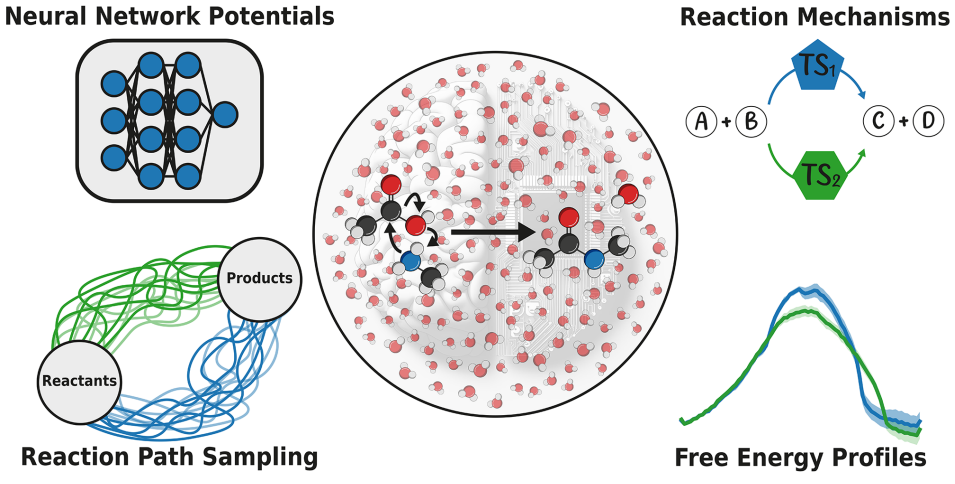
Comptes Rendus
de l'Académie des sciences

Chimie



Volume 27, Special Issue S5, 2024

Special Issue / *Numéro spécial*

French Network on Solvation (GDR 2035 SolvATE)

Réseau thématique sur la solvatation (GDR 2035 SolvATE)

Guest Editor / *Rédactrice en chef invitée*

Francesca Ingrosso

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Académie des sciences — Paris

ISSN: 1878-1543 (electronic)

CENTRE MERSENNE

Comptes Rendus

Chimie

Objective of the journal

Comptes Rendus Chimie is an international peer-reviewed electronic journal, covering all areas of chemistry.

It publishes special issues, original research articles, review articles, accounts, historical perspectives, pedagogical texts or conference proceedings, of unlimited length, in English or in French and in as flexible a format as necessary (figures, associated data, etc.).

Comptes Rendus Chimie has been published since 2020 with the centre Mersenne pour l'édition scientifique ouverte (Mersenne Center for open scientific publishing), according to a virtuous Diamond Open Access policy, free for authors (no author processing charges nor publishing fees) as well as for readers (immediate and permanent open access).

About the journal

Comptes Rendus Chimie is published exclusively in electronic format.

All information on the journal, as well as the full text of all articles, is available on its website at: https://comptes-rendus.academie-sciences.fr/chimie/.

Author enquiries

For any inquiries about submitting a manuscript, please refer to the journal's website: https://comptes-rendus.academie-sciences.fr/chimie/.

Contact

Académie des sciences
23 quai de Conti
75006 Paris (France)
cr-chimie@academie-sciences.fr

Comptes Rendus
Chimie
2024, 27, nº S5

Contents / *Sommaire*

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 1
https://doi.org/10.5802/crchim.438

French Network on Solvation (GDR 2035 SolvATE)

Guest Editor

Francesca Ingrosso joined the University of Lorraine faculty in 2008, after obtaining her PhD at the University of Pisa in 2005, a post-doctoral appointment at Colorado State University (Branka Ladanyi's group), and two postdocs at the École Normale Supérieure in Paris (Casey Hynes' group), one of which was funded by a Marie Curie Individual Fellowship. Her research focuses on modeling solvation at a molecular level, with a special interest in complex environments and time-dependent phenomena. Molecular properties and reactivity are studied by means of quantum chemistry, molecular dynamics as well as multi-scale techniques. Since 2018, she has been the director of the French Network on solvation, SolvATE.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 3-4
https://doi.org/10.5802/crchim.437

Foreword

French Network on Solvation (GDR 2035 SolvATE)

French network on solvation (GDR 2035 SolvATE): Foreword

Francesca Ingrosso

It is fitting to begin this historical overview with an anecdote. The first time the possibility of developing a national network dedicated to solvation was raised by the two future members of the SolvATE leadership dates back to December 2014, on the occasion of a PhD defense in Nancy. The location was particularly appropriate, as we were in the premises of the former laboratory of Professors Jean Barriol and Jean-Louis Rivail. The memory of those discussions never faded, and the idea of establishing a CNRS Research Network (GDR being the French acronym) gradually took shape.

In May 2017, following an initial email contact about solvation research from the perspective of both French theorists and experimentalists, we quickly moved toward organizing a "Perspectives Meeting" in Paris. Twenty-five researchers, representing about twenty laboratories (from the Paris region, Lille, Nancy, Dijon, Toulouse, Rennes, Clermont-Ferrand, Lyon, Strasbourg, and Poitiers), discussed in detail the scientific goals of the GDR, as well as its budget structure and governance. A detailed report was then sent to all interested laboratories, enabling the integration of additional suggestions and comments into the final version of the project, submitted in September 2017, together with the individual laboratory forms signed by the various institutional bodies.

In the autumn of 2017, further discussions with the CNRS Institute of Chemistry (INC), following the submission of the project, helped identify potential weaknesses and avenues for improvement. In particular, exchanges with Professor Philippe Hapiot led to contact with other French laboratories with significant research activity in the field of solvation. A meeting at CNRS headquarters in Paris with Professors Claire-Marie Pradier and Daniel Borgis provided valuable insight into the need to better structure the scientific project and clarify its governance. All of these highly constructive remarks led to the creation of SolvATE (GDR 2035), which was officially announced in January 2018. At the same time, the SolvATE website[1] and mailing list were created and have since served as communication tools within the network.

An essential mission of the SolvATE network is to bring forward the broad issue of solvation as a major transversal field in chemistry and physical chemistry. Our aim is to promote exchanges and interactions among French researchers, both theoreticians and experimentalists, who investigate solvent effects at the molecular level to understand chemical processes from different yet complementary perspectives, and to encourage collaborations with high level international groups. We are proud to see that, after eight years of shared activity, we are a highly dynamic community conducting research of international standing. Through our meetings and scientific actions, we stimulated the emergence of cutting-edge projects in order to build competitive funding applications for national and international calls,

[1] https://solvate.cnrs.fr/

to increase the visibility of the community to partners outside academia. We have actively involved young scientists (PhD students and postdoctoral researchers), and organized initiatives for outreach and research training, the details of which can be found on our website.

