemble due to the potential interference of the thermostats in self-diffusivity prediction. Uncertainty is reduced by running multiple independent simulations, allowing for a more thorough sampling of the system's possible states. Multiple runs at different system sizes can be used to extrapolate the infinite system size limit prediction for self-diffusivity. Atom positions are recommended to be outputted 1000 times over a production run, however, the user can choose to output less frequently to reduce file size or more frequently for potentially increased accuracy. To ensure simulations are run long enough for the dynamics of the system to be fully emulated, the user can run a series of simulations at differing lengths, and observe deviation in estimated self-diffusivity with changes in simulation time. The system's degree of exploration of configuration space can be estimated by calculating the MSD of the molecules and comparing it to the radius of gyration and box length. The MSD should be greater than the radius of gyration, and ideally, on the order of the box length. In post-simulation analysis, best practice improves precision by averaging the velocity autocorrelation function over all molecules and over multiple time origins. Some judgment by the user is necessary to decide where the slope is measured for the Einstein approach, and it is important that the user communicate the approach used and justify how the decision was made. Measures should be taken to rigorously estimate the precision of the self-diffusivity prediction, a method of bootstrapping these uncertainties is detailed in this work.

For viscosity, the Green-Kubo approach is recommended, although the Einstein method may be preferred with certain systems. It is worth noting that NEMD simulations are highly preferred for moderate to high viscosity materials (more than 20 cP or so). The NVE ensemble is suggested, but some success has been found using the NVT ensemble [6, 13]. Because it is a collective property, viscosity requires significantly more data than self-diffusivity. The simulation length of each trajectory should typically be at least an order of magnitude greater than the Green-Kubo integral plateau time. This can be more precisely determined by comparing the Green-Kubo integrals of varying simulation lengths. Due to their slower dynamics, more viscous materials require longer simulation times. Stress tensor components are recommended to be outputted every 5-10 fs. System size seems to be of little impact on viscosity prediction (see Figures 6-7), however, it is still recommended that the user justify their choice of system size by plotting $N^{-1/3}$ versus predicted viscosity. Any trend observed would suggest system size effects, and an infinite system size result could be estimated from a linear fit. In post-simulation data processing, it is recommended to average over all six independent components of the stress tensor to enhance precision. The number of replicates a system needs can vary greatly depending on the compound, number of molecules, temperature, and simulation length. Replicates required can be investigated by plotting variance in viscosity versus number of replicates (on the order of 10-30), wherein a trend may indicate that more replicates are needed. It is recommended that the viscosity is estimated from the Green-Kubo equation by fitting the running integral to a double-exponential function as specified by Zhang et al. [34] (see Section 6.2.1).

The focus of the current work is outlining best practices in EMD for self-diffusivity and viscosity prediction. This work may be expanded on in the future with NEMD techniques or best practices for other transport properties.

8 Acknowledgments

Funder and other information can be given here.

References

[1] Athanassios Panagiotopoulos, The University of Princeton Class Notes;. Accessed: 2018-01-01. http://paros.princeton.edu/cbe422/md2.pdf; http://paros.princeton.edu/cbe520/Transport.pdf.

[2] David Kofke, The University of Buffalo Class Notes;. Accessed: 2018-01-01. http://www.eng.buffalo.edu/ kofke/ce530/Lectures/Lecture12.ppt.pdf.

[3] Ed Maginn, The University of Notre Dame Class Notes;. Accessed: 2018-01-01. https://www3.nd.edu/ ed/notes.pdf.

[4] Scott Shell, The University of California Santa Barabara Class Notes;. Accessed: 2018-01-01. https://engineering.ucsb.edu/ shell/che210d/Computing_properties.pdf.

[5] **Allen MP**, Tildesley DJ. Computer simulation of liquids. Second ed. Oxford England: Oxford University Press; 2017.

[6] **Basconi JE**, Shirts MR. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. Jouornal of Chemical Theory and Computation. 2013 JUL; 9(7):2887–2899. doi: {10.1021/ct400109a}.

