A Comparative Study of Diffusion Dynamics in Methane (CH_4) and Fluoroform (CHF_3) using Molecular Dynamics

Raghav Mathur[1]

Received October 26, 2025
Accepted May 9, 2026
Electronic access June 15, 2026

This study investigated the diffusion behaviour of Methane (CH_4) and Fluoroform (CHF_3) in an aqueous environment, using a Molecular Dynamics simulation with Groningen Machine for Chemical Simulations (GROMACS). Using the OPLS-AA force field, solvated systems were simulated to understand the rate of diffusion in a polar solvent like water, both as pure water and as a saline solution. The diffusion coefficient for each molecule was calculated using the Einstein relation. This paper offers insight into the comparative diffusivity of relatively small molecules in water and concludes that because CH_4 is lighter than CHF_3, its diffusion coefficient is higher than that of the latter. The result of the study also shows that, generally, a molecule will diffuse faster in pure water than in saline solutions. Molecular mass affects diffusivity. However, in solution, diffusion reflects solute–solvent interactions and also has contributions from effective hydrodynamic size and solvation structure. CH_4 and CHF_3 molecules are different not only in terms of mass but also in terms of polarity, which affects their interaction with water molecules and leads to the observed differences in diffusion coefficients.

Introduction

Diffusion is a process involving the net movement of a substance from a region of higher concentration to a region of lower concentration, driven by a gradient of Gibbs free energy or chemical potential[1]. This movement of particles is observed everywhere in nature. For example, diffusion occurs in our lungs when we breathe and affects the rates of chemical reactions. It is understood that in order for a chemical reaction to occur, atoms/molecules that are known as the reactants must collide together with sufficient energy. The frequency of the collisions of the reactant molecules is determined by how fast those molecules diffuse, which results in the variation of the rate of reaction, as it is dependent on the collision frequency.

Molecular dynamics is a way to intuitively understand the interactions between tiny molecules, their collisions, and their net movement, known as diffusion, which will be the theme of this paper[2]. Molecular dynamics helps us to capture the mass and transport mechanisms at the atomic scale.

The benefit of these simulations is that they are able to predict the trajectories of molecules and use theoretical equations and principles to give ideal experimental values. This is done by applying fundamental laws of Newtonian mechanics to a large number of microscopic molecules to explain macroscopic effects such as pressure, temperature, and volume[3]. Molecular dynamics simulation software, such as GROMACS[4], allows for reduced experimental error when carrying out investigations.

[1] *Premier Academy, Nairobi, Kenya*

The two molecules that this investigation was carried out on were Methane and Fluoroform:

Methane, the simplest hydrocarbon, was chosen for the study because of its fundamental structure. Its smaller molecular size affects factors such as the diffusion coefficient[5]. Fluoroform has a heavier molecule than Methane. This helps us compare the effect of the mass of a molecule to its diffusion rate (diffusion coefficient)[6].

Methodology

GROMACS uses the `gmx msd` function in order to calculate the diffusion coefficient of a molecule. MSD, a quantity commonly used in Statistical mechanics, stands for Mean Square Displacement. MSD is the average distance of a particle from its starting point over a short period of time. "Equation of motion is calculated using leapfrog algorithm with an integration time step" (Chen et al. 2018)[7]. In our case, this time step was 2 fs. "All diffusion coefficients are calculated with the mean-squared-displacement (MSD) method which is based on the Einstein relation that MSD is linear with diffusion time." (Zhao & Jin, 2020)[8].

$$D = \frac{1}{2d} \lim_{t\to\infty} \frac{d}{dt} \left\langle \left|\mathbf{r}_{\mathrm{CM}}(t) - \mathbf{r}_{\mathrm{CM}}(0)\right|^2 \right\rangle \tag{1}$$

Einstein relation for diffusion, but $\mathbf{r}_{\mathrm{CM}}(0) = 0$, therefore:

$$D = \frac{1}{2d} \frac{d}{dt} \left\langle \Delta r_{\mathrm{CM}}^2(t) \right\rangle \approx \frac{\text{slope of MSD}(t)}{2d} \tag{2}$$

where d is the number of dimensions. Since there are 3 dimensions in the simulation, $d = 3$.

