

# In silico all-atom approach to thermodiffusion in dilute aqueous solutions

Alejandro Diaz-Marquez and Guillaume Stirnemann\*

*CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Université de Paris, 13 rue Pierre et Marie Curie, 75005, Paris, France*

E-mail: stirnemann@ibpc.fr

## Abstract

Thermodiffusion (or thermophoresis) is the phenomenon by which the spatial distributions of constituents of liquid or gas phases become inhomogeneous in response to a temperature gradient. It has been evidenced in a variety of systems and has many practical applications, as well as implications in the context of the origins of life. A complete molecular picture of thermophoresis is still missing and phenomenological approaches are often employed to account for the experimental observations. In particular, the amplitude of the resulting concentration-gradients (quantified by the Soret coefficient) depends on many factors that are not straightforwardly rationalized. All-atom molecular dynamics simulations appear as an exquisite tool to shed light on the molecular origins for this phenomenon in molecular systems, but the practical implementation of thermophoretic settings in silico poses significant challenges. Here, we propose a robust approach to tackle thermophoresis in dilute realistic solutions at the molecular level. We rely on a recent enhanced heat-exchange algorithm to generate temperature-gradients. We carefully assess the convergence of thermophoretic simulations in dilute aqueous solutions. We show that simulations typically need to be propagated on long timescales (hundreds of nanoseconds). We find that the magnitude of the temperature gradient and the box sizes have little effect on the measured Soret coefficients. Practical guidelines

are derived from such observations. Provided with this reliable setup, we discuss the results of thermophoretic simulations on several examples of molecular, neutral solutes, which we find in very good agreement with experimental measurements regarding the concentration-, mass-, and temperature-dependence of the Soret coefficient.

## Introduction

Thermodiffusion (or Soret effect) is the phenomenon by which the spatial distributions of constituents of liquid or gas phases become inhomogeneous in response to a gradient of temperature. It was discovered some 150 years ago in aqueous salt solutions,<sup>1,2</sup> and it has since been evidenced in a broad gamut of contexts, ranging from liquid mixture of small organic molecules to large colloids in water.<sup>2-4</sup> While the term "thermophoresis" is usually employed to refer to thermodiffusion of large particles in a solvent, it is sometimes also used for molecular systems as well.<sup>4</sup> In the following, these two terms will be used indistinctively.

In addition to the ordinary, Fickian particle diffusion, the motion of particles under a temperature gradient is phenomenologically described in terms of a drift velocity proportional to the temperature gradient through a thermal diffusion coefficient. Note that this thermal diffusion coefficient should not be confused with a Fickian diffusion coefficient in isothermal conditions and it does not have the units of a diffu-

sion coefficient. The ratio between ordinary and thermal diffusion coefficients defines the Soret coefficient (SC), whose sign dictates whether molecules will accumulate on the cold or on the hot side, and whose amplitude governs the resulting concentration gradient. Whereas for electrophoresis, the direction of particle motion only depends on its charge, the SC sign is sensitive to temperature, pressure, and to the mixture composition: a given molecule can accumulate on the cold or hot side depending on one or several of these factors.<sup>3-5</sup> Because in a mixture, particles of different natures have distinct SCs, thermophoresis has been used as a polymer separation technique,<sup>5,6</sup> and more recently, it has been measured and characterized in aqueous suspensions of polymer beads,<sup>7</sup> proteins<sup>8</sup> or nucleic acids.<sup>9</sup> Technological applications include the manipulation and the polymerization of biomolecules in thermal traps,<sup>10,11</sup> the measurement of protein-ligand affinity,<sup>12</sup> applications in microfluidics,<sup>13</sup> or the tuning of colloid surfaces.<sup>5</sup>

A particularly striking implication of this phenomenon pertains to its relevance in the accumulations of reactants in the context of prebiotic chemistry.<sup>14</sup> While in modern cells, interacting molecules and objects are confined by membranes in a volume in the  $10^{-14}$ -litre range, resulting in concentrations that critically enhance reaction rates, it is very hard to conceive efficient chemistry in dilute, open conditions. Pioneer experiments from the Braun group have shown that thermally-driven processes, in particular thermophoresis, can accumulate otherwise dilute precursors at localized spots acting as thermal traps for RNA strands oligomerization.<sup>10,11</sup> Such realizations are not just lab inventions: natural thermophoretic settings occur e.g. in the pores of rocks of hydrothermal vents on the ocean’s floor that may have played a key role in the appearance of biological chemistry.<sup>14,15</sup>

The nature of thermophoresis is still not completely understood and it remains highly debated, with a variety of theoretical models and sometimes contradictory pictures, especially regarding the ability of concepts from equilibrium-thermodynamics to explain such an

out-of-equilibrium process.<sup>16</sup> As of today, thermophoresis remains often regarded from a phenomenological point of view,<sup>3,4</sup> with no definitive and universal physical model able to explain it. Another limitation in our current understanding of thermophoresis is that we cannot fully predict the behavior of any given molecule under a thermal gradient: what is the amplitude of its SC, its sign, and how does it depend on thermodynamic conditions?

Simulation approaches could bring very valuable information about the microscopic origin of thermophoresis,<sup>4,17</sup> and these strategies have been employed at several levels of descriptions, ranging from all-atom molecular dynamics (aaMD) simulations<sup>18-21</sup> to mesoscale<sup>22,23</sup> and finite elements approaches.<sup>24,25</sup> The main interest of aaMD approaches is that they do not require any major assumptions about the physical rules that are relevant or not for the phenomenon, and contain all physical interactions in the system of interest, which evolves following Newtonian dynamics. More coarse-grained descriptions, such as Langevin or Brownian dynamics, readily require assumptions (most of all, treating the solvent as a continuum), which may not be appropriate to properly describe thermophoresis.

The main limitation of aaMD is that they are currently limited to systems of small to moderate sizes, typically to molecules smaller than a few nanometers, and for limited timescales. AaMD has been e.g. applied for mixtures of ideal LJ particles,<sup>26-28</sup> apolar solutes,<sup>19,29</sup> or aqueous solutions of ions<sup>20,30</sup> or of small molecular solutes.<sup>18,31</sup> The study of thermophoresis and the determination of the Soret coefficients (SC) for a given molecular solute in dilute aqueous solutions, is of particular relevance in the context of reactant accumulation in aqueous environment; however, it poses several key challenges in practice. In particular, the associated SC are on the order of  $1-10 \times 10^{-3} \text{ K}^{-1}$ , implying that a 10-K temperature difference leads to a 1-10% change in concentration only, which might be hard to converge when only a handful of solute molecules are present inside the simulation box. Due to their finite diffusion timescale, each solute should sample sev-

eral times the total lengthscale of the temperature gradient so that an accurate steady-state can be determined, especially if the resulting concentration-gradient is small because of small values of the SC. While this might not be critical for concentrated binary mixtures, due to the large number of solute molecules, this is critical for dilute solutions, as we shall see in detail.

The concentration gradients can be made more prominent by imposing larger temperature-differences across the simulation box;<sup>19,20</sup> however, this could result in temperature-gradients that are orders of magnitude larger as compared to the experiments. An indirect consequence is that the employed molecular force-fields, and that of water in particular, should be chosen with care<sup>19</sup> to remain valid on a large temperature-range.