[7] **Borodin O**, Smith GD, Kim H. Viscosity of a Room Temperature Ionic Liquid: Predictions from Nonequilibrium and Equilibrium Molecular Dynamics Simulations. The Journal of Physical Chemistry B. 2009; 113(14):4771–4774. https://doi.org/10.1021/jp810016e, doi: 10.1021/jp810016e, pMID: 19275203.

[8] **Camley BA**, Lerner MG, Pastor RW, Brown FLH. Strong influence of periodic boundary conditions on lateral diffusion in lipid bilayer membranes. The Journal of Chemical Physics. 2015; 143(24):243113. https://doi.org/10.1063/1.4932980, doi: 10.1063/1.4932980.

[9] **Chen T**, Smit B, Bell AT. Are pressure fluctuation-based equilibrium methods really worse than nonequilibrium methods for calculating viscosities? The Journal of Chemical Physics. 2009; 131(24):246101. https://doi.org/10.1063/1.3274802, doi: 10.1063/1.3274802.

[10] **Daivis PJ**, Evans DJ. COMPARISON OF CONSTANT-PRESSURE AND CONSTANT VOLUME NONEQUILIBRIUM SIMULATIONS OF SHEARED MODEL DECANE. Journal of Chemical Physics. 1994; 100(1):541–547. 19 AMER INST PHYSICS MN426.

[11] **Daivis PJ**, Evans DJ. Transport Coefficients of Liquid Butane Near the Boiling Point by Equilibrium Molecular Dynamics. Journal of Chemical Physics. 1995; 103(10):4261–4265.

[12] **Evans DJ**, Morriss GP. Statistical Mechanics of Nonequilibrium Liquids. Academic press, London; 1990.

[13] **Fanourgakis GS**, Medina JS, Prosmiti R. Determining the Bulk Viscosity of Rigid Water Models. The Journal of Physical Chemistry A. 2012; 116(10):2564–2570. http://dx.doi.org/10.1021/jp211952y, doi: 10.1021/jp211952y, pMID: 22352421.

[14] **Fernández GA**, Vrabec J, Hasse H. Shear viscosity and thermal conductivity of quadrupolar real fluids from molecular simulation. Molecular Simulation. 2005; 31(11):787–793. https://doi.org/10.1080/08927020500252599, doi: 10.1080/08927020500252599.

[15] **Frenkel D**, Smit B. Understanding molecular simulation : from algorithms to applications. 2nd ed. Computational science series, San Diego: Academic Press; 2002.

[16] **GROMACS**. GROMACS Reference Manual; 2016.

[17] **Haile JM**. Molecular Dynamics Simulation: Elementary Methods. New York: John Wiley and Sons, Inc.; 1992.

[18] **Hess B**. Determining the shear viscosity of model liquids from molecular dynamics simulations. The Journal of Chemical Physics. 2002; 116(1):209–217. http://aip.scitation.org/doi/abs/10.1063/1.1421362, doi: 10.1063/1.1421362.

[19] **Kikugawa G**, Nakano T, Ohara T. Hydrodynamic consideration of the finite size effect on the self-diffusion coefficient in a periodic rectangular parallelepiped system. J Chem Phys. 2015 JUL 14; 143(2). doi: {10.1063/1.4926841}.

[20] **Kondratyuk ND**, Norman GE, Stegailov VV. Rheology of liquid n-triacontane: Molecular dynamics simulation. Journal of Physics: Conference Series. 2016; 774(1):012039. http://stacks.iop.org/1742-6596/774/i=1/a=012039.

[21] **LAMMPS**. LAMMPS Users Manual. Sandia National Laboratories; 2017.

[22] **Leach AR**. Molecular modelling : principles and applications. 2 ed. Pearson Prentice Hall; 2001.

[23] **Liu H**, Maginn E, Visser AE, Bridges NJ, Fox EB. Thermal and Transport Properties of Six Ionic Liquids: An Experimental and Molecular Dynamics Study. Industrial & Engineering Chemistry Research. 2012; 51(21):7242–7254. http://dx.doi.org/10.1021/ie300222a, doi: 10.1021/ie300222a.

