Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics [Article v1.0]

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 January 6, 2019

Abstract

The ability to predict transport properties (e.g., diffusivity, viscosity, and 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 multicomponent diffusivity.

***For correspondence:**

ed@nd.edu (EM); daniel.roe@nih.gov (DR); elliot1@uakron.edu (JRE); richard.messerly@nist.gov (RAM); daniel.j.carlson@byu.edu (DC)

Received: 22 June 2018
Accepted: 29 November 2018

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

1 Introduction

Transport properties describe the rates at which mass, momentum, heat or charge move through a given substance. They involve the mean squared displacement (MSD) 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.

Some molecular simulation packages include built-in post-simulation analysis tools that automatically estimate transport properties from an EMD or NEMD simulation (e.g., Refs. [1–6]). In addition, several stand-alone trajectory analysis tools are available which are intended to be simulation code agnostic (e.g., Refs. [7–9]). However, 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," as no amount of post-processing can compensate for a poorly designed simulation. Following best practices for both the molecular simulation set-up and data analysis 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) Allen and Tildesley, *Computer simulation of liquids*, pages 73–79, 274–281, and 292–296. [10]
   (b) Frenkel and Smit, *Understanding molecular simulation: from algorithms to applications*, pages 87–90 and 509–523. [11]
   (c) Leach, *Molecular modelling: principles and applications*, pages 374–382. [12]
   (d) Haile, *Molecular dynamics simulation: elementary methods*, pages 189–192 and 485–488. [13]
2. Class notes:
   (a) Panagiotopoulos. [14]
   (b) Kofke. [15]
   (c) Maginn. [16]
   (d) Shell. [17]
3. Peer-reviewed articles:
   (a) Chen et al., *Are pressure fluctuation-based equilibrium methods really worse than nonequilibrium methods for calculating viscosities?* [18]
   (b) Hess, *Determining the shear viscosity of model liquids from molecular dynamics simulations.* [19]
   (c) Nieto-Draghi et al., *A general guidebook for the theoretical prediction of physicochemical properties of chemical for regulatory purposes*, pages 13139–13140. [20]
   (d) Ungerer et al., *Molecular simulation of the thermophysical properties of fluids: From understanding toward quantitative predictions.* [21]
4. Software manuals:
   (a) LAMMPS. [1]
   (b) GROMACS. [2]
   (c) AMBER. [3]

Most textbooks 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 mechanical variable associated with the particular transport property under consideration, and $\dot{\xi}$ signifies a time derivative. Integrals of the form given by Equation

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

1 are known as "Green-Kubo" integrals. It is trivial to show that an integrated form of Equation 1 results in an equivalent expression for γ known as the "Einstein" formula

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

where the derivative form is often preferred.

For self-diffusivity, ξ is the Cartesian atom position (r_α) and the time correlation function, $\dot{\xi}$, in Equation 1 is of the molecular velocities (v_α). For shear viscosity, the integral in Equation 1 is of the time correlation of the off-diagonal elements of the stress tensor. For thermal conductivity, the integral is over the energy current, and for 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. 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

An overview of the checklist items for each property (D and η) and method (Green-Kubo and Einstein) is found on the next page. Secs. 4-6 provide a detailed discussion for each checklist item.

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, rather than the canonical (*NVT*: constant number of molecules, volume, temperature) or isothermal–isobaric (*NPT*: constant number of molecules, pressure, temperature) ensembles. 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 the system density has properly equilibrated
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

It is important that each step consist of both an equilibration and production period. The average pressure and temperature for the *NVE* production run are computed and should be close to the input P and T of the original *NPT* run (although it is typical to observe some deviation between the average and prescribed values of P and T). These average pressures and temperatures must be reported along with the self-diffusivity and/or viscosity.

Note that, although the best practice is to use the *NVE* ensemble (Steps 1 to 3), it is common to see values reported using the *NPT* ensemble (just Step 1) or *NVT* ensemble (Steps 1 and 2). 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 [22].

A study by Basconi and Shirts [23] reached a similar conclusion, and provides guidelines for how thermostats should be applied when computing transport properties. To summarize their results, *NVE* ensemble estimates for D and η are statistically indistinguishable from those obtained by *NVT* simulations with velocity-scaling thermostats (e.g., Berendsen [24], stochastic rescaling [25], and Nosé-Hoover [26] thermostats) and a wide range of thermostat coupling strengths (e.g., time constants of 0.1, 1, and 10 ps, where stronger coupling is achieved with a smaller time constant). By contrast, velocity-randomizing algorithms (e.g., Andersen thermostat [27] and Langevin dynamics [28]) with strong coupling (time constants of 0.1 and 1 ps) dramatically decrease D and increase η relative to the *NVE* ensemble values. Therefore, we recommend using either the *NVT* or *NVE* ensemble, with *NVE* being preferred. If *NVT* simulations are used, we recommend judiciously selecting the thermostat and coupling strength by consulting Refs. [22, 23].

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

CHECKLIST FOR COMPUTING SELF-DIFFUSIVITY

- ☐ **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 each replicate simulation.
    - ☐ Increase the output frequency. Einstein: see Sec. 5.2.1. Green-Kubo: see Sec. 5.3.1.
    - ☐ Ensure that simulations are sufficiently long. Einstein: see Sec. 5.2.2. Green-Kubo: see Sec. 5.3.2.
  - ☐ Check for system size effects. See Sec. 5.1.2
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of D.
  - ☐ Improve precision by averaging over:
    - ☐ Multiple time origins. See Sec. 4.2.1.
    - ☐ N molecules. See Sec. 4.2.1.
    - ☐ Three dimensions (x, y, z). See Sec. 4.2.1.
    - ☐ Several replicate simulations. See Sec. 4.1.2.
  - ☐ Clearly communicate how D is obtained from Equation 1 (Green-Kubo) or Equation 2 (Einstein). See Sec. 4.2.2. Einstein: see Secs. 5.1.1 and 5.2.3. Green-Kubo: see Secs. 5.1.1 and 5.3.3.
  - ☐ Correct for system size effects, e.g., Equation 3. See Sec. 5.1.2.
  - ☐ Report the uncertainty in D:
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis. Einstein: see Sec. 5.2.3. Green-Kubo: see Sec. 5.3.3.
- ☐ **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 VISCOSITY

- ☐ **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 each replicate simulation.
    - ☐ Increase the output frequency. See Sec. 6.1.1.
    - ☐ Ensure that simulations are sufficiently long. See Sec. 6.1.2.
  - ☐ Check for system size effects. See Sec. 6.1.3.
- ☐ **Common pitfalls.** Double-check that your results are not plagued by one of the common pitfalls. See Sec. 4.3.
- ☐ **Post-simulation data analysis.** Data analysis is key for obtaining reproducible and meaningful estimates of η.
  - ☐ Improve precision by averaging over multiple:
    - ☐ Time origins. See Sec. 4.2.1.
    - ☐ 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 (Green-Kubo) or Equation 2 (Einstein). See Sec. 4.2.2. Green-Kubo: see Sec. 6.2.1. Einstein: see Sec. 6.3.1.
  - ☐ Report the uncertainty in η:
    - ☐ Bootstrap replicate simulations. See Sec. 4.2.3.
    - ☐ Perform sensitivity analysis. Green-Kubo: see Sec. 6.2.1. Einstein: see Sec. 6.3.1.
- ☐ **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.

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

Table 1. Equilibrium molecular dynamics equations.