Another problem pertains to the choice of a reliable non-equilibrium simulation technique in order to generate a temperature-gradient in a MD simulation box. Most studies to date that have focused on aqueous solutions have relied on the thermostating of two regions of the simulation box;<sup>20,30,31</sup> while this has been shown to be efficient, this usually imposes, considering the current implementation of thermostats in simulation codes, to restrain solvent positions in the thermostatted regions, which can more generally affect fluid motions in the box. Moreover, the simulation ensemble is not well-defined. Another approach consists in imposing a heat flux<sup>32,33</sup> that in turn results in a temperature gradient.<sup>19,26,34</sup> The heat flux is imposed by exchanging momentum<sup>33</sup> or kinetic energy<sup>32</sup> between two defined regions of the simulation box, or by symmetrically rescaling the particle velocities in these regions.<sup>35</sup> The simulations are ideally performed in the micro-canonical ensemble, which poses the question of total energy conservation when using finite time-steps, in addition to problems inherent to the exchange algorithms themselves.<sup>36</sup>

Our goal here is to address these challenges and to propose a robust aaMD approach to tackle thermophoresis in dilute realistic solutions at the molecular level. We propose to rely on a recent enhanced heat-exchange algorithm to generate temperature-gradients,<sup>36</sup> which, to

the best of our knowledge, has not yet been applied in this context. We carefully assess the convergence of thermophoretic simulations in dilute aqueous solutions, which is critical for dilute and realistic solutions. We show that simulations typically need to be propagated on very long timescales (hundreds of nanoseconds) to generate reliable concentration-gradients. We find that the magnitude of the temperature gradient and the box sizes have little effect on the measured SC. Practical guidelines are derived from such observations. Provided with this reliable setup, we discuss the results of thermophoretic simulations on several examples of molecular, neutral solutes, which we find in very good agreement with experimental measurements regarding the concentration-, mass-, and temperature-dependence of the SC.

## Theoretical considerations and simulation strategy

### Thermophoresis and Soret coefficient

In the following, we will consider a linear and uniform temperature gradient  $\nabla T$  along the  $z$  direction. Thermophoresis of a compound  $s$  in a solvent  $w$  results from the competition between ordinary diffusion, characterized by a diffusion coefficient  $D^s$ , and thermal diffusion, related to a phenomenological thermal diffusion coefficient  $D_T^s$ . By taking the notations of refs.,<sup>7,16</sup> the total current of (dilute) solute particles present at a concentration  $c_s(z)$  at a position  $z$  can be written as a sum of the diffusion current  $j_D(z)$  and of the thermal diffusion current  $j_{TD}(z)$ , such that

$$j^s(z) = j_D(z) + j_{TD}(z), \quad (1)$$

with

$$j_D(z) = -D^s(z)\nabla c_s|_z \quad (2)$$

and

$$j_{TD}(z) = -D_T^s(z)c_s(z)\nabla T. \quad (3)$$

Experiments show a steady-state can be reached, which implies that  $j^s(z) = 0$  at each

position. We also note that other derivations consider the mass currents in place of the particle currents; for dilute solutions, both approaches are equivalent, and non-ideal effects can also be neglected. In the steady-state, combining these 3 equations leads to

$$\left. \frac{d \ln c_s}{dT} \right|_z = -\frac{D_T^s(z)}{D^s(z)}, \quad (4)$$

or, equivalently, into

$$\left. \frac{d \ln c_s}{dT} \right|_x = -S_T^s(z), \quad (5)$$

where  $S_T^s(z) = D_T^s(z)/D^s(z)$  is defined as the Soret coefficient (SC). In principle, each position is associated with a given concentration  $c_s(z)$  and temperature  $T(z)$ , and  $S_T$  depends on both the concentration and the 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  $T$ , suggesting that  $S_T$  can be considered locally independent of  $T$  and  $c_s$  under these conditions.

## In silico thermophoretic set-up

In order to generate a temperature gradient along one chosen direction  $z$ , parallelepipedic boxes are considered, such that  $L_x = L_y$  and where  $L_z$  is typically larger than  $L_x$  or  $L_y$ , and centered in  $(0, 0, 0)$ . We then define cold and hot slabs along  $z$ , that are symmetrically positioned at  $z_c$  and  $z_h$  respectively, and of thickness  $\delta_z$ . These slabs can be taken at the edges of the simulation box, i.e.,  $z_c \approx -L_z/2$  and  $z_h \approx L_z/2$ , leading to a temperature-gradient along  $L_z$ ; however, because of periodic boundary conditions, this would result in a discontinuity and a much larger gradient as the cold region is directly adjacent to the hot region of its image along the  $z$  direction. We thus prefer to position the slabs symmetrically, i.e.,  $z_c = -L_z/4$  and  $z_h = L_z/4$ . In that case, the temperature-gradient is established over  $L_z/2$ , and is occurring twice in the box (Figure 1). Unless otherwise specified, the width of the thermostatted regions  $\delta_z$  was set to 4 Å and  $L_x = L_y = 25$  Å.

Creating cold and hot slabs can be achieved by thermostating solvent molecules belonging to these two regions. However, current implementations of thermostats in simulation softwares require to define particles to be thermostatted a priori, i.e., it cannot be directly updated on the fly as molecules belonging to the cold/hot regions leave/enter the slab. As a consequence, harmonic restraints are usually employed to constrain solvent molecules in the thermostatted regions.<sup>20,31</sup> Even when not all the molecules in the slab are thermostatted, it can clearly affect molecular motions along the direction of the gradient.

As already mentioned in the introduction, several methods exist that impose a heat flux within a molecular dynamics simulation box, by swapping e.g. kinetic energy,<sup>32</sup> momenta,<sup>33</sup> or by performing velocity rescaling moves in pre-assigned cold and hot regions.<sup>35</sup> Here, we will use a heat-exchange algorithm (HEX), where, at given time intervals, heat  $-Q$  is "pumped" out of what is decided to be the cold region, and injected  $+Q$  into the corresponding hot region.<sup>32</sup> Formally, such simulations are thus performed in the microcanonical ensemble, and should conserve energy (Note that energy conservation in this case both depends on the exchange algorithm itself, but also on independent factors that would also lead to poor conservation in equilibrium simulations, such as the simulation timestep.). However, it has been shown that many versions of this algorithm can lead to energy-conservation problems, which was only recently solved in an enhanced version of the algorithm (eHEX).<sup>36</sup> It is not clear whether other algorithms that swap momenta or perform velocity rescaling moves would present the same energy conservation issues when propagated on tens of nanoseconds, but they probably suffer from the same truncation errors identified in ref<sup>36</sup> when written in terms of an integration scheme. We rely on the implementation of eHEX into the LAMMPS simulation package, and the reader is referred to the corresponding work for details about the algorithm for the propagation of the equations of motion that allow for better energy conservation. Heat exchange was performed at each simulation time-

step (we verified that exchanging heat less frequently did not affect our results).

## Simulation details and forcefields

All simulations employed the LAMMPS package (version 07Aug19), starting from configurations generated using Packmol.<sup>37</sup> The TIP4P-05 water model,<sup>38</sup> which reproduces water structural, dynamical and thermodynamical properties well on a wide temperature range, was employed in most cases, except when a comparison was made with the SPC/E water model<sup>39</sup> when investigating the effect of the solvent forcefield on the solute thermophoretic behavior. Because of the presence of a dummy atom, TIP4P-05 requires to use particular options in LAMMPS, including a specific version of the solver for long-range interactions (PPPM).<sup>40</sup> The real space cutoff for Coulombic interactions was set at 8.5 Å, and the Lennard-Jones cutoff to 9 Å (which is slightly larger as compared to the electrostatic interaction cutoff because of the difference in the localization of the negative charge and of the mass of the oxygen atom). The effect of the simulation timestep was investigated, as detailed later, and fixed to 1 fs in most of our simulations. The solutes were described using state-of-the-art forcefields from the literature (TMAO,<sup>41,42</sup> urea<sup>43</sup> (bonded terms from<sup>44</sup>), methanol<sup>45</sup> and glucose<sup>46</sup>).

## Simulation strategy

The data used to measure the steady-state behavior of a given aqueous solutions under a set thermal gradient is usually averaged over tens of independent trajectories. Each trajectory is propagated in the presence of the temperature gradient after it had been prepared and equilibrated in the absence of gradient as follows, and as schematically indicated in Table 1 and Figure 1.

