



HAL
open science

Molecular basis of thermophoresis

Alejandro Díaz Márquez

► **To cite this version:**

Alejandro Díaz Márquez. Molecular basis of thermophoresis. Theoretical and/or physical chemistry. Université Paris Cité, 2021. English. NNT : 2021UNIP7067 . tel-03626360

HAL Id: tel-03626360

<https://theses.hal.science/tel-03626360>

Submitted on 31 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

École doctorale de Chimie Physique et de Chimie Analytique de Paris-Centre (ED 388)

Laboratoire de Biochimie Théorique, UPR 9080

Institut de Biologie Physico-Chimique

Molecular basis of thermophoresis

Par Alejandro Díaz Márquez

Thèse de doctorat de chimie théorique et modélisation

Dirigée par Guillaume Stirnemann

Présentée et soutenue publiquement le 10.12.2021

Devant un jury composé de :

- Guillaume Galliéro, PR, Université de Pau et des Pays de l'Adour, FR, Rapporteur.
- Margarida Costa Gomes, DR, ENS de Lyon, FR, Rapporteur.
- Benjamin Rotenberg, DR, Sorbonne Université, FR, Président du jury.
- Guillaume Stirnemann, CR, CNRS, Paris, FR, Directeur de thèse.



Except where otherwise noted, this is work licensed under <https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

I would like to dedicate this thesis to my uncle, Luis Márquez Abadín.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains less than 65,000 words including appendices, bibliography, footnotes, tables and equations and has less than 150 figures.

Alejandro Díaz Márquez

October 2021

Acknowledgements

At the end of this road, I would like to acknowledge those people that have been part of these last three years and have contributed personally and professionally to my PhD and the thesis you are reading now. I apologize in advance in case that I forgot to mention somebody.

First and foremost, I would like to thank my supervisor Guillaume Stirnemann for his scientific guidance, advices, motivation and support. I have learnt a lot from you, which has made me grown as a person and as a professional. Thank you very much for giving me the opportunity to be part of your group during these three years.

I would also like to thank the members of my jury for accepting to evaluate my thesis: Guillaume Galliéro, Margarida Costa and Benjamin Rottenberg. In particular, I would like to thank to my rapporteurs for having reviewed my manuscript and for the quality of their comments.

A big thank to every past and current LBT member, whose generate a fun and comfortable working environment. In particular, Élise, Marina and Jérôme, thanks for your professional advices and fruitful discussions. Nicolas, Astrid and Štěpán, thank you very much for your unconditional help during the first steps of my thesis, solving technical and bureaucratic difficulties. Olivier and Matthias, thank you very much for the good times we have spent during these three years as PhD students. Antonio, thank you for all the discussions about physics. Special thanks also go to all the younger PhD students: Antoine, Arthur, André, Aimeric, Marie, Emeline, Julie, Zakarya and Sélène (merci pour ton aide avec le résumé!). Many thanks to Victoria, Yves and Geoffrey, their work makes the research possible.

I have to thank all the other people from the IBPC institute for all the good times that we shared. In particular, Théo and Verónica, thank you for the great moments and nice experimental/theoretical discussions.

I would also like to thanks all the friends who have been around in the last years. Justine and Wissem, thank for our mutual support during the pandemic lockdown. Amara, thank you for

all the videocalls that we shared during the pandemic. Luis, Selim and Jordi, thank you for your support from Groningen, I hope that we can share more “lekkel” moments together. Marta and Raph, thank you very much for your support and friendship along these years. Léa, thank you for the good moments and for showing me the French culture.

Of course, I have to thank my close friends, who have been present in my life for a long time. Roser, Judit, Anna and Cucu, thank you for all the great moments and trips during the university and later years that we spent together. Roger, Cano, Albert, Fera and Urpi, thank you very much for being with me every day and supporting me in the happiest and saddest days. Cristina, I am very grateful for having met you and I have great memories of many conversations, dinner and movies moments. Maria, I have no words to express my gratitude for all these years of friendship, almost 30 years and counting, I just want to thank you from the bottom of my heart from being there all this time. I would like to thank you all for coming to my thesis defence.

Finally, I am most grateful for the unconditional support of my parents Leandro and Josefa, and my sister Rebeca. I am so lucky to be part of this family, which has always believed in me, without you I would not have been the person I am today. You are part of this.

RÉSUMÉ

La thermophorèse ou thermodiffusion est un processus thermique par lequel la composition d'un mélange devient inhomogène sous l'effet d'un gradient de température, ce phénomène est connu sous le nom d'effet Soret. La nature de la thermophorèse n'est pas encore complètement comprise et reste très débattue, par contre est un phénomène assez présent dans la nature. Un cadre de recherche important où la thermophorèse avec la convection thermique pourrait expliquer l'accumulation de précurseurs dans des conditions prébiotiques, un processus crucial pour l'émergence de la vie en l'absence de toute compartimentation.

L'objectif principal de ce projet est de comprendre les bases moléculaires de la thermophorèse dans les solutions aqueuses diluées. La thermophorèse est par essence un phénomène de non-équilibre, et certaines théories moléculaires se sont concentrées sur l'interprétation des gradients de concentration observés en termes de coefficients de transport. Néanmoins, expériences récentes ont suggéré que la thermophorèse en conditions d'état stationnaire peut-être comprise en prenant des considérations thermodynamiques dans équilibre.

Les simulations dans le cadre de dynamique moléculaire sont un outil idéal pour déchiffrer ces questions sur de petits systèmes modèles. Nous nous sommes particulièrement intéressés aux petites molécules amphiphiles dans l'eau, qui englobent les caractéristiques moléculaires de base des solutés avec des parties hydrophiles et hydrophobes. Tout d'abord, nous avons proposé une approche de dynamique moléculaire robuste et fiable pour simuler la thermophorèse dans des solutions diluées réalistes en utilisant des simulations de dynamique moléculaire tout-atome. Ensuite, nous avons recherché des multiples propriétés physiques qui pourraient être corrélées avec le coefficient de Soret (c'est-à-dire l'ampleur du gradient de concentration en réponse à un gradient de température). Finalement, nous avons vérifié la viabilité de ce processus thermique qui pourrait être compris en termes de thermodynamique ou de cinétique à l'équilibre, comme c'est suggéré précédemment dans la littérature.

Mots clés : Hors équilibre, coefficient de Soret, dynamique moléculaire, diffusion, énergie libre de solvation

ABSTRACT

Thermophoresis or thermodiffusion is a thermally-driven process whereby the composition of a mixture becomes inhomogeneous under a temperature gradient, a phenomenon known as the Soret effect. The nature of thermophoresis is still not completely understood and it remains highly debated, though it is a phenomenon fairly present in the nature. One important framework field is that thermophoresis together with thermal convection could unveil the precursor accumulation in prebiotic conditions, a crucial process for the emergence of life in the absence of any compartmentalization.

The main goal of this project is to understand the molecular bases for thermophoresis in dilute aqueous solutions. Thermophoresis is in essence a non-equilibrium phenomenon, and some molecular theories have focused on the interpretation of the observed concentration gradients in terms of transport coefficients. However, recent experiments have suggested that thermophoresis steady-states could be understood using equilibrium thermodynamic considerations.

Molecular dynamics simulations are an ideally suited tool to decipher these questions on small model systems. We especially scoped small and amphiphilic molecules in water, which encompasses the basic molecular features of solutes with hydrophilic and hydrophobic moieties. Firstly, we proposed a robust and reliable molecular dynamic approach to undertake thermophoresis in dilute realistic solutions using all-atom molecular dynamics simulations. Secondly, we enquired into multiple physical properties that could be correlated with the Soret coefficient (i.e., the extent of the concentration-gradient in response to a temperature-gradient). Finally, we checked the viability of this thermally-driven process could be understood in terms of equilibrium thermodynamics of kinetics, as suggested before in the literature.

Keywords: Non equilibrium, Soret coefficient, molecular dynamics, diffusion, solvation free-energy

RESUMEN

La termoforesis o termodifusión es un proceso térmico por el cuál la composición de una mezcla se vuelve inhomogénea bajo un gradiente de temperatura, este fenómeno es conocido como efecto Soret. La naturaleza de la termoforesis aún no se entiende del todo y sigue siendo muy debatida, aunque en realidad es un fenómeno bastante común en la naturaleza. Un campo de investigación importante es dónde la termoforesis junto con la convección térmica podrían explicar la acumulación de precursores en condiciones prebióticas, un proceso crucial para el origen de la vida en ausencia de cualquier compartimentación.

El objetivo principal de este proyecto es entender las bases moleculares de la termoforesis en soluciones acuosas diluidas. La termoforesis es en esencia un fenómeno de no equilibrio y algunas teorías moleculares se han centrado en la interpretación de los gradientes de concentración observados en términos de coeficientes de transporte. Sin embargo, experimentos recientes han sugerido que la termoforesis en condiciones de estado estacionario podría entenderse tomando consideraciones termodinámicas en el equilibrio.

Las simulaciones de dinámica molecular son una herramienta muy útil para estudiar estos fenómenos utilizando pequeños sistemas. En esta tesis nos hemos centrado específicamente en pequeñas moléculas anfifílicas en el agua, estas moléculas abarcan características moleculares básicas con partes hidrofílicas e hidrofóbicas. En primer lugar, propusimos una metodología de dinámica molecular robusta y fiable para llevar a cabo la termoforesis en soluciones diluidas realistas, mediante simulaciones de dinámica molecular de “all atom”. En segundo lugar, investigamos las múltiples propiedades físicas que podrían correlacionarse con el coeficiente de Soret (es decir, la magnitud del gradiente de concentración en respuesta a un gradiente de temperatura). Por último, comprobamos la viabilidad de que este proceso térmico pudiera entenderse en términos de termodinámica y cinética en equilibrio, como la literatura ha sugerido anteriormente.

Palabras clave: No equilibrio, coeficiente de Soret, dinámica molecular, difusión, energía libre de solvatación

Contents

Declaration	v
Acknowledgements	vii
RÉSUMÉ	ix
ABSTRACT	xi
RESUMEN	xiii
Contents	xv
List of Figures	xix
List of Tables	xxiii
Prelude	25
1. Introduction	27
1.1 Abiogenesis	30
1.2 Precursor accumulation	32
2. Objective	35
3. The Soret Coefficient	36
3.1 Generalities	36
3.2 From experimental measurements to phenomenological description	38
3.2.1 External factors (concentration, temperature and pressure)	39
3.2.2 Mass effect	41
3.3 Molecular model and theories	45
3.3.1 Temperature dependence of the solvation free energy	45
3.3.2 Temperature dependence of diffusion	49
3.4 Overview	51
4. Molecular dynamics simulations	53
4.1 Brief introduction	53
4.2 Steady state	56
4.3 Molecular models	57
4.4 Temperature gradient	60
5. Thermophoresis simulations	67

5.1	Thermophoretic set-up	67
5.2	Convergence.....	72
5.2.1	Energy conservation.....	72
5.2.2	Convergence of the temperature gradient	74
5.2.3	Convergence of the concentration gradient	78
5.3	Robustness of the results	85
5.3.1	Dependence on the box length (Lz).....	85
5.3.2	Dependence on the gradient amplitude (∇T).....	87
5.3.3	Testing eHEX exchange rate algorithm.....	89
5.3.4	Different force fields.....	91
6.	Intrinsic and thermodynamic conditions modulating the Soret coefficient.....	93
6.1	Low-sensitivity of pressure effect	94
6.2	Temperature effect	96
6.2.1	Ethanol molecule	96
6.2.2	Urea molecule	98
6.2.3	TMAO molecule	100
6.2.4	Empirical equations	102
6.3	Concentration effect	106
6.4	Solute effect.....	109
6.5	Mass effect	118
6.5.1	TMAO molecule	119
6.5.2	Urea molecule	124
6.5.3	Methanol molecule.....	129
6.5.4	Water molecule	135
6.5.5	Mass effect conclusion.....	138
7.	Molecular origins of thermophoresis	141
7.1	Thermophoresis as a result of temperature dependence solvation free energy	141
7.1.1	Simulation details.....	144
7.1.2	Results.....	147
7.2	Diffusion effect (in process).....	154
7.2.1	Simulation details.....	155
7.2.2	Diffusion regime results.....	158
7.2.3	Ballistic regime	163

8. Conclusions.....	165
8.1 In silico thermophoresis set-up	165
8.2 Intrinsic and thermodynamic contributions.....	167
8.3 Molecular models.....	170
9. Résumé de la thèse.....	174
9.1 Introduction	174
9.2 L'effet Soret et les dynamiques de la thermophorèse	176
9.3 L'effet de Soret : Propriétés des molécules et conditions thermodynamiques	183
9.4 Origins moléculaires de la thermophorèse	188
References.....	191

List of Figures

Figure 1.1: Electrophoresis vs thermophoresis	27
Figure 1.2: Binary mixture under a thermophoretic gradient	28
Figure 1.3: Hydrothermal vent at the seabed. ⁴⁷	31
Figure 1.4: Braun description of precursor accumulation	33
Figure 3.1: Hydration free energy	45
Figure 3.2: Schematic representation of the Eastman idea	46
Figure 4.1: Solutes	57
Figure 4.2: Water models.....	60
Figure 4.3: Thermostats	61
Figure 4.4: eHEX algorithm.	62
Figure 5.1: Schematic representation of flat concentration profile.....	68
Figure 5.2: Thermostatted regions	70
Figure 5.3: Energy conservation study	73
Figure 5.4: Convergence of the water temperature-gradient	77
Figure 5.5: Convergence of the TMAO concentration-gradient (temperature).....	79
Figure 5.6: Convergence of the TMAO concentration-gradient (concentration)	81
Figure 5.7: Convergence of the TMAO concentration-gradient (length)	82
Figure 5.8: Convergence of the TMAO concentration-gradient (size).....	84
Figure 5.9: Box size dependence	86
Figure 5.10: Gradient amplitude dependence	88
Figure 5.11: Heat exchange rate dependence	90
Figure 5.12: Force fields dependence	91
Figure 6.1: Median pressure dependence.....	95
Figure 6.2: Median temperature dependence of ethanol.....	97
Figure 6.3: Median temperature dependence of urea.....	99
Figure 6.4: Median temperature dependence of TMAO.....	101

Figure 6.5: Empirical fits	103
Figure 6.6: Average concentration dependence of TMAO.....	107
Figure 6.7: Concentration gradient representations	110
Figure 6.8: Solute effect.....	112
Figure 6.9: Dipole moment dependence of urea	114
Figure 6.10: Mass effect of TMAO	119
Figure 6.11: Mass ratio effect of TMAO	122
Figure 6.12: Mass effect of urea	124
Figure 6.13: Mass ratio effect of urea	126
Figure 6.14: Water mass effect of urea	128
Figure 6.15: Mass effect of methanol	130
Figure 6.16: Water mass effect of methanol	132
Figure 6.17: Urea and methanol mass effect	135
Figure 6.18: Mass effect of isotopic water.....	136
Figure 6.19 : Mass effect summary.....	138
Figure 7.1: (Replot) Hydration free energy	142
Figure 7.2: Schematic representation of our desolvation process used in LAMMPS	145
Figure 7.3 : Free energy methods	147
Figure 7.4: Soret coefficient methods	149
Figure 7.5: Free energy of desolvation vs temperature for different solutes: methanol (yellow), TMAO (magenta) and urea (cyan)	150
Figure 7.6: Free energy of desolvation vs temperature for methanol	151
Figure 7.7: Eastman model	153
Figure 7.8: Uni-dimensional mean squared displacement (MSD) vs time	156
Figure 7.9: MSD for the TMAO solution	157
Figure 7.10: Diffusion coefficients	158
Figure 7.11: Prigogine model	160
Figure 7.12: Mass correction on the Prigogine model.....	162
Figure 7.13: Ballistic regime.....	163
Figure 8.1: Mass effect summary.....	170
Figure 8.2: Eastman model	171

Figure 8.3: Prigogine model	172
Figure 8.4: Ballistic regime.....	173
Figure 9.1: Solutés	180
Figure 9.2: Régions thermostatées	182
Figure 9.3: Résumé de l'effet de la masse	187
Figure 9.4: Modèle d'Eastman.....	188
Figure 9.5: Modèle de Prigogine	189
Figure 9.6: Régime balistique	190

List of Tables

Table 5.1: Summary of the generation of a steady-state thermophoretic setting.....	71
Table 6.1 : Empirical fits.	104
Table 6.2: Solute effect	113
Table 6.3: Mass effect of TMAO.....	120
Table 6.4: Mass ratio effect of TMAO.	123
Table 6.5: Mass effect of urea.....	125
Table 6.6: Mass ratio effect of urea	127
Table 6.7: Water mass effect of urea	129
Table 6.8: Mass effect of methanol.....	131
Table 6.9: Water mass effect of methanol	133
Table 6.10: Urea and methanol mass effect.....	134
Table 6.11: Mass effect of isotopic water	137
Table 6.12: Empirical fit mass effect.....	139

Prelude

The present manuscript will expose the work carried out during my three years PhD, under the supervision of Dr. Guillaume Stirnemann with the grant ERC-ABIOS (Grant agreement ID: 757111). The entire work has been done at the LBT (Laboratoire de Biochimie Théorique) from the IBPC (Institut de Biologie Physico-Chimique) starting in October 2018.

This study is part of a larger project which focuses on the emergence of life, which is one of the most fascinating and yet unsolved questions of natural science. The ultimate goal is to understand the synthesis paths of prebiotic systems under prebiotic conditions. Mainly, the focus is on RNA synthesis in the absence of biological catalysis, where this leads to thermodynamically forbidden reactions, rearrangements and polymerisation paths.

My thesis focuses on a thermally-driven process that is considered as a good candidate in order to explain precursor accumulation in the absence of compartmentalization. While this phenomenon has been experimentally shown to be efficient in the context of prebiotic chemistry, it is still poorly understood at a theoretical level. The goal of this PhD was to provide a better comprehension of the phenomenon in aqueous solutions, starting with simple solutes.

1. Introduction

Thermophoresis is a thermally-driven process, where a mixture of components becomes inhomogeneous under a temperature gradient. It was evidenced for the first time 150 years ago by C. Ludwig,^{1,2} and later by C. Soret.³ This phenomenon manifests itself for a wide range of system sizes. Thermally-driven movements have been studied for molecular systems ranging from the nanometre scale, to systems of the size of microscopic particles,⁴ where thermally-driven processes are still present at millimetric lengths.^{5,6}

This effect is observed because different particles exhibit different responses to the thermal gradient. Perhaps more familiar is the phenomenon of electrophoresis, where the particles move under the influence of a uniform electrical field. Positively charged molecules drift towards the cathode (negatively charged), negatively charged molecules drift towards the anode (positively charged), and neutral molecules are not affected by the electric field since they do not possess a net charge.

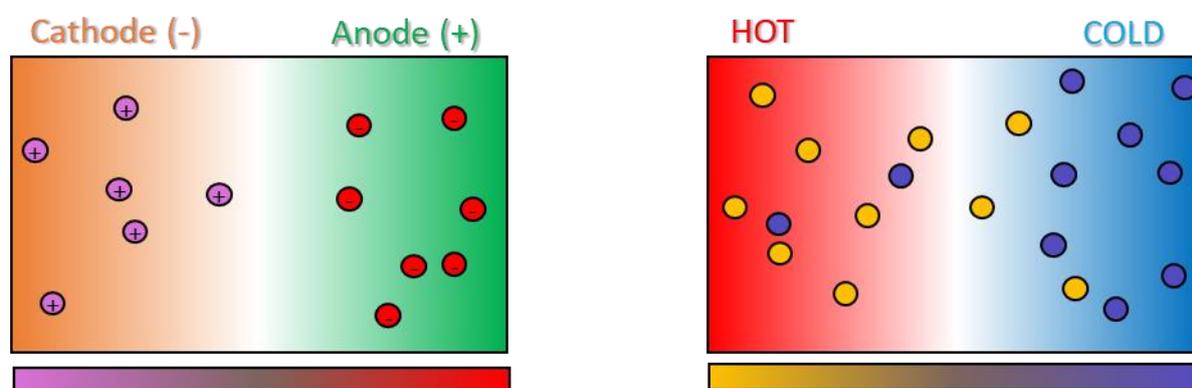


Figure 1.1: Electrophoresis vs thermophoresis. The left graph shows a schematic representation of electrophoresis, cathode (-) region in the left and anode (+) region in the right. Red (-) and pink (+) particles move along the electric field. The right graph shows a schematic representation of thermophoresis, hot region in the left and cold region in the right. Yellow and blue particles move along the thermal gradient. Bottom bars represent the concentration colour of each particle along thermophoretic axis.

Similar effects apply for the thermophoresis, where the particles move according to a thermal field rather than an electrical field as in electrophoresis. However, the “thermal affinity” is not

a simple property as the charge of the particles in electrophoresis. There are several intrinsic properties in the molecule that can affect this thermally-driven process (e.g. mass, size, charge, polarity...) as well as some external factors (e.g. concentration and temperature).⁷ Our work studied some of these effects to better understand the “thermal affinity” of a binary solution. The thermophoresis of a molecule can be achieved simply by inducing a temperature gap between two connected areas.

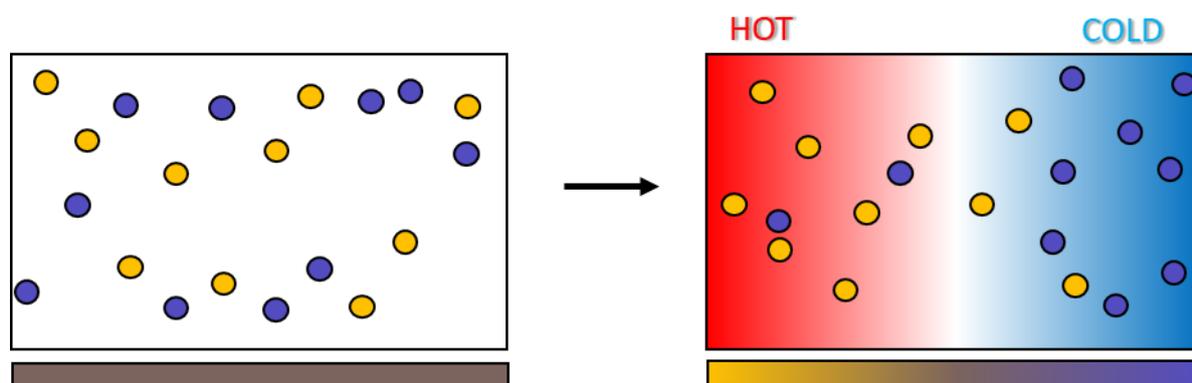


Figure 1.2: Binary mixture under a thermophoretic gradient. The left graph shows the homogeneity of a binary mixture at constant temperature. The right image shows how the binary mixture becomes inhomogeneous under a thermal gradient, with the hot region on the left, and the cold region on the right. The yellow and blue particle represents different molecules.

Figure 1.2 represents a schematic binary system (blue and yellow particles), where the system (a) is in equilibrium at a constant temperature and evolves towards the system (b) once a thermal-gradient is turned on. Since the coloured particles possess different physico-chemical properties, each set of particles behaves differently under the same thermal gradient, thus inducing a non-homogeneous distribution. In this example, yellow and blue particles accumulate in the hot and the cold region respectively. As we can imagine, thermophoresis is present in most processes where spatial variations of temperature are present in the system, leading to a thermal-gradient.

Nevertheless, most of the thermodynamic systems are represented in equilibrium conditions, and their thermodynamic properties are time independent. Under equilibrium conditions, the median temperature is constant and the particle motion follows the Boltzmann distribution:

$$\bar{K} = \frac{n_D}{2} k_B T, \quad (1.1)$$

where \bar{K} is the averaged ensemble kinetic energy, n_D is the number of degrees of freedom, k_B is the Boltzmann constant, and T is the temperature. However, particles do not follow the Boltzmann distribution during thermophoresis as their median temperature is not the same throughout the system. Once a thermal gradient is present in the system, thermally-driven processes disrupt the prior thermodynamic equilibrium, and the system is out of equilibrium. After a transient regime, the system usually reaches a particular state, known as the steady state, in which the system properties do not evolve anymore. The Braun group⁸ has shown in experiments that thermophoresis for moderate temperature gradients follows the Boltzmann distributions once the system reaches the steady state.

In recent years, thermodiffusion has attracted great interest both academically and industrially due to its importance in many scientific and engineering applications. One example is the microscale thermophoresis (MST),^{9,10} which analyses biomolecular systems based on the directed movement of particles in a thermal-gradient, as used e.g. by the Nanotemper company. Temperature control allows real-time and immobilization-free measurements of interaction in liquid molecular complexes at a lower cost and sample consumption compared with similar methods,¹¹ such as surface plasmon resonance (SPR)¹² or isothermal microcalorimetry (IMC).^{13,14}

Another example in the industrial application field is that thermodiffusion, together with gravitational effects, could play an important role in the compositional grading of oil reservoirs^{15,16} induced by the geothermal gradient.^{16,17} Compositional grading has a strong impact on the estimation of the initial (before production) hydrocarbon in place and is important for selecting the best reservoir development scenario. Further studies quantitatively confirmed that thermodiffusion in multi-component mixtures noticeably impacts the distribution of species in a petroleum reservoir and should be taken into account to determine accurately the initial state of a reservoir.^{18–20}

On the other hand, several physico-chemical processes can be studied in the context of life science from a theoretical point of view. There are key new applications for this effect,⁷ including:

- Monitoring protein binding reactions through the sensitivity of thermodiffusion to complex formation.²¹
- Accumulation of a component in microfluidic devices through a combination of thermodiffusion and convection.²²

Therefore, thermophoresis is gaining particular interest in the origins of life context.²³ Recent studies have demonstrated that this thermally-driven process combined with thermal convection could act as a thermal trap and lead to an accumulation of precursors,^{24–26} which is crucial for understanding certain aspects of the origin of life (abiogenesis). The following section will give a brief overview of the origin of life and will explain the fundamental role of thermophoresis in this role.

1.1 Abiogenesis

What is life? How did it come to be? What was the first living organism? Is it possible to find extra-terrestrial life? These questions are still highly debated, without a clear answer, and multiple hypotheses have been proposed.²⁷ Life is a characteristic of biological systems, which distinguishes them from the inorganic matter, but this distinction becomes more vague as the biological system becomes simpler.²⁸ For example, the NASA defined life broadly as:²⁹

'Life is a self-sustained chemical system capable of undergoing Darwinian evolution.'

Over the years, life's definition was expanded:²⁸

'A system which is self-sustaining by utilizing external energy/nutrients owing to its internal process of component production and coupled to the medium via adaptive changes which persist during the time history of the system.'

Even if several definitions can be found in the literature,²⁸ the self-replication process is a key point that differentiates living organism from non-living matter. Self-replication is the ability of dynamical system to replicate itself.³⁰ In general terms, abiogenesis (origin of life) is the process in which life was generated from non-living matter, such as small molecular organic compounds, reaching the capacity to self-sustain and replicate. This process and the required

environmental conditions for life generation are still highly debated,²⁷ with several hypotheses trying to explain this phenomenon.

One of these is the Panspermia hypothesis,³¹ which proposes that life was generated outside the Earth, and that extra-terrestrial living organisms reached Earth vectorised by meteorites. Other hypotheses describe abiogenesis as an intra-terrestrial process. It has been suggested several times that life originated about 3.5 billion years ago.³²⁻³⁴ The best-known theories on the origins of organic compounds are based on the ‘primordial soup’ idea proposed by Oparin in 1924,³⁵ and later supported by the Miller experiment in 1953,^{36,37} where the possibility of generating amino acid from simple molecules available on the early Earth (CH₄, NH₃, H₂O and H₂) was demonstrated by applying an electric discharge. However, the atmospheric composition of the early Earth makes it unlikely that the origins of life can be explained by the yield of Miller experiment.^{38,39}

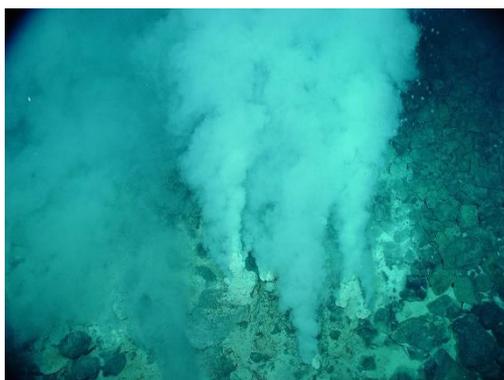


Figure 1.3: Hydrothermal vent at the seabed.⁴⁰

The hydrothermal origin of life theory gained particular interest since the discovery of thermophilic organisms in deep-sea hydrothermal systems in the late 1970s.⁴¹ The perceived benefits afforded to primitive life in this environment include protection from intense asteroid bombardment and ultraviolet radiation, and a source of thermal and chemical energy, along with potentially catalytic minerals.⁴² These optimal conditions led to a new idea that life might have originated in hydrothermal systems on the primitive Earth,⁴³⁻⁴⁵ where hydrothermal vents were proposed as a suitable habitat for the emergence of life,^{43,46,47} and a sustained prebiotic synthesis.

Although the theory of hydrothermal vents gives the optimal conditions for primitive life, an important part of this theory is the accumulation of precursors, where prebiotic molecules accumulate in a defined region inside the hydrothermal vents and the chemical reactions of abiogenesis take place. In the next section, the process of precursor accumulation and its relevance to the origins of life will be explained in more detail.

1.2 Precursor accumulation

One crucial aspect in the origins of life hypothesis is the precursor accumulation, because any chemical reaction requires molecules to come in contact with each other. In modern cells, this question is easily answered, since they are confined by a membraned system in a reduced volume in the 10^{-14} litre range, with high concentrations that critically enhance the reaction rates, and further assisted by an enzymatic system that catalyses the reaction. How can we imagine such efficient chemistry leading to the emergence of life, where the concentrations of precursor molecules are significantly lower and without a complex enzymatic catalyser, since the enzymes were developed long after over the emergence of life?

Consequently, the precursor accumulation in early Earth conditions must have occurred by some physico-chemical processes without the intervention of any biological mechanism, which is present in actual cells. The thermally-driven processes could have played an important role in precursor accumulation, which is essential for some abiogenesis theories. This process does not depend on any structured system, since it requires only an inhomogeneous distribution of temperatures. In 1987, Gaeta *et al.*²³ introduced a connection between precursor accumulation and temperature gradients. However, this hypothesis was not extensively developed until the early 2000s, when pioneering experiment from Braun's group have demonstrated the viability of dilute solute accumulation under thermally-driven processes^{48,49} acting as thermal traps for the oligomerization of RNA strands.^{26,50-52}

The Braun group described the precursor accumulation process as a combination between thermophoresis and thermal convection, which can lead to a precursor migration in a specific region of the system:

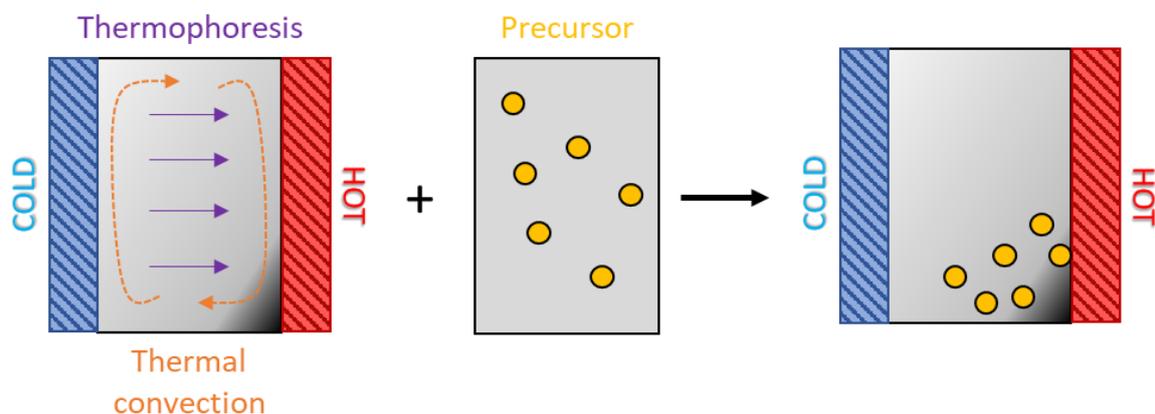


Figure 1.4: Braun description of precursor accumulation. Schematic representation of the precursor (yellow dots) introduced into a system in which thermal convection (orange arrows) and thermophoresis (purple arrows) are generated by a thermal gradient, which results in a precursor accumulation in a specific region of the space, acting as a thermal trap.

These simple systems are not simply physico-chemical lab-inventions, but rather occur in natural thermophoretic settings. One example is the hydrothermal vents that have already been introduced in one of the abiogenesis hypotheses. Hydrothermal vents are systems located at the bottom of the ocean, and are generated by volcanic activity. These systems gain a particular interest due to the large temperature gradients that can be achieved over relatively short distances. Hydrothermal vents are the result of geothermally heated water (400 °C) that is discharged through a hydrothermal system to the oceanic water (2 °C), causing large temperature differences. In addition, the porous rocks of hydrothermal vents on the ocean's floor²⁵ may have also played a key role in the emergence of biological chemistry,^{25,43,53,54} acting as precursor accumulator and catalyst.

The idea of precursor accumulation driven by thermophoretic motion and thermal convection opened a door to new research areas, supported by the hydrothermal vent hypothesis. This particular system has been applied in the context of several life's origins related studies, e.g. trapping nucleotides,^{50,52,55,56} RNA⁵⁷⁻⁶¹ and DNA^{24,26,51,62,63} accumulation and replication, lipid phase transition,^{64,65} as well as other studies outside the origins of life, e.g. colloidal systems,^{48,66-68} charged particles,^{63,68,69} etc.

Currently, the thermally-driven accumulation process of a given molecule remains poorly understood, and the molecular origins of thermophoresis are still unknown. Several theoretical

models have been proposed to describe thermophoresis phenomenologically by explaining the dependence on several molecular properties (e.g. mass, size, charge, polarity...)⁷. Understanding the molecular basis of thermophoresis could shed some light on the process of precursor accumulation in the context of the origins of life, and this thesis makes a contribution in that regard.

2. Objective

Our main goal is to understand the nature of thermophoresis in dilute aqueous solutions. We study aqueous solutions of small amphiphilic molecules under a thermal gradient in order to understand thermally-driven effects that are crucial for precursor accumulation in the origins of life context. We decide to work with such solute molecules in order to avoid pure hydrophobic effects (solute aggregation) or very hydrophilic molecules that strongly interact with water, which we use as solvent. Moreover, the considered molecules are neutral, which prevents electrophoretic effects on top of thermophoretic ones.

The first part of our work consists in establishing a good set-up for thermophoretic simulations. We study the evolution and energy conservation in a system that contains a the thermal gradient and a posterior concentration gradient, by using a non-equilibrium molecular dynamics algorithm based on heat exchange (eHEX).⁷⁰ We focus on the effects of box length, thermal gradient amplitude, heat exchange rate, and the force field dependence in order to test the robustness of our thermophoretic simulations.

In the second part, we analyse thermodynamic conditions (median temperature, pressure, and concentration) and intrinsic molecular properties (mass, charge and solute) that could have some effect in the molecular accumulation in a specific region. These effects are quantified by means of the Soret coefficient, and we compare our results with some phenomenological theories available in the literature.

In the final part, we check the viability of two plausible models, which could be attributed to the molecular origins of the thermophoresis. The first model tries to relate the thermally-driven effect of a molecule to a temperature-dependence of a solvation free-energy process.⁷¹ The second model conjectures that the thermophoretic motion of a molecule is related to the activation energies for the diffusion of the solute and solvent molecules.^{72–76}

3. The Soret Coefficient

3.1 Generalities

The thermophoresis of a binary mixture, e.g. a solute in an aqueous system, results from a competition between the ordinary particle diffusion, described by the diffusion coefficient (D),⁷⁷ and the motion of particles under a thermal gradient, which is phenomenologically described in terms of drift velocity proportional to the temperature gradient through a thermal diffusion coefficient (D_T). The ratio between Fick's diffusion and the thermal diffusion defines the Soret coefficient (S_T), whose sign dictates whether molecules will accumulate in the cold or the hot region, and its value determines the amplitude of the concentration gradient:

$$S_T = \frac{D_T}{D}. \quad (3.1)$$

In this work, we consider a uni-dimensional linear and uniform thermal gradient ∇T along the z axis. The total flux of solute particles moving at a z position $j^s(z)$ can be phenomenologically written as a sum of the diffusion current $j_D(z)$ and thermal diffusion current $j_{D_T}(z)$,^{8,78} such that:

$$j^s(z) = j_D(z) + j_{D_T}(z). \quad (3.2)$$

The first term is the diffusion current, and takes the first Fick law form for dilute concentrations:

$$j_D(z) = -D^s(z)\nabla c^s|_z, \quad (3.3)$$

where D^s is the solute diffusion coefficient, and $\nabla c^s|_z$ is the solute concentration gradient along z . The second term of equation (3.2) is the thermal diffusion current, which for dilute solutions can be written:

$$j_{D_T}(z) = -D_T^s(z)c^s\nabla T, \quad (3.4)$$

where D_T^s is the phenomenological solute thermal diffusion coefficient, c^s is the solute concentration, and ∇T is the thermal gradient.

Experiments show that a steady-state can be reached (more details in Section 4.2), where ordinary diffusion and thermal diffusion are balanced, implying that $j^s(z) = 0$:

$$j_D(z) = -j_{D_T}(z). \quad (3.5)$$

Combining equations (3.3) and (3.4):

$$\left. \frac{\nabla c^s}{c^s} \right|_z = - \frac{D_T^s(z)}{D(z)} \nabla T. \quad (3.6)$$

Since we consider a uni-dimensional thermal gradient, we can thus rewrite the equation as a finite derivative:

$$\left. \frac{dc^s}{c^s} \right|_z = - \frac{D_T^s(z)}{D(z)} dT. \quad (3.7)$$

Rearranging the equation, we obtain the following form:

$$\left. \frac{d \ln c^s}{dT} \right|_z = - \frac{D_T^s(z)}{D(z)}, \quad (3.8)$$

or, equivalently, into:

$$\left. \frac{d \ln c^s}{dT} \right|_z = - S_T^s(z), \quad (3.9)$$

where $S_T^s(z) = - D_T^s(z)/D(z)$ is the uni-dimensional Soret coefficient, which determines the magnitude of thermodiffusion in the steady state. In principle, each position is associated to a given concentration $c_s(z)$ and temperature $T(z)$, and S_T depends on both the concentration and temperature. In the thermophoretic simulations described below, even under pretty significant temperature gradients but more moderate concentration gradients, $\ln c_s$ is often a linear

function of the temperature (T), suggesting that S_T can be considered constant under these thermal gradient conditions. Therefore, if diffusion and thermodiffusion coefficients are taken along the constant gradient, this leads to an exponential depletion law:

$$\frac{c^s}{c_0^s} = \exp(-S_T(T - T_0)), \quad (3.10)$$

with the normalized concentration c^s/c_0^s , which depends on the temperature difference $T - T_0$ only.

3.2 From experimental measurements to phenomenological description

Experimental results show that the thermodiffusion behaviour of a particle depends on a large number of molecular factors.⁷ Mass, size and moment of inertia influence the Soret effect and are independent of particle interactions. On the other hand, heat of transfer, hydrogen bonding, cross interactions, ionic strength and interfacial tension depend on the particle interaction in the solution. Moreover, temperature, pressure and concentration are external factors independent of the nature of the solution that also have an influence in the Soret effect.^{49,79} The complexity of the Soret effect makes the understanding of its molecular origin very arduous.

In this work, we have focused on few key parameters influencing the Soret coefficient and we study their trends for dilute aqueous solutions, where only small amphiphilic neutral solutes are considered. We first analyse the effect of the external factors (temperature, concentration and pressure) that may influence the Soret effect for these systems and compared our results with phenomenological descriptions available in the literature. Furthermore, the mass effect is extensively studied in this work, and we check if some interesting phenomenological approaches are in agreement with our results.

3.2.1 External factors (concentration, temperature and pressure)

It has been observed in many aqueous systems that the Soret coefficient increases with increasing median temperature, often a sign change is observed from a negative $S_T < 0 \text{ K}^{-1}$ (thermophilic behaviour) to a $S_T > 0 \text{ K}^{-1}$ positive value (thermophobic). Moreover, the increase is abrupt at lower median temperatures and becomes flat at high median temperatures. One phenomenological description comes from Iacopini and Piazza,^{79,80} who proposed an empirical equation to model the temperature dependence of the Soret coefficient:

$$S_T(T) = S_T^\infty \left[1 - \exp\left(\frac{T^* - T}{T_i}\right) \right], \quad (3.11)$$

where S_T^∞ represents the limiting value of $S_T(T)$ at high temperatures, T^* the temperature at which the Soret coefficient changes the sign, i.e. $S_T(T) = 0 \text{ K}^{-1}$, and T_i embodies the strength of temperature effect. This equation describes accurately the thermodiffusion of macromolecules in dilute aqueous solutions,^{81,82} but fails with low-molecular-weight mixtures, usually at high concentrations.⁸³ Moreover, this equation also does not describe properly an ethanol/water mixture in a dilute regime,⁸⁴ and molecular dynamics simulations found a monotonous decrease of S_T with the increment of the temperature in agreement with experiments.⁸⁵⁻⁸⁷

Further studies have demonstrated some particularities on the $S_T(T)$ evolution as a function of median temperature. Often there is a sign change from a negative S_T (thermophilic behaviour) to a positive S_T (thermophobic behaviour) for some solutes at high concentrations,⁸⁸⁻⁹⁰ as well as specific concentration points where the S_T becomes independent of the median temperature.⁹⁰ Therefore, concentration effects are also relevant for the explanation of the Soret coefficient. Wittko and Köhler⁹¹ proposed an empirical equation, splitting the Soret coefficient in two additive terms. First, the chemical contribution (S_T^{ch}) is caused by the interactions between the particles in the solution, which depends neither on T nor on c . Second, concentration-dependent function $\alpha_c(c)$ augmented with a temperature-dependent multiplicative amplitude factor $\beta_c(T)$:

$$S_T(c, T) = \alpha_c(c)\beta_T(T) + S_T^{ch}, \quad (3.12)$$

Where S_T^{ch} is the chemical contribution for the Soret coefficient of the different species, $\alpha_c(c)$ and $\beta_T(T)$ are polynomial empirical fits, fitting the concentration and the median temperature respectively:

$$\begin{aligned} \alpha_c(c) &= a_0 + a_1c + a_2c^2 + a_3c^3 + \dots, \\ \beta_T(T) &= 1 + b_1(T - T_0) + b_2(T - T_0)^2 + b_3(T - T_0)^3 + \dots, \end{aligned} \quad (3.13)$$

where T_0 is the reference temperature. The temperature and concentration dependence of aqueous solutions can also be described with these equations, but a_i and b_i fitting values do not explain the physical effect behind these dependences.

The Wiegand group⁹² has recently proposed a modified version for eq. (3.11). The authors assumed that the temperature and the concentration dependence of S_T for aqueous solutions is directly related to the formation and breaking of hydrogen bonds:

$$S_T(c, T) = S_T^\infty - C_H \exp(-A_H T), \quad (3.14)$$

where S_T^∞ are the thermal properties of the molecule, possible charges, and so forth, excluding the presence of hydrogen bonds, C_H is a measure of the number of hydrogen bonds, and the parameter $A_H > 0$ measures the temperature-dependent strength of a hydrogen bond. The temperature-dependent exponential factor describes the diminishing contribution of hydrogen bonds as they weaken with increasing temperature.

Other studies reported phenomenologically that thermophoresis is a consequence of a pressure change generated by the temperature gradient. Semenov and Schimpf⁹³ described the thermophoretic movement of the solute molecule as a net force from the combined action of the local pressure gradient around the molecule and the force caused by the macroscopic pressure gradient. The thermal diffusion coefficient is defined as an empirical ratio between the isobaric thermal expansion and the dynamic viscosity for dilute solutions:

$$D_T = -\frac{8}{27} \frac{A r_s^2}{v_o \eta} \alpha_T, \quad (3.15)$$

where A is Hamaker constant that defines its interaction with the solvent, r_s is the solute radius, v_o is the specific volume of one particle, η is the dynamic viscosity. The authors empirically associated the temperature dependence as an indirect effect of the isobaric thermal expansion coefficient α_T for the thermal diffusion coefficient:

$$\alpha_T \equiv \frac{1}{v} \left(\frac{\partial v}{\partial T} \right)_p = -\frac{1}{\rho} \left(\frac{\partial \rho}{\partial T} \right)_p, \quad (3.16)$$

where v is the specific volume, and ρ is the density. Even if thermophoretic movement described with a thermal expansion coefficient works for non-polar solutes, eq. (3.16) failed for aqueous solutions.⁷ The explanation is due to the thermal expansion coefficient is significantly small for aqueous solutions, which could be related to the stronger hydrogen bonds of aqueous systems,⁹⁴ so associated pressure effects in the Soret coefficient are relatively small for the aqueous systems.

3.2.2 Mass effect

Previously, we introduced the fact that the mass, the moment of inertia and the size are intrinsic properties of the molecule that do not depend on particle interactions. In 1939, Clusius and Dickel⁹⁵ used a continuously thermal convective method to the separation of isotopes. The authors observed that mass has a direct effect to the Soret coefficient by comparing the molar masses of different isotopes. Galliéro *et al.*⁹⁶ justified this behaviour as a density effect, where a thermal gradient causes a concentration gradient in the opposite direction, and the heavier molecules, which occupy the same volume as the lighter ones, participate more efficiently in the generation of the strong density zone and the heavier molecules tend to migrate towards the cold areas more than the lighter ones. Chapman⁹⁷ showed that for equimolar mixtures, it is possible to expand the Soret coefficient in powers of relative mass difference $\delta M = (M_1 - M_2)/(M_1 + M_2)$ for a binary gas mixture:

$$S_T = \frac{S_T^0 \delta M}{T} + \dots \quad (3.17)$$

The mixtures of many linear n-alkanes were also studied,⁹⁸⁻¹⁰⁰ and it was observed that their Soret coefficients depends weakly on the concentration, and their values for equimolar mixtures show a linear evolution as a function of this relative mass difference $\delta M = (M_1 - M_2)/(M_1 + M_2)$. However, the Wiegand group¹⁰¹ has also demonstrated a linear correlation between $\delta I = (I_1 - I_2)/(I_1 + I_2)$ and the Soret coefficient for linear alkanes, making it more difficult to separate the mass and moment of inertia contributions. The explanation of such contribution can be explained in the same manner as the one deduced for the mass alone. Galliéro *et al*⁹⁶ justified this effect as a rotational process, where the molecules with the bigger moment of inertia will rotate more slowly than the other ones and so will tend to migrate, compared to the other species, toward the area of strong density, which is located in the cold part of the simulation box. The Köhler group^{76,102} previously observed that the Soret coefficient can be split in two independent and additive terms for non-polar molecules ($S_t^{ch} + S_t^{iso}$):

$$S_T = S_T^{ch} + S_T^{iso} \quad (3.18)$$

where S_T^{iso} is the isotopic contribution that depends only on mass and moment of inertia, and S_T^{ch} is the chemical contribution, which depends on molecular interactions. These two contributions are independent, which means that identical particles with different masses could exhibit a variation in their respective Soret coefficients. Even if there are several formulations for describing S_T^{iso} term,¹⁰³ the most common is:

$$S_T^{iso} = a_M \frac{M_2 - M_1}{M_2 + M_1} + b_I \frac{I_2 - I_1}{I_2 + I_1} \quad (3.19)$$

where M_i and I_i are the mass and the inertia moment of each particle, and a_M and b_I are empirical coefficients. However, the “chemical” contribution description is harder to define. Artola and Rousseau⁹⁰ computationally described the “chemical” contribution of the Soret coefficient in terms of the Lennard-Jones 12-6 (LJ) potential:

$$LJ_{ij}(r_{ij}) = 4 \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (3.20)$$

where ε_{ij} is the depth of the LJ potential, σ_{ij} is the distance at which the particle-particle potential energy is zero, and r_{ij} is the distance between the i and j particles. Artola and Rousseau demonstrated that the “chemical” contribution of the Soret coefficient is due purely to the nature of these interactions. More precisely, this effect depends on the strength of direct and cross interaction energies between particles, and the interaction energy parameter between different particles (ε_{12}) can be obtained by their own modified version of the Lorentz-Berthelot mixing rule:

$$\varepsilon_{12} = k_{12} \sqrt{\varepsilon_{11} \varepsilon_{22}}, \quad (3.21)$$

where k_{12} controls the strength of the interaction between different particles, the original Lorentz-Berthelot mixing rule¹⁰⁴ is the case that $k_{12} = 1$. Furthermore, the authors included a second parameter which is the energy ratio $\psi_\varepsilon = \varepsilon_{22}/\varepsilon_{11}$, and this parameter indicates the difference of the thermodynamic properties. The Wiegand group¹⁰⁵ experimentally correlated the thermophilicity of peptide model systems to its hydrophilicity, and demonstrated a higher complexity for polar systems as compared to non-polar systems due to the cross interaction effects among these peptides by just including the energy ratio (ψ_ε) of Artola and Rousseau as the “chemical” contribution:

$$S_T^i = a_M \delta M + b_I \delta I + c_\varepsilon \Delta \psi_\varepsilon, \quad (3.22)$$

where $\delta I = (I_2 - I_1)/(I_2 + I_1)$. Even if the Wiegand group determined a_M and b_I for amides, the authors did not find a solution which also covers aqueous solutions of ethanol and glucose. Therefore, the Wiegand group assumed that non-ideal structural changes (e.g., excess effects) of the polar systems are not fully covered by this simple approach.

In our work, we study the mass effect of the Soret coefficient using another strategy. Instead of using molecules chemically different, we focus on the use of isotopes. Basically, isotopes are atoms that differs only in the number of neutrons in the atomic nucleus, and hence atomic properties remains largely unchanged. Therefore, isotopic mixtures allow us to study

chemically equivalent systems at different masses. The first isotopic thermodiffusive experimental studies were made with binary mixtures of deuterated and non-deuterated water (D₂O-H₂O) systems.^{95,106,107} Moreover, several studies have also demonstrated an isotopic thermodiffusive effect in mixtures of benzene of various degrees of deuteration.^{108–112} The Köhler group^{76,91,102,113} has widened the isotopic effect in the Soret coefficient for organic isotopic mixtures of different liquids. The authors have been shown that the isotope substitution (deuteration) of one of the components leads to a constant shift of the Soret coefficient depending neither on concentration nor on the mixing partner.

Most of these systems were organic based mixtures due to the high amount of hydrogen atoms in the molecule that could be deuterated, and hence a wide range of masses could be explored. However, the range of masses are quite limited for small molecules. High isotopic atoms make them often unstable i.e., radioactive isotopes. Even if there are stable isotopes such as ²H, ¹³C, ¹⁵N... their masses only differ from 1 to 2 atomic units of their respective most abundant isotopes i.e., ¹H, ¹²C, ¹⁴N. As a consequence, experimental studies are rather limited by the nature of their systems in order to study similar molecular systems which only differs on mass for small solutes.

On the other hand, molecular dynamics simulations allow us to more freely explore the mass “isotopic” effect in the Soret coefficient. Since atomic mass is just an input parameter, we can artificially modify the mass value without the experimental constrains. Several computational studies have focused on the isotopic substitution’s effect on the Soret coefficient, but are mainly focused on Lennard-Jones model particles,^{86,114–116} which is a convenient representation of an isotopic binary/ternary mixture of noble gases. All their results pointed out to a direct correlation between the relative mass and an increment of the Soret coefficient. In this work, we analyse the viability to describe the mass effect to the Soret coefficient for some small solutes in aqueous solutions by the following equation:

$$S_T = S_T^0 + a_M \frac{M_2 - M_1}{M_2 + M_1}, \quad (3.23)$$

where S_T^0 is the molecular contribution of the Soret coefficient, which encompasses the “chemical” and moment of inertial terms ($S_T^0 = S_T^{ch} + b_I \delta I$) of eq. (3.18) .

3.3 Molecular model and theories

The temperature/concentration and mass dependence studies of the Soret coefficient evidenced some connexion between these factors and the amplitude of the thermal gradient, leading to phenomenological descriptions, as described above. In the literature, we can find some models based on thermodynamic or kinetic theories that try to explain the molecular basis of thermophoretic motion. Two are of particular interest and relevance for molecular systems, but come from different perspectives. The first model tries to relate the thermally-driven effect of a molecule to a temperature dependence of a solvation free-energy process,⁷¹ and the second model conjectures that the thermophoretic motion of a molecule is related to the activation energies for the diffusion of the solute and solvent molecules.⁷²⁻⁷⁶

3.3.1 Temperature dependence of the solvation free energy

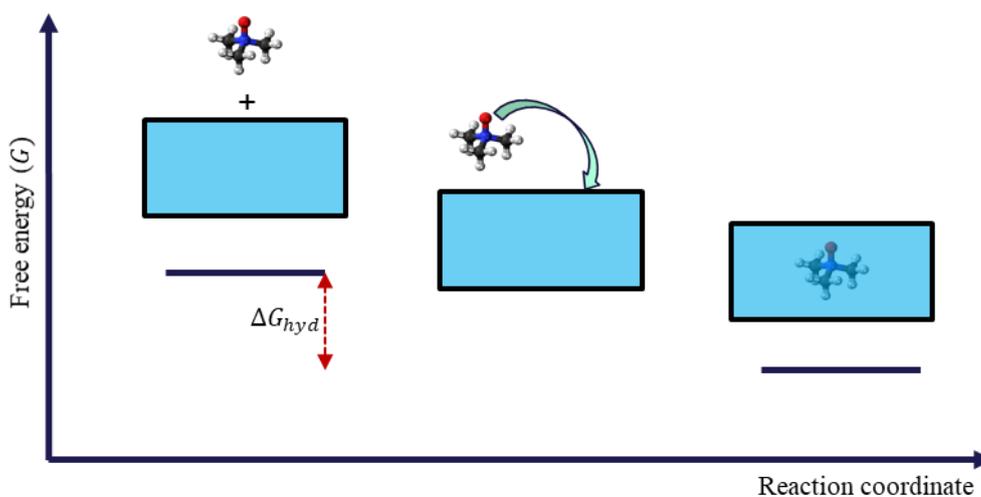


Figure 3.1: Hydration free energy. Schematic representation of the hydration free energy process of the trimethyl ammine N-oxide molecule (TMAO). The blue box represents the liquid phase and outer box the gas phase. ΔG_{hyd} is the hydration free energy of the process.

An appealing idea that was first suggested by Eastman is that the Soret coefficient can be understood in terms of equilibrium thermodynamic considerations, relating the Soret coefficient to the temperature-dependence of the solute hydration free-energy. The hydration free-energy (ΔG_{hyd}) is the difference of energy between a molecule in the gas phase and the same molecule in a liquid phase, which is surrounded by solvent, see Figure 3.1.

In an early approach, Eastman⁷¹ considered that the probability of a particle moving along a temperature gradient is related to the corresponding entropy change, which is absorbed or released by the surrounding liquid. Recalling that the Soret effect is a “steady state” condition in which a concentration gradient arises from a temperature gradient. The transfer of any particle from one region towards another one can only be considered as an equilibrium process, if no net change in entropy results from this virtual transfer. Eastman divided the total entropy variation of the system in two parts; the entropy of the reservoirs (S_R) and the entropy of the system (S_{Sys}). Therefore, the author described this virtual transfer under equilibrium conditions by the following equation:

$$\delta S_R + \delta S_{Sys} = 0, \quad (3.24)$$

The removal of a solute from one region at a certain temperature (T_1) produce an infinitesimal change in the concentration, see Figure 3.2.