The simulation times were kept short and limited sampling was carried out as this study is focused mainly on trend comparison rather than precision[9]. The investigation comprised three different simulations:

Methane in water (referred to as CH_4), Fluoroform in water (CHF_3) and CH_4 in salt-water (CH_4 SW).

Throughout all the simulations, the OPLS-AA force field was used[10].

The structure file of CH_4 was created using ChemSketch[11] which was then imported to GROMACS. The structure files can be viewed using software such as Avogadro[12] and VMD[13], as shown in Figure 2.1 and 2.2. Next, topology files - specifically the `.top` and `.gro` files - were generated. This was done by giving it the input of the `.pdb` file or SMILES code using OpenBabel[14]. The SMILES code for the CHF_3 molecule's topology is: `FC(F)F`, as given by LigParGen[15]. Additionally, `.mdp` files were needed to configure the parameters of the simulations.

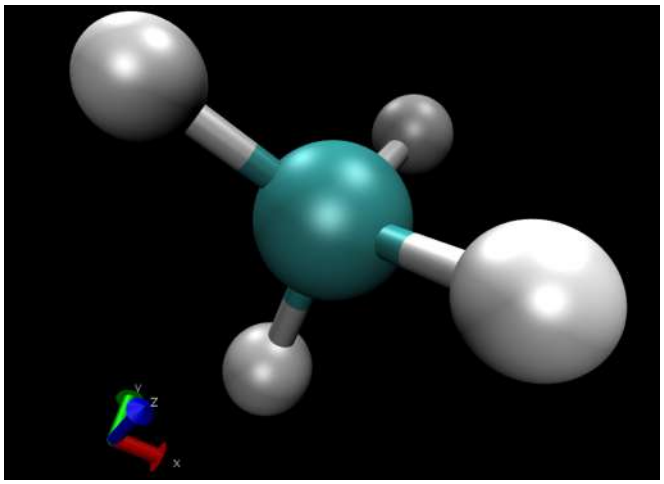

Fig 2.1 CH_4 molecule

Next, a simulation box was defined using `gmx editconf`:

```
gmx editconf -f CH4.gro -o newbox.gro
-c -d 1.0 -bt cubic
```

In this case, a cubic box having periodic boundary conditions was created with the dimensions $2.56\,\text{nm} \times 2.56\,\text{nm} \times 2.56\,\text{nm}$. The two molecules were placed in relatively larger boxes to avoid possible solute–solute interactions. Additionally, using a smaller box results in greater finite size effects acting on it, causing suppressed long-range correlations and leads to an underestimation of molecular movement. This decreases the accuracy of the calculated diffusion coefficient.

Next, the box was solvated by using the command `gmx solvate`. To solvate the CH_4 molecules, Na^+ and Cl^- ions

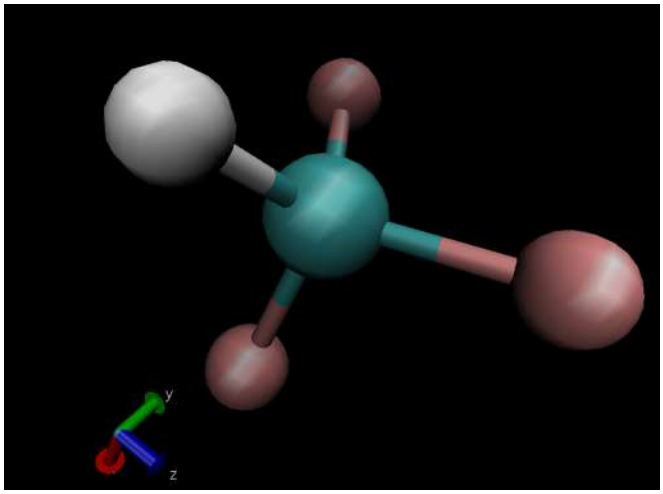

Fig 2.2 CH_3 molecule

were introduced at a concentration 0.5 M, using the command `gmx genion` as shown in Figure 2.3. A saline environment of 0.5 M was used because it is similar to that of seawater, keeping a simple composition that is often assumed in simulations.