Solvent molecules are randomly inserted within the simulation box. In order to avoid any strong heterogeneity in the (dilute) solute distribution in the simulation box, we insert solute molecules uniformly along the thermal

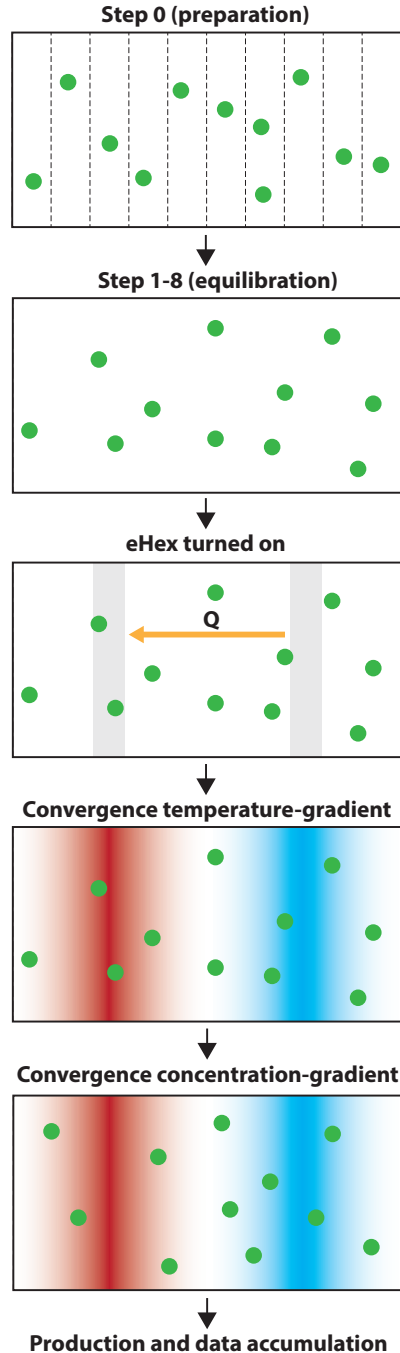


Figure 1: Summary of the generation of a steady-state thermophoretic setting. After preparation (step 0), the system is equilibrated (steps 1 to 8) before the heat-exchange algorithm is activated. Convergence of first the temperature-gradient and then the concentration-gradient are assessed before a production run can be propagated. This is typically repeated over tens of trajectories that are generated and equilibrated independently starting from step 0.

gradient direction  $z$ , and randomly along the perpendicular dimensions (step 0). The system is then minimized using a Polak-Ribiere<sup>47</sup> version of the conjugate gradient algorithm until relative changes in energy and forces fall below cut-off values equal to  $10^{-4}$  and  $10^{-6}$ , respectively (step 1). Velocities are then set to correspond to the target median temperature (step 2), and a NPT equilibration of 0.4 ns is performed at this median temperature (step 3). All simulations in this work have been prepared using a pressure of 1 atm, using a Nose-Hoover barostat with a damping time of 1 ps, that is only applied along the  $x$  and  $y$  dimensions (as we do not want the thermal gradient direction to vary in length). The average volume under these thermodynamic conditions is determined, and the box rescaled accordingly (step 4). The system is then equilibrated in the NVT ensemble for 0.2 ns at this equilibrium volume (step 5), using the Nose-Hoover thermostat with a characteristic damping time of 100 fs. The last frame of this second equilibration is then used for a first, short propagation in the microcanonical ensemble (step 6) for 1 ps. This portion of the trajectory is used to compute the average energy and compare it with the average energy of the NVT trajectory. The velocities of the system particles in the last NVE frame are then rescaled by the ratio between these two average energies (step 7), such that a second propagation in the microcanonical ensemble for 0.2 ns then leads to the correct target temperature (step 8). Finally, the heat-exchange algorithm eHex is turned on and the simulation is propagated for 10 ns under the non-equilibrium conditions of the temperature-gradient but formally in the microcanonical simulation ensemble.

## Data analysis

The solution properties along the gradient first involve the definition of regularly-spaced slabs along the  $x$  and  $y$  directions, and perpendicular to the direction of the gradient. We typically divide  $L_z$  into 15 to 30 slabs. A larger number of slabs tend to give smoother data but at the expense of increased statistical noise,

which turns out to be problematic when studying properties with poor statistics (such as the solute distributions). Temperature was measured within each slab by averaging over all solute molecules and taking into account the restraints of the water model. Such a functionality is directly implemented in LAMMPS and we checked that an independent calculation a posteriori gave the same results. Molality (sometimes referred to as concentration here) was defined in each slab by dividing the average number of solutes centers of mass found in this slab by the average mass of water solvent molecules in the same slab, and is given in units of molal m, which are moles of solutes by kilogram of solvent. Standard-deviations of the molality profiles were estimated from a set of typically 20–40 independent trajectories, taking into account the last 10 ns (as discussed in the main text). Soret coefficients (SC) were determined using Equation 5 on the average concentration profile. Because individual trajectories lead to very noisy concentration profiles, the determination of the error bars on SC based on the estimation of the SC for each separate trajectory would be meaningless and lead to unrealistic uncertainties. In that case, we therefore used bootstrapping with 50 samples of 10 trajectories each one to estimate the statistical noise.

## Results and discussion

### Convergence of the temperature-gradient and energy conservation

We first assess the ability of the thermophoretic simulation setup to generate temperature gradients and to conserve the total energy over the simulation timescale that is required to lead to a reliable solute concentration gradient, which is on the order of a few nanoseconds for each trajectory.

As noted earlier, the original implementations of the heat exchange algorithm<sup>32,33</sup> faced important energy conservation problems due to the truncation error in the propagation operator splitting, which was solved by adding an extra integration step in the algorithm propa-

**Table 1: Summary of the generation of a steady-state thermophoretic setting.**

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	$\tau_T \approx 100$ ps
10	Concentration-gradient convergence	NVE	$\tau_C \approx 2$ ns
11	Production	NVE	$\tau_P = 10$ ns

gating the equations of motion.<sup>36</sup> We first verify the quality of energy conservation for the eHex algorithm over a 10-ns timescale, depending on the employed timestep (we tested 0.5, 1, 2 or 5 fs). In the first three cases, the relative energy drift is barely noticeable, and typically in the range of  $1\text{--}3 \times 10^{-6} \text{ ns}^{-1}$ . For a timestep of 5 fs, the measured drift is about 100 times larger. We have verified that the NVE/eHex simulation exhibits a behavior very similar to that of a standard NVE simulation, and does not lead to additional drifts. This can be compared to drifts that are orders of magnitude larger for the original algorithm implementations.<sup>36</sup> While a timestep of 2 fs would be reasonable when all bonds involving hydrogen atoms are maintained rigid (as it is usually the case for most water forcefields), we chose to employ a smaller  $\delta t = 1$  fs timestep in order to allow for fully-flexible solute forcefields.

We now turn to the kinetics of the transition to a steady-state temperature-gradient once the eHex algorithm has been switched on (Figure 2), with a value of  $Q$  such that  $\Delta T = 60$  K. In agreement with previous studies on model Lennard-Jones systems,<sup>48</sup> a stable and steady-state temperature-gradient is reached on short timescales. For pure water with  $L_z = 50$  Å and thus  $\nabla T = 2.4 \text{ K}/\text{Å}$ , a steady-state is reached after  $\approx 10 - 20$  ps for a median temperature of  $T_{\text{ref}} = 300$  K (Figure 2A) or 330 K (Figure 2B). Considering that the water ther-

mal diffusivity  $\alpha$  in this temperature range<sup>49</sup> is close to  $\approx 0.14 - 0.15 \text{ mm}^2/\text{s}$ , we expect the heat flux to propagate on a  $(L_z/4)/\alpha$  timescale, which for a box length of  $L_z = 50$  Å, corresponds to  $\approx 11$  ps, which is fully consistent with our observations. When the box length along the direction of the gradient is doubled such that  $L_z = 100$  Å and  $\nabla T = 1.2 \text{ K}/\text{Å}$ , the equilibration time increases to  $\approx 50 - 100$  ps (Figure 2C). This quadratic dependence upon the box length was indeed expected from our back-of-the-envelope estimation based on water thermal diffusivity. The addition of solute molecules at a dilute concentration (Figure 2D) has no noticeable effect on the equilibration timescale when compared a pure-water system in the same conditions (Figure 2B). We can therefore consider that for systems in which the temperature-gradient spans over a few nanometers, a steady-state temperature-gradient is reached within a few tens of picoseconds at most. This timescale would of course increase if longer simulation boxes were considered; however, as now studied in detail, the Soret effect and the establishment of a steady solute concentration-gradient occur on significantly longer timescales.