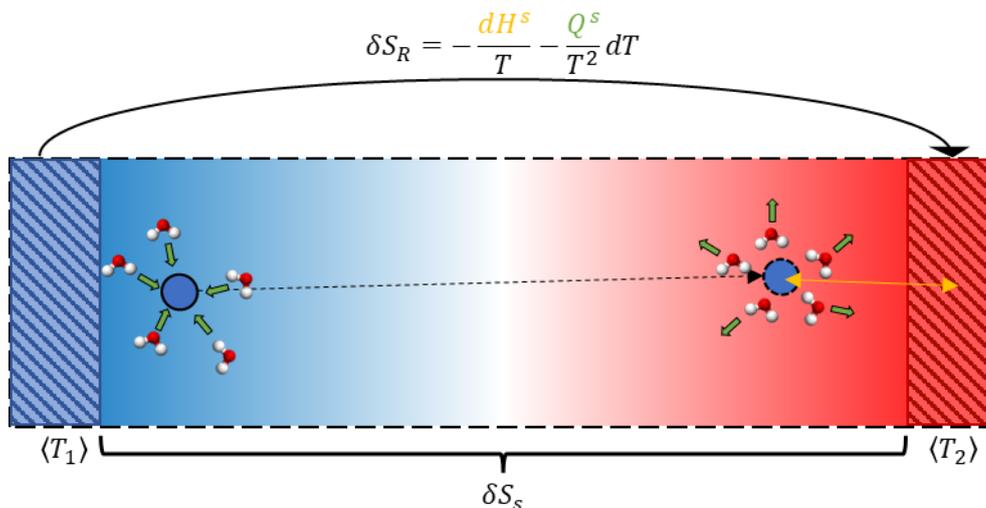


Figure 3.2: Schematic representation of the Eastman idea. The blue particle represents the virtual transfer from the cold region to the hot region. The green arrows represent the heat effect (Q^s) of the solution rearrangement. The yellow arrow represents the partial molar heat (dH^s) exchanged with the reservoir.

This change may be assumed to give rise to a corresponding change in the intensity of intermolecular forces in general terms. The rearrangements resulting from the change in forces may result in a finite heat effect (Q^s). Moreover, there is a heat absorbed by the system from

the reservoirs in this process in partial molal heat terms (dH^s), since the virtual transferred particle has to reach the corresponding kinetic energy of the temperature (T_2) of the new region.

The total change in entropy for the reservoirs is then:

$$\delta S_R = -\frac{dH^s}{T} - \frac{Q^s}{T^2} dT = -\left(\frac{1}{T}\right) \left[\left(\frac{\partial H^s}{\partial T}\right)_{P,N} dT + \left(\frac{\partial H^s}{\partial N^s}\right)_{P,T} dN^s \right] - \frac{Q^s}{T^2} dT, \quad (3.25)$$

and the entropy change of the system is:

$$\delta S_{sys} = dS^s = \left(\frac{\partial S^s}{\partial T}\right)_{P,N} dT + \left(\frac{\partial S^s}{\partial N^s}\right)_{P,T} dN^s. \quad (3.26)$$

Replacing both terms of the equation (3.24) yields to:

$$dG^s = \left(\frac{\partial G^s}{\partial T}\right)_{P,N} dT + \left(\frac{\partial G^s}{\partial N^s}\right)_{P,T} dN^s = -\frac{Q^s}{T} dT, \quad (3.27)$$

where G^s represents a solute single-particle free enthalpy. Eq. (3.27) shows that the solutions whose internal dynamic conditions are independent of the concentration ($(\partial G^s / \partial N^s)_{P,T} = 0$) at a constant temperature ($(\partial G^s / \partial T^s)_{P,N} = 0$), the exchanged heat (Q^s) is zero. In these ideal solutions the Soret coefficient, governed by Eq. (3.27) ($dG^s / dT = -Q^s / T$), is therefore zero. In the case of non-ideal solutions, Eastman considered that Q^s must reach a maximum at some finite concentration, which implies that the partial free energy term of concentration dependence is non-zero ($(\partial G^s / \partial N^s)_{P,T} \neq 0$). The author associated this effect to an irreversible process, which defines the Soret coefficient value. Eastman justified this irreversible process as a pressure effect for gaseous solutions. However, the author noted that liquid solutions present a more difficult problem, since their irreversible effects may be less important but still relevant. Therefore, Eastman defined the Soret effects as a combination of a reversible (S_T^r) and irreversible (S_T^i) parts:

$$S_T = S_T^r + S_T^i. \quad (3.28)$$

Eastman generalized the Eq. (3.27) to define the reversible and irreversible Soret coefficients, as follows:

$$S_T^r = \frac{(dG^s)_r}{dT} = -\frac{Q^s}{T}, \quad (3.29)$$

$$S_T^i = \frac{(dG^s)_i}{dT}, \quad (3.30)$$

where $(dG^s)_r$ and $(dG^s)_i$ are the reversible and irreversible free energies respectively. Replacing both terms of the eq. (3.33) yields to:

$$S_T = \frac{(dG^s)_r}{dT} + \frac{(dG^s)_i}{dT}. \quad (3.31)$$

In 2005, the Braun group⁴⁹ extended this previous idea and their experimental results shown that the thermophoretic steady states could be described by a Boltzmann distribution in local thermodynamics equilibrium of the equation (3.10) ($\exp(-S_T(T - T_0))$). Even if thermophoresis is a local non-equilibrium effect, the Braun group demonstrated that for moderate temperature gradients the thermal fluctuations of the molecule are still a local equilibrium.⁸ This allows the description of the thermodiffusion in a steady state by a succession of local Boltzmann laws:

$$\frac{c^s}{c_0^s} = \exp\left(-\frac{(G^s(T) - G^s(T_0))}{k_B T}\right). \quad (3.32)$$

Their experiments demonstrated that the Soret coefficients of carboxyl-modified beads of different radii scales with its particle surface, resulting in a direct relation between the Soret coefficient (S_T) and the solute particle free enthalpy (G^s), combining equations (3.10) and (3.32):

$$\frac{D_T^s}{D^s}(T - T_0) = \frac{(G^s(T) - G^s(T_0))}{RT}, \quad (3.33)$$

which can be rewritten as $S_T = D_T^s/D^s$ and as a function of partial derivatives dG^s/dT

$$S_T = \frac{1}{RT} \frac{dG^s}{dT}. \quad (3.34)$$

This relation can be generalized by locally applying the thermodynamic relation $dG \equiv -SdT + Vdp + \mu dN$. For single molecules at a constant pressure, the Soret coefficient is related to the solvation entropy of the solute particle in the solvent ($S_T = -S/RT$).

3.3.2 Temperature dependence of diffusion

A very different approach was initiated by Prigogine.^{72,73} The thermal diffusion process is described as an activated process in a thermal gradient, where the mass fluxes are expressed as a function of the activation energies for the Fick's diffusion of the different species. Their temperature dependence can be represented as an Arrhenius form:

$$D^s = D_0^s \exp\left(-\frac{E_a^s}{RT}\right), \quad (3.35)$$

where E_a^s is the activation energy for diffusion. In Prigogine's approach, this energy is seen as the sum of the energy needed to form a hole in which the molecule will be placed, and the energy needed for the particle to detach itself from its neighbours. The diffusive flux of the solute molecule can be described as:

$$j^s(z) = -\frac{1}{N} D_0^s \exp\left(-\frac{E_a^s}{RT}\right) \frac{\partial N^s}{\partial z}. \quad (3.36)$$

In the steady state of thermodiffusion, Prigogine defined that the diffusive flux of any molecule in the system is zero, and therefore the motion of a molecule from one position to another is just defined by a difference in the activation energies between the molecules in the solution:

$$\frac{1}{N^s N^w} \frac{\partial N^s}{\partial z} = \frac{E_a^w - E_a^s}{RT^2} \frac{\partial T}{\partial z}. \quad (3.37)$$

Recalling the equation (3.9), and expressing the number of molecules (N^s) in terms of the solute concentration (c^s), the Soret coefficient of a binary mixture can be written as a simple relation between the activation energies of the two components:

$$S_T = \frac{E_a^s - E_a^w}{RT^2}. \quad (3.38)$$

We note that the original Prigogine equation presents a sign error, because the Prigogine's Soret coefficient is defined as $(E_a^w - E_a^s)/RT^2$. This description yields to a negative Soret coefficient value for the solute. However, the solute should present a positive Soret coefficient, since the activation energy of the solute is usually higher than the solvent. Therefore, we switch the activation energies $(E_a^s - E_a^w)/RT^2$ of equations (3.37) and (3.38) to be in accordance with the thermophilic or thermophobic character of the solute. Furthermore, eq. (3.38) does not take into account explicitly the mass effect of the Soret coefficient. Artola, Rousseau and Galliéro⁷⁵ rescaled the temperature in terms of the ideal part of the partitionial function means to displace the position of the diffusion transition state as a barycentre of the relative masses. The authors defined a positive mass flux as:

$$J^+(z) \propto c^w \left(z - \frac{dz}{2} \right) c^s \left(z + \frac{dz}{2} \right) \exp\left(-\frac{E_a^w}{R(T - \xi^w dT)}\right) \exp\left(-\frac{E_a^s}{R(T + \xi^s dT)}\right), \quad (3.39)$$

where ξ^i which defines the transition state position, is given by:

$$\xi^w = \frac{M^w}{M^s + M^w}, \quad (3.40)$$

and the same for ξ^s . In the steady state of thermodiffusion, the mass fluxes compensate ($J^+(z) = -J^-(z)$), and the authors obtained the following expression of the Soret coefficient:

$$S_T = \frac{E_a^s - E_a^w}{RT^2} + \frac{E_a^s + E_a^w}{RT^2} \frac{M_s - M_w}{M_s + M_w}, \quad (3.41)$$

where the first term is the Prigogine model description of the Soret coefficient, eq. (3.38), which may describe the chemical contribution (S_T^{ch}). After a discussion with Guillaume Galliéro, the third author of this model, he agrees that the sign problem of the Prigogine's approach is still present in their article.⁷⁵ For this reason, we also switch the activation energies for the eq. (3.41). Although Artola, Rousseau and Galliéro included the mass effect in Prigogine's approach, which improved the description of S_T as an activated process, it is shown that their predictions of the Soret coefficient were underestimated for water-ethanol mixtures.⁷⁵

3.4 Overview

The Soret coefficient has been studied from different perspectives, but there is not a general consensus on its molecular origins. The following insights can be extracted:

Phenomenological descriptions

- The Soret coefficient exhibits a clear dependence on **temperature** and **concentration**, but a clear molecular picture is missing. For example, the Soret coefficient values reach a certain saturation point at higher temperatures, and under certain temperature and concentration conditions, the molecules could change from the cold region to the hot region, which means that the molecule is no longer thermophobic and becomes thermophilic:^{79,80}

$$S_T(T) = S_T^\infty \left[1 - \exp\left(-\frac{T^* - T}{T_i}\right) \right]. \quad (3.42)$$

- A **chemical contribution** that is neither concentration nor temperature dependent (S_T^{ch}) often appears in some descriptions and is usually associated with solute and solvent interactions.⁹⁰
- The **mass effect** seems to play an important role in the Soret coefficient, which depends on the mass ratio of the components of the binary mixture.^{76,102,103}

$$S_T = S_T^0 + a_M \frac{M_2 - M_1}{M_2 + M_1}. \quad (3.43)$$

Molecular models

- Eastman's thermodynamic approach defines the molecular origins of the Soret coefficient as a temperature-dependent **free energy solvation**:⁷¹

$$S_T = \frac{1}{RT} \frac{dG^s}{dT}. \quad (3.44)$$

- Prigogine's kinetic point describes the molecular origins of the Soret coefficient as an **activation process of molecular diffusion**:^{72,73}

$$S_T = \frac{E_a^s - E_a^w}{RT^2}. \quad (3.45)$$

This overview summarizes the most relevant molecular properties and external factors study in this work as well as introduces two plausible molecular models that could describe the Soret coefficient. The next chapter will discuss the technical part of the thermophoretic study.

4. Molecular dynamics simulations

4.1 Brief introduction

Molecular dynamics (MD) is an important computational technique, used to study condensed matter *in silico* at the molecular level. The idea of MD is simple: atoms or molecules are considered as point particles, interacting with each other via molecular interactions. Their dynamical interactions are solved numerically by integrating the equations of motion of Newtonian dynamics in discrete time steps. The basic ingredients of MD can be summarized as follow

- Model describing the interaction between particles (force field)
- Calculation of energies and forces from the model
- Integration of the equations of motions

There are several types of MD simulations depending on the system size and the accuracy level. *Ab-initio* MDs¹¹⁷ are high-level precision simulations, where the forces between atoms are calculated at the quantum level, commonly by the density functional theory (DFT).¹¹⁸. The DFT method works in terms of electronic density rather than discretized electrons, and thus reduces computational cost. However, these simulations are limited to a few hundred of picoseconds and short length scales, because they still require high computational cost. On the opposite side, coarse-grained simulations represent groups of atoms as a single particle, with an averaged chemical character of a chunk of atoms. This leads to a reduction in the number of interactions among the many-particles system, allowing faster simulations exploration of large systems, such as biological systems: membranes or protein folding. Nevertheless, since our main goal is understanding thermophoresis at the molecular level, the target molecules must translate, rotate, and interact classically with the solvent and other particles. For these reasons, all-atom MD simulations are at the optimal accuracy level for our purposes, representing the system at the atomistic level, whereby each atom is represented as a particle. Simulations in this range can easily reach the microsecond timescale, with many-particle systems made of several thousands of interacting particles.

Molecular dynamics is a powerful tool for studying the physico-chemical processes that occur in nature at the molecular level. Some of these processes are for example: the diffusion of a particle, the viscosity of the medium, thermal conductivity, adsorption processes, dielectrical properties of the system... Most of these properties are usually studied at equilibrium and are time-independent. Unlike thermophoresis, which is a phoretic response of particles to a thermal gradient, moving particles from one region to another as a function of time. Consequently, thermophoresis can only be observed outside equilibrium.¹¹⁹ Therefore, non-equilibrium molecular dynamics is a suitable method for physical systems that are not in thermodynamic equilibrium. Time-dependent effects or chemical reactions rates are some examples that are studied outside equilibrium. In what follows, we inspect whether thermophoresis could be studied as a time-independent phenomenon.

To answer this question, we need to understand what is the main difference between equilibrium and non-equilibrium thermodynamics. This difference can be described thermodynamically, by resorting the second law of thermodynamics, which introduces the thermodynamic variable called the entropy S :

$$\frac{dS}{dt} \geq 0. \quad (4.1)$$

Equation (4.1) shows that any time-dependent change generates a change in entropy higher or equal than zero. A non-equilibrium process is subject to an increase in entropy, and the evolution between two states becomes irreversible. On the other hand, equilibrium thermodynamics is the systematic study of transformation of matter and energy at constant total entropy. Once the system is in thermodynamic equilibrium, all the processes inside this system are reversible, and entropy remains constant.¹²⁰ Consequently, the properties of system at “equilibrium” do not depend on time:

$$\frac{dS}{dt} = 0. \quad (4.2)$$

Statistical mechanics defines entropy as a number of microscopic configurations (Ω) that represent a mechanical system defined by macroscopic variables. Maxwell-Boltzmann

theory¹²⁰ defines the relation between entropy and possible microstates for a mechanical system as:

$$S = k_B \ln \Omega_m , \quad (4.3)$$

where Ω_m is the number of possible microstates. This equation assumes that all microstates have the same probability of occurrence. Thermodynamic ensembles are a particular variety of statistical ensemble that can be described a finite set of thermodynamic variables (i.e., energy, number of particles, temperature, pressure, volume...). One example is the microcanonical ensemble (NVE), which represents possible states of a mechanical system whose number of particles (N), volume (V), and total energy (E) are constant. The entropy un such a system is thus a function of those variables:

$$S = S(N, V, E) . \quad (4.4)$$

Sometimes the total energy of the system is unknown, while the temperature is known. This situation is in fact more common in reality than knowing the energy of the system. These systems are described by another ensemble, which is the canonical ensemble (NVT) that fixes temperature instead of energy, and allows the energy to fluctuate. Furthermore, another common ensemble exists that fixes the pressure instead of the volume, and it is known as the isothermal–isobaric ensemble (NPT). In this work, we make us of all three ensembles; NVE, NVT and NPT.

Within the framework of a statistical-mechanical ensemble, the connection between macroscopic variables and microscopic configurations are described statistically with their respective partition functions. For a canonical ensemble (NVT), which is classical and discrete, the canonical partition function (Z) is defined as:

$$Z = e^{-\frac{F}{RT}} , \quad (4.5)$$

where $F(NVT)$ is the Helmholtz free energy. Once Z is known, it is straightforward to compute the average energy and other thermodynamic quantities.

In fact, the idea behind equilibrium MD simulations is precisely that we can study the average property of a many-particle system in the phase space by averaging the quantity of interest over a sufficiently long time (t):¹²¹

$$\bar{a}_i = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' a_i(t'), \quad (4.6)$$

where a_i is a time-dependent quantity, and \bar{a}_i its temporal average. Moreover, the ergodic hypothesis¹²² enables averaging a stochastic process as a function of a large collection of random samples (N), instead of averaging over the time:

$$\langle a_i \rangle = \frac{1}{N} \sum_{i=1}^N a_i, \quad (4.7)$$

where N is the number of samples. This can be done, because there is a one-to-one correspondence between the initial phase space coordinates of a system and those specified the state of the system at late time (t'). Therefore, according to the ergodic principle, averaging over the time evolved phase space coordinates is equivalent to averaging over all initial phase space coordinates. Combining time-average and sample average, we can enhance the sampling, by the following expression:

$$\bar{a}_i = \frac{\sum_N \left(\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' a_i(t') \right)}{N}. \quad (4.8)$$

4.2 Steady state

The ergodic hypothesis is no longer valid for non-equilibrium MD because system properties are expected to vary along simulation time. Therefore, some thermodynamic states become more complex to define, as they are not constant, and the system cannot be studied in a thermodynamic ensemble. A non-equilibrium system can reach a particular state that is close enough to thermodynamic equilibrium, and its non-equilibrium properties can be measured locally using equilibrium measurement techniques.

To define properties in non-equilibrium thermodynamics, the assumption of local equilibrium is essential. It allows the application of the statistical ensemble as well as ergodicity. Due to Onsager's variational principle,¹²³ the entropy must be minimised. The non-equilibrium steady state is one in which the least amount of total entropy is produced. The steady state can be reached after a period of relaxation as long as the external perturbation is constant over time. In the steady state, macroscopic parameters such as temperature, concentration and pressure at each point of the thermodynamically open or closed system are time-independent values. For example, if a constant heat flux is applied across two separate spatial regions in an aqueous solution, the parameters of the system will approach time-invariant values. Thus, the equilibrium theory can be used in order to provide a thermodynamic framework for determining the composition gradient across a thermal gradient.

4.3 Molecular models

Our work focuses on dilute aqueous solutions, and to avoid working with very large systems, we have selected small polar solutes. Unlike completely non-polar solutes, these types of molecules usually do not aggregate at low concentrations, which is crucial for determining variations in concentration gradient during thermophoresis, because are solely due to the temperature-gradient and are not further enhanced by non-thermophoretic aggregation. A natural choice for such polar solutes are amphiphilic molecules. These molecules contain a hydrophilic head and a hydrophobic body. Moreover, we seek solutes sufficiently different in order to observe different behaviours under the thermal gradient, and hence have significantly different Soret coefficients.

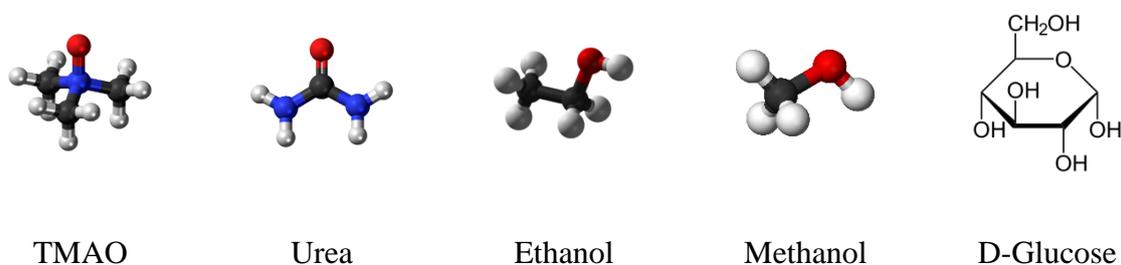


Figure 4.1: Solutes. Three-dimensional structures from left to right for trimethylamine N-oxide (TMAO), urea, ethanol and methanol. Lewis's structure of D-Glucose.

Our main solute is trimethylamine N-oxide (TMAO), which has interesting properties in the biochemical processes related to deep ocean conditions.^{124,125} It is a small solute that offers both a very polarized hydrophilic moiety, as well as bulky hydrophobic region. In addition, TMAO is large enough to show a clear accumulation behaviour under thermal gradient, and it is known not to aggregate unless very high concentrations, much higher than those used in this study, are used.^{126,127} The Kast TMAO force field model^{128,129} is employed in our simulations. It is a non-polarisable force field and has demonstrated high stability over a wide range of temperature and pressure. In addition, its results come in close comparison to the *ab-initio* molecular dynamics simulations and experimental data.¹³⁰

Urea is a planar organic molecule, widely used as a protein denaturant in aqueous solutions, as well as in studies on protein folding and stability.¹³¹ Urea is also a well-known solubilizer,¹³² and it might have been useful to some molecular mechanisms linked to the origins of life,¹³³ acting in an urea-mediated phosphoryl transfer mechanism.¹³⁴ Here, we select a new nonpolarizable force field for mixtures of urea and water available in the literature. The Smith group^{135,136} parametrized it to reproduce the experimental Kirkwood-Buff (KB) integrals, as the KB integrals were observed to be most sensitive to charge distribution employed and the authors experimentally well reproduced urea changes in his activity. The force field parameters from the Smith group¹³⁵ are employed for our urea molecule, but this version of the urea force field constrained the bonds with the SHAKE algorithm, and we want a fully unconstrained molecule. For this reason, the bond stretching parameters are obtained from a posterior work from the Shimizu group,¹³⁶ which are also parametrized by experimental Kirkwood-Buff (KB) integrals. We analysed both versions of the force field, and we did not observe any relevant difference if the bonded parameters are constrained or not.

Ethanol and methanol are selected because they are the smallest amphiphilic organic molecules, and they were well-suited for observing the limits of thermophoretic calculations. The CHARMM General Force Field (CGenFF) is the chosen force field for ethanol,¹³⁷ and an OPLS based force field is selected for methanol.¹³⁸ We consider that these general force fields describe simple molecules such as ethanol and methanol with sufficient accuracy.

Glucose is an important carbohydrate, relevant both for its biological functions¹³⁹ and for industrial processes.¹⁴⁰ As a monomer of cellulose, the most abundant biopolymer, it is the

most important source of energy in all organisms as well as an alternative feedstock for fuels and chemicals in the biorefinery framework.^{139,141} In general, most force fields predict self-aggregation of solute molecules due to the overestimation of solute–solute interactions for aqueous solutions of carbohydrates.¹⁴² To alleviate this effect, The Vlugt group optimized nonbonded interactions of the OPLS Force Field.¹⁴³ The authors scaled the nonbonded interaction parameters of sucrose, a disaccharide which is composed by a fructose and a glucose. The scaling factors were chosen in such a way that experimental thermodynamic and transport properties of aqueous solutions of sucrose are accurately reproduced. The Vlugt group found excellent agreement between experimental and computed liquid densities, thermodynamic factors, shear viscosities, self-diffusion coefficients, and Fick (mutual) diffusion coefficient.

In the case of water, it is of vital importance to choose a force field that is valid over a wide range of temperature and pressure, which accurately describes the physico-chemical properties of water, and which is compatible with other force fields of molecules of interest. TIP4P/2005¹⁴⁴ is the principal water model for our work, which is a rigid four-site model, consisting of three fixed point charges and one Lennard-Jones centre, see Figure 4.2. As an important detail, we noted that this model leads to Coulombic cut-off extension of a distance of $2 * (OM \text{ distance})$ (Oxygen-Massless dummy charge),¹⁴⁴ due to the extra distance of the fictional oxygen charge in this water model. Thus, it is typically best sense to extend Lennard-Jones potential by the following equation to improve efficiency: $LJ_{cutoff} = Coul_{cutoff} + 2 * (O - M)$.

Moreover, the TIP4P/2005 model has a good performance regarding several physico-chemical characteristics of water;¹⁴⁴ thermodynamic properties of the liquid and solid phases, phase diagram involving condensed phases, properties at melting and vaporization, dielectric constant, pair distribution function, and self-diffusion coefficient. These properties cover a temperature range from 123 to 573 K and pressures up to 40 000 bar. Among all the four-site models, it is probably the best at reproducing water thermodynamical, structural and dynamical properties, at least when one does not deviate too much from the ambient conditions.



Figure 4.2: Water models. The left water representation is the 4-site water model (TIP4P 2005). The right water representation is the 3-site water model (SPC/E).

We have occasionally performed some extra calculations with the SPC/E water model,¹⁴⁵ a 3-site model, Figure 4.2, which is an older but popular water model.

4.4 Temperature gradient

Our simulations require a stable and well-defined temperature gradient in order to understand the thermophoretic forces applied to particles, and to study their motion. It is well known that equilibrium MD simulations can be run at constant temperature by the implementation of thermostats for several ensembles, i.e., canonical (NVT) or isothermal-isobaric (NPT). While there are several thermostat implementations (Berendsen,¹⁴⁶ Andersen¹⁴⁷ and Langevin¹⁴⁸), the most commonly used is the Nose-Hoover thermostat,^{121,149} which defines an external bath, where particle motion inside the system is affected by the following Hamiltonian:

$$H_{Nose} = \sum_{i=1}^N \frac{p_i^2}{2m_i s^2} + U(r^N) + g k_B T \ln(s) + \left(\frac{p_s^2}{2Q} \right), \quad (4.9)$$

where H_{Nose} is the Hamiltonian of the extended system of N particles, p_i is the momentum and m_i the mass of the particle, s is an extra degree of freedom for the heat bath, $U(r^N)$ is the potential energy, $g = 3N$ for canonical ensembles, k_B is the Boltzmann constant, T is the temperature, p_s is the heat bath momentum, and Q its imaginary mass.

We can split the equation (4.9) in two parts, where the first part describes the motion of particles, and the second part is the extra motion effect due to the external bath. This method was designed to keep temperature throughout the system equal by applying an external thermal

bath for all particles. However, thermophoretic simulations require a spatially resolved thermostat, where particles can be discretized depending on their motion and space region in the system, because different particles must be at different temperature at the same time.

For this reason, at least two regions in the system must be generated, one at lower temperature (cold region), and the other at high temperature (hot region). Once these two regions are well defined, cold and hot slabs can be achieved by thermostatting particles belonging to these two regions.

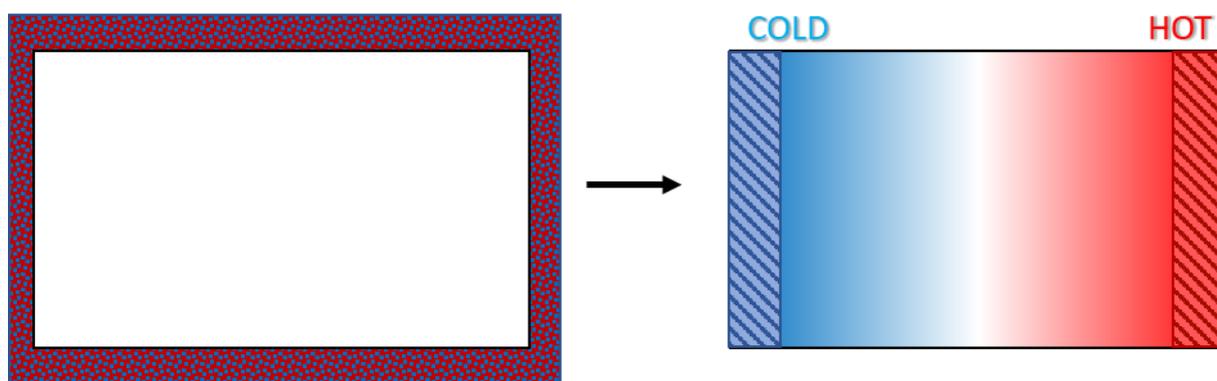


Figure 4.3: Thermostats. The left picture is a schematic representation of a simulation box under a thermal bath. The right picture represents two thermostatted regions.

Current implementations of thermostats applied to different space regions require the definition of a set of thermostatted particles before running the simulation, which cannot be directly updated each simulation step. Since some of these molecules will leave or enter the slab at a given timestep, these algorithms are not feasible. Consequently, harmonic restraints are usually employed to constrain solvent particles in the thermostatted regions.^{79,150,151} Even though not all particles in the thermostatted region are affected, the constrained particles can still affect molecular motions along the direction of the gradient, producing artefacts.

Another possible approach is the heat-exchange algorithm (HEX) introduced by Hafskjold and Ikesoji,¹⁵² where an algorithm transfers kinetic energy in the form of heat instead of controlling temperature inside each slab. This algorithm extracts a certain amount of heat $-\Delta Q_{\Gamma_{cold}}$ from the designed cold region, and injects the same amount of heat $\Delta Q_{\Gamma_{hot}}$ into the corresponding hot region. As a consequence, the temperature difference between the hot/cold regions is a consequence of heat transfer instead of a thermostatic effect. Hence, modulating the

amount/frequency of heat-exchanged transfers leads to a larger or smaller temperature gradient indirectly.

In these simulations, no temperature control is applied since the simulations are performed in the microcanonical ensemble (NVE), posing an energy conservation problem. For instance, adding the HEX algorithm on top of independent factors that could lead to poor conservation in equilibrium simulation in the microcanonical ensemble could aggravate total energy conservation. However, it has been shown that many versions of this algorithm evoke energy-

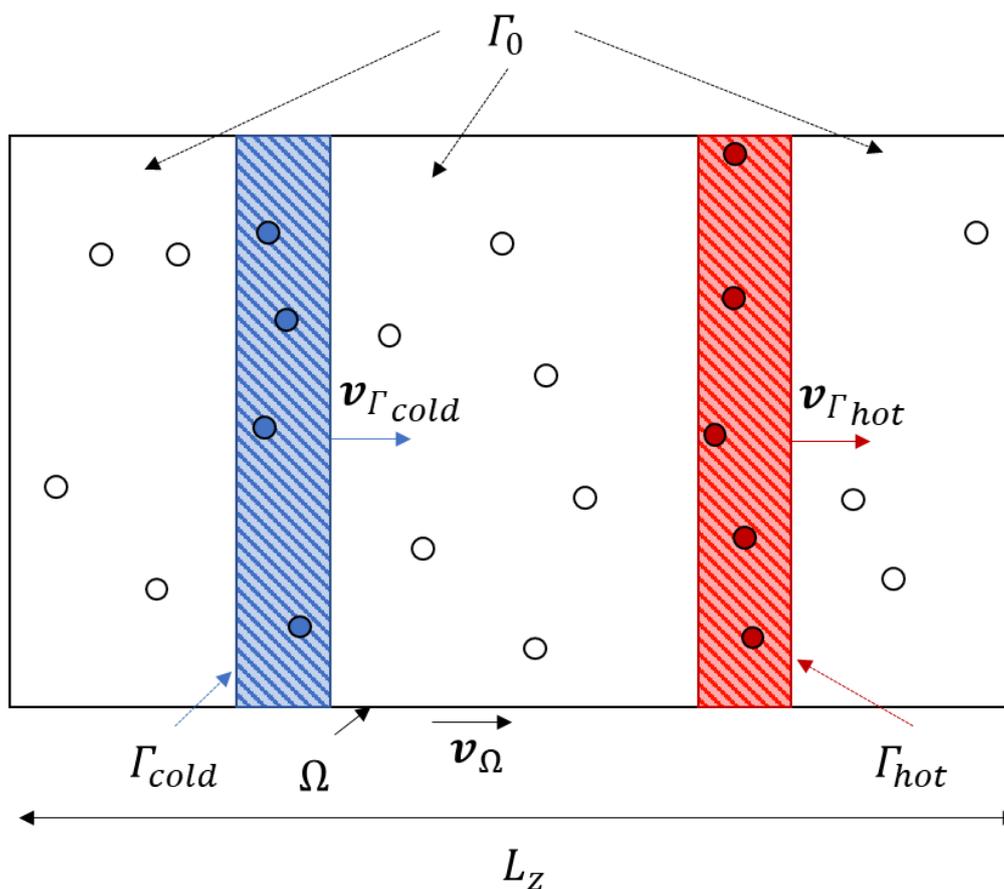


Figure 4.4: eHEX algorithm. Illustration of the simulation box, Ω , with non thermostatted regions, Γ_0 , a hot region, Γ_{hot} (red), and a cold region, Γ_{cold} (blue). The centre of mass velocities of Ω , Γ_{hot} and Γ_{cold} are \mathbf{v}_{Ω} , $\mathbf{v}_{\Gamma_{hot}}$ and $\mathbf{v}_{\Gamma_{cold}}$, respectively. Particles are represented by red/blue circles, if they are located in the hot/cold region and by empty circles otherwise. Readapted from eHEX article ⁷⁰.

conservation problems, which was only recently solved by an enhanced version of the algorithm by Wirnsberger *et al.*,⁷⁰ known as the enhanced Heat EXchange algorithm (eHEX), see Figure 4.4. Their improvement focuses on energy conservation compared with the previous

HEX algorithm. The authors identify the problem as a truncation error in the propagation operator splitting, which was solved by adding an extra integration step in the propagation of equations of motion.

Wirnsberger *et al.* compared both HEX and eHEX algorithms in order to study energy conservation. First, the authors described the basis of the kinetic exchange algorithm, where two spatial disjoint subdomains (Γ_k) are defined. Hot spatial region (Γ_{hot}) is where a certain amount of heat is injected (ΔQ_{Γ_k}), and cold spatial region (Γ_{cold}) is where the same amount of heat is subtracted ($-\Delta Q_{\Gamma_k}$). Moreover, the authors assume that there is no net energy flux applied to the simulation box (Ω):

$$\sum_k \Delta Q_{\Gamma_k} = 0. \quad (4.10)$$

The system will approach a steady-state in which heat fluxes are established between the two subdomains, leaving other regions (Γ_0) of simulation box unthermostatted. One crucial aspect of the heat-exchange algorithm is the special treatment of kinetic motion in the thermostatted regions, whereby the algorithm exchanges only non-translational kinetic energy. Energy is added or removed by rescaling velocities (v_i) of all the particles in the active regions (Γ_k), but leaving the centre of mass velocities (v_{Γ_k}) of these regions unchanged. The non-translational kinetic energy in one region can then be described as:

$$K_{\Gamma_k} = \sum_{i \in \gamma_k} \frac{m_i v_i^2}{2} - \frac{m_{\Gamma_k} v_{\Gamma_k}^2}{2}, \quad (4.11)$$

where m_{Γ_k} is the total mass contained in Γ_k and m_i is the mass of the particle. The time-dependent index set γ_k comprises all particles which are located in Γ_k . An important point is that particles outside the thermostatted region are not affected by this procedure. The velocity upgrade of each particle inside thermostatted regions can be formulated as:¹⁵³

$$v_i \rightarrow \bar{v}_i = \xi_k v_i + (1 - \xi_k) v_{\Gamma_k}, \quad (4.12)$$

where \bar{v}_i is the particle velocity after applying the algorithm, and ξ_k is the rescaling factor, which is defined for the HEX algorithm as:

$$\xi_k = \sqrt{1 + \frac{\Delta Q_{\Gamma_k}}{K_{\Gamma_k}}}. \quad (4.13)$$

Previous equations satisfy the following relations: (i) the updated non-translational kinetic energy (\bar{K}_{Γ_k}) is equal to the sum of the non-translational kinetic energy (K_{Γ_k}) and the amount of heat (ΔQ_{Γ_k}) added in the thermostatted region, (ii) the updated center of mass velocity (\bar{v}_{Γ_k}) is conserved:

$$\bar{K}_{\Gamma_k} = K_{\Gamma_k} + \Delta Q_{\Gamma_k}, \quad (4.14)$$

$$\bar{v}_{\Gamma_k} = v_{\Gamma_k}. \quad (4.15)$$

In this setup, the system is described by 6N-dimensional vector $x = (r, v)$ in the phase space. In order to study time evolution, Wirnsberger *et al.*⁷⁰ propose the following steps in their algorithm:

$$\bar{v}_i^n = \xi_{k(r_i)} v_i^n + (1 - \xi_{k(r_i)}^n) v_{\Gamma_{k(r_i)}}, \quad (4.16)$$

$$\bar{v}_i^{n+\frac{1}{2}} = \bar{v}_i^n + \frac{\Delta t}{2m_i} f_i^n, \quad (4.17)$$

$$r_i^{n+1} = r_i^n + \Delta t \bar{v}_i^{n+\frac{1}{2}}, \quad (4.18)$$

$$f_i^{n+1} = -\nabla U(r)|_{r=r^{n+1}}, \quad (4.19)$$

$$\bar{v}_i^{n+1} = \bar{v}_i^{n+\frac{1}{2}} + \frac{\Delta t}{2m_i} f_i^{n+1}, \quad (4.20)$$

$$v_i^{n+1} = \bar{\xi}_{k(r_i)}^{n+1} \bar{v}_i^{n+1} + (1 - \bar{\xi}_{k(r_i)}^{n+1}) \bar{v}_{\Gamma_{k(r_i)}}^{n+1}, \quad (4.21)$$

where the time evolution is defined as $t = n\Delta t$ with n as a superscript and Δt the timestep. $U(r)$ is the potential energy and f is the force acting on each particle i . Wirnsberger *et al.* applied their formulation of the HEX algorithm for a Lennard-Jones fluid and to the SPC/E

water model, and the authors observed an energy drift scaling as $O(\Delta t^2)$, which becomes restrictive for long simulations. Therefore, the authors improved the description of the HEX algorithm. The energy drift was identified to be due to higher order truncation terms, which are not taken into account in the adapted velocity Verlet time integration. Therefore, the authors reformulated the HEX algorithm as a Trotter factorisation of Liouville operator (iL), following Tuckerman *et al.*¹⁵⁴ who demonstrated that reversible integrators can be created by a Trotter factorisation. The exact evolution of the system as described by a 6N-dimensional vector $x(r, v)$ in space is:

$$x_{ex}(t) = e^{t i L} x(0), \quad (4.22)$$

where the Liouville operator is split in two parts $iL = iL_1 + iL_2$. In order to solve the integration step, the Strang splitting¹⁵⁵ is applied to approximate time evolution:

$$x(t) = \left[e^{\frac{\Delta t}{2} i L_1} e^{\frac{\Delta t}{2} i L_2} e^{\frac{\Delta t}{2} i L_1} \right]^P x(0), \quad (4.23)$$

where L_1 and L_2 describes the partial derivative of velocities and positions. The Strang splitting has the following local truncation error:¹⁵⁶

$$x(\Delta t) - x_{ex}(\Delta t) = \Delta t^3 \mathcal{E} x_{ex}(0) + O(\Delta t^4), \quad (4.24)$$

where \mathcal{E} is the leading-order error term from the Liouville operator splitting. Wirnsberger *et al.*⁷⁰ found that it is sufficient to consider a coordinate correction in the form of eq. (4.24). One of the reasons is that velocity Verlet integration is less accurate and its local truncation error is $O(\Delta t^3)$, which is larger than local truncation error from trotter factorisation $O(\Delta t^4)$. Consequently, the authors ignored the additional Verlet truncation error and all other correction terms which affect the velocities.

Finally, a coordinate correction in the time integration sequence from the previous definition of the HEX algorithm (equations (4.16) to (4.21)) was included, which is redefined as the enhanced heat exchange algorithm (eHEX):

$$r_i^{n+1} = \bar{r}_i^n - \Delta t^3 \mathcal{E} \bar{r}_i^{n+1}. \quad (4.25)$$

The eHEX demonstrates a significant reduction of the energy loss as compared to the HEX algorithm. The simulations of SPC/E water systems demonstrate an energy stability for a nanosecond simulation and timesteps up to 3.5 fs. For that reason, we rely on the implementation of this algorithm within the LAMMPS simulation package^{157,158} in order to perform simulations with a thermal gradient.

5. Thermophoresis simulations

5.1 Thermophoretic set-up

To generate a unidimensional temperature gradient along one chosen direction z , parallelepipedic boxes are considered, where the axes orthogonal to the thermophoretic gradient are kept equidistant $L_x = L_y$ and the L_z distance is typically larger than L_y and L_x . The box is centered in $(0,0,0)$, with a right equiangular distribution $(90^\circ, 90^\circ, 90^\circ)$, such as square cuboid.

The solvent particles are randomly introduced in the box with the Packmol package.¹⁵⁹ The solutes particles are inserted uniformly along the z -direction of the thermal gradient, resulting in a flat concentration profile as the initial configuration. For the z -direction, we simply divide the box width (L_z) by the number of particles (n_{solute}):

$$\delta_z = \frac{L_z}{n_{solute}}, \quad (5.1)$$

where δ_z is the amplitude of the z -direction slab. For the axes x and y , we define a small area for each solute between $5^2 \leftrightarrow 7^2 \text{ \AA}^2$, and we randomly insert the particles inside these areas. This strategy allows us to gain more control during the insertion process, which generates a better homogeneous concentration distribution in the xy plane, and avoids the overlap between two consecutive particles at high concentrations.

In Figure 5.1, we show the distribution of 10 solute particles in the plane xz and xy . The 10 solute molecules are equi-distributed along L_z distance. However, the 10 solute molecules are randomly placed along L_x and L_y lengths. In addition, the blue area represents the predefined region in which each solute is randomly inserted.

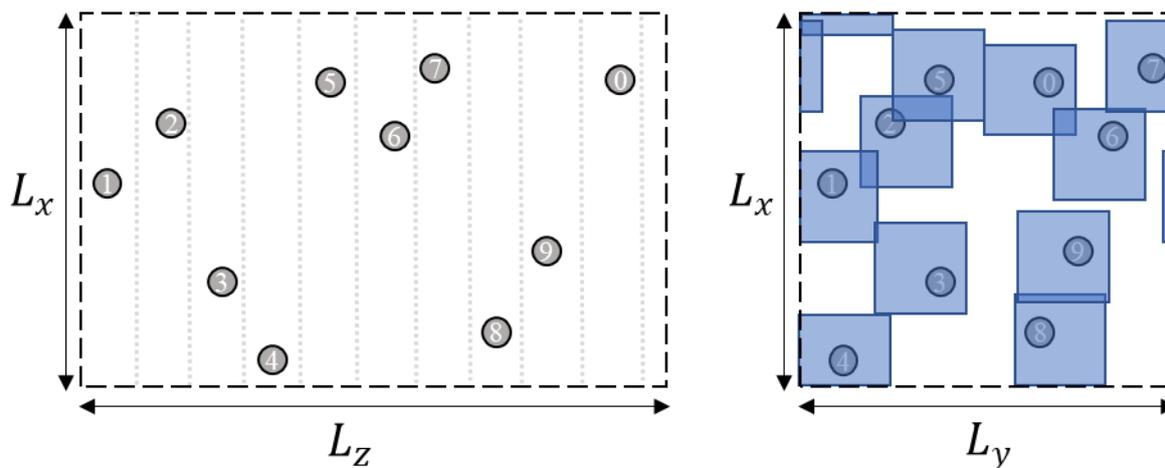


Figure 5.1: Schematic representation of flat concentration profile. Left picture represents the plane xz , where the particles are randomly distributed along L_z distance, but equidistributed along L_z distance. The right picture represents the plane xy , where the blue areas are defined for each solute and the solutes are randomly placed within this area.

In our simulations, the real space cutoff of Coulombic interaction is 8.5 \AA ,¹⁴⁴ although the Lennard-Jones potential is 9 \AA . The particle-particle particle-mesh (PPPM) solver¹⁶⁰ has been employed for the calculation of the long-range electrostatic forces, including the analytic differentiation approach which smooths the PPPM long range electrostatic forces. We follow a similar procedure as the eHEX thermal gradient activation work⁷⁰ for the equilibration process of our systems and the posterior temperature gradient generation, following these general steps:

1. Minimization process
2. Velocity rescaling: driving the system to the target temperature
3. NPT simulation
4. Box rescaling
5. NVT simulation: with the correct volume
6. 1st NVE simulation
7. Adjusting the energy of the last configuration
8. 2nd NVE simulation: with the correct temperature

As may be seen, this process requires the use of several ensembles in order to ensure suitable conditions for the posteriori set-up of thermal gradient. The first three steps are commonly used

for bringing the system into our working conditions. The system is first minimized using a Polak-Ribiere¹⁶¹ version of the conjugate gradient algorithm until relative changes in energy and forces fall below cut-off values of energy tolerance limit = 10^{-4} kcal · mol⁻¹ and force tolerance limit = 10^{-6} kcal · mol⁻¹ (step 1). Velocities are then set to correspond to the target median temperature (step 2), and a NPT equilibration of 400 ps is performed using the Nose-Hoover¹⁶²⁻¹⁶⁴ thermostat and barostat (step 3), with the damping times of $P_{damp} = 1000$ fs (barostat) and $T_{damp} = 100$ fs (thermostat). We want to emphasise that NPT barostat only applies for the plane xy and keeping $L_x = L_y$, since the z axis is going to be the thermal gradient axis, we keep invariant L_z distance to gain control of the thermal gradient length, as we do not want the thermal gradient direction to vary in length.

A critical point is the transition between the NPT and the NVT ensembles, steps 3 and 5 respectively, in which the volume must be fixed for the rest of the simulation. Therefore, the average volume is determined once the NPT reached pressure and temperature equilibrium (step 4). The box is modified to ensure the correct volume for the given conditions by rescaling the plane xy . The system is then equilibrated in the NVT ensemble for 200 ps at this equilibrium volume (step 5), using the Nose-Hoover thermostat with the same damping time of 100 fs as step 3.

After the NVT simulation, the total energy of its last step could differ from the median temperature of the system. Therefore, the last frame of this second equilibration is then used for a first, short propagation in the NVE ensemble (step 6) for 1 ps, and the total energy of the box is calculated during the 1st NVE simulation (E_{NVE}), and thus compare with the previous total energy average of the NVT simulation (E_{NVT}), the E_{NVT} is defined by the median temperature of the system. The energy difference between these two ensembles ($E_{diff} = E_{NVT} - E_{NVE}$) is reintroduced in the system by rescaling the velocities of the system particles in the last 1st NVE frame (step 7). In this case, we apply the eHEX algorithm to add or remove this energy difference in the system. Finally, a second NVE simulation of 200 ps prepares the system for the temperature gradient activation process at the previously fixed the total energy (step 8), which defines the target median temperature of the system. It should be noted that steps 4 and 7 are crucial rescaling processes in volume and energy respectively. These steps ensure the desired conditions for subsequent thermophoretic simulations under NVE ensemble.

Once the equilibration part is done, we define the cold and hot slab along z -axis. These two regions are symmetrically positioned at z_c and z_h respectively, with a thickness of $\delta_z = 4 \text{ \AA}$. One strategy would be to include these regions at the edges of the simulation box, i.e., $z_c \approx -L_z/2$ and $z_h \approx L_z/2$, leading a temperature gradient along L_z . However, this would generate a discontinuity in the transition between two periodic simulations boxes because of the periodic boundary conditions and would lead to an extreme temperature-gradient. Thus, we prefer to place these two regions symmetrically at: $z_c \approx -L_z/4$ and $z_h \approx L_z/4$. In this case, the temperature-gradient over the half of the box size $L_z/2$, and a second symmetrical gradient along the other half of the box. The second advantage is that we double the thermal-gradient for one simulation.

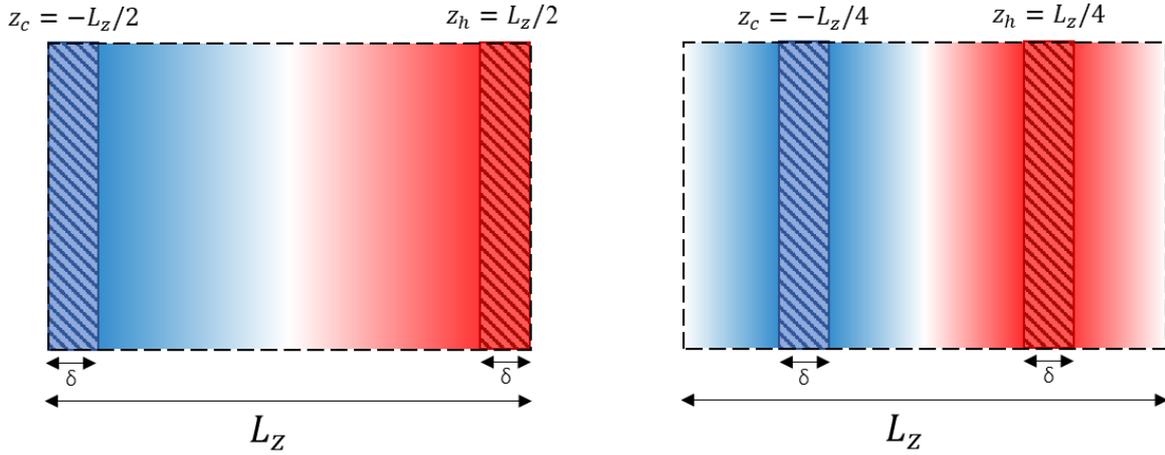


Figure 5.2: Thermostatted regions. The left figure shows a schematic representation of thermostatted regions at $\pm L_z/2$, where a temperature gradient of L_z distance is generated, but leading a temperature discontinuity at the edges. The right figure shows a schematic representation of thermostatted regions at $\pm L_z/4$, where two temperature gradients of $L_z/2$ distance are generated, without a temperature discontinuity at the edges.

Once both regions are well defined, the next step is the thermal-gradient generation by activating the eHEX algorithm.⁷⁰ As have mentioned before, not defining a set of thermostatted particles before running the simulation is one of the key points of this algorithm. The fact that we do not need to fix thermostatted molecules prevents having non-natural “obstacles” to the fluid flow in the thermostatted regions. Furthermore, water molecules are the only ones affected by the eHEX algorithm. We exclude solute molecules even if they enter in the reservoirs to have the most natural representation of the Soret effect. In addition, our systems are based on dilute aqueous solutions, which are largely dominated by water, so sufficient number of

molecules inside reservoirs will avoid any problems in the heat exchange process. This is one of the reasons why we choose a thickness of $\delta = 4 \text{ \AA}$ for each region, and not a thinner slab.

After the eHEX algorithm is activated, the system requires some simulation time until the steady state condition is reached. First, a temperature gradient starts to emerge (step 9), which takes approximately 100 ps to stabilize (step 9). This is followed by a concentration gradient as a consequence of the previous temperature gradient, which takes 2 ns to converge (step 10).

Step	Action	Ensemble	Length
0	Preparation	-	-
1	Energy minimization	-	-
2	Velocity scaling	-	-
3	Equilibration	NPT	400 ps
4	Volume rescaling	-	-
5	Equilibration	NVT	200 ps
6	Energy average	NVE	1 ps
7	Energy rescaling	-	-
8	Equilibration	NVE	200 ps
9	Temperature-gradient convergence	NVE	≈ 100 ps
10	Concentration-gradient convergence	NVE	≈ 2 ns
11	Production	NVE	= 10 ns

Table 5.1: Summary of the generation of a steady-state thermophoretic setting.

After this point, the system reaches the steady-state condition, where both solute and solvent concentrations gradients will remain invariant for the rest of the simulation (step 11). Therefore, the concentration profiles as a function of the temperature gradient are independent on time, and the Soret coefficient value can be determined. In the next section, we will discuss the convergence of these two last steps in more detail.

During the analysis, if we represent logarithmic concentration as a function of the temperature gradient, we can determine the Soret coefficient as its linear regression, see eq. (3.9). Moreover,

we overlap both gradients in order to improve the statistics, since the concentration gradients are double represented for one simulation box.

The Soret coefficient errors are calculated by the bootstrap method,¹⁶⁵ which is a statistical technique for estimating quantities about a population by averaging estimates from multiple small data samples. This method consists in taking $N_{samples}$ (in our case, 50), where each sample comprises of a certain number of independent trajectories $N_{trajectories}$ (in our case, 10) randomly selected among our simulated trajectories that are 20 independent trajectories for most simulations. The error is thus determined by the standard deviation of these $N_{samples}$. In the next section, we will discuss the energy conservation over the simulation time, and the time required for the establishment of the temperature and concentration gradients.

5.2 Convergence

We first assess the ability of the thermophoretic simulation set-up to generate a temperature gradient for the timescale of the gradient onset. Our aim of this section is to investigate the conservation of energy throughout the simulation, whether the temperature and concentration gradients are adequately generated for any system that we are interested in studying.

5.2.1 Energy conservation

The original implementation of heat exchange algorithm^{166,167} faces an important problem in the energy conservation due to a truncation error in the operator splitting, which leads an energy drift that scales $O(\Delta t^2)$. Wirnsberger *et al*⁷⁰ eHEX algorithm largely improved the truncation error by adding an coordinate correction in the velocity Verlet propagation algorithm, which reduces the error to $O(\Delta t^4)$. However, their work focuses on relatively short timescales for simple models (SPC/E and Lennard-Jones fluids), not longer than 1 ns, which becomes unpractical for our simulations that requires a longer time for equilibration and production runs. Consequently, we first assess the energy conservation for this algorithm over longer timescales (10 ns), and its robustness at different time integration steps.

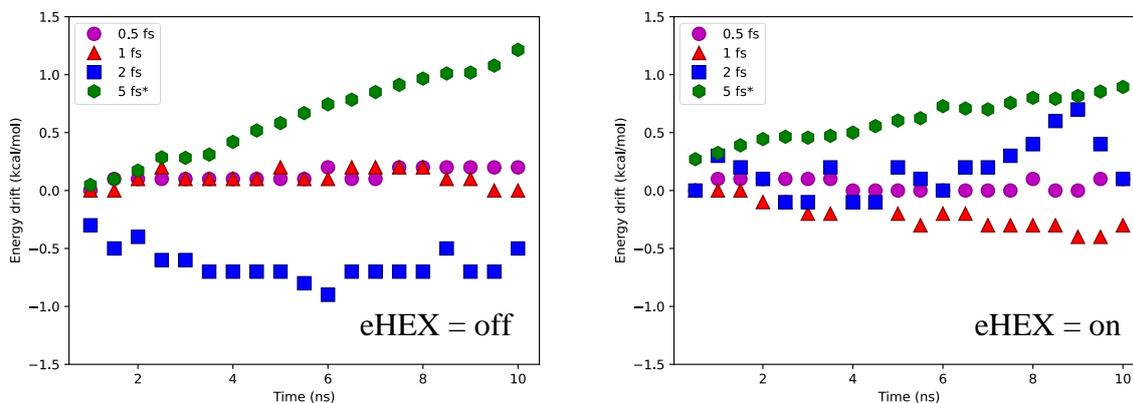


Figure 5.3: Energy conservation study. One single trajectory is represented. Data are shown for four different timesteps; $\Delta t_1 = 0.5$ fs, $\Delta t_2 = 1$ fs, $\Delta t_3 = 2$ fs and $\Delta t_4 = 5$ fs, in a pure aqueous solution. The energy drift of $\Delta t_4 = 5$ fs is divided by 50. $P = 1$ bar. $T_0 = 300$ K. $L_z = 50$ Å. The left graph represents the energy drift over 10 ns simulation without eHEX algorithm. The right graph represents the energy drift over 10 ns simulation with eHEX algorithm, and the energy flux was set to $\Delta Q_{r_k} = 0.0375$ kcal \cdot mol $^{-1}$ fs $^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60$ K.

In Figure 5.3, we compare the energy drift of NVE simulations for pure water systems at 300 K composed of 1064 water molecules in a box with the length of $L_z = 50$ Å. The left graph represents the NVE simulation without the implementation of eHEX algorithm, and the right graph represents the same NVE simulation once the eHEX is active. We do not observe any relevant difference on the energy drift, prior and post the activation of the kinetic exchange algorithm. The results show that energies are conserved for timesteps 0.5 fs (magenta), 1 fs (red) and 2 fs (blue) along 10 ns simulation, which follows the same energy conservation line as shown by Wirnsberger *et al.*⁷⁰ However, the energy drift is much more pronounced at 5 fs (green), therefore, we divide the energy drift value by 50 in Figure 5.3.

The 0.5-fs simulations show a high energy conservation for both simulations, which indicates the good energy conservation of the NVE simulations with and without the eHEX algorithm for a timescale of 10 ns. The 1-fs simulation has the same energy conservation than 0.5 fs simulation for the NVE simulation without the eHEX algorithm, but it has a very small constant energy drift once the eHEX is activated, -0.05 kcal \cdot mol $^{-1}$ \cdot ns $^{-1}$, causing the total energy drift for the NVE simulation with the active eHEX is lower than -0.5 kcal \cdot mol $^{-1}$. We also observe some energy conservation deviation for the 2-fs simulations. The left graph shows an

energy drift smaller than $1 \text{ kcal} \cdot \text{mol}^{-1}$ on the first nanoseconds of the simulation, but after some steps the energy becomes constant. The right graph does not show any energy drift, but it has a small energy peak on the last nanoseconds (8-10 ns). Finally, the energy completely drifts for the 5-fs simulation. Consequently, the lower integration timestep, the better the energy conservation, because we observe that the integration step of 0.5 fs keeps better the total energy.