Our activities are structured in three research axes. In Axis 1 (Solvents in chemistry: toward a sustainable future), we focus on developing experimental and theoretical techniques to elucidate solvent effects on chemical reactivity, in a context of innovation aimed at environmentally responsible technological and industrial applications.

Within Axis 2 (Solvation and interfaces/surfaces, nanoconfined environments), we study chemical processes occurring at interfaces/surfaces and within nanoconfined media display specific features that can be exploited to construct a new chemistry in solution. In phenomena underlying many chemical processes (distillation, extraction, materials development) as well as in fields such as electrochemistry and microfluidics, a microscopic understanding of the behavior of solvated molecules in the presence of interfaces is essential.

Finally, in Axis 3 (Solvation in systems of biological, pharmaceutical, and agri-food interest), we engage both in fundamental research and in applications related to the design of pharmaceutical and agri-food processes.

In conclusion, the establishment and maturation of the SolvATE network has led to the consolidation of a vibrant and cohesive scientific community, now firmly recognized at the national level for its expertise in solvation science. The structuring efforts undertaken over the past years have created a solid foundation for future developments. Building on this momentum, the community is now well positioned to broaden its scope, foster new strategic alliances, and engage more actively with European partners. This special issue thus marks not an endpoint, but rather a decisive step toward expanding the visibility and impact of SolvATE beyond national borders, with the ambition of contributing to a wider European research landscape in solvation science.

Francesca Ingrosso
Université de Lorraine and CNRS
Laboratoire de Physique et Chimie Théoriques
UMR 7019, Nancy, F-54000
France
Francesca.Ingrosso@univ-lorraine.fr

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 5-14
https://doi.org/10.5802/crchim.368

Research article

French Network on Solvation (GDR 2035 SolvATE)

Thermal, dielectric, and electrochemical study of decanoic acid–tetrabutylammonium chloride deep eutectic solvent

Mohammad Nadim Kamar [a], Ludovic Paquin [*,b], Emmanuelle Limanton [b], Corinne Lagrost [*,b] and Denis Morineau [*,a]

[a] Institut de Physique de Rennes, CNRS-Université de Rennes, UMR 6251, F-35042 Rennes, France

[b] Institut des Sciences Chimiques de Rennes, CNRS-Université de Rennes, UMR 6226, F-35042 Rennes, France

E-mails: ludovic.paquin@univ-rennes.fr (L. Paquin), corinne.lagrost@univ-rennes.fr (C. Lagrost), denis.morineau@univ-rennes.fr (D. Morineau)

Abstract. A mixture based on decanoic acid (DA) and tetrabutylammonium chloride (TBACl) is a simple and prototypical deep eutectic solvent (DES) useful for extracting compounds that are poorly soluble in water or in electrochemical applications. The most widely studied composition is DA–TBACl with a molar ratio equal to 2:1. The composition of DESs has a strong impact on their physicochemical properties. Herein, a comparative study of thermal, dielectric, and ionic conductivities and electrochemical properties of DA–TBACl is carried out with varying molar compositions 2:1, 1:1, and 1:2. All the molar compositions lead to a stable fluid whose properties depend on the composition, providing better understanding and further insights into the chemical interactions that prevail in those materials.

Keywords. Deep eutectic solvents, Dielectric spectroscopy, Electrochemistry.

Funding. Rennes Metropole and European Union (ERDF—CPER PRINT2TAN), CNRS—Network SolvATE (GDR 2035).

Manuscript received 27 June 2024, revised 26 September 2024, accepted 26 November 2024.

1. Introduction

Since pioneering studies in the early 2000s [1], deep eutectic solvents (DESs) have developed as a new class of alternative solvents, complementing the potential offered by other unconventional media such as ionic liquids. They are promising in many applications [2,3], notably as extracting media for natural compounds or for pesticide recovery or even for CO_2 capture to name a few [4–8].

Conservatively, a DES is defined as a mixture of two or more H-bonded compounds that exhibits an eutectic point and large non-ideal mixing effects [9]. The non-ideal character of DES that relates to specific interactions between a hydrogen-bond donor (HBD) and a hydrogen-bond acceptor (HBA) significantly enhances the temperature depression of the melting point at the eutectic point. Interestingly, this effect has significantly expanded the number of candidates as ingredients for the formulation of these new solvents, as it allows the inclusion of compounds that are normally found in solid form at room

*Corresponding authors

temperature in their pure state. Beyond melting, non-ideality also implies that DESs present properties that differ from those of their constituents when considered independently, thus offering new opportunities to design solvents for specific applications.

The DESs have been divided into five classes. Notably, class III is of interest in this work where the HBA is an ionic constituent of an organic salt, typically a quaternary ammonium halide similarly to ionic liquids [10]. Type III systems have been largely represented in the DES panel, and are especially interesting for electrochemical applications. Most DESs proposed so far have been of hydrophilic nature although increasing their hydrophobicity has become an important goal in order to expand their application scope [11,12]. For instance, higher hydrophobicity is interesting because it provides DESs the ability to dissolve natural products with limited water solubility like carotenoids for instance, also helping preserve their antioxidant properties [7] and enabling better performance in CO_2 capture [8].

These characteristics are promoted by the combination of organic salts comprising long alkyl chains with poorly water miscible HBD molecules such as fatty acids. Since a pioneering study in 2015, such systems have been commonly classified as "hydrophobic DESs" due to their low water content and low ion leaching after mixing with water [12].

They have generally higher viscosities than hydrophilic DESs and exhibit lower to very weak conductivities. Few of them could be used as electrolytes in electrochemistry since the "hydrophobic DESs" often belong to type V, that is, composed solely of molecular substances [13,14]. Yet, for application in electrochemistry, the composition of "hydrophobic DESs" necessarily involves an organic salt. However, a question arises about the evolution of polarity and conductivity as a function of molar composition and temperature. Herein, we address this point by focusing on one of the first reported and since widely studied "hydrophobic DESs" consisting of a mixture of decanoic acid (DA) as the HBD and tetrabutylammonium chloride (TBACl) as the HBA (Figure 1). We present a systematic study of thermal, dielectric, and ionic conductivities and electrochemical properties of the DA–TBACl mixture as a function of molar composition for three compositions (1:2, 1:1, 2:1) in order to cover the range of interest related to previous studies, including the most widely studied one

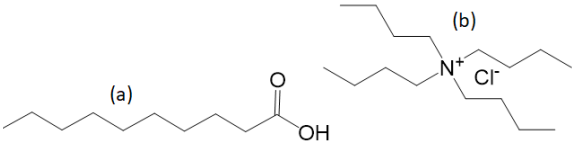


Figure 1. Chemical structure of (a) decanoic acid and (b) tetrabutylammonium chloride.