Property (γ)	Mechanical variable (ξ)	Green-Kubo (Equation 1)	Einstein (Equation 2)
Self-diffusivity (D)	r_α	$\frac{1}{d_\alpha}\int_0^\infty dt \left\langle \frac{1}{N}\sum_{i=1}^{N} v_{\alpha,i}(t)v_{\alpha,i}(0)\right\rangle_{t_0}$	$\frac{1}{2d_\alpha}\lim_{t\to\infty}\frac{d}{dt}\left\langle \frac{1}{N}\sum_{i=1}^{N}\left\lvert r_{\alpha,i}(t)-r_{\alpha,i}(0)\right\rvert^2\right\rangle_{t_0}$
Shear viscosity (η)	$r_\alpha v_\beta$	$\frac{V}{k_\mathrm{B}T}\int_0^\infty dt \left\langle \tau_{\alpha\beta}(t)\tau_{\alpha\beta}(0)\right\rangle_{t_0}$	$\frac{V}{2k_\mathrm{B}T}\lim_{t\to\infty}\frac{d}{dt}\left\langle \left(\int_0^t dt'\tau_{\alpha\beta}(t')\right)^2\right\rangle_{t_0}$

t = time
r = position
v = velocity
α, β = x, y, or z Cartesian coordinates of the atoms or molecule center of mass
d_α = dimensionality (1, 2, or 3)
N = number of atoms or molecules (see Sec. 4.2.2)
V = volume
T = temperature
k_B = Boltzmann constant
$\tau_{\alpha\beta}(t) = \frac{1}{V}\sum_{i=1}^{N}\left(mv_{\alpha,i}(t)v_{\beta,i}(t) + r_{\alpha,i}(t)f_{\beta,i}(t)\right), \alpha \neq \beta$
$f_{\beta,i}$ = the force acting on particle i in direction β
m = mass of particle
$\langle\cdots\rangle_{t_0}$ = an average over time origins (see Sec. 4.2.1)

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 seeds to initialize velocities). The primary advantage of performing replicates as opposed to one longer simulation is the computational speed-up. Figure 2 in Ref. [29] 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 real time is reduced, although the CPU time may be the same or more. 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). The uncertainty (specifically, the standard error of the mean) is inversely proportional to the square root of the number of replicates (see Figure 7 of Ref. [30] and Figure 8 of Ref. [31]). Therefore, increasing the number of replicates is a simple, fast, and direct way to reduce the uncertainty. For example, the fluctuations in η are much smaller for the average of 10 replicates compared to that of a single longer simulation (see Figure 2 of Ref. [29]). As fluctuations in η are typically much larger than D, more replicate simulations are required for estimating viscosity (see Sec. 6.1.4).

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 squared 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 numbers 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 $D_{xx} = \frac{1}{2}\lim_{t\to\infty}\frac{d}{dt}\left\langle \frac{1}{N}\sum_{i=1}^{N}\left\lvert x_i(t)-x_i(0)\right\rvert^2\right\rangle$. 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. We also encourage the user to verify that the off-diagonal

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

terms are approximately zero. Note that for inhomogeneous systems, the diagonal terms will not necessarily be equivalent and the off-diagonal terms may not be zero.

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*, *xz*) 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 (see Ref. [32] for details regarding correlation time).

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 [30]. 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. 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

Providing meaningful estimates of uncertainty is a pivotal, but often overlooked, step in reporting molecular simulation results. For this reason, in addition to the discussion provided below, we recommend a close study of Refs. [33] and [32]. Although force field deficiencies can lead to large systematic deviations with respect to the "true" experimental value, we limit our discussion to random uncertainties associated with fluctuations in the simulation output.

Replicates can provide a rigorous uncertainty assessment. We recommend bootstrapping the uncertainties by randomly sampling which replicates are included in the data analysis procedure [34]:

1. Randomly select (with replacement) a subset 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 through 3 several hundred times
5. Generate distribution of D or η values from step 4
6. Determine lower and upper uncertainty bounds of D or η at desired confidence level, 1 – α

In step 1, the resampled set should be the same size as the original set of replicates, i.e., $N_{\text{boots}} = N_{\text{reps}}$. For this reason, replacement after each random selection is key to ensure that a different set is resampled for each cycle (step 4). Note that this approach is typically most reliable when $N_{\text{reps}} \geq 20$. However, we recommend this approach even if only a few replicate simulations are performed to at least have a rough estimate of the uncertainty.

Step 6 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 $\alpha/2 \times 100\%$ 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. Performing simulations with single precision is one of the main causes for poor conservation of energy due to the accumulation of round-off errors. For this reason, we recommend the use of full double precision or double/fixed precision.

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 [13] 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

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

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 traversed far enough to sample a significant amount of configuration space.

4.4 General transport: Validation

Validation is an important step to verify 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) [35]. "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 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 [36–38].

Sec. 5.1 discusses self-diffusivity checklist items that apply to both the Einstein and Green-Kubo approaches. Secs. 5.2 and 5.3 discuss checklist items that are specific to either the Einstein or Green-Kubo approaches, respectively, for estimating the self-diffusion coefficient. Sec. 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 and 2 of Ref. [38]). 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. [38]). 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 is 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. Calculated self-diffusivities increase with increasing system size, as can be seen in Figure 1 from Ref. [39] 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.

The first approach is to perform simulations with progressively larger system sizes, i.e., by increasing the number of molecules, N, and the box length, L, such that the density is constant. The computed self-diffusivities are then plotted as a function of $N^{-1/3}$. As shown in Figure 1, such a plot is approximately linear, and extrapolating to when $N^{-1/3} = 0$ (i.e., $N \to \infty$) gives an estimate of the self-diffusivity (although note that some studies, such as Ref. [40], extrapolate D with respect to N^{-1}). 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 [41]. 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,

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

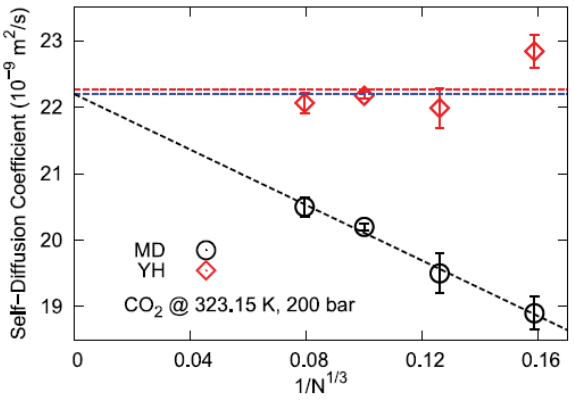

Figure 1. Self-diffusivity obtained with Einstein approach demonstrates significant system size dependence. Reproduced with permission from J. Chem. Phys. 145, 074109 (2016). Copyright 2016 AIP Publishing [39]. Blue dashed lines are obtained by extrapolating the molecular dynamics (MD) results to the infinite system size, i.e., $N^{-1/3} \rightarrow 0$. Red diamonds are the values of D after applying the Yeh-Hummer (YH) finite-size correction, i.e., Equation 3. The red dashed line is an average of these corrected values of D. For further details, see Ref. [39]. Note that 1 bar = 10^5 Pa.

k_B is the Boltzmann constant, T is the absolute temperature, η is the shear viscosity, and ξ = 2.837298 is a dimensionless constant determined by an Ewald-like summation of a periodic lattice. The shear viscosity must be computed separately but fortunately, η is not typically a strong function of system size (see Sec. 6.1.3).

As can be seen in Figure 1, 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 [42]. Also note that a different correction may be more appropriate for anisotropic condensed-phase systems (e.g., those containing membranes), discussed in Sec. 5.4.

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

The necessary simulation length depends on the number of molecules, where fewer molecules require 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 2). As mentioned in Sec. 4.3, another heuristic is whether the square root of the MSD is sufficiently large, i.e., larger than the radius of gyration of the molecule at the low end and larger than 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 to approximate the long-time slope. 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 log-log plot, borrowed from Ref. [36], is provided in Figure 2, where the linear regression to the "middle" region is included. From visual inspection, the "ballistic" short-time interval ranges from the beginning of the simulation to approximately 100 ps. The "middle" region is identified by the linear regime with a slope of 1 (for a log-log plot) spanning from approximately 100 ps to 1000 ps. Note that the noisy "long-time" simulation data (beyond 1000 ps) are not depicted in Figure 2 and are excluded from the linear fit.

