

Simulation and Analysis of EGM Molecule in Different Phases

Guido Putignano^a, Lorenzo Tarricone^a

^aDepartment of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

Abstract

In the realm of scientific research, the experimental analysis of unknown molecules often presents substantial challenges, both in terms of cost and time. These challenges can lead research activities towards a trial-and-error approach, resulting in a limited number of molecules being explored. As a response to these constraints, computational methods have emerged as a valuable tool for the study and analysis of molecules. In this study, we employ computational techniques to investigate the properties of the EGM molecule. After constructing the molecular topology and conducting simulations of the EGM molecule, our investigation commenced with a comprehensive exploration of its conformational dynamics in response to varying states and environmental contexts. Subsequently, we conducted a rigorous analysis of the molecular trajectory and assessed the convergence of our simulation data. Finally, we engaged in a critical examination of the fundamental assumption underpinning our analysis: the transferability of parameters. We noticed that the agreement with experimental data is quite good concerning the density but falls short when it comes to vaporization heat. We attribute this discrepancy to the fact that our approximation provides a reasonably accurate estimation of mass (and therefore density), but modeling the vaporization heat heavily depends on achieving an accurate representation of molecule-molecule interactions, a level of accuracy we do not attain with this method. Our aim in this study is to gain comprehensive insights into the behavior and properties of the EGM molecule through computational analysis, contributing to the advancement of molecular understanding.

Keywords: Ethylene glycol monoacetate, EGM, Classical Simulation, Biomolecular Systems.

1. Conformational Behavior

The first element we went deeper in our study was the conformation behaviour of EGM's confromation.

1.1. Conformational differences between states

The variability in the conformational dynamics of Ethyl Group Methanesulfonate (EGM) across distinct environmental contexts is predominantly driven by disparities in intermolecular interactions, solvation phenomena, and the molecule's intrinsic degrees of freedom within each medium. These influencing factors assume a pivotal role in shaping and orchestrating the structural and behavioral attributes of EGM under different conditions.

In the gas phase, the EGM molecule exhibits remarkable stability in its conformational state, especially within the initial carbon groups, where the improper dihedral block is situated. Notably, the angle between the hydroxyl group and the remainder of the molecule emerges as the most mutable parameter. This phenomenon is attributed to the elevated conformational freedom stemming from the absence of neighboring molecules and the minimization of steric repulsion effects. Furthermore, upon a focused examination of a single EGM molecule, an observable shift in the center of mass is discernible, proceeding with a nearly constant velocity. Once again, this behaviour is ascribed to the diminished intermolecular interactions inherent to the gas phase environment



Figure 1: Representation in Gas Phase

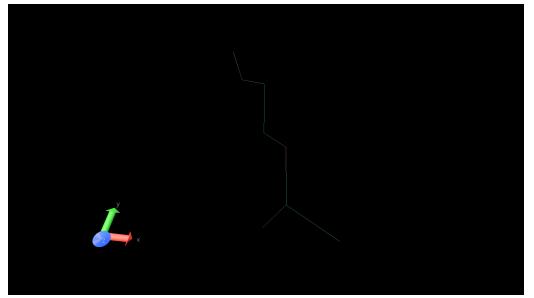


Figure 2: Single molecule in Gas Phase

In the liquid state, the visual field is already densely populated in standard visualization compared to the gas phase, and the center of mass exhibits significantly reduced velocity, often approaching zero, contrasting with the higher velocities observed in the gaseous state. At this juncture, individual EGM molecules undergo movement in a 'diffusive fashion,' where bond angles display reduced oscillation. This observed behavior can be predominantly attributed to the closer proximity of neighboring atoms.

In the liquid phase, EGM molecules find themselves encompassed by neighboring molecules, thereby curbing their rotational freedom and constricting the spectrum of available conformations. In this milieu, intermolecular interactions, notably van der Waals forces and dipole-dipole interactions, emerge as influential factors. These interactions tend to promote a more compact and organized spatial arrangement of EGM molecules.

As a result, EGM molecules in the liquid phase tend to adopt conformations that minimize steric clashes with neighbouring molecules. This preference leads to a more restricted and orderly structural configuration, distinguishing it from the more dynamic and flexible behavior exhibited in the gaseous phase.