(2:1) [11,12]. The mixing of the corresponding eutectic mixtures with a large amount of water is also qualitatively investigated according to the molar ratios of eutectic components. The results show the strong impact of the hydrophobic DA component with respect to the ammonium salt to design a stable solvent of low polarity having good conductivity, reasonable viscosity, and good electrochemical property.

2. Materials and methods

The DA–TBACl liquid mixtures were prepared for three different compositions denoted as DA–TBACl (n:m), with n:m corresponding to the molar stoichiometry. DA and TBACl were purchased from Acros Organics and were used without further purification. DA (melting point 31.5 °C) and TBACl (melting point 83 °C–86 °C) were mixed in molar ratios 1:1, 2:1, and 1:2 at 80 °C under stirring for 2 h until a homogeneous and transparent liquid was obtained. Then, the mixtures were allowed to cool at room temperature and stored under ambient conditions. The colorless fluids are visually stable over time (more than 1 year). The water content of the resulting mixtures was determined by Karl Fischer titration (831 KF Coulometer with a generator electrode with diaphragm, Metrohm) with an average of three measurements: DA–TBACl (1:1) has 2.8% of water (w/w); DA–TBACl (2:1) has 1.7% of water (w/w); DA–TBACl (1:2) has 3.4% of water (w/w).

For differential scanning calorimetry (DSC) experiments, the weighted samples were sealed in Tzero© aluminum hermetic pans. The measurements were taken with a Q20 TA instrument equipped with a liquid nitrogen cooling system. The standard calibration of temperature and heat flux was performed by measuring the melting transition of an indium sample. Thermograms were acquired on cooling, followed by a heating ramp in the temperature range from −120 °C to 35 °C with the same scanning rate

of 5 °C·min^{-1}. For DA–TBACl (2:1), an additional measurement of the heating ramp at 5 °C·min^{-1} was recorded after a fast thermal quench at the maximum cooling rate (approximately 200 °C·min^{-1}) to avoid crystallization (see the discussion in Section 3 for details).

For dielectric spectroscopy experiments, the samples were injected with a pipette between two stainless steel electrodes maintained by Teflon spacers in order to form a parallel plate capacitor geometry with a diameter of 20 mm and a spacing of 260 µm. During this operation, both the liquids and the sample cell were heated up to about 60 °C to decrease viscosity and ensure fast and complete filling of the cell by the action of capillary forces. Then, the cell was placed in a cryostat and maintained under a dry nitrogen atmosphere. The complex impedance of the as-prepared capacitor was measured from 1 Hz to 10^6 Hz with a Novocontrol high-resolution dielectric Alpha Analyzer with an active sample cell. The measurements were taken at thermal equilibrium along a cooling branch and a subsequent heating branch with a temperature step of 2 °C, and typically covering the temperature range from −120 °C to 60 °C. The temperature of the samples was controlled by a Quatro temperature controller (Novocontrol) with nitrogen as the heating/cooling agent providing a temperature stability within 0.1 °C. The temperature scan rate, although discontinuous, was 0.3 °C·min^{-1} on average.

For electrochemical experiments, the measurements were performed by using a home-made three-electrode cell [15]. A glassy carbon disk electrode (Ø 1 mm), a platinum wire, and a silver wire were used as working, counter, and quasi-reference electrodes, respectively. The working electrode was carefully polished with SiC paper and diamond paste (Struers) and then rinsed with ultrapure water and dried with argon flow prior to experiments. Cyclic voltammetry (CV) was carried out with an Autolab PGSTAT30 potentiostat/galvanostat (Metrohm Autolab BV).

3. Results and discussion

3.1. *Differential scanning calorimetry*

The phase behavior of the three different samples was determined by DSC. For the three systems, the same thermal cycling was first applied. It consists in cooling from 35 °C to −120 °C and heating up to 35 °C at 5 °C·min^{-1}. For DA–TBACl (1:2) and DA–TBACl (1:1), no crystallization was observed, neither on cooling nor on heating as presented in Supplementary Information (Figures S1 and S2). This means that for these compositions, DA–TBACl mixtures can be easily supercooled and form very good glass-forming systems. Indeed, a glass transition was clearly demonstrated by a jump in the heat capacity. In contrast for DA–TBACl (2:1), crystallization occurred during cooling as indicated by an exothermic peak in Figure S3. This crystallization concerned only a fraction of the sample. In fact, the remaining liquid phase performed a glass transition and eventually fully crystallized during the subsequent heating (Figure S3). In order to form a glass that was free from partial crystallization, the DA–TBACl (2:1) system was quenched by cooling at the maximum rate (200 °C·min^{-1}). Under these conditions, the absence of an exothermic signal during cooling indicated the successful formation of a pure amorphous glassy state.

On heating above the glass transition temperature T_g, the supercooled liquid of DA–TBACl (2:1) crystallized at −55 °C (cold crystallization) and then melted on a broad temperature with two main endothermic peaks centered at −25 °C and −10 °C. These features are characteristic of the melting of eutectic forming binary systems at a composition that differs from the eutectic point [16]. On the other hand, it has been shown that the DES exhibits a better glass-forming tendency for compositions neighboring the eutectic one as seen for the other two DA–TBACl systems. The precise determination of the eutectic point, though interesting, would require a systematic study of the entire phase diagram, which lies out of the scope of the present study dedicated to liquid phase properties. Moreover, due to its very good glass-forming capability on a broad range of compositions, the precise determination of the eutectic point would probably be elusive.