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. Note that compared to the position information required by the Einstein approach, velocity information must be stored at a much higher frequency for the Green-Kubo approach. This is because the velocity autocorrelation function (VACF) that

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

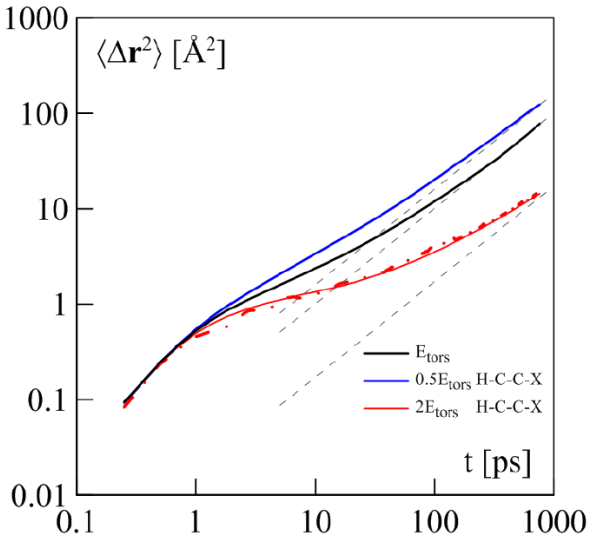

Figure 2. Log-log plot of MSD with respect to time is used to identify the "middle" region. Reproduced with permission from J. Phys. Conf. Ser. 774 (2016) 012039, under the Creative Commons Attribution 3.0 license [36]. The line colors correspond to different torsional energies (E_{tors}). The dash-dotted and solid red lines represent the same simulation results averaged over one and sixty time origins, respectively. The gray dashed lines are linear fits to the corresponding diffusive regimes, as determined by the authors. For further details, see Ref. [36]. Note that 1 Å = 10^{-10} m.

must be integrated decays very rapidly and fine time resolution is needed to perform an accurate numerical integration. For this reason, we recommend saving the velocities at least every 4 to 5 fs. Be warned that this will result in trajectory files that are significantly larger than the positional trajectories required for the Einstein approach, which are saved at lower frequencies (see Sec. 5.2.1).

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 (see Sec. 6.1.2 for a more thorough discussion regarding the relationship between simulation times and plateau times).

A trade-off exists when deciding the elapsed time between subsequent time origins (often referred to as the "lag"). Specifically, if the lag is too short the VACF will not converge to 0 and, thus, the Green-Kubo integral will not converge to the appropriate finite value. By contrast, if the lag is too long the long-time VACF "tail" will introduce considerable noise in the integration. A good rule of thumb for ensuring that the calculated VACF has been sampled adequately is to perform simulations that are approximately a factor of 20 times longer than the estimated decay-to-zero time. The lag between time origins should then be greater than the decay-to-zero time but not exceed half the total simulation time. For example, water models at liquid conditions typically exhibit a decay-to-zero time on the order of 1 ps, such that a simulation of 20 ps with a lag time between 1 and 10 ps should yield a well converged VACF (see Figures 12-15 of Ref. [43]).

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, cut-off times, etc.). The running integral versus time is calculated and the self-diffusivity is estimated from the plateau value. Since the long-time "tail" of the VACF can contribute significantly to the integral [2, 10], care must be taken to ensure that the VACF properly converges to 0 at long times. As with the Einstein approach, a cut-off time should be chosen to determine 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

5.4.1 Considerations for membrane systems

Although molecular simulation is well-suited for predicting diffusivity coefficients in membranes, several issues arise that require special attention. For example, the standard non-bonded long-range cut-off corrections are not straightforward in a heterogeneous system. For this reason, it is common to modify the non-bonded interactions such that tail corrections are not needed, e.g., cut-and-shift, force-switch [2, 5]. However, it is important to investigate the impact modifying the non-bonded potential has on the diffusion coefficients.

Furthermore, since membranes require anisotropic pressure control it is important to use barostats/thermostats that maintain the correct isobaric/isothermal ensemble. Venable *et al.* have shown that as long as a high-quality barostat and thermostat are used (such as those implemented via so-called "extended system" methods [26, 44]), diffusion coefficients determined from *NVE* and *NPT* simulations are quite similar [45].

Finally, correcting diffusion coefficients for finite size effects (i.e., when periodic boundary conditions are employed) in membrane systems requires some additional consideration. Camley *et al.* develop a method for determining diffusion coefficients in membrane systems which uses the immersed-boundary approach in the context of the the Saffman-Delbrüuck model [46]. This method is available via https://diffusion.lobos.nih.gov.

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

6 Viscosity

Although the popularity of NEMD methods for predicting shear viscosity has increased in recent years, Ref. [18] 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 20×10^{-3} Pa-s, although EMD has been successfully implemented for systems near 50×10^{-3} Pa-s. 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. The Green-Kubo approach appears to be the most popular EMD method found in the literature and, more importantly, less arbitrary data analysis methods exist that improve the reliability and reproducibility (see Sec. 6.2.1).

We should note, however, that Ref. [19] states 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 integrals over shorter times." Although this observation is true (see Figure 9), several algorithmic advances have been implemented with the Green-Kubo approach since 2002 (when Ref. [19] was published). Most of these improvements rely on performing large amounts of replicate simulations (see Sec. 6.1.4). Therefore, we only recommend utilizing the Einstein approach when replicate simulations are too computationally expensive (see Sec. 6.3.1).

Sec. 6.1 discusses shear viscosity checklist items that apply to both the Einstein and Green-Kubo approaches. Secs. 6.2 and 6.3 discuss checklist items specific to the Green-Kubo and Einstein approaches, respectively, for estimating viscosity. Sec. 6.4 provides a brief discussion of some topics that are relevant in certain applications.

6.1 Viscosity: General

6.1.1 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 recommend writing the stress tensor every 5 to 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.

6.1.2 Simulation length

Since viscosity is a collective property, i.e., the stress tensor depends on all N molecules, about ten times more data are required to compute viscosity than diffusivity. 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 Sec. 4.3 to determine the length of the simulation required.

Figure 3, borrowed from Ref. [30], 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 *a priori*, we recommend performing an analysis similar to that shown in Figure 3 to ensure adequate sampling.

It is important not to confuse the Green-Kubo integration time (the abscissa for the top panel of Figure 3) with the simulation length (the different color lines in both panels of Figure 3). 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 which the Green-Kubo integral reaches a plateau (around 100 ps in the top panel of Figure 3) 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, e.g., Figure 3 demonstrates that trajectories of 1 ns or longer are required for a plateau time of approximately 100 ps.

Figure 4, borrowed from Ref. [30], 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.

6.1.3 Finite size effects

Finite size effects can arise in small, dense systems due to limited space for configurational rearrangements [47]. Kim et al. [47] demonstrate that a complex oscillatory relationship exists between the shear viscosity of dense fluids and $V^{-1/3}$, where the oscillations dampen with increasing system size (see Figures 8 and 9 of Ref. [47]). However, extrapolation to the infinite system size viscosity is not feasible due to the non-linear scaling behavior with respect to $V^{-1/3}$.