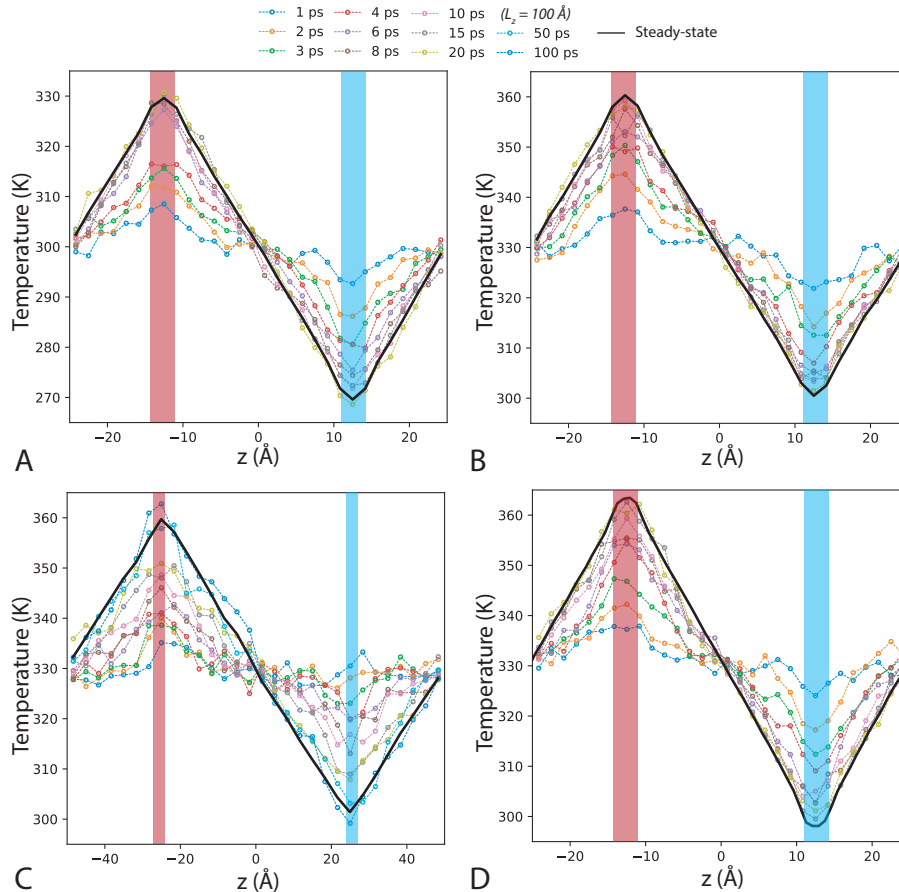


Figure 2: Convergence of the water temperature-gradient. The energy flux was set to  $0.0375 \text{ kcal.mol}^{-1}.\text{fs}^{-1}$  ( $L_z = 50 \text{ Å}$ ) and  $0.01875 \text{ kcal.mol}^{-1}.\text{fs}^{-1}$  ( $L_z = 100 \text{ Å}$ ). Data was averaged over 20 independent trajectories. The blue and red bars indicate the regions where heat is pumped and injected, respectively; their width is at scale. (A) Pure water,  $L_z = 50 \text{ Å}$  and  $T_{\text{ref}} = 300 \text{ K}$ ; (B) Pure water,  $L_z = 50 \text{ Å}$  and  $T_{\text{ref}} = 330 \text{ K}$ ; (C) Pure water,  $L_z = 100 \text{ Å}$  and  $T_{\text{ref}} = 330 \text{ K}$ ; (D) Aqueous solution of TMAO (2.2 m),  $L_z = 50 \text{ Å}$  and  $T_{\text{ref}} = 330 \text{ K}$ .

## Convergence of the concentration-gradient

Provided with a reliable set-up to generate a temperature-gradient across a water simulation box, we now turn to the core topic of the current work, namely the study of dilute aqueous solutions. As noted earlier, the concentration gradients equilibrate on much longer timescales as compared to the solvent temperature gradients.<sup>20,48</sup> This is even more critical in dilute solutions, since there are only a few solute molecules in the simulation box, leading to poor statistics. Moreover, solute molecules need to diffuse at least locally to be able to sense the temperature-gradient. For a typical, molecular size solute, the diffusion coefficient under ambient conditions is on the order

of  $0.01 - 0.1 \text{ Å}^2/\text{ps}$ , meaning that one solute molecule will sample 1 nm (which covers less than half the typical length of the employed gradient) over  $\approx 0.5 - 5 \text{ ns}$ .

In the following, we will mostly focus on a particular solute, trimethylamine *N*-oxide (TMAO). TMAO has been considered as a paradigm, small molecular solute containing both a hydrophilic and a hydrophobic character, and it behaves as a protein and polymer protective osmolyte.<sup>50-52</sup> In particular, it has been used extensively to assess the effect of hydrophobic groups on water structure and dynamics,<sup>53-56</sup> as purely hydrophobic molecules are not soluble in water. Its large dipole moment prevents aggregation up to high concentrations.<sup>57</sup> This aspect is critical since we

want to avoid spurious effect from aggregation that could enhance solute accumulation in thermophoretic settings.

In Figure 3, we show the evolution of the concentration-gradient for the model TMAO solute in water varying the average temperature and the concentration. An accurate determination of the time-dependence of the concentration profile in such "dilute" solutions is not straightforward, since a given simulation frame only reports on a few dozens solute molecules. Even when averaged on tens of frames and when repeated over tens of trajectories, data remain noisy, especially at short times where the average is necessarily performed on shorter time windows. However, the data shown in Figure 3 readily provides some very useful information about the involved timescales. For a concentration of 2.2 m, the concentration profile is seen to converge on a  $\approx 1$  ns timescale for a reference temperature of 300 K (Figure 3A), and equilibrates twice as fast at 330 K (Figure 3B). When the concentration is doubled (Figure 3C), convergence is faster as well, which is expected since more molecules are present to sample local temperature environments.