In this work, a 1-fs integration step is the best choice for our thermophoretic simulations and appears a good compromise, 0.5-fs demonstrated a better energy conservation but a 1-fs timestep significantly reduces the computational cost. We would also like to point out that thermophoretic simulations have been attempted using 2-fs, giving similar concentration gradients as in 1-fs simulations, but we decided to stay at 1-fs to be able to study flexible molecules due to energy drift of 2-fs simulation in NVE.

5.2.2 Convergence of the temperature gradient

A crucial part for thermophoretic simulations is the establishment of a stable and constant thermal gradient during the whole simulation time, because the system will only be able to reach the steady-state condition if we first ensure a constant thermal gradient independent of time. The energy conservation study from the previous section confirmed that the eHEX algorithm allow us to study a non-equilibrium thermally-driven effect from the NVE ensemble, since the total energy of the system is conserved in a nanoscale time.

The median temperature (T_0) is established during the NPT and posterior NVT equilibration steps by using the Nose-Hoover thermostat.¹⁶²⁻¹⁶⁴ On the other hand, the temperature gradient (ΔT) is directly proportional to the amount of non-translational kinetic energy transferred from the cold region to the hot region of the simulation box every timestep. Therefore, the quantity of heat transfer process (ΔQ_{T_k}) will determine the amplitude of the temperature gradient.

The amount of heat required to generate a specific thermal gradient depends on the solvent molecules (water) affected by the eHEX algorithm, the length and the width of the simulation box, and the thermalised regions. In this work, the main system is composed of 1024 water molecules, and 20 to 80 solutes, because we want to remain in a dilute regime. Even if we expect that the partial volume of the solvent is not symmetric between both regions, we do not

take into account this difference for determining the amount of kinetic energy necessary. Moreover, we have analysed the thermal gradient amplitude by adding 80 molecules of TMAO in 1024 water molecules, and we do not observe a substantial difference on the thermal gradient, after the equilibration and the posterior generation of the thermal gradient, compared with the pure water box (1024 water).

Our main parallelepipedic simulation box has a total fixed length of $L_z = 50 \text{ \AA}$, but variable widths around $L_x = L_y \approx L_z/2$, depending on the number of solutes, the temperature and the pressure, because the barostat is applied in the x and y directions. In the stationary state, the heat flux between the reservoirs is given by:⁷⁰

$$J_{Q(z)} = \frac{\Delta F_{\Gamma_k}}{2\Delta t L_x L_y}, \quad (5.2)$$

and the volume of the reservoirs is:

$$V_{\Gamma_k} = \delta \cdot L_x \cdot L_y, \quad (5.3)$$

where $\delta = 4 \text{ \AA}$ is the thickness of the reservoir. The imposed amount of heat per timestep for this system is $\Delta Q_{\Gamma_k} = \Delta F_{\Gamma_k} / \Delta t = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$. However, longer systems (i.e., $L_z = 75, 100 \text{ \AA}$) with the same width requires lower heat transfer, as the distance between the hot and cold reservoir increases. Longer distances require less energy to keep the same temperature gradient, because the thermal conduction of the solution needs to propagate along longer distances to equilibrate the two reservoirs due to the same heat flow rate (φ_q):

$$\varphi_q = \frac{\Delta F_{\Gamma_k}}{2\Delta t} = -k_B L_x L_y \frac{\Delta T(z)}{\Delta z}. \quad (5.4)$$

In order to capture the spatial variation of the temperature, we divide the z -axis into N_b bins:

$$T_j = \frac{2K_j}{(N_j f - 3)k_B}, \quad (5.5)$$

where j is one bin, T_j is the instantaneous kinetic energy, N_j is the number of atoms contained in the bin, f is the number of degrees of freedom per atom, and K_j is the total non-translational kinetic energy. The total number of bins is calculated by:

$$N_b = \frac{\delta_b}{L_z}, \quad (5.6)$$

where $\delta_b = 0.6 \text{ \AA}$. We thus verify the convergence of the temperature gradient in our simulations, which will induce a posterior concentration gradient in response of the generated thermally-driven forces, and we also check the ability to generate temperature gradient for the timescale of the gradient onset. Several thermophoretic systems have been taken into account, represented in Figure 5.4.

We have analysed the evolution of the temperature gradient for a simulation time of 100 ps, except 200 ps for Figure 5.4.D. Temperature evolution is analysed by blocks from 1 to 5 ps, depending on the thermal gradient timestep. Once the simulations reached the thermal steady-state, the thermal gradients are represented by a black line. The resulting amplitude of the thermal gradient is thus $\Delta T \approx 60 \text{ K}$ for all the thermophoretic systems.

Figure 5.4.A and Figure 5.4.B represent a pure water box of 1024 molecules, and we show the thermal gradient growth as a function of time for temperatures of 300 K and 330 K respectively. We observe that these thermal gradients progressively increase for both simulations, and they converge in less than 10 ps (purple line (10 ps) reaches the shape of the black line (steady-state)). This first analysis shows that the median temperature (300 K or 330 K) does not have an impact on the eHEX algorithm during the thermal gradient generation, which means that the amplitude of the thermal gradient does not depend on the median temperature of the system.

Figure 5.4.B and Figure 5.4.C represent a pure water box of 1024 and 2048 molecules respectively at 330 K. In this case, we double the system size from Figure 5.4.B to Figure 5.4.C, and we observe instead a slow gradient convergency, which takes almost 5 to 10 times longer to reach the steady-state expected for our target thermal gradient ($\Delta T \approx 60 \text{ K}$). Two effects can explain this slower convergence which is also justified by the eq. (5.4). First, doubled-size system implies that heat flow (φ_q) needs to double the distance along the z-axis ($\Delta t \propto \frac{\Delta F_{T_k}}{\Delta T} \Delta z$),

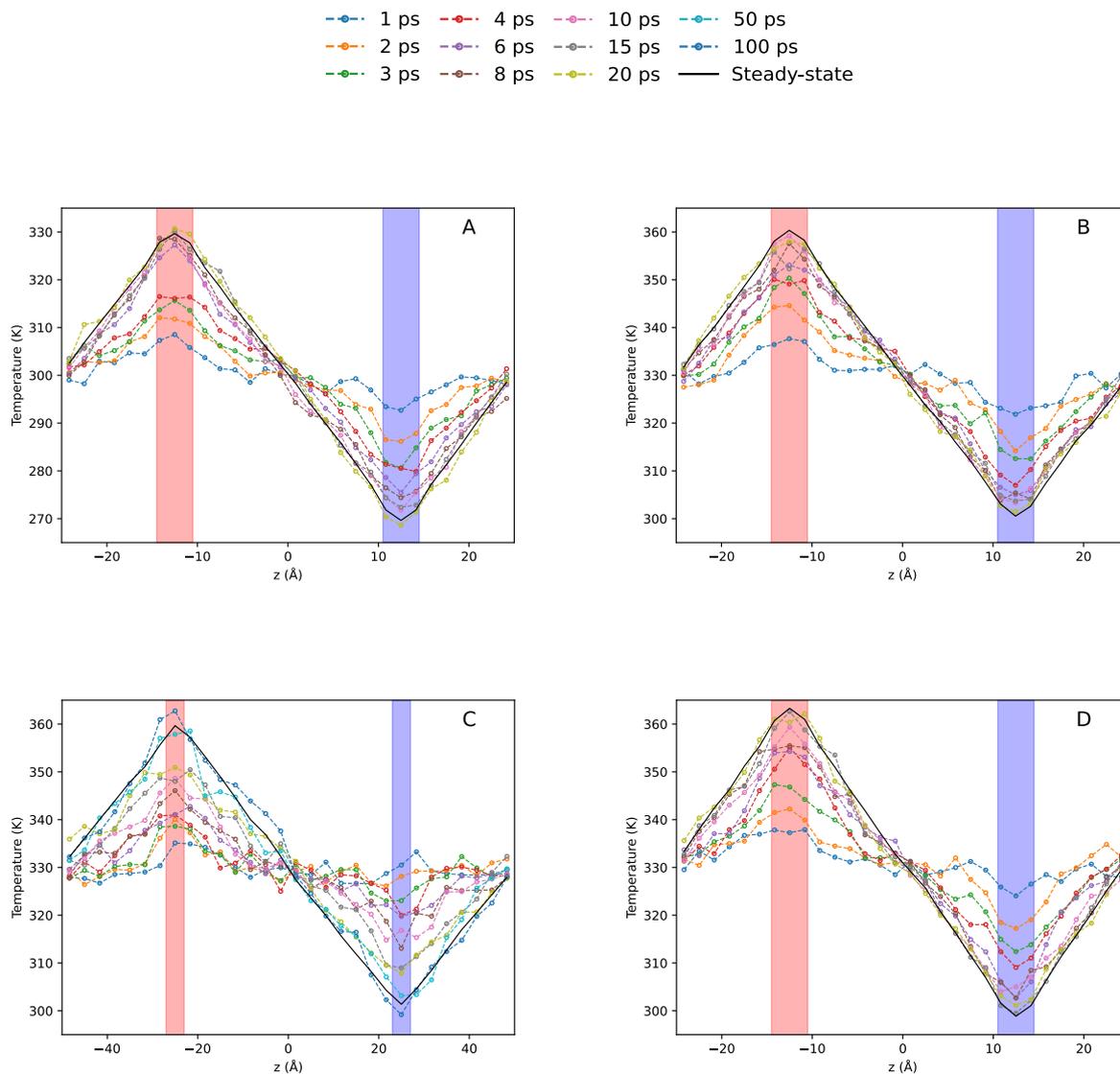


Figure 5.4: Convergence of the water temperature-gradient. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k}^{50\text{\AA}} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$ ($L_z = 50 \text{ \AA}$) and $\Delta Q_{\Gamma_k}^{100\text{\AA}} = 0.01875 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$ ($L_z = 100 \text{ \AA}$) in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. (A) Pure water (1024 molecules), $L_z = 50 \text{ \AA}$, $T_0 = 300 \text{ K}$. (B) Pure water (1024 molecules), $L_z = 50 \text{ \AA}$, $T_0 = 330 \text{ K}$. (C) Pure water (2048 molecules), $L_z = 100 \text{ \AA}$, $T_0 = 330 \text{ K}$. (D) Aqueous solution of TMAO (2.17 m) (40 TMAO/1024 water), $L_z = 50 \text{ \AA}$, $T_0 = 330 \text{ K}$.

since the total length of the box (L_z) increases from 50 \AA to 100 \AA . Secondly, the heat transfer is reduced for longer systems ($\Delta Q_{\Gamma_k}^{100\text{\AA}} = \Delta Q_{\Gamma_k}^{50\text{\AA}}/2$), since they require less energy to generate the same amplitude of the temperature gradient ($\Delta T \approx 60 \text{ K}$), but reducing the amount of heat transfer slows the thermal gradient generation process.

Figure 5.4.B and Figure 5.4.D represent the pure water box of 1024 molecules and the aqueous solution of TMAO 2.17 m (40 TMAO /1024 water) respectively. In this case, we check if the addition of solute molecules has any impact on the heat transfer timescale. Comparing both systems do not reveal any relevant difference on the timescale thermal gradient generation. The first reason is that the solute is not directly affected by the eHEX algorithm, which means that the kinetic energy of each solute is given by the thermal gradient generated by the solvent. The second reason is that the TMAO aqueous solution is in a dilute regime, where the heat transfer from solvent to solute is relatively small enough to alter the amplitude of the thermal gradient. The third reason is that the amount of TMAO molecules inside the reservoirs is not high enough to displace the minimal quantity of water molecules required for a correct operation of the eHEX algorithm.

We show that the equilibration timescale of 100 ps for 5 nm box length systems and 200 ps for 10 nm box length systems are enough to reach the thermal gradient ($\Delta T \approx 60$ K) in steady state condition. Moreover, we also confirm that solutes added in our systems do not affect the thermal gradient.

5.2.3 Convergence of the concentration gradient

We already set-up the timescale for a proper and constant temperature gradient generated along the simulation box, and we reach a thermal steady state from the perspective of the temperature gradient. We can now focus on the concentration gradient profile for our dilute aqueous solutions. As we can imagine, the concentration gradients require much longer timescales to equilibrate as compared to the temperature gradients.¹⁵¹ The thermal gradients work in terms of heat transfer, but the concentration gradients generated in the thermophoretic simulations are instead thermally-driven diffusion processes in response to these thermal gradients, which takes longer to reach the diffusive steady state. Furthermore, our work focuses on dilute solutions, since there are only a few solute molecules inside the simulation box, these solutes must diffuse all over the box to fully sample the thermophoretic effect. For a typical molecular size solute, the diffusion coefficient under ambient conditions is on the order of 0.01-0.1 Å²/ps, so one solute molecule will sample 1 nm over $\approx 0.5 - 5$ ns. For this reason, we average 40 independent trajectories of the same simulation for a total duration of 10 ns to get better concentration profiles. We also divide the z-axis into the same number of bins (N_b) as a

function of the box size, see eq. (5.6), in order to capture the spatial variation of the concentration, same bins as the temperature gradient. We now investigate the evolution of the concentration gradient for the model TMAO solute in a dilute aqueous solution at different median temperatures, concentrations, and box lengths.

Schematic representation of the simulation boxes

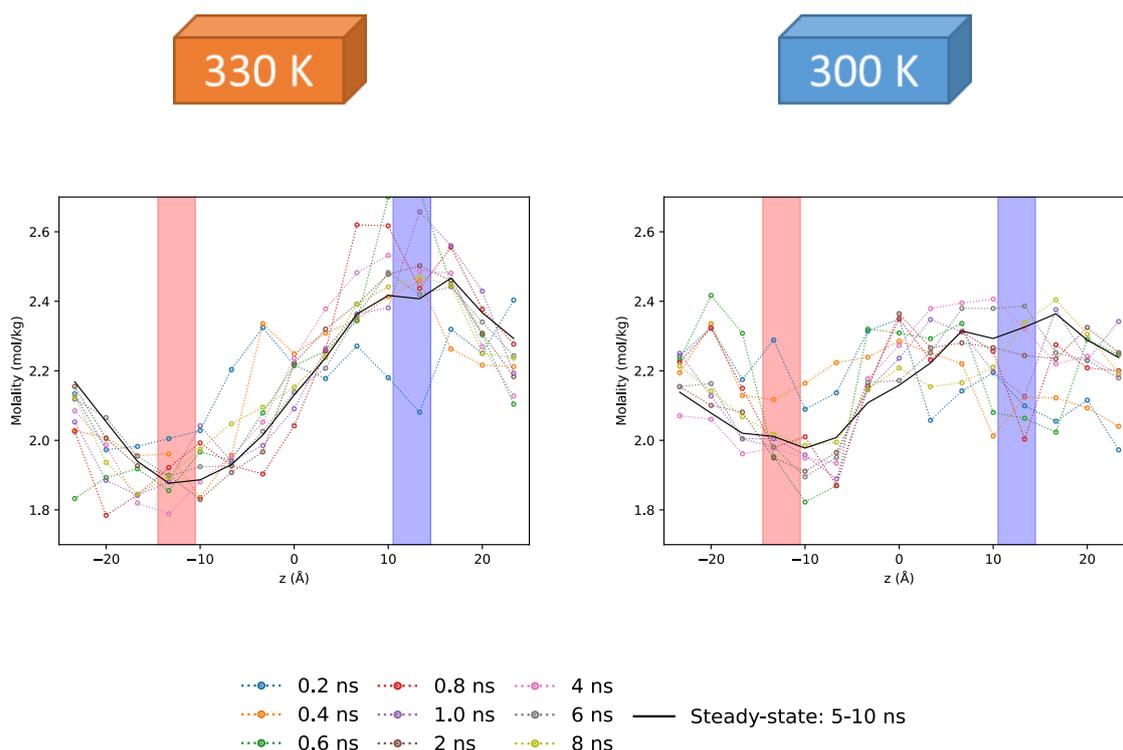


Figure 5.5: Convergence of the TMAO concentration-gradient (temperature). Data was averaged over 40 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. The blue and red zones represent the cold and hot regions, where the heat is pumped and injected. The left graph is a 2.17-m TMAO average concentration at $T_0 = 330 \text{ K}$. The right graph is a 2.17-m TMAO average concentration at $T_0 = 300 \text{ K}$. The simulation boxes above the graph shows the difference between both graphs.

The concentration profiles of the 2.17-m TMAO solutions are shown in Figure 5.5 at the median temperatures of 300 K and 330 K respectively. The concentration profile is seen to converge on a 2 ns (magenta line) timescale for the reference median temperature of 300 K, and 1 ns (green line) timescale for the median temperature of 330 K. These convergence time difference can be associated the corresponding TMAO diffusion coefficient (D^S) at the

corresponding temperatures of 300 K and 330 K, which are equal to 0.049 and 0.101 Å²/ps, respectively. While a proper determination of a scaling law would require more extensive tests involving other solutes and concentrations, a rule of thumbs that we can derive here is that the equilibration time (τ_{eq}) for N solute molecules along a gradient spanning $L_z/2$ is on the order of:

$$\tau_{eq} \approx \frac{L_z^2}{D^s \cdot N}. \quad (5.7)$$

While we do not expect this relationship to remain valid varying extensively the solute or the simulation conditions, this does provide a rough estimate of the required timescale before reaching a steady-state in these systems. If we assume that solute molecules do not directly interact with each other in sufficiently dilute solutions, this timescale also imposes a lower limit to the average concentrations that can be studied in practice. At the maximum dilution of 1 solute molecule in the simulation box, simulations involving similar system and gradient sizes would likely require 50 ns to converge and then provide very poor statistics anyway. As a consequence, studying thermophoresis for dilute solutions within this simulation framework requires a reasonable trade-off between the number of solute molecules in the system and convergence of the concentration-gradient on accessible timescales. Moreover, we also notice that the amplitude of the concentration profile of 330 K is larger than 300 K, which might indicate a higher thermophobicity for the TMAO molecule as the median temperature of the system increases. This aspect will be quantified by the Soret coefficient calculations in the next chapter.

Schematic representation of the simulation boxes

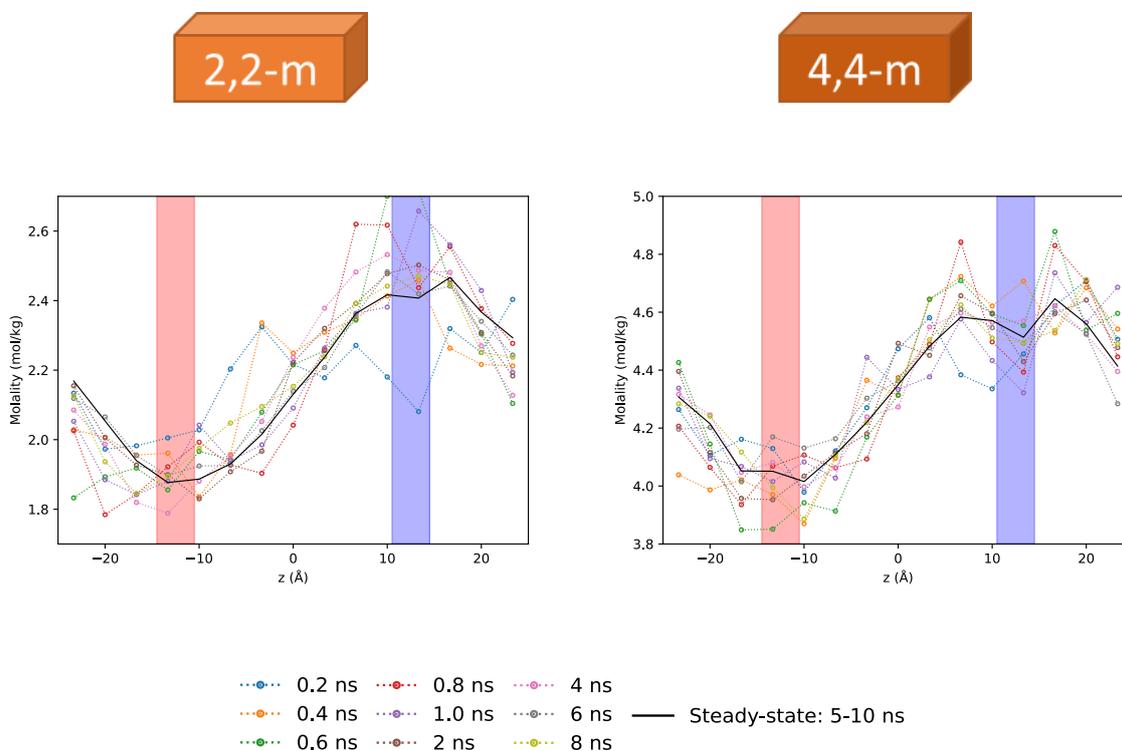


Figure 5.6: Convergence of the TMAO concentration-gradient (concentration). Data was averaged over 40 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. The blue and red zones represent the cold and hot regions, where the heat is pumped and injected. The left graph is a 2.17-m TMAO average concentration. The right graph is a 4.4-m TMAO average concentration. The simulation boxes above the graph shows the difference between both graphs.

The concentration profiles of 2.17-m and 4,4-m TMAO solutions are shown in Figure 5.6 respectively at a median temperature of 330 K. The concentration profile is seen to converge on a 1 ns (green line) timescale for the median temperature of 330 K for the higher concentration system, which is the same as the 2.17-m TMAO solution timescale previously discussed. These results manifest that solute concentration does not alter the convergence time. Moreover, once the steady state is reached for both simulations, the amplitude of the concentration gradients are very similar for both simulations, which indicates that the Soret effect might not depend on the concentration. This aspect will be also quantified by the Soret coefficient calculations in the next chapter.

Schematic representation of the simulation boxes

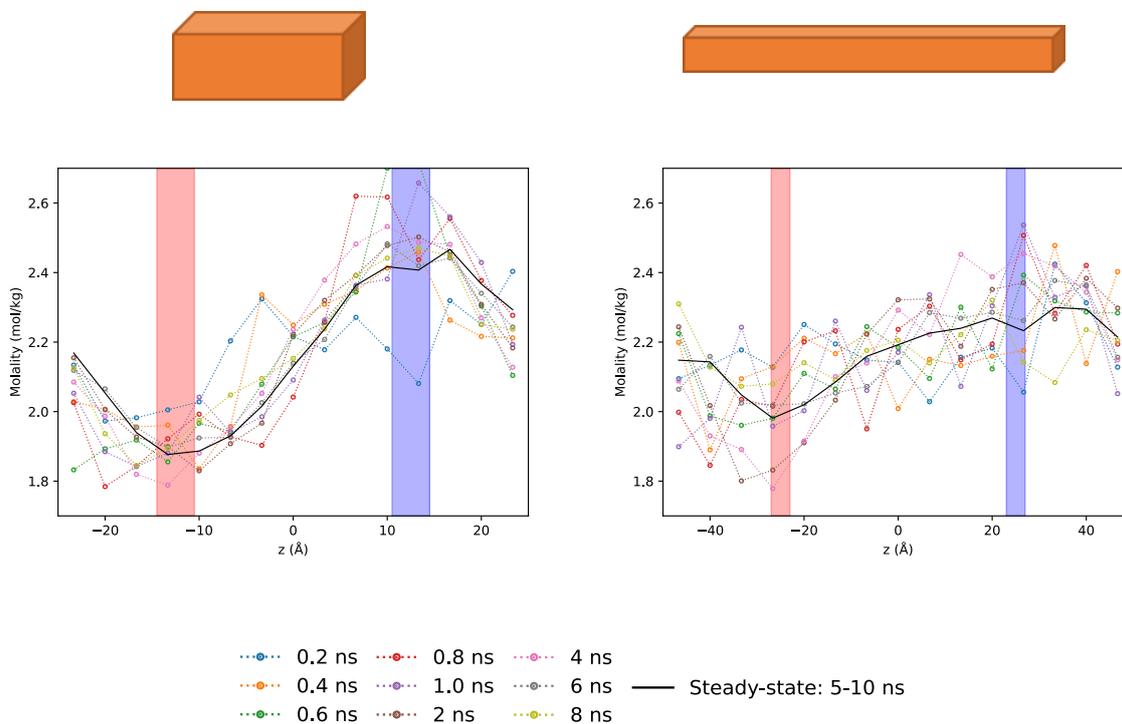


Figure 5.7: Convergence of the TMAO concentration-gradient (length). Data was averaged over 40 independent trajectories. $P = 1$ bar. $T_0 = 330$ K. The blue and red zones represent the cold and hot regions, where the kinetic exchange is active. The left graph is a 2.17-m TMAO average concentration, $L_z = 50$ Å and $\Delta Q_{\Gamma_k}^{50} = 0.0375$ kcal · mol⁻¹fs⁻¹. The right graph is a 2.17-m TMAO average concentration, $L_z = 100$ Å and $\Delta Q_{\Gamma_k}^{100^*} = 0.009375$ kcal · mol⁻¹fs⁻¹. The simulation boxes above the graph shows the difference between both graphs.

The concentration profiles of the 2.17-m TMAO solutions are shown in Figure 5.7 at the box lengths of 50 Å to 100 Å respectively along the thermal gradient direction z . The volume is the same ($V_E = V_B$), which means that the right graph simulation box has $L_x = L_y \approx 18$ Å lengths perpendicular to the thermal gradient. The box area ($L_x \cdot L_y$) is reduced by 2 while the distance between the two reservoirs is multiplied by 2. Therefore, the imposed heat transfer (ΔQ_{Γ_k}) is set to 0.009375 kcal · mol⁻¹fs⁻¹, which is four times lower than heat transfer of the left graph to keep constant the heat flow rate (φ_q), see eq. (5.4). One problem that arises on extending the box length in the z -direction is that the same amount of TMAO molecules have to sample a distance which is twice as long in the gradient direction. As we have mentioned before, concentration gradients profiles are generated by sampling the solutes molecules moving along the thermal gradient. The concentration gradient in the left graph is represented by 40 TMAO

molecules (N_{TMAO}) in a box of length 50 \AA , which means that its uni-dimensional z -density is equal to $\rho_z = N_{TMAO}/L_z = 0.8 \text{ \AA}^{-1}$. On the other hand, the concentration gradient in the right graph is represented by 40 TMAO molecules in a box of length 100 \AA , so its uni-dimensional z -density is equal to $\rho_z = 0.4 \text{ \AA}^{-1}$. The reduction of the heat transfer between the reservoirs together with the lower z -density (ρ_z) increases and hinders the concentration gradient convergence on the time. The fact that the concentration gradient in the left graph is different as compared to the right graph might be because this system does not converge during a 10 ns simulation. Therefore, longer simulations should be done to correctly describe the concentration gradient for the system in the right graph. However, we have not performed longer simulations for this system as this is not our main objective, and would require higher computational costs. We investigate instead the evolution of the concentration gradient for a larger system size in Figure 5.8.

The concentration profiles of the 2.17-m TMAO solutions are shown in Figure 5.8 at the box lengths of 50 \AA to 100 \AA respectively along the thermal gradient direction z . The volume is the double ($V_F = 2 \cdot V_B$), which means that the axes perpendicular to the thermal gradient remained unaltered, but the length of the z -axis was doubled ($L_z = 100 \text{ \AA}$) in the right graph. The box area ($L_x \cdot L_y$) is conserved while the distance between the two reservoirs is multiplied by 2. Consequently, it is observed that the concentration profile converges on $\sim 2 \text{ ns}$, which is 2 times longer than that of the left graph. This result is in agreement with the quadratic part of a diffusive process ($\tau_{eq} \approx L_z^2/(D^s \cdot N)$), which will require four times longer to sample a doubled distance. However, once the steady state is reached for the concentration profile in the left graph, the resulting concentration gradient is very similar to the concentration gradient of the right graph, which indicates that the Soret effect does not depend on the system size.

Schematic representation of the simulation boxes

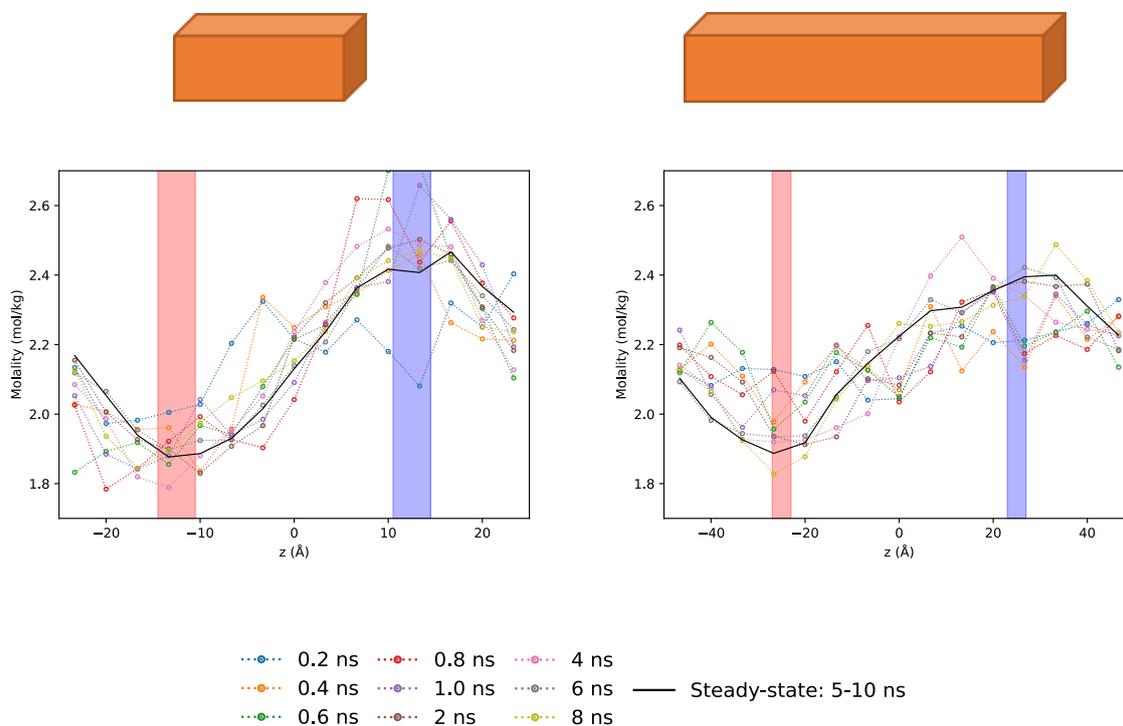


Figure 5.8: Convergence of the TMAO concentration-gradient (size). Data was averaged over 40 independent trajectories. $P = 1$ bar. $T_0 = 330$ K. The blue and red zones represent the cold and hot regions, where the kinetic exchange is active. The left graph is a 2.17-m TMAO average concentration, $L_z = 50$ Å and $\Delta Q_{T_k}^{50} = 0.0375$ kcal · mol⁻¹fs⁻¹. The right graph is a 2.17-m TMAO average concentration, $L_z = 100$ Å and $\Delta Q_{T_k}^{100} = 0.001875$ kcal · mol⁻¹fs⁻¹. The simulation boxes above the graph shows the difference between both graphs.

As we have mentioned before, even though the ideal dilute solution should be just a single solute molecule in the simulation box, these simulations would require an enormous amount of time to converge, because the solute molecule must sample the entire concentration gradient profile and thus travel through the simulation box several times. Considering that solute molecules do not interact directly with each other at sufficiently dilute solutions, a larger number of molecules in the same system would improve the sampling process and thus reduce the time scale of the simulation. Moreover, we will obtain better estimates of average concentration gradients, and reduce the statistical errors. We find that 2.17-m TMAO solution is a good compromise, because it allows to accelerate convergence while being in the dilute regimes where solute-solute interactions are negligible. For such systems, we estimate that a 2-ns timescale can safely be chosen as the time required to reach the steady state for the solute

concentration profile as long as the conditions for the simulation and the system sizes do not significantly differ from the ones studied here.

5.3 Robustness of the results

In this section, we will analyse different aspects of thermophoretic systems once they reach steady state equilibrium. First, we will focus on the effect of the box size and of the gradient amplitude. Second, we will investigate the effect of the rate at which heat is exchanged between the reservoirs as well as testing different force fields available in the literature.

5.3.1 Dependence on the box length (L_z)

A first important check is to verify that the Soret coefficient does neither critically depend on the box size nor the amplitude of the thermal gradient. We have repeated our simulations on the model TMAO system. The thermal gradient z -axis (L_z) was increased from 50 Å to 100 Å, the other two distances ($L_y = L_x$) perpendicular to the thermal gradient remained invariant ($L_x^{50 \text{ Å}} \approx L_x^{100 \text{ Å}}$) and the number of molecules was rescaled in order to keep the z -density constant ($\rho_z = 0.8 \text{ Å}^{-1}$). We have seen before that the main consequence of increasing the box size is the equilibrium time scale, with a doubled equilibration time for a doubled box size (see above). However, the results in Figure 5.8 suggest that the achieved concentration profiles are independent of the box length. In this section, we have analysed the concentration gradient as a function of thermal-gradient length (L_z) in more detail. We study three simulations box lengths (L_z) of 50 Å, 75 Å and 100 Å, and their Soret coefficient value is determined for these systems. The previous equation (3.10) described the Soret coefficient ($S_T = D_T^S/D$) as an exponential depletion law, and we can thus rewrite the expression as:

$$\ln\left(\frac{c^S}{c_0^S}\right) = -S_T(T - T_0), \quad (5.8)$$

where the Soret coefficient is then defined as the slope of the normalised logarithmic concentration of the solute as a function of the temperature difference from the median temperature. Moreover, our simulations are composed by two zones under thermal gradients,

they are combined in order to double the statistics of our logarithmic concentration gradients as a function of the temperature.

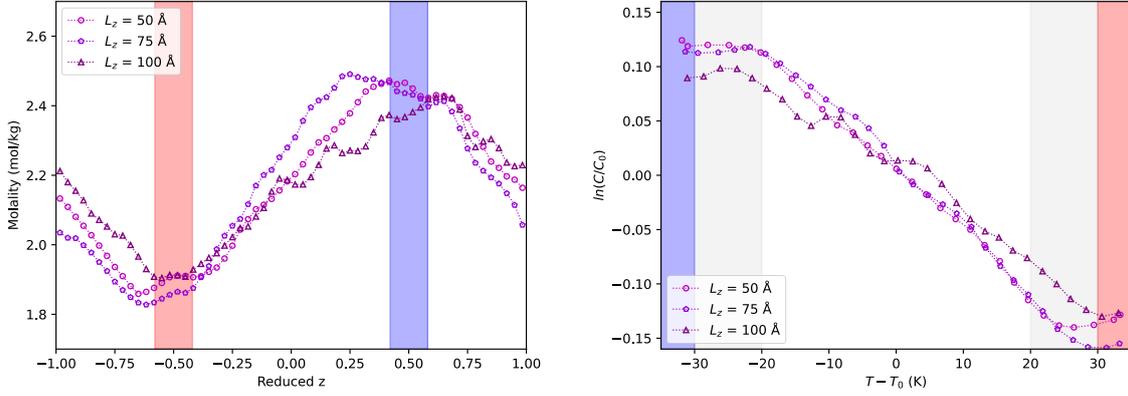


Figure 5.9: Box size dependence. Data was averaged over 20 independent trajectories (40 for $L_z = 100 \text{ \AA}$). The energy flux was set to $\Delta Q_{\Gamma_k}^{50\text{\AA}} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$, $\Delta Q_{\Gamma_k}^{75\text{\AA}} = 0.025 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$ and $\Delta Q_{\Gamma_k}^{100\text{\AA}} = 0.01875 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $T_0 = 330 \text{ K}$. The blue and red bars indicate the regions where heat is pumped and injected, respectively; their width is not at scale. Data is shown for three different box lengths in a 2.17-m TMAO solution of $L_z = 50, 75, 100 \text{ \AA}$. The left graph represents the molality vs reduced box length. The right graph represents the same data on a log-scale for the molality and as a function of temperature.

The concentration profiles of the 2.17-m TMAO solutions are shown in the left graph of Figure 5.9 for the box lengths of 50 \AA , 75 \AA and 100 \AA . The box length L_z is represented in reduced units to observe the overlap among the three concentration gradients. The reduced units represent each box length ($50, 75$ and 100 \AA) in an interval from -1 to 1 . While the concentration profiles are not exactly similar, the amplitude of the concentration difference between the two extremes in the hot and cold regions are the same within error bars. The right graph represents the $\ln(C/C_0)$ as a function of the temperature difference ($T - T_0$), which might indicate lower slope for largest simulation length ($L_z = 100 \text{ \AA}$). Therefore, the corresponding Soret coefficients are: $S_T^{50\text{\AA}} = 5 \pm 0.9 \cdot 10^{-3} \text{ K}^{-1}$, $S_T^{75\text{\AA}} = 5.6 \pm 1.3 \cdot 10^{-3} \text{ K}^{-1}$, $S_T^{100\text{\AA}} = 3.6 \pm 1.5 \cdot 10^{-3} \text{ K}^{-1}$. Even if the Soret coefficient from the largest simulation is noticeably lower than the other two simulations, its associated error is large enough to be in the range of the Soret coefficient value from the shorter length simulation ($L_z = 50 \text{ \AA}$). However, the simulations of

$L_z = 50 \text{ \AA}$ and $L_z = 75 \text{ \AA}$ exhibit very similar Soret coefficient values. We also observe that the Soret coefficient associated error increases with the system size. We detailed above those longer simulations lead to longer equilibration times, which could explain that our estimations may be limited for the largest system due to limited statistics. Even if more statistics would be better to improve the description of these systems, we can conclude that the concentration gradient is practically insensitive to the size of the system box, once the steady state condition for the concentration gradient is reached at a given median temperature. Therefore, the smallest box system of 50 \AA is the best option in terms of computational resources and lower statistic errors.

5.3.2 Dependence on the gradient amplitude (∇T)

Another aspect that could affect the Soret coefficient value is the thermal gradient amplitude. Larger thermal gradient amplitudes will induce stronger thermally-driven effect, and hence a stronger response of the solute molecules inside the thermal gradient. The objective of this chapter is to check whether there is a variation in the slope of the concentration profiles as a function of the temperature for various temperature gradient amplitudes, the range of these thermal amplitudes (∇T) is set from 0.8 to $3.6 \text{ K} \cdot \text{\AA}^{-1}$.

The concentration profiles of the 2.17-m TMAO solutions are shown in Figure 5.10 at different thermal gradient amplitudes (∇T) of 0.8 , 1.2 , 2.4 and $3.6 \text{ K} \cdot \text{\AA}^{-1}$. The left graph represents the concentration gradient profile for a simulation box (L_z) of 50 \AA at a median temperature of 330 K . The temperature difference (ΔT) increases from 20 K to 90 K . The shaded areas indicate the standard deviations estimated from 20 independent trajectories for the largest and the smallest thermal gradients. We observe that the difference in TMAO concentration between hot and cold regions is accentuated as the amplitude of the thermal gradient increases, as expected from eq. (3.9) ($d \ln c^s / dT|_z = -S_T^s(z)$). We notice that the concentration profile gets smoother as the gradient amplitude increases.

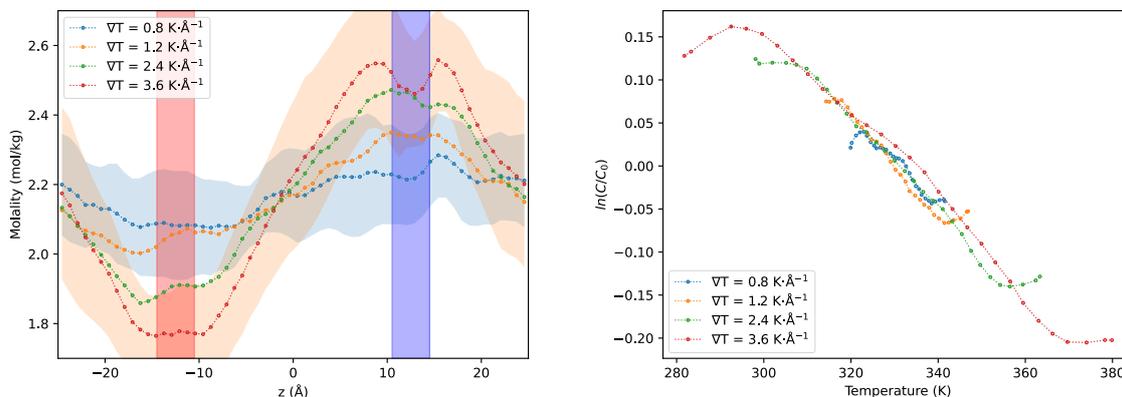


Figure 5.10: Gradient amplitude dependence. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{F_k}^{0.8\text{K}\cdot\text{\AA}^{-1}} = 0.0125 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$, $\Delta Q_{F_k}^{1.2\text{K}\cdot\text{\AA}^{-1}} = 0.01875 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$, $\Delta Q_{F_k}^{2.4\text{K}\cdot\text{\AA}^{-1}} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$ and $\Delta Q_{F_k}^{3.6\text{K}\cdot\text{\AA}^{-1}} = 0.05625 \text{ kcal} \cdot \text{mol}^{-1}\text{fs}^{-1}$. Data are shown for four different gradients ∇T , ranging from 0.8 to $3.6 \text{ K} \cdot \text{\AA}^{-1}$, in a 2.17-m TMAO solution. $P = 1 \text{ bar}$. $T_0 = 330 \text{ K}$. $L_z = 50 \text{ \AA}$. The left graph represents the TMAO molality as a function of z . The shaded areas indicate the standard deviations estimated from 20 independent trajectories at the two extreme gradients. The blue and red bars indicate the regions where heat is pumped and injected, respectively. The right graph represents the same data on a log-scale for the molality and as a function of temperature. The Soret coefficient at the reference temperature can be interpreted as the slope of the curves, which are basically identical at the centre of the gradient.

We confirm that the absolute concentration differences increase as ΔT increase. However, the Soret coefficient values around the median temperature remain constant. This effect can be observed by plotting the concentrations as a function of the temperature, which are represented in the right graph of Figure 5.10. We observe similar concentration gradient variation for the different thermal amplitudes as long as they are outside the thermostatted regions. Therefore, a critical effect of the gradient amplitude is to increase the signal-to-noise ratio. While the smallest investigated gradient of $0.8 \text{ K} \cdot \text{\AA}^{-1}$, which corresponds to $\Delta T = 20 \text{ K}$ here, readily leads to a noticeable concentration-gradient, the resulting profile suffer from statistical uncertainties that are almost as large as the concentration-difference themselves. For the largest investigated gradient ($\nabla T = 3.6 \text{ K} \cdot \text{\AA}^{-1}$, i.e., $\Delta T = 90 \text{ K}$), the concentration-difference is much larger than the statistical uncertainties. A compromise has to be found between the accuracy of the measured concentration-gradient, that should be large enough to exceed the statistical uncertainties, while employing a reasonable temperature-gradient where the water phase lies

within, or not too far, from its equilibrium liquid phase domain (namely, between 273 and 373 K). In the following, we chose to employ the $2.4 \text{ K} \cdot \text{\AA}^{-1}$ gradient, which allows to satisfy these conditions in most cases.

5.3.3 Testing eHEX exchange rate algorithm

The enhanced kinetic exchange algorithm (eHEX) allows us to adapt the amount of non-translational energy transferred from the cold region to the hot region, as well as its periodicity (τ_{eHEX}), which is the equivalent of the damping parameter of the thermostats or the barostats. However, the thermal gradient in the previous simulations was generated by transferring a certain amount of heat (ΔQ_{Γ_k}) at each timestep ($\tau_{eHEX} = \Delta t$). In this section, we analyse the effect of decreasing this heat transfer frequency, but we increased the amount of heat transferred to keep the same amplitude of the thermal gradient. We thus decrease the eHEX frequency ($\tau_{eHEX} = 10 \cdot \Delta t$), so the eHEX algorithm will transfer 10 times more heat every 10 timesteps. Our aim is to see if the concentration gradient is affected by the way in which the eHEX algorithm is applied. The simulation timestep (Δt) is 1 fs, so the eHEX algorithm is applied every 10 fs.

First, we have checked that the total energy is conserved along the whole simulation, and the thermal gradient profiles are the same for both simulations. The concentration profiles of the 2.17-m TMAO solutions are shown in the left graph of Figure 5.11. The concentration gradients at different τ_{eHEX} have the same amplitude, and they are not really affected by the eHEX algorithm. Moreover, the Soret coefficients values extracted are really similar. This result demonstrates the consistency of the eHEX algorithm for our systems and their respective thermalised regions, since reservoirs contain sufficient amount of water molecules inside, and hence the exchanged non-translational kinetic energy can be always distributed among the water molecules inside the reservoirs.

Moreover, the concentration gradient in the left graph of Figure 5.11 shows a higher deformation inside the reservoirs for the system with a larger eHEX periodicity ($\tau_{eHEX} = 10 \text{ fs}$), this particularity can also be observed in the concentration gradient profile of the simulation at the thermal gradient amplitude of ($\Delta T = 90 \text{ K}$) from Figure 5.10. This effect

could be explained due to the amount of heat transferred per time. The simulation $\tau_{eHEX} = 10$ fs transfers a large amount of heat every 10 fs, however the simulation $\tau_{eHEX} = 1$ fs

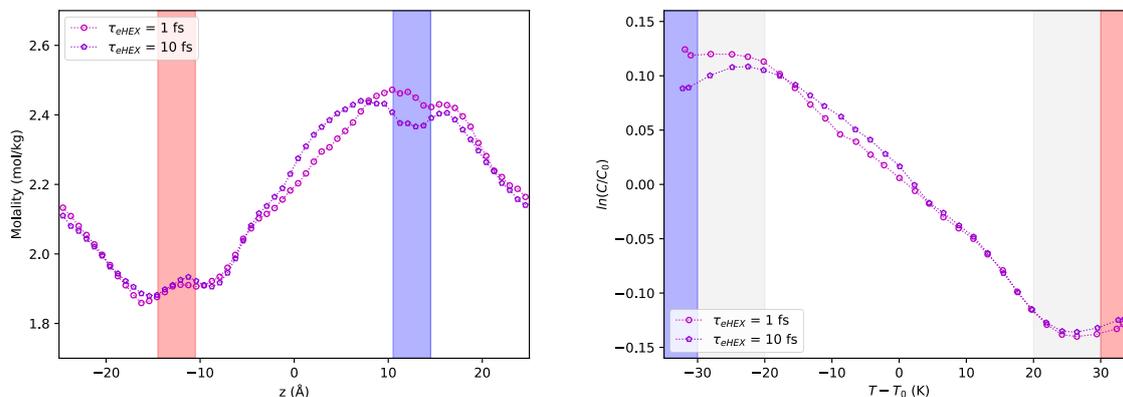


Figure 5.11: Heat exchange rate dependence. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60$ K. Data are shown for two different heat exchange rates τ_{eHEX} for 1 fs and 10 fs, in a 2.17-m TMAO solution. $P = 1 \text{ bar}$. $T_0 = 330 \text{ K}$. $L_z = 50 \text{ \AA}$. The left graph represents the TMAO average concentration as a function of the box length. The blue and red bars indicate the regions where heat is pumped and injected, respectively. The right graph represents the same data on a log-scale for the molality and as a function of temperature.

transfers smaller amounts of heat in shorter timesteps of 1 fs. Therefore, the frequency of heat transfer might play an important role in the molecular concentration inside the reservoir, but it does not really affect the concentration profile outside the thermalised regions. For this reason, we always determine the Soret coefficient from the concentration profile in the white region, as shown in the right graph of Figure 5.11. The blue and red zones are the thermalised regions, and the grey zone is a transition zone and it is also discarded, because it could have some side effects from the implementation of the eHEX algorithm in the concentration gradient. The exclusion grey zones are the first 10 K of the thermal gradient after the temperatures for both reservoirs (T_{hot} and T_{cold}). For example, our thermal gradient has 60 K and it varies between 300 K 360 K. The white zone is suitable for the calculation of the Soret coefficient, and comprises the temperature range from 310 K to 350 K, which is a thermal gradient amplitude of 40 K. In the following, we chose to employ the $\tau_{eHEX} = 1$ fs, which gives us better concentration gradient profiles in the thermalised regions.

5.3.4 Different force fields

Most force fields are parametrized at constant temperature and pressure. However, our systems exhibit large thermal gradients in a small box at the same time. In this section, we test different force field combinations for the solute (TMAO) and the solvent (water) to the possible effects in the TMAO concentration profile and hence a difference in the Soret coefficient. We choose two different force field for each molecule. The TMAO molecule is simulated with the Kast model,^{128,129} which is the one used during the previous sections, and we have compared it with the CgenFF36,¹³⁷ which is a Charmm general force field. The water molecule is represented by two known models; the 4-site TIP4P-2005 model,¹⁴⁴ and the 3-site SPC/E model.¹⁴⁵

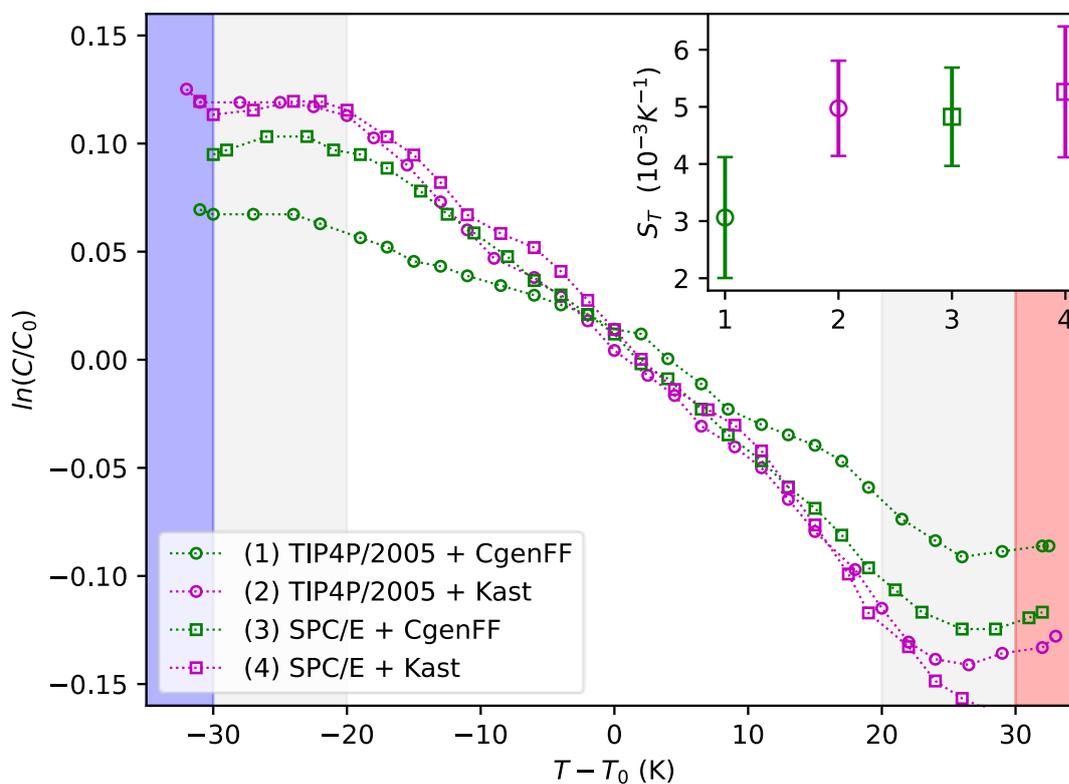


Figure 5.12: Force fields dependence. TIP4P/2005+CgenFF (green circle), TIP4P/2005+Kast17 (magenta circle), SPC/E+CgenFF (green square) and SPC/E+Kast17 (magenta square), in a 2.17-m TMAO solution. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{T_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs logarithmic pressure ($\log_{10}(P)$), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{\text{sample}} = 50$, $N_{\text{trajectories}} = 10$).

Figure 5.12 represents the $\ln(C/C_0)$ as a function of the temperature difference ($T - T_0$) for four possible force field combinations; (1) TIP4P/2005+CgenFF36, (2) TIP4P/2005+Kast, (3) SPC/E+CgenFF36 and (4) SPC/E+Kast. Colours represent the TMAO model force field, and point shape represents the different water model force fields. We have checked that the total energy is conserved for all the force field couples, and the thermal gradient profiles are the same for all the simulations. We do not observe any particular behaviour that might indicate an incompatibility between each pair of force fields. Moreover, we also represent the Soret coefficient for each force field combination in the inset graph in Figure 5.12.

The concentration profiles as a function of the temperature are pretty linear in the white regions of the graph. These results demonstrate a certain stability, which is independent of the force field combination applied. However, the first force field combination (TIP4P/2005+CgenFF36) has a lower Soret coefficient compared with the rest of the combinations. This lower value of the Soret coefficient might be caused by the charge distribution, we have analysed the charges of the CgenFF36 force field and the charge difference between the oxygen and the nitrogen of the TMAO molecules is ($q_O - q_N = 0.46$), and the Kast model is ($q_O - q_N = 1.42$). Even if the global molecule is neutral, the dipole moment associated to the TMAO for the Kast model is higher than the TMAO CgenFF36 model, which might explain this difference. We have analysed in more detail the TMAO force fields, and the Kast model was parametrized first with TIP3P water mater,¹²⁸ and it was later improved with the TIP4P/2005 for higher pressures,¹²⁹ which sustains the combination of the Kast model with the TIP4P/2005 water model, but it could not be compatible with the SPC/E water model.

On the other hand, the third force field combination (SPC/E+CgenFF36) has the same Soret coefficient as the second force field combination (TIP4P/2005+Kast) as well as the fourth force field combination (SPC/E+Kast). Usually, CgenFF36 are parametrized with the SPC/E or TIP3P water models, which can explain the good agreement with the other results. As a conclusion, three of four force field combinations give us the same results, which indicates a good consistency for the force field combination between the TIP4P/2005 water model and the Kast TMAO model.

6. Intrinsic and thermodynamic conditions modulating the Soret coefficient

Previously, we observed that the amplitude of the concentration gradient in response to a temperature gradient was little sensitive to the following technical simulation details; the box size, the amplitude of thermal gradient, the non-translational kinetic exchange rate, and the force field. We now focus on the system dependence of the Soret coefficient. Several heuristic descriptions in the literature exist, which try to decompose the Soret coefficient into different contributions to gain a better understanding of this effect. Each description focuses on a few contributions and tries to describe a specific system, but often fails when it is applied to another system. The goal of this chapter is to see whether similar conclusions can be reached by our simulations.

In this chapter, we mainly take the example of a dilute TMAO solution as our reference molecule to study the Soret coefficient. However, other solutes (ethanol, urea, methanol and glucose) are also included in some sections of this work to support the discussion of a particular phenomenological aspect. The first three sections, we focus on external thermodynamic factors, which are pressure, median temperature and solute concentration. Furthermore, we study some intrinsic properties of solutes that can affect the Soret coefficient. We analyse the nature of the solute by comparing solutes which are “chemically” different. Finally, we perform an extensive study of the effect of mass on the Soret effect. We analyse different mass ratios between the solute and the solvent. We check the sensibility of the mass effect for different molecules and we carry out an “isotopic” study of water molecules with different masses in the same system.

6.1 Low-sensitivity of pressure effect

In this section, we test whether median pressure makes a relevant contribution to the Soret effect. We are interested to see if pressure has a strong impact or not on the Soret coefficient. The different median pressures are established during the equilibration process. Each median pressure of our system is defined during the NPT simulation process, which is the third step in the equilibration part, before the activation of the eHEX thermal gradient. The resulting volume of the box together with the median temperature dictates the median pressure for the NVE ensemble. Although the generation of a thermal gradient induces local pressure differences among the thermal slabs, these condensed aqueous phase systems have relatively low thermal expansion coefficients, so we do not expect large local pressures variations along the thermal gradient. Therefore, the median pressure of the whole system is the only pressure effect studied in this section, and we do not evaluate local pressures along the thermal gradient in our simulation system.

We prepare a 2.17 m dilute aqueous solution of TMAO with a box length (L_z) equal to 50 Å, and the other lengths equal ($L_z \approx L_x = L_y$). These simulations are performed at a median temperature (T_0) of 330 K, a thermal gradient amplitude (ΔT) of 60 K and we analyse a range of median pressures from 1 bar to 1 kbar. All Soret coefficients are determined for a median temperature of 330 K.