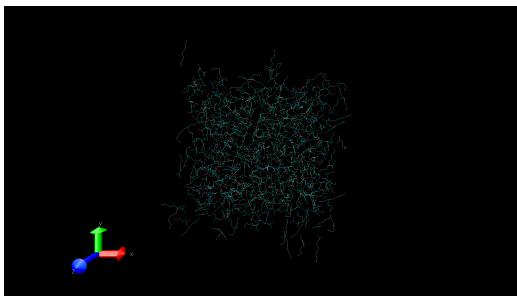


Figure 3: Representation of Liquid state

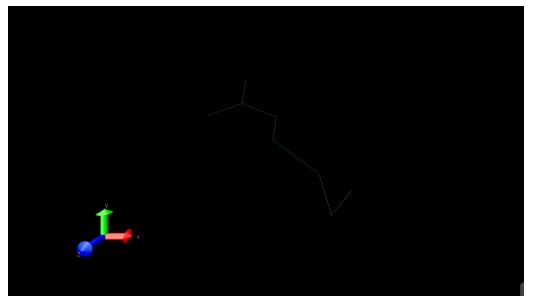


Figure 4: Single molecule in liquid state

In an aqueous environment, the EGM molecule exhibits significantly reduced motion, with its center of mass displaying an average velocity close to zero. This quiescent behavior can be ascribed to the potent interactions between EGM and water molecules, primarily driven by hydrogen bonding.

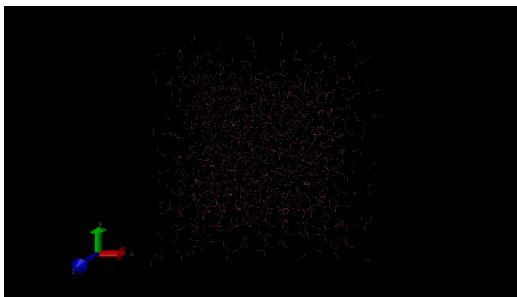


Figure 5: Representation among water molecules



Figure 6: Single molecule among water molecules

When examining the behavior of EGM molecules in different states of matter, a distinct contrast becomes evident. In the gaseous phase, characterized by minimal intermolecular interactions and molecular isolation,

EGM molecules adopt extended and flexible conformations. This flexibility manifests in diverse torsional angles and bond rotations, resulting in a significantly expanded conformational space and uninhibited movement of individual molecules within this environment.

Conversely, in the liquid phase, the presence of substantial intermolecular forces, such as van der Waals interactions and dipole-dipole attractions, induces a more structured and constricted configuration for EGM molecules. Consequently, the spatial mobility of individual molecules is considerably limited when compared to their gaseous phase counterparts.

In the aqueous environment, where EGM molecules establish robust interactions with water molecules, predominantly through hydrogen bonding, a distinctive structural shift occurs. The overall configuration becomes more compact than that observed in the gaseous phase. However, when examining the behavior of an individual EGM molecule in this aqueous milieu, it aligns more closely with the characteristics of the liquid phase, as it continues to be influenced by intermolecular interactions that impart constraints on its spatial mobility.

1.2. Problems with Molecular Visualisations

Direct visualisation of molecules can be useful to understand how the system is behaving, at the same time, it may not be helpful when finding some mathematical ways to predict and understand it. For this reason, there could have been multiple studies and analysis to obtain this result

Table 1 Description of possible quantitative studies

Statistical Tool	Significance
Dihedral angles	Dihedral angles are a measure of the rotation around a bond that can change depending on the conformations that the molecule is adopting.
Radius of gyration	The radius of gyration is a measure of the size and compactness of a molecule to understand how much the molecule is spreading out or contracting.
End-to-end distance	The end-to-end distance is the distance between the two most distant atoms in a molecule to get a sense of how elongated or stretched out the molecule is.
Hydrogen bond count	The hydrogen bond count is the number of hydrogen bonds that a molecule is forming. By calculating the hydrogen bond count of a molecule, we can get a sense of how strongly the molecule is interacting with its surroundings.
Solvent accessible surface area (SASA)	The solvent accessible surface area (SASA) is the surface area of a molecule that is accessible to solvent molecules. By calculating the SASA of a molecule, we can get a sense of how exposed the molecule is to its surroundings.
RMSD between the atoms in the gas phase and in the liquid phase	The RMSD can allow the possibility of visualising the differences in positions of other parameters between two different states, where we consider the average of one of them
Potential energy, kinetic energy, and total energy over time	Potential Energy, Kinetic Energy and Total energy can allow us to understand changes between different environments

2. Convergence Analysis

Our analysis later moved to trajectories

2.1. Potential energy and the density of the trajectories

When considering the analysis of the vaporization enthalpy and the density of the trajectories, there are considerable differences.