In order to focus on the long-time liquid dynamics, the glass transitions of the three samples are compared in Figure 2. On increasing the relative amount of organic salt (TBACl) in the mixture, the position of the heat capacity jump systematically shifted to a higher temperature, with the glass transition temperatures being respectively T_g = −83.8 °C, −76 °C, and

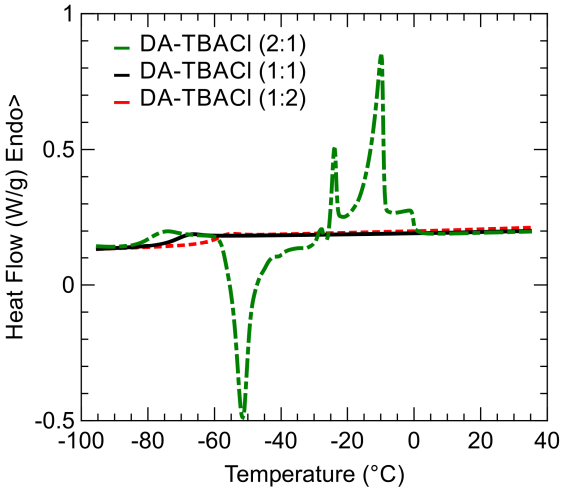


Figure 2. Thermograms measured during heating at 5 °C·min^{-1} of DA–TBACl mixtures with compositions 2:1 (dash-dotted green), 1:1 (solid black), and 1:2 (dashed red). These heating ramps were acquired after cooling at the same rate except for DA–TBACl (2:1), where a thermal quench was applied.

−63.1 °C. This means that for this range of compositions, the relaxation dynamics of the mixture slow down by adding TBACl to DA. This is in agreement with the observed increase in viscosity of the solvents at room temperature.

This dependence of T_g on the composition can be interpreted as resulting from an increase in intermolecular correlations between the different species of the mixtures due to enhanced H-bonds and electrostatic interactions between DA and TBACl. Such interactions are often invoked in DESs and also related to unusual thermodynamic or structural features, such as the formation of supramolecular species [17–22]. However, it is worth pointing out that mixing effects on the dynamical properties of DESs are not systematically observed. For instance, opposite effects of adding choline chloride into neat polyols have been reported for glyceline and ethaline [23]. As a whole, the different dependences of the glassy dynamics on the composition of the liquid mixture illustrate the complexity of intermolecular correlations in DESs that result from the balance between H-bonds, electrostatic interactions, and herein, hydrophobic interactions.

The calorimetric glass transition offers a limited view of the DES dynamics, typically restricted to a relaxation timescale of the order of 10^2 s. This corresponds to a temperature at which the system is extremely viscous. In order to link the glass transition to the actual dynamics in the fluid liquid state, a complementary study by spectroscopic methods is valuable.

3.2. *Dielectric spectroscopy*

Dielectric spectroscopy has been demonstrated to be a very powerful method for DES studies [16,23–25]. It covers an extended dynamical range bridging the gap from molecular to calorimetric timescales while providing insights into both the ionic transport (conductivity) and dipolar relaxation (polarization).

The complex dielectric function of the sample $\varepsilon^*(f) = \varepsilon'(f) - i\varepsilon''(f)$ was measured for the three samples by cooling steps of 2 °C, where f denotes the frequency of the electric field that ranges from 0.1 Hz to 1 MHz, ε' and ε'' are the real and loss parts of the complex dielectric function, and i symbolizes the imaginary unit [26]. They are illustrated in Figure 3 for DA–TBACl (1:1) and for a selection of temperatures. Different contributions can be identified as a function of frequency and temperature. First, a considerable increase in the real part of permittivity was observed at high temperature and low frequency. This effect is classically attributed to electrode polarization induced by the accumulation of ions at the surface of the blocking electrodes [26]. This phenomenon does not reveal any useful physical information about liquid properties, and its contribution to the total intensity could be simply accounted for by a phenomenological power law function.

The most remarkable observed feature is the prominent dipolar relaxation process that is illustrated in Figure 3a by the jump of $\varepsilon'(f)$ from static permittivity ε_s to high-frequency permittivity ε_∞. This relaxation process is also apparent in the loss part $\varepsilon''(f)$ although its corresponding peak (see the dashed line in Figure 3b) is overwhelmed by conductivity that additionally contributes to $\varepsilon''(f)$.

The dielectric relaxation and the ionic conductivity were analyzed quantitatively at each temperature by fitting a model comprising a Havriliak and Negami

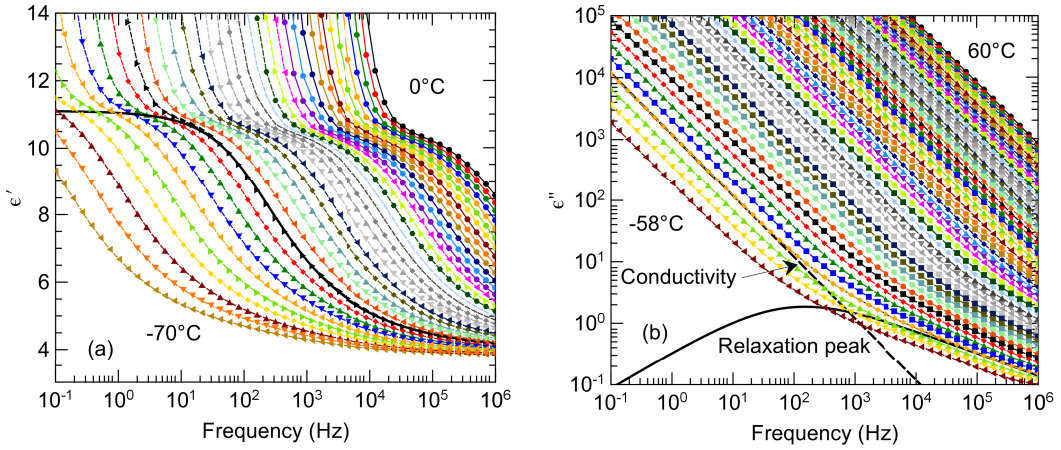


Figure 3. (a) Real and (b) loss parts of the complex dielectric function of DA–TBACl (1:1) as a function of frequency for a selection of temperatures regularly spaced by steps of 2 °C (the lowest and highest temperature values are indicated in each panel). The total fitted functions (thin dashed lines) are virtually indistinguishable from the experimental data points (symbols). The individual contributions from dipolar relaxation (thick black solid line) and conductivity (thick dashed line) to the total fitted function are illustrated for the temperature $T = -52$ °C.

function (HN-model) [27] and a dc-conductivity term according to Equation (1).

$$\varepsilon^*(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{(1+(\mathrm{i}\omega\tau_{\mathrm{HN}})^{\alpha_{\mathrm{HN}}})^{\beta_{\mathrm{HN}}}} - \mathrm{i}\frac{\sigma}{\omega\varepsilon_0} \qquad (1)$$

In this model, $\omega = 2\pi f$, ε_∞ is the sample permittivity in the limit of high frequency, and $\Delta\varepsilon = \varepsilon_s - \varepsilon_\infty$ and τ_{HN} are, respectively, the dielectric strength and the HN relaxation time; σ stands for the dc-conductivity of the sample and ε_0 the permittivity of vacuum. According to the formalism of the HN-model, the exponents α_{HN} and β_{HN} ($0 < \alpha_{\mathrm{HN}}$; $\alpha_{\mathrm{HN}}\beta_{\mathrm{HN}} \leq 1$) are fractional parameters describing, respectively, the symmetric and asymmetric broadening of the complex dielectric function with respect to the Debye model.