Fortunately, the error introduced by neglecting finite size effects is significantly smaller than that for self-diffusivity (recall Figure 1). For example, Figures 5-6 from Refs. [39]

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

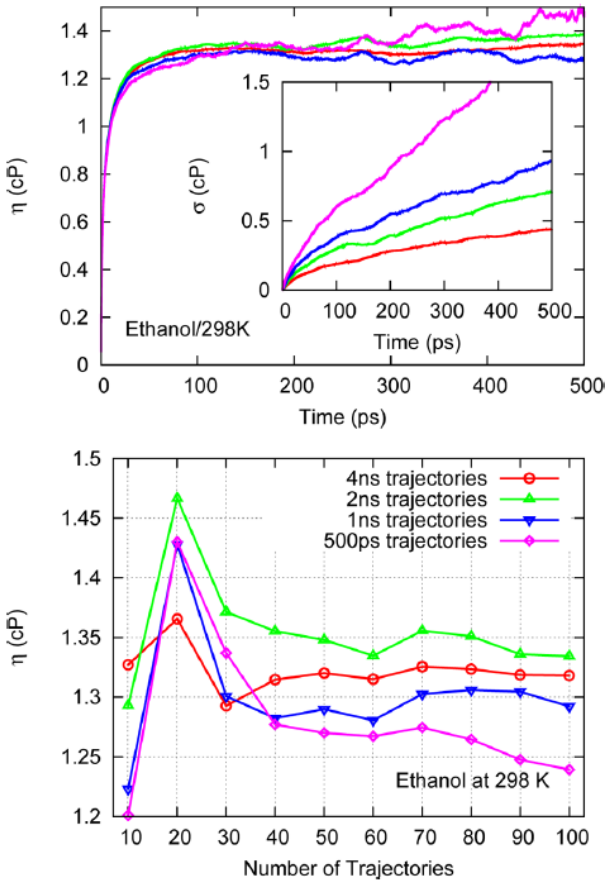

Figure 3. Viscosity (η) and corresponding standard deviation (σ) depend on simulation length. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society Ref. [30]. Different lines and symbols correspond to different simulation lengths, i.e., trajectory times. The inset in the top panel plots the standard deviation, σ. For further details, see Ref. [30]. Note that 1 cP = 10^{-3} Pa-s.

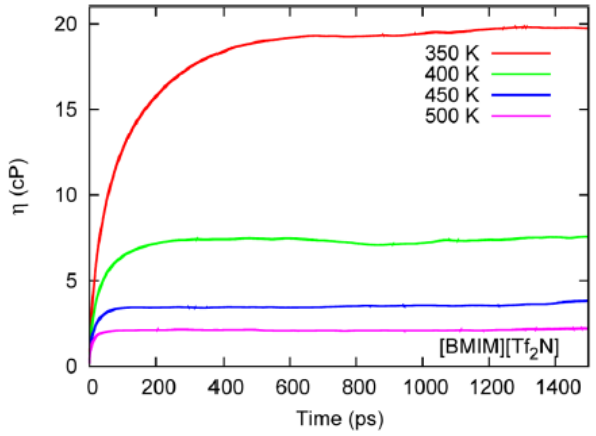

Figure 4. Plateau time increases with increasing viscosity (decreasing temperature). Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society Ref. [30]. Different lines correspond to different temperatures and, thus, different viscosities. For further details, see Ref. [30]. Note that 1 cP = 10^{-3} Pa-s.

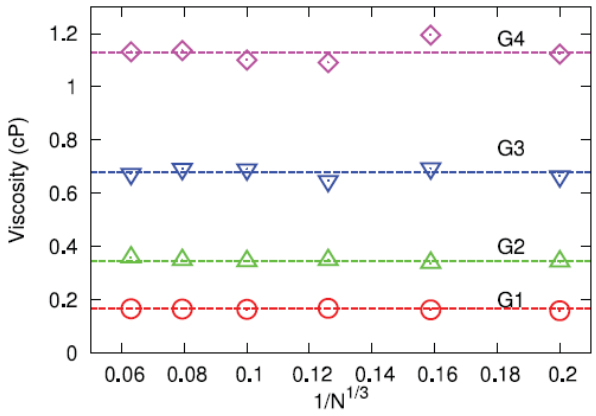

Figure 5. Finite size effects are negligible for viscosity obtained with Green-Kubo approach. Reproduced with permission from J. Chem. Phys. 145, 074109 (2016). Copyright 2016 AIP Publishing [39]. Different symbols correspond to different types of glymes (G*i*). Dashed lines are average value for each glyme from various system sizes (*N*). For further details, see Ref. [39]. Note that 1 cP = 10^{-3} Pa-s.

and [30], 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 [40], also report that shear viscosity has a weak dependence on system size (see Figure 4 of Ref. [40]). 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 $N^{-1/3}$ (or $V^{-1/3}$), 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.

6.1.4 Improved precision

To improve statistical averaging, it is common to include multiple terms from the stress tensor. For example, Figure 7, borrowed from Ref. [19], 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 [48, 49], which uses all six independent components of the stress tensor. Details are given in the Appendix of Ref.

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

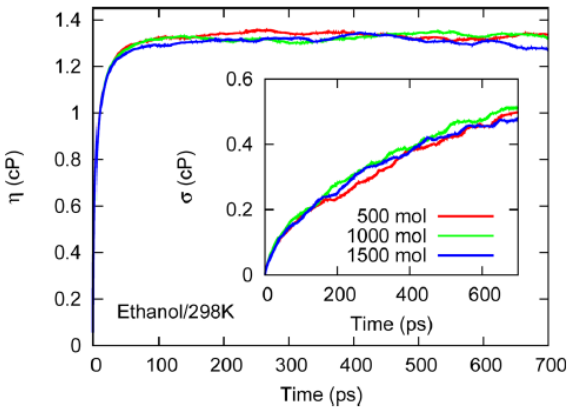

Figure 6. System size does not affect viscosity (η) obtained with Green-Kubo approach or the corresponding standard deviation (σ). Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society Ref. [30]. Different colors correspond to different numbers of molecules. The inset plots the standard deviation, σ. For further details, see Ref. [30]. Note that 1 cP = 10^{-3} Pa-s.

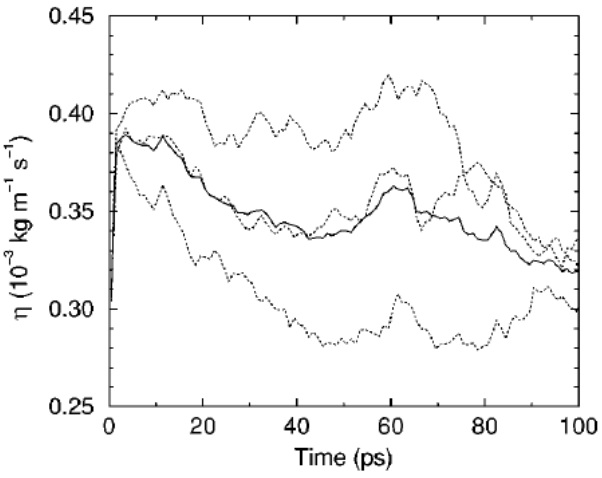

Figure 7. Averaging three off-diagonal elements of pressure tensor improves precision of Green-Kubo viscosity estimate (η). Reproduced from J. Chem. Phys., 2002, 116(1):209–217, with the permission of AIP Publishing [19]. Dashed lines represent a single off-diagonal element of the pressure tensor while solid line is the average of the three off-diagonal elements. For further details, see Ref. [19].

[49]. This generalized integral is given by

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

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

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

where δ is the unit tensor. Note that the factor of 10 in the denominator of Equation 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 [37, 38, 50] (although some authors have argued for an equal weighting [18], which would modify the normalization factor in the denominator of Equation 4 from 10 to 9). The equivalent generalization of the Einstein relation is

$$\eta = \frac{V}{20k_BT}\lim_{t\to\infty}\frac{d}{dt}\sum_i\sum_j\left\langle\left(\int_0^t \tau_{ij}^{os}(t')dt'\right)^2\right\rangle_{t_0} \quad (6)$$

We are not aware of any studies that rigorously quantify the improvement in precision obtained by using all six terms. Figure 8, borrowed from Ref. [18], demonstrates that the average viscosity and fluctuations are nearly identical when using the three off-diagonal terms or when using six terms. Therefore, although we recommend including all six terms, it is typically sufficient to utilize just the three off-diagonal terms. Regardless, it is important to clearly state which terms are included when computing viscosity.