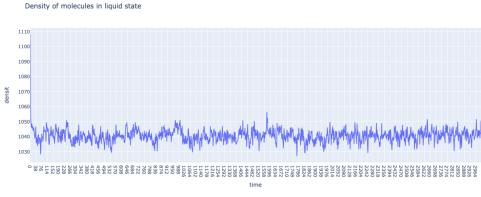


Figure 7: Vaporization enthalpy of molecule in liquid state

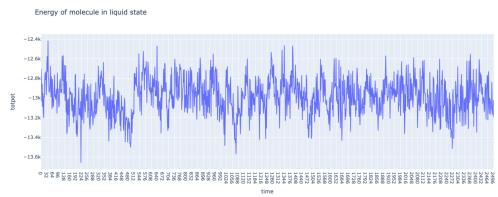


Figure 8: Density of molecule in liquid state

In the context of vaporization enthalpy, measurements are denominated in KJ/mole, with time being quantified in picoseconds (ps). Observing the system from this perspective, it becomes evident that the system undergoes relatively minor alterations compared to its average state. Analyzing the graph representing the heat of vaporization in the liquid state, where energy is plotted as a function of time, a noteworthy observation emerges. The mean energy value remains remarkably stable over time, exhibiting minimal variation. Moreover, the variance associated with this mean value tends to maintain a consistent pattern throughout the entirety of the simulation. Conversely, a divergent trend is observed when examining the density of molecules in the liquid state. In this scenario, one can discern a noticeable decline in density initially, followed by a stabilization at a plateau after approximately 10 simulation steps. This phenomenon can be attributed to the initial disordered state of the molecules, driven by the provided initial conditions, resulting in their substantial separation. This effect is likely due to intermolecular forces that come into play facilitating the formation of bonds among molecules and ultimately culminating in the establishment of a more stable and structured configuration

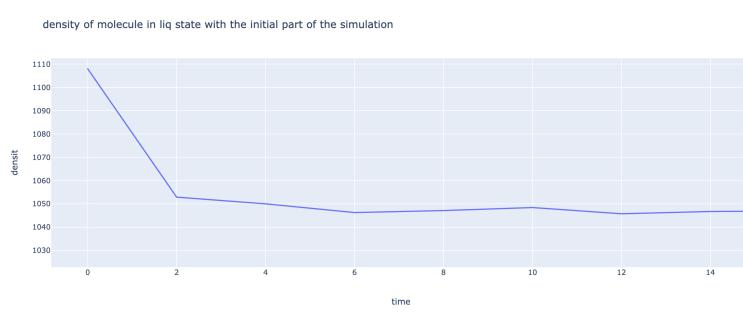


Figure 9: Zoom in of the first time points of the simulation

2.2. Mean vaporisation and density

When evaluating the simulation results depicting the mean vaporization enthalpy and density as a function of simulation length, several noteworthy observations come to the forefront. In the case of mean vaporization enthalpy as a function of simulation length, a significant trend emerges. It becomes evident that as the number of considered trajectories increases, even by just a mere two, the standard deviation shows a discernible decrease. This trend is particularly valuable as it underscores the importance of conducting a greater number of simulations to mitigate standard deviations, thereby enhancing the reliability and precision of the results. On the other hand, the situation is distinct when examining the mean density. In this context, the relationship between the number of trajectories and standard deviation is less straightforward. Increasing the number of trajectories does not consistently lead to a reduction in standard deviation. The impact of additional trajectories on standard deviation in the context of mean density exhibits more variability, and it cannot be universally claimed that more trajectories will consistently yield lower standard deviations.