First, we consider the static dielectric permittivity ε_s, which is an important indicator of the polar character of solvents. For DA–TBACl (2:1), we found $\varepsilon_s = 6.5 \pm 0.5$, with negligible temperature effects in the range studied while larger values were found for the other two compositions ($\varepsilon_s = 11 \pm 1$). The increase in ε_s with the addition of TBACl when going from (2:1) to (1:1) is most probably associated with the large polar character of the ionic component TBACl. However, the saturation of ε_s with salt content when further increasing the fraction of TBACl from (1:1) to (1:2) demonstrates that not only the individual dipoles of species in the mixture but also their relative spatial arrangement determines the dielectric permittivity of the solvent. Indeed, the Kirkwood–Fröhlich formalism accounts for the role of angular correlations between the dipoles of different molecules present in the liquid. In this framework, the evolution of ε_s among the three samples would be consistent with the gradual formation of DA–TBACl supramolecular arrangements that promote the antiparallel dipolar configurations of TBACl ions [28]. This possibility definitely necessitates complementary structural characterizations as accessible by optical spectroscopy or diffraction methods.

Beyond static information gained from ε_s, we discuss now the liquid dynamics. The average relaxation time, which is classically related to the maximum peak position in the loss part of the complex dielectric function, was evaluated by Equation (2) [29].

$$\tau = \tau_{\mathrm{HN}} \sin\left(\frac{\pi\alpha_{\mathrm{HN}}}{2+2\beta_{\mathrm{HN}}}\right)^{-1/\alpha_{\mathrm{HN}}} \sin\left(\frac{\pi\alpha_{\mathrm{HN}}\beta_{\mathrm{HN}}}{2+2\beta_{\mathrm{HN}}}\right)^{1/\alpha_{\mathrm{HN}}} \qquad (2)$$

The temperature dependence of the relaxation time is illustrated in Figure 4 in Arrhenius coordinates. Note that for DA–TBACl (2:1), the accessible temperature range was reduced due to crystallization on cooling at about −36 °C. Deviation from the Arrhenius law was observed for the three samples. This

phenomenon is typical for supercooled liquids, and it has also been reported for many DESs [16,23–25]. It is often associated with the emergence of cooperativity, which leads to an increase in the apparent activation energy on approaching the glass transition. Very good fits to the data were achieved with the VFT law as illustrated by the solid line. The extrapolation of the relaxation time using the VFT law toward $\tau = 10^2$ s provides an estimate of the glass transition temperature that is in perfect agreement with the calorimetric temperature (see the symbol in Figure 3). This demonstrates that the dipolar relaxation measured by dielectric spectroscopy is directly coupled to the main structural relaxation of the liquid. A systematic slowdown of the relaxation dynamics is obtained when increasing the fraction of TBACl into the liquid mixture. This agrees with the conclusion made from the DSC part. In addition, we have used the Vogel–Fulcher–Tammann (VTF) fits to compute the fragility index m, which is a measure of deviation from the Arrhenius behavior [30]. The obtained values are in the range $m = 58$–81, and they increase with increasing TBACl fraction. They are located between values obtained for choline chloride based DESs and their aqueous solutions (m = 40–60) [23,24], which are classified as intermediate liquids on the one hand, and hydrophobic DESs based on menthol–thymol mixtures (m = 77–86), which are classified as fragile (i.e., showing larger deviation from the Arrhenius law) on the other hand [16].

In addition to the temperature dependence of the average relaxation time, another important feature is the deviation of the dipolar relaxation function from a simple Debye process. This salient behavior can be expressed in time domain by a stretched exponential function $e^{-(t/\tau)^{\beta}}$, which is also known as the Kohlrausch–Williams–Watts (KWW) law. While simple Debye relaxation is recovered for $\beta = 1$, a stretched relaxation process is obtained for lower values of the β exponent. We evaluated the value of β using HN fractional exponents obtained from the fit in the frequency domain and the numerical ansatz $\beta = (\alpha_{HN}\beta_{HN})^{1/1.23}$ [31].

On the temperature range studied, the stretching exponent was $\beta = 0.7 \pm 0.05$ for DA–TBACl (2:1), which indicates relatively weak deviation from the Debye law. On the contrary, much lower values of the KWW exponent ($\beta = 0.45 \pm 0.05$) were found for the other two compositions having a larger fraction of ionic species, namely, (1:1) and (1:2). For glass-forming liquids, the non-Debye relaxation behavior is often attributed to dynamic heterogeneity. For DESs, dynamic heterogeneity possibly stems from the association of different components of mixtures, which has been shown to result in the formation of mesoscopic domains [20–22]. In this context, the broader distribution of relaxation times can arise from molecules experiencing different local environments.

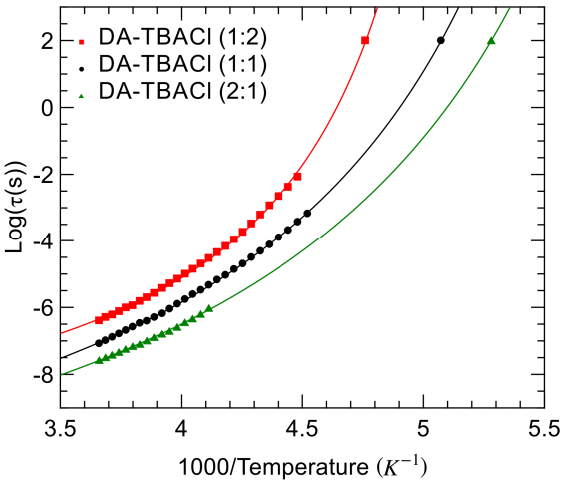


Figure 4. Arrhenius plot of the dipolar relaxation time of the studied DA–TBACl mixtures with compositions 2:1 (green triangles), 1:1 (black circles), and 1:2 (red squares). The calorimetric glass transition is indicated by the symbol located at $\tau = 10^2$ s. VTF fits are illustrated by solid lines.