We now investigate the pressure dependence of the Soret effect. In Figure 6.1, we show the concentration gradient evolution as a function of the temperature for a TMAO solution at different median pressures: $P_1 = 1$ bar, $P_2 = 10$ bar, $P_3 = 100$ bar and $P_4 = 1$ kbar. We do not observe a great difference in behaviour as median pressure increases. All logarithmic concentrations profiles exhibit similar response under the thermal gradient and the TMAO molecule tends to accumulate in the cold region regardless of the pressure exerted in the system. The inset graph represents the Soret coefficient values as a function of the decimal logarithm of the pressure. Inside this graph, we can observe the pressure effect in more detail. $P_1 = 1$ bar and $P_2 = 10$ bar have almost the same Soret coefficient value, which is around $S_T^{P_1} \approx S_T^{P_2} \approx 5 * 10^{-3} \text{K}^{-1}$. The only difference is that P_2 has a slightly smaller error in its Soret coefficient, but it is not possible to conclude that higher pressures improve the concentration profile by this

difference. However, we observe a small increase of the Soret coefficient for the median pressures of $P_3 = 100$ bar and $P_4 = 1000$ bar, which are around $S_T^{P_3} \approx 6 \cdot 10^{-3} \text{K}^{-1}$ and $S_T^{P_4} \approx 7 \cdot 10^{-3} \text{K}^{-1}$ respectively:

$$S_T^{P_1} \approx S_T^{P_2} < S_T^{P_3} < S_T^{P_4}. \quad (6.1)$$

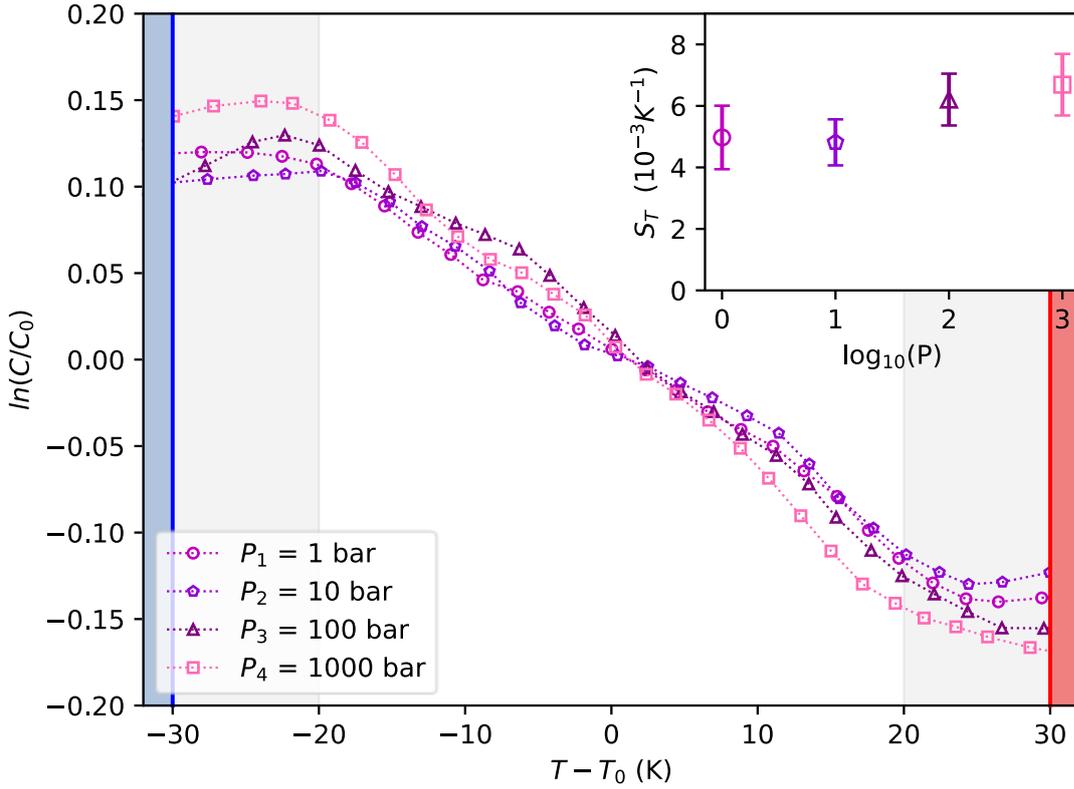


Figure 6.1: Median pressure dependence. $P_1 = 1$ bar, $P_2 = 10$ bar, $P_3 = 100$ bar and $P_4 = 1000$ bar, in a 2.17-m TMAO solution. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60$ K. $L_z = 50$ Å. $T_0 = 330$ K. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs logarithmic pressure ($\log_{10}(P)$), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{\text{sample}} = 50$, $N_{\text{trajectories}} = 10$).

Although we observe an increase of the Soret coefficient as a function of the median pressure, the slope is very low ($dS_T/dP \approx 2 \cdot 10^{-6} \text{K}^{-1} \text{bar}^{-1}$). Therefore, large pressures must be applied to observe a significant effect in the Soret coefficient for dilute aqueous solutions. We can thus ensure that the median pressure does not have an important effect in thermophoresis,

and the median pressure and local pressure differences that may be generated along the temperature gradient should not greatly affect the Soret coefficient value.

6.2 Temperature effect

In this section, we study the contribution of the median temperature to the Soret coefficient. As seen before, the amplitude of thermal gradient does not have a significant effect on the behaviour of solute accumulation. However, several experiments and computational studies have demonstrated that the Soret coefficient does actually have a median temperature dependence^{80,91,92}. However, there is no clear consensus on how the Soret coefficient changes as a function of the median temperature, as different mixtures exhibit different behaviours.

For example, ethanol/water mixtures⁸⁴ have a thermophobic behaviour $S_T > 0$ at lower concentrations consistent with our previous simulations of TMAO aqueous solutions, but it was found that the ethanol thermophobicity decreases with temperature $\partial S_T / \partial T < 0$. On the other hand, urea/water mixtures⁷⁹ exhibit a thermophobic behaviour for all possible concentrations $S_T > 0$, but its evolution as a function of temperature depends on the urea concentration. The Soret coefficient increases with the temperature $\partial S_T / \partial T > 0$ at lower concentrations ($< 30\%$ weight fraction) (~ 5.4 m), but decreases $\partial S_T / \partial T < 0$ at higher concentrations ($> 30\%$ weight fraction) (~ 5.4 m).⁷⁹

Therefore, we analyse whether our results show these temperature trends for urea and ethanol, and we see how the TMAO molecule behaves as a function of the median temperature. It should be considered that simulations at 270 K and 390 K are slightly outside the condensed liquid phase, and are just included to support to our simulations in the 300-360 K range.

6.2.1 Ethanol molecule

We now investigate the temperature dependence of the Soret effect for the ethanol molecule. In Figure 6.2, we show the concentration gradient evolution as a function of the temperature for an ethanol solution at different median temperatures: $T_0^{(1)} = 270$ K, $T_0^{(2)} = 300$ K, $T_0^{(3)} = 330$ K and $T_0^{(4)} = 360$ K. The logarithmic concentrations for median temperatures between 270 K and 330 K exhibit similar response under the thermal gradient and the ethanol molecule

tends to accumulate in the cold region. However, the simulation at a median temperature of 360 K shows a practically flat concentration profile, which might indicate that the ethanol molecule at this median temperature ($T_0^{(4)}$) does not have a preference for the cold or the hot region. Moreover, we observe an inversion of the concentration gradient in the grey regions as they approach the thermostatted regions, which might indicate a higher sensibility to the eHEX algorithm.

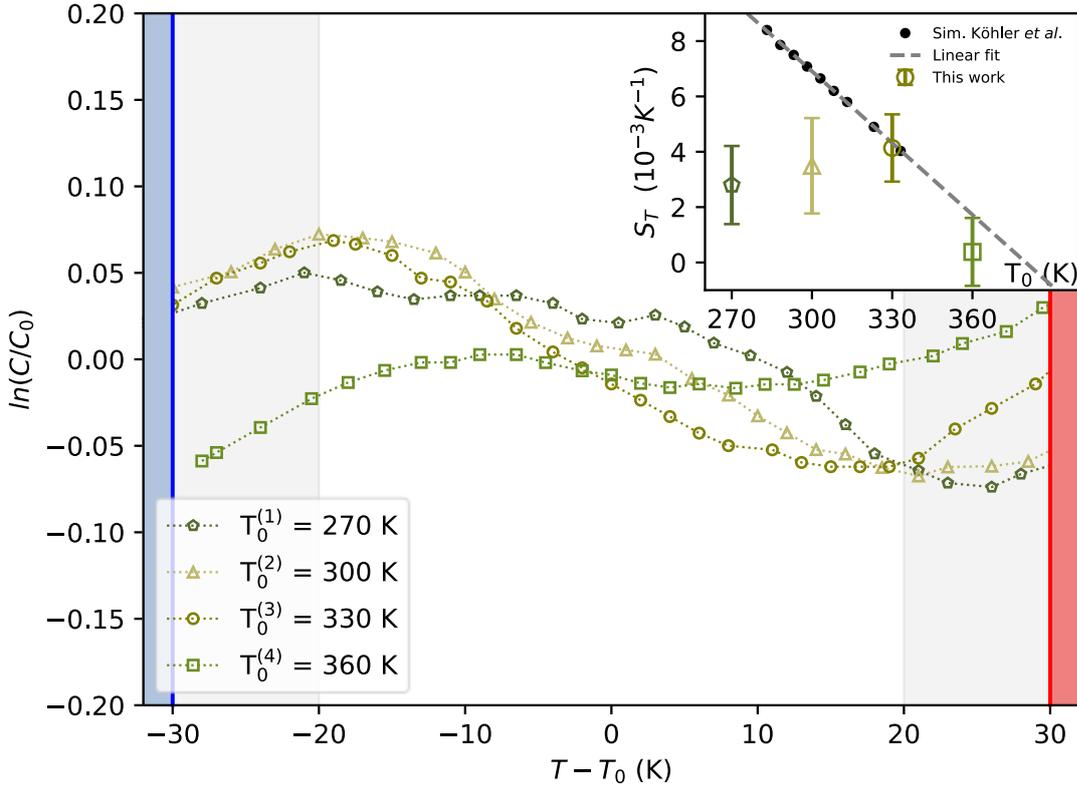


Figure 6.2: Median temperature dependence of ethanol. $T_0^{(1)} = 270$ K, $T_0^{(2)} = 300$ K, $T_0^{(3)} = 330$ K and $T_0^{(4)} = 360$ K, in a 2.17-m ethanol solution. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375$ kcal \cdot mol $^{-1}$ fs $^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60$ K. $P = 1$ bar. $L_z = 50$ Å. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs median temperature (T_0), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$). The black dots are the experimental data of Köhler *et al.*⁸⁴ and the grey line is the linear fit.

Furthermore, the inset graph represents the Soret coefficient values as a function of the median temperature. We observe a small increase of the Soret coefficient along the median temperatures of $T_0^{(1)}$, $T_0^{(2)}$ and $T_0^{(3)}$, which are around $S_T^{270K} \approx 2.8 \cdot 10^{-3}K^{-1}$, $S_T^{300K} \approx 3.5 \cdot$

10^{-3}K^{-1} and $S_T^{330\text{K}} \approx 4.1 * 10^{-3}\text{K}^{-1}$ respectively, but the Soret coefficient completely drops to zero ($S_T^{360\text{K}} \approx 0 \text{K}^{-1}$) at the highest median temperature ($T_0^{(4)} = 360 \text{K}$).

$$S_T^{270\text{K}} < S_T^{300\text{K}} < S_T^{330\text{K}} \gg S_T^{360\text{K}} . \quad (6.2)$$

The Köhler group⁸⁴ experimentally measured several Soret coefficients for the ethanol/water mixtures at several temperatures and concentrations, and supported their results with previous experimental data.^{168–170} Our results show similar values in the Soret coefficient for the median temperature at 330 K, $S_T^{330\text{K}}$ (this work) $\approx S_T^{330\text{K}}$ (Köhler). However, our Soret coefficients at the lowest temperatures ($T_0^{(1)}$ and $T_0^{(2)}$) underestimate their experimental values, but they are still at the same order of magnitude. Unfortunately, we do not have experimental data for their Soret coefficient at the median temperature of 360 K, but their results show that the thermophobicity of ethanol decreases with temperature, $\partial S_T / \partial T < 0$, as we have mentioned before. We have thus extrapolated their trend, and we found that the Soret coefficient is zero at a median temperature of 380 K, $S_T^{380\text{K}}$ (Köhler) $\approx 0 \text{K}^{-1}$, which is a temperature difference of 20 K between our results and this extrapolation.

6.2.2 Urea molecule

We now investigate the temperature dependence of the Soret effect for the urea molecule. In Figure 6.3, we show the concentration gradient evolution as a function of the temperature for a urea solution at different median temperatures: $T_0^{(1)} = 270 \text{K}$, $T_0^{(2)} = 300 \text{K}$, $T_0^{(3)} = 330 \text{K}$ and $T_0^{(4)} = 360 \text{K}$. All concentrations for median temperatures between 270 K and 360 K exhibit similar response under the thermal gradient, and the urea molecule tends to accumulate in the cold region. The simulation at a median temperature of 270 K shows a small slope in concentration profile, which might indicate that the urea molecule at this median temperature ($T_0^{(4)}$) does have a relatively small thermophobic behaviour. However, no inversion of the concentration gradient is observed in the grey regions in contradiction with the previous study of ethanol, which may indicate that urea is less sensitive to the eHEX algorithm in the reservoirs.

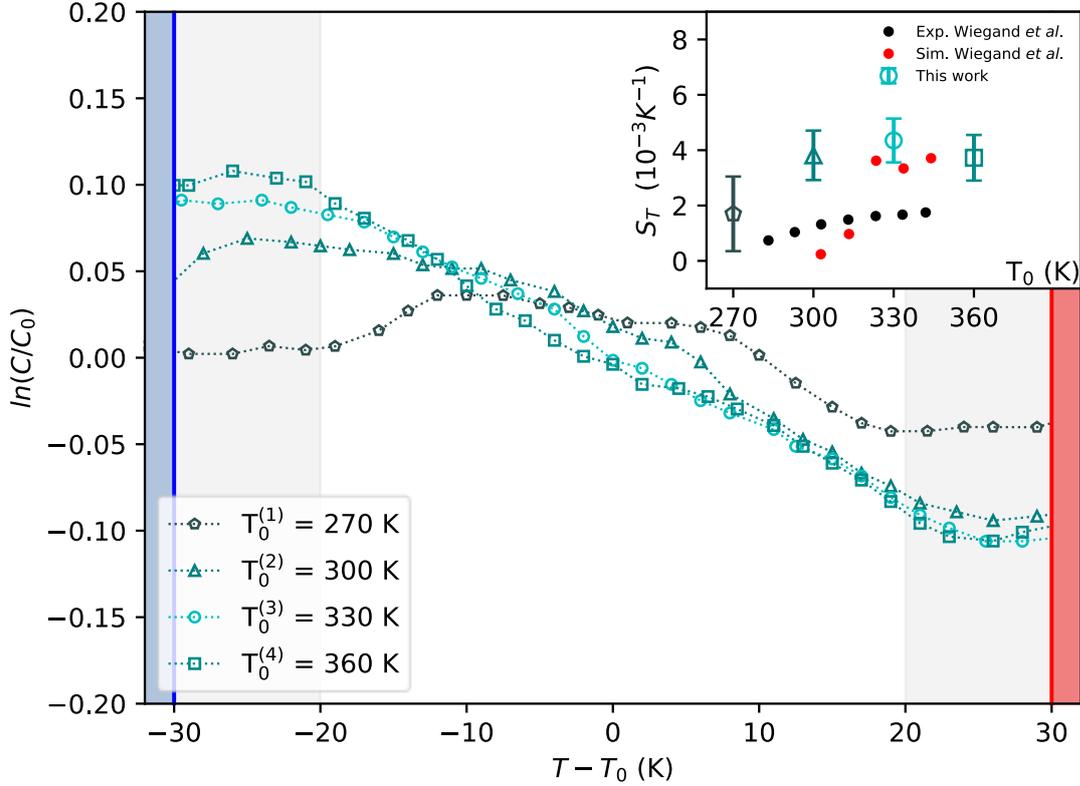


Figure 6.3: Median temperature dependence of urea. $T_0^{(1)} = 270$ K, $T_0^{(2)} = 300$ K, $T_0^{(3)} = 330$ K and $T_0^{(4)} = 360$ K, in a 2.17-m urea solution. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375$ kcal \cdot mol $^{-1}$ fs $^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60$ K. $P = 1$ bar. $L_z = 50$ Å. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs median temperature (T_0), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$). The dots are the experimental (black) and computational (red) data of Wiegand *et al.*

79

The inset graph represents the Soret coefficient values as a function of the median temperature. We observe a progressive small increase of the Soret coefficient along the median temperatures of $T_0^{(1)}$, $T_0^{(2)}$ and $T_0^{(3)}$, which are around $S_T^{270K} \approx 1.7 \cdot 10^{-3} \text{K}^{-1}$, $S_T^{300K} \approx 3.8 \cdot 10^{-3} \text{K}^{-1}$, $S_T^{330K} \approx 4.3 \cdot 10^{-3} \text{K}^{-1}$, but the Soret coefficient is slightly reduced, $S_T^{360K} \approx 3.7 \cdot 10^{-3} \text{K}^{-1}$ at the highest median temperature ($T_0^{(4)} = 360$ K). Even if we observe small variations between 300 K and 360 K, these differences are inside their error bars, and we can consider that the Soret coefficient between 300 K and 360 K is in practice the same:

$$S_T^{270K} < S_T^{300K} \approx S_T^{330K} \approx S_T^{360K} . \quad (6.3)$$

Wiegand *et al.*⁷⁹ experiments (IR-TDFRS) and simulations (NEMD) calculated several Soret coefficients for the urea/water mixtures at several temperatures and concentrations. Our Soret coefficient results overestimate their experimental (black dots) results by a difference of $\Delta S_T = S_T$ (this work) $- S_T$ (Wiegand) $\sim 2.5 * 10^{-3} K^{-1}$. However, our results follow a similar trend of the Soret coefficient as a function of the median temperature. On the other hand, our results are comparable to their Soret coefficient (red dots) computationally obtained at median temperature of $T_0^{(3)} = 330$ K (S_T^{330K} (this work) $\approx S_T^{330K}$ (Wiegand)), but our Soret coefficients are overestimated at lower median temperature (S_T^{300K} (this work) $\gg S_T^{300K}$ (Wiegand)). In conclusion, our results are approximately in agreement with the trends observe by experimental works and we obtain similar Soret coefficient values for some temperatures in relation to these non-equilibrium studies.

6.2.3 TMAO molecule

We have noted in the last two sections that both ethanol/water and urea/water mixtures exhibit positive Soret coefficients, so that ethanol and urea tend to accumulate in the cold region. Our simulations show a positive and progressive evolution of the Soret coefficient for both solutes. However, ethanol/water and urea/water solutions start to behave differently at higher temperatures ($T_0^4 = 360$ K); the Soret coefficient of urea seems to stabilise at higher temperatures, but the Soret coefficient of ethanol decays to zero. These preliminary results for urea and ethanol are qualitatively in agreement with the experimental results of the Köhler group⁸⁴ and Wiegand *et al.*⁷⁹ Therefore, we now proceed to the study of temperature dependence for our main solute TMAO.

In the previous section, we were already able to observe some aspects of the TMAO/water solutions that are relevant for the study of their Soret coefficient. These simulations showed a TMAO accumulation in the cold region, which is similar to the thermophobic behaviour of the urea/water and ethanol/water solutions. Therefore, a positive Soret coefficient value is expected for these dilute TMAO solutions, even if no experimental data is currently available.

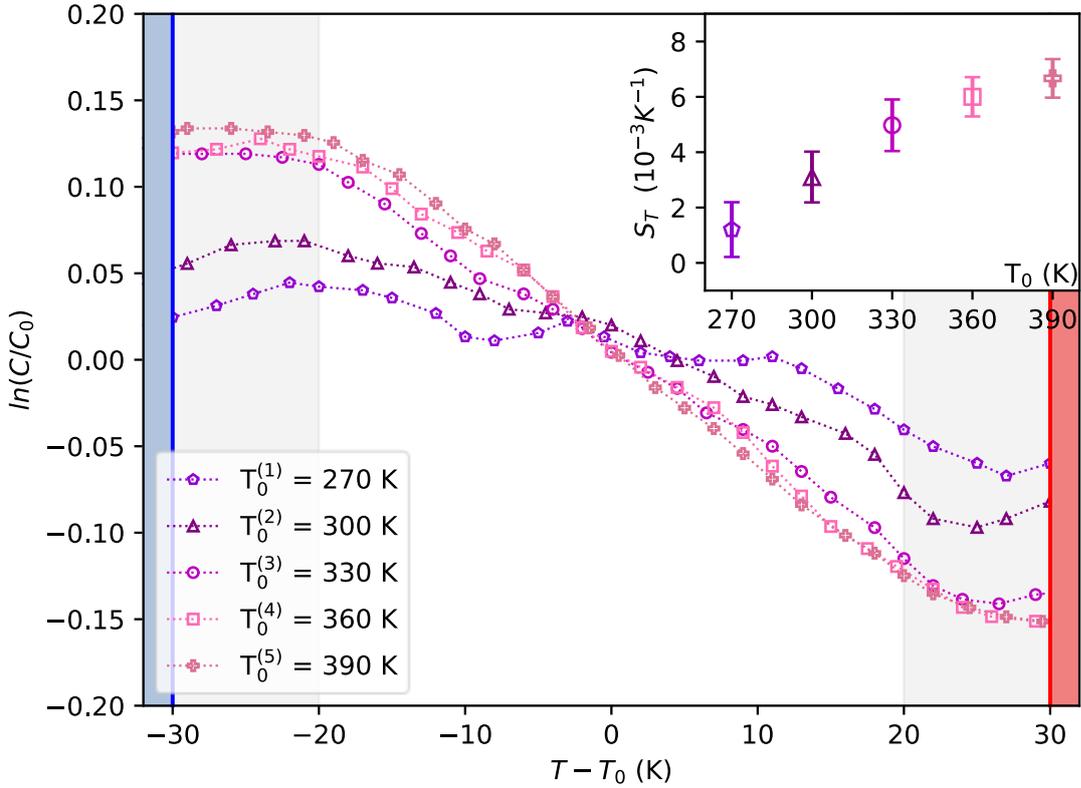


Figure 6.4: Median temperature dependence of TMAO. $T_0^{(1)} = 270$ K, $T_0^{(2)} = 300$ K, $T_0^{(3)} = 330$ K, $T_0^{(4)} = 360$ K and $T_0^{(5)} = 390$ K, in a 2.17-m TMAO solution. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60$ K. $P = 1$ bar. $L_z = 50 \text{ \AA}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs median temperature (T_0), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{\text{sample}} = 50$, $N_{\text{trajectories}} = 10$).

We now investigate the temperature dependence of the Soret effect for the TMAO molecule. In Figure 6.4, we show the concentration gradient evolution as a function of the temperature for a TMAO solution at different median temperatures: $T_0^{(1)} = 270$ K, $T_0^{(2)} = 300$ K, $T_0^{(3)} = 330$ K, $T_0^{(4)} = 360$ K and $T_0^{(5)} = 390$ K. A. We observe that higher temperatures show a stronger TMAO thermophobic behaviour, since the concentration increases in the cold region as the median temperature increases. Moreover, we notice that the concentration gradients at higher median temperature are smoother. For example, 270 K and 300 K median temperatures have larger fluctuations on their concentration profiles than higher temperature. One possible explanation of these fluctuations could be that the TMAO molecule aggregates at these

temperatures. However, several studies have demonstrated that TMAO does not easily aggregate,¹²⁶ where concentrations up to 5,6 M do not exhibit any sign of aggregation,¹²⁷ and we are largely below these concentrations. Moreover, we have analysed in more detail these simulation boxes, and we do not find any indication of aggregation. Furthermore, simulations at lower concentrations also exhibit these fluctuation effects at a median temperature of 300 K. We conclude that these observations are merely a fluctuation effect, since the concentration gradient decreases with the temperature, and it makes more difficult to get well defined linear gradients. Another possible reason could be that lower temperatures translate into less kinetic motion, and hence the TMAO molecules require more time to sample all the possible positions in the box system, but the statistical errors in the Soret coefficient do not really differ from the highest temperature ones. This particularity also appears in the ethanol and urea systems from Figure 6.2 and Figure 6.3 respectively.

The Soret coefficient of TMAO in the inset graph does not present any abrupt change in the Soret coefficient evolution. This result lets us think that the TMAO/water mixture behaviour is more alike to urea solutions than ethanol ones, since it always exhibits $\partial S_T / \partial T > 0$. We thus observe a progressive increase of the Soret coefficient along all median temperatures, which are around $S_T^{270K} \approx 1.2 * 10^{-3} K^{-1}$, $S_T^{300K} \approx 3.1 * 10^{-3} K^{-1}$, $S_T^{330K} \approx 5.0 * 10^{-3} K^{-1}$, $S_T^{360K} \approx 6.0 * 10^{-3} K^{-1}$ and $S_T^{390K} \approx 6.7 * 10^{-3} K^{-1}$:

$$S_T^{270K} < S_T^{300K} < S_T^{330K} < S_T^{360K} < S_T^{390K} . \quad (6.4)$$

This behaviour confirms a decrease in the thermophobic character of the TMAO molecule with increasing the median temperature, and these Soret coefficients of TMAO follows an exponential depletion form, which might be fitted with the phenomenological equations described in the previous chapters.

6.2.4 Empirical equations

We analyse this thermal evolution of the Soret coefficient in more detail for the three solutions studied before. We replot the Soret coefficients obtained from Figure 6.2, Figure 6.3 and Figure 6.4.

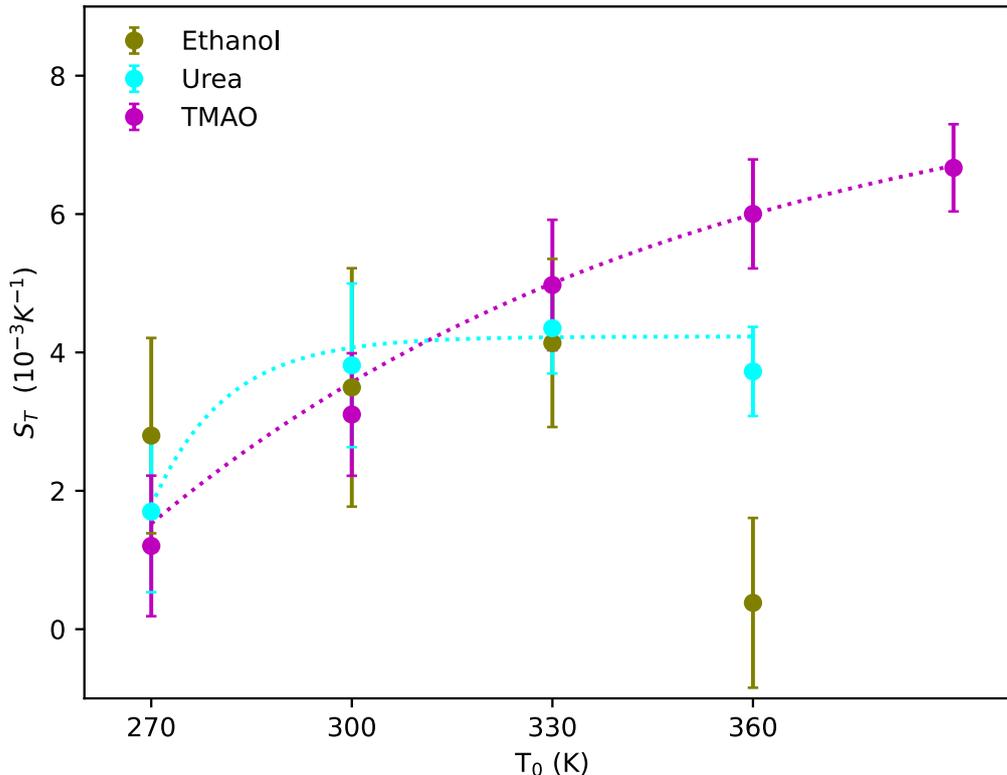


Figure 6.5: Empirical fits. Ethanol (olive), urea (cyan) and TMAO (magenta) solutions. Data was averaged over 20 independent trajectories. $P = 1$ bar. $L_z = 50$ Å. $T_0 = 330$ K. The errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$). The dashed lines are fittings of the exponential function $S_T = S_T^\infty [1 - \exp(T^* - T)/T_i]$ for urea and TMAO.

We also investigate the temperature dependence of the Soret effect for the three solutes. In Figure 6.5, we show the Soret coefficient vs the median temperature (T_0) for the ethanol (olive), urea (cyan) and TMAO (magenta) solutions and their respective error bars. We also include the respective exponential function fittings ($S_T = S_T^\infty [1 - \exp(T^* - T)/T_i]$) for urea and TMAO. We do not fit the ethanol Soret coefficients, because we already observed that ethanol does not follow same exponential depletion form from the literature⁸⁴ and our results confirmed this trend at high median temperatures. We observe that the Soret coefficient values are practically the same at the lowest median temperatures for the three solutes (270 K and 300 K), but they start to diverge from 330 K onwards, and finally they are completely different at the median temperature of 360 K.

As discussed before, the absolute value of the Soret coefficient depends on a large number of properties, which makes it difficult to get a direct comparison among the different solutes. In fact, the temperature dependence of the Soret coefficient is determined by the interactions between the solute and water. Wiegand *et al.*⁷⁹ discussed that the chemical potential varies in a linear way, increasing with temperature for the urea, hence indicating a decrease in solubility, which might explain the different temperature behaviours of these solutes. However, we have not gone deeper into this aspect, since we have focused on getting a more global perspective of the different contributions on the Soret effect.

Empirical equations		Urea		TMAO
		This work	Wiegand <i>et al.</i> ⁷⁹ (exp)	This work
$S_T(T) =$	$S_T^\infty (10^{-3} \text{K}^{-1})$	4.02	1.9	8.33
$S_T^\infty \left[1 - \exp \frac{(T^* - T)}{T_i} \right]$ ⁸⁰	$T^* (\text{K}^{-1})$	264	269	253
	$T_i (\text{K}^{-1})$	11	29	84
$S_T(c, T) =$	$S_T^\infty (10^{-3} \text{K}^{-1})$	4.02	1.74	8.33
$S_T^\infty - C_H \exp(-A_H T)$ ⁹²	$C_H (\text{K}^{-1})$	$2.75 \cdot 10^{10}$	$2.14 \cdot 10^6$	167
	$A_H (\text{K}^{-1})$	0.086	0.051	0.012

Table 6.1 : Empirical fits. Phenomenological empirical fits^{80,92} for the urea and TMAO molecules.

The table above represents the fitted Iacopini/Piazza and Wiegand phenomenological descriptions for our TMAO and urea molecules, and we also include the urea's experimental results of Wiegand *et al.*⁷⁹ The equation of Iacopini and Piazza⁸⁰ described the Soret coefficient by these factors: S_T^∞ represents the limiting value of $S_T(T)$ at high temperatures, T^* the temperature at which the Soret coefficient changes the sign, i.e. $S_T(T) = 0 \text{ K}^{-1}$, and T_i embodies the strength of temperature effect. The equation of the Wiegand group⁹² described the Soret coefficient by these other factors: S_T^∞ are the thermal properties of the core material, possible charges, and so forth, excluding the presence of hydrogen bonds, C_H is a measure of the contribution of the hydrogen bonds, and the parameter $A_H > 0$ measures the temperature-dependent strength of a hydrogen bond.

The limiting Soret coefficient value for the Iacopini and Piazza⁸⁰ equation at high temperatures was $S_T^\infty = 0.00402 \text{ K}^{-1}$ for the urea solution, which indicates the maximum point of thermophobicity that urea/water mixture can reach at this given concentration 2.17 m. The $S_T^\infty = 0.0019 \text{ K}^{-1}$ experimental result from Wiegand *et al.*⁷⁹ is two times smaller, which is not surprising since we are overestimating the Soret coefficient of urea in our simulations compared with these experimental results. However, the Soret coefficient sign changes at a median temperature of $T^* = 264 \text{ K}$ and is comparable with the experimental results of $T^* = 269 \text{ K}$, so we should expect a thermophobic behaviour over a wide temperature range for urea systems in dilute aqueous solutions ($T_0 > 270 \text{ K}$). Finally, $T^i = 11 \text{ K}$ is the strength of the temperature effect for our results and $T^i = 29 \text{ K}$ for Wiegand results.

In the case of the TMAO molecule, the limiting Soret coefficient value at high temperatures was $S_T^\infty = 0.00833 \text{ K}^{-1}$, which indicates the maximum point of thermophobicity that TMAO/water mixture can reach, and TMAO thermophoretic strength is the double as urea. The Soret coefficients at 330 K and 360 K have relatively high values compared to this limit S_T^∞ , which may indicate that TMAO could reach his higher thermophobicity at not relatively high temperatures. Furthermore, the sign changes at the median temperature of $T^* = 253 \text{ K}$, so we must denote thermophobic behaviour over a wide temperature range for TMAO systems in dilute aqueous solutions similar to urea. Finally, $T^i = 84 \text{ K}$ was the strength of the temperature effect, which is bigger than the urea solution.

The phenomenological description of the Soret coefficient in the Wiegand group⁹² shows similar limiting Soret coefficients (S_T^∞) as compared with the Iacopini and Piazza⁸⁰ exponential fit. The Wiegand group defined that the contribution of hydrogen bonds is expected to be the main cause of temperature dependence of the Soret coefficient, so the preexponential factor (C_H) describes describes the diminishing contribution of hydrogen bonds as they weaken with increasing temperature. Our results show that the urea molecule has a large preexponential factor, but the TMAO molecule temperature dependence is not really relevant ($C_H^{urea} \gg C_H^{TMAO}$). Moreover, hydrogen bonds weaken more strongly with increasing temperature for larger values of A_H . Then we should expect that TMAO hydrogen bonds ($A_H = 0.012 \text{ K}^{-1}$) remain stable with increasing temperature, but the urea hydrogen bonds strength ($A_H = 0.085 \text{ K}^{-1}$) decreases with the median temperature. Moreover, the Wiegand group found a

linear correlation between $\ln(C_H)$ and A_H , which is a really practical, because reduces sensitivity of the strength of hydrogen bonds to one parameter instead of two, but this concept needs to be tested for a larger number of systems.⁷

6.3 Concentration effect

The previous studies analysing the effect of the median temperature have also reported a concentration effect in their Soret coefficients.^{79,84} Often concentration and temperature effects are studied together, but they do not usually include the concentration effect of the solute in their empirical equations to describe the Soret coefficient. We found an example that includes a concentration effect, which is the Wittko and Köhler⁹¹ phenomenological equation, $S_T(c, T) = \alpha_c(c)\beta_T(T) + S_T^{ch}$, where $\alpha_c(c)$ and $\beta_c(T)$ are polynomial empirical fits. However, this equation is still purely empirical and does not attempt to inquire into the nature of the Soret coefficient.

Usually, the Soret coefficient remains unchanged in sufficiently dilute solutions, but at high concentrations, the Soret coefficient exhibits stronger variation with concentration. These non-ideality and association effects has been widely studied during the 1950s by Prigogine^{72,73} and later by Drickamer.¹⁷¹ The authors often observed a change in the sign of the Soret coefficient with concentration, where particles exhibit thermophobic behaviour at low concentrations, but the same particles in the system become thermophilic at high concentrations. Moreover, Kolodner¹⁶⁸ and later Köhler⁸⁴ described a detailed transition in the thermal diffusion of the ethanol/water mixture. Kolodner and Köhler observed a transition point of the Soret coefficient at a mass fraction of 0.3, independently of the median temperature.

These studies are carried out for a wide range of concentrations, to describe the evolution of the Soret coefficient as a function of concentration, with mass fractions ranging from to 1. However, our project does not explicitly address the concentration issue, since we are mostly interested by the dilute regime. Our main concern is that we could perform the simulations at high enough concentrations such that convergence of the concentration gradient occurs on reasonably accessible timescales (hundreds of ns), while being in a sufficiently dilute regime such that the Soret coefficient is almost concentration independent. Therefore, we study the stability of the Soret coefficient for the TMAO dilute concentrations in the mass fraction range

of 0.08-0.25. This range of concentrations would allow us to confirm whether or not there is an effect for our main concentration of 0.14 mass fraction (2.17 m). Furthermore, the dilute aqueous solutions guarantee that enough water molecules are thermalised, and avoids additional considerations for the eHEX algorithm, as it will not alter the temperature gradient, nor we will apply the eHEX algorithm for the TMAO molecules.

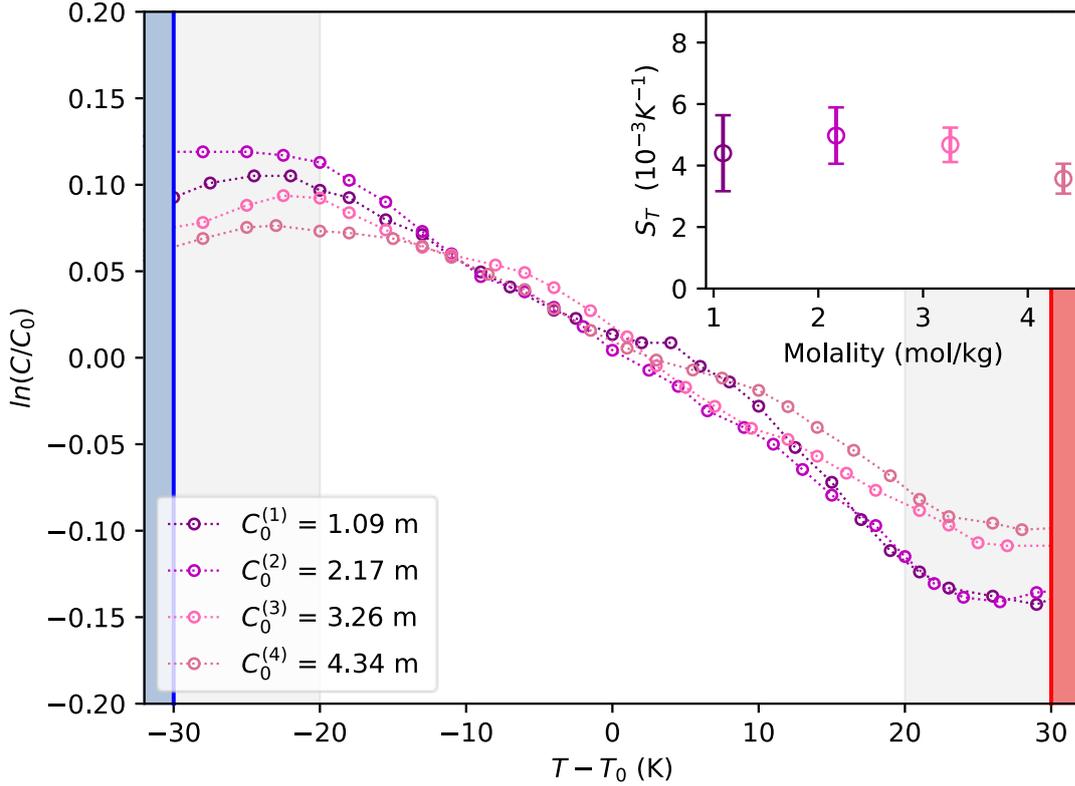


Figure 6.6: Average concentration dependence of TMAO. $C_0^{(1)} = 1.09 \text{ mol} \cdot \text{kg}^{-1}$, $C_0^{(2)} = 2.17 \text{ mol} \cdot \text{kg}^{-1}$, $C_0^{(3)} = 3.26 \text{ mol} \cdot \text{kg}^{-1}$ and $C_0^{(4)} = 4.34 \text{ mol} \cdot \text{kg}^{-1}$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{T_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs average concentration (C_0), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{\text{sample}} = 50$, $N_{\text{trajectories}} = 10$).

We now investigate the concentration dependence of the Soret effect. In Figure 6.6, we show the concentration gradient evolution as a function of the temperature for a TMAO solution at different molalities: $C_0^{(1)} = 1.09 \text{ mol} \cdot \text{kg}^{-1}$, $C_0^{(2)} = 2.17 \text{ mol} \cdot \text{kg}^{-1}$, $C_0^{(3)} = 3.26 \text{ mol} \cdot$

kg^{-1} and $C_0^{(4)} = 4.34 \text{ mol} \cdot \text{kg}^{-1}$. As we have mentioned before, several studies have demonstrated that TMAO does not easily aggregate,¹²⁶ with the first coordination shell of TMAO that is slightly reduced with concentration. Moreover, it is known that concentrations lower than $C_0 < 6 \text{ mol/kg}$ do not exhibit any sign of aggregation,¹²⁷ and our systems are below these concentrations.

The Soret coefficients of TMAO in the inset graph does not exhibit a strong dependence with concentration. The Soret coefficients for median concentrations no larger than $C_0^{(3)} = 3.26 \text{ mol} \cdot \text{kg}^{-1}$ are the same within error bars. This may indicate that in sufficiently dilute solutions, the Soret coefficient is not concentration dependent. This effect could be understood by the fact that the TMAO molecules practically do not interact among themselves, and their respective water solvation shell depends only on the temperature gradient, and thus keeping the “chemical” interaction solute-solvent invariant.¹²⁶ However, a small deviation is observed for the Soret coefficient at the largest concentration ($C_0^{(4)} = 4.34 \text{ mol} \cdot \text{kg}^{-1}$), which might show a variation in the solvation shell between the molecules located in the hot and cold slab.

Increasing the total TMAO concentration makes solute-solute interactions more predominant, then concentration differences between the hot and cold regions may accentuate a larger differentiation in their respective “chemical” contributions. This reduction in the Soret coefficient may be caused by a change in the chemical contribution along the concentration gradient that directly competes with the effect of thermophoresis, making TMAO molecules less likely to accumulate in the cold region. The molal concentration of $C_0^{(4)} = 4.34 \text{ mol} \cdot \text{kg}^{-1}$ is a relatively high concentration, this represents 12.8 water molecules per TMAO, which is larger than the hydration shell: therefore, hydration shells overlap and TMAO molecules are in close contact, and we start observing some effects on the Soret coefficient for this reason:

$$S_T^{C_0^{(1)}} \approx S_T^{C_0^{(2)}} \approx S_T^{C_0^{(3)}} > S_T^{C_0^{(4)}}. \quad (6.5)$$

Furthermore, the Soret coefficient error decreases with increasing TMAO concentration. Therefore, we have opted for the median concentration of $C_0^{(2)} = 2.17 \text{ mol} \cdot \text{kg}^{-1}$, since it contains enough TMAO molecules to obtain good statistics for the determination of the Soret

coefficient while being in a regime that is dilute enough such that it does not affect too much the Soret coefficient.

6.4 Solute effect

We have already seen that external effects contribute to the Soret coefficient, like the median temperature, pressure or concentration. However, the Soret coefficient depends also on several physico-chemical properties of the molecule in itself, e.g., mass, size, moment of inertia or polarity.⁷

In the literature, the Soret coefficient is usually phenomenologically divided into two independent terms ($S_T = S_T^{ch} + S_T^{iso}$):^{76,102} S_T^{iso} corresponds to “isotopic contribution”, which is the additive part related to the difference in mass and moment of inertia of the components, and S_T^{ch} correspond to “chemical contribution”, which reflects the actual chemical activity of the compounds. Typically, the isotopic term of the Soret coefficient has been proposed in some empirical formulations, which describe the effect of mass and inertia with limited accuracy. However, the “chemical” term is still described in vague terms, because it tries to describe the interactions between solute-solvent as well as solute-solute along the thermal gradient, and as it already mentioned, it could be related to the chemical potential of the solution.⁷⁹ Consequently, the S_T^{ch} depends on the particle environment and is therefore sensitive to temperature and concentration. Understanding it is a key point of thermophoresis for dilute aqueous solutions.

In this section, we study different solutes that have similar characteristics in comparison to the TMAO molecule. These molecules are methanol, ethanol, urea and glucose. Our objective is to unveil some trends of their Soret coefficients, even if we are restricted in precision by the statistical error. First, we validated each step of the simulation for all solutes, using a similar procedure as for TMAO. The total energy is conserved during all the stages for all solutes, the temperature gradient is constant and linear along the box, and we confirm the absence of any aggregation for all solutes under the following conditions; $P = 1$ bar, $T_0 = 330$ K, $\Delta T = 60$ K, $L_z = 50$ Å, $C_0 = 2.17$ mol · kg⁻¹ (40 solute molecules in 1024 water).

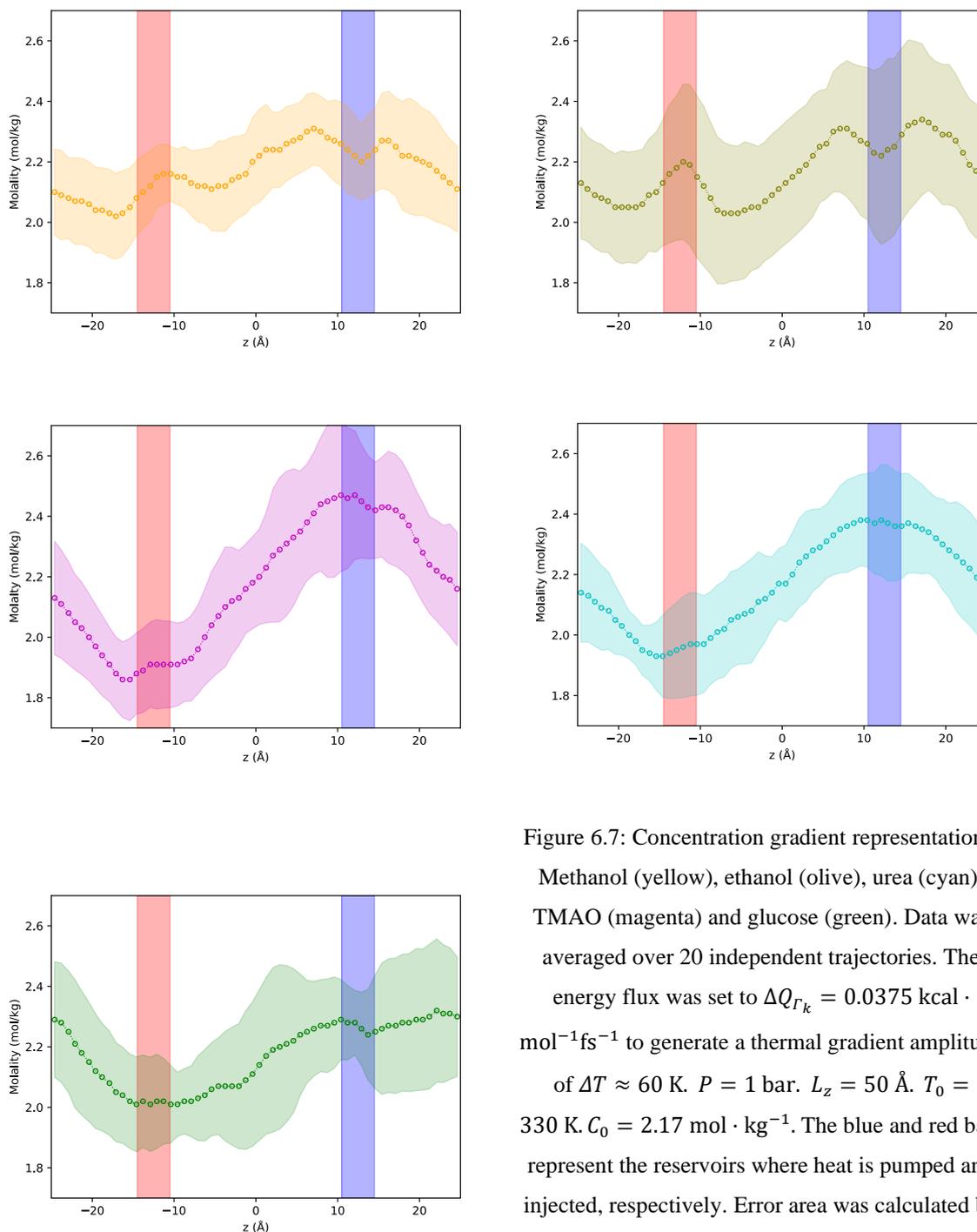


Figure 6.7: Concentration gradient representations. Methanol (yellow), ethanol (olive), urea (cyan), TMAO (magenta) and glucose (green). Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Error area was calculated by the standard deviation from 20 independent trajectories.

In Figure 6.7, we show the concentration gradient for each solute as a function of the box length L_z . The blue and red areas represent the cold and hot regions, where the kinetic exchange is active $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1}$. The error area is the standard deviation of 20 independent

trajectories. All solutes exhibit thermophobic behaviour, as they accumulate in the cold region (blue area), and therefore we expect a positive value for their Soret coefficients. This behaviour was already observed in the previous sections for TMAO (magenta), ethanol (olive) and urea (cyan), and we also have the same effect for methanol (yellow) and glucose (green).

The error areas are similar for all solutes in their respective plots in Figure 6.7, indicating that methanol and glucose can be studied under the same simulation conditions as we the previously studied molecules of TMAO, ethanol and urea, and therefore they are able to sample the phase-space region for a 10 nanoseconds simulation. However, we observe a loss of definition of the concentration gradient for methanol and glucose, which could be due to the lower amplitude gradient for these solutes. As we have seen before, larger concentrations amplitudes exhibit clearer linear concentration gradient, such as TMAO or urea.

Small abrupt variations can be also observed in the regions subject to the heat exchange, where the eHEX algorithm is active. One clear example is the concentration gradient of the ethanol molecule. The nature of this effect remains unclear. It may be due to the constant alteration of the kinetic energy in that region, which prevents reaching thermodynamic equilibrium and leads to an inversion of the Soret coefficient, and the ethanol or methanol solutions are more sensitive to this effect than urea or TMAO. For this reason, this affected area and its surroundings are excluded for the determination of the Soret coefficient and only the central part of the concentration gradient part is considered.

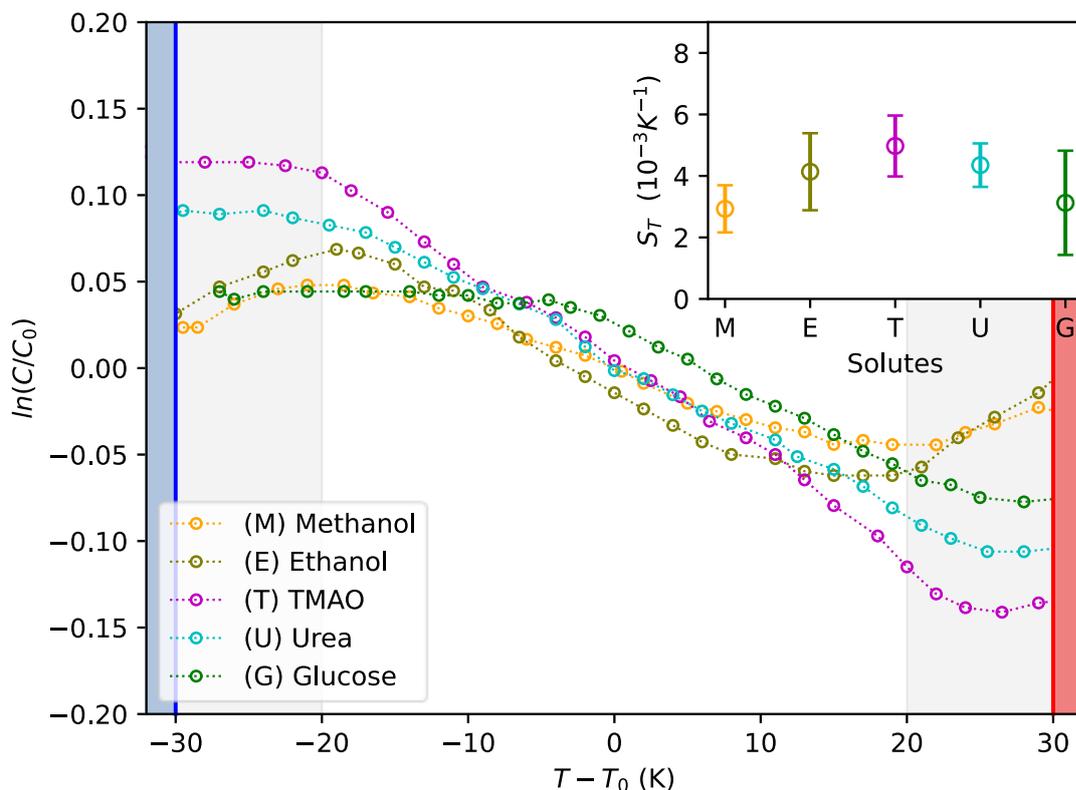


Figure 6.8: Solute effect. methanol (yellow), ethanol (olive), TMAO (magenta), urea (cyan) and glucose (green). Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs different solutes, the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{\text{sample}} = 50$, $N_{\text{trajectories}} = 10$).

We now investigate the “chemical” dependence of the Soret effect. In Figure 6.8, we show the concentration gradient evolution as a function of temperature for the solutes; methanol, ethanol, urea, TMAO and glucose. The Soret coefficients for each solute are represented in the inset graph in order to gain a more detailed description of the understanding of the “chemical” and isotopic contributions. These results are in agreement with previous studies^{79,84,172} since we are working on dilute solutions and the solute molecules are larger than the solvent one. Moreover, we notice that their Soret coefficients remain in the same order of magnitude, because these molecules exhibit similar characteristics of mass, size, and polarity.

	Molar mass ($g \cdot mol^{-1}$)	Dipole moment (D)	Molecular volume (\AA^3)	Soret coefficient ($10^{-3} K^{-1}$)
Water	18	2.4	30	-
Methanol	32	2.3	66	2.9 ± 0.7
Ethanol	46	2.4	119	4.1 ± 1.0
Urea	60	4.5	77	4.4 ± 1.0
TMAO	75	6.4	124	5.0 ± 0.7
Glucose	180	1.0	196	3.1 ± 1.1

Table 6.2: Solute effect. Molar mass, dipole moments, molecular volume and Soret coefficients for the water, methanol, ethanol, urea, TMAO and glucose molecules.

Table 6.2 shows some intrinsic molecular properties for our solutes to better understand their respective Soret coefficient values. We determined the dipole moment of each molecule in the gas phase for each force field to qualitatively check if there is a correlation between the dipole moment and the Soret coefficient, and we compared our dipole moment values with the literature results.^{173–176} Even if we obtained dipole moments 30% larger than those observed in the literature, we observed that all solutes have same trend except for glucose, which corresponds to the solvation effect since our dipole moments are determined in the gas phase. Molecular volumes are calculated by the volume difference between a pure water box composed of 1024 water molecules and a solution of the respective 40 solutes and 1024 water molecules. These simulations are carried out at $P = 1$ bar and $T_0 = 330$ K in the NPT ensemble, and the average volume is determined along 200 ps. Even if we do not take into account the excess volume for each system, these molecular volumes approximations are enough to decipher some trends.

The following formula orders the Soret coefficient as a function of the solute:

$$S_t^{Methanol} \approx S_t^{Glucose} < S_t^{Ethanol} \approx S_t^{Urea} < S_t^{TMAO} . \quad (6.6)$$

As we have mentioned before, “chemical” contributions of the Soret coefficient are mainly dominated by the interactions between solute and solvent for dilute solutions. Hydrophilic parts of the molecule are the predominant interaction part in aqueous solutions, such as dipole-dipole forces. Therefore, we compare the Soret coefficient for each molecule with the calculated dipole moments in the gas phase:

$$\mu_{Glucose} < \mu_{Ethanol} \approx \mu_{Methanol} < \mu_{Urea} < \mu_{TMAO}. \quad (6.7)$$

Broadly speaking, we observe larger Soret coefficients for molecules with larger dipole moments such as TMAO or urea.