Although fluctuations in η are significantly reduced by including multiple terms from the stress tensor, the key to

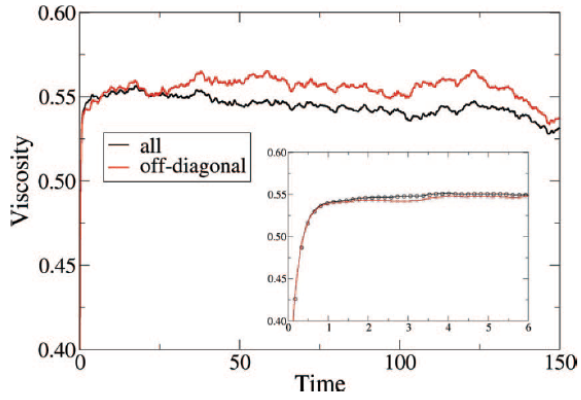

Figure 8. Green-Kubo viscosity is practically equivalent when averaging all six or just the three off-diagonal pressure tensor elements. Results are for Lennard-Jones (LJ) fluid (in reduced LJ units, i.e., $\eta^* = \frac{\eta\sigma^2}{\sqrt{m\epsilon}}$ and $t^* = t\sqrt{\frac{\epsilon}{m\sigma^2}}$, where σ and ϵ are the respective LJ size and energy parameters and m is the mass of the LJ particle). Reproduced from J. Chem. Phys., 131, 246101 (2009), with the permission of AIP Publishing [18]. The inset focuses on the short-time interval where the initial plateau is observed. 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. [18].

improved precision of viscosity estimates is to average several replicate simulations. For N_{reps} replicates, the Green-Kubo equation is

$$\eta = \frac{V}{10k_BTN_{reps}}\sum_{n=1}^{N_{reps}}\sum_i\sum_j\int_0^\infty \left\langle \tau_{ij,n}^{os}(0)\tau_{ij,n}^{os}(t)\right\rangle_{t_0} dt \quad (7)$$

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

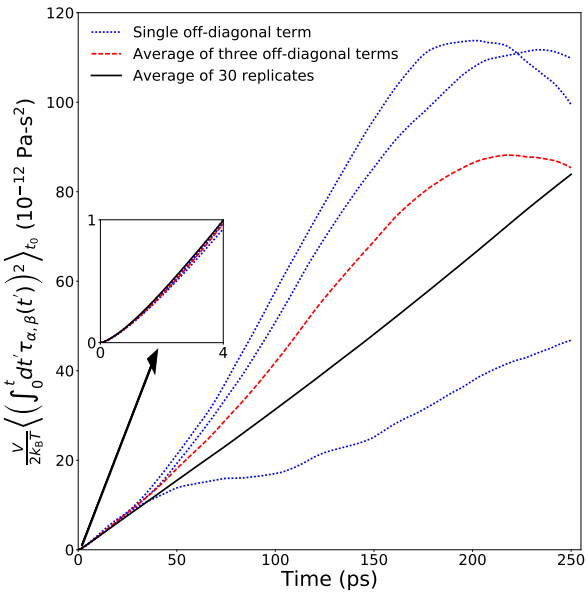

Figure 9. Einstein approach for viscosity is improved by averaging three stress tensor terms and 30 replicate runs. Simulations were performed with Gromacs 2018 for saturated liquid ethane at 137 K using the TraPPE-UA model [51]. Simulation details: velocity Verlet integrator [52], 2 fs time-step, 6 fs pressure tensor output frequency, 12 time origins, 1 ns equilibration time, 1 ns production time, 400 molecule system, 1.4 nm non-bonded cut-off distance with analytic tail corrections, Nosé-Hoover thermostat with time constant of 1 ps [44], bond-lengths constrained with a linear constraint solver (LINCS) algorithm [2, 53, 54].

and the Einstein relation becomes

$$\eta = \frac{V}{20 k_B T N_{\text{reps}}} \lim_{t\to\infty} \frac{d}{dt} \sum_{n=1}^{N_{\text{reps}}} \sum_i \sum_j \left\langle \int_0^t \tau_{ij,n}^{\text{os}}(t')dt' \right\rangle_{t_0} \quad (8)$$

For example, Figure 9 demonstrates that averaging three stress tensors is not sufficient to obtain a reliable Einstein slope as $t \to \infty$. By contrast, a near linear trend at high time is observed by averaging a large number of replicates (N_{reps} = 30).

The number of replicates used in the literature varies widely. In their study of the shear viscosity of alkanes, Payal and co-workers [29] used 10 replicates, whereas Zhang *et al.* [30] performed a systematic investigation of the minimal number of replicates required for convergence. They observed that a value of 30 to 40 replicates was statistically equivalent to 100 replicates for their system. However, the necessary number of replicates depends on the system. Specifically, 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 to 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 discuss each approach below but recommend the latter methodology as it requires fewer arbitrary user decisions and is more robust than the other two methods.

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 to 100 ps) is typically the region of choice, see Refs. [18, 22]. However, it is important to explain how this time interval was selected (e.g., visual inspection, test of convergence, magnitude of fluctuations) and to quantify how much the estimated viscosity changes if the time interval were modified. For example, in Figures 7 and 8 the reported viscosity would likely be the average from approximately 5 to 15 ps and 10 to 25 dimensionless time units, respectively. Clearly, the estimated viscosity would vary significantly if the average included long-time data.

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 [2, 55]

$$\frac{S_{\text{ACF}}^{\text{f}}(t)}{S_{\text{ACF}}^{\text{f}}(0)} = (1 - C)\cos(\omega t)\exp\left(-t/\tau_f\right)^{\beta_f} + C\exp\left(-t/\tau_s\right)^{\beta_s} \quad (9)$$

where C, ω, τ_f, τ_s, β_f, β_s (and sometimes $S_{\text{ACF}}^{\text{f}}(0)$) are fitting parameters. Specifically, 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, $S_{\text{ACF}}^{\text{f}}(t)$ is the stress autocorrelation function at time t, and $S_{\text{ACF}}^{\text{f}}(0)$ is the initial (time-zero) autocorrelation function.

Figure 10, from Ref. [22], demonstrates that Equation 9 has the correct shape to fit the stress autocorrelation function for this system. However, notice the significant deviation between the model fit ($S_{\text{ACF}}^{\text{f}}$) and the raw simulation output

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

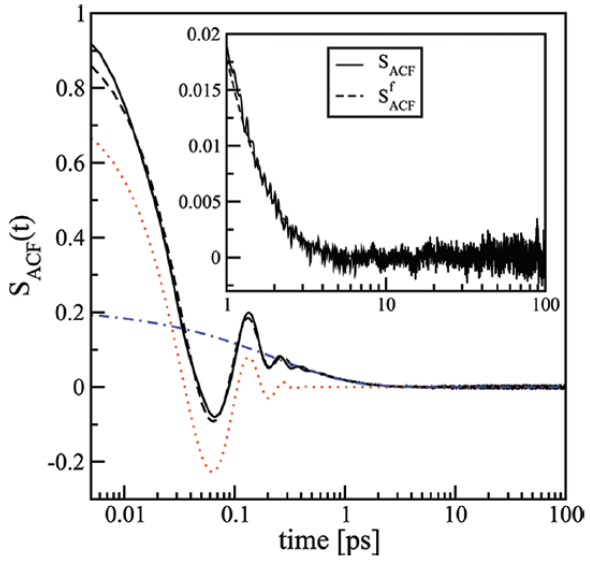

Figure 10. Equation 9 provides reliable fit of autocorrelation function. Reproduced with permission from J. Phys. Chem. A., 2012, 116 (10), pp 2564–2570. Copyright (2012) ACS Publications [22]. S_{ACF} and S^f_{ACF} correspond to the raw autocorrelation function and the fit to Equation 9, respectively. 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 9. The inset highlights the large long-time fluctuations. For further details, see Ref. [22].