3.3. *Ionic dc-conductivity*

Complementary to dipolar relaxation, dielectric spectroscopy experiments provide useful information about liquid dynamics from dc-conductivity. As shown in Figure 3b, the conductivity of DESs appears as an intense component of the loss part of the dielectric function, which is inversely proportional to frequency,

The temperature dependence of conductivity, as determined by the fitting of Equation (1), is illustrated in Figure 5 in Arrhenius coordinates. For all samples, a super-Arrhenius behavior is obtained, which again could be well reproduced by a VTF

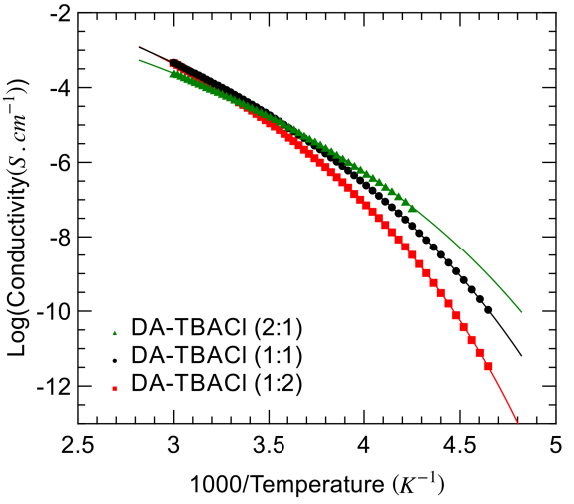


Figure 5. Arrhenius plot of the dc-conductivity of the studied DA–TBACl mixtures with compositions 2:1 (green triangles), 1:1 (black circles), and 1:2 (red squares). VTF fits are illustrated by solid lines.

model. This temperature behavior is close to that of dipolar relaxation, which suggests that both properties are actually linked. A possible explanation is that both processes reflect the temperature dependence of viscosity. On the one hand, dielectric relaxation relates to a large extent to the rotational dynamics of dipolar species. Assuming the validity of classical hydrodynamic laws, the relaxation time should scale with the viscosity η according to the Stokes–Einstein–Debye equation $\tau \propto 8\pi r\eta/(kT)$, with r being the hydrodynamic radius. On the other hand, the ionic conductivity should be inversely proportional to the viscosity, assuming that both the Nernst–Einstein and Stokes–Einstein relations apply. This predicted behavior was actually confirmed for DA–TBACl (2:1) and (1:1) as illustrated in Figure S4. Contrariwise, this scaling law was only partly obeyed for DA–TBACl (1:2), which showed deviation in the high temperature limit.

In the literature, different situations have been recently reported about the (de-)coupling between rotation and translation dynamics in DESs. In the case of the prototypical ionic DES ethaline, translation–rotation decoupling was observed neither for the neat mixture nor for its moderately hydrated variants [23–25]. For the non-ionic menthol–thymol DES instead, a power law $\sigma \propto \tau^{-\alpha}$ was observed, with a fractional exponent α that was close to unity for equimolar composition but increasingly deviated as the fraction of thymol in the mixture increased [16]. This partial decoupling was interpreted as a possible hint for the development of spatial dynamic heterogeneities in this range of compositions. Similarly, for reline, the deviation from the Walden rule that links ionic conductivity to viscosity was reported [25]. The presence of complex supramolecular structures was invoked as a possible origin of this unusual charge transport. Although based solely on dynamical properties, the observation made for DA–TBACl by increasing the amount of salt also points toward the possible formation of supramolecular entities involving both ionic (TBACl) and H-bond donors (DA).

Finally, the inversion in the order of conductivity values, which intersect approximately at room temperature, reflects the different impacts of TBACl on the liquid property. On the one hand, going from (2:1) to (1:2), DA–TBACl increases both the concentration of ionic mobile species and water content, which contributes to high conductivity at high temperature. On the other hand, it increases liquid fragility, which has a negative impact on conductivity at sub-ambient temperature since viscosity increases more drastically on cooling with TBACl stoichiometry.

3.4. *Electrochemistry*

The electrochemical behavior of the three mixtures is displayed in Figure 6. The CVs have been recorded using a glassy carbon electrode with anodic and cathodic potential limits arbitrarily chosen. In agreement with the results above, all the mixtures are sufficiently conducting to serve as electrolytes. The electrochemical window (i.e., the potential range between anodic and cathodic limits) is relatively shrunk compared to ionic liquids. The potential windows are about 2 V while ionic liquids show typical windows of 4.5–5 V [32]. This is probably due to the presence of water (2–3 wt%) and oxygen from air that accompanies water. However, the corresponding electrochemical windows are still larger than that of aqueous electrolytes.

Another interesting point is the shape of the CVs. The CV shape for DA–TBACl (1:2) is different from

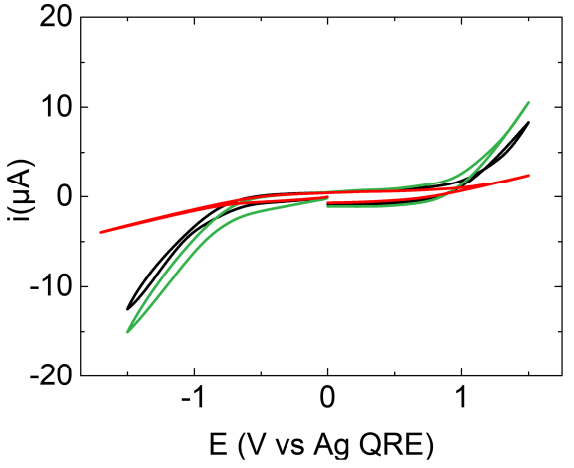


Figure 6. Cyclic voltammetries of the three DA–TBACl mixtures at 0.2 V·s^{-1} at a glassy carbon disk electrode: DA–TBACl (1:1) (black), DA–TBACl (2:1) (green), and DA–TBACl (1:2) (red).

that for DA–TBACl (1:1) and DA–TBACl (2:1). On the one hand DA–TBACl (1:1) and DA–TBACl (2:1) show similar behavior, but on the other hand the CV corresponding to DA–TBACl (1:2) is much more flattened than the former two. This indicates a more resistive behavior of DA–TBACl (1:2) than those of the other two compositions. This observation is in fair agreement with the ionic dc-conductivity analyses.