Mean vaporisaiton heat and error as function of the simulation lenght (descending order)

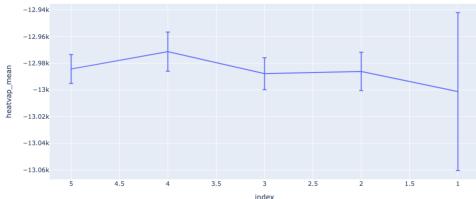


Figure 10: Mean vaporisaiton heat and error as function of the simulation length

Mean density and error as function of the simulation lenght (descending order)

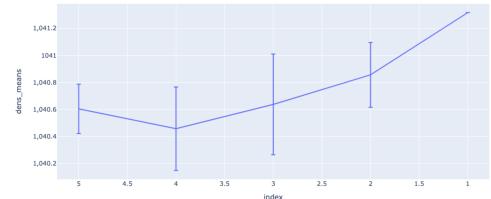


Figure 11: Mean density and error as a function of simulating length

2.3. Hydration Enthalpy

In order to estimate the hydration enthalpy ΔH_{wat} based on the solute-solvent non-bonded energy, we can use the total non-bonded energy (totnonbonded = ENER[10]) from the ‘ene_ana.md++.lib’ file. The ‘totnonbonded’ energy represents the cumulative non-bonded interactions among all atoms in a molecule, encompassing van der Waals forces, electrostatic interactions, and hydrogen bonds.

This allows us to calculate the intermolecular energy resulting from interactions within the solution. In this context, we consider the intermolecular interactions between the solute and the solvent. However, for a comprehensive assessment, it is essential to account not only for the interactions between the solute and the solvent (leading to a decrease in the total energy of the system) but also the energy increase arising from water-water hydrogen bonds that would have formed in the absence of the EGM molecule in the solvent. The complete formula is as follows:

$$H_{\text{wat}} = H_{\text{solv-solv}} - H_{\text{solv-solut}}$$

3. Quality of the Model

3.1. ρ and H_{vap}

To address this issue correctly, it would be advisable to conduct an additional simulation utilizing the same number of water molecules without the presence of the EGM molecule. This supplementary simulation would enable the extraction of valuable data regarding the solvent-solvent non-bonded interactions. Subsequently, the acquired information could be used to compute and incorporate the first addendum on the right-hand side (RHS) of the equation. The average values and relative error bars of density and vaporization heat obtained for the liquid-state compound at room temperature are 1040.600 ± 18 and -12984.31 ± 10.80 , respectively. We observe that the agreement with experimental data is quite good concerning the density but falls short when it comes to vaporization heat. We attribute this discrepancy to the fact that our approximation provides a reasonably accurate estimation of mass (and therefore density), but modeling the vaporization heat heavily depends on achieving an accurate representation of molecule-molecule interactions, a level of accuracy we do not attain with this method.

3.2. transferability

The assumption of parameter transferability might not be the best in this case. An example of why this could raise an issue is when studying the partial charges on the molecule.

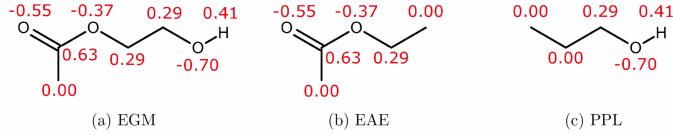


Figure 12: Atomic Partial Charge

While on the RHS of EAE and the LHS of PLL the partial charge is zero (and therefore we can assume are not involved in Coulomb interaction with surrounding molecules) the assigned charges on the central part of the EGM molecule are non-zero. This might potentially lead to other non-bonded interactions on the molecule with the surrounding ones. Another striking difference concerns the proper dihedral torsion potential assignment of the group 3-5-6-7.

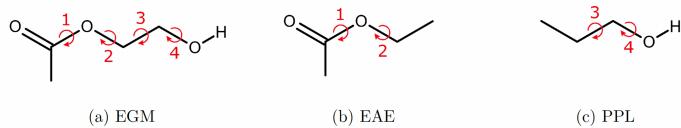


Figure 13: Proper Dihedral Torsion Potentials

Here we assign the parameter for a group of the kind C-C-C-O to one of the kind O-C-C-O. These torsion potentials are clearly different and this might lead to differences in the final conformation and therefore of the final energy of all the system.