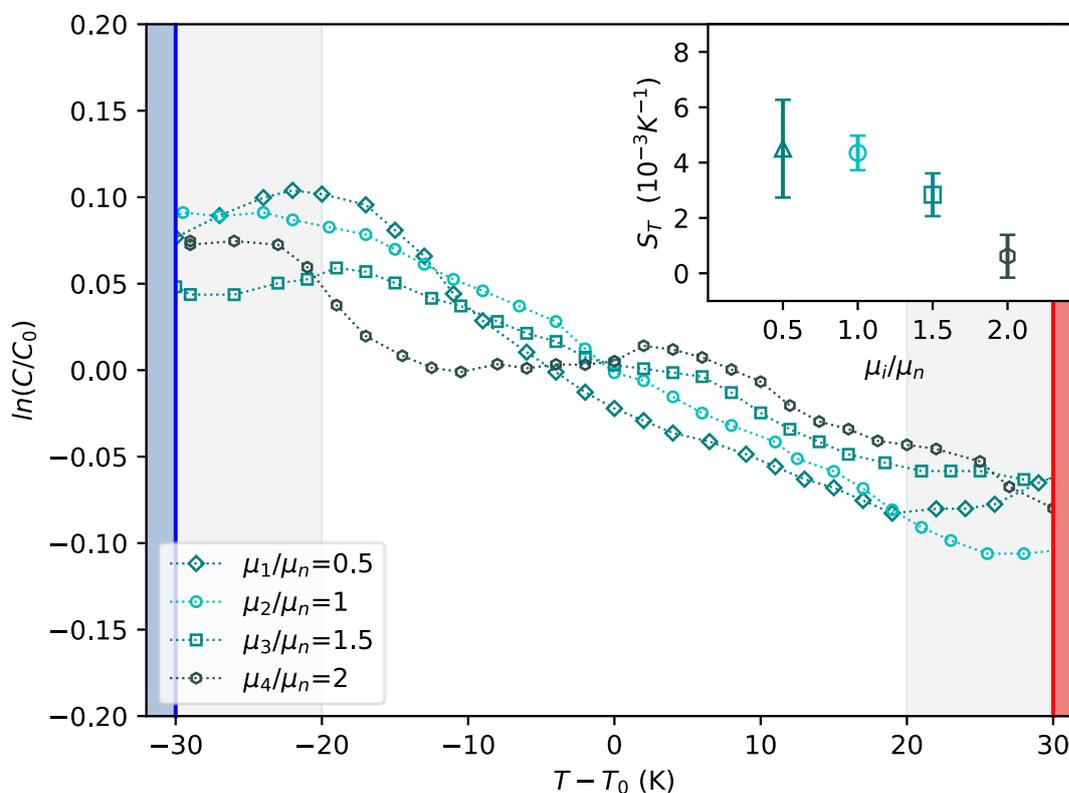


Figure 6.9: Dipole moment dependence of urea. $\mu_1/\mu_n = 0.5$, $\mu_2/\mu_n = 1$, $\mu_3/\mu_n = 1.5$ and $\mu_4/\mu_n = 0.5$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative dipole moment (μ_i/μ_n), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

To investigate the effect of the solute dipole moment independently of other effects (mass, volume, etc), we repeated our simulation of the urea solution after rescaling the atomic charges of each atom by the factors 0.5, 1.5 and 2.0. Rescaling all charges by the same factor allow us to get different total dipole moments, conserving the neutrality of the molecule.

We now investigate the dipole moment dependence of the Soret effect for the urea molecule. In Figure 6.9, we show the concentration gradient evolution as a function of relative dipole moments: $\mu_1/\mu_n = 0.5$, $\mu_2/\mu_n = 1$, $\mu_3/\mu_n = 1.5$ and $\mu_4/\mu_n = 2$. In all cases, urea remains thermophobic and any urea dipole variation does not show any indication to accumulate in the hot region. Surprisingly, the Soret coefficients of urea in the inset graph decreases as the dipole moment of the modified urea increases. This result contradicts the trends previously discussed regarding the dipole moment for different solutes. Previous studies already discussed this higher complexity for polar systems as compared to non-polar systems,¹⁰⁵ where the authors phenomenologically included a new term which is the energy ratio $\psi_\varepsilon = \varepsilon_{22}/\varepsilon_{11}$ ⁹⁰ in the description of the Soret coefficient ($S_T^i = \alpha_M \delta M + b_I \delta I + c_\varepsilon \Delta \psi_\varepsilon$). From our results, we can only predict that $c_\varepsilon \Delta \psi_\varepsilon$ has a negative effect on the Soret coefficient, seeing how urea behaves as its polarity increases, but the authors assumed that non-ideal structural changes (e.g., excess effects) of the polar systems are not fully covered by this simple approach.¹⁰⁵

Furthermore, we observe a certain instability on the concentration gradient, when the atomic charges of the urea are manipulated. Concentration profiles from Figure 6.9 show large fluctuations for simulations that are not the natural dipole moment of urea ($\mu_1/\mu_n = 1$), and posterior calculations of radial distribution functions have shown differences among these systems. Usually, force-fields are parametrized to describe the inter-molecular interactions as an equilibrium between the Lennard-Jones potential together with the Coulombic form, so altering the Coulombic forces might lead to unrealistic description of the inter-molecular forces, and hence invalidate the force-field. In conclusion, further studies must be done in order to understand the evolution of the Soret coefficient as a function of the dipole moment of a molecule.

The “isotopic” contributions of the Soret coefficient are also analysed by comparing the molar masses and the molecular volumes of these solutes. A priori, the evolution of the Soret

coefficient might indicate a direct correlation between his value and the molar mass of each solute:

$$M_{Methanol} < M_{Ethanol} < M_{Urea} < M_{TMAO} < M_{Glucose} \cdot \quad (6.8)$$

A progressive increase of the Soret coefficient is observed, as the molar mass of the molecules increases. However, some discrepancies have been found. Glucose molecule shows a concentration gradient lower as expected, even if it is the largest molecules $M_{Glucose} \approx 180$ g/mol in our set of molecules, his Soret coefficient value ($S_t^{Glucose} \approx 3 * 10^{-3} K^{-1}$) is significantly lower than TMAO ($S_t^{TMAO} \approx 5 * 10^{-3} K^{-1}$), which has a molecular mass of $M_{TMAO} \approx 75$ g · mol⁻¹. This result emphasises the idea that the Soret coefficient is not only dependent on the molar mass of the solute, and the mass effect actually acts as an additive term in the Soret coefficient together with other factors. Furthermore, we qualitatively check the evolution of the Soret coefficient as a function of the size of molecules:

$$V_{Methanol} < V_{Urea} < V_{Ethanol} < V_{TMAO} < V_{Glucose} \cdot \quad (6.9)$$

Some interesting points can be extracted from this point. We observe a direct relation between the Soret coefficient and the molecular volume of each molecule, except for glucose. This trend indicates that particle size could be relevant for the determination of the Soret coefficient, which is in agreement with previous studies that have reported a size dependency in colloidal solutions,^{49,68,177} but the dependence of the Soret coefficient on the radius of the solute particle is still controversial.⁷ It is worth noting that this rough approach does not take into account the shape, space-distribution and types of atoms that conforms the molecules. These aspects are relevant for the inertial moments of the molecule, which forms part of the “isotopic” contribution of some phenomenological descriptions in the literature,^{90,103} so more detailed description should be made in case one wants to go deeper into this effect.

In general terms, we observe that the lowest values in their Soret coefficients are associated with alcohol groups (methanol, ethanol and glucose). These organic molecules are composed of hydroxyl groups as functional groups (polar moiety) and a carbon chain as hydrophobic part. The difference between methanol and ethanol is only one methyl group, but the resulting Soret coefficient of the ethanol molecule is higher than methanol ($S_t^{Methanol} < S_t^{Ethanol}$). Table 6.2

show that their dipole moments are essentially the same ($\mu_{Ethanol} \approx \mu_{Methanol}$), but the molar mass ($M_{Methanol} < M_{Ethanol}$) and molecular volume ($V_{Methanol} < V_{Ethanol}$) of ethanol are bigger than methanol due to its extra carbon in his chain. These results indicate that the hydrophobic part of the molecule also plays a role in the thermal accumulation, as larger molecules could lead to a higher “isotopic” contribution. On the other hand, the glucose molecule does not follow this trend. In Table 6.2, we show that the molar mass ($M_{Ethanol} < M_{Glucose}$) and the molecular volume ($V_{Ethanol} < V_{Glucose}$) of glucose are bigger than ethanol, but the Soret coefficient ($S_t^{Ethanol} > S_t^{Glucose}$) and dipole moment ($\mu_{Ethanol} > \mu_{Glucose}$) of glucose are lower than ethanol. Glucose is a big molecule with several hydroxyl groups, so an equilibrium between hydrophobic and hydrophilic parts of the molecule might explain the reduction the thermophoretic motion toward the cold region. Nevertheless, high fluctuations are present in the concentration gradients for ethanol and glucose, and hence the Soret coefficient associated errors are higher, it is thus harder to get a clear conclusion from these systems.

The TMAO molecule has the highest Soret coefficient value ($S_t^{TMAO} = 5 \cdot 10^{-3} K^{-1}$), which is not surprising as this molecule contains the highest dipole moment and if we disregard the glucose molecule, it also has the highest molecular mass and molecular volume. The high dipole moment value ($\mu_{TMAO} = 6.4$ D) comes from his amino oxide functional group, which forms a strong hydrophilic head. Moreover, the TMAO molecules also contains a large hydrophobic moiety provided by three methyl groups which confers a high molecular mass ($M_{TMAO} = 75$ g/mol) and a high molecular volume ($V_{TMAO} = 124 \text{ \AA}^3$). All these intrinsic properties of TMAO could explain the relatively high value of its Soret coefficient compared to the other solutes, since the high dipole moment could be compensated by his mass and size.

We also observe that ethanol and urea have similar Soret coefficients ($S_t^{Ethanol} \approx S_t^{Urea}$). Urea is a completely polar molecule with two amine groups as well as carboxyl group, which justifies a higher dipole moment compared to the ethanol molecule ($\mu_{Ethanol} < \mu_{Urea}$). On the other hand, the urea molecule lacks a hydrophobic body, which explains its lower molecular volume compared to the ethanol molecule ($V_{Ethanol} > V_{Urea}$). Focusing on the dipole moment and size, we should expect the Soret coefficient of urea to be smaller than that of ethanol, but they are quite similar. On the other hand, the molecular masses between ethanol and urea are

similar ($M_{Urea} \approx M_{Ethanol}$), which can explain the similar Soret coefficients between ethanol and urea. Therefore, the mass effect as an intrinsic property of the molecule might determine the Soret coefficient rather than the size or the dipole moment.

6.5 Mass effect

Previously, we discussed the fact that the molar mass has a direct contribution in the Soret coefficient when comparing the masses of different solutes. The results have shown that an increase in molecular mass of the solute ($M_{Methanol} < M_{Ethanol} < M_{Urea} < M_{TMAO}$) generally resulted in a higher Soret coefficient ($S_t^{Methanol} < S_t^{Ethanol} \approx S_t^{Urea} < S_t^{TMAO}$). However, this trend is difficult to appreciate as there are other factors intrinsic to the molecule that also affect the Soret coefficient (size and polarity). In this section, we go one step further. The masses of some solutes are rescaled in order to gain a better understanding in the mass effect to the Soret coefficient, and all other things being equal.

The mass effect for a particular molecule can be studied by rescaling each atomic mass that forms part of the target molecule. M^n represents the molar mass of the natural molecule without any mass rescaling. In the case that the new molecular mass is larger than the natural molecular mass of the rescaled molecule, all atomic masses are multiplied by the same factor. However, when the new molecular mass is smaller than the natural molecular mass, special consideration must be taken into account to their hydrogen atoms. Small hydrogen mass could affect the stability of the molecular dynamics with energy conservation issues. In these cases, we only rescale the atomic masses of the atoms heavier than hydrogen in the molecule and retained the natural hydrogen mass. In addition, we distribute the percentage of hydrogen mass that should lose among the rest of heavier atoms in the molecule. This method allows us to study lighter molecules without constraining the bonds, angles and dihedrals of the whole molecule. However, we rescale the hydrogen masses of the water molecule, since this molecule has its bonds and angle already constrained and would not affect the dynamic of the system.

6.5.1 TMAO molecule

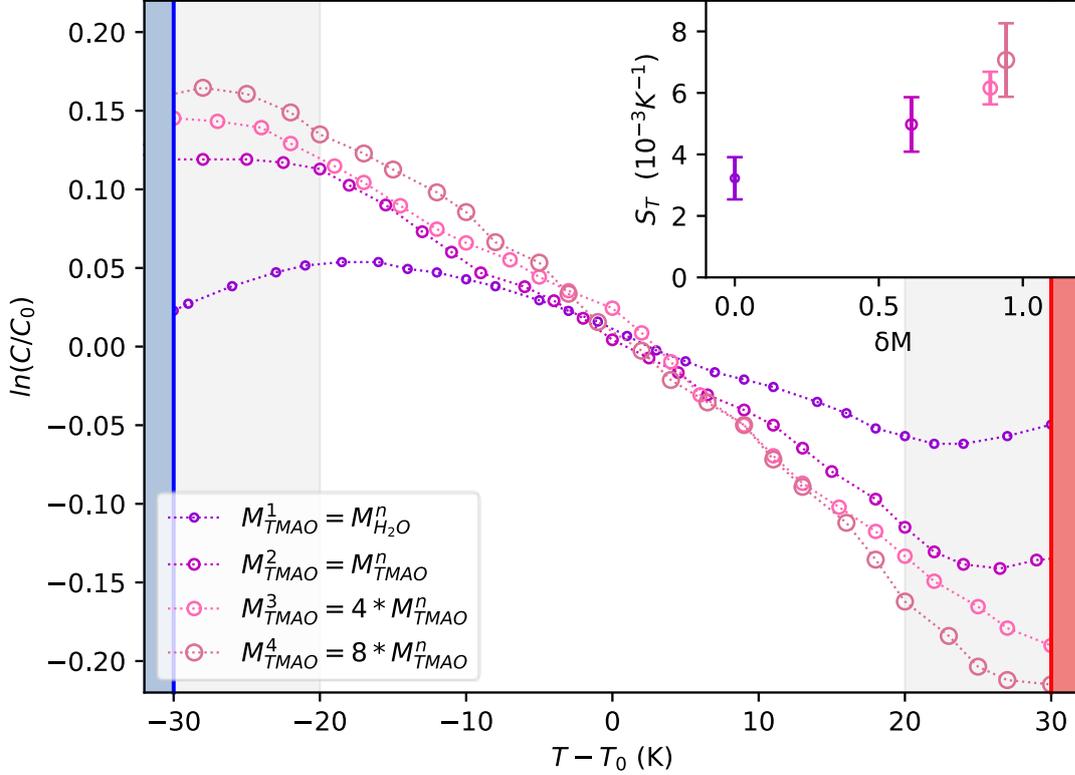


Figure 6.10: Mass effect of TMAO. $M_{TMAO}^1 = M_{water}^n$, $M_{TMAO}^2 = M_{TMAO}^n$, $M_{TMAO}^3 = 4 \cdot M_{TMAO}^n$ and $M_{TMAO}^4 = 8 \cdot M_{TMAO}^n$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative mass difference (δM), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

We now investigate the mass ratio dependence of the Soret effect for the TMAO molecule. In Figure 6.10, we show the concentration gradient evolution as a function of the temperature for a TMAO solution at different molar masses; $M_{TMAO}^1 = 18 \text{ g} \cdot \text{mol}^{-1}$, $M_{TMAO}^2 = M_{TMAO}^n = 75 \text{ g} \cdot \text{mol}^{-1}$, $M_{TMAO}^3 = 300 \text{ g} \cdot \text{mol}^{-1}$ and $M_{TMAO}^4 = 601 \text{ g} \cdot \text{mol}^{-1}$. A small increase in the concentration gradient can be perceived as the mass increases, as compared to normal mass and a smaller concentration gradient is observed when the mass of TMAO (M_{TMAO}^1) is equal to the mass of water (M_{water}^n). These early indications suggest that mass plays an active role in

thermophoresis. As we have mentioned previously, Galliéro *et al.*⁹⁶ justified this behaviour as a density effect. The TMAO molecule participates more efficiently in the generation of the strong density zone as its mass increases, because heavier molecules tend to migrate towards the cold areas more than the lighter ones, once the system is in steady state equilibrium.

The Soret coefficients of TMAO in the inset graph increase as a function of the mass ratio ($\delta M = (M_{TMAO}^i - M_{water}^n)/(M_{TMAO}^i + M_{water}^n) \rightarrow i = 1,2,3,4$), which their values are represented in this table.

	Molar mass solute ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3} K^{-1}$)
M_{TMAO}^1	18	18	3.2 ± 0.7
M_{TMAO}^2	75	18	5.0 ± 0.9
M_{TMAO}^3	300	18	6.2 ± 0.6
M_{TMAO}^4	601	18	7.1 ± 1.1

Table 6.3: Mass effect of TMAO. Molar masses of TMAO and water and its Soret coefficient value.

The Soret coefficient values in Table 6.3 shows the following trend as the mass of the TMAO molecule increases:

$$S_T^{M_{TMAO}^1} < S_T^{M_{TMAO}^2} < S_T^{M_{TMAO}^3} < S_T^{M_{TMAO}^4} . \quad (6.10)$$

This trend shows the thermophobic character of TMAO increases with its mass. The Soret coefficients of TMAO progressively grow with the mass ratio. An interesting point is that the TMAO solute with the water molar mass ($M_{TMAO}^1 = M_{water}^n$) shows a positive contribution in its Soret coefficient ($S_t^{M_{TMAO}^1} \approx 3 * 10^{-3} K^{-1}$). As discussed above in Chapter 3, the Köhler group^{76,102} defined the Soret coefficient as a contribution of two independent terms ($S_T = S_T^{ch} + S_T^{iso}$), and the isotopic part of the Soret coefficient is described in terms of mass and moment of inertia ratios ($S_T^{iso} = a_M \delta M + b_I \delta I$). In the case where the mass ratio is zero ($\delta M = 0$), the Soret coefficient ($S_T^{M_{TMAO}^1}$) can be redefined as:

$$S_T^{\delta M=0} = S_T^{ch} + b_I \delta I, \quad (6.11)$$

$S_t^{\delta M=0}$ could be the key to deciphering the other intrinsic properties effects of the molecule that affect the Soret coefficient. The moment of inertia depends on the mass and size of the molecule. Assume for a moment that these molecules behave as solid spherical particles. Thus, the inertial moments of these solid spherical particles can be defined by the following expression:

$$I = \frac{2}{5} MR^2, \quad (6.12)$$

where I is the inertial moment, M is the molecular mass and R is the molecular radius, and we can rewrite the isotopic contribution of the Soret coefficient from as:

$$S_T^{iso} = a_M \frac{M_{TMAO} - M_{water}}{M_{TMAO} + M_{water}} + b_I \frac{2 M_{TMAO} R_{TMAO}^2 - M_{water} R_{water}^2}{5 M_{TMAO} R_{TMAO}^2 + M_{water} R_{water}^2}. \quad (6.13)$$

If we consider that both components have the same mass ($M = M_{TMAO} = M_{water}$), the isotopic Soret coefficient is simplified to:

$$S_T^{iso} = b_I \frac{2 R_{TMAO}^2 - R_{water}^2}{5 R_{TMAO}^2 + R_{water}^2}, \quad (6.14)$$

and the volume of a solid sphere is defined as $V = \frac{4}{3} \pi R^3$, the isotopic Soret coefficient as a function of the molecular volume takes the form:

$$S_T^{iso} = b_I^* \frac{V_{TMAO}^{2/3} - V_{water}^{2/3}}{V_{TMAO}^{2/3} + V_{water}^{2/3}}, \quad (6.15)$$

where b_I^* is the spherical inertial factor. Thus, the Soret coefficient for the spherical particles of the binary mixture with the same mass ($S_t^{\delta M=0}$) can be understood in terms of molecular volume and "chemical" contribution. However, this approach needs to be studied in more detail as it does not take into account the shape of the molecules and therefore does not accurately describe the moments of inertia of each molecule.

On the other hand, we have also studied a system in which the mass of the water ($M_{water} \approx 144 \text{ g} \cdot \text{mol}^{-1}$) is eight times its natural mass, while the mass of the TMAO keeps its natural mass (M_{TMAO}^n). The result shows a Soret coefficient of $S_T \approx 2 \cdot 10^{-3} \text{ K}^{-1}$. Therefore, we observe that the Soret coefficient of the TMAO molecule decreases as the mass of water (solvent) increases, which indicates that the water mass also plays a role in the Soret coefficient value of the solute molecules. In the next step, we analyse whether the effect of mass on the Soret coefficient is due to the absolute value of the molecular mass of the solute and solvent molecules or rather to the difference in mass between them.

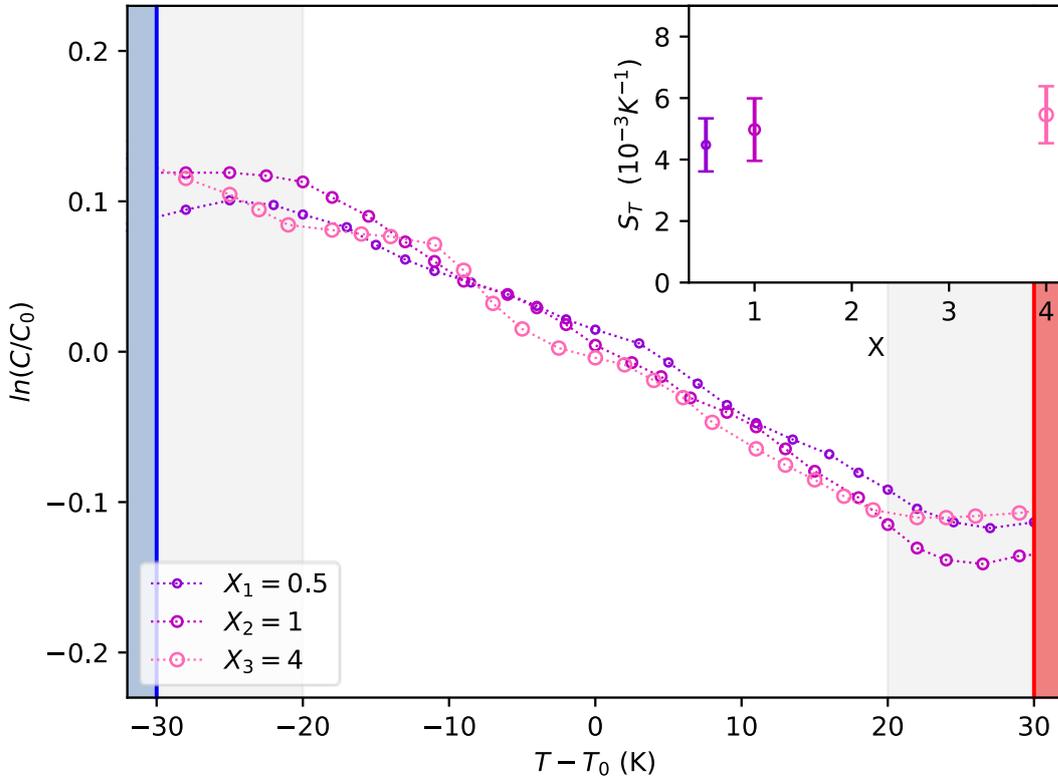


Figure 6.11: Mass ratio effect of TMAO. multiplicative factors are $X_1 = 0.5$, $X_2 = 1$ and $X_3 = 4$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{ fs}^{-1}$, $\Delta Q_{\Gamma_k}^{X_1} = 0.0530 \text{ kcal} \cdot \text{mol}^{-1} \text{ fs}^{-1}$, $\Delta Q_{\Gamma_k}^{X_2} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{ fs}^{-1}$ and $\Delta Q_{\Gamma_k}^{X_3} = 0.01875 \text{ kcal} \cdot \text{mol}^{-1} \text{ fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs multiplicative factor (X), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

We now investigate the total mass dependence of the Soret effect for the TMAO molecule. In Figure 6.11, we show the concentration gradient evolution as a function of the temperature for a TMAO solution at different molar masses for both TMAO and water. We rescale the mass of the TMAO and water molecules by the following multiplicative factors: $X_1 = 0.5$, $X_2 = 1$ and $X_3 = 4$. We observe that the Soret coefficient is little sensitive to a rescaling of both particles by the same factor in the case of binary mixture of TMAO and water. All TMAO concentration profiles for these multiplicative factors exhibit almost the same response under the thermal gradient and the TMAO molecule clearly tends to accumulate in the cold region.

	Molar mass solute ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3}K^{-1}$)
X_1	38	9	4.5 ± 0.7
X_2	75	18	5.0 ± 0.9
X_3	300	36	5.5 ± 1.2

Table 6.4: Mass ratio effect of TMAO. Molar masses of TMAO and water for each multiplicative factor and its Soret coefficient value.

The Soret coefficients of the TMAO molecule represented in the inset graph of Figure 6.11 and in Table 6.4 do not vary significantly as we increase overall mass of the system ($S_T^{X_1} \approx S_T^{X_2} \approx S_T^{X_3}$) as long as the mass ratio between solute and solvent is preserved. This result emphasizes the fact that the Soret coefficient of a solute depends on both its intrinsic properties and the intrinsic properties of its solvent.

6.5.2 Urea molecule

Furthermore, we have studied the mass effect of the urea molecule. We want to observe whether the trends obtained for TMAO are the same as for urea, a smaller and fully polar solute.

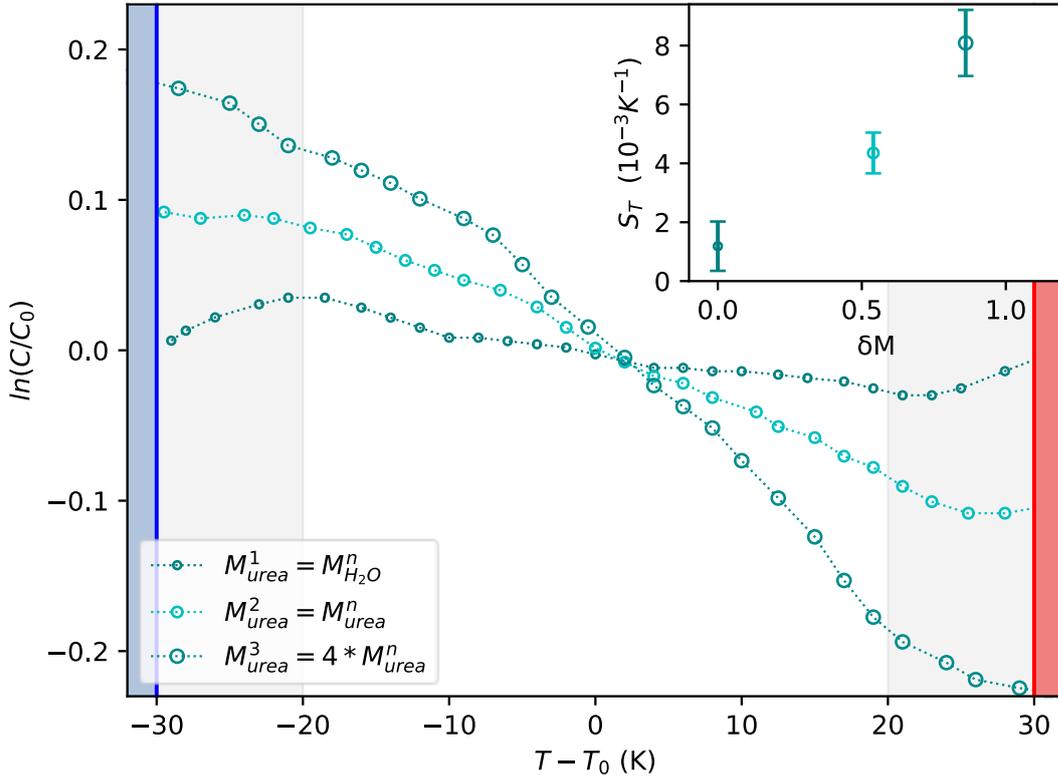


Figure 6.12: Mass effect of urea. $M_{urea}^1 = M_{water}^n$, $M_{urea}^2 = M_{urea}^n$ and $M_{urea}^3 = 4 \cdot M_{urea}^n$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative mass difference (δM), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

We now investigate the mass ratio dependence of the Soret effect for the urea molecule. In Figure 6.12, we show the concentration gradient evolution as a function of the temperature for a urea solution at different masses. The rescaled molar masses for urea are represented; $M_{urea}^1 = 18 \text{ g} \cdot \text{mol}^{-1}$, $M_{urea}^2 = 60.024 \text{ g} \cdot \text{mol}^{-1}$ and $M_{urea}^3 = 240.096 \text{ g} \cdot \text{mol}^{-1}$. We observe a progressive increment of the concentration in the cold side as we increase the mass of the urea, which is similar to TMAO in Figure 6.10. The Soret coefficients of urea in the inset

graph highly increase as a function of the mass ratio, which their values are represented in this table.

	Molar mass solute ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3} K^{-1}$)
M_{urea}^1	18	18	1.2 ± 0.8
M_{urea}^2	60	18	4.4 ± 0.7
M_{urea}^3	240	18	8.1 ± 1.1

Table 6.5: Mass effect of urea. Molar masses of urea and water and its Soret coefficient value

In fact, we notice that the Soret coefficient value increases as a function of the mass ratio for the urea molecule, and is higher than that observed for the TMAO molecule. When solutes have the mass of water ($M_{water}^n = M_{TMAO}^1 = M_{urea}^1$), the TMAO molecule has a higher Soret coefficient than the urea molecule ($S_T^{M_{TMAO}^1} > S_T^{M_{urea}^1}$), which is consistent with the previous results, where the Soret coefficient is larger for TMAO rather than urea as we can see by comparing the Soret coefficients at their natural mass. However, when solute masses are four times higher than their natural masses, we observe a rebound in the thermophobicity of urea, exceeding that of TMAO ($S_T^{M_{TMAO}^3} < S_T^{M_{urea}^3}$), even though the mass of TMAO is still higher than the mass of urea ($M_{TMAO}^3 > M_{urea}^3$). This highlights that the mass effect with respect the migration to colder regions depends in itself on the nature of the solute, which means that the empirical factor for mass dependence (a_M) as well as inertial empirical factor (b_I) are then actually defined by the nature of the solute.

We thus proceed to analyse the effect of the mass ratio once both water and urea masses are rescaled by the same factors ($X = 0.5, 1, 4$), as done previously with TMAO molecule (Figure 6.11).

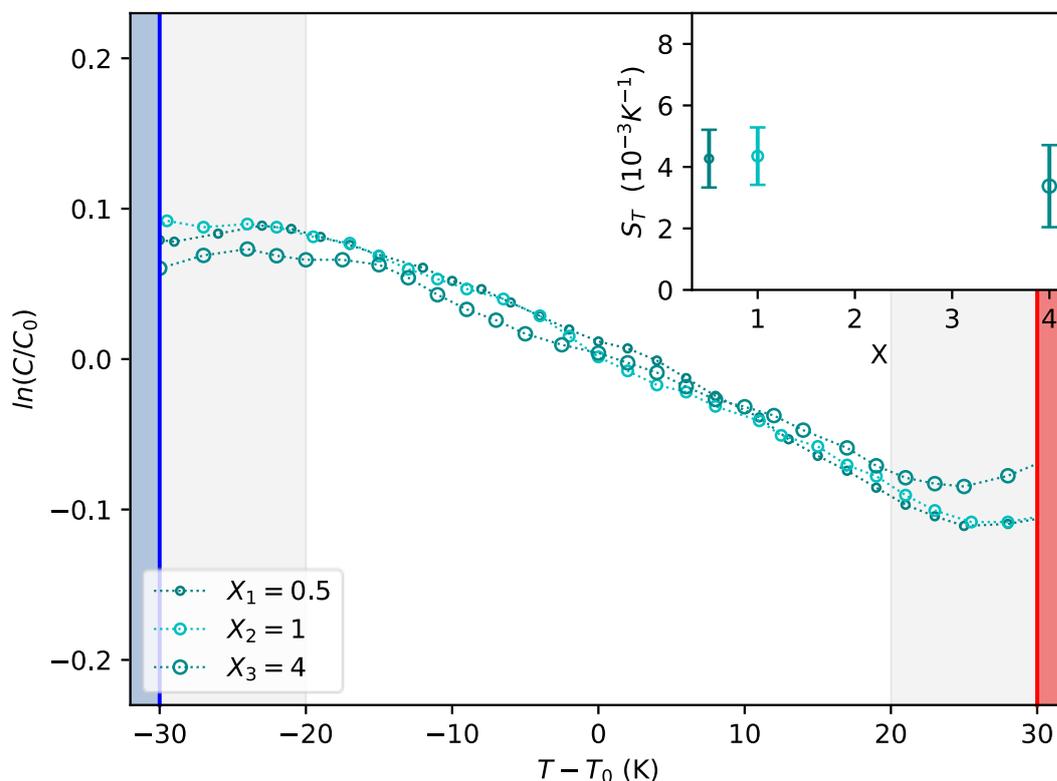


Figure 6.13: Mass ratio effect of urea. multiplicative factors are $X_1 = 0.5$, $X_2 = 1$ and $X_3 = 4$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$, $\Delta Q_{\Gamma_k}^{X_1} = 0.0530 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$, $\Delta Q_{\Gamma_k}^{X_2} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ and $\Delta Q_{\Gamma_k}^{X_3} = 0.01875 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs multiplicative factor (X), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{\text{sample}} = 50$, $N_{\text{trajectories}} = 10$).

We now investigate the total mass dependence of the Soret effect for the urea molecule. In Figure 6.13, we show the concentration gradient evolution as a function of the temperature for a urea solution at different molar masses for both urea and water. We rescale the mass of the urea and water molecules by the following multiplicative factors: $X_1 = 0.5$, $X_2 = 1$ and $X_3 = 4$. We observe the same behaviour for the urea molecule as the one observed for the TMAO molecule, same mass ratios (solute/solvent) exhibit a similar Soret coefficient value, which reinforces the idea that mass effect is due to the relation between solute/solvent masses and not just a pure mass effect of one component.

	Molar mass solute ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3} K^{-1}$)
X_1	30	9	4.3 ± 0.8
X_2	60	18	4.4 ± 0.9
X_3	240	36	3.3 ± 1.7

Table 6.6: Mass ratio effect of urea. Molar masses of urea and water for each multiplicative factor and its Soret coefficient value.

Table 6.6 shows the Soret coefficients for the different multiplicative factors. We observe that the Soret coefficient of X_3 is slightly lower than the other two results. However, the inset graph in Figure 6.13 shows that this Soret coefficient has higher error bars which comprises the other values, so we do not consider as a different Soret coefficient value.

The last part of the urea study is to examine what is the effect of the water mass instead of the urea mass. We noticed that the simulation with a urea mass equal to the mass of water ($M_{urea}^1 = M_{water}^n$) has a very small Soret coefficient value ($S_T^{M_{urea}^1} = 1.2 \cdot 10^{-3} K^{-1}$), which means that the urea molecule with a small mass has a modest thermophobic behaviour. A roughly extrapolation from our results predicts that the urea molecule could reach $S_T^{urea} = 0 K^{-1}$ at $M_{urea} = 10.5 g \cdot mol^{-1}$, but it is not recommended to reduce the mass by that much without constraining the molecule. For this reason, we modify the water masses of the solution instead, as we have already seen that the mass effect on the Soret coefficient depends on the difference in mass between the solute and the solvent and not on the absolute mass. We prepared a set of four systems, where the urea mass is always the mass of water ($M_{urea}^3 = M_{water}^n$), and we rescale the mass of water by these factors (0.5, 1, 2, 4).

We now investigate the water mass dependence of the Soret effect for the urea molecule. In Figure 6.14, we show the concentration gradient evolution as a function of the temperature for a urea solution at different water molar masses: $M_{water}^1 = 9 g \cdot mol^{-1}$, $M_{water}^2 = 18 g \cdot mol^{-1}$, $M_{water}^3 = 36 g \cdot mol^{-1}$ and $M_{water}^4 = 144 g \cdot mol^{-1}$. All these urea concentration profiles exhibit really flat concentration gradients. The urea molecule with the water molar mass ($M_{urea} = 18 g/mol$) does not have a clear trend to accumulate in the cold or the hot

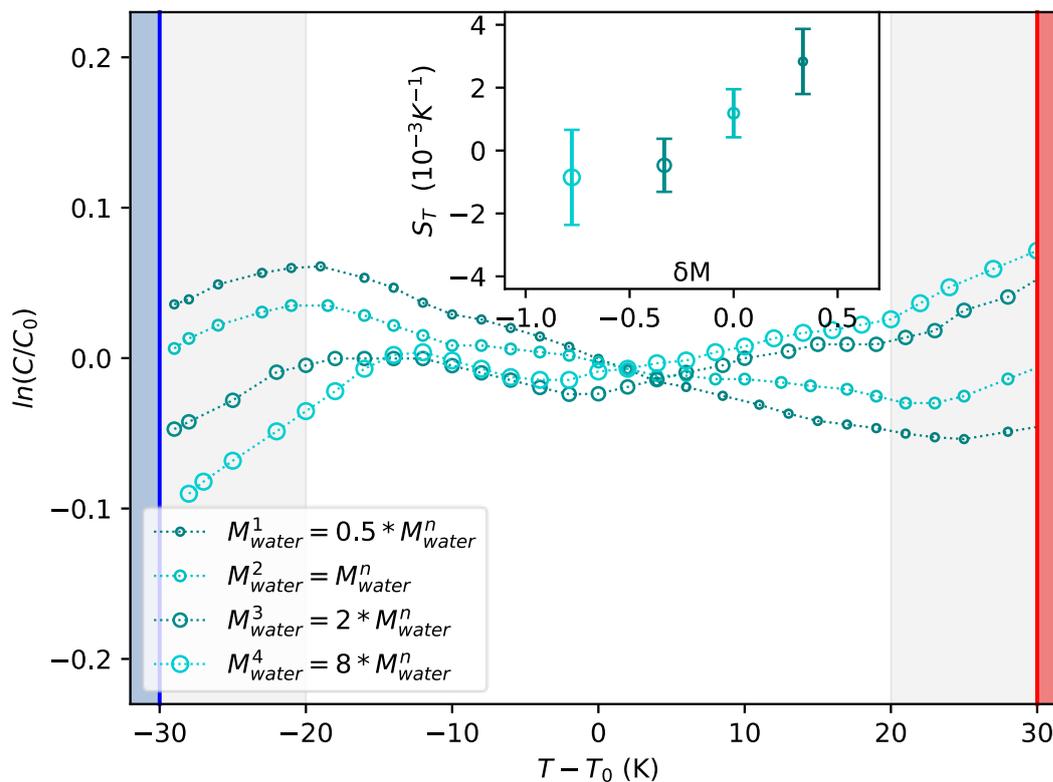


Figure 6.14: Water mass effect of urea. $M_{water}^1 = 0.5 \cdot M_{water}^n$, $M_{water}^2 = M_{water}^n$, $M_{water}^3 = 2 \cdot M_{water}^n$ and $M_{water}^4 = 8 \cdot M_{water}^n$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k}^{M_{water}^1} = 0.0530 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$, $\Delta Q_{\Gamma_k}^{M_{water}^2} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$, $\Delta Q_{\Gamma_k}^{M_{water}^3} = 0.0265 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ and $\Delta Q_{\Gamma_k}^{M_{water}^4} = 0.0133 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative mass difference (δM), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

region. Furthermore, we observe a small decrease in the concentration gradient as the water mass (solvent) increases. It can be barely perceived that the urea molecule shows thermophobic behaviour for water masses equal to or less than $18 \text{ g} \cdot \text{mol}^{-1}$, but the urea molecule becomes thermophilic for water masses higher than $18 \text{ g} \cdot \text{mol}^{-1}$. The Soret coefficients for the urea molecule represented in the inset graph confirm these trends and the values are represented in the following table.

	Molar mass solute ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3}K^{-1}$)
M_{water}^1	18	9	2.8 ± 1.0
M_{water}^2	18	18	1.2 ± 0.7
M_{water}^3	18	36	-0.5 ± 1.0
M_{water}^4	18	144	-0.9 ± 1.9

Table 6.7: Water mass effect of urea. Molar masses of urea and water for each multiplicative factor and its Soret coefficient value.

The Soret coefficient for urea is positive for water masses lower or equal than the isotopic mass of urea (M_{water}^1 and M_{water}^2). However, we observe an inversion of the sign in the Soret coefficient for the urea molecule, once the water mass becomes larger (M_{water}^3 and M_{water}^4) than the isotopic mass of urea, which confirms that urea starts to accumulate in the hot region. These results support the idea that heavier molecules tend to accumulate in the cold region, since they participate more efficiently to generate the area of strong density. Consequently, the water molecules (solvent) for systems M_3^{water} and M_4^{water} become heavier than the urea molecules, and displace the urea molecules towards the hot region. We try to reproduce the negative Soret coefficient value for the TMAO molecule, but we do not achieve the thermophilic behaviour for the TMAO molecule, even if we increase the mass of water eight times. A strong “chemical” contribution and a larger size could explain that the TMAO molecule does not lose the thermophobic behaviour and makes it stay in the colder region, even if the mass of the water (solvent) is larger.

6.5.3 Methanol molecule

In this section, we are going to see in more detail the mass effect for the methanol molecule, which is the solute with the lowest Soret coefficient value among the molecules studied in the section 6.4. First, we studied the effect of mass for the methanol molecule, thus we performed the same simulations as previously performed for the TMAO and urea solutes, see Figure 6.10 and Figure 6.12 respectively.

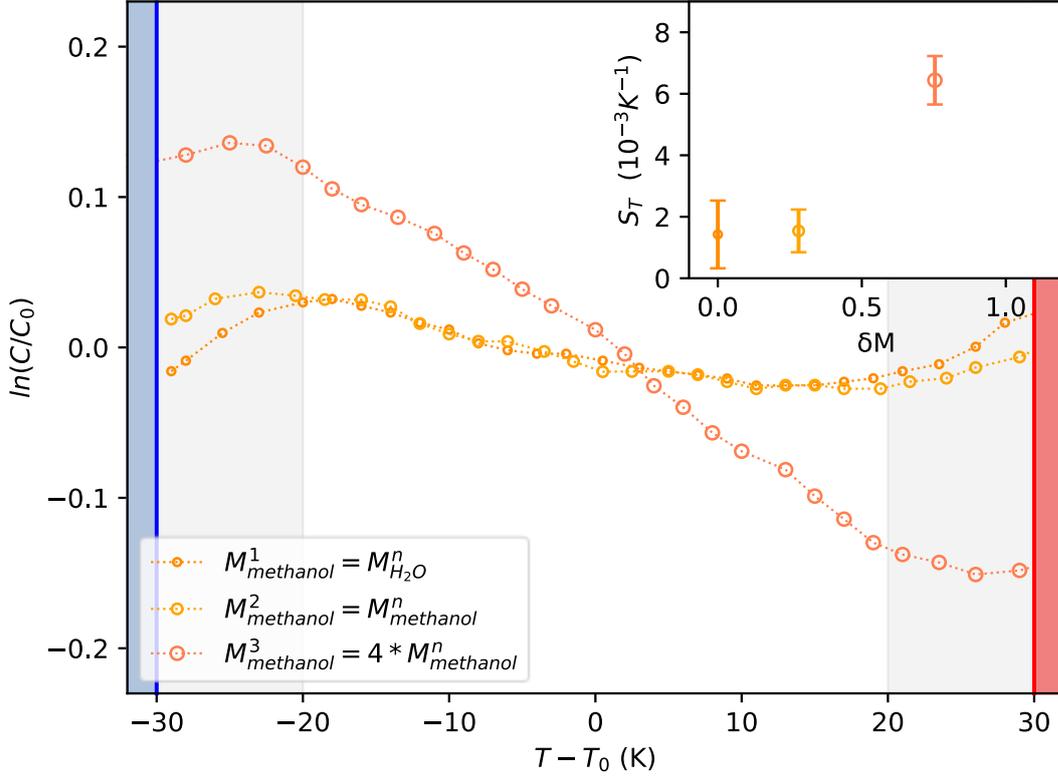


Figure 6.15: Mass effect of methanol. $M_{methanol}^1 = M_{H_2O}^n$, $M_{methanol}^2 = M_{methanol}^n$ and $M_{methanol}^3 = 4 \cdot M_{methanol}^n$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative mass difference (δM), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

We now investigate the mass dependence of the Soret effect for the methanol molecule. In Figure 6.15, we show the concentration gradient evolution as a function of the temperature for a methanol solution at different masses. The rescaled molar masses for methanol are represented: $M_{methanol}^1 = 18 \text{ g} \cdot \text{mol}^{-1}$, $M_{methanol}^2 = 32.04 \text{ g} \cdot \text{mol}^{-1}$ and $M_{methanol}^3 = 128.16 \text{ g} \cdot \text{mol}^{-1}$. The first two methanol concentration profiles ($M_{methanol}^1$ and $M_{methanol}^2$) exhibit almost flat concentration gradients. However, the methanol molecule tends to accumulate in the cold region the third concentration profile ($M_{methanol}^3$), where its mass is significantly higher. Therefore, we always observe a progressive increase of the concentration in the cold region as we increase the mass for any kind of molecule regardless of its nature.

The Soret coefficients of methanol in the inset graph shows a high value for the heaviest methanol system while the other two Soret coefficient values are practically zero. These values are represented in this table.

	Molar mass solute ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3} K^{-1}$)
$M_{methanol}^1$	18	18	1.4 ± 1.1
$M_{methanol}^2$	32	18	1.5 ± 0.7
$M_{methanol}^3$	128	18	6.4 ± 0.8

Table 6.8: Mass effect of methanol. Molar masses of methanol and water and its Soret coefficient value.

One interesting point is that the first two systems exhibit the same value of the Soret coefficient ($S_T^{M_{methanol}^1} = S_T^{M_{methanol}^2}$) despite the fact that their mass difference is almost twice as large ($M_{methanol}^1 < M_{methanol}^2$). On the other hand, we confirm that the mass effect ($M_{methanol}^1 < M_{methanol}^2 \ll M_{methanol}^3$) in the Soret coefficient value for the last simulation ($S_T^{M_{methanol}^1} = S_T^{M_{methanol}^2} \ll S_T^{M_{methanol}^3}$) is similar to the mass effects previously observed for urea and TMAO.

The Soret coefficient evolution as a function of the methanol mass ratio is actually similar to that observed for the urea molecule rather than the TMAO molecule behaviour. We notice that urea and methanol have relative similar molecular size as compared to TMAO ($V_{urea} \approx V_{methanol} < V_{TMAO}$), see Table 6.2. This might suggest that smaller molecules have a larger mass effect than bigger molecules in terms of size. Moreover, the simulations of these solutes at mass of 18 g/mol ($M_{TMAO}^1 = M_{urea}^1 = M_{methanol}^1$) reinforce this idea. Urea and methanol exhibit the same Soret coefficient value but it is not the same for the TMAO molecule ($S_T^{M_{methanol}^1} = S_T^{M_{urea}^1} < S_T^{M_{TMAO}^1}$). However, when the masses of the solutes are four times higher than their natural masses (M^3), we observe the same rebound in the thermophobicity of methanol as for urea, but in this case the Soret coefficient is the same as that of TMAO ($S_T^{M_{TMAO}^3} = S_T^{M_{methanol}^3}$), and the mass of TMAO is still higher than that of methanol ($M_{TMAO}^3 > M_{methanol}^3$). On the other hand, the Soret coefficient of methanol is lower than urea ($S_T^{M_{urea}^3} >$

$S_T^{M_{methanol}^3}$), because urea is significantly heavier ($M_{urea}^3 > M_{methanol}^3$). These results emphasise that the mass effect with respect the migration to colder regions is itself dependent on the nature of the solute and the size effect, which means that the moment of inertia empirical factor dependence (b_I) is defined in somehow by the nature of the solute, and gains importance depending on the size of the molecule.

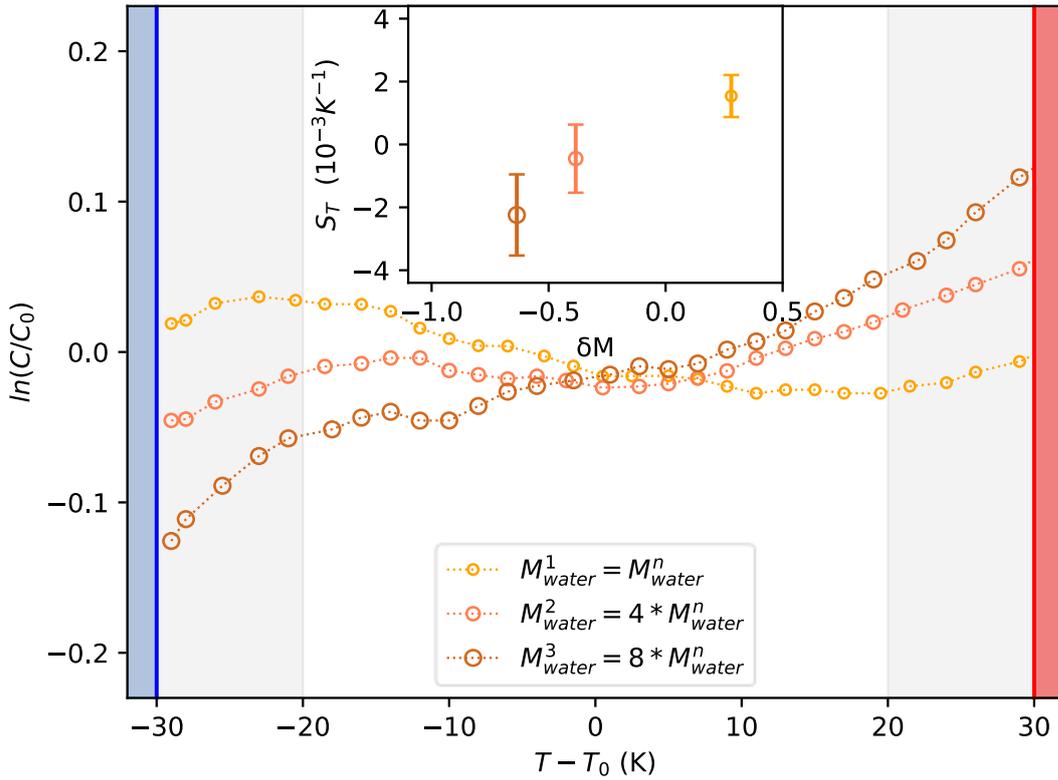


Figure 6.16: Water mass effect of methanol. $M_{water}^1 = 0.5 \cdot M_{water}^n$, $M_{water}^2 = 4 \cdot M_{water}^n$ and $M_{water}^3 = 8 \cdot M_{water}^n$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k}^{M_{water}^1} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$, $\Delta Q_{\Gamma_k}^{M_{water}^2} = 0.0188 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ and $\Delta Q_{\Gamma_k}^{M_{water}^3} = 0.0133 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative mass difference (δM), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

The last part of the methanol study is to examine what is the effect of the water mass instead of the urea mass. We noticed that the simulation with the methanol natural mass ($M_{methanol}^2$)

has a very small Soret coefficient value ($S_T^{M_{methanol}^1} = 1.5 \cdot 10^{-3} K^{-1}$), which means that the methanol molecule has a small thermophobic behaviour. In this case, we modify the water masses of the solution by these factors (1,4, 8).

We now investigate the water mass dependence of the Soret effect for the methanol molecule. In Figure 6.16, we show the concentration gradient evolution as a function of the temperature for a methanol solution at different water molar masses: $M_{water}^1 = 18 \text{ g} \cdot \text{mol}^{-1}$, $M_{water}^2 = 72 \text{ g} \cdot \text{mol}^{-1}$, $M_{water}^3 = 144 \text{ g} \cdot \text{mol}^{-1}$. The methanol concentration profiles show a progressive gradient reversal as the water mass increase. This progressive effect can be observed in the Soret coefficients in the inner graph, and their values are represented in the following table.

	Molar mass solute ($\text{g} \cdot \text{mol}^{-1}$)	Molar mass water ($\text{g} \cdot \text{mol}^{-1}$)	Soret coefficient ($10^{-3} K^{-1}$)
M_{water}^1	32	18	1.5 ± 0.7
M_{water}^2	32	72	-0.5 ± 1.1
M_{water}^3	32	144	-2.2 ± 1.6

Table 6.9: Water mass effect of methanol. Molar masses of methanol and water for each multiplicative factor and its Soret coefficient value.

We observe that the Soret coefficient of methanol reaches a negative value ($S_T^{M_{water}^1} < 0 < S_T^{M_{water}^2} < S_T^{M_{water}^3}$) once the water mass is larger than the mass of methanol ($M_{water}^1 < M_{methanol} < M_{water}^2 < M_{water}^3$). This trend is in agreement with our previous results obtained for urea in Figure 6.14, which exhibit a positive Soret coefficient for systems where the solute mass is higher than solvent mass ($M_{solute} > M_{water}$), and a negative Soret coefficient ($S_T < 0$) for systems where the solute mass is lower than the solvent mass ($M_{solute} < M_{water}$).

	Urea (18 g/mol)		Methanol (32 g/mol)	
	Mass ratio (δM)	Soret coefficient ($10^{-3}K^{-1}$)	Mass ratio (δM)	Soret coefficient ($10^{-3}K^{-1}$)
M_{water}^1	0.33	2.8 ± 1	0.28	1.5 ± 0.8
M_{water}^2	0	1.2 ± 0.7	-0.38	-0.5 ± 0.9
M_{water}^3	-0.33	-0.5 ± 1	-0.63	-2.2 ± 1.7
M_{water}^4	-0.78	-0.9 ± 1.6		

Table 6.10: Urea and methanol mass effect. Mass ratios of the urea and methanol solutions and its respective Soret coefficient value. The errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

Table 6.10 summarizes the Soret coefficient values of urea and methanol of Table 6.7 and Table 6.9 in terms of mass ratios. We observe a fairly good correlation between the Soret coefficients of urea and methanol with a similar mass ratio (the pairs are bolded in blue, red and violet). Although the violet comparison does not show exactly the same Soret coefficient value between urea and methanol, we observe that the Soret coefficients of systems with heavy water mass have larger errors. The following graph represents the results of the Table 6.10.

In Figure 6.17, we show the Soret coefficients for urea and methanol at different mass ratios. We observe that close mass ratios exhibit similar Soret coefficient values. Moreover, we observe that both Soret coefficient values increase in parallel as the mass ratio increases.

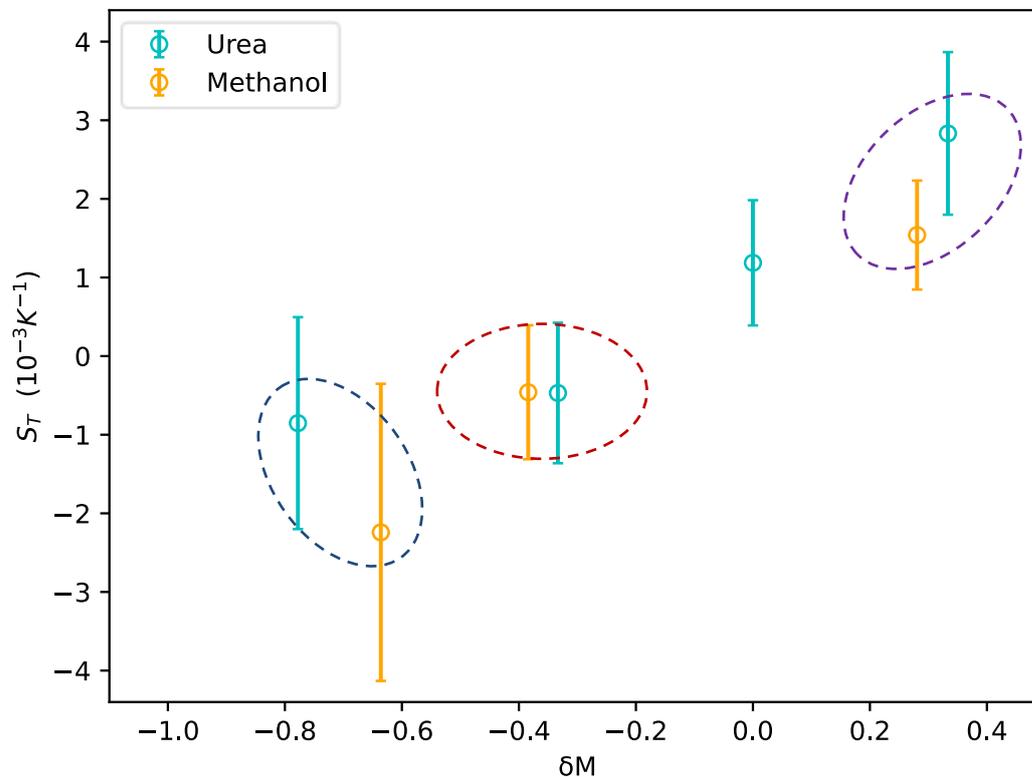


Figure 6.17: Urea and methanol mass effect. Urea (cyan) and methanol (magenta) solutions. Data was averaged over 20 independent trajectories. $P = 1$ bar. $L_z = 50 \text{ \AA}$. $T_0 = 330$ K. The errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$). The Soret coefficient of urea are the inset graph values from Figure 6.14. The Soret coefficient of methanol are the inset graph values from Figure 6.16.