[24] **Ma J**, Zhang Z, Xiang Y, Cao F, Sun H. On the prediction of transport properties of ionic liquid using 1-n-butylmethylpyridinium tetrafluoroborate as an example. Molecular Simulation. 2017; 43(18):1502–1512. https://doi.org/10.1080/08927022.2017.1321760, doi: 10.1080/08927022.2017.1321760.

[25] **Mondello M**, Grest GS. Viscosity calculations of n-alkanes by equilibrium molecular dynamics. The Journal of Chemical Physics. 1997; 106(22):9327–9336. https://doi.org/10.1063/1.474002, doi: 10.1063/1.474002.

[26] **Moultos OA**, Zhang Y, Tsimpanogiannis IN, Economou IG, Maginn EJ. System-size corrections for self-diffusion coefficients calculated from molecular dynamics simulations: The case of CO2, n-alkanes, and poly(ethylene glycol) dimethyl ethers. The Journal of Chemical Physics. 2016; 145(7):074109. http://dx.doi.org/10.1063/1.4960776, doi: 10.1063/1.4960776.

[27] **Nieto-Draghi C**, Fayet G, Creton B, Rozanska X, Rotureau P, de Hemptinne JC, Ungerer P, Rousseau B, Adamo C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. Chemical Reviews. 2015; 115(24):13093–13164. http://dx.doi.org/10.1021/acs.chemrev.5b00215, doi: 10.1021/acs.chemrev.5b00215, pMID: 26624238.

[28] **Payal RS**, Balasubramanian S, Rudra I, Tandon K, Mahlke I, Doyle D, Cracknell R. Shear viscosity of linear alkanes through molecular simulations: quantitative tests for n-decane and n-hexadecane. Molecular Simulation. 2012; 38(14-15):1234–1241. https://doi.org/10.1080/08927022.2012.702423, doi: 10.1080/08927022.2012.702423.

[29] **Rey-Castro C**, Vega LF. Transport Properties of the Ionic Liquid 1-Ethyl-3-Methylimidazolium Chloride from Equilibrium Molecular Dynamics Simulation. The Effect of Temperature. The Journal of Physical Chemistry B. 2006; 110(29):14426–14435. http://dx.doi.org/10.1021/jp062885s, doi: 10.1021/jp062885s, pMID: 16854152.

[30] **Shen SDWKWP V K**, Hatch E H W, NIST Standard Reference Simulation Website, NIST Standard Reference Database Number 173, National Institute of Standards and Technology, Gaithersburg MD, 20899;. Accessed: 2018-04-25. http://doi.org/10.18434/T4M88Q.

[31] **Ungerer P**, Nieto-Draghi C, Rousseau B, Ahunbay G, Lachet V. Molecular simulation of the thermophysical properties of fluids: From understanding toward quantitative predictions. Journal of Molecular Liquids. 2007; 134(1):71 – 89. http://www.sciencedirect.com/science/article/pii/S016773220600331X, doi: https://doi.org/10.1016/j.molliq.2006.12.019, eMLG/JMLG 2005 Special Issue.

[32] **Venable RM**, Ingólfsson HI, Lerner MG, Perrin BS, Camley BA, Marrink SJ, Brown FLH, Pastor RW. Lipid and Peptide Diffusion in Bilayers: The Saffman–Delbrück Model and Periodic Boundary Conditions. The Journal of Physical Chemistry B. 2017;

121(15):3443–3457. https://doi.org/10.1021/acs.jpcb.6b09111, doi: 10.1021/acs.jpcb.6b09111, pMID: 27966982.

[33] **Yeh IC**, Hummer G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. The Journal of Physical Chemistry B. 2004; 108(40):15873–15879. http://dx.doi.org/10.1021/jp0477147, doi: 10.1021/jp0477147.

[34] **Zhang Y**, Otani A, Maginn EJ. Reliable Viscosity Calculation from Equilibrium Molecular Dynamics Simulations: A Time Decomposition Method. Journal of Chemical Theory and Computation. 2015; 11(8):3537–3546. http://dx.doi.org/10.1021/acs.jctc.5b00351, doi: 10.1021/acs.jctc.5b00351, pMID: 26574439.