3.3. Molecular Interactions

The assignment of atomic partial charges indeed holds fundamental importance in the determination of molecule-molecule interactions, and consequently, it significantly influences the total energy, vaporization enthalpy, and density of the system. These partial charges play a pivotal role in shaping the electrostatic interactions between molecules, which are a key component of the overall potential energy in the system. Furthermore, variations in the torsion potential parameters can impart differing degrees of flexibility to individual molecules. This, in turn, can lead to the emergence of various stable states for the entire system, resulting in measurements that more closely align with observed data. Care should be taken when modifying torsion potential parameters, as excessive alterations can potentially alter the inherent nature of the molecule, affecting its structural and conformational characteristics. It is worth noting that bond stretching and bond angle bending potentials are typically derived quantum mechanically and play a crucial role in preserving the fundamental structural features

of the molecule. Excessive modifications to these potentials may indeed lead to changes in the molecule's nature, potentially deviating from its actual behavior. Therefore, any adjustments to these potentials should be made judiciously to ensure that they accurately reflect the molecule's true characteristics while achieving the desired simulation results.

4. Thinking Questions

4.1. Distribution of partial charges

The distribution of the partial charges aligns remarkably well with chemical intuition. Indeed, upon calculating the average partial charges for the various types of atoms, we observe an approximate value of -0.54 for oxygen, 0.30 for carbon, and 0.41 for hydrogen. These findings are entirely consistent with the electronegativity values associated with these atoms, which stand at 3.54 for oxygen, 2.55 for carbon, and 2.20 for hydrogen. This close correspondence between calculated partial charges and electronegativity values underscores the reliability and accuracy of the charge assignment process, reinforcing the chemical soundness of the simulation results.

4.2. Lennard-Jones interaction

Molecule	Molecule	C_{12}	C_6
CH2	CH2	3.396558e-05	7.468416e-03
CH3	CH2	3.008414e-05	8.473481e-03
CH3	CH3	2.664624e-05	9.613802e-03

4.2.1. CH2 - CH2

Given: $-C_{12} = 3.396558 \times 10^{-5} \text{ kJ/mol}\cdot\text{nm}^{12}$ - $C_6 = 7.468416 \times 10^{-3} \text{ kJ/mol}\cdot\text{nm}^6$

To calculate r^* , we use the formula:

$$r^* = \left(\frac{2C_6}{C_{12}} \right)^{1/6}$$

Now, we'll plug in the values:

$$r^* = \left(\frac{2 \times 7.468416 \times 10^{-3}}{3.396558 \times 10^{-5}} \right)^{1/6}$$

Simplify the expression inside the parentheses:

$$r^* = \left(\frac{1.4936832 \times 10^{-2}}{3.396558 \times 10^{-5}} \right)^{1/6}$$

Now, perform the division:

$$r^* = (439.750003)^{1/6}$$

Take the sixth root:

$$r^* \approx 2.75763 \text{ nm}$$

So, r^* is approximately 2.75763 nanometers when using the provided values of C_{12} and C_6 in $\text{kJ/mol}\cdot\text{nm}^{12}$ and $\text{kJ/mol}\cdot\text{nm}^6$. r^* represents the characteristic distance for the Lennard-Jones potential in this context.

To calculate ϵ , we use the formula:

$$\epsilon = \frac{C_6^2}{4C_{12}}$$

Now, we'll plug in the values:

$$\epsilon = \frac{(7.468416 \times 10^{-3})^2}{4 \cdot 3.396558 \times 10^{-5}}$$

Simplify the expression:

$$\epsilon \approx 0.4106 \text{ kJ/mol}$$

So, ϵ is approximately 0.4106 kJ/mol when using the provided values of C_{12} and C_6 in $\text{kJ/mol}\cdot\text{nm}^{12}$ and $\text{kJ/mol}\cdot\text{nm}^6$ units. ϵ represents the depth of the Lennard-Jones potential energy well in this context.

4.2.2. $\text{CH}_2 - \text{CH}_3$

Given: $-C_{12} = 3.008414 \times 10^{-5} \text{ kJ/mol}\cdot\text{nm}^{12}$ - $C_6 = 8.473481 \times 10^{-3} \text{ kJ/mol}\cdot\text{nm}^6$

To calculate r^* , we use the formula:

$$r^* = \left(\frac{2C_6}{C_{12}} \right)^{1/6}$$

Now, we'll plug in the values:

$$r^* = \left(\frac{2 \cdot 8.473481 \times 10^{-3}}{3.008414 \times 10^{-5}} \right)^{1/6}$$

Simplify the expression:

$$r^* \approx 2.8738 \text{ nm}$$

So, r^* is approximately 2.8738 nanometers when using the provided values of C_{12} and C_6 in $\text{kJ/mol}\cdot\text{nm}^{12}$ and $\text{kJ/mol}\cdot\text{nm}^6$ units. r^* represents the characteristic distance for the Lennard-Jones potential in this context.