(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^f_{ACF}, to infinite time [23, 56, 57]. The advantage of this hybrid (combined) 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 hybrid integration approach is especially preferred when S_{ACF} is highly oscillatory, such as that shown in Figure 7 of Ref. [57], where Equation 9 is likely inadequate.

The time where the Green-Kubo integration switches from using S_{ACF} to S^f_{ACF}, referred to as the switch-time (t_s), should be after the "fast" autocorrelation component has dissipated (the first term in Equation 9 and the red dotted line in Figure 10). As this time depends on the system and user judgment, examples of t_s in the literature range widely. For example, t_s = 5 ps in Refs. [23, 56] while the switch-time value is only 0.015 ps in Ref. [57].

A simpler exponential decay function than Equation 9 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 $t > t_s$. For example, it is common to fit S_{ACF} values for $t > t_s$ to a single exponential term, e.g., $S^f_{ACF} = \exp(-(t/\tau)^\beta)$ [56], $S^f_{ACF} = \exp(-t/\tau)$ [23], or $S^f_{ACF} = a\exp(-t/b)$ [57], where τ, β, a, and b are fitting parameters.

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 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 9) is that uncertainties in the model fit do not propagate through the integration.

For example, Refs. [58] and [30] recommend fitting the "running integral" to a double-exponential function

$$\eta(t) = A\alpha\tau_1\,(1 - \exp\,(-t/\tau_1)) + A(1-\alpha)\tau_2\,(1 - \exp\,(-t/\tau_2)) \quad (10)$$

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. [50] proposes an alternative model by integrating the slow stretched-exponential function (second term in Equation 9) which results in the expression

$$\eta(t) = \eta_\infty(1 - \exp(-(t/\tau_s)^{\beta_s})) \quad (11)$$

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 10 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 11. That being said, the η_∞ estimates obtained with Equations 10 and 11 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 10 or 11 is implemented, it is important

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

to include a description of how the fit is performed, e.g., the objective function, range of data included.

Ref. [30] recommends that the data be weighted by the inverse of the standard deviation (σ) from the replicate simulation values. As σ increases significantly with respect to time, Ref. [30] fit σ to the 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 11, borrowed from Ref. [30], 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 obtained from the σ fit. Note that Ref. [58] recommends a value of b = 2. Ref. [30] demonstrates that for b = 2 the estimated value of η for an ionic liquid ([BMIM][Tf_2N]) at 350 K is approximately 11 $\times$ 10^{-3} Pa-s (compared to $\approx$ 19 $\times$ 10^{-3} Pa-s in the bottom panel of Figure 11).

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 10 with each value of b from the distribution generated in the previous step.

Ref. [30] also suggests that a cut-off time be implemented to improve the fit. They provide a heuristic 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, since the magnitude of variability depends strongly on the system, we recommend that the author quantify the cut-off time dependence.

Furthermore, Ref. [30] recommends excluding short-time data from the fitting procedure. In Figure 12, borrowed from Ref. [30], 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 region from the model fitting.

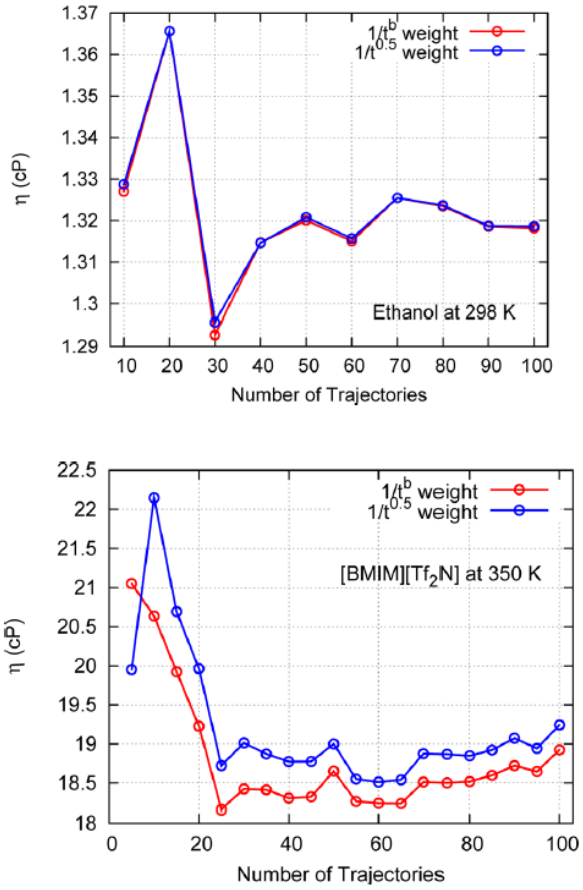

Figure 11. The Green-Kubo viscosity estimate depends on the exponent of the weighting model, b, which is approximately 0.5 for certain systems. Note that b = 0.52 for ethanol at 298 K, top panel, while b is between 0.60 to 0.73 for [BMIM][Tf_2N] at 350 K, bottom panel. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society. For further details, see Ref. [30]. Note that 1 cP = 10^{-3} Pa-s.

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. Unfortunately, by contrast with self-diffusivity, the long-time trend from a single run is often non-linear (recall Figure 9). Replicate simulations are typically necessary to obtain a well-behaved long-time trend. For example, Figure 9 demonstrates that the replicate-averaged Einstein integral is approximately linear over a large time interval (using 30 replicate simulations). Therefore, if sufficient replicates are used it is possible to compute a reliable slope (viscosity) at the long-time limit.

However, as observed in the inset of Figure 9, the Einstein integral becomes nearly linear for a single simulation after a few ps. For these reasons, it is common to fit the slope from a single simulation over an intermediate time interval, e.g.,

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

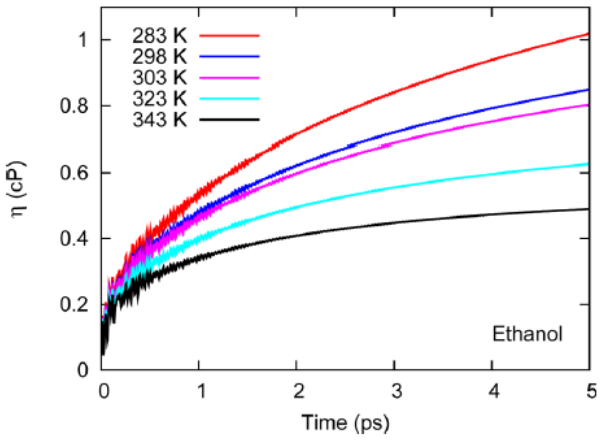

Figure 12. Green-Kubo "running integral" demonstrates large fluctuations at very short time. Reprinted with permission from J. Chem. Theory Comput., 2015, 11 (8), pp 3537–3546. Copyright (2015) American Chemical Society. For further details, see Ref. [30]. Note that 1 cP = 10^{-3} Pa-s.

5 to 50 ps. We only recommend calculating the slope from an intermediate time interval if performing a large number of replicate simulations is too computationally expensive. As mentioned in Sec. 6, the Einstein approach is likely preferred over the Green-Kubo approach only in this scenario. With a single simulation (or a small number of replicate simulations), we recommend that the author explain why the slope was calculated using a given intermediate time interval and how much variability is introduced if a different region is selected.

For example, Figure 13 helps visualize the uncertainty in η due to the time interval used to compute the slope. Panels a) and b) are obtained from the respective single and 30 replicate simulation results presented previously in Figure 9. Panel a) demonstrates that determining η from a single run depends strongly on the time interval, while Panel b) shows that η is much less dependent on the time interval when obtained from 30 replicate simulations.