Before the production simulations, energy minimisation (EM) was performed to relax the system to a mechanically stable configuration.

Molecular dynamics simulations were then conducted under the NVT ensemble, in which the number of particles and volume were kept constant while the temperature was controlled using a thermostat. NVT abbreviates the following: N is the number of particles; V is the Volume; T is Temperature. Pressure is not controlled, therefore the box size stays constant.

This was followed by the NPT ensemble, where both temperature and pressure were regulated, allowing the system volume to reach the target pressure. NPT abbreviates the following: N is the Number of particles; P is the Pressure; T is Temperature. The box's volume changes to maintain the desired pressure. The thermostat coupling constant was 0.1 ps and the barostat coupling constant was 2.0 ps.

The NVT, NPT and production molecular dynamics (Production MD) simulations were carried out at different time lengths for different molecules in different conditions. This was due to some systems stabilising before others, in terms of temperature and pressure.

Energy Minimisation

With the solvated, neutral system created, the next step was energy minimization. Energy minimisation is an important step to ensure that the system has a stable structure[16]. We can initiate the simulation by using the `gmx grompp` command:

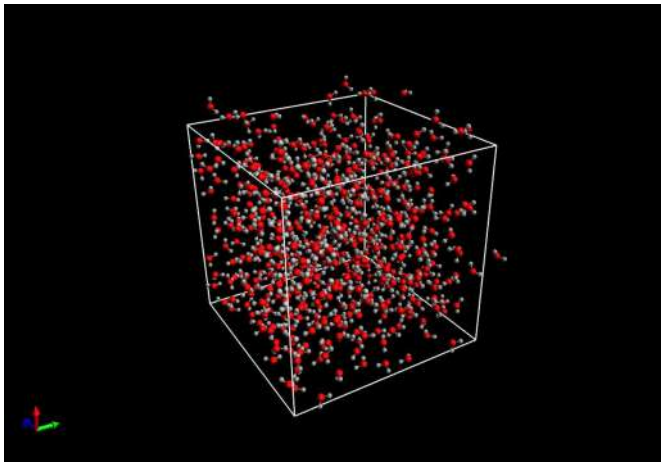

Fig 2.3 Solvated Box in Avogadro for CH_4 in water

```
gmx grompp -f minim.mdp -c solv.gro -p
topol.top -o em.tpr
```

Figure 3.1 shows the Energy Potential Graph, obtained by the command `gmx energy`. As the energy potential decreases, the system becomes more stable.

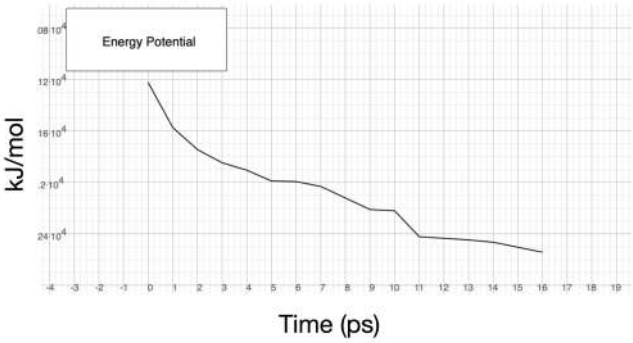

Fig 3.1

Table 3.1 shows the maximum force threshold for the energy minimisation and the number of steps taken for this.

Table 3.1 EM Parameters.

Parameter	Value
Integrator	steep (steepest descent)
Maximum force (`emtol`)	1000.0 kJ mol^{-1} nm^{-1}
Minimization step size (`emstep`)	0.01 nm
Maximum steps (`nsteps`)	50000
Cutoff scheme	Verlet
Neighbor search type	grid
Coulomb type	PME
Coulomb cutoff	1.0 nm
van der Waals cutoff	1.0 nm
Periodic boundary conditions	xyz

NVT Equilibration

The NVT Equilibration simulations utilised a thermostat to regulate and maintain the system temperature at a set point. For the NVT ensemble, a V-Rescale thermostat was used to control the velocities of each individual particle depending on the deviation of the temperature of the system from the set point (298 K). The V-Rescale thermostat adds to the Berendsen thermostat by adding accurate temperature fluctuations and it is a rather sensitive thermostat[17]. The simulation can be initiated by using the `gmx grompp` command:

```
gmx grompp -f nvt.mdp -c em.gro -r
em.gro -p topol.top -o nvt.tpr
```

Table 4.1 NVT Equilibration Parameters.