To calculate ϵ , we use the formula:

$$\epsilon = \frac{C_6^2}{4C_{12}}$$

Now, we'll plug in the values:

$$\epsilon = \frac{(8.473481 \times 10^{-3})^2}{4 \cdot 3.008414 \times 10^{-5}}$$

Simplify the expression:

$$\epsilon \approx 0.5978 \text{ kJ/mol}$$

So, ϵ is approximately 0.5978 kJ/mol when using the provided values of C_{12} and C_6 in $\text{kJ/mol}\cdot\text{nm}^{12}$ and $\text{kJ/mol}\cdot\text{nm}^6$ units. ϵ represents the depth of the Lennard-Jones potential energy well in this context.

4.2.3. CH₃ - CH₃

Given: - $C_{12} = 2.664624 \times 10^{-5}$ kJ/mol·nm¹² - $C_6 = 9.613802 \times 10^{-3}$ kJ/mol·nm⁶
To calculate r^* , we use the formula:

$$r^* = \left(\frac{2C_6}{C_{12}} \right)^{1/6}$$

Now, we'll plug in the values:

$$r^* = \left(\frac{2 \cdot 9.613802 \times 10^{-3}}{2.664624 \times 10^{-5}} \right)^{1/6}$$

Simplify the expression:

$$r^* \approx 2.9949 \text{ nm}$$

So, r^* is approximately 2.9949 nanometers when using the provided values of C_{12} and C_6 in kJ/mol·nm¹² and kJ/mol·nm⁶ units. r^* represents the characteristic distance for the Lennard-Jones potential in this context.

To calculate ϵ , we use the formula:

$$\epsilon = \frac{C_6^2}{4C_{12}}$$

Now, we'll plug in the values:

$$\epsilon = \frac{(9.613802 \times 10^{-3})^2}{4 \cdot 2.664624 \times 10^{-5}}$$

Simplify the expression:

$$\epsilon \approx 0.8664 \text{ kJ/mol}$$

So, ϵ is approximately 0.8664 kJ/mol when using the provided values of C_{12} and C_6 in kJ/mol·nm¹² and kJ/mol·nm⁶ units. ϵ represents the depth of the Lennard-Jones potential energy well in this context.

4.3. General Considerations

Molecule Pair	r^* (nm)
CH ₂ - CH ₂	2.75763
CH ₂ - CH ₃	2.8738
CH ₃ - CH ₃	2.9949

Given two molecules we can notice that the Lennard-Jones distance increases when we consider molecules with more and more atoms, due to the likely higher volume occupied in space

4.4. average translational velocity of a gas-phase EGM

To estimate the average translational velocity of a gas-phase molecule at a given temperature (298.15 K), you can use the root-mean-square (RMS) speed formula, which is based on the kinetic theory of gases. The formula for calculating the RMS speed (v) of gas molecules is given by:

$$v_{\text{RMS}} = \sqrt{\frac{3kT}{m}}$$

Where:

v_{RMS} = Root Mean Square Speed

k = Boltzmann's constant ($1.380649 \times 10^{-23} \text{ J K}^{-1}$)

T = Absolute temperature

m = Mass of a gas molecule

To calculate the molar mass of ethylene glycol monoacetate (EGM), you can use the chemical formula $\text{C}_4\text{H}_8\text{O}_3$

So, the molar mass of ethylene glycol monoacetate (EGM) is approximately 104.12 g/mol.