It is useful to compare these timescales to the corresponding solute diffusion coefficient at the corresponding temperature, which are equal to 0.049 and 0.101  $\text{\AA}^2/\text{ps}$ , respectively. While a proper determination of a scaling law would require more extensive tests involving other solutes and concentrations, a rule of thumb that we can derive here is that the equilibration time for  $N$  solute molecules along a gradient spanning  $L_z/2$  is on the order of  $\approx L_z^2/(D \times N)$ . While we do not expect this relationship to remain valid varying extensively the solute or the simulation conditions, this gives 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 others in sufficiently dilute solutions, this timescale also sets 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

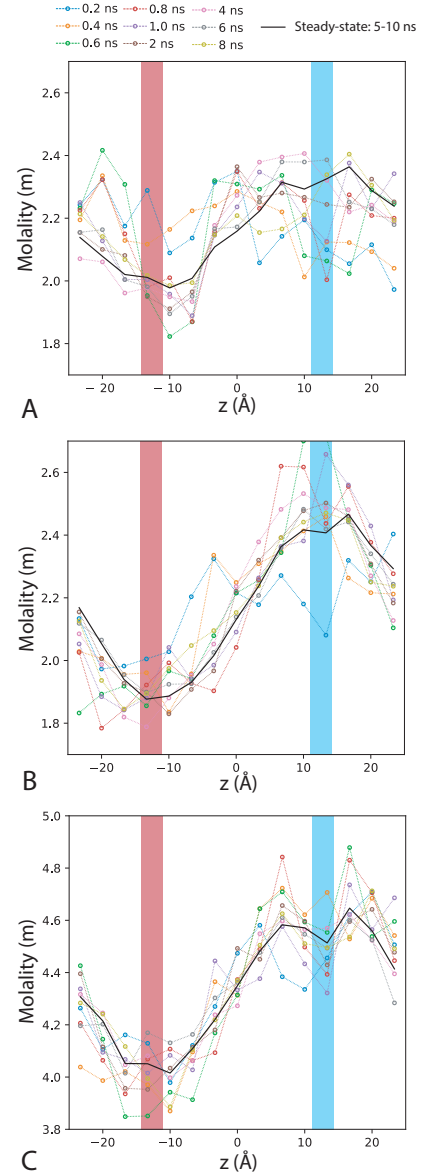


Figure 3: Convergence of the TMAO concentration-gradient. The energy flux was set to 0.0375 kcal.mol<sup>-1</sup>.fs<sup>-1</sup> in all cases. Data was averaged over 40 independent trajectories. The blue and red bars indicate the regions where heat is pumped and injected, respectively; their width is at scale. (A) 2.2-m average concentration,  $L_z = 50$   $\text{\AA}$  and  $T_{\text{ref}} = 300$  K; (B) 2.2-m average concentration,  $L_z = 50$   $\text{\AA}$  and  $T_{\text{ref}} = 330$  K; (C) 4.4-m average concentration,  $L_z = 50$   $\text{\AA}$  and  $T_{\text{ref}} = 330$  K.

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.

The above study of the simulation convergence in terms of temperature- and concentration-gradients allows us to derive a successful strategy in order to compute Soret coefficients in dilute aqueous solutions. In order to get detailed estimates of the average concentration-gradients and of the statistical errors, many trajectories (typically 20) are run in parallel following the equilibration procedure presented before.

The first 2 ns after eHex has been turned on, corresponding to the transient temperature- and concentration-gradient regimes, are typically discarded (which is conservative enough for all investigated systems but might need to be adapted for solutes diffusing more slowly). Another point that we found important was that because of the diffusion timescale discussed above, a box that would initially exhibit strong heterogeneities in the spatial distribution of solute molecules along the direction of the gradient would need much longer timescales to equilibrate. One way to solve this issue would be to randomly select initial frames for each trajectory from a very long preliminary simulation to cover a range of realizations of concentration heterogeneities; but considering the size of the simulation box and the small concentrations, together with the timescale for solute diffusion, this could require tens of nanoseconds of simulation, and the transient regime would be much longer for those configurations with an initial distribution very far from the final, steady-state one. In order to limit as much as possible the timescale of the transient regime, we turned to an alternate strategy where each trajectory starts from a configuration where the distribution of solute molecules is homogeneous in the direction of the gradient, but random along the perpendicular directions (step 0).

### Dependence upon box size and gradient amplitude

As compared to experimental setups, simulation approaches require both larger temperature differences between the hot and cold re-

gions (for the simulated concentration difference should be larger than the large, stochastic fluctuations) and much smaller systems, typically in the 1–10 nm range. Both factors contribute to simulated temperature-gradients that are much larger than their experimental counterparts. While experimental values are far from reach for all-atom MD simulations, we can nonetheless address the question whether the amplitude of the gradient has any impact on the measured SC at a given median temperature and concentration.

We have studied the effect of the gradient amplitude on the model TMAO solute in water at a fixed median temperature of 330 K and a fixed concentration of 2.2 m. The gradient was first varied by adjusting  $L_z$  between 5 and 10 nm while keeping  $L_x = L_y$  constant (which required us to adjust the number of solute molecules in order to maintain a fixed concentration). The amount of exchanged heat  $Q$  was adjusted to lead to a temperature difference between the hot and cold regions of 60 K. As seen in Figure 4, the concentration-gradient is virtually insensitive to the box size and thus the corresponding temperature-gradient.

Second, we have tested several temperature-gradients, ranging from 0.8 to 3.6 K.Å<sup>-1</sup>, for a fixed box length of  $L_z = 50$  Å (Figure 5A and B). The absolute concentration differences obviously increase when  $\Delta T$  increases, but as shown using the representation of Figure 5B, similar variations are seen outside the thermostatted regions, leading to very similar SC values around the median temperature. However, a critical effect of the gradient amplitude is to increase the signal-to-noise ratio. While the smallest investigated gradient of 0.8 K.Å<sup>-1</sup>, which corresponds to  $\Delta T = 20$  K here, readily leads to a noticeable concentration-gradient, the resulting profile suffers from statistical uncertainties that are almost as large as the concentration-difference themselves. For the largest investigated gradient (3.6 K.Å<sup>-1</sup>, i.e.,  $\Delta T = 90$  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 sta-

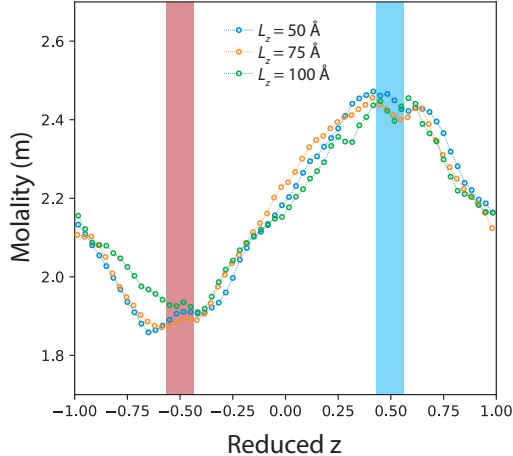
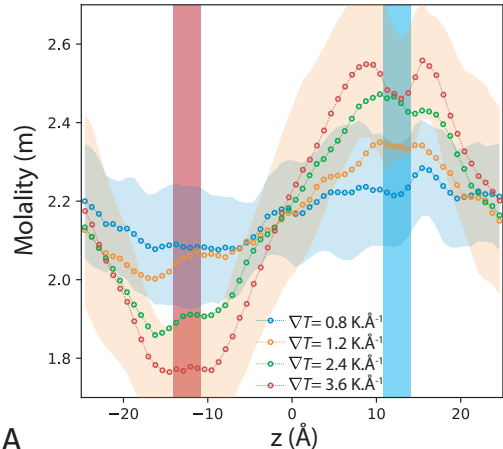


Figure 4: Box size dependence. The heat flux was set to  $0.0375 \text{ kcal.mol}^{-1}.\text{fs}^{-1}$  ( $L_z = 50 \text{ Å}$ ),  $0.025 \text{ kcal.mol}^{-1}.\text{fs}^{-1}$  ( $L_z = 75 \text{ Å}$ ) and  $0.01875 \text{ kcal.mol}^{-1}.\text{fs}^{-1}$  ( $L_z = 100 \text{ Å}$ ), and  $T_{\text{ref}} = 330 \text{ K}$ . Data was averaged over 20 independent trajectories  $L_z = 50 \text{ Å}$ , and 40 for  $L_z = 75 \text{ Å}$  and  $L_z = 100 \text{ Å}$ . 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 2.2-m TMAO solution:  $L_z = 50 \text{ Å}$  (blue),  $L_z = 75 \text{ Å}$  (orange) and  $L_z = 100 \text{ Å}$  (green).