Parameter	Value
Integrator	md
Number of steps	50000
Time step (`dt`)	0.002 ps
Constraint algorithm	LINCS
Constraints	h-bonds
Cutoff scheme	Verlet
Coulomb cutoff	1.0 nm
van der Waals cutoff	1.0 nm
Electrostatics	PME
Thermostat	V-rescale (300 K)

Table 4.1 shows the parameters that determine the length and accuracy of our NVT Equilibration.

Figures A1, A2, and A3 in the supplementary material show the graphs of temperature equilibration during the NVT ensemble.

NPT Equilibration

During NPT equilibration, both the temperature and pressure were controlled using a V-Rescale thermostat and a Berendsen barostat, respectively. A barostat regulates the pressure of a system around a setpoint by changing the volume of the box gradually, while constantly checking the pressure at intervals. The Berendsen barostat is great for a quick stabilisation, but not very useful for analysis as it smooths out the fluctuations, which makes the system behave unnaturally. As a result, it does not accurately sample the NPT ensemble, and therefore it is used only in the equilibration steps. This simulation can be initiated by using the `gmx grompp` command:

```
gmx grompp -f npt.mdp -c nvt.gro -r
nvt.gro -t nvt.cpt -p topol.top -o
npt.tpr
```

Table 5.1 shows the parameters that determine the length and accuracy of our NPT Equilibration.

Table 5.1 NPT Equilibration Parameters.

Parameter	Value
Integrator	md
Number of steps	50000
Time step (`dt`)	0.002 ps
Constraint algorithm	LINCS
Constraints	h-bonds
Cutoff scheme	Verlet
Coulomb cutoff	1.0 nm
van der Waals cutoff	1.0 nm
Electrostatics	PME
Thermostat/Barostat	V-rescale (300 K) / Parrinello–Rahman (1 bar)

Figures B1, B2, and B3 in the supplementary material show the graphs of pressure equilibration during the NPT equilibration.

Production MD Run

Once the equilibration phase was complete, the simulation moved to the Production Molecular Dynamics stage. This was the final stage of the simulation that produced the final result for data collection [17].

The simulation was initiated by using the `gmx grompp` command:

```
gmx grompp -f md.mdp -c npt.gro -t
npt.cpt -p topol.top -o md.tpr
```

The simulation was subsequently initiated with a small runtime and limited sampling was carried out. This is because the study is focused mainly on comparing trends rather than obtaining precise data.

Figure 6.1 shows the molecule CH_4 in Avogadro at the end of the production MD simulation. It is the visual representation of the final simulation.

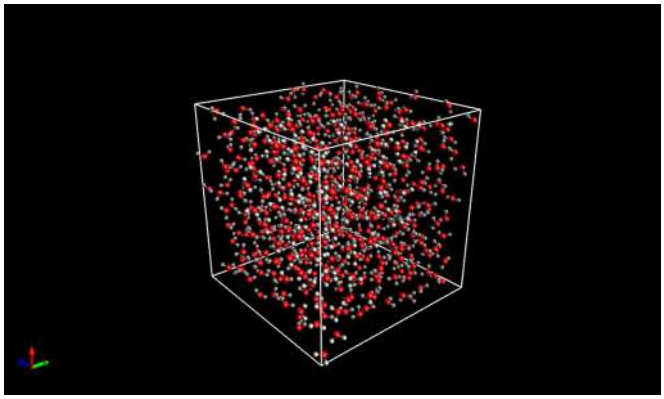

Fig 6.1 Production MD Box

Table 6.1 contains the different simulation parameters for the MD Run, obtained from a simulation file. These parameters are the same to all 3 simulations except for the simulation lengths.