From here, we can obtain a theoretical value of: $v_{\text{rms}} \approx 413.9 \text{ m/s}$

We consider a $d_{\text{gen}} = 500 \text{ nm}$, and a $d_{\text{cut}} = 1.4 \text{ nm}$. Given a $t_{\text{sim}} = 300 \text{ ps}$

$$d_{\text{max}} = 124.17 \text{ nm}$$

In the worst case scenario, where molecules are generated and move in the same direction, we can consider a $d_{\text{lim}} = 248.34 \text{ nm}$. $d_{\text{gen}} \approx 2 \cdot d_{\text{lim}}$. For this reason, can highly conclude the molecules will never collide

4.5. Discussion on $\langle H \rangle = \langle K + U \rangle$ and RT

The reason we use the average total potential energy ($\langle U \rangle$) and not the average total energy ($\langle H \rangle$) to calculate the enthalpy of vaporization is because [the process of vaporization happens at a well-defined temperature and] the kinetic energy ($\langle K \rangle$) is the same in the liquid and gas phases at the same temperature. This is because the average kinetic energy of a molecule is only dependent on its temperature, not its state of matter. [1]

When considering enthalpy, it's important to note that it doesn't only involve the internal energy (U), but also the pressure-volume work (PV) term. Indeed, the enthalpy of vaporization is the amount of energy required to convert one mole of a liquid to a gas at a constant temperature and pressure. It is equal to the change in the internal energy of the system plus the work done against the external pressure.

The potential energy of the liquid phase is higher than the potential energy of the gas phase, where the particles are further apart and interact less strongly with each other. In the liquid phase, the particles are close together and interact strongly with each other.

When a liquid evaporates, the work done is either to overcome the intermolecular forces that are holding the molecules together or to work done against the external pressure, which is equal to the product of the pressure and the change in volume of the system. For this reason, the product RT represents the energy contribution due to the ideal gas behavior of the molecules [since $PV = nRT$ for ideal gases and we are working with one mole of compound].

4.6. Temperature

Using the standard enthalpy of vaporization at the standard temperature of 298.15 K (25°C) rather than at the boiling point (T_b) can be useful for several reasons. Firstly, and perhaps most importantly, it eliminates temperature effects. Temperature would influence the kinetic energy of the molecules. Since the total energy is the sum of potential and kinetic energy, maintaining a consistent temperature simplifies multiple calculations and ensures comparability of results across different systems and simulations. Secondly, measuring the enthalpy of vaporization at the boiling point (T_b) of a liquid may not always be practical or convenient, especially when dealing with liquids that have exceptionally high boiling points. Experimental measurements under such conditions can be challenging. By using the standard temperature of 298.15 K, experimental data can be more readily obtained in laboratory settings. Lastly, it is worth noting that many experimental measurements are frequently conducted at or around room temperature, making the standard temperature of 298.15 K a commonly used reference point for practicality and consistency.

5. Appendix

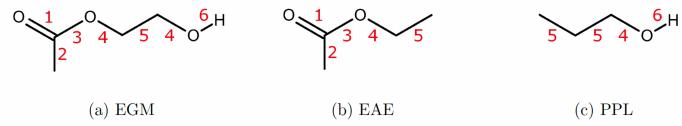


Figure 14: Bond Stretching Potentials

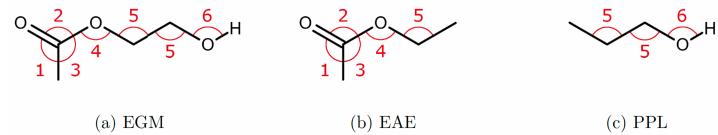


Figure 15: Bond Angle bending Potentials

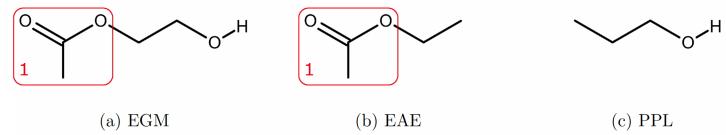


Figure 16: Improper Dihedral Distortion Potentials

CGC

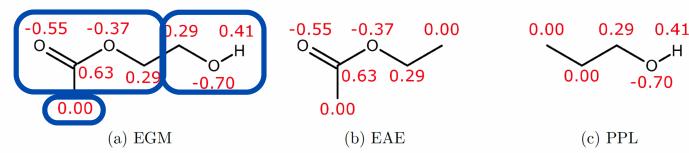


Figure 3: Atomic Partial Charges

Figure 17: Atomic Partial Charges

References

- [1] LibreTexts, Heat of vaporization, 2023. 15/10/2023.