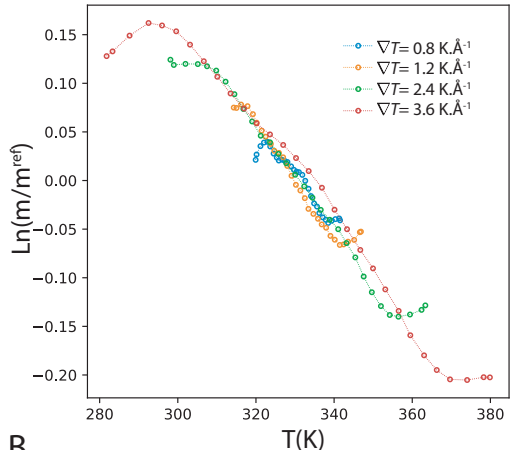
tistical uncertainties, while employing a reasonable temperature-gradient where the water phase lies within, or not too far, from its equilibrium liquid phase domain (namely, between 273 and 373 K). In the following, we chose to employ the  $2.4 \text{ K.Å}^{-1}$  gradient, which allows to satisfy these conditions in most cases.

### TMAO solutions: temperature-dependence and forcefield sensitivity

Provided with a reliable computational thermophoretic set-up, we now turn the investigation of the temperature-dependence of the Soret coefficients in aqueous solutions, and we investigate the effect of the employed water and solute forcefields. We chose to focus on TMAO because it represents a paradigm and minimal amphiphilic system, containing both a hydrophobic moiety and a hydrophilic head. Unfortunately, no experimental data could be found regarding thermophoresis in aqueous TMAO so-



A



B

Figure 5: Gradient amplitude dependence. Data was averaged over 20 independent trajectories. The blue and red bars indicate the regions where heat is pumped and injected, respectively. Data is shown for four different gradients  $\nabla T$  in a 2.2-m TMAO solution, with  $T_{\text{ref}} = 330 \text{ K}$ . (A) TMAO molality as a function of  $z$ . The shaded areas indicate the standard-deviations estimated from 20 independent trajectories at the two extreme gradients. (B) Same data represented on a log-scale for the molality and as a function of temperature. The SC at the reference temperature can be interpreted as the slope of the curves, which are basically identical at the center of the gradient.

lutions; however, we can make a direct connection with experimental work on solutes of similar sizes and with similar chemical groups,<sup>4,31</sup> from which the following observations were made: first, the SC is on the order of  $1\text{--}10 \times 10^{-3} \text{ K}^{-1}$  in dilute solutions; second, in

dilute solutions, it usually decreases with decreasing temperature, which could even lead to a sign inversion.

We have considered 2 solute forcefields as well as two different water models, and we have compared 4 different combinations; first, a reference TMAO forcefield that has been specifically parametrized (Kast)<sup>41,42</sup> and the TIP4P-2005 water model, which was already presented in the results mentioned above; second, the same Kast TMAO with the SPC/E water model; finally, a TMAO forcefield generated using the Charmm-CgenFF<sup>58</sup> procedure, together with TIP4P-2005, or SPC/E.

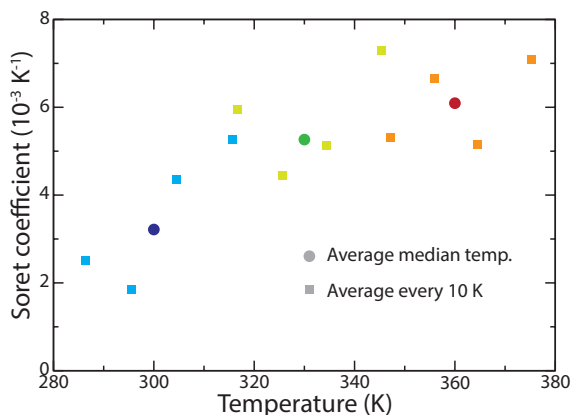


Figure 6: Temperature dependence of the Soret coefficient in a 2.2-m TMAO solution. Data was averaged over 20 independent trajectories. Plain circle indicate the average SC calculated from simulations performed at a median temperature of  $T_{\text{med}} = 300$  K (blue), 330 K (green) and 360 K (red), with a temperature gradient of  $\approx 60$  K. Squares indicate the SC calculated on 10-K intervals between  $T_{\text{ref}} - 20$  and  $T_{\text{ref}} + 20$  K in each set of simulations (same color code).

We selected two water models that are today widely employed because they offer some of the best agreement with experimental data for relatively simple 3- and 4-site non-polarizable models. They were also shown to reproduce the thermal conductivity of water reasonably well on a wide temperature and density range, suggesting that explicit polarizability or flexibility of water molecules was not required for a proper description of thermodiffusion in aqueous solutions.<sup>59</sup> TIP4P-2005<sup>38</sup> has become a reference

model because of its ability to reproduce a large number of features of the water phase diagram and exhibits a melting temperature close to the experimental value. SPC/E,<sup>39</sup> which is usually considered to be a decent model as well but is less recent, suffers from limitations in that respect, and its melting temperature is too low as compared to the experimental values. However, the behavior of these two models at temperatures above melting is expected to be pretty similar, as they both give excellent agreement with experimental structural and dynamical data in the liquid state above 273 K. Indeed, no noticeable change was observed when we varied the water model (Table 2). The obtained Soret coefficients are the same within error bars.

In contrast, a larger effect is observed when changing the TMAO forcefield (Table 2). In that case, a different solute forcefield, especially one that has not been specifically tuned to reproduce thermodynamical, structural and/or dynamical data, may interact differently with water. Here, the Charmm-Cgenff is leading to a SC what is reduced by 40%. Quite surprisingly, a value more consistent with that obtained with the reference Kast forcefield is recovered when the water model was switched to SPC/E. As already argued, solute–water interactions have a direct effect on the amplitude of the SC,<sup>31</sup> and while the obtained SC seems to be quite robust, its exact value is sensitive to the choice of the employed solute and solvent forcefield.

We now switch to the temperature-dependence of the SC, which can be estimated using two different strategies (the Kast/TIP4P-05 combination is employed). The first one consists in dividing the simulation box, for a given median temperature, into a few different regions for which the concentration profiles are analyzed separately and the corresponding SC attributed to the region’s median temperature. Otherwise, one can also vary the simulation median temperature and fit the overall concentration profile along the gradient to get an average SC at this median temperature.

In Figure 6, we show the temperature-dependence of the SC as calculated with both approaches, from 3 simulations ran at median

**Table 2: Measured Soret coefficients at 330 K for 2.2-m aqueous solutions of several molecular solutes, together with their molar mass and average molecular volume.**

Solute	$M$ (g.mol <sup>-1</sup> )	$V$ (Å <sup>3</sup> )	$S_T$ (10 <sup>-3</sup> K <sup>-1</sup> )
Methanol	32.0	66.1	2.9 ± 0.7
Urea	60.1	77.7	4.4 ± 1.0
TMAO (Kast/TIP4P-05)	75.1	122.8	5.0 ± 0.8
TMAO (Kast/SPCE)	75.1	119.9	5.3 ± 0.8
TMAO (Cgenff/TIP4P-05)	75.1	105.7	3.1 ± 1.3
TMAO (Cgenff/SPCE)	75.1	103.0	4.8 ± 0.9
Glucose	180.2	147.1	3.1 ± 1.1

temperatures of 300, 330, and 360 K respectively, with a temperature difference of 60 K. These two approaches lead to very similar results along the investigated temperature-range. As expected, the SC decreases as with temperature, and a sign inversion is expected around 260–270 K, which is consistent to what was e.g. observed experimentally for aqueous solutions of urea at similar concentrations.<sup>31</sup>

Both the magnitude and the temperature-dependence of the SC thus appear to be correctly predicted by our thermophoretic simulations.