6.5.4 Water molecule

The last aspect of the mass effect on the Soret coefficient is the analysis of a pure water system. Experimental works have already demonstrated a mass effect in a binary mixtures of water with deuterated water.^{95,106,107} We find that the best way to see if there is indeed an isotopic effect on the Soret coefficient is to perform simulations with the same substance at different masses where the "chemical" contribution term and the size of the Soret coefficient become irrelevant. Moreover, the water molecule is a small molecular system with almost all of its mass centred on one atom (oxygen), so the moment of inertia should not have a large impact on the Soret coefficient compared to the effect of mass.

In this case, we modify the mass of a small number of water molecules to preserve the idea of dilute aqueous solutions for a large range of masses. In addition, we exclude the isotopic water molecules from the eHEX algorithm, which allows us to have a better control of the temperature gradient, since it depends on the amount of energy exchanged between the two regions and the mass of the molecules that are affected. We validated each step of the simulation for this pure water systems, using a similar procedure as the previous solute systems.

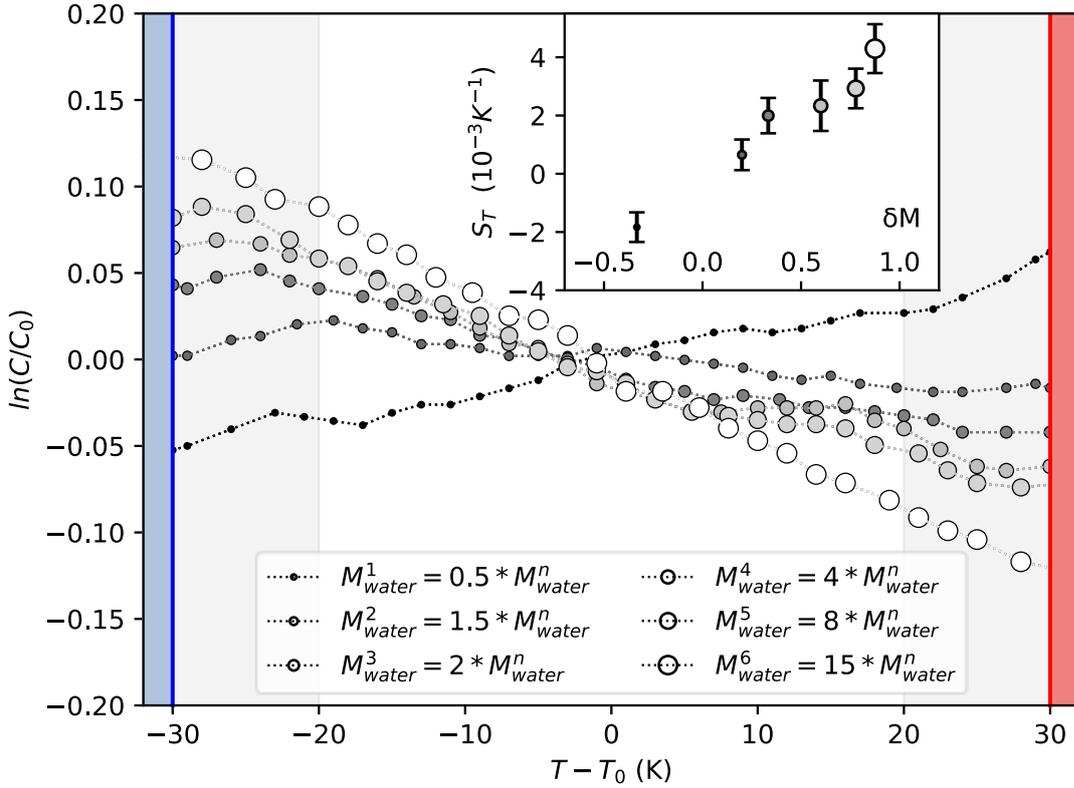


Figure 6.18: Mass effect of isotopic water. Isotopic water at different molar masses; $M_{water}^{iso} = X \cdot M_{water}$ and $X = 0.5, 1, 1.5, 2, 4, 8, 15$. Data was averaged over 20 independent trajectories. The energy flux was set to $\Delta Q_{\Gamma_k} = 0.0375 \text{ kcal} \cdot \text{mol}^{-1} \text{fs}^{-1}$ in order to generate a thermal gradient amplitude of $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red bars represent the reservoirs where heat is pumped and injected, respectively. Inset graph: Soret coefficient (S_T) vs relative mass difference (δM), the Soret coefficient values are calculated using the data outside the grey regions, and errors are calculated by the bootstrap method ($N_{sample} = 50$, $N_{trajectories} = 10$).

We confirmed that the total energy does not vary for all steps of the simulation. The temperature gradient is constant with an amplitude of 60 K. These simulations are performed at a median temperature of 330 K. Finally, the system consists of 40 isotopic water molecules

and 1024 water solvent molecules, which maintain the same number of particles as the previous solute studies.

We now investigate the mass dependence of the Soret effect for the water molecule. In Figure 6.18, we show the concentration gradient evolution as a function of the temperature for a water system at different isotopic water molecules. The mass has been rescaled for a certain number (40 water molecules) of isotopic waters ($M_{water}^{iso} = X \cdot M_{water}$) by the following factors $X = 0.5, 1.5, 2, 4, 8, 15$. A progressive increase in the concentration gradient towards the cold region can be observed as the mass of the isotopic waters increases. Moreover, the Soret coefficients, represented in the inner graph, show a gradual growth as a function of the mass ratio between the isotopic waters and natural ones. This result confirms that the mass effect is an additive factor in the global Soret coefficient. Furthermore, the isotopic water with its water mass reduced by the half (M_{water}^1) exhibit a thermophilic behaviour instead the rest of systems, where their isotopic water molecules are higher than the natural water mass.

	Molar mass isotopic water ($g \cdot mol^{-1}$)	Molar mass water ($g \cdot mol^{-1}$)	Soret coefficient ($10^{-3} K^{-1}$)
M_{water}^1	9	18	-1.8 ± 0.5
M_{water}^2	27	18	0.7 ± 0.5
M_{water}^3	36	18	2 ± 0.6
M_{water}^4	72	18	2.3 ± 0.9
M_{water}^5	144	18	2.9 ± 0.7
M_{water}^6	270	18	4.3 ± 0.8

Table 6.11: Mass effect of isotopic water. Molar masses of the isotopic water (solute) and water (solvent) molecules and its Soret coefficient value.

Table 6.11 shows the molar mass of each isotopic water and its resulting Soret coefficient value. We confirm that a higher mass ratio leads to higher thermophobic behaviour. Moreover, the inversely proportional mass ratios (M_{water}^1 and M_{water}^2) also have the same Soret coefficient value but have opposite sign ($S_T^{M_{water}^1} \approx -S_T^{M_{water}^2}$), which confirms that the Soret

coefficient for “chemically” identical particles depends only on the difference in mass between the solute and the solvent, therefore, the thermophobic or thermophilic behaviour of the particle is simply a mass effect.

6.5.5 Mass effect conclusion

In this section we have studied the effect of mass on the Soret coefficient for different solutes. We have been modifying the mass of the solute, the solvent and both in different simulations.

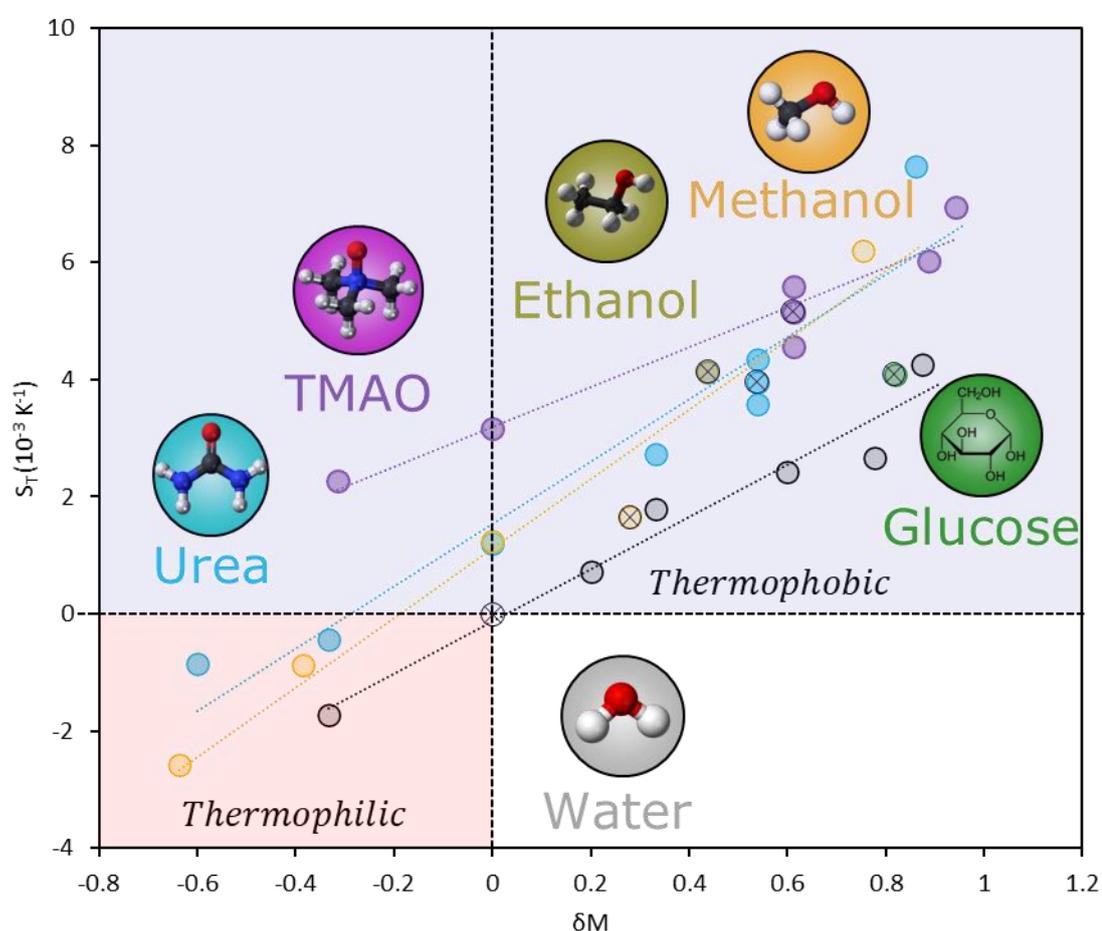


Figure 6.19 : Mass effect summary. Methanol (yellow), ethanol (olive), urea (cyan), TMAO (magenta), glucose (green), and water (grey). Data was averaged over 20 independent trajectories. $\Delta T \approx 60 \text{ K}$. $P = 1 \text{ bar}$. $L_z = 50 \text{ \AA}$. $T_0 = 330 \text{ K}$. $C_0 = 2.17 \text{ mol} \cdot \text{kg}^{-1}$. The blue and red zones represent thermophobic and thermophilic regions. The cross dots are the simulations at natural mass for both solute and solvent. The dashed lines are linear regressions of the equation $S_T = S_T^0 + a_M \delta M$.

In Figure 6.19, we show all the Soret coefficients obtained above for methanol (yellow), ethanol (olive), urea (cyan), TMAO (magenta), glucose (green), and isotopic water (grey). The dots with crosses represent the simulations when both the solute (M_{solute}^n) and the solvent (M_{water}^n) have their natural mass. The three dots at the same mass ratio (δM) represent the TMAO and urea molecules in the simulations in which the mass has been rescaled for both molecules in the simulation (solute and water). The blue part shows the thermophobic ($S_T > 0 \text{ K}^{-1}$) behaviour for the solutes in certain simulations, and the red part shows the thermophilic ($S_T < 0 \text{ K}^{-1}$) behaviour. We observe that all molecules show a thermophobic character for positive mass ratios. On the other hand, urea, methanol and water acquire a thermophilic character at certain negative mass ratios.

Even if the solutes studied are larger than the water molecule (solvent), we should expect a moment of inertia effect in the isotopic Soret coefficient. However, the Soret coefficients evolution for the methanol, urea, TMAO solutes show a constant increase as a function of the mass ratio ($S_t \propto \delta M$). Therefore, we have employed the description of the Soret coefficient:

$$S_T = S_T^0 + a_M \frac{M_2 - M_1}{M_2 + M_1}, \quad (6.16)$$

where the isotopic moments of inertia effects are neglected and included in the constant Soret coefficient term $S_T^0 = S_T^{ch} + b_I \delta I$.

	S_T^0 (10^{-3} K^{-1})	a_M	Coefficient of determination (R^2)
Water	0.0	4.5	0.96
Methanol	1.1	5.9	0.97
Urea	1.5	5.3	0.92
TMAO	3.2	3.4	0.94

Table 6.12: Empirical fit mass effect. Constant values of the linear fit of $S_T = S_T^0 + a_M \delta M$ from Figure 6.19 for the water, methanol, urea and TMAO molecules.

The coefficients of the linear regression of equation (6.16) are shown in Table 6.12 for water, methanol, urea and TMAO. It should be noted that $S_T^0 \approx 0 \text{ K}^{-1}$ for the isotopic water simulations. This value can be explained by the fact that these simulations are pure water solutions in which we only modify the mass of certain water molecules, and therefore we just expect a pure mass effect in the Soret coefficient. These simulations are isotopic water in water, so we do not have a chemical contribution of the Soret coefficient ($S_T^{ch} = 0$), and the moment of inertia contribution ($b_I \delta I = 0$) should not affect the Soret coefficient for two reasons. Previously, we discussed the fact that the moment of inertia contribution of the Soret coefficient for the spherical particles of the binary mixture with the same mass ($S_t^{\delta M=0}$) can be understood in terms of molecular volume ($\delta I = b_I^* \delta V^{(2/3)}$), which would be $\delta I = 0$ for the same molecule. Moreover, the water barycentre is mostly centred on the oxygen atom so the effect of the moment of inertia should be very small as the mass increases.

We note that S_T^0 is larger for the TMAO molecule as compared to urea or methanol, which indicate a higher “chemical” contribution or moment of inertial effect of the Soret coefficient. However, a more extensive study should be done to know the moment of inertia effect to isolate the “chemical” contribution.

Finally, we observe that all coefficients of determination are higher than 0.9 for all solutes, which indicates a trend of the Soret coefficient as a function of mass ratio and reinforces the approximation made in equation (6.16).

7. Molecular origins of thermophoresis

In the previous chapter, we discussed some intrinsic properties of the solutes, as well as some external factors that affect the thermophoretic motion of these solutes under a thermal gradient and the Soret coefficient value. In addition, we analysed several phenomenological descriptions suggested in the literature that describe particular thermophoretic trends for certain systems, but they do not reveal the molecular origin of thermophoresis and the Soret coefficient.

On the other hand, some molecular models have been proposed to explain the molecular basis of thermophoresis. In this section, we will test the feasibility of two plausible models, which are of particular interest and relevance for molecular systems, but which come from different perspectives. The first model is the Eastman's approach,⁷¹ which tries to relate the thermally-driven effects on a molecule to a temperature dependence of its solvation free-energy. The second model is the Prigogine's model which conjectures that the thermophoretic motion of a molecule is related to the difference in the activation energies for the diffusion of the solute and solvent molecules.⁷²⁻⁷⁶

7.1 Thermophoresis as a result of temperature dependence solvation free energy

Eastman defined the thermophoretic motion of a particle as a thermodynamic effect. In a system at a non-constant temperature, the movement of a particle from one region to another is associated with a change in entropy. The author described this entropic change as a rearrangement of particles when a particle moves away and leaves an empty space and the energy required to heat or cool that particle in the new medium which is at a different temperature. Therefore, Eastman justified that the Soret coefficient can be understood in terms of equilibrium thermodynamic considerations, relating the Soret coefficient to the temperature-dependence of the solute hydration free-energy. The hydration free-energy (ΔG^S) is the difference in free energy between a molecule in the gas phase and the same molecule in a liquid phase, which is surrounded by solvent.

The process of solvation describes the rearrangement of the molecules in a system when a new molecule is added. Normally, water is the most commonly used solvent in chemistry, and hence the solvation process particular to aqueous systems is commonly known as hydration, which is the case here. Depending on the properties of the solvent and the solute, their interactions are relatively thermodynamically stable. Experimental measurements are usually carried out under constant temperature and pressure conditions, and hence the thermodynamic property in which the solvation process will be described is the Gibbs free energy (symbol G).

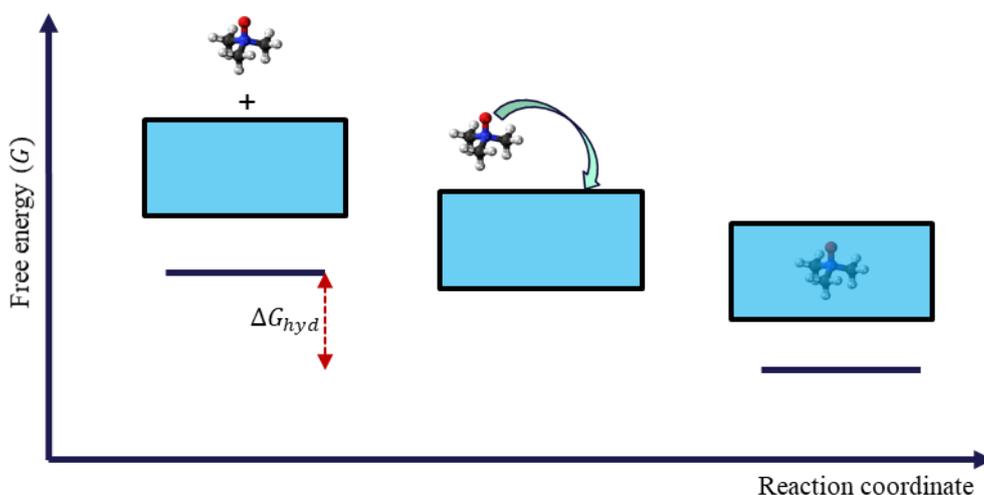


Figure 7.1: (Replot) Hydration free energy. Schematic representation of the hydration free energy process of the trimethyl ammine N-oxide molecule (TMAO). The blue box represents the liquid phase and outer box the gas phase. ΔG_{hyd} is the hydration free energy of the process.

Figure 7.1 is the same as Figure 3.1. We replot this figure to visualize the solvation process. The hydration free energy ($\Delta G^s = \Delta G_{hyd}$) is a difference between two states. First, a solute molecule (TMAO in Figure 7.1) is outside the system, and this solute molecule does not interact with the rest of the molecules in the system, this state is commonly referred as the gas phase of the solute. The last step of the reaction coordinate represents the solute completely solvated in the solution, which is known as the liquid phase. The difference in energy between the solute in the gas phase and the solute in the liquid phase is the hydration free energy (red arrow), which takes into account the solute-solvent and solute-solute interactions as well as the rearrangement of the solution due to solute insertion. Furthermore, the hydration free energy sign will determine whether the process would be favourable (exergonic $\Delta G < 0$) or unfavourable (endergonic $\Delta G > 0$). The same process can be carried out in the other direction,

since the free energy is a thermodynamic state, and does not depend on the path but on the two final states. The inverse process is known as the dehydration process, which is equivalent to the negative free energy value from the hydration process ($\Delta G_{hyd} = -\Delta G_{dehyd}$).

Molecular dynamics can evaluate the hydration free energy process by just decoupling the solute interactions from the solution. The free energy perturbation (FEP) method is one technique that was introduced by Zwanzig in 1954.¹⁷⁸ According to this method, the free energy dehydration difference is obtained from the following equation:

$$\Delta G_{dehyd}(l \rightarrow l^*) = G_{l^*} - G_l = -k_B T \ln \left\langle \exp \left(-\frac{E_{l^*} - E_l}{RT} \right) \right\rangle_l, \quad (7.1)$$

where T is the temperature, k_B is the Boltzmann constant, the triangular brackets denote an average over a simulation run for the liquid state transition ($l \rightarrow l^*$), E_l is the total energy of the system in the liquid phase, and E_{l^*} is the total energy of the decoupled system in the liquid phase. In practice, a normal simulation is run for the liquid state (l), but the energy of the decoupled system in the liquid state (l^*) is also calculated. Moreover, the free energy perturbation calculations only converge properly as the difference between two states is small enough; therefore, it is necessary to divide a perturbation in smaller "windows", which are computed independently. One technique consists of progressively modifying the intermolecular interactions between our target solute molecule and the other molecules in the solution. The "windows" are discretized by a decoupling factor or λ , which varies between one and zero. In the case that λ is equal to one, the intermolecular interactions (Lennard-Jones or Coulombic) between the solute molecule and the rest of the molecules are full active, but once λ is equal to zero, the intermolecular interactions are deactivated. Therefore, we can thus monitor the level of decoupling by simply controlling the value of λ , and the hydration ($\lambda_{0 \rightarrow 1}$) or dehydration ($\lambda_{1 \rightarrow 0}$) processes can be studied by simply modifying this parameter (λ).

There are several methods in order to quantify the free energy differences between two states, apart from the free energy perturbation method (FEP).¹⁷⁸ The other methods used in this work are; the Bennet's acceptance ratio method (BAR),¹⁷⁹ the thermodynamic integration method (TI),¹²¹ the Multistate Bennett Acceptance Ratio method (MBAR).¹⁸⁰ Most of them compare free energy differences between two states, except the MBAR method, which calculates the

(relative) free energies of several states. It essentially reduces to the BAR method when only two states are involved. Henceforth, the total free energy dehydration ($\Delta_1^0 G$) for all these methods can be calculated by just adding these free energy differences for each λ :

$$\Delta_1^0 G = \sum_{i=0}^{n-1} \Delta_{\lambda_i}^{\lambda_{i+1}} G . \quad (7.2)$$

7.1.1 Simulation details

In this work, the desolvation process ($\Delta_1^0 G^s$) is studied for different molecular systems to determine the free energy of dehydration at several temperatures and thus check the viability of viewing the thermophoresis as a thermodynamic process. In addition, the solvation process has also been tested with results similar to those of the desolvation process. The Coulombic cut-off is put at 8.5 Å with the particle-particle particle-mesh (PPPM) solver¹⁶⁰ for the long-range electrostatic force, and the Lennard-Jones cut-off is put at 9 Å, the same as the thermophoretic conditions. For the free energy calculations, a cubic box of size $L_x = L_y = L_z \approx 33.15$ Å, centred at (0,0,0), with equiangular and perpendicular distribution (90°,90°,90°) was considered. The solvent molecules are randomly inserted into the simulation box using the Packmol software¹⁵⁹ without any spatial constraint. The system is then minimized using a Polak-Ribiere¹⁶¹ version of the conjugate gradient algorithm (energy tolerance limit = 10^{-4} kcal · mol⁻¹ and force tolerance limit = 10^{-6} kcal · mol⁻¹). The velocities are set to correspond to the target median temperature, and after a NPT equilibration of 200 ps is performed at this median temperature, with the Nose-Hoover¹⁶²⁻¹⁶⁴ barostat and thermostat. The damping parameters for the barostat and the thermostat are $P_{damp} = 1000$ fs and $T_{damp} = 100$ fs respectively. All simulations are performed under the NPT ensemble at 1 atm of pressure, which is vital for the determination of the free Gibbs energy, because this ensemble controls the temperature and pressure of the system.

Moreover, we use a modified version of the Lennard-Jones and Coulombic potentials, which are known as the soft core potentials.¹⁸¹ The soft-core potentials slightly modify the standard potentials in order to avoid the asymptotic functions during the decoupling process. These potentials have a soft repulsive core, tunable by the parameter λ , in order to avoid singularities

during free energy calculations when sites are created or annihilated. When λ tends to 0, the pair interaction vanishes with a soft repulsive core. When λ tends to 1, the pair interaction approaches the normal, non-soft potential.

The next step is to progressively decrease these Coulombic and Lennard-Jones interactions along 40 steps. First, we switch off Coulombic interactions, and we decrease the Coulombic part (λ^{Coul}) over 20 free energy steps ($d\lambda^{Coul} = -0.05$). Once the Coulombic term is completely decoupled, we proceed to decrease the Lennard-Jones potential in the same way as the Coulombic, the Lennard-Jones part (λ^{LJ}) decreases over 20 free energy steps ($d\lambda^{LJ} = -0.05$). Moreover, we impose that the new λ_{i+1} starts with the previous λ_i configuration, and we always perform an equilibration process of 200 ps and the following production of 200 ps for each λ . We first switch off the Coulombic interactions to avoid pure attractive or repulsive interactions due to the electrostatic forces, and thus prevent the system from collapsing. As long as the Lennard-Jones forces are active, the particles will maintain the minimum energy distance.

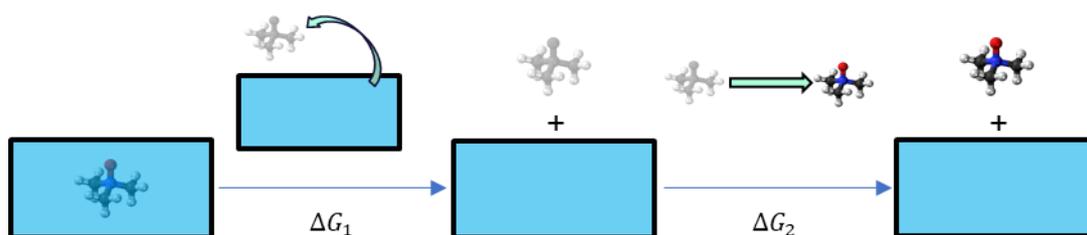


Figure 7.2: Schematic representation of our desolvation process used in LAMMPS. ΔG_1 is the free energy corresponding to the desolvation process of the target molecule represented with a TMAO molecule, the TMAO molecule with greyish colours represents that none of its atoms have charges. ΔG_2 is the same process as ΔG_1 , but the target molecule is isolated. The sum of both energies results in the free energy of dehydration ($\Delta G_{dehyd} = \Delta G_1 + \Delta G_2$).

The LAMMPS software (version 07Aug19) is used with the FEP-package (Agilio Padua). One of the problems we face in this package is in the implementation when modifying the Coulombic interactions. The FEP-package decouple the Coulombic interactions by directly rescaling the charges of the target molecule, instead of decoupling the Coulombic pairs with the other molecules of the solute. Consequently, the Coulombic intramolecular interactions of the target molecule also progressively vanish with this technique, which makes all atoms of the

target molecule completely uncharged at the end of the decoupling process, see ΔG_1 in Figure 7.2.

Therefore, we have performed a second free energy calculation with the same molecule in the gas phase to quantify the corresponding free energy associated with the intramolecular Coulombic, which is represented as ΔG_2 in Figure 7.2. These simulations are performed in the NVT ensemble to keep the same size of the system as the liquid phase, as we are now working just with the target molecule.

Additional simulations are performed with the GROMACS package (version 2019.4) to support the LAMMPS results. The Coulombic and Lennard-Jones cut-offs are 9 Å, the particle-mesh Ewald (PME) solver¹⁸² is used for the long-range electrostatics. The same cubic box system is used and 50000 steps are set for the minimisation process. The velocities are set to correspond to the target mean temperature. A 200 ps NPT equilibration is performed at the respective median temperatures, with the Nose-Hoover thermostat,^{162–164} and the Parrinello-Rahman barostat.¹⁸³ The damping parameters for barostat and thermostat are set at 2 ps, and the barostat compressibility is set at $4.46 \times 10^{-5} \text{ bar}^{-1}$. All simulations are performed under the NPT ensemble at 1 bar of pressure. The Coulombic and Lennard-Jones interactions are decoupled over 14 steps. First, we decouple the Coulombic interactions, decreasing λ^{Coul} over 4 free energy steps ($d\lambda^{Coul} = -0.25$). Once the Coulombic term is completely decoupled, we proceed to decrease the Lennard-Jones potential in the same way as the Coulombic, where λ^{LJ} decrease over 10 steps ($d\lambda^{LJ} = -0.1$). Moreover, we impose that the new λ_{i+1} starts with the previous λ_i configuration, and we always perform an equilibration process of 200 ps and the following production of 200 ps for each λ , which is the same procedure as the LAMMPS free energy calculations. The implementation of the GROMACS free energy directly decouples the Coulombic pair interactions instead of modifying the charges, therefore the second step depicted in Figure 7.2 is not necessary.

The subsequent free energy analysis has been performed by the Alchemical Analysis tool,¹⁸⁴ which is compatible with the GROMACS software, and this tool handles, the analysis through a number of free energy methods, including BAR, MBAR and TI. However, LAMMPS free energy calculations are not included in this tool. First, they have been determined by the tool available in the USER FEP-package of the LAMMPS software. In addition, we have re-adapted

the LAMMPS inputs to be able to calculate the BAR and MBAR free energy methods, and we have modified the resulting free energy outputs to be readable by the Alchemical Analysis tool.

7.1.2 Results

First, we check the robustness of the free energy desolvation process for a TMAO molecule in a 2.17 m aqueous solution. We plot the evolution of the free energy as a function of the inverse of the temperature for both programs (LAMMPS and GROMACS) and several subsequent analysis techniques (TI, FEP, BAR, MBAR). The TI method is only analysed for the GROMACS simulations by the Alchemical Analysis, and the FEP method is only analysed for the LAMMPS simulations by the python script FEP-package available.

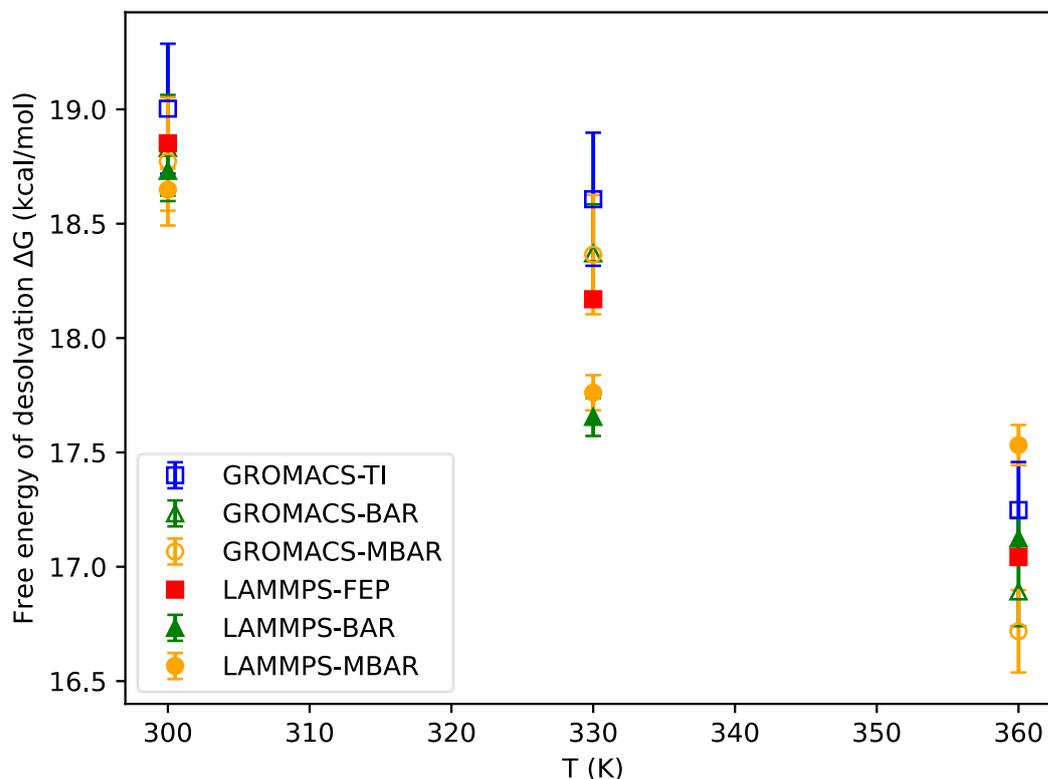


Figure 7.3: Free energy methods. Free energy desolvation process vs temperature for TMAO with different software and several post-analysis methods (TI (empty blue), BAR (empty green) and MBAR (empty orange) for GROMACS, and FEP (fill red), BAR (fill green) and MBAR (fill orange) for LAMMPS. Range of temperatures between 290-360 K. $P = 1$ bar. TMAO molal concentration of 2.17 m.

The desolvation free energies plotted in Figure 7.3 show that the trends are well reproduced by the two programs (GROMACS and LAMMPS) as well as among all the methods (TI, FEP, BAR and MBAR). However, we note that the calculations of the free energies for a temperature of 330 K do not fully coincide, but we observe that these differences are lower than 1 kcal/mol. One of the reasons could be due to the different implementations between LAMMPS and GROMACS, so that they do not reproduce exactly the same conditions. We have tried to reproduce as well as possible the same conditions but we have found some particular limitations for each programme. A clear example is how the Coulombic intermolecular interactions are managed in the free energy calculation for each programme. On the other hand, the long range electrostatic interactions are calculated with the particle-mesh Ewald technique (PME)¹⁸² for the GROMACS simulations, and the particle-particle particle-mesh solver (PPPM)¹⁶⁰ for the LAMMPS simulations. The GROMACS simulations are performed with the leap-frog integrator,¹⁵⁴ and the LAMMPS simulations are performed with the velocity Verlet algorithm.¹⁸⁵ The GROMACS NPT simulations are performed with the Nose-Hoover¹⁶²⁻¹⁶⁴ thermostat, and the Parrinello-Rahman¹⁸³ barostat, but the LAMMPS NPT simulations are performed, with the Nose-Hoover¹⁶²⁻¹⁶⁴ barostat and thermostat. All these differences make comparisons between the two programmes more complex and could explain these differences in energy.

Furthermore, we observe a clear trend as a function of temperature in Figure 7.3. The free energy of desolvation increases as the temperature decreases. These results are in agreement with the solvation energies of aqueous solutions with polar molecules, for example the solvation of methanol in water.¹⁸⁶ Lower temperatures make the TMAO molecule more stable in the aqueous solution and thus require a higher energy to desolvate. However, as the temperature increases, the intermolecular interactions between the TMAO molecule and the solution are weaker, and the free energy required to desolvate a TMAO molecule is thus lower.

Eastman's approximation⁷¹ predicted that the Soret coefficient can be estimated by calculating free energies as a function of temperature. Later, the Braun group^{72,73} developed this idea and proposed the following expression:

$$S_T = \frac{1}{RT} \frac{dG^s}{dT}. \quad (7.3)$$

Even if thermophoresis is a non-equilibrium effect, the Braun group demonstrated that for moderate temperature gradients, the spatial distribution of molecules follow local equilibrium thermodynamics.⁸ Our next step is to determine the Soret coefficients for the desolvation free energies in Figure 7.3. by using the Braun's equation described above, see eq. (7.3). The following graph represents the Soret coefficient values for each method shown in Figure 7.3, and we also include the TMAO Soret coefficient value obtained by the thermophoretic simulations. Moreover, all these Soret coefficients are represented at a median temperature of 330 K.

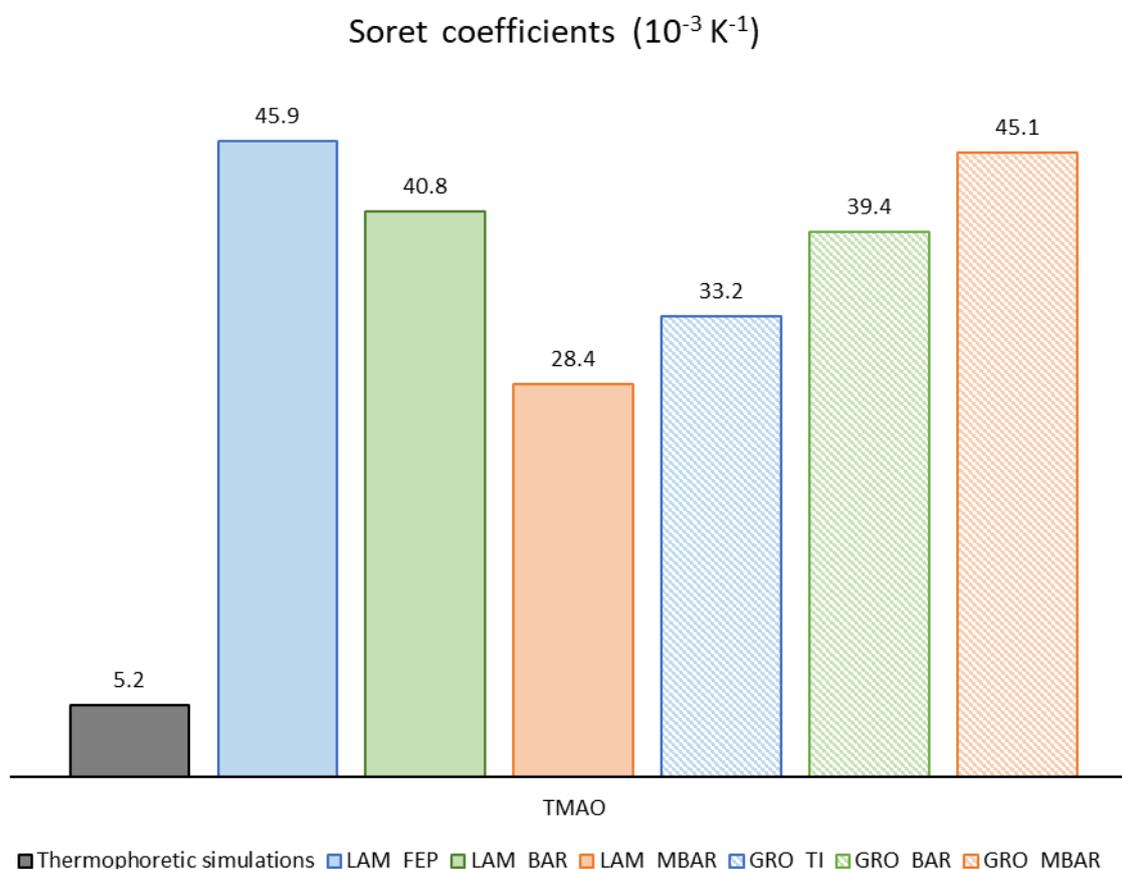


Figure 7.4: Soret coefficient methods. Soret coefficient values at 330 K using several methods (FEP, TI, BAR and MBAR). The LAMMPS results are represented by solid blocs and the GROMACS results are represented by weaved blocks. The methods are FEP or TI (blue), BAR (green) and MBAR (orange). The black column represents the Soret coefficient determined by the thermophoretic calculations.

The TMAO Soret coefficient represented in Figure 7.4 shows the values for each free energy methods. We do not perceive that any post analysis method gives a completely different result

compared to the other methods. However, the Soret coefficients for the different methods are quite disparate and without any clear trend among them. For example, the MBAR method has the lowest Soret coefficient for the LAMMPS simulation, but the highest for the GROMACS simulation. To improve these results, more statistics should be considered for each solvation free energy calculations at a determined temperature and include more temperatures to improve the description of (dG/dT) . We do not consider it necessary to modify the λ decoupling process by more than 20 steps, neither to increase the equilibration or production time for each λ . For this level of accuracy, we observe that the Soret coefficients calculated by the desolvation free energy as a function of the temperature are between $28 - 46 \cdot 10^{-3} \cdot \text{K}^{-1}$, which are one order of magnitude larger than thermophoretic one ($S_T = 5 \cdot 10^{-3} \cdot \text{K}^{-1}$). In addition, we have studied the solvation free energy for urea and methanol. The following graph represents the solvation free energy as a function of the inverse of temperature for TMAO, urea and methanol.

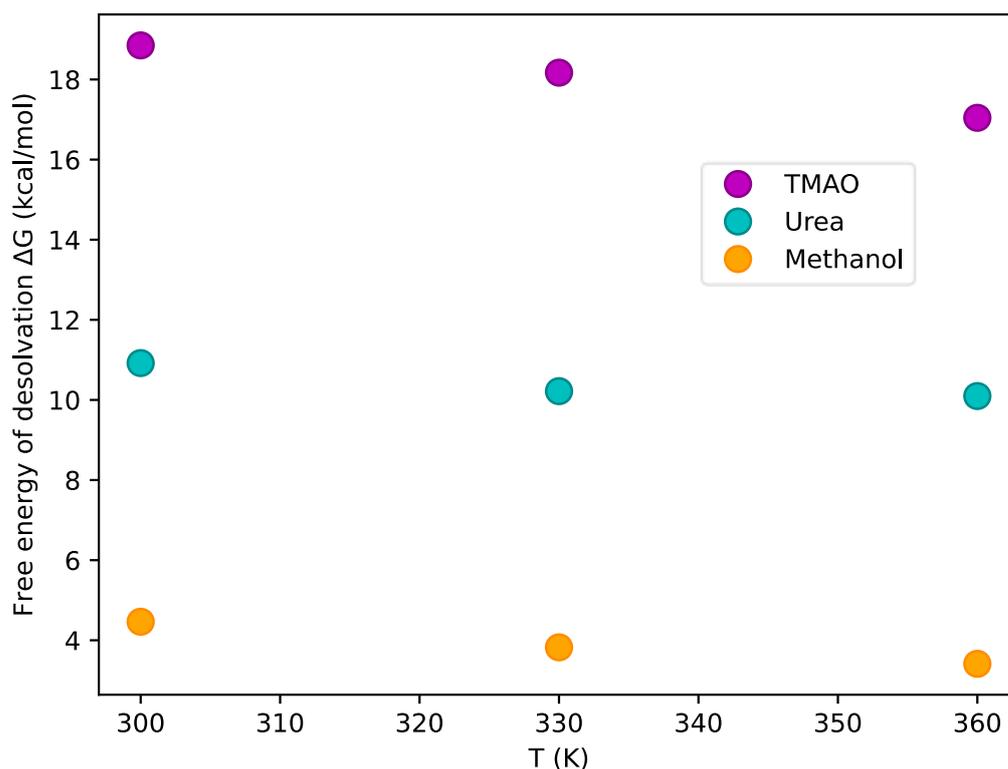


Figure 7.5: Free energy of desolvation vs temperature for different solutes: methanol (yellow), TMAO (magenta) and urea (cyan). Range of temperatures between 290-360 K. $P = 1$ bar. Molal concentration of 2.17 m. Free energies obtained with the LAMMPS software and the FEP as post-analysis method.

We have only plotted LAMMPS free energy calculations with the FEP post-analysis method. However, we have also analysed these results with the BAR and MBAR methods and their results are similar to those obtained with the FEP method.

We now investigate the desolvation free energies as a function of the inverse of the temperature. In Figure 7.5, we show the free energies of TMAO, urea and methanol for the temperatures; 300 K, 330 K and 360 K. The free energies for urea and methanol show similar trends compared to the free energies of TMAO. In addition, we observe that higher temperatures require lower free desolvation energies for all these solutes, which means that molecules are more easily desolvated as the temperature increases. We also observe that TMAO has the largest desolvation free energies, and the methanol molecule has the lowest desolvation free energies. These results manifest that the energy required to desolvate a TMAO molecule is thus larger than the methanol. Some factors could explain this trend such as the intermolecular interactions or the molecular volume, but it will require a further analysis to understand the nature of these free energy trends for the different solutes, and the posterior relation with thermophoresis.

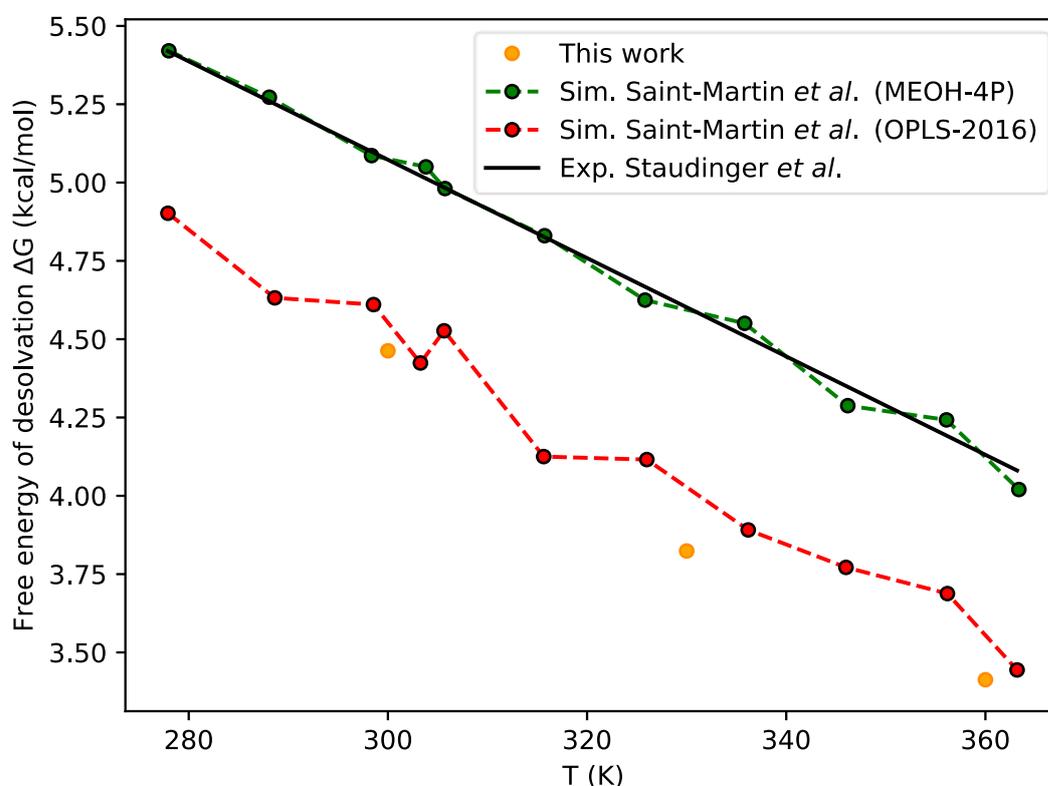


Figure 7.6: Free energy of desolvation vs temperature for methanol. Range of temperatures between 280-360 K. $P = 1$ bar. Molal concentration of 2.17 m. The green (MEOH-4P) and red (OPLS-2016) dots are computational data of Saint-Martin *et al.*¹⁸⁶ The black line is the experimental linear fit of Staudinger *et al.*¹⁸⁷

Moreover, we compare the free energy results for the methanol molecule in Figure 7.5, with the solvation free energy data available in the literature.^{186,187} In Figure 7.6, we show the free energies of methanol for a temperature range from 270 K to 370 K. The yellow dots represent the free energies of methanol, already observed in Figure 7.5. Saint-Martin *et al.*¹⁸⁶ computational free energy results are represented by the dashed lines. The red dots are the free energies of the methanol molecule for simulations with a non-polarisable forcefield (OPLS-2016). The green dots are the free energies of the methanol molecule for simulations with its polarisable forcefield (MEOH-4P). The black line represents the experimental results of Staudinger *et al.*¹⁸⁷ Our free energy results for the methanol molecule underestimates the experimental values and the Saint-Martin *et al.*¹⁸⁶ values with the polarizable force field by 1 kcal/mol of difference, but our results are in good agreement with the free energies results with a non-polarizable force field. This difference could be explained by the fact that the computational free energies¹⁸⁶ were determined with a polarisable four-site potential for methanol (MEOH-4P), which describes better the charge distribution of the methanol molecule and hence improves the free energy calculations. However, the force field used in this work for methanol, which is OPLS based force field,¹³⁸ exhibit the same values as the results of Saint-Martin *et al.*¹⁸⁶ for the non-polarizable results. Actually, our work does not focus in the absolute desolvation free energy values, but on its variations as a function of temperature, and we observe that our trend is the same as the polarizable methanol force field and the experimental results. For this reason, we do not contemplate a polarizable force field for our simulations in this work, and we observe the same trends for this accuracy level. Therefore, we focus on the dependence of the free energy on temperature (dG/dT) to check the validity of the Braun's group model for urea and methanol. We represent the correlation between the Soret coefficient determined by the Braun's model and the Soret coefficients previously calculated during the thermophoresis simulations, see Figure 7.7.

We note that the Soret coefficients determined by free energy calculations clearly overestimate the Soret coefficient calculated by molecular dynamics under a thermal gradient for urea, methanol and TMAO. Therefore, the Soret coefficient cannot be understood only in terms of an entropic solvation effect for this kind of systems, since it largely overestimates the thermophoretic value obtained previously.

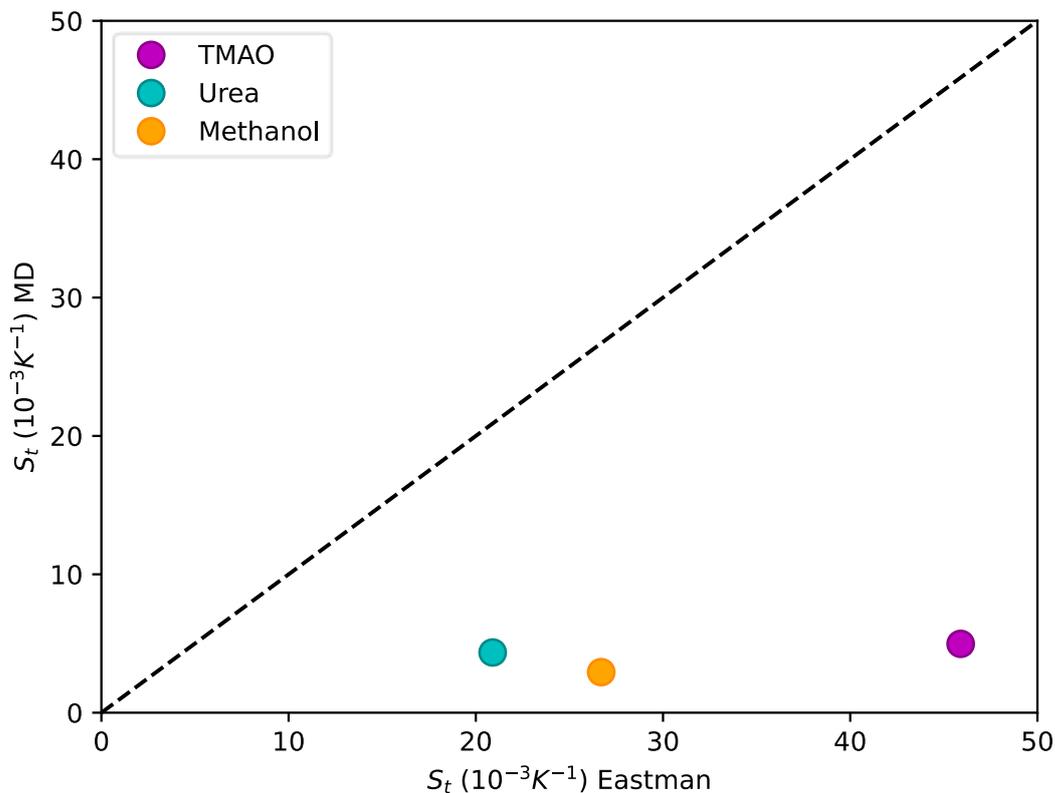


Figure 7.7: Eastman model. The Soret coefficient obtained by the Eastman model vs the thermophoretic Soret coefficient at 330 K for TMAO (magenta), urea (cyan) and methanol (yellow).

The theoretical approach of Braun remains valid for moderate temperature gradients, because the thermal fluctuations of the molecule are still a local equilibrium and then the thermophoretic depletion follows a Boltzmann distribution,⁸ which relates small concentration changes with small Gibbs-free energy differences. Although, our thermophoresis simulations are performed at large thermal gradients amplitudes, $\Delta T = 60$ K, which corresponds a temperature gradient of $\nabla T = 2.4 \text{ K} \cdot \text{\AA}^{-1}$ for a box length of $L_z = 50 \text{ \AA}$, we remember that our system is composed by two thermal gradients. The Braun group validated this approach ($S_T = \frac{1}{k_B T} \frac{dG^s}{dT}$) only if the temperature gradient ∇T is below a threshold $\nabla T < (aS_T)^{-1}$, which is given by the particle fluctuations with the hydrodynamic radius (a) and Soret coefficient (S_T), and therefore, the temperature gradient should be really small to respect this inequality. We obtain that the left part of the previous inequality is one order of magnitude larger than the temperature gradient studied, we obtain at least, $(aS_T)^{-1} > 40 \text{ K} \cdot \text{\AA}^{-1}$, for all the solutes studied in this work. The value is determined by approaching the molecular volume of Table

6.2 as the hydrodynamic radius. We should increase the thermal gradient amplitude, $\Delta T > 1000$ K, to invalidate the Braun's group approach. For this reason, we consider that the thermophoretic depletion of our systems follows a Boltzmann distribution, and can be studied in terms of local equilibrium.

A few precisions should be considered that may explain this difference in the Soret coefficient order of magnitude for our system. The Braun's group applied this model for polystyrene beads of diameter 1.1, 0.5, and 0.2 μm and 1000-bp (base pair) DNA molecules,⁴⁹ which are much larger than our solutes. Moreover, we should mention that one of the intrinsic properties that affect the Soret effect is the mass effect, as we already ratified in the previous chapter and was experimentally confirmed.⁹⁸⁻¹⁰⁰ However, Neither Eastman or the Braun group consider the mass effect in their thermodynamic models and the solvation free energy does not depend on the mass of the solute. In conclusion, further considerations on this theory must be applied to describe the thermophoretic behaviour in terms of solvation free energies for small molecular systems.

7.2 Diffusion effect (in process)

The last part of this project tries to understand the thermophoresis effect from a kinetic point of view. As we have mentioned before, the Soret coefficient is defined as a ratio of two diffusive effects once the system reaches the static equilibrium; D^S is the diffusion coefficient of the solute, and D_T^S is the phenomenological thermal diffusion coefficient of the solute. Prigogine's approach^{72,73} is an appealing theory, because the author proposes that the Soret coefficient can be understood in terms of activation processes of the diffusion between the solute and the solvent:

$$S_T = \frac{E_a^S - E_a^W}{RT^2}, \quad (7.4)$$

where the activation process of the diffusion is seen as the sum of the energy needed to form a hole in which the molecule will be placed, and the energy needed for the particle to detach itself from its neighbours. Furthermore, these activation energies can be extracted from the

temperature dependence of the Fick's diffusion for the different species, which, if we assume an Arrhenius temperature-dependence, is:

$$D^s = D_0^s \exp\left(-\frac{E_a^s}{RT}\right), \quad (7.5)$$

where D_0^s is the pre-exponential term and describes some factors such as the collision frequency and the orientation between the particles. However, diffusion itself does not depend on the mass of the particles, so the Prigogine model does not take into account the effect of mass, which is an intrinsic property affecting the Soret coefficient already observed in the previous chapter. Therefore, Artola, Rousseau and Galliéro⁷⁵ have proposed an extended version the Prigogine model to explicitly include the mass effect in the Soret coefficient determination.

$$S_T = \frac{E_a^s - E_a^w}{RT^2} + \frac{E_a^s + E_a^w}{RT^2} \frac{M_s - M_w}{M_s + M_w} \quad (7.6)$$

where the authors expressed the Soret coefficient in form of different contributions, which takes a similar form as introduced by the Köhler group^{76,102} ($S_T = S_T^{ch} + S_T^{iso}$). The first term is the usual Prigogine model for thermal diffusion and the second term is the mass term, which is described in terms of the activation energies of the different species and their mass ratio. In this section, we will review the feasibility of both models for our particular aqueous systems, as well as the mass effect from a kinetic point of view.

7.2.1 Simulation details

The diffusion for several systems has been studied to determine the Soret coefficient. All simulations are performed with the LAMMPS package. The same cubic box ($L_x = L_y = L_z \approx 33.15 \text{ \AA}$) is considered from the previous free energy calculations, which is centred at (0,0,0), with equiangular and perpendicular distributions ($90^\circ, 90^\circ, 90^\circ$). The equilibration part is replicated from thermophoretic simulations. The same steps have been followed up to the second NVE simulation, which are the steps 1 to 8. The Coulombic cut-off is 8.5 \AA with particle-particle particle-mesh (PPPM) solver¹⁶⁰ for the long-range electrostatic forces, and the Lennard-Jones cut-off is 9 \AA , which are the same as the thermophoretic conditions. The diffusion analysis has been studied under the microcanonical ensemble (NVE), which has no

barostat or thermostat. We have used the NVE ensemble for these two reasons; the constant volume avoids changes in the size of the box, which make it difficult to analyse diffusive movements across the boundaries; and the absence of thermostat avoids any kinetic perturbation due to the action of an external thermostat.

The diffusion constants have been determined by measurements of mean squared displacements (MSD):

$$\text{MSD} = \sum_{i=1,n} \langle (x_i(t) - x_i(0))^2 \rangle = 2nD^s t, \quad (7.7)$$

where $\langle (x_i(t) - x_i(0))^2 \rangle$ is the uni-dimensional mean squared displacement, and n is the number of dimensions. The MSD gives us the information on the position of a particle relative to a reference position over time.