Using the slope for an intermediate time interval is less theoretically rigorous and depends strongly on user judgment. Therefore, when computationally feasible, we recommend averaging the Einstein integral for multiple replicate simulations, i.e., Equation 8. 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 3) 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.

Similar to the Green-Kubo recommendation, we also recommend bootstrapping the uncertainty for the Einstein approach (see Sec. 4.2.3). This is done by randomly sampling

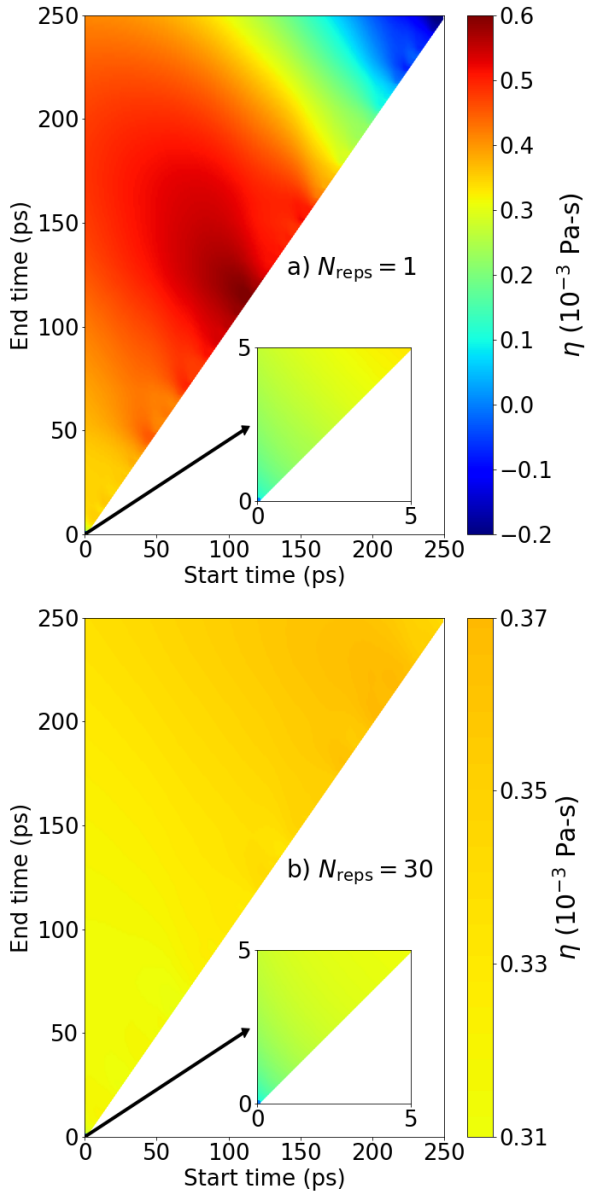

Figure 13. Einstein viscosity is highly sensitive to the time interval used to compute the slope. For example, "start time" = 50 ps and "end time" = 200 ps computes the slope from t = 50 to 200 ps of $\frac{V}{2k_B T N_{reps}} \sum_{n=1}^{N_{reps}} \left\langle \int_0^t \tau_{\alpha\beta,n}(t')dt' \right\rangle_{t_0}$. Panels a) and b) correspond to a single simulation and 30 replicate simulations, respectively. The same color scale is used in both panels, but Panel b) also includes a zoomed-in color bar. Simulation data are the same as those presented in Figure 9.

which replicates are included in the replicate-averaged Einstein integral, calculating the viscosity from the slope, and producing a distribution of these viscosity values from hundreds or thousands of different random replicate sets.

In addition, we recommend bootstrapping the time in-

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

terval uncertainty, i.e., computing the slope for hundred or thousands of different time intervals. This approach is especially important if analyzing intermediate time intervals from a single simulation. For example, from Figure 13 Panel a) we would report the distribution of η values obtained from randomly selected time intervals with a "start time" greater than 5 ps and an "end time" less than 50 ps.

6.4 Viscosity: Special topics

6.4.1 Considerations for cut-off length

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." [2, 19] Therefore, when computing viscosities for systems with electrostatics, it is extremely important to investigate the effect of cut-off distance. This can be done by performing simulations with variable Coulombic cut-off lengths, typically 0.9, 1.2, and 1.4 nm. We also recommend investigating whether the estimated viscosity depends strongly on the algorithm for computing the electrostatics, e.g., particle mesh Ewald. [2, 59]

Similarly, we suspect that the viscosity also depends on the van der Waals cut-off length and/or if a cut-and-shift, force-switch, or truncated (with or without tail corrections) potential is implemented [2, 60, 61]. For these reasons, we strongly recommend reporting how the non-bonded (electrostatics and van der Waals) interactions are computed.

7 Conclusions

Molecular simulation is commonly used to predict transport properties. However, without careful simulation design and acute analysis, results may not be meaningful. This work outlines the best practices in the design and analysis of equilibrium molecular dynamics simulations for self-diffusivity and viscosity prediction. It is worth noting that non-equilibrium molecular dynamics simulations are highly preferred for moderate to high viscosity materials (more than 20×10^{-3} to 50×10^{-3} Pa-s). We anticipate future studies discussing best practices for NEMD techniques or additional transport properties.

In liquid systems, we suggest production simulations use either the *NVE* ensemble or the *NVT* ensemble, after judiciously selecting the thermostat and coupling strength [22, 23]. By contrast, we strongly discourage the use of the *NPT* ensemble due to the potential interference of the barostat with the system dynamics. We recommend averaging over multiple time origins to improve precision of the self-diffusivity and viscosity estimates. To ensure simulations are run long enough for the system dynamics to be fully emulated, we recommend the user run a series of simulations at differing lengths, and observe deviations in estimated self-diffusivity or viscosity with changes in simulation time. In addition, the degree of configuration space exploration can be estimated by calculating the mean squared displacements of the molecules and comparing to the radius of gyration and box length. The root MSD should be greater than the radius of gyration, and ideally, on the order of the box length. We also recommend running multiple independent simulations to more thoroughly sample the possible states. Measures should be taken to rigorously estimate the uncertainty of the self-diffusivity and viscosity prediction, where we recommend bootstrapping the replicate simulation results.

For self-diffusivity, we recommend employing the Einstein method. We recommend that the atomic positions be output 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. We discuss two methods to correct for system size effects, which are significant in self-diffusivity calculations. In post-simulation analysis, we recommend improving precision by averaging the velocity autocorrelation function over all molecules and the diagonal self-diffusivity dimensional components. 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.

For viscosity, we recommend the Green-Kubo approach, although the Einstein method may be preferred with certain systems, i.e., if replicate simulations are too computationally expensive. As a collective property, viscosity requires significantly more data and replicates than self-diffusivity. We recommend outputting the stress tensor components every 5 to 10 fs. The number of replicates a system needs can vary greatly depending on the compound, number of molecules, and temperature. The simulation length of each trajectory should typically be at least an order of magnitude greater than the Green-Kubo integral plateau time. Due to their slower dynamics, more viscous materials require longer simulation times. System size seems to be of little impact on viscosity prediction (see Figures 5-6), however, we still recommend that the user justify their choice of system size by plotting $N^{-1/3}$ versus predicted viscosity. In post-simulation data processing, we recommend averaging over either the three off-diagonal or all six independent components of the stress tensor to enhance precision. We recommend fitting the Green-Kubo running integral to a double-exponential function and extrapolating to the infinite-time limit [30].

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324

Acknowledgments

We are grateful for the internal review provided by Alta Y. Fang and Arno R. Laesecke of the National Institute of Standards and Technology (NIST). This research was performed while R.A.M. held a National Research Council (NRC) Postdoctoral Research Associateship at NIST.

Disclaimer

Commercial equipment, instruments, or materials are identified only in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by NIST, nor does it imply that the products identified are necessarily the best available for the intended purpose.