3.5. *Behavior of mixtures with water: qualitative examination*

Hydrophobicity of the DES could be qualitatively assessed by the appearance of a phase separation after being mixed with a large amount of water. After vigorous stirring at ambient temperature, the three compositions tend to phase-separate (Figure 7) just after mixing with 29 wt% of water. These behaviors are in good agreement with the dielectric measurements above, suggesting that all the compositions give non-polar fluids. However, after 3 days, another equilibrium is reached and only the composition DA–TBACl (2:1) still exhibits a phase separation with water while the other two compositions remain mixed with water (Figure 7). Considering that TBACl is a hydrophilic and DA a hydrophobic component, these observations show that the hydrophobicity character obviously stems from DA, which is in molar excess for this mixture composition. Of course, leaching of the hydrophilic component could occur in the water-rich phase as already observed [5]. The clear mixing of hydrophobic DES with water is known to be sensitive to the amount of water. For instance, the upper limit is found around 10 wt% for a DES formed with decanoic acid and tetrabutylammonium bromide (2:1) [8]. A higher value of the water content upper limit is nevertheless found herein although tetrabutylammonium bromide is less hydrophilic than its corresponding chloride counterpart. Interestingly, we also observe that water mixing is also strongly dependent on the molar ratio composition. A more precise microscopic description of the DES would be interesting to better understand the phenomenon. However, we may rationalize the observations on the basis of hydrogen-bonding effects since these interactions are of major importance in the preparation of these mixtures. Indeed in the DES, TBACl is likely to be hydrogen-bonded with DA to form a stable structure even if a small amount of water is part of the structure, probably through the formation of hydrogen bonds between the eutectic and water. When increasing the water content, a competition may occur between the formation of hydrogen bonds of water with the HBD and HBA components of the eutectic mixture and the formation of hydrogen bonds between the two components that preserve the eutectic structure. At a certain point, and for a large amount of water, the hydrogen bonds arise mainly from the association of water-isolated components, hence destroying the fluid structure to enable efficient mixing with water. In the case of choline chloride based DESs, a cross-over separating two distinct thermodynamic behaviors was observed for a water content of 30 wt% from a “water-in-DES” to “DES-in-water” situation although no macroscopic liquid–liquid phase separation occurred at room temperature due to their strong hydrophilic character compared to DA–TBACl [33–35]. A larger amount of HBD, namely, DA helps maintain the structure even if the eutectic point is not necessarily reached. Hydrophobic eutectic mixtures can form stable mixing systems with a certain molar fraction of water depending on their own molar composition.

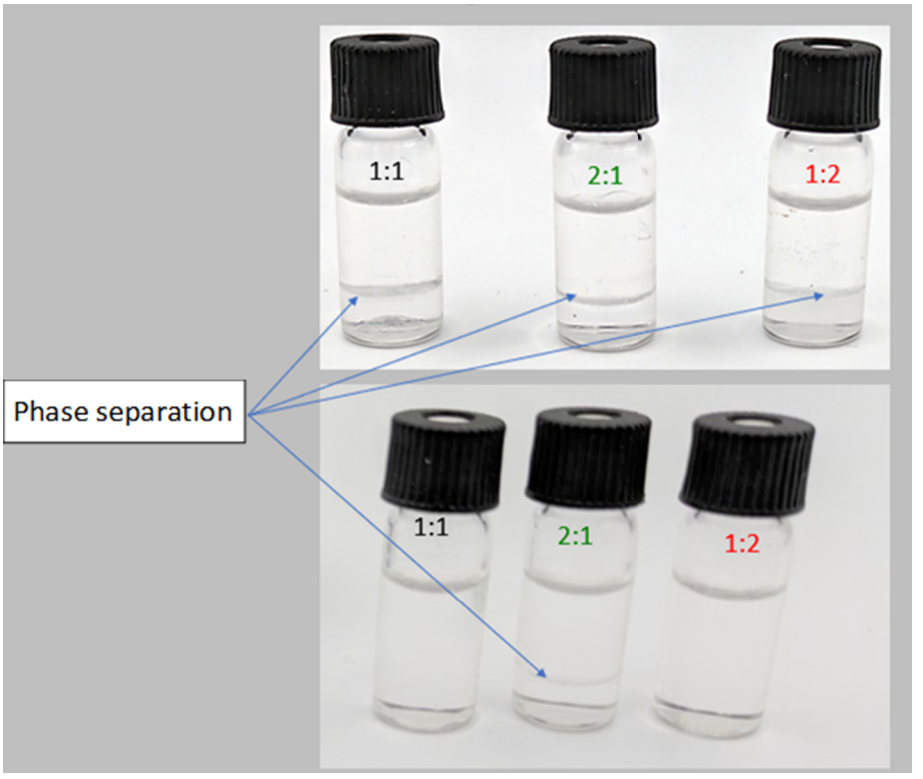


Figure 7. Pictures of 800 µL DA–TBACl (1:1), DA–TBACl (2:1), and DA–TBACl (1:2) just after mixing with 300 µL of water (top) and after 3 days (bottom).

4. Conclusion

In the quest for new solvents with designed properties, hydrophobic DESs are promising candidates for meeting important needs for the extraction of compounds that are poorly soluble in water or for electrochemical applications. The present experimental study highlights the specific interest in DESs based on decanoic acid and tetrabutylammonium chloride.

Combining experimental approaches, we demonstrate the low polarity of this solvent, which has suitable viscosity, good conductivity, and electrochemical property. Moreover, comparing three different compositions (n:m) 2:1, 1:1, and 1:2, we reveal the central impact of stoichiometry on these properties. This regards thermal stability, which is virtually unlimited for $n \leq m$ due to the suppression of crystallization. Furthermore, an acceleration by two orders of magnitude of molecular dynamics, as evaluated from dipolar relaxation and glass transition, is achieved by increasing the DA content. This also results in different ionic conductivities and electrochemical activities. Finally, the different long-time evolutions of DESs after water addition also indicate, at least at a qualitative level, that stoichiometry deeply influences the resistance of the DES structure to hydration. Overall, these findings support the interest in DA–TBACl as a solvent with real potential to fine-tune its properties by varying stoichiometry in order to adapt it to specific application requirements.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Acknowledgments

Support from Rennes Metropole and European Union (ERDF—CPER PRINT2TAN) is acknowledged.