### Extension to other aqueous solutions

We finally extend our approach to other solutes, at the same dilute concentration. Experiments have shown that in general, the SC increases for larger, heavier, and more hydrophobic solutes.<sup>4</sup> Our goal here is not to precisely assess the details of each of these effects, but rather to test whether aaMD simulations with conventional forcefields are able to capture these experimental observations and measurements. Results are shown in Table 2 for methanol, urea, TMAO, and glucose, together with their average volume and their molar mass. State-of-the-art solute forcefields were chosen in each case (see Methods), together with the TIP4P-2005 water model.

These solutes were chosen so that they do not aggregate at this concentration, which could lead to spurious effects in their thermophoretic behavior by enhancing solute accumulation. As a consequence, all these solutes contain at least

one or several hydrophilic groups that interact favorably with water molecules by engaging in intermolecular hydrogen-bonds. TMAO also contains a large hydrophobic moiety, but is very soluble because of its large dipole moment, which contributes to prevent aggregation.

Overall, the results presented in Table 2 suggest that larger/heavier solutes tend to exhibit larger Soret coefficients, with a value that is almost doubled when the mass/volume of the solute doubles when going from methanol to TMAO. Glucose appears as an outlier in this series, with a SC close to that of the much smaller methanol. While the detailed investigation of this result is not the focus of this work, we can comment on the possible origins of this phenomenon. For large solutes, diffusion gets slower and therefore the equilibration of the concentration gradient could take longer than a few nanoseconds. To check that this SC value was well-converged, we doubled the set of trajectories (40 instead of 20 as employed for the other systems), but we did not observe any significant statistical difference in the estimation of the SC. A second explanation could be that despite its larger mass and size, the molecular interactions between glucose and water are different from that of methanol such that what has been attributed to the "chemical" contribution<sup>26,28</sup> to the SC might compensate for the mass/volume effect; another cause could be the employed forcefield, which, as exemplified for the TMAO case, can have a significant impact on the measured SC. While it is not possible to conclude solely based on one concentration with one given solute forcefield, this observation is

consistent with the fact that, under given thermodynamic conditions, the SC depends on multiple factors<sup>4,5</sup> (size, mass, chemistry) that can lead to some degree of compensation regarding the thermophoretic effect.

## Conclusions

In this work, we develop an all-atom MD strategy to reproduce thermophoretic settings *in silico* and to quantify the amplitude of solute accumulations in dilute aqueous solutions. While such simulations have been reported before for a variety of molecular systems,<sup>17–20,27,28,30,31,33,34</sup> dilute aqueous systems pose significant challenges that were carefully addressed and discussed. In particular, the solute concentration-gradients that result from the temperature-gradient converge on timescales on the order of a few nanoseconds, and because of the poor statistics resulting from the low solute concentration, they have to be averaged over tens of trajectories to lead to reliable estimates of the Soret coefficients. An additional novelty of our work is to apply a recent heat-exchange algorithm<sup>36</sup> that lead to excellent energy conservation and that does not require to employ thermostats with restrained solvent molecules and an ill-defined simulation ensemble. This approach leads to temperature-gradients that converge on short timescales ( $< 100$  ps) as compared to the concentration-gradients, and allows to perform simulations in the microcanonical ensemble.

Provided with this simulation strategy on a paradigm aqueous solution, we show that the SC is independent of the employed system sizes and temperature-gradient amplitudes, at least within the ranges that were studied here. The obtained values are consistent with existing experimental data on other dilute solutes in water. We demonstrate that the SC is to some extent sensitive to the employed solute and water forcefields, which is expected since it has been shown to depend on solute–water interactions that may differ when switching forcefields. We have used two different approaches to measure the temperature-dependence of the SC within

the temperature domain of liquid water. Both approaches lead to similar results and recover the temperature-dependence of the SC that is usually observed experimentally in such solutions, i.e., a decreasing SC as temperature is lowered and premises of a sign inversion around 270–280 K.<sup>4,5</sup>

Finally, we extend our approach to other solutes. In agreement with experimental data,<sup>4,5</sup> the SC is generally larger for bigger/heavier solutes but is also seen to depend on the solute chemistry, i.e., its interactions with solvent water molecules. A much more detailed and systematic study would be required to extensively discuss these aspects, but our results readily demonstrate that such a simulation strategy can be deployed and trusted in order to reproduce and study thermophoresis in dilute aqueous solutions *in silico* at an atomistic level. It would be interesting to compare the results of simulations performed in such thermophoretic settings, with that of the approach suggested by Galliero et al.<sup>60</sup> and later refined by Lusebrink et al.,<sup>23</sup> which proposes to quantify thermodiffusion by measuring the force acting on harmonically restrained particle in a solvent (infinite dilution limit). We also hope that more and more experimental data can be accumulated on these systems, as they are most often performed on binary mixtures (that can be seen as concentrated solutions), and as different techniques might lead to conflicting results.<sup>3,61</sup> Combined with a computational effort along the lines presented here, and as others have also pursued before, these could contribute to a better understanding of the molecular origin of thermophoresis in dilute solutions.

**Acknowledgement** The research leading to these results has received funding from the European Research Council under the European Union’s Eighth Framework Program (H2020/2014-2020)/ERC Grant Agreement No. 757111 (G.S.). This work was also supported by the "Initiative d’Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01 to GS). The simulations presented here benefited from a local computing platform administered by G. Letessier, and was

granted access to the HPC resources of TGCC under the allocation A0070811005 made by GENCI (Grand Equipement National de Calcul Intensif).

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- (1) Ludwig, C. Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösungen. *Ber Akad. Wiss. Wien Math-Naturw. KI.* **1856**, *20*, 539.
- (2) Rahman, M. A.; Saghir, M. Z. Thermodiffusion or Soret effect: Historical review. *Int. J. Heat Mass Transf.* **2014**, *73*, 693–705.
- (3) Köhler, W.; Morozov, K. I. The Soret Effect in Liquid Mixtures - A Review. *J. Non-Equil. Thermodyn.* **2016**, *41*, 151–197.
- (4) Niether, D.; Wiegand, S. Thermophoresis of biological and biocompatible compounds in aqueous solution. *J. Phys.: Condens. Matter* **2019**, *31*, 503003.
- (5) Piazza, R. Thermophoresis: moving particles with thermal gradients. *Soft Matter* **2008**, *4*, 1740.
- (6) Messaud, F. A.; Sanderson, R. D.; Runyon, J. R.; Otte, T.; Pasch, H.; Williams, S. K. R. An overview on field-flow fractionation techniques and their applications in the separation and characterization of polymers. *Prog. Pol. Sci.* **2009**, *34*, 351–368.
- (7) Duhr, S.; Braun, D. Thermophoretic depletion follows Boltzmann distribution. *Phys. Rev. Lett.* **2006**, *96*, 168301.
- (8) Iacopini, S.; Piazza, R. Thermophoresis in protein solutions. *Europhys. Lett.* **2007**, *63*, 247–253.
- (9) Duhr, S.; Braun, D. Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19678–19682.
- (10) Mast, C. B.; Braun, D. Thermal trap for DNA replication. *Phys. Rev. Lett.* **2010**, *104*, 188102.
- (11) Mast, C. B.; Schink, S.; Gerland, U.; Braun, D. Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 8030–5.
- (12) Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Commun.* **2010**, *1*, 100.
- (13) Vigolo, D.; Rusconi, R.; Stone, H. A.; Piazza, R. Thermophoresis: microfluidics characterization and separation. *Soft Matter* **2010**, *6*, 3489.
- (14) Braun, D.; Libchaber, A. Thermal force approach to molecular evolution. *Phys. Biol.* **2004**, *1*, 1–8.
- (15) Martin, W.; Russell, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2007**, *362*, 1887–1925.
- (16) Würger, A. Is Soret equilibrium a non-equilibrium effect? *C. R. - Mec.* **2013**, *341*, 438–448.
- (17) Artola, P.-A.; Rousseau, B. Thermal diffusion in simple liquid mixtures: what have we learnt from molecular dynamics simulations? *Mol. Phys.* **2013**, *111*, 3394–3403.
- (18) Nieto-Draghi, C.; Valos, J. B.; Rousseau, B. Computing the Soret coefficient in aqueous mixtures using