Potentially Conflicting Interests

The authors declare no competing financial interests.

Author Information

ORCID:

Edward J. Maginn: 0000-0002-6309-1347
Richard A. Messerly: 0000-0003-1321-4246
Daniel J. Carlson: 0000-0002-6610-414X
Daniel R. Roe: 0000-0002-5834-2447
J. Richard Elliott: 0000-0002-1159-4786

References

[1] **LAMMPS**. LAMMPS Users Manual. Sandia National Laboratories; 2017, http://lammps.sandia.gov/doc/Manual.html.

[2] **GROMACS**. GROMACS Reference Manual; 2018, http://manual.gromacs.org/documentation/2018/index.html.

[3] **Case DA**, Brozell SR, Cerutti DS, Cheatham TE, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, et al., AMBER 2018, University of California, San Francisco.;. Accessed: 2018-05-4. http://ambermd.org/.

[4] **Phillips JC**, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, , Schulten K. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry. 2005; 26(1):1781 –1802. https://doi.org/10.1002/jcc.20289, NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. http://www.ks.uiuc.edu/Research/namd/.

[5] **Brooks BR**, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, et al. CHARMM: The biomolecular simulation program. Journal of Computational Chemistry. 2009; 30(10):1545–1614. https://doi.org/10.1002/jcc.21287.

[6] **Rutkai G**, Köster A, Guevara-Carrion G, Janzen T, Schappals M, Glass CW, Bernreuther M, Wafai A, Stephan S, Kohns M, Reiser S, Deublein S, Horsch M, Hasse H, Vrabec J. ms2: A molecular simulation tool for thermodynamic properties, release 3.0. Computer Physics Communications. 2017; 221:343 – 351. https://doi.org/10.1016/j.cpc.2017.07.025.

[7] **Roe DR**, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation. 2013; 9(7):3084–3095. https://doi.org/10.1021/ct400341p.

[8] **McGibbon RT**, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophysical Journal. 2015; 109(8):1528 – 1532. https://doi.org/10.1016/j.bpj.2015.08.015.

[9] **Humphrey W**, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996; 14(1):33 – 38. https://doi.org/10.1016/0263-7855(96)00018-5.

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

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

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

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

[14] 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.

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

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

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

[18] **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.

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

[20] **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. https://doi.org/10.1021/acs.chemrev.5b00215.

[21] **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. https://doi.org/10.1016/j.molliq.2006.12.019.

[22] **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. https://doi.org/10.1021/jp211952y.

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

[24] **Berendsen HJC**, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics. 1984; 81(8):3684–3690. https://doi.org/10.1063/1.448118.

[25] **Bussi G**, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. The Journal of Chemical Physics. 2007; 126(1):014101. https://doi.org/10.1063/1.2408420.

[26] **Nosé S**. A unified formulation of the constant temperature molecular dynamics methods. The Journal of Chemical Physics. 1984; 81(1):511–519. https://doi.org/10.1063/1.447334.

[27] **Andersen HC**. Molecular dynamics simulations at constant pressure and/or temperature. The Journal of Chemical Physics. 1980; 72(4):2384–2393. https://doi.org/10.1063/1.439486.

[28] **Hess B**, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. Journal of Chemical Theory and Computation. 2008; 4(3):435–447. https://doi.org/10.1021/ct700301q.

[29] **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.

[30] **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. https://doi.org/10.1021/acs.jctc.5b00351.

[31] **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.

[32] **Grossfield A**, Patrone PN, Roe DR, Schultz AJ, Siderius DW, Zuckerman DM. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations: v1.0. Living Journal of Computational Molecular Science. 2018; A LiveCoMS Best Practices Guide. https://github.com/dmzuckerman/Sampling-Uncertainty/blob/master/main.pdf.

[33] **Taylor BN**, Kuyatt CE. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results (1994 Edition). NIST Technical Note 1297; 1994.

[34] **Efron B**. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics. 1979; 7(1):1–26. https://doi.org/10.1214/aos/1176344552.

[35] Shen VK, Siderius DW, Krekelberg WP, Hatch HW, editors. NIST Standard Reference Simulation Website, NIST Standard Reference Database Number 173. Gaithersburg, MD, 20899: National Institute of Standards and Technology; 2006. https://doi.org/10.18434/T4M88Q.

[36] **Kondratyuk ND**, Norman GE, Stegailov VV. Rheology of liquid n-triacontane: Molecular dynamics simulation. Journal of Physics: Conference Series. 2016; 774(1):012039.

[37] **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. https://doi.org/10.1021/ie300222a.

[38] **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.

[39] **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. https://doi.org/10.1063/1.4960776.

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

[41] **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. https://doi.org/10.1021/jp0477147.

[42] **Kikugawa G**, Nakano T, Ohara T. Hydrodynamic consideration of the finite size effect on the self-diffusion coefficient in a periodic rectangular parallelepiped system. The Journal of Chemical Physics. 2015; 143(2). https://doi.org/{10.1063/1.4926841}.

[43] **Stillinger FH**, Rahman A. Improved simulation of liquid water by molecular dynamics. The Journal of Chemical Physics. 1974; 60(4):1545–1557. https://doi.org/10.1063/1.1681229.

[44] **Hoover WG**. Canonical dynamics: Equilibrium phase-space distributions. Physical Review A. 1985; 31:1695–1697. https://doi.org/10.1103/PhysRevA.31.1695.

[45] **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.

[46] **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.

[47] **Kim KS**, Han MH, Kim C, Li Z, Karniadakis GE, Lee EK. Nature of intrinsic uncertainties in equilibrium molecular dynamics estimation of shear viscosity for simple and complex fluids. The Journal of Chemical Physics. 2018; 149(4):044510. https://doi.org/10.1063/1.5035119.

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

[49] **Daivis PJ**, Evans DJ. Comparison of constant pressure and constant volume nonequilibrium simulations of sheared model decane. The Journal of Chemical Physics. 1994; 100(1):541–547.

[50] **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.

[51] **Martin MG**, Siepmann JI. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. The Journal of Physical Chemistry B. 1998; 102(14):2569–2577. https://doi.org/10.1021/Jp972543+.

[52] **Swope WC**, Andersen HC, Berens PH, Wilson KR. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. The Journal of Chemical Physics. 1982; 76(1):637–649. https://doi.org/10.1063/1.442716.

[53] **Hess B**, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry. 1998; 18(12):1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[54] **Hess B**. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal of Chemical Theory and Computation. 2008; 4(1):116–122. https://doi.org/10.1021/ct700200b.

[55] **Guo GJ**, Zhang YG, Refson K, Zhao YJ. Viscosity and stress autocorrelation function in supercooled water: a molecular dynamics study. Molecular Physics. 2002; 100(16):2617–2627. https://doi.org/10.1080/00268970210133477.

[56] **van der Spoel D**, van Maaren PJ, Berendsen HJC. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. The Journal of Chemical Physics. 1998; 108(24):10220–10230. https://doi.org/10.1063/1.476482.

[57] **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.

[58] **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. https://doi.org/10.1021/jp062885s.

[59] **Darden T**, York D, Pedersen L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 1993; 98(12):10089–10092. https://doi.org/10.1063/1.464397.

[60] **Levitt M**, Hirshberg M, Sharon R, Daggett V. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Computer Physics Communications. 1995; 91(1):215 – 231. https://doi.org/10.1016/0010-4655(95)00049-L.

[61] **van der Spoel D**, van Maaren PJ. The Origin of Layer Structure Artifacts in Simulations of Liquid Water. Journal of Chemical Theory and Computation. 2006; 2(1):1–11. https://doi.org/10.1021/ct0502256.

https://doi.org/10.33011/livecoms.1.1.6324
Living J. Comp. Mol. Sci. 2019, 1(1), 6324