Table 6.1 MD Run Parameters

Parameter	Value
Integrator	md
Number of steps	500000
Timestep	0.002 ps
Constraints	h-bonds
Cutoff scheme	Verlet
Coulomb cutoff	1.0 nm
van der Waals cutoff	1.0 nm
Electrostatics	PME
Thermostat type	V-rescale
Barostat type	Parrinello–Rahman

Figures C1, C2, and C3 in the supplementary material show the graphs of the temperature equilibration, while Figures D1, D2 and D3 show the graphs of pressure equilibration during the MD production run.

Results and Discussion

The diffusion coefficient for each of the simulations was then calculated using the function `gmx msd`. From the Einstein relation for diffusion, it can be deduced that:

$$D \approx \frac{\text{slope of MSD}(t)}{6} \tag{3}$$

The mean square displacement (MSD) graphs were plotted and the diffusion coefficients were determined using the `gmx msd` function:

```
gmx msd -f md.xtc -s md.tpr -n
index.ndx -o msd.xvg
```

A log–log graph of MSD against time was used to determine the linear portions of the graph. This is shown by the graphs in Figure 7.1 with non-linear scales.

After plotting these graphs, the flags `-beginfit` and `-endfit` were used to add the limits of the linear portion and redo the simulation. This made the calculated diffusion coefficient more accurate. The diffusion coefficients obtained from GROMACS are shown in Table 7.1.

Table 7.1 compares the diffusion coefficients achieved by the simulations being run compared to literature. However, due to the limited literature values on CHF_3 in water and CH_4 in salt water, they are given according to widely accepted or estimated values. The literature value for CH_4 in water is given by Chen et al. (2018) [7].

Table 7.1 Diffusion Coefficients

Molecule Name	Diffusion Coefficient ($\times 10^{-5}$ cm^2 s^{-1})	Literature Diffusion Coefficient ($\times 10^{-5}$ cm^2 s^{-1})
CH_4	8.98 ± 4.28	1.44 ± 0.11
CHF_3	3.89 ± 1.24	1.2 (estimated)
CH_4 SW	3.58 ± 0.36	1.4 (estimated)

Running simulations using various initial velocity seeds and calculating block average for the mean square displacement will definitely increase the statistical reliability and reduce uncertainty. In this paper, only single trajectories for calculation were used due to computational limitations. Even with single trajectories, the simulations show expected qualitative trends for molecular diffusion. These include slower diffusion for heavier solutes and less mobile behaviour in saline solutions. The lack of block averaging and uncertainty analysis is recognised as a limitation of this study. The manuscript clarifies that high uncertainty in the values of the diffusion coefficients arises from trajectory length and sampling and contains a plan for next stage of work. This will also involve block averaging among other error analysis methods.

Molecular mass affects diffusivity, however, in solution, diffusion reflects solute–solvent interactions and also has contributions from effective hydrodynamic size and solvation structure. CH_4 and CHF_3 molecules are different not only in terms of mass but also in terms of polarity. This could affect their interaction with water molecules and lead to different results being obtained. The current findings are therefore to be considered as a result of the effect of both molecular mass and solvation-related interactions.

NVT and NPT Equilibration, and Production MD

During NVT equilibration, NPT equilibration and production MD, the temperature is stabilised to around 298 K and the pressure stabilised to about 1 bar. This is demonstrated by the graphs in the supplementary material, which have a 10 ps running average that fluctuates about the set-point at an acceptable amount to be considered as stabilised.

Mean Square Displacement Graphs

When analysing the simulations to measure diffusion, the mean square displacement (MSD) of the molecules in the simulations must be considered. In order to obtain accurate values of the diffusion coefficients, a graph of MSD against time with non-linear scales must be plotted. This is done by plotting the logarithm of MSD against the logarithm of time. Note that the logarithms are to the base 10. In Figure 7.1, the linear portions of the log MSD – log time graphs can be observed in order to choose the limits for the accurate calculation of the diffusion coefficient. The MSD graphs were also plotted with a linear scale [18].