- p>boundary driven nonequilibrium molecular dynamics.
- J. Chem. Phys.*
- 2005**
- ,
- 122*
- , 114503.
- (19) Artola, P. A.; Rousseau, B.; Galliéro, G. A new model for thermal diffusion: Kinetic approach. *J. Am. Chem. Soc.* **2008**, *130*, 10963–10969.
  - (20) Römer, F.; Wang, Z.; Wiegand, S.; Bresme, F. Alkali halide solutions under thermal gradients: Soret coefficients and heat transfer mechanisms. *J. Phys. Chem. B* **2013**, *117*, 8209–8222.
  - (21) Calero, C.; Sibert, E.; Rey, R. Self-thermophoresis at the nanoscale using light induced solvation dynamics. *Nanoscale* **2020**, *12*, 7557–7562.
  - (22) Luettmer-Strathmann, J. Lattice model for thermodiffusion in polymer solutions. *Int. J. Thermophys.* **2005**, *26*, 1693–1707.
  - (23) Lüsebrink, D.; Yang, M.; Ripoll, M. Thermophoresis of colloids by mesoscale simulations. *J. Phys.: Condens. Matter* **2012**, *24*, 284132.
  - (24) Baaske, P.; Weinert, F. M.; Duhr, S.; Lemke, K. H.; Russell, M. J.; Braun, D. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9346–9351.
  - (25) Niether, D.; Afanasenkau, D.; Dhont, J. K.; Wiegand, S. Accumulation of Formamide in hydrothermal pores to Form prebiotic nucleobases. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 4272–4277.
  - (26) Reith, D.; Müller-Plathe, F. On the nature of thermal diffusion in binary Lennard-Jones liquids. *J. Chem. Phys.* **2000**, *112*, 2436.
  - (27) 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 Ph. Equilibria* **2003**, *208*, 171–188.
  - (28) Artola, P. A.; Rousseau, B. Microscopic interpretation of a pure chemical contribution to the Soret effect. *Phys. Rev. Lett.* **2007**, *98*, 125901.
  - (29) Polyakov, P.; Zhang, M.; Müller-Plathe, F.; Wiegand, S. Thermal diffusion measurements and simulations of binary mixtures of spherical molecules. *J. Chem. Phys.* **2007**, *127*, 014502.
  - (30) Di Lecce, S.; Albrecht, T.; Bresme, F. The role of ion-water interactions in determining the Soret coefficient of LiCl aqueous solutions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 9575–9583.
  - (31) 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.* **2018**, *20*, 1012–1020.
  - (32) Ikeshoji, T.; Hafskjold, B. Non-equilibrium molecular dynamics calculation of heat conduction in liquid and through liquid-gas interface. *Mol. Phys.* **1994**, *81*, 251–261.
  - (33) Müller-Plathe, F. A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. *J. Chem. Phys.* **1997**, *106*, 6082.
  - (34) Nieto-Draghi, C.; Avalos, J. B. Non-equilibrium momentum exchange algorithm for molecular dynamics simulation of heat flow in multicomponent systems. *Mol. Phys.* **2003**, *101*, 2303–2307.
  - (35) Kuang, S.; Gezelter, J. D. A gentler approach to RNEMD: Nonisotropic velocity scaling for computing thermal conductivity and shear viscosity. *J. Chem. Phys.* **2010**, *133*, 164101.
  - (36) Wirnsberger, P.; Frenkel, D.; Dellago, C. An enhanced version of the heat exchange

- algorithm with excellent energy conservation properties. *J. Chem. Phys.* **2015**, *143*, 124104.
- (37) Martinez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
  - (38) Abascal, J. L.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*.
  - (39) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
  - (40) Hockney, R. W.; Eastwood, J. W. *Bristol: Hilger*; 1988.
  - (41) 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* **2003**, *107*, 5342–5351.
  - (42) Hözl, C.; Kibies, P.; Imoto, S.; Frach, R.; Suladze, S.; Winter, R.; Marx, D.; Horinek, D.; Kast, S. M. Design principles for high-pressure force fields: Aqueous TMAO solutions from ambient to kilobar pressures. *J. Chem. Phys.* **2016**, *144*, 144104.
  - (43) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Mixtures of Urea and Water. *J. Phys. Chem. B* **2003**, *107*, 3891–3898.
  - (44) 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* **2016**, *120*, 7714–7723.
  - (45) Gonzalez-Salgado, D.; Vega, C. A new intermolecular potential for simulations of methanol: The OPLS/2016 model. *J. Chem. Phys.* **2016**, *145*, 034508.
  - (46) Jamali, S.; Hossein, van Westen, T.; 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. Theo. Comput.* **2018**, *14*, 6690–6700.
  - (47) Polak, E.; Ribiere, G. Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* **1969**, *3*, 35–43.
  - (48) Bonella, S.; Ferrario, M.; Ciccotti, G. Thermal Diffusion in Binary Mixtures: Transient Behavior and Transport Coefficients from Equilibrium and Nonequilibrium Molecular Dynamics. *Langmuir* **2017**, *33*, 11281–11290.
  - (49) James, D. W. The thermal diffusivity of ice and water between -40 and + 60 ° C. *J. Mat. Sci.* **1968**, *3*, 540–543.
  - (50) Mondal, J.; Halverson, D.; Li, I. T. S.; Stirnemann, G.; Walker, G. C.; Berne, B. J. How osmolytes influence hydrophobic polymer conformations: A unified view from experiment and theory. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 9270–5.
  - (51) Mondal, J.; Stirnemann, G.; Berne, B. J. When Does Trimethylamine N-Oxide Fold a Polymer Chain and Urea Unfold It? *J. Phys. Chem. B* **2013**, *117*, 8723–8732.
  - (52) Canchi, D. R.; García, A. E. Cosolvent effects on protein stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293.
  - (53) Stirnemann, G.; Hynes, J. T.; Laage, D. Water hydrogen bond dynamics in aqueous solutions of amphiphiles. *J. Phys. Chem. B* **2010**, *114*, 3052–3059.
  - (54) Laage, D.; Stirnemann, G.; Hynes, J. T. Why water reorientation slows without

iceberg formation around hydrophobic solutes. *J. Phys. Chem. B* **2009**, *113*, 2428–2435.

- (55) Hunger, J.; Tielrooij, K. J.; Buchner, R.; Bonn, M.; Bakker, H. J. Complex formation in aqueous trimethylamine-N-oxide (TMAO) solutions. *J. Phys. Chem. B* **2012**, *116*, 4783–4795.
- (56) Rezus, Y. L.; Bakker, H. J. Observation of immobilized water molecules around hydrophobic groups. *Phys. Rev. Lett.* **2007**, *99*, 148301.
- (57) Stirnemann, G.; Sterpone, F.; Laage, D. Dynamics of water in concentrated solutions of amphiphiles: Key roles of local structure and aggregation. *J. Phys. Chem. B* **2011**, *115*, 3254–3262.
- (58) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. 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.* **2009**, *31*, 671–690.
- (59) Römer, F.; Lervik, A.; Bresme, F. Nonequilibrium molecular dynamics simulations of the thermal conductivity of water: A systematic investigation of the SPC/E and TIP4P/2005 models. *J. Chem. Phys.* **2012**, *137*, 074503.
- (60) Galliero, G.; Volz, S. Thermodiffusion in model nanofluids by molecular dynamics simulations. *J. Chem. Phys.* **2008**, *128*, 064505.
- (61) Platten, J. K.; Bou-Ali, J. K.; Blanco, P.; Madariaga, J. A.; Santamaria, C. Soret Coefficients in Some Water-Methanol, Water-Ethanol, and Water-Isopropanol Systems. *J. Phys. Chem. B* **2007**, *111*, 11524–11530.