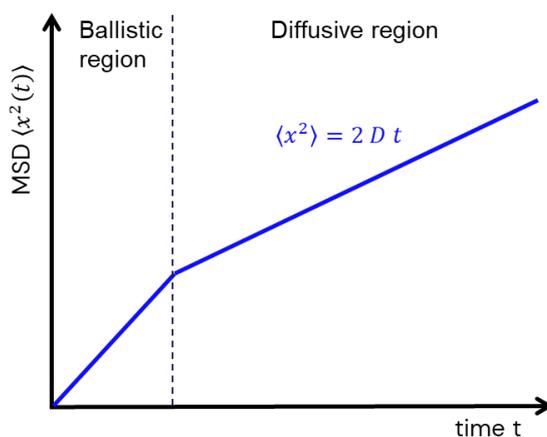


Figure 7.8: Uni-dimensional mean squared displacement (MSD) vs time. The first steps of the MSD correspond to the ballistic region, and after some steps the motion goes towards the diffusive region, where the diffusion constant can be obtained. Dotted points correspond to the transition regime that is approximately to 170 fs.

The motion of a particle can be divided in two regions, see Figure 7.8 . The first steps are dominated by the ballistic region, which is the mean free path that the particle can travel freely in the solution, before a collision, which could change its momentum. After a few time steps, there is a transition process whereby the slope of the MSD as a function of time decreases ($d\text{MSD}/dt$). At this moment the particle starts to collide with the other particles in the solution,

so their trajectory would be modified. After this transition regime, the mean squared displacement will describe the diffusive part of the molecule in the system. This part of the molecular motion is known as diffusive region, and its slope as a function of time is considered as the diffusion constant of the particle in this system, see Figure 7.8.

The MSD analysis has been performed with a modified version of the PyLAT analysis tool¹⁸⁸. The total NVE production time is 5 ns and the molecular coordinates are printed each 1 ps. The following graph represents the MSD for the first 100 ps.

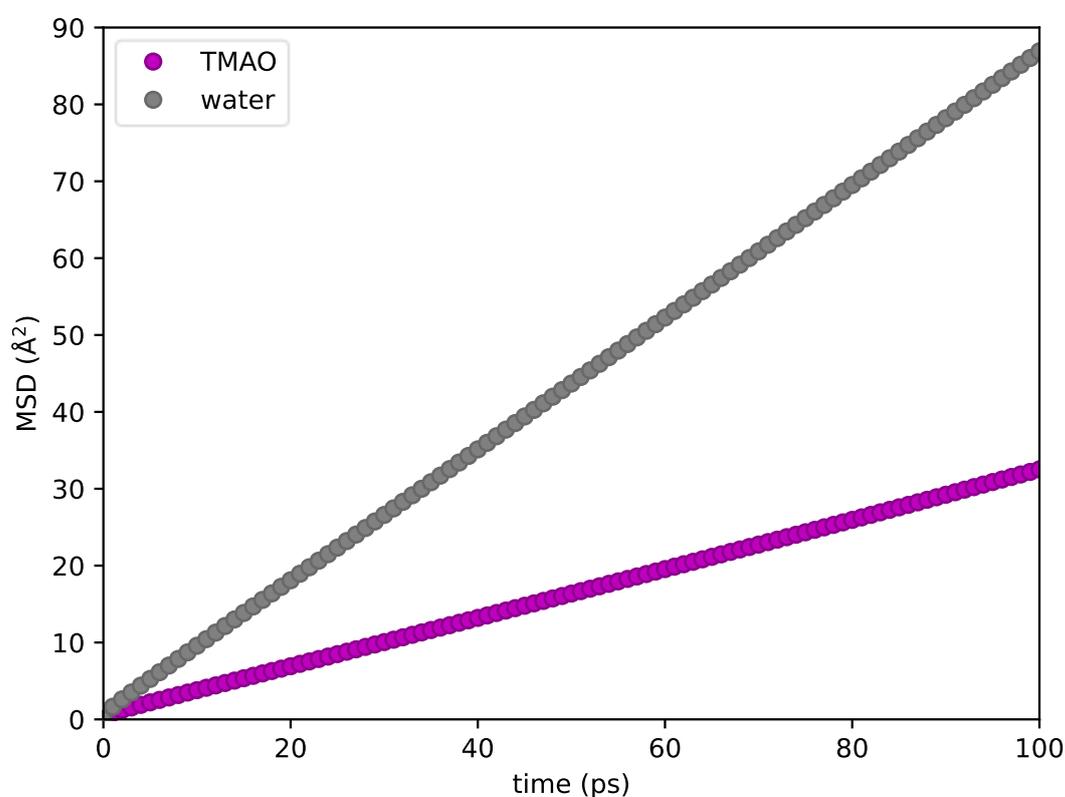


Figure 7.9: MSD for the TMAO solution. Mean square displacement vs time. The magenta dots represent the MSD displacement of the TMAO molecule. The grey dots represent the MSD displacement of the water molecule. 2.17 m molal concentration.

In Figure 7.9, we show the MSD for the TMAO and water molecules at 300 K. All simulations are performed at 2.17 m concentration. We observe the slope of water is larger than the slope of TMAO, which indicates a larger diffusion coefficient. In the next section, we will check the viability of both models (Prigogine and Artola) for our molecular systems in the diffusive regime.

7.2.2 Diffusion regime results

The diffusion coefficients are calculated for the TMAO, urea, methanol and glucose solutes and their respective waters in the solution. Activation energies are determined by the logarithm of the diffusion coefficient as a function of the inverse of temperature for a range of temperatures between 280-360 K:

$$\ln D^s = \ln D_0^s - \frac{E_a^s}{RT}. \quad (7.8)$$

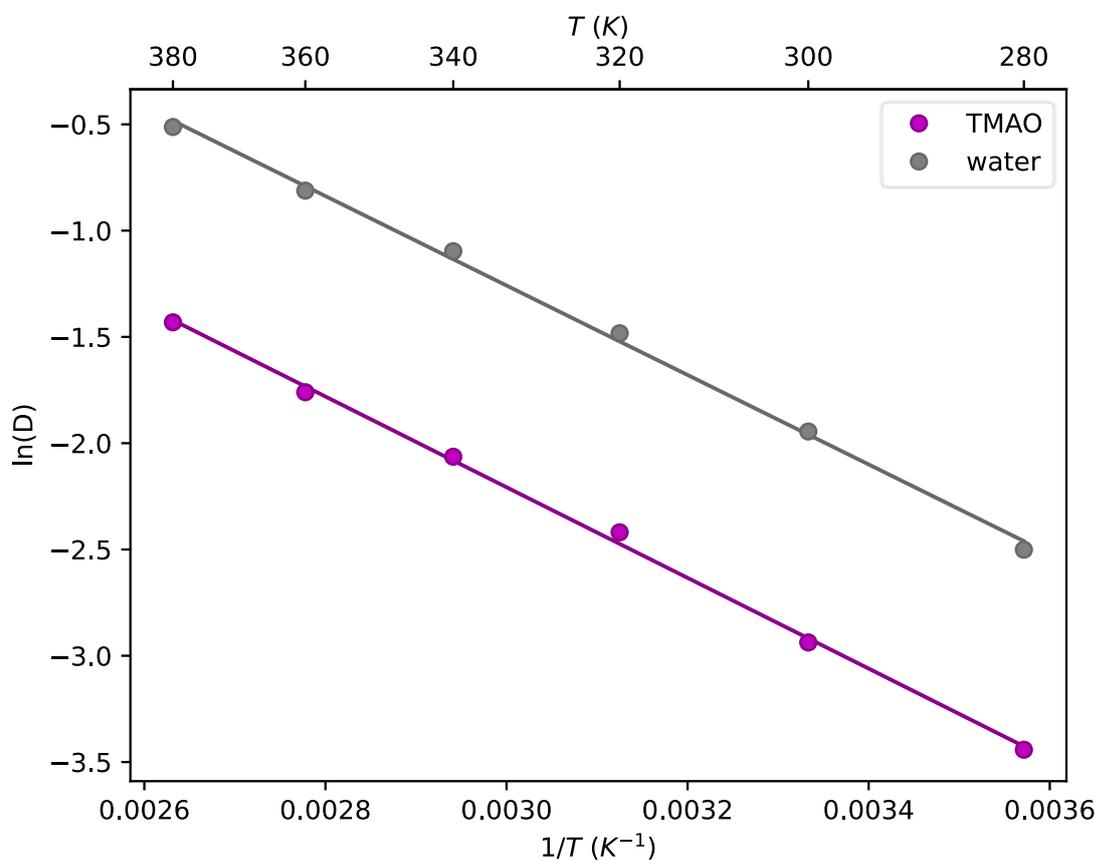


Figure 7.10: Diffusion coefficients. Logarithmic diffusion vs inverse of the temperature for the TMAO molecule and the waters in the solution. Range of temperatures between 280-380 K. P = 1 bar. TMAO molal concentration of 2.17-m.

First, we check the robustness of the diffusion process for the TMAO molecules in a 2.17 m aqueous solution. We plot the evolution of the logarithmic diffusion as a function of the inverse of the temperature for the TMAO and water molecules inside the solutions, see Figure 7.10.

The slope of these trend will determine the activation energy for both molecules in this system conditions.

The logarithmic diffusion coefficients determined by the MSD calculations at different temperatures are represented in Figure 7.10. We perceive that the water diffusion is higher than the TMAO diffusion, and both molecules exhibit the same trend of $\ln(D)$ as a function of the inverse of the temperature. Moreover, we observe that $\ln(D)$ decreases with temperature, which means that the diffusion coefficient for TMAO and water increases with the temperature. This result agrees with the thermodynamics principles on which an increase in temperature leads to a higher molecular kinetic energy, and hence higher diffusion rates for all molecules in the system.

On the other hand, the activation energy is defined as the minimum energy that a system needs to be able to initiate a certain process. In our case, the Arrhenius equation defines this activation energy as the energy needed for a given particle to diffuse into the medium. Therefore, higher activation energies lead to lower diffusions. Thus, we expect the TMAO molecules to have a higher activation energy than the water molecule. Therefore, we have calculated the difference between the activation energies from their respective slopes in Figure 7.10. The energy difference between the solute and the solvent in this system is $\Delta E_a^s - \Delta E_a^w \approx 0.4$ kcal/mol and the Soret coefficient according to the Prigogine model is $S_T \approx 2 \cdot 10^{-3} \cdot \text{K}^{-1}$, which is on the same order of magnitude of the Soret coefficient obtained by thermophoresis, $S_T \approx 5 \cdot 10^{-3} \cdot \text{K}^{-1}$, but we do not get the same value. We have also calculated the Soret coefficient from the Prigogine model for other solutes presented below.

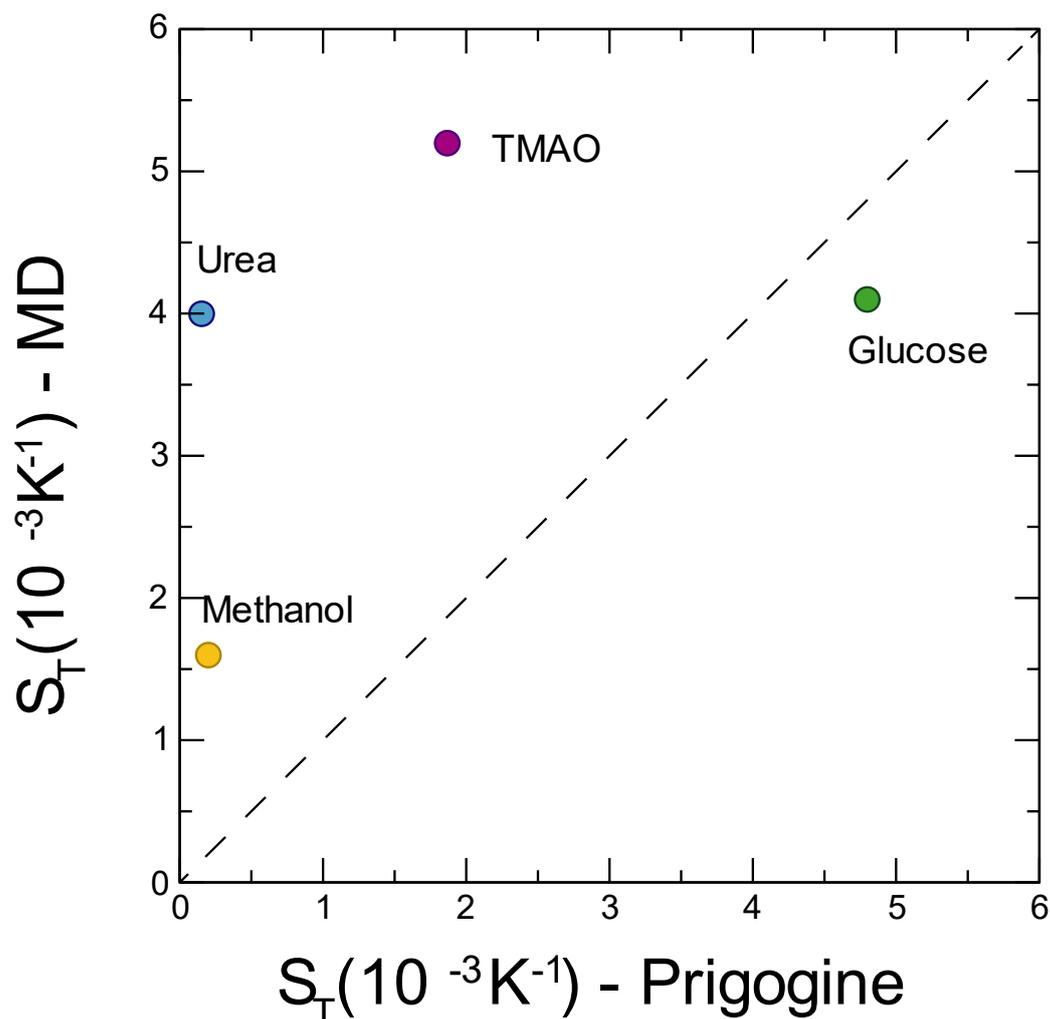


Figure 7.11: Prigogine model. The Soret coefficient obtained by the Prigogine model vs the thermophoretic Soret coefficient at 330 K for TMAO (magenta), urea (cyan), methanol (yellow) and glucose (green).

The Soret coefficients determined by the thermophoresis simulations and by the diffusive Prigogine model are represented in Figure 7.11. We note that the Soret coefficient for the TMAO molecular system calculated by the Prigogine model is smaller than the Soret coefficient determined by thermophoresis, as described above. However, the glucose molecule has a higher Soret coefficient by the Prigogine model than by thermophoresis. On the other hand, both Soret coefficients for urea and methanol calculated by the Prigogine model are one order of magnitude lower than those previously obtained by the thermophoresis simulations. From these results we do not observe any clear trend with respect to the solutes studied. However, the mass and size of the molecule could have an effect on their activation energies,

since larger or heavier molecules would make their movement more costly in terms of energy. Therefore, these intrinsic properties of the solutes could have an impact in the resulting Soret coefficient calculated by the Prigogine's model, and could explain that these Soret coefficients by the Prigogine's model increases as the size and mass of the solute increases.

However, we have already confirmed that mass is one of the intrinsic properties of the molecule that affects solute accumulation under a thermal gradient. We observed a progressive increase of the Soret coefficient of the thermophoresis simulations as we increase the mass ratio between the solute and the solvent. However, the Prigogine model does not take into account the mass effect. Artola, Rousseau and Galliéro⁷⁵ included the mass effect in the Prigogine's approach, see eq. (7.6). The authors rescaled the temperature in terms of the ideal part of the partitional function means to displace the position of the diffusion transition state as a ratio of the relative masses.

Therefore, we have performed isotopic simulations of TMAO under the same conditions as TMAO in its natural mass. The first simulation is with half the mass of TMAO and we have doubled the mass of TMAO in the second simulation. The activation energies have been obtained from the same Arrhenius expression, see eq. (7.8).

Different Soret coefficients for a TMAO solution are shown in Figure 7.12 as a function of relative mass ($\delta M = \frac{M_s - M_w}{M_s + M_w}$). The Soret coefficients determined by the thermophoresis simulations are represented by circles. The Soret coefficients determined by the Prigogine model are represented by triangles, see eq. (7.4). The Soret coefficients determined by the Artola, Rousseau and Galliéro model are represented by squares, see eq. (7.6). The Prigogine model shows no effect on mass as we increase the mass of TMAO. Indeed, diffusion does not depend on mass, since we do not observe any noticeable change in the activation energy. Therefore, the activation free energies are mass independent, which leaves the solute effect observer in Figure 7.11 for the Soret coefficients calculated by the Prigogine model to a particle size effect.

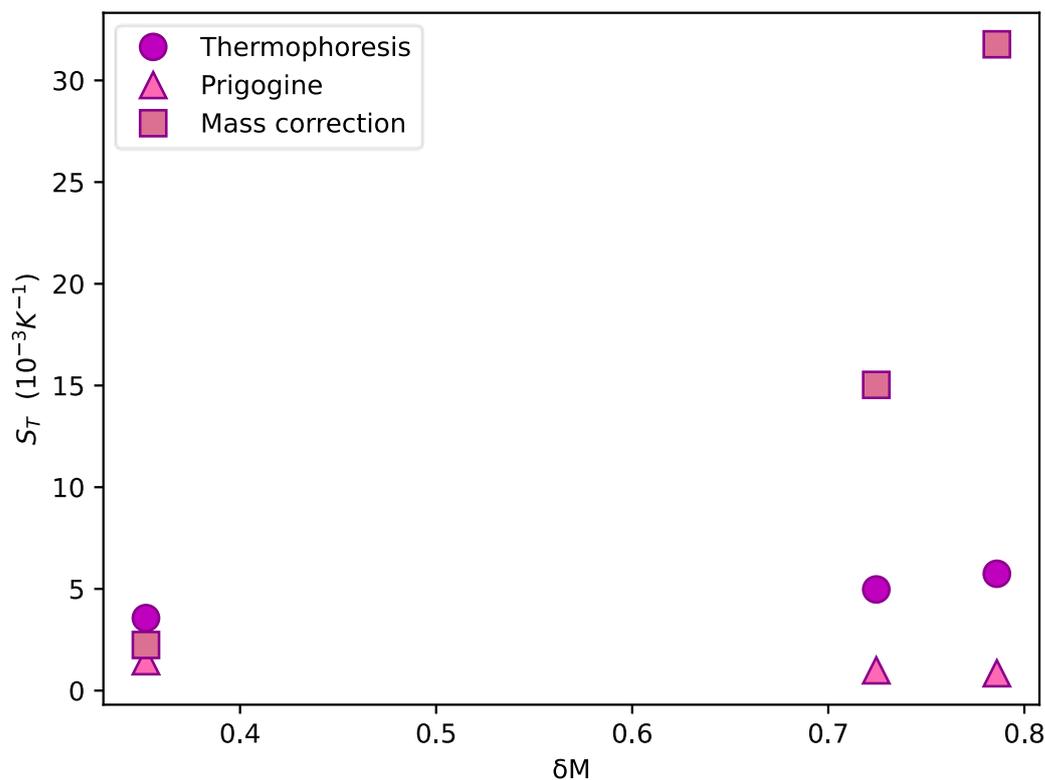


Figure 7.12: Mass correction on the Prigogine model. Soret coefficient vs relative mass difference (δM) for TMAO at 330 K. Thermophoretic simulations (circles), Prigogine's diffusion hypothesis (triangles) and Prigogine's diffusion with mass effect hypothesis (squares).

On the other hand, adding mass effect results in an exponential increase of the Soret coefficient value as a function of the mass ration. However, this increase does not correspond to the mass effect observed in the thermophoretic results. The kinetic model proposed by Artola, Rousseau and Galliéro ⁷⁵ describes the mass effect for mass ratios of $M_2/M_1 = 1$ and $M_2/M_1 = 2$, which is lower than our mass ration, which are between the TMAO and water $M_{TMAO}/M_{water} \approx (2, 4, 8)$, so the mass effect described in the equation (7.6) could be no longer valid for a large mass difference. Moreover, the authors discussed about some additional effects that should be taken into account to properly describe the Soret coefficient and its mass effect. Inertial moments in the diffusive process, which could be part of the isotopic effect, should also be studied. Furthermore, the authors formula seems to predict composition-independent Soret coefficients, but we already observed that the Soret coefficient depends on the molality.

7.2.3 Ballistic regime

One of the last points of our research is that we observe a correlation between the thermophoretic Soret coefficient and the Soret coefficient determined by the molecular motion in the ballistic region. We analysed the MSD calculation for shorter timesteps of 10 fs for a total simulation time of 1 ns to describe the ballistic region. In addition, we include a jump parameter of 1000 for the time averaging to avoid a correlation between too close time steps, which means that the MSD for 10 fs is averaged over the distance differences of the following time steps; 0-10, 1000-1010, 2000-2010....

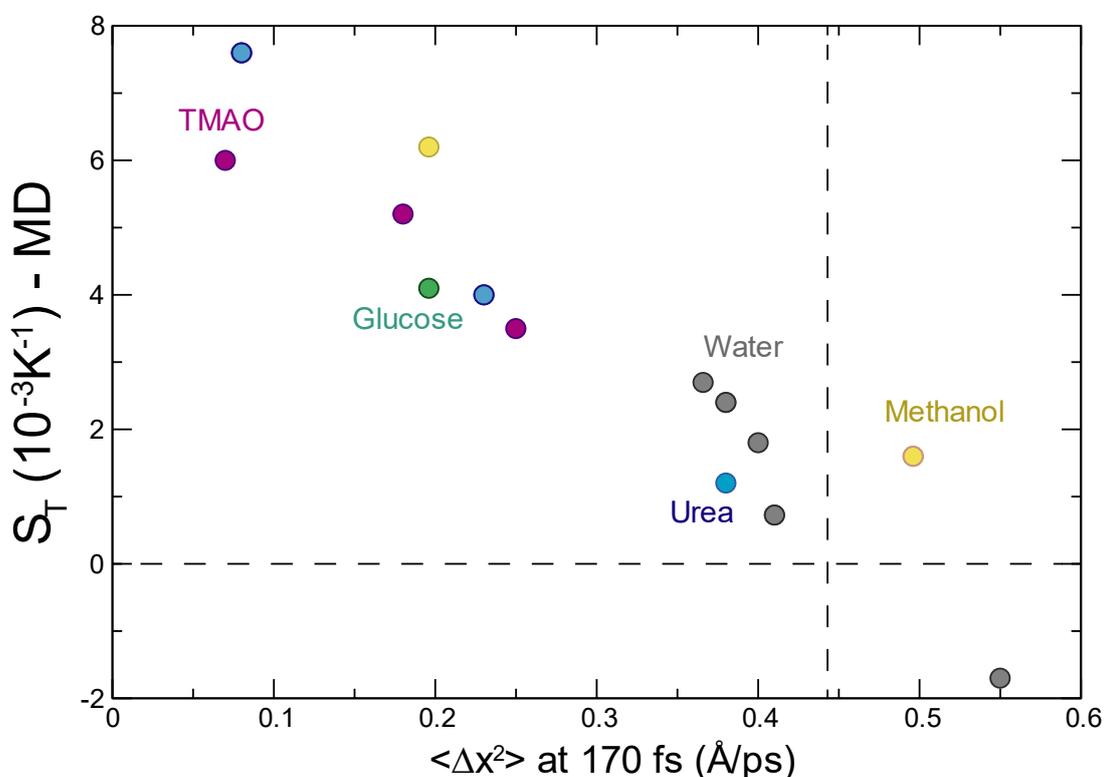


Figure 7.13: Ballistic regime. The thermophoretic Soret coefficient vs the MSD ($\langle \Delta x^2 \rangle$) displacement at 170 ($\text{\AA}/\text{ps}$) for several solutes at 330 K: methanol (yellow), urea (cyan), TMAO (magenta), glucose (green) and water (grey).

The ballistic MSD is determined for several molecules, as well as some isotopic masses of certain molecules. We consider 170 fs to be the limit of the ballistic region, because the first derivative of the MSD reaches the maximum at this point ($\partial \langle \Delta x^2 \rangle / \partial t = 0$). The Soret

coefficients are represented in Figure 7.13 as a function of the ballistic MSD. We observe a linear correlation between the Soret coefficient calculated by the thermophoresis simulations and the mean squared displacement of each molecule after 170 fs. These results could indicate that the thermophoretic forces applied on the molecules under a thermal gradient are actually on very short timescales, and hence the Soret coefficient could be understood in terms of pure kinetic effect of the particles rather than a diffusive effect. However, these results are really preliminary due to the lack of time in this project, and more studies should be carried out on this idea to better understand this relationship.

8. Conclusions

This thesis seeks to understand the effect of thermophoresis for dilute aqueous solutions, as this phenomenon could explain the precursor accumulation in the context of the origins of life. However, its nature is still not completely understood, and hence several hypotheses have been proposed in the literature.

For this purpose, we presented a robust all atom molecular dynamic approach to undertake thermophoresis in dilute realistic solutions at the molecular level. A homogeneous thermal gradient prior to concentration gradient is vital for the proper evolution of the thermally-driven processes. The recent heat-exchange algorithm (eHEX)⁷⁰ lead to an excellent energy conservation for nanoscale simulation range and integration steps no larger than 2-fs. The advantage of this temperature gradient generator is that it does not require the use of thermostats with restrained solvent molecules and an ill-defined simulation ensemble.

8.1 In silico thermophoresis set-up

The first part of our work was to ensure that the total energy is conserved throughout the simulation time. We have ensured that the total energy does not diverge over 10 ns as long as the timestep is no longer than 2 fs. We assumed that the equilibration timescale of 100 ps for 5 nm box length systems and 200 ps for 10 nm box length systems are enough to reach the thermal gradient ($\Delta T \approx 60$ K) in the steady state condition. Furthermore, we considered that solute molecules do not interact directly with each other in sufficiently dilute solutions. We found that the 2.17-m TMAO solution was a good compromise, because they allow to accelerate convergence while being in the dilute regimes where solute-solute interactions are negligible. For such systems, we estimated that a 2-ns timescale can safely be chosen as the time required to reach the steady state for the solute concentration profile as long as the conditions for the simulation and the system sizes do not significantly differ from the ones studied here.

Moreover, we explored several factors that could affect the stability of the thermal and concentration gradient and the robustness of the results. We first studied the system size effect of the system, and we concluded that the concentration gradient is practically insensitive to the size of the system box, because the Soret coefficient expression is, by definition, an intensive variable. Once the steady state condition for the concentration gradient is reached at a given median temperature. However, we observed that the Soret coefficient associated error increases with the system size. For this reason, the smallest box system of 50 Å was the best option in terms of computational resources and lower statistic errors.

Secondly, we studied the effect of the amplitude gradient for a temperature gradient range from 0.8 to 3.6 K · Å⁻¹. We did not observe any effect on the Soret coefficient, but we found that larger is the amplitude of the thermal gradient lower are the statistical uncertainties. However, a compromise had to be found between the accuracy of the measured concentration-gradient, that should be large enough to exceed the statistical uncertainties, while employing a reasonable temperature-gradient where the water phase lies within, or not too far, from its equilibrium liquid phase domain (namely, between 273 and 373 K). For these reasons, we chose to employ the 2.4 K · Å⁻¹ gradient, which satisfies these conditions in most cases.

The third aspect was the frequency of heat transfer, which plays an important role in the molecular concentration inside the reservoir, but it does not really affect the concentration profile outside the thermalised regions. For this reason, we always determined the Soret coefficient from the concentration profile far apart of the thermalised regions.

Finally, we tested different force fields for the TMAO solute and the water molecule, and a certain stability has been observed for the same system by using different force fields available in the literature, just one force field set gave a disparity in the concentration gradient. As a conclusion, three of four force field combinations gave us the same results, which indicates a good consistency for the force field combination between the TIP4P/2005 water model and the Kast TMAO model.

8.2 Intrinsic and thermodynamic contributions

There are several heuristic descriptions in the literature, which try to decompose the Soret coefficient into different contributions. Here, we enquired into some of these physical properties most relevant for the Soret effect. The first part was the study of the external factors such as pressure, temperature and concentration that could affect the Soret coefficient.

We observed a small increase of the Soret coefficient for high median pressures of 100 bar and 1000 bar, but we did not notice any difference for relatively low median pressures of 1 bar and 10 bar in their respective Soret coefficients:

$$S_T^{1 \text{ bar}} \approx S_T^{10 \text{ bar}} < S_T^{100 \text{ bar}} < S_T^{1 \text{ kbar}} . \quad (8.1)$$

Although we observed an increase of the Soret coefficient as a function of the median pressure, the effect is very low. Therefore, we did not consider that the Soret coefficient depends on the median pressure, since large pressures must be applied to observe a significant effect in the Soret coefficient for dilute aqueous solutions.

We also investigated the median temperature dependence of the Soret effect for ethanol, urea and TMAO. One result was that the Soret coefficient of TMAO increases as a function of the median temperature:

$$S_T^{270\text{K}} < S_T^{300\text{K}} < S_T^{330\text{K}} < S_T^{360\text{K}} < S_T^{390\text{K}} . \quad (8.2)$$

We observed that the Soret coefficients for TMAO and urea followed the exponential depletion form described phenomenologically by the equation of Iacopini and Piazza^{79,80} ($S_T = S_T^\infty [1 - \exp(T^* - T)/T_i]$), but the ethanol molecules did not follow the same exponential depletion form. This was not surprisingly, as this Soret coefficients results for ethanol are in agreement with the previous literature results.⁸⁴ We thus concluded that temperature has an effect on the Soret coefficient, but its effect depends on the nature of the solute. In general terms, we observed a progressive increase of the Soret coefficient as a function of the temperature for low median temperatures, which means that thermophobicity increases with the temperature.

In general terms, solute molecules try to accumulate in the cold region as the median temperature increases in dilute aqueous solutions.

Furthermore, we analysed the concentration effect of the Soret coefficient, and we observed that the Soret coefficients of TMAO does not exhibit a strong dependence with concentration:

$$S_T^{1.1 \text{ m}} \approx S_T^{2.2 \text{ m}} \approx S_T^{3.3 \text{ m}} > S_T^{4.3 \text{ m}} . \quad (8.3)$$

This trend indicated that in sufficiently dilute solutions, the Soret coefficient is not concentration dependent in dilute aqueous solutions. This effect could be understood by the fact that the TMAO molecules practically do not interact among themselves, and their respective water solvation shell depends only on the temperature gradient, and thus keeping the “chemical” interaction solute-solvent invariant.¹²⁶ However, the molal concentration of $C_0^{(4)} = 4.34 \text{ mol/kg}$ is a relatively high concentration, where the hydration shells overlap and TMAO molecules are in close contact. For this reason, we started to observe some effects on the Soret coefficient, and we opted for the median concentration of $C_0^{(2)} = 2.17 \text{ mol/kg}$. This concentration contains enough TMAO molecules to obtain good statistics for the determination of the Soret coefficient while being in a regime that is dilute enough such that it does not affect too much the Soret coefficient.

The second part was the study of the intrinsic properties such as the dipole moment, molecular size or mass effect of different molecules (ethanol, methanol, urea, TMAO, glucose):

$$\mu_{Glucose} < \mu_{Ethanol} \approx \mu_{Methanol} < \mu_{Urea} < \mu_{TMAO} , \quad (8.4)$$

$$V_{Methanol} < V_{Urea} < V_{Ethanol} < V_{TMAO} < V_{Glucose} , \quad (8.5)$$

$$M_{Methanol} < M_{Ethanol} < M_{Urea} < M_{TMAO} < M_{Glucose} . \quad (8.6)$$

The following formula orders the Soret coefficient as a function of the solute:

$$S_t^{Methanol} \approx S_t^{Glucose} < S_t^{Ethanol} \approx S_t^{Urea} < S_t^{TMAO} . \quad (8.7)$$

Broadly speaking, we observed larger Soret coefficients for molecules with larger dipole moments such as TMAO or urea. However, a further analysis at different dipole moments for

urea demonstrated that its Soret coefficient decreases as the dipole moment of the modified urea increases. Previous studies already discussed this higher complexity for polar systems as compared to non-polar systems,¹⁰⁵ the authors assumed that non-ideal structural changes (e.g., excess effects) must be considered to understand the effect of the polarity in the Soret coefficient. Moreover, the modification of a force field by rescaling charges could lead to an unrealistic description of the inter-molecular forces and thus invalidate the force field. Therefore, further studies are required to understand the evolution of the Soret coefficient as a function of the dipole moment of a molecule.

In general terms, we also observed that the lowest values in the Soret coefficients are associated to small molecular systems in terms of mass and size. A progressive increase of the Soret coefficient is observed as the molar mass and the size of the molecules increases, except for glucose. These trends are in agreement with previous studies that have reported a size dependence in colloidal solutions,^{49,68,177} and a mass effect in linear n-alkanes.^{98–100} However, all these intrinsic contributions combined make it difficult to understand the physical meaning of a particular contribution in the Soret effect. For this reason, we examined the effect of isotopic molecules in the Soret coefficient in order to observe the pure mass effect.

In the literature, the Soret coefficient is usually phenomenologically divided into two independent terms ($S_T = S_T^{ch} + S_T^{iso}$),^{76,102} where S_T^{iso} is the isotopic contribution that depends only on mass and moment of inertia. In this work, we analysed the viability to describe the mass effect to the Soret coefficient for small solutes in aqueous solutions by this equation:

$$S_T = S_T^0 + a_M \frac{M_2 - M_1}{M_2 + M_1}. \quad (8.8)$$

We assumed that the moment of inertia effect does not significantly affect the isotopic Soret coefficient as the mass ratio increases.

Our results showed that all molecules studied exhibit a thermophobic behaviour for positive mass ratios at their natural mass. Furthermore, urea, methanol and water acquired a thermophilic character at certain negative mass ratios. We thus confirmed that heavier molecules have higher Soret coefficient values, and we demonstrated a linear relation between the Soret coefficient and the mass ratio (δM), the dotted lines show the linear fit of eq. (8.8).

We also evidenced that the Soret

coefficient depends only on the mass ratio between the solute and the solvent and it is not really affected on the absolute mass of the solution. These results are represented by three vertical points for TMAO and urea in Figure 8.1. Moreover, we wanted to highlight that $S_T^0 \approx 0 \text{ K}^{-1}$ for the isotopic water simulations. This value can be explained by the fact that these simulations are pure water solutions in which we only modify the mass of certain water molecules, and therefore we just expected a pure mass effect in the Soret coefficient. Finally, we noted that S_T^0 was larger for the TMAO molecule as compared to urea or methanol, which indicated a higher “chemical” contribution or moment of inertial effect of the Soret coefficient. However, an extensive study should be done to know the moment of inertia effect to isolate the “chemical” contribution.

8.3 Molecular models

The final part of this work was testing the feasibility of two plausible models, which were of particular interest and relevance for molecular systems, but which came from different perspectives.

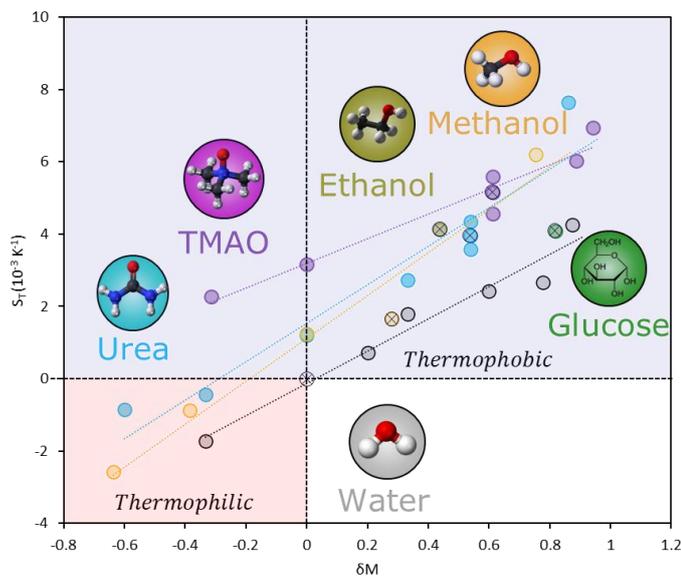


Figure 8.1: Mass effect summary. Methanol (yellow), ethanol (olive), urea (cyan), TMAO (magenta), glucose (green), and water (grey).

The first model was the Eastman's approach,⁷¹ which tried to relate the thermally-driven effect of a molecule to a temperature dependence of a solvation free-energy process, and the Braun group⁴⁹ developed this previous idea and their experimental results shown that the thermophoretic steady states could be described by a Boltzmann distribution in local thermodynamics equilibrium of the equation:

$$S_T = \frac{1}{RT} \frac{dG^s}{dT}. \quad (8.9)$$

Therefore, we represented the correlation between the Soret coefficient determined by the Braun's model and the Soret coefficients previously calculated during the thermophoresis simulations.

We noted that the Soret coefficients determined by free energy calculations clearly overestimated the Soret coefficient calculated by molecular dynamics under a thermal gradient for all solutes. Furthermore, we should mention that this model does not take into account the mass effect already ratified and experimentally demonstrated,⁹⁸⁻¹⁰⁰ and we concluded that the Soret coefficient cannot be understood only in terms of an entropic solvation effect for dilute aqueous solutions, since it largely overestimated the thermophoretic value obtained previously.

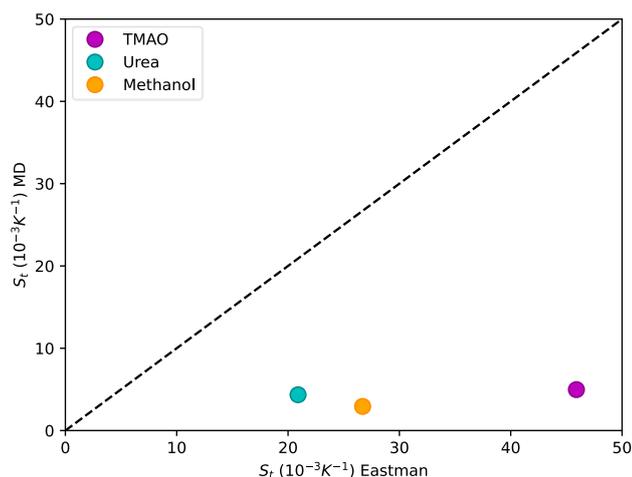


Figure 8.2: Eastman model. The Soret coefficient obtained by the Eastman model vs the thermophoretic Soret coefficient at 330 K for TMAO (magenta), urea (cyan) and methanol (yellow).

The second model was the Prigogine's model, which conjectures that the thermophoretic motion of a molecule is related to the activation energies for the diffusion of the solute and solvent molecules.⁷²⁻⁷⁶

The Soret coefficients determined by the thermophoresis simulations and by the diffusive Prigogine model are represented in Figure 8.3. The results did not show any clear trend with respect to the solutes studied. One explanation could be that mass and size could have an effect on their activation energies, since larger or heavier molecules would make their movement more costly in terms of energy.

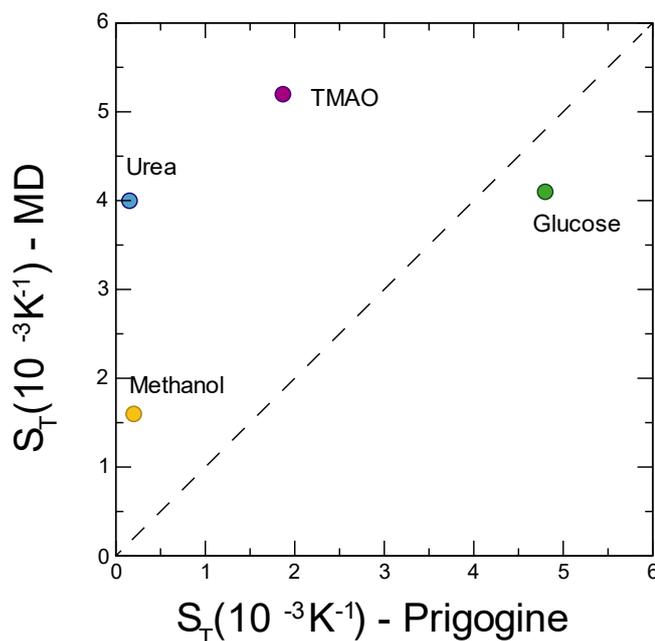


Figure 8.3: Prigogine model. The Soret coefficient obtained by the Prigogine model vs the thermophoretic Soret coefficient at 330 K for TMAO (magenta), urea (cyan), methanol (yellow) and glucose (green).

Furthermore, we analysed the Artola, Rousseau and Galliéro⁷⁵ model that include the mass effect in the Prigogine's approach. However, the mass effect resulted in an exponential increase of the Soret coefficient value as a function of the mass ration, which did not correspond to the mass effect observed in the thermophoretic results. The authors discussed about some additional effects that should be taken into account to properly describe the Soret coefficient and its mass effect. Inertial moments in the diffusive process, which could be part of the isotopic effect, should also be studied. In conclusion, a more detailed study should be done in order to understand thermophoresis as an activated process of diffusion.

The last point of this work was to check whether the Soret coefficient can be understood as a ballistic effect of molecular motion.

The ballistic MSD was determined for several molecules, as well as some isotopic masses of certain molecules. We considered 170 fs to be the limit of the ballistic region. We observed a linear correlation between the Soret coefficient calculated by the thermophoresis simulations and the ballistic MSD at 170 fs. These results could indicate that the thermophoretic forces applied on the molecules under a thermal gradient are actually on very short timescales, and hence the Soret coefficient could be understood in terms of pure kinetic effect of the particles rather than a diffusive effect. However, these results were really preliminary due to the lack of time in this project, and more studies should be carried out on this idea to better understand this relationship.

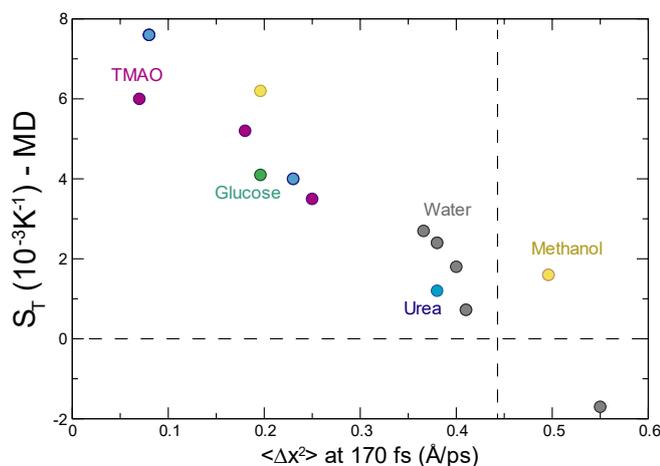


Figure 8.4: Ballistic regime. The thermophoretic Soret coefficient vs the MSD ($\langle \Delta x^2 \rangle$) displacement at 170 ($\text{\AA}/\text{ps}$) for several solutes at 330 K: methanol (yellow), urea (cyan), TMAO (magenta), glucose (green) and water (grey).

9. Résumé de la thèse

Dans les pages suivantes, nous donnerons un aperçu des principaux sujets de cette thèse et des résultats obtenus. Nous suivrons le même ordre que dans le texte principal, qui est divisé en quatre parties liées aux principaux domaines de contenu. Le premier chapitre présente le contexte général de ce travail et les objectifs. Le deuxième chapitre décrira le terme du coefficient de Soret et le processus suivi pour générer les conditions de la thermophorèse. Dans le troisième chapitre, nous étudierons certaines propriétés intrinsèques des molécules et les conditions thermodynamiques qui peuvent affecter l'effet Soret. Enfin, nous discuterons de certains modèles qui pourraient expliquer les origines moléculaires de la thermophorèse dans le dernier chapitre.

9.1 Introduction

La thermophorèse est un processus thermique, dans lequel un mélange de composants devient inhomogène sous l'effet d'un gradient de température. Mis en évidence pour la première fois il y a 150 ans par C. Ludwig^{1,2} puis plus tard par C. Soret,³ ce phénomène se manifeste pour une large gamme de tailles de systèmes. Ces mouvements thermodynamiques, dont l'amplitude peut atteindre plusieurs millimètres,^{5,6} ont été étudiés pour des systèmes moléculaires allant de l'échelle nanométrique à microscopiques.⁴

Ces dernières années, la thermodiffusion a suscité un grand intérêt, au niveau académique et industriel, en raison de son importance dans de nombreuses applications scientifiques et techniques. Un exemple est la thermophorèse à micro-échelle (MST),^{9,10} qui analyse les systèmes biomoléculaires en se basant sur le mouvement dirigé des particules dans un gradient thermique, comme l'utilise la société Nanotemper. Un autre application que l'on retrouve dans l'industrie est la thermodiffusion qui, associée aux effets gravitationnels, pourrait jouer un rôle important dans la séparation des composées par gradient géothermique^{5,6} dans les réservoirs de pétrole.^{15,16}

En outre, la thermophorèse suscite un intérêt particulier dans le contexte des origines de la vie.²³ Depuis la découverte d'organismes thermophiles et des systèmes hydrothermaux dans les profondeurs océaniques à la fin des années 1970,⁴¹ l'idée selon laquelle la vie aurait pu naître au sein des systèmes hydrothermaux de la Terre primitive⁴³⁻⁴⁵ a gagné en popularité. En effet, les avantages de cet environnement sont multiples : sous-marin, il procure une protection efficace contre le bombardement intense d'astéroïdes et le rayonnement ultraviolet. Par ailleurs, l'activité volcanique est source de minéraux potentiellement catalytiques,⁴² et d'énergie chimique et thermique. Les gradients thermiques résultants induisent la thermophorèse, qui, combinée à la convection thermique, pourrait agir comme un piège et conduire à une accumulation de précurseurs.²⁴⁻²⁶ Or la concentration des réactifs est une condition nécessaire à toute synthèse prébiotique soutenue, et donc à l'apparition de la vie.^{43,46,47}

La thermophorèse serait donc un processus clef de la théorie de l'origine hydrothermale. Plusieurs travaux ont porté spécifiquement sur ces applications aux origines de la vie comme par exemple le piégeage des nucléotides,^{50,52,55,56} ou encore l'accumulation et la réplication de l'ARN⁵⁷⁻⁶¹ et de l'ADN.^{24,26,51,62,63}

Cependant, le processus d'accumulation thermique d'une molécule donnée reste incompris, et les origines moléculaires de la thermophorèse sont encore inconnues. Plusieurs modèles théoriques ont été proposés pour décrire la thermophorèse de manière phénoménologique afin d'expliquer la dépendance à plusieurs propriétés des molécules du système (la masse, la taille, la charge, la polarité, etc.).⁷ Une meilleure compréhension de cette contribution moléculaire permettrait in fine d'éclairer le processus d'accumulation des précurseurs abiotiques.

Notre objectif principal est de comprendre la nature de la thermophorèse pour les solutions aqueuses diluées. Nous étudions ici des solutions aqueuses de petites molécules amphiphiles sous un gradient thermique. Les solutés étudiés ont été choisis pour éviter les effets hydrophobes purs (agrégation de solutés), et les interactions hydrophiles trop fortes avec l'eau (notre solvant). De plus, les molécules considérées sont neutres, ce qui évite les effets électrophorétiques en plus des effets thermophorétiques.

La première partie de notre travail consiste à établir une bonne configuration pour les simulations thermophorétiques. Nous étudions l'évolution du gradient thermique, du gradient de concentration qui en résulte, et la conservation de l'énergie via un algorithme de dynamique

moléculaire hors-équilibre basé sur l'échange thermique (eHEX).⁷⁰ De manière à vérifier la robustesse de nos simulations thermophorétiques, nous testons les effets liés à la longueur de la boîte, à l'amplitude du gradient thermique, au taux d'échange de chaleur et à la dépendance du champ de force.

Dans la deuxième partie, nous analysons les conditions thermodynamiques (température, pression et concentration moyennes) et les propriétés moléculaires intrinsèques (masse, charge et soluté) qui pourraient avoir un effet sur l'accumulation de la molécule dans une région spécifique. Ces effets sont quantifiés par un coefficient appelé coefficient de Soret, et nous comparons nos résultats avec certaines théories phénoménologiques disponibles dans la littérature.

Dans la dernière partie, nous vérifions la viabilité de deux modèles, auxquels on pourrait attribuer les origines moléculaires de la thermophorèse. Le premier modèle tente de relier l'effet thermique d'une molécule à la dépendance en température d'un processus d'énergie libre de solvation.⁷¹ Le second modèle conjecture que le mouvement thermophorétique d'une molécule est lié aux énergies d'activation pour la diffusion des molécules de soluté et de solvant.⁷²⁻⁷⁶

9.2 L'effet Soret et les dynamiques de la thermophorèse

La thermophorèse d'un mélange binaire résulte d'une compétition entre la diffusion ordinaire des particules, décrite par le coefficient de diffusion (D),⁷⁷ et leur mouvement sous un gradient thermique. Ce dernier est décrit phénoménologiquement en termes de vitesse de dérive proportionnelle au gradient de température par un coefficient de diffusion thermique (D_T). Le rapport entre la diffusion de Fick et la diffusion thermique définit le coefficient de Soret (S_T). Le signe d'un tel coefficient dicte si les molécules s'accumuleront dans la région froide ou chaude, et sa valeur détermine l'amplitude du gradient de concentration :

$$S_T = \frac{D_T}{D}, \quad (9.1)$$

Dans ce travail, nous considérons un gradient thermique uni-dimensionnel linéaire et uniforme ∇T le long de l'axe z . Le flux total de particules de soluté se déplaçant à une position z $j^s(z)$ peut être phénoménologiquement écrit comme une somme du courant de diffusion $j_D(z)$ et du courant de diffusion thermique $j_{D_T}(z)$:^{8,78}

$$j^s(z) = j_D(z) + j_{D_T}(z) . \quad (9.2)$$

Le premier terme est le courant de diffusion, et est décrit par la loi de Fick pour les concentrations diluées :

$$j_D(z) = -D^s(z)\nabla c^s|_z , \quad (9.3)$$

où D^s est le coefficient de diffusion du soluté, et $\nabla c^s|_z$ est le gradient de concentration du soluté le long de z . Le deuxième terme de l'équation (3.2) est le courant de diffusion thermique, qui, pour les solutions diluées, peut s'écrire :

$$j_{D_T}(z) = -D_T^s(z)c^s\nabla T , \quad (9.4)$$

où D_T^s est le coefficient phénoménologique de diffusion thermique du soluté, c^s est la concentration du soluté, et ∇T est le gradient thermique. Les expériences montrent qu'un état stationnaire peut être atteint. Un tel état se caractérise par le fait que ses variables sont considérées comme indépendantes du temps, tout en étant hors équilibre. Ainsi, les techniques de mesures développées pour les systèmes à l'équilibre thermodynamiques peuvent être appliquées pour caractériser ses propriétés.

Cet état stationnaire peut être atteint après une période de relaxation, à condition que les perturbations externes et que les paramètres macroscopiques (température, concentration et pression en chaque point du système) soient des valeurs indépendantes du temps. Par exemple, si un flux thermique constant est appliqué à travers deux régions spatiales distinctes dans une solution aqueuse, le système tendra vers un état stationnaire. Ainsi, la théorie peut être utilisée pour déterminer le gradient de concentration à travers un gradient thermique via le coefficient de Soret.

De plus, lorsque l'état stationnaire est atteint, la diffusion ordinaire et la diffusion thermique sont équilibrées, ce qui implique que $j^s(z) = 0$. En combinant les équations (9.12) et (9.13), le coefficient de Soret dans une dimension est décrit par le logarithme de la concentration en fonction de la température :

$$\left. \frac{d \ln c^s}{dT} \right|_z = -S_T^s(z), \quad (9.5)$$

Par contre, les résultats expérimentaux montrent que le comportement de thermodiffusion d'une particule dépend d'un grand nombre de facteurs moléculaires qui ne sont pas pris en compte dans cette équation.⁷ Il faudrait considérer d'une des facteurs indépendants des interactions entre les particules comme la masse, la taille et le moment d'inertie. D'autre part, la chaleur de transfert, les liaisons hydrogènes, les interactions croisées, la force ionique et la tension interfaciale sont autant de paramètres dépendants de l'interaction des particules qui influencent aussi le coefficient de Soret. Enfin, la température, la pression et la concentration sont des facteurs externes indépendants de la nature de la solution qui ne peuvent être négligés.^{49,79} La complexité de l'effet Soret rend la compréhension de son origine moléculaire très ardue. Dans ce travail, nous avons étudié quelques contributions de l'effet de Soret dans les solutions aqueuses.

Descriptions phénoménologiques

- Le coefficient de Soret présente une dépendance à la température. Nous étudions notamment un modèle disponible dans la littérature basé sur ce paramètre, qui donne comme particularité que les valeurs du coefficient de Soret atteignent un certain point de saturation à des températures plus élevées.^{79,80}

$$S_T(T) = S_T^\infty \left[1 - \exp\left(-\frac{T^* - T}{T_i}\right) \right]. \quad (9.6)$$

- Le coefficient de Soret présente une dépendance à la concentration. Cette influence est par exemple visible lorsque, sous l'effet d'une variation de concentration, les molécules passent de la région froide à la région chaude. La molécule passe ainsi de l'état thermophile à thermophobe.^{79,80}

- Une contribution chimique qui ne dépend ni de la concentration ni de la température (S_T^{ch}) apparaît souvent dans certaines descriptions et est généralement associée aux interactions entre soluté et solvant.⁹⁰
- L'effet du rapport de masse des composants du mélange binaire masse semble aussi jouer un rôle important dans le coefficient de Soret :^{76,102,103}

$$S_T = S_T^0 + a_M \frac{M_2 - M_1}{M_2 + M_1}. \quad (9.7)$$

Modeles moleculaires

De plus, nous pouvons trouver dans la littérature quelques modèles moléculaires basés sur des théories thermodynamiques ou cinétiques qui tentent d'expliquer fondamentalement le principe du mouvement thermophorétique. Deux d'entre eux sont particulièrement intéressants et pertinents, et proviennent de perspectives différentes.

- L'approche thermodynamique d'Eastman établit les origines moléculaires du coefficient de Soret comme l'énergie libre de solvation en fonction de la température :⁷¹

$$S_T = \frac{1}{RT} \frac{dG^s}{dT}. \quad (9.8)$$

- Le point cinétique de Prigogine établit les origines moléculaires du coefficient de Soret comme un processus d'activation de la diffusion moléculaire :^{72,73}

$$S_T = \frac{E_a^s - E_a^w}{RT^2}. \quad (9.9)$$

Notre étude se concentre sur les solutions aqueuses diluées. Afin d'éviter de travailler avec de très grands systèmes, nous avons choisi de petits solutés polaires. Contrairement aux solutés totalement apolaires, ce type de molécules ne s'agrègent généralement pas à de faibles concentrations. Ce comportement est crucial pour pouvoir déterminer les variations du gradient de concentration uniquement dues au gradient de température et non renforcées par l'agrégation.

De plus, nous choisissons des molécules amphiphiles. En effet, les petites molécules purement polaires présentent un coefficient de Soret quasi nul en solution aqueuse. Pour observer une sensibilité à la thermophorèse, nous choisissons des molécules contiennent une tête hydrophile et un corps hydrophobe dont le coefficient de Soret est non négligeable.

Enfin, ces solutés doivent être suffisamment différents les uns des autres afin d'observer des comportements différents sous le gradient thermique, et donc des coefficients de Soret significativement différents.

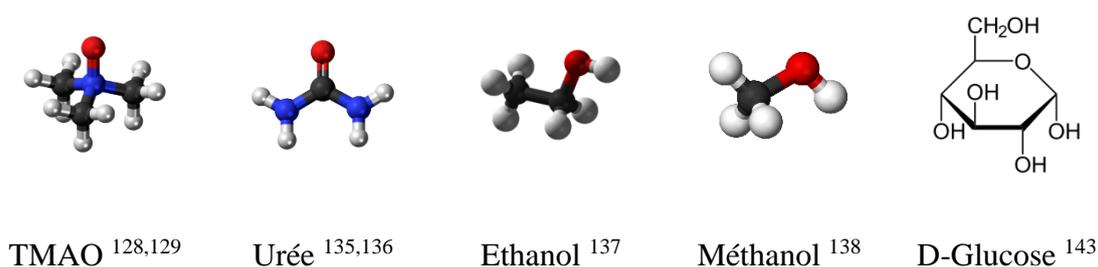


Figure 9.1: Solutés. Structures tridimensionnelles de gauche à droite pour le N-oxyle de triméthylamine (TMAO), l'urée, l'éthanol et le méthanol. Structure de Lewis du D-Glucose.