Diffusion Coefficients

The diffusion coefficients were calculated by GROMACS, as listed in Table 7.1. They indicate that CH_4 in water has the highest rate of diffusion out of the three simulations, followed by CHF_3 in water, and lastly, CH_4 in salt water. This is due to the larger mass of a CHF_3 molecule as compared to CH_4 and the slower diffusion of molecules in salt water as compared to pure water. CH_4 is a molecule that is much lighter than CHF_3. CH_4 has a mass of approximately 16.0425 g mol^{-1}, while CHF_3 has a mass of approximately 70.0138 g mol^{-1}. The diffusion coefficient of CH_4, as given in Table 7.1, is higher than the diffusion coefficient of CHF_3. This means that the less the (molar) mass of a molecule, the faster it can diffuse in water. If NaCl salt is added to water of concentration 0.5 M, we observe another decrease in the diffusion coefficient of CH_4. This means that, generally, a molecule will diffuse faster in pure water than in water with salt [19].

Conclusion

This study investigated the diffusion behaviour of Methane (CH_4) and Fluoroform (CHF_3) in a water environment, using a molecular dynamics simulation. There were 3 separate sets of molecular dynamics simulations that were carried out. The first two simulations were a comparison of the mass of a molecule to its rate of diffusion, while the last and first simulations were a comparison of the rate of diffusion of a molecule to the presence of ions. The diffusion coefficient for each of CH_4 and CHF_3 was calculated using the Einstein relation for diffusion [20]. The simulations carried out confirm the claim that due to CH_4 being much lighter than CHF_3, the diffusion coefficient of CH_4 is higher than the diffusion coefficient of CHF_3. The simulations also indicate that, generally, a molecule will diffuse faster in pure water than in salt water. The findings are to be considered as the result of both molecular mass and solvation-related interactions. Future work will involve block-averaging with several simulations that run for longer times in order to increase the statistical reliability.