Pour le l'éluant, le modèle TIP4P/2005¹⁴⁴ a été le principal modèle d'eau pour notre travail, mais aussi nous avons effectué quelques calculs supplémentaires avec le modèle d'eau SPC/E.¹⁴⁵

Pour l'étude de la thermophorèse, nous considérons des boîtes parallélépipédiques, pour générer un gradient de température unidimensionnel le long d'une direction choisie z . Les axes orthogonaux au gradient thermophorétique sont maintenus équidistants $L_x = L_y$ et la distance L_z est typiquement plus grande que L_y et L_x . La boîte est centrée en $(0,0,0)$, avec une distribution équiangulaire droite ($90^\circ, 90^\circ, 90^\circ$), comme un cuboïde carré. Les particules de soluté sont insérées uniformément le long de la direction z du gradient thermique, ce qui donne un profil de concentration plat comme configuration initiale avec le logiciel Packmol.¹⁵⁹ En revanche, les particules de soluté sont insérées aléatoirement dans les autres deux directions (x, y) .

Dans nos simulations, les interaction coulombiennes sont tronquées à $8,5 \text{ \AA}$,¹⁴⁴ bien que le potentiel de Lennard-Jones soit tronqué à 9 \AA . Le PPPM (particule-particule-particule-

maille)¹⁶⁰ est utilisé pour le calcul des forces électrostatiques à longue portée, y compris l'approche de différenciation analytique qui lisse les forces électrostatiques à longue portée PPPM. Le système est d'abord minimisé en utilisant une version Polak-Ribiere¹⁶¹ de l'algorithme du gradient conjugué (limite de tolérance de l'énergie = 10^{-4} kcal · mol⁻¹ et limite de tolérance de la force = 10^{-6} kcal · mol⁻¹). Les vitesses sont ensuite réglées pour correspondre à la température médiane cible, et les équilibrages NPT/NVT sont effectués avec le barostat (en NPT) et thermostat (en NVT) Nose-Hoover.¹⁶²⁻¹⁶⁴ Les paramètres d'amortissement pour le barostat et le thermostat sont respectivement $P_{damp} = 1000$ fs et $T_{damp} = 100$ fs.

La prochaine étape est la génération d'un gradient de température. Nos simulations nécessitent un gradient de température stable et bien défini pour comprendre les forces thermophorétiques appliquées à nos particules et étudier leur mouvement. Nous utilisons un mécanisme d'échange de chaleur introduit pour la première fois par Hafskjold and Ikesoji.¹⁵² Cet algorithme présente toutefois des problèmes de conservation d'énergie. Une nouvelle version, connu sous le nom de « enhanced Heat EXchange algorithm » (eHEX) et lancée par Wirnsberger *et al.*,⁷⁰ a néanmoins permis de dépasser ce problème de conservation d'énergie.

Nous suivons une procédure similaire à celle du travail d'activation du gradient thermique d'eHEX⁷⁰ pour le processus d'équilibrage de nos systèmes et la génération du gradient thermique postérieur, en suivant ces étapes générales :

1. Minimization
2. Remise à l'échelle de la vitesse : conduire le système à la température cible
3. 1^{ère} simulation dans le NPT ensemble
4. Remise à l'échelle de la boîte
5. 2^{ème} simulation dans le NPT ensemble avec le volume correct
6. 1^{ère} simulation dans le NVE ensemble
7. Ajustement de l'énergie de la dernière configuration
8. 2^{ème} simulation NVE : avec la température correcte

Il convient de noter que les étapes 4 et 7 sont des processus cruciaux de remise à l'échelle du volume et de l'énergie respectivement. Ces étapes garantissent les conditions souhaitées lors des simulations d'ensemble NVE.

Une fois que la partie d'équilibrage est faite, nous définissons les régions froides et chaudes le long de l'axe z . Nous plaçons ces deux régions de façon symétrique à $z_c \approx -L_z/4$ et $z_h \approx L_z/4$ et nous appliquons l'algorithme eHEX dans ces régions pour les molécules d'eau.

La prochaine partie de notre travail consiste à s'assurer que l'énergie totale est conservée tout au long du temps de simulation. Nous nous sommes assurés que l'énergie totale ne diverge pas sur 10 ns pour un pas de temps inférieur à 2 fs. L'échelle de temps d'équilibrage de 100 ps pour les systèmes de 5 nm de longueur de boîte et de 200 ps pour les systèmes de 10 nm de longueur de boîte est suffisante pour atteindre le gradient thermique ($\Delta T \approx 60$ K) dans l'état d'équilibre.

En outre, la concentration choisie pour les solutions de TMAO est de 2,17-m, de manière à atteindre un bon compromis entre dilution suffisante pour négliger les interactions soluté-soluté, et concentration suffisante pour diminuer le temps de convergence. Pour de tels systèmes, nous estimons qu'une échelle de temps de 2-ns peut être choisie sans risque. En effet, 2ns correspond au temps nécessaire pour atteindre l'état d'équilibre du profil de concentration du soluté (tant que les conditions de la simulation et les tailles des systèmes ne diffèrent pas significativement).

De plus, nous explorons plusieurs facteurs qui pourraient affecter la stabilité des gradients thermique et de concentration, et la robustesse des résultats. Nous avons d'abord étudié l'effet de la taille du système, et nous concluons que le gradient de concentration est pratiquement insensible à la taille de la boîte. L'expression du coefficient de Soret est, par définition, une variable intensive une fois que la condition d'équilibre du gradient de concentration est atteinte pour une température médiane donnée. Cependant, nous observons que l'erreur associée au coefficient de Soret augmente avec la taille du système. Pour cette raison, le plus petit système

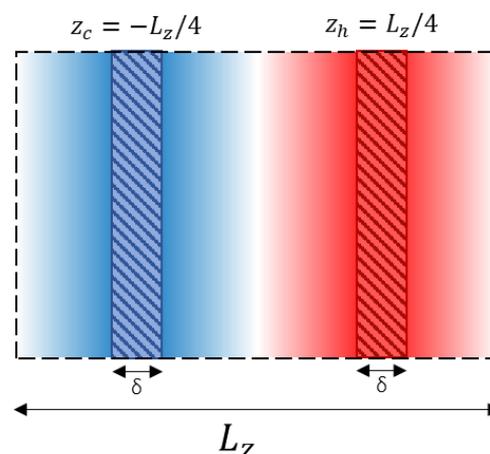


Figure 9.2: Régions thermostatées. Représentation schématique des régions thermostatées à $\pm L_z/4$ figure de droite, où deux gradients de température de distance $L_z/2$ sont générés, sans transition de discontinuité de température aux bords.

de boîtes de 50 Å est la meilleure option en termes de ressources de calcul et d'erreurs statistiques.

Deuxièmement, nous étudions l'effet de l'amplitude du gradient de température, dans un intervalle allant de 0,8 à 3,6 K · Å⁻¹. Nous n'observons pas d'effet sur le coefficient de Soret, mais nous constatons que, plus l'amplitude est grande, meilleure est la précision de la mesure du gradient de concentration engendré. Cependant, les températures visitées doivent se trouver dans, ou pas trop loin, de l'intervalle de phase liquide à l'équilibre (à savoir, entre 273 et 373 K). Pour ces raisons, nous avons choisi d'utiliser le gradient de 2,4 K · Å⁻¹ qui satisfait à ces conditions dans la plupart des cas.

Le troisième aspect est la fréquence du transfert de chaleur, qui joue un rôle important dans la concentration moléculaire à l'intérieur du réservoir. Dans les régions thermalisées, le transfert provoque de fortes variations dans les profils de concentration, c'est pourquoi nous déterminons toujours le coefficient de Soret à partir du profil de concentration à l'extérieur des régions thermalisées.

Enfin, nous testons différents champs de force pour le soluté TMAO et la molécule d'eau. Nous constatons une certaine reproductibilité pour un système donné en utilisant différents champs de force disponibles dans la littérature, à l'exception d'une combinaison de champs qui donne une disparité dans le gradient de concentration. La combinaison de champs de force du modèle d'eau TIP4P/2005¹⁴⁴ avec le modèle TMAO de Kast^{128,129} est retenue pour le reste de cette étude.

9.3 L'effet de Soret : Propriétés des molécules et conditions thermodynamiques

Nous avons présenté dans le paragraphe précédent plusieurs descriptions heuristiques issue de la littérature, qui tentent de décomposer le coefficient de Soret en différentes contributions. Dans le cadre de notre étude, nous avons choisi d'examiner plus en détail certaines de ces propriétés physiques, considérées comme les plus pertinentes pour l'effet Soret. Dans une première partie, nous nous concentrons sur des facteurs externes tels que la pression, la température et la concentration qui pourraient affecter le coefficient de Soret.

Nous observons une légère augmentation du coefficient de Soret pour les pressions moyennes élevées de 100 bar et 1000 bar, mais nous ne remarquons pas de différence pour les pressions moyennes relativement faibles de 1 bar et 10 bar.

$$S_T^{1 \text{ bar}} \approx S_T^{10 \text{ bar}} < S_T^{100 \text{ bar}} < S_T^{1 \text{ kbar}} \quad (9.10)$$

Bien que nous observions une augmentation du coefficient de Soret en fonction de la pression médiane, l'effet est très faible. Par conséquent, nous ne considérons pas que le coefficient de Soret dépend de la pression moyenne, puisque de grandes pressions doivent être appliquées pour observer un effet significatif.

Nous étudions également la dépendance de la température médiane de l'effet Soret pour l'éthanol, l'urée et le TMAO. Un des résultats est que le coefficient de Soret du TMAO augmente en fonction de la température médiane :

$$S_T^{270\text{K}} < S_T^{300\text{K}} < S_T^{330\text{K}} < S_T^{360\text{K}} < S_T^{390\text{K}} . \quad (9.11)$$

Nous observons que les coefficients de Soret pour le TMAO et l'urée vérifie une déplétion exponentielle décrite phénoménologiquement par l'équation de Iacopini et Piazza^{79,80} ($S_T = S_T^\infty [1 - \exp(T^* - T)/T_i]$), au contraire des molécules d'éthanol. Ces résultats sont en accord avec la littérature.⁸⁴ Nous en déduisons que la température a un effet sur le coefficient de Soret, mais que cet effet dépend de la nature du soluté. De manière générale, nous observons une augmentation progressive du coefficient de Soret en fonction de la température pour des températures moyennes basses, ce qui signifie que la thermophobie augmente avec la température. Les molécules de soluté tentent de s'accumuler dans la région froide lorsque la température médiane augmente dans les solutions aqueuses diluées.

Par ailleurs, nous analysons l'effet de la concentration sur le coefficient de Soret, et nous observons que les coefficients de Soret du TMAO ne présentent pas de forte dépendance avec la concentration :

$$S_T^{1,1 \text{ m}} \approx S_T^{2,2 \text{ m}} \approx S_T^{3,3 \text{ m}} > S_T^{4,3 \text{ m}} . \quad (9.12)$$

Cette tendance indique que dans les solutions aqueuses suffisamment diluées, le coefficient de Soret ne dépend pas de la concentration. Cet effet pourrait être expliqué par le fait que les molécules de TMAO n'interagissent pratiquement pas entre elles. A ce niveau de dilution, l'interaction "chimique" soluté-solvant reste invariante, et la coquille de solvation ne dépend pas du gradient de concentration.¹²⁶

Cependant, la concentration molaire $C_0^{(4)} = 4,34 \text{ mol} \cdot \text{kg}^{-1}$ est relativement élevée, où les coquilles d'hydratation se recouvrent et où les molécules de TMAO sont en contact étroit. Pour cette raison, nous commençons à observer certains effets sur le coefficient de Soret, et nous optons pour la concentration médiane $C_0^{(2)} = 2,17 \text{ mol} \cdot \text{kg}^{-1}$. Cette dernière contient suffisamment de molécules de TMAO pour obtenir de bonnes statistiques pour la détermination du coefficient de Soret tout en étant dans un régime suffisamment dilué.

Dans la deuxième partie de notre étude, nous nous intéressons aux propriétés intrinsèques telles que le moment dipolaire, la taille moléculaire ou l'effet de masse de différentes molécules (éthanol, méthanol, urée, TMAO, glucose) :

$$\mu_{Glucose} < \mu_{Ethanol} \approx \mu_{Methanol} < \mu_{Urea} < \mu_{TMAO} , \quad (9.13)$$

$$V_{Methanol} < V_{Urea} < V_{Ethanol} < V_{TMAO} < V_{Glucose} , \quad (9.14)$$

$$M_{Methanol} < M_{Ethanol} < M_{Urea} < M_{TMAO} < M_{Glucose} . \quad (9.15)$$

La formule suivante ordonne le coefficient de Soret en fonction du soluté :

$$S_t^{Methanol} \approx S_t^{Glucose} < S_t^{Ethanol} \approx S_t^{Urea} < S_t^{TMAO} . \quad (9.16)$$

De manière générale, nous observons des coefficients de Soret plus importants pour les molécules ayant des moments dipolaires plus grands, comme le TMAO ou l'urée. Cependant, une analyse plus poussée où nous imposons différents moments dipolaires pour l'urée démontre que son coefficient de Soret diminue lorsque le moment dipolaire de l'urée modifiée augmente. Des travaux précédents ont déjà discuté de cette plus grande complexité pour les systèmes polaires par rapport aux systèmes non polaires.¹⁰⁵ Les auteurs ont supposé que les changements structurels non idéaux (par exemple, les effets d'excès) doivent être pris en compte pour

comprendre l'effet de la polarité dans le coefficient de Soret. Par ailleurs, nous estimons que la remise à l'échelle des charges du champ de force induite par les modifications du moment dipolaire pourrait conduire à une description irréaliste des forces intermoléculaires et donc invalider la pertinence de nos résultats. Par conséquent, des études supplémentaires sont nécessaires pour comprendre l'évolution du coefficient de Soret en fonction du moment dipolaire d'une molécule.

De manière générale, nous observons que les valeurs les plus faibles des coefficients de Soret sont associées à de petits systèmes moléculaires en termes de masse et de taille. Plus précisément, une augmentation progressive du coefficient de Soret est observée lorsque la masse molaire et la taille des molécules augmentent, sauf pour le glucose. Ces tendances sont en accord avec les études précédentes qui ont rapporté une dépendance de la taille dans les solutions colloïdales,^{49,68,177} et un effet de masse dans les n-alcane linéaires.⁹⁸⁻¹⁰⁰

Cependant, la corrélation et l'intrication de ces contributions intrinsèques rend difficile l'interprétation physique de leurs contributions respectives pour l'effet Soret. Pour cette raison, nous examinons l'effet des molécules isotopiques dans le coefficient de Soret afin de mettre en évidence l'effet de masse.

Dans la littérature, le coefficient de Soret est généralement divisé de manière phénoménologique en deux termes indépendants ($S_T = S_T^{ch} + S_T^{iso}$),^{76,102} où S_T^{iso} est la contribution isotopique qui ne dépend que de la masse et du moment d'inertie. Dans ce travail, la description de l'effet de masse pour les petits solutés en solutions aqueuses est approximée par cette équation :

$$S_T = S_T^0 + a_M \frac{M_2 - M_1}{M_2 + M_1}. \quad (9.17)$$

Où $\delta M = M_2 - M_1 / M_2 + M_1$. Nous supposons que l'effet du moment d'inertie n'affecte pas de manière significative le coefficient de Soret isotopique lorsque le rapport de masse augmente.

Nos résultats montrent que toutes les molécules étudiées présentent un comportement thermophile pour des rapports de masse entre soluté et solvant positifs à leur masse naturelle. De plus, l'urée, le méthanol et l'eau ont acquis un caractère thermophile pour certains rapports de masse négatifs. Nous confirmons donc que les molécules plus lourdes ont des valeurs de coefficient de Soret plus élevées, et nous démontrons une relation linéaire entre le coefficient de Soret et le rapport de masse (δM) (les lignes pointillées dans la Figure 9.3 montrent l'ajustement linéaire de l'éq. (8.8)).

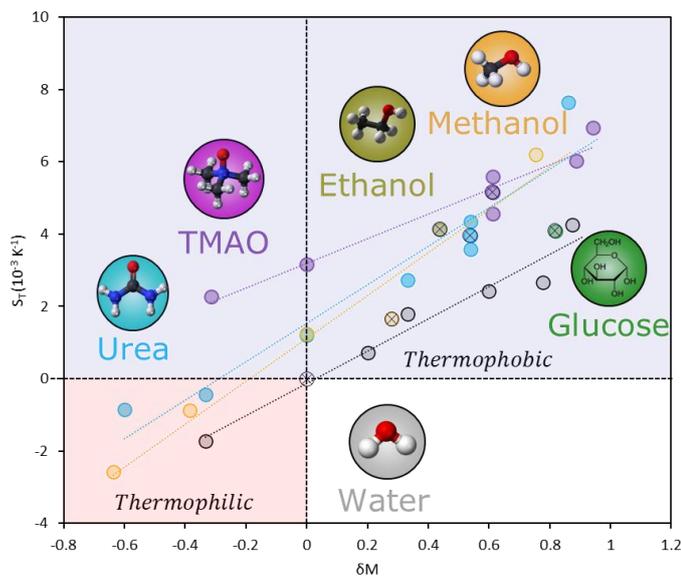


Figure 9.3: Résumé de l'effet de la masse. méthanol (jaune), éthanol (olive), urée (cyan), TMAO (magenta), glucose (vert), and eau (gris).

Si le coefficient de Soret dépend bien du rapport de masse, nous avons montré qu'il n'est pas vraiment affecté par la masse absolue de la solution. Ces résultats sont représentés par trois points verticaux pour le TMAO et l'urée dans la Figure 8.1.

De plus, nous avons voulu mettre en évidence que $S_T^0 \approx 0 \text{ K}^{-1}$ pour les simulations de l'eau seule. Pour étudier les effets de masse de l'eau dans l'eau, nous modifions la masse de certaines molécules d'eau, et calculons le coefficient de Soret de ces isotopes. Nous constatons un effet de masse pure, attestant de la valeur nulle du S_T^0 de l'eau.

Enfin, nous observons que S_T^0 est plus grand pour la molécule de TMAO par rapport à l'urée ou au méthanol, ce qui indique une plus grande contribution "chimique" de celui-ci, ou un moment d'effet. Cependant, une étude approfondie doit être menée pour connaître l'effet du moment d'inertie afin d'isoler cette contribution "chimique".

9.4 Origins moléculaires de la thermophorèse

La dernière partie de ce travail consiste à tester la faisabilité de deux modèles plausibles, qui présentent un intérêt et une pertinence particuliers pour les systèmes moléculaires, mais qui proviennent de perspectives différentes.

Le premier modèle est l'approche d'Eastman,⁷¹ qui tente de relier l'effet thermodynamique d'une molécule à la dépendance de la température d'un processus d'énergie libre de solvation. Le groupe de Braun,⁴⁹ parti cette même idée, développa et testa expérimentalement un modèle plus aboutit. Leurs résultats montrent que les états stables thermophorétiques pourraient être décrits par une distribution de Boltzmann dans l'équilibre thermodynamique local de l'équation :

$$S_T = \frac{1}{RT} \frac{dG^s}{dT}. \quad (9.18)$$

La corrélation entre le coefficient de Soret déterminé par le modèle de Braun et les coefficients de Soret précédemment calculés lors des simulations de thermophorèse est représenté ci-dessous :

Les coefficients de Soret déterminés par les calculs d'énergie libre surestiment clairement le coefficient de Soret calculé par dynamique moléculaire sous un gradient thermique pour tous les solutés. De plus, ce modèle ne tient pas compte de l'effet de masse déjà ratifié et démontré expérimentalement.⁹⁸⁻¹⁰⁰ Ainsi, le coefficient de Soret ne peut pas être compris uniquement en

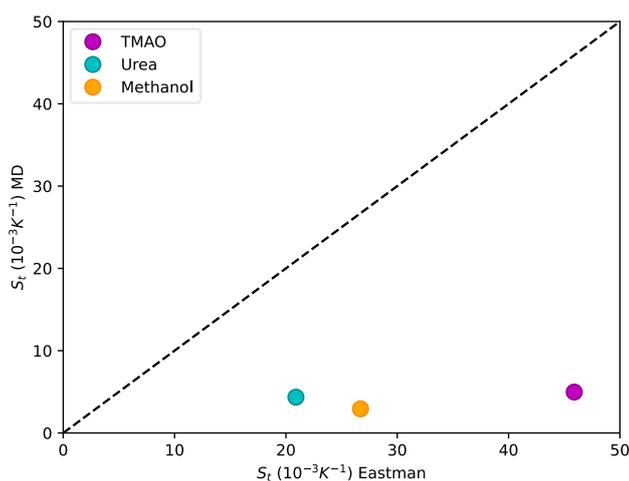


Figure 9.4: Modèle d'Eastman. Le coefficient de Soret déterminé à partir du modèle d'Eastman vs le coefficient de Soret déterminé à partir de la thermophorèse à 330 K et pour les solutés : TMAO (magenta), urée (cyan) et méthanol (jaune).

termes d'effet de solvation entropique pour les solutions aqueuses diluées, puisqu'il surestimerait largement la valeur de thermophorèse obtenue précédemment.

Le deuxième modèle est celui de Prigogine, qui suppose que le mouvement thermophorétique d'une molécule est lié aux énergies d'activation pour la diffusion des molécules de soluté et de solvant.⁷²⁻⁷⁶ Le coefficient de diffusion est obtenu à partir du déplacement quadratique moyenne (MSD).

Les coefficients de Soret déterminés par le modèle diffusif de Prigogine sont comparés dans la Figure 8.3 à ceux obtenus par les simulations de thermophorèse. Aucune tendance ne se dégage : la masse et la taille pourraient avoir un effet sur leurs énergies d'activation, puisque des molécules plus grandes ou plus lourdes rendraient leur mouvement plus coûteux en termes d'énergie.

Enfin, nous analysons la correction au modèle proposée par Artola, Rousseau et Galliéro,⁷⁵ qui ajoute l'effet de masse à l'approche de Prigogine. Nous observons que la prise en compte de ce terme supplémentaire entraîne une augmentation exponentielle de la valeur du coefficient de Soret en fonction du rapport de masse, ce qui ne correspond pas à l'effet de masse observé dans les résultats de thermophorèse.

Cet écart pourrait être expliqué par la non prise en de certains effets, notamment mentionnés par les auteurs à la fin de leur article. Les moments d'inertie dans le processus de diffusion, qui pourraient faire partie de l'effet isotopique, devraient également être étudiés. En conclusion,

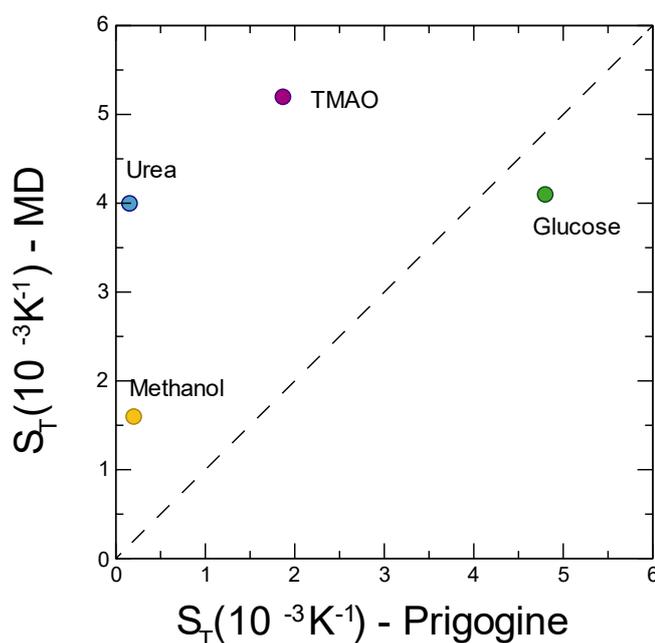


Figure 9.5: Modèle de Prigogine. Le coefficient de Soret déterminé à partir pour le modèle de Prigogine vs le coefficient de Soret déterminé à partir de la thermophorèse à 330 K, pour les solutés : TMAO (magenta), urée (cyan), méthanol (jaune) et glucose (vert).

une étude plus détaillée doit être menée afin de comprendre la thermophorèse comme un processus de diffusion activé.

Le dernier axe de ce travail propose de considérer le coefficient de Soret comme un effet balistique du mouvement moléculaire. Ce mouvement peut être déterminé via les déplacements moléculaires à temps très courtes de la dynamique avec des calculs MSD.

Le coefficient de Soret dans le régime balistique est déterminé pour plusieurs molécules, prises à différentes masses. Nous considérons que 170 fs comme la limite de la région balistique. Nous observons une corrélation linéaire entre le coefficient de Soret calculé par les simulations de thermophorèse et la MSD balistique à 170 fs. Cette corrélation est présente que dans le régime balistique.

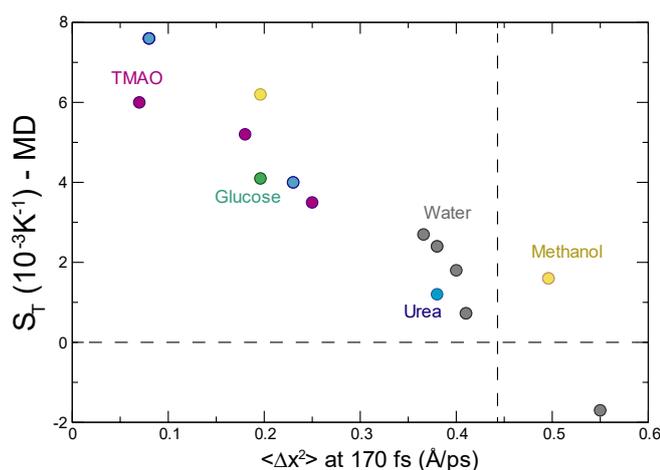


Figure 9.6: Régime balistique. Coefficient de Soret déterminé à partir de la thermophorèse vs MSD ($\langle \Delta x^2 \rangle$) déplacement jusqu'à 170 (Å/ps) à 330 K et pour les solutés : TMAO (magenta), urée (cyan), méthanol (jaune), glucose (vert) et eau (gris).

Ces résultats pourraient indiquer que les forces thermophorétiques sont en fait appliquées sur des échelles de temps très courtes. Ainsi, le coefficient de Soret devrait être compris comme un effet cinétique et non comme un effet diffusif. Cependant, ces résultats doivent être considérées comme préliminaires, et d'autres études seront nécessaires pour interpréter la relation entre coefficient de Soret et MSD balistique.

References

1. Ludwig, C. *Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösung*. (Aus der KK Hof-und Staatsdruckerei, in Commission bei W. Braumüller ..., 1856).
2. Rahman, M. A. & Saghir, M. Z. Thermodiffusion or Soret effect: Historical review. *Int. J. Heat Mass Transf.* **73**, 693–705 (2014).
3. Duparc, L. Charles Soret. *Arch. Des Sci. Phys. Nat.* **4**, (1904).
4. Calero, C., Sibert, E. L. & Rey, R. Self-thermophoresis at the nanoscale using light induced solvation dynamics. *Nanoscale* **12**, 7557–7562 (2020).
5. Crookes, W. XV. On attraction and repulsion resulting from radiation. *Philos. Trans. R. Soc.* **164**, 501–527 (1873).
6. Kavokine, N. *et al.* Ultrafast photomechanical transduction through thermophoretic implosion. *Nat. Commun.* **11**, 50 (2020).
7. Niether, D. & Wiegand, S. Thermophoresis of biological and biocompatible compounds in aqueous solution. *J. Phys. Condens. Matter* **31**, 503003 (2019).
8. Duhr, S. & Braun, D. Thermophoretic depletion follows boltzmann distribution. *Phys. Rev. Lett.* **96**, 168301 (2006).
9. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077**, 101–113 (2014).
10. Gupta, A. J., Duhr, S. & Baaske, P. Microscale Thermophoresis (MST). in *Encyclopedia of Biophysics* (eds. Roberts, G. & Watts, A.) 1–5 (Springer Berlin Heidelberg, 2018). doi:10.1007/978-3-642-35943-9_10063-1
11. XanTec bioanalytics. Comparison of biomolecular interaction techniques SPR, ITC, MST and BLI. 4 (2015).

12. Zeng, S., Baillargeat, D., Ho, H. P. & Yong, K. T. Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. *Chem. Soc. Rev.* **43**, 3426–3452 (2014).
13. Wadsö, I. Design and Testing of a Micro Reaction Calorimete. *Acta Chem. Scand.* **22**, 927–937 (1968).
14. Suurkuusk, J. & Wadsö, I. Design and testing of an improved precise drop calorimeter for the measurement of the heat capacity of small samples. *J. Chem. Thermodyn.* **6**, 667–679 (1974).
15. Whitson, C. & Belery, P. Compositional Gradients in Petroleum Reservoirs. in *Proceedings of University of Tulsa Centennial Petroleum Engineering Symposium* 443–459 (Society of Petroleum Engineers, 1994).
16. Montel, F., Bickert, J., Lagisquet, A. & Galliéro, G. Initial state of petroleum reservoirs: A comprehensive approach. *J. Pet. Sci. Eng.* **58**, 391–402 (2007).
17. Ghorayeb, K., Firoozabadi, A. & Anraku, T. Interpretation of the unusual fluid distribution in the Yufutsu gas-condensate field. *SPE J.* **8**, 114–123 (2003).
18. Touzet, M. *et al.* Thermodiffusion: From microgravity experiments to the initial state of petroleum reservoirs. *Comptes Rendus - Mec.* **339**, 318–323 (2011).
19. Galliero, G. *et al.* Impact of Thermodiffusion on the Initial Vertical Distribution of Species in Hydrocarbon Reservoirs. *Microgravity Sci. Technol.* **28**, 79–86 (2016).
20. Galliero, G. *et al.* Thermodiffusion in multicomponent n-alkane mixtures. *npj Microgravity* **3**, 20 (2017).
21. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1**, 100 (2010).
22. Chen, J. *et al.* Thermal gradient induced tweezers for the manipulation of particles and cells. *Sci. Rep.* **6**, 35814 (2016).
23. Gaeta, F. S., Bencivenga, U., Canciglia, P., Rossi, S. & Mita, D. G. Temperature

- gradients and prebiological evolution. *Cell Biophys.* **10**, 103–125 (1987).
24. Braun, D., Goddard, N. L. & Libchaber, A. Exponential DNA replication by laminar convection. *Phys. Rev. Lett.* **91**, 158103 (2003).
 25. Braun, D. & Libchaber, A. Thermal force approach to molecular evolution. *Phys. Biol.* **1**, P1-8 (2004).
 26. Mast, C. B. & Braun, D. Thermal trap for DNA replication. *Phys. Rev. Lett.* **104**, 188102 (2010).
 27. Kitadai, N. & Maruyama, S. Origins of building blocks of life: A review. *Geosci. Front.* **9**, 1117–1153 (2018).
 28. Luisi, P. L. About various definitions of life. *Orig. Life Evol. Biosph.* **28**, 613–622 (1998).
 29. About | Life Detection | Research | Astrobiology. Available at: <https://astrobiology.nasa.gov/research/life-detection/about/>. (Accessed: 6th September 2021)
 30. Gilbert Walter. The RNA world Superlattices point ahead. *Nature* **319**, 618 (1986).
 31. Henrique Rampelotto, P. Panspermia: A promising field of research. in *38th COSPAR Scientific Assembly* **38**, 2 (2010).
 32. Schopf, J. W. Fossil evidence of Archaean life. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 869–885 (2006).
 33. Schopf, J. W., Kudryavtsev, A. B., Czaja, A. D. & Tripathi, A. B. Evidence of Archean life: Stromatolites and microfossils. *Precambrian Res.* **158**, 141–155 (2007).
 34. Djokic, T., VanKranendonk, M. J., Campbel, K. A., Walter, M. R. & Ward, C. R. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. *Nat. Commun.* **8**, 15263 (2017).
 35. Oparin, A. I. *El origen de la vida*. (Akal, 2015).
 36. Miller, S. L. Production of Some Organic Compounds under Possible Primitive Earth

- Conditions. *Science* (80-.). **117**, 528–529 (1953).
37. Bada, J. L. & Lazcano, A. Perceptions of science: Prebiotic soup - Revisiting the Miller experiment. *Science* (80-.). **300**, 745–746 (2003).
 38. Abelson, H. P. Chemical events on the primitive earth. *Proc. Natl. Acad. Sci.* **55**, 1365–1372 (1966).
 39. Chyba, C. F. Rethinking earth's early atmosphere. *Science* (80-.). **308**, 962–963 (2005).
 40. champagne_vent_hirez.jpg (3000×2250). Available at: https://oceanexplorer.noaa.gov/explorations/04fire/logs/hirez/champagne_vent_hirez.jpg. (Accessed: 6th September 2021)
 41. Corliss, J. B. *et al.* Submarine Thermal Springs on the Galapagos. *Science* (80-.). **203**, 1073–1083 (1979).
 42. Baross, J. A. & Hoffman, S. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig. Life* **15**, 327–345 (1985).
 43. Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805–814 (2008).
 44. Takai, K. & Nakamura, K. Archaeal diversity and community development in deep-sea hydrothermal vents. *Curr. Opin. Microbiol.* **14**, 282–291 (2011).
 45. Sousa, F. L. *et al.* Early bioenergetic evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20130088 (2013).
 46. Früh-green, G. L. *et al.* 30,000 Years of Hydrothermal Activity at the Lost City Vent Field. *Am. Assoc. Adv. Sci.* **301**, 341–342 (1881).
 47. Akanuma, S. *et al.* Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11067–11072 (2013).
 48. Duhr, S. & Braun, D. Two-dimensional colloidal crystals formed by thermophoresis

- and convection. *Appl. Phys. Lett.* **86**, 131921 (2005).
49. Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci.* **103**, 19678–19682 (2006).
 50. Mast, C. B., Schink, S., Gerland, U. & Braun, D. Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci.* **110**, 8030–8035 (2013).
 51. Braun, D. & Libchaber, A. Trapping of DNA by Thermophoretic Depletion and Convection. *Phys. Rev. Lett.* **89**, 188103 (2002).
 52. Agerschou, E. D., Mast, C. B. & Braun, D. Emergence of Life from Trapped Nucleotides? Non-Equilibrium Behavior of Oligonucleotides in Thermal Gradients. *Synlett* **28**, 56–63 (2017).
 53. Martin, W. & Russell, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 1887–1925 (2007).
 54. Wächtershäuser, G. Before Enzymes and Templates: Theory of Surface Metabolism. *Microbiol. Rev.* **52**, 452–484 (1988).
 55. Islam, S., Bučar, D. K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).
 56. Šponer, J. E. *et al.* Prebiotic synthesis of nucleic acids and their building blocks at the atomic level-merging models and mechanisms from advanced computations and experiments. *Phys. Chem. Chem. Phys.* **18**, 20047–20066 (2016).
 57. Salditt, A. *et al.* Thermal Habitat for RNA Amplification and Accumulation. *Phys. Rev. Lett.* **125**, 48104 (2020).
 58. Baaske, P. *et al.* Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9346–9351 (2007).
 59. Morasch, M., Braun, D. & Mast, C. B. Heat-Flow-Driven Oligonucleotide Gelation Separates Single-Base Differences. *Angew. Chemie Int. Ed.* **55**, 6676–6679 (2016).

-
60. Obermayer, B., Krammer, H., Braun, D. & Gerland, U. Emergence of information transmission in a prebiotic RNA reactor. *Phys. Rev. Lett.* **107**, 018101 (2011).
 61. Krammer, H., Möller, F. M. & Braun, D. Thermal, autonomous replicator made from transfer RNA. *Phys. Rev. Lett.* **108**, 238104 (2012).
 62. Reineck, P., Wienken, C. J. & Braun, D. Thermophoresis of single stranded DNA. *Electrophoresis* **31**, 279–286 (2010).
 63. Reichl, M., Herzog, M., Götz, A. & Braun, D. Why charged molecules move across a temperature gradient: The role of electric fields. *Phys. Rev. Lett.* **112**, 198101 (2014).
 64. Morasch, M. *et al.* Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nat. Chem.* **11**, 779–788 (2019).
 65. Kudella, P. W. *et al.* Fission of Lipid-Vesicles by Membrane Phase Transitions in Thermal Convection. *Sci. Rep.* **9**, 18808 (2019).
 66. Piazza, R. & Parola, A. Thermophoresis in colloidal suspensions. *J. Phys. Condens. Matter* **20**, 153102 (2008).
 67. Talbot, E. L., Kotar, J., Parolini, L., Di Michele, L. & Cicuta, P. Thermophoretic migration of vesicles depends on mean temperature and head group chemistry. *Nat. Commun.* **8**, 15351 (2017).
 68. Dhont, J. K. G., Wiegand, S., Duhr, S. & Braun, D. Thermodiffusion of charged colloids: Single-particle diffusion. *Langmuir* **23**, 1674–1683 (2007).
 69. Di Lecce, S. & Bresme, F. Thermal Polarization of Water Influences the Thermoelectric Response of Aqueous Solutions. *J. Phys. Chem. B* **122**, 1662–1668 (2018).
 70. Wirnsberger, P., Frenkel, D. & Dellago, C. An enhanced version of the heat exchange algorithm with excellent energy conservation properties. *J. Chem. Phys.* **143**, 124104 (2015).
 71. Eastman, E. D. Thermodynamics of non-isothermal systems. *J. Am. Chem. Soc.* **48**, 1482–1493 (1926).

-
72. Prigogine, I., De Brouckere, L. & Amand, R. Recherches sur la thermodiffusion en phase liquide. *Physica* **16**, 577–598 (1950).
73. Prigogine, I., De Brouckere, L. & Amand, M. R. Recherches sur la thermodiffusion en phase liquide. *Physica* **16**, 851–860 (1950).
74. Artola, P.-A. & Rousseau, B. Thermal diffusion in simple liquid mixtures: what have we learnt from molecular dynamics simulations? *Mol. Phys.* **111**, 3394–3403 (2013).
75. Artola, P. A., Rousseau, B. & Galliéro, G. A new model for thermal diffusion: Kinetic approach. *J. Am. Chem. Soc.* **130**, 10963–10969 (2008).
76. Debuschewitz, C. & Köhler, W. Molecular origin of thermal diffusion in benzene+cyclohexane mixtures. *Phys. Rev. Lett.* **87**, 055901 (2001).
77. Fick, A. V. On liquid diffusion. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **10**, 30–39 (1855).
78. Würger, A. Is Soret equilibrium a non-equilibrium effect? *Comptes Rendus - Mec.* **341**, 438–448 (2013).
79. Niether, D., Di Lecce, S., Bresme, F. & Wiegand, S. Unravelling the hydrophobicity of urea in water using thermodiffusion: Implications for protein denaturation. *Phys. Chem. Chem. Phys.* **20**, 1012–1020 (2018).
80. Iacopini, S. & Piazza, R. Thermophoresis in protein solutions. *Europhys. Lett.* **63**, 247–253 (2003).
81. Iacopini, S., Rusconi, R. & Piazza, R. The ‘macromolecular tourist’: Universal temperature dependence of thermal diffusion in aqueous colloidal suspensions. *Eur. Phys. J. E* **19**, 59–67 (2006).
82. Kishikawa, Y., Wiegand, S. & Kita, R. Temperature dependence of soret coefficient in aqueous and nonaqueous solutions of pullulan. *Biomacromolecules* **11**, 740–747 (2010).
83. Niether, D., Afanasenkau, D., Dhont, J. K. G. & Wiegand, S. Accumulation of Formamide in hydrothermal pores to Form prebiotic nucleobases. *Proc. Natl. Acad.*

- Sci. U. S. A.* **113**, 4272–4277 (2016).
84. Königer, A., Meier, B. & Köhler, W. Measurement of the Soret, diffusion, and thermal diffusion coefficients of three binary organic benchmark mixtures and of ethanol/water mixtures using a beam deflection technique. *Philos. Mag. Philos. Mag. Lett.* **89**, 907–923 (2009).
85. Polyakov, P., Müller-Plathe, F. & Wiegand, S. Reverse nonequilibrium molecular dynamics calculation of the Soret coefficient in liquid heptane/benzene mixtures. *J. Phys. Chem. B* **112**, 14999–15004 (2008).
86. Yeganegi, S. Temperature dependence of thermal diffusion factor for isotopic binary mixtures by non-equilibrium molecular dynamics simulation. *J. Phys. Soc. Japan* **72**, 2260–2264 (2003).
87. Simon, J. M., Dysthe, D. K., Fuchs, A. H. & Rousseau, B. Thermal diffusion in alkane binary mixtures: A molecular dynamics approach. *Fluid Phase Equilib.* **150**, 151–159 (1998).
88. Rousseau, B., Nieto-Draghi, C. & Avalos, J. B. The role of molecular interactions in the change of sign of the Soret coefficient. *Europhys. Lett.* **67**, 976–982 (2004).
89. Nieto-Draghi, C., Ávalos, J. B. & Rousseau, B. Computing the Soret coefficient in aqueous mixtures using boundary driven nonequilibrium molecular dynamics. *J. Chem. Phys.* **122**, 114503 (2005).
90. Artola, P. A. & Rousseau, B. Microscopic interpretation of a pure chemical contribution to the soret effect. *Phys. Rev. Lett.* **98**, 125901 (2007).
91. Wittko, G. & Köhler, W. On the temperature dependence of thermal diffusion of liquid mixtures. *Europhys. Lett.* **78**, 46007 (2007).
92. Niether, D. *et al.* Role of Hydrogen Bonding of Cyclodextrin-Drug Complexes Probed by Thermodiffusion. *Langmuir* **33**, 8483–8492 (2017).
93. Semenov, S. & Schimpf, M. Thermophoresis of dissolved molecules and polymers: Consideration of the temperature-induced macroscopic pressure gradient. *Phys. Rev. E*

- *Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **69**, 011201 (2004).
94. Saslawsky, J. J. Die Dichte der Flüssigkeit und deren Temperatur. *Zeitschrift für Phys. Chemie* **109U**, 111–135 (1924).
95. Clusius, K. & Dickel, G. Das Trennrohrverfahren bei Flüssigkeiten. *Naturwissenschaften* **27**, 148–149 (1939).
96. Galliéro, G., Duguay, B., Caltagirone, J. P. & Montel, F. Thermal diffusion sensitivity to the molecular parameters of a binary equimolar mixture, a non-equilibrium molecular dynamics approach. *Fluid Phase Equilib.* **208**, 171–188 (2003).
97. Chapman, S. The characteristics of thermal diffusion. *Proc. R. Soc. Lond.* **177**, 38–62 (1940).
98. Madariaga, J. A., Santamaría, C., Bou-Ali, M. M., Urteaga, P. & Alonso De Mezquia, D. Measurement of thermodiffusion coefficient in n -alkane binary mixtures: Composition dependence. *J. Phys. Chem. B* **114**, 6937–6942 (2010).
99. Blanco, P. *et al.* Thermodiffusion coefficients of binary and ternary hydrocarbon mixtures. *J. Chem. Phys.* **132**, 114506 (2010).
100. Alonso de Mezquia, D., Mounir Bou-Ali, M., Madariaga, J. A. & Santamaría, C. Mass effect on the Soret coefficient in n-alkane mixtures. *J. Chem. Phys.* **140**, 84503 (2014).
101. Polyakov, P., Luettmer-Strathmann, J. & Wiegand, S. Study of the thermal diffusion behavior of alkane/benzene mixtures by thermal diffusion forced rayleigh scattering experiments and lattice model calculations. *J. Phys. Chem. B* **110**, 26215–26224 (2006).
102. Wittko, G. & Köhler, W. Universal isotope effect in thermal diffusion of mixtures containing cyclohexane and cyclohexane-d₁₂. *J. Chem. Phys.* **123**, 14506 (2005).
103. Köhler, W. & Morozov, K. I. The Soret Effect in Liquid Mixtures - A Review. *J. Non-Equilibrium Thermodyn.* **41**, 151–197 (2016).
104. Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* **248**, 127–136 (1881).

-
105. Niether, D., Kriegs, H., Dhont, J. K. G. & Wiegand, S. Peptide model systems: Correlation between thermophilicity and hydrophilicity. *J. Chem. Phys.* **149**, 44506 (2018).
 106. Korsching, H. & Wirtz, K. Zur Trennung von Flüssigkeitsgemischen im Clusiusschen Trennrohr (Trennung der Zinkisotope). *Naturwissenschaften* **27**, 367–368 (1939).
 107. Prigogine, I., De Brouckere, L. & Buess, R. Recherches sur la thermodiffusion en phase liquide: thermodiffusion de l'eau lourde. *Physica* **18**, 915–920 (1952).
 108. Korsching, H. Trennung von Benzol und Hexadeuterobenzol durch Thermodiffusion in der Flüssigkeit. *Naturwissenschaften* **31**, 348–349 (1943).
 109. Ning-Yuan, R. M. & Beyerlein, A. L. The $2H$ thermal diffusion isotope effect in benzene and methanol. *J. Chem. Phys.* **78**, 7010–7011 (1983).
 110. Rutherford, W. M. Separation of isotopically substituted liquids in the thermal diffusion column. *J. Chem. Phys.* **59**, 6061–6069 (1973).
 111. Rutherford, W. M. Effect of mass distribution on the isotopic thermal diffusion of benzene. *J. Chem. Phys.* **86**, 5217–5218 (1987).
 112. Rutherford, W. M. Effect of carbon and hydrogen isotopic substitutions on the thermal diffusion of benzene. *J. Chem. Phys.* **90**, 602–603 (1989).
 113. Wittko, G. & Köhler, W. Influence of isotopic substitution on the diffusion and thermal diffusion coefficient of binary liquids. *Eur. Phys. J. E* **21**, 283–291 (2006).
 114. Galliero, G., Bugel, M., Duguay, B. & Montel, F. Mass effect on thermodiffusion using molecular dynamics. *J. Non-Equilibrium Thermodyn.* **32**, 251–258 (2007).
 115. Artola, P.-A. & Rousseau, B. Isotopic Soret effect in ternary mixtures: Theoretical predictions and molecular simulations. *J. Chem. Phys.* **143**, 174503 (2015).
 116. Reith, D. & Müller-Plathe, F. On the nature of thermal diffusion in binary Lennard-Jones liquids. *J. Chem. Phys.* **112**, 2436–2443 (2000).
 117. Iftimie, R., Minary, P. & Tuckerman, M. E. Ab initio molecular dynamics: Concepts,

- recent developments, and future trends. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6654–6659 (2005).
118. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B871 (1964).
119. Liang, S., Busiello, D. M. & Rios, P. D. L. The intrinsic non-equilibrium nature of thermophoresis. 1–16 (2021).
120. Sharp, K. & Matschinsky, F. Translation of Ludwig Boltzmann’s paper ‘on the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium’ Sitzungberichte der kaiserlichen akademie d. *Entropy* **17**, 1971–2009 (2015).
121. Frenkel, D. B. S. *Understanding molecular simulation: from algorithms to applications. Vol. 1. Elsevier.* (Academic Press, 2001).
122. Birkhoff, G. D. Proof of the Ergodic Theorem. *Proc. Natl. Acad. Sci.* **17**, 656–660 (1931).
123. Onsager, L. Reciprocal Relations in Irreversible Processes. I. *Phys. Rev.* **37**, 405–426 (1931).
124. Mondal, J., Stirnemann, G. & Berne, B. J. When Does Trimethylamine N-Oxide Fold a Polymer Chain and Urea Unfold It? *J. Phys. Chem. B* **117**, 8723–8732 (2013).
125. Laage, D., Stirnemann, G. & Hynes, J. T. Why water reorientation slows without iceberg formation around hydrophobic solutes. *J. Phys. Chem. B* **113**, 2428–2435 (2009).
126. Paul, S. & Patey, G. N. Why tert-butyl alcohol associates in aqueous solution but trimethylamine-N-oxide does not. *J. Phys. Chem. B* **110**, 10514–10518 (2006).
127. Fiore, A., Venkateshwaran, V. & Garde, S. Trimethylamine N-oxide (TMAO) and tert-butyl alcohol (TBA) at hydrophobic interfaces: Insights from molecular dynamics simulations. *Langmuir* **29**, 8017–8024 (2013).
128. Kast, K. M., Brickmann, J., Kast, S. M. & Berry, R. S. Binary phases of aliphatic N-

- oxides and water: Force field development and molecular dynamics simulation. *J. Phys. Chem. A* **107**, 5342–5351 (2003).
129. Hölzl, C. *et al.* Design principles for high–pressure force fields: Aqueous TMAO solutions from ambient to kilobar pressures. *J. Chem. Phys.* **144**, 144104 (2016).
130. Courtenay, E. S., Capp, M. W., Anderson, C. F. & Record, M. T. Vapor pressure osmometry studies of osmolyte-protein interactions: Implications for the action of osmoprotectants in vivo and for the interpretation of ‘osmotic stress’ experiments in vitro. *Biochemistry* **39**, 4455–4471 (2000).
131. Anfinsen. Folding of Protein Chains. *Science (80-.)*. **181**, 223–230 (1973).
132. Booth, J. J., Abbott, S. & Shimizu, S. Mechanism of hydrophobic drug solubilization by small molecule hydrotropes. *J. Phys. Chem. B* **116**, 14915–14921 (2012).
133. Powner, M., Sutherland, J. & Szostak, J. The Origins of Nucleotides. *Synlett* **14**, 1956–1964 (2011).
134. Lohrmann, R. & Orgel, L. E. Prebiotic synthesis: Phosphorylation in aqueous solution. *Science (80-.)*. **161**, 64–66 (1968).
135. Weerasinghe, S. & Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *J. Phys. Chem. B* **107**, 3891–3898 (2003).
136. Chiba, S., Furuta, T. & Shimizu, S. Kirkwood-Buff Integrals for Aqueous Urea Solutions Based upon the Quantum Chemical Electrostatic Potential and Interaction Energies. *J. Phys. Chem. B* **120**, 7714–7723 (2016).
137. VANOMMESLAEGH, K. *et al.* CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **31**, 671–690 (2009).
138. Gonzalez-Salgado, D. & Vega, C. A new intermolecular potential for simulations of methanol: The OPLS/2016 model. *J. Chem. Phys.* **145**, 034508 (2016).
139. Franklin, B. H. & J., H. P. Reaction of Monosaccharides with Proteins: Possible Evolutionary Significance. *Science (80-.)*. **213**, 222–224 (1981).

-
140. Hull, P. *Glucose syrups: technology and applications*. (John Wiley & Sons, 2010).
141. Garrett, R. H. & Grisham, C. M. *Biochemistry*. (Cengage Learning, 2016).
142. Batista, M. L. S., Pérez-Sánchez, G., Gomes, J. R. B., Coutinho, J. A. P. & Maginn, E. J. Evaluation of the GROMOS 56A Force Field for the Calculation of Structural, Volumetric, and Dynamic Properties of Aqueous Glucose Systems. *J. Phys. Chem. B* **119**, 15310–15319 (2015).
143. Jamali, S. H., Westen, T. Van, Moulton, O. A. & Vlugt, T. J. H. Optimizing Nonbonded Interactions of the OPLS Force Field for Aqueous Solutions of Carbohydrates: How to Capture Both Thermodynamics and Dynamics. *J. Chem. Theory Comput.* **14**, 6690–6700 (2018).
144. Abascal, J. L. F. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123**, 234505 (2005).
145. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
146. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
147. Andersen, H. C. Molecular Dynamics Simulations at Constant Temperature and Pressure. *Comput. Simul. Mater. Sci.* **72**, 2384 (1980).
148. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17**, 1302–1322 (1978).
149. Evans, D. J. & Holian, B. L. The Nose-Hoover thermostat. *J. Chem. Phys.* **83**, 4069–4074 (1985).
150. Di Lecce, S., Albrecht, T. & Bresme, F. A computational approach to calculate the heat of transport of aqueous solutions. *Sci. Rep.* **7**, 44833 (2017).
151. Wang, Z., Wiegand, S. & Bresme, F. Alkali Halide Solutions under Thermal Gradients: Soret Coefficients and Heat Transfer Mechanisms. *J. Phys. Chem. B* **117**,

- 8209–8222 (2013).
152. Hafskjold, B., Ikeshoji, T. & Ratkje, S. K. On the molecular mechanism of thermal diffusion in liquids. *Mol. Phys.* **80**, 1389–1412 (1993).
 153. Webb III (Sandia National Laboratories, Albuquerque, NM), E. B. *et al.* A robust, coupled approach for atomistic-continuum simulation. (2004).
 154. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).
 155. Strang, G. On the Construction and Comparison of Difference Schemes. *SIAM J. Numer. Anal.* **5**, 506–517 (1968).
 156. LeVeque, R. J. & Olinger, J. Numerical Methods Based on Additive Splittings for Hyperbolic Partial Differential Equations. *Math. Comput.* **40**, 469–497 (1983).
 157. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
 158. LAMMPS Molecular Dynamics Simulator. Available at: <https://www.lammps.org/index.html>. (Accessed: 22nd June 2021)
 159. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).
 160. Hockney, R. W. & Eastwood, J. W. *Computer simulation using particles*. Bristol: Hilger (1988).
 161. POLAK, E. & RIBIERE, G. Note sur la convergence de méthodes de directions conjuguées. *Dunod* **16**, 35–43 (1969).
 162. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
 163. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

-
164. Martyna, G. J., Tuckerman, M. E., Tobias, D. J. & Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **87**, 1117–1157 (1996).
165. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap*. (Chapman and Hall/CRC, 1994).
166. Müller-Plathe, F. A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. *J. Chem. Phys.* **106**, 6082–6085 (1997).
167. Nieto-Draghi, C. & Avalos, J. B. Non-equilibrium momentum exchange algorithm for molecular dynamics simulation of heat flow in multicomponent systems. *Mol. Phys.* **101**, 2303–2307 (2003).
168. Kolodner, P., Williams, H. & Moe, C. Optical measurement of the Soret coefficient of ethanol/water solutions. *J. Chem. Phys.* **88**, 6512–6524 (1988).
169. Zhang, K. J., Briggs, M. E., Gammon, R. W. & Sengers, J. V. Optical measurement of the Soret coefficient and the diffusion coefficient of liquid mixtures. *J. Chem. Phys.* **104**, 6881–6892 (1996).
170. Wiegand, S., Ning, H. & Kriegs, H. Thermal diffusion forced Rayleigh scattering Setup optimized for aqueous mixtures. *J. Phys. Chem. B* **111**, 14169–14174 (2007).
171. Tichacek, L. J., Kmak, W. S. & Drickamer, H. G. Thermal Diffusion in Liquids; The Effect of Non-ideality and Association. *J. Phys. Chem* **60**, 660–665 (1956).
172. Platten, J. K., Bou-Ali, M. M., Blanco, P., Madariaga, J. A. & Santamaria, C. Soret coefficients in some water - Methanol, water - Ethanol, and water - Isopropanol systems. *J. Phys. Chem. B* **111**, 11524–11530 (2007).
173. Pogorelov, V., Yevglevsky, A., Doroshenko, I., Berezovchuk, L. & Zhovtobryuch, Y. Nanoscale molecular clusters and vibrational relaxation in simple alcohols. *Superlattices Microstruct.* **44**, 571–576 (2008).
174. Brown, R. D., Godfrey, P. D. & Storey, J. The microwave spectrum of urea. *J. Mol. Spectrosc.* **58**, 445–450 (1975).
175. Greenberg, A., Green, A. R. & Liebman, J. F. Computational Study of Selected Amine

- and Lactam N-Oxides Including Comparisons of N-O Bond Dissociation Enthalpies with Those of Pyridine N-Oxides. *Molecules* **25**, 3703 (2020).
176. Mazurkiewicz, J. & Tomasik, P. Effect of external electric field upon charge distribution, energy and dipole moment of selected monosaccharide molecules. *Nat. Sci.* **4**, 276–285 (2012).
177. Würger, A. Thermal non-equilibrium transport in colloids. *Reports Prog. Phys.* **73**, 126601 (2010).
178. Zwanzig, R. W. HighTemperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **22**, 1420–1426 (1954).
179. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
180. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
181. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. & van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529–539 (1994).
182. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
183. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
184. Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput. Aided. Mol. Des.* **29**, 397–411 (2015).
185. Verlet, L. Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).
186. Martínez-Jiménez, M. & Saint-Martin, H. A Four-Site Molecular Model for Simulations of Liquid Methanol and Water-Methanol Mixtures: MeOH-4P. *J. Chem. Theory Comput.* **14**, 2526–2537 (2018).

187. Staudinger, J. & Roberts, P. V. A critical compilation of Henry's law constant temperature dependence relations for organic compounds in dilute aqueous solutions. *Chemosphere* **44**, 561–576 (2001).
188. Humbert, M. T., Zhang, Y. & Maginn, E. J. PyLAT: Python LAMMPS Analysis Tools. *J. Chem. Inf. Model.* **59**, 1301–1305 (2019).