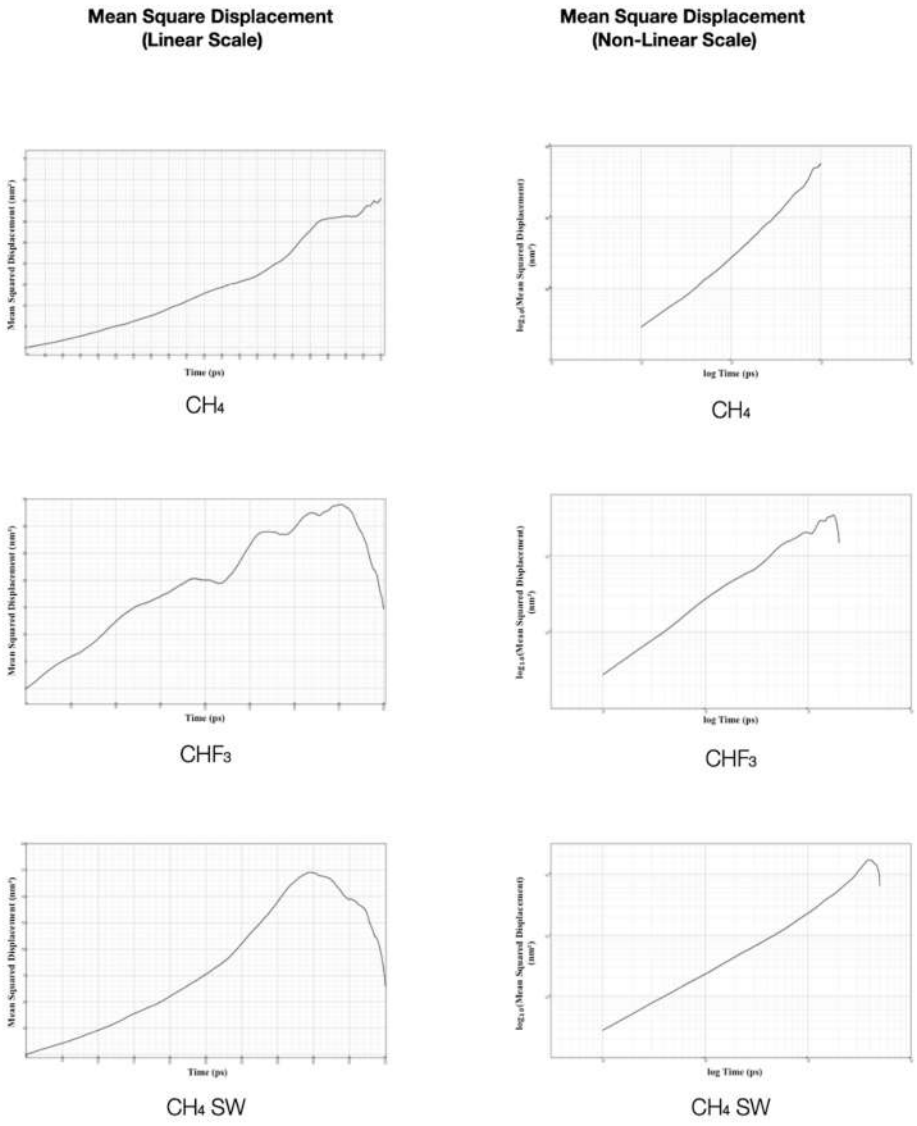

Fig 7.1 Mean square displacement graphs

References

1 P. Naeiji, T. K. Woo, S. Alavi and R. Ohmura, *Title not provided*, 2020, 10.1063/5.0008114.

2 G. Battimelli and G. Ciccotti, *Title not provided*, 2018, 10.1140/epjh/e2018-90027-5.

3 A. D. Astuti, R. Refianti and A. Mutiara, *Molecular Dynamics Simulation on Protein Using Gromacs*, 2011.

4 M. J. A. et al., *GROMACS – Groningen Machine for Chemical Simulations*, 2019, 10.5281/zenodo.14846130.

5 M. Wahlen, *Title not provided*, 1993, 10.1146/annurev.ea.21.050193.002203.

6 P. A. Witherspoon and D. N. Saraf, *Diffusion of Methane, Ethane, Propane, and n-Butane in Water from 25 to 43°*, 1965, 10.1021/j100895a017.

7 Y.-A. Chen, C.-K. Chu, Y.-P. Chen, L.-S. Chu, S.-T. Lin and L.-J. Chen, *Measurements of diffusion coefficient of methane in water/brine under high pressure*, 2018, 10.3319/TAO.2018.02.23.02.

8 X. Zhao and H. Jin, *Title not provided*, 2020, 10.1016/j.applthermaleng.2020.114941.

9 V. Marry, P. Turq, T. Cartailler and D. Levesque, *Microscopic simulation of structure and dynamics of water and counterions in a monohydrated montmorillonite*, 2002, 10.1063/1.1493186.

10 W. L. Jorgensen and J. Tirado-Rives, *Title not provided*, 2005, 10.1073/pnas.0408037102.

11 Advanced Chemistry Development, Inc., *ChemSketch*, 2022, Available at: www.acdlabs.com.

12 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform*, 2012, 10.1186/1758-2946-4-17.

13 W. Humphrey, A. Dalke and K. Schulten, *VMD: Visual molecular dynamics*, 1996, 10.1016/0263-7855(96)00018-5.

14 N. M. O'Boyle, M. Banck and C. A. J. et al., *Open Babel: An open chemical toolbox*, 2011, 10.1186/1758-2946-3-33.

15 L. S. Dodda, I. C. de Vaca, J. Tirado-Rives and W. L. Jorgensen, *LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands*, 2017, 10.1093/nar/gkx312.

16 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *The missing term in effective pair potentials*, 1987, 10.1021/j100308a038.

17 I. Shvab and R. J. Sadus, *Thermodynamic properties and diffusion of water + methane binary mixtures*, 2014, 10.1063/1.4867282.

18 H. Moradi, H. Azizpour, H. Bahmanyar, M. Mohammadi and M. Akbari, *Prediction of methane diffusion coefficient in water using molecular dynamics simulation*, 2020, 10.1016/j.heliyon.2020.e05385.

19 T. Funazukuri, *Title not provided*, 2018, 10.1016/j.supflu.2017.11.035.

20 M. J. W. Frank, J. A. M. Kuipers and W. P. M. van Swaaij, *Diffusion Coefficients and Viscosities of CO2 + H2O, CO2 + CH3OH, NH3 + H2O, and NH3 + CH3OH Liquid Mixtures*, 1996, 10.1021/je950157k.

Supplementary information

The online version contains supplementary material available at https://nhsjs.com/?p=44597