The authors are grateful to the CNRS—Network SolvATE (GDR 2035) for financial support and fruitful discussions.

Supplementary data

Supporting information for this article is available on the journal's website under https://doi.org/10.5802/crchim.368 or from the author.

References

[1] A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed, V. Tambyrajah, *Chem. Commun.*, 2003, 70-71.
[2] B. B. Hansen, S. Spittle, B. Chen *et al.*, *Chem. Rev.*, 2021, **121**, 1232-1285.
[3] Q. Zhang, K. D. O. Vigier, S. Royer, F. Jérôme, *Chem. Soc. Rev.*, 2012, **41**, 7108-7146.
[4] C. H. J. T. Dietz, D. J. G. P. van Osch, M. C. Kroon, G. Sadowski, M. van Sint Annaland, F. Gallucci, L. F. Zubeir, C. Held, *Fluid Phase Equilibria*, 2017, **448**, 94-98.
[5] C. Florindo, L. C. Branco, I. M. Marrucho, *Fluid Phase Equilibria*, 2017, **448**, 135-142.
[6] L. Percevault, E. Limanton, P. Nicolas, L. Paquin, C. Lagrost, *ACS Sustain. Chem. Eng.*, 2021, **9**, 776-784.
[7] A. Viñas-Ospino, A. Rita Jesus, A. Paiva, M. J. Esteve, A. Frígola, J. Blesa, D. López-Malo, *Food Chem.*, 2024, **442**, article no. 138530.
[8] Q. Zheng, F. Yang, H. Tan, X. Wang, *J. Chem. Thermodyn.*, 2024, **195**, article no. 107306.
[9] M. A. R. Martins, S. P. Pinho, J. A. P. Coutinho, *J. Solution Chem.*, 2019, **48**, 962-982.
[10] A. P. Abbott, *Curr. Opin. Green Sustain. Chem.*, 2022, **36**, article no. 100649.
[11] D. J. G. P. van Osch, C. H. J. T. Dietz, S. E. E. Warrag, M. C. Kroon, *ACS Sustain. Chem. Eng.*, 2020, **8**, 10591-10612.
[12] D. J. G. P. Osch, L. F. van Zubeir, A. van den Bruinhorst, M. A. A. Rocha, M. C. Kroon, *Green Chem.*, 2015, **17**, 4518-4521.
[13] L. Bahadori, M. H. Chakrabarti, F. S. Mjalli, I. M. AlNashef, N. S. A. Manan, M. A. Hashim, *Electrochim. Acta*, 2013, **113**, 205-211.
[14] S. Ruggeri, F. Poletti, C. Zanardi *et al.*, *Electrochim. Acta*, 2019, **295**, 124-129.
[15] G. Loget, S. Chevance, C. Poriel, G. Simonneaux, C. Lagrost, J. Rault-Berthelot, *ChemPhysChem*, 2011, **12**, 411-418.
[16] C. D'Hondt, D. Morineau, *J. Mol. Liq.*, 2022, **365**, article no. 120145.
[17] O. S. Hammond, D. T. Bowron, K. J. Edler, *Green Chem.*, 2016, **18**, 2736-2744.
[18] S. Kaur, A. Gupta, H. K. Kashyap, *J. Phys. Chem. B*, 2016, **120**, 6712-6720.
[19] E. Posada, N. López-Salas, R. J. J. Riobóo, M. L. Ferrer, M. C. Gutiérrez, F. del Monte, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17103-17110.
[20] V. Alizadeh, D. Geller, F. Malberg, P. B. Sánchez, A. Padua, B. Kirchner, *ChemPhysChem*, 2019, **20**, 1786-1792.
[21] L. Percevault, A. Jani, T. Sohier, L. Noirez, L. Paquin, F. Gauffre, D. Morineau, *J. Phys. Chem. B*, 2020, **124**, 9126-9135.
[22] S. Kaur, M. Kumari, H. K. Kashyap, *J. Phys. Chem. B*, 2020, **124**, 10601-10616.
[23] D. Reuter, C. Binder, P. Lunkenheimer, A. Loidl, *Phys. Chem. Chem. Phys.*, 2019, **21**, 6801-6809.
[24] A. Jani, B. Malfait, D. Morineau, *J. Chem. Phys.*, 2021, **154**, article no. 164508.
[25] D. Reuter, P. Münzner, C. Gainaru, P. Lunkenheimer, A. Loidl, R. Böhmer, *J. Chem. Phys.*, 2021, **154**, article no. 154501.
[26] F. Kremer, A. Schönhals, *Broadband Dielectric Spectroscopy*, Springer, Berlin, Heidelberg, 2003.
[27] S. Havriliak, S. Negami, *J. Polym. Sci. C*, 1966, **14**, 99-117.
[28] H. Fröhlich, *Trans. Faraday Soc.*, 1948, **44**, 238-243.
[29] R. Díaz-Calleja, *Macromolecules*, 2000, **33**, 8924.
[30] R. Böhmer, K. L. Ngai, C. A. Angell, D. J. Plazek, *J. Chem. Phys.*, 1993, **99**, 4201-4209.
[31] F. Alvarez, A. Alegra, J. Colmenero, *Phys. Rev. B*, 1991, **44**, 7306-7312.
[32] P. Hapiot, C. Lagrost, *Chem. Rev.*, 2008, **108**, 2238-2264.
[33] M. J. Roldán-Ruiz, R. J. Jiménez-Riobóo, M. C. Gutiérrez, M. L. Ferrer, F. del Monte, *J. Mol. Liq.*, 2019, **284**, 175-181.
[34] A. Jani, T. Sohier, D. Morineau, *J. Mol. Liq.*, 2020, **304**, article no. 112701.
[35] B. Malfait, A. Jani, D. Morineau, *J. Mol. Liq.*, 2022, **349**, article no. 118488.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 15-27
https://doi.org/10.5802/crchim.326

Research article

French Network on Solvation (GDR 2035 SolvATE)

Ion-specific effects in polyelectrolyte solutions: chain–chain interactions, chain rigidity and dynamics

Claire Hotton [a], Yasine Sakhawoth [a], Anne-Laure Rollet [a], Juliette Sirieix-Plénet [a], Lingsam Tea [a], Sophie Combet [b], Melissa Sharp [c], Ingo Hoffmann [d], Frédéric Nallet [e] and Natalie Malikova [*, a]

[a] Laboratory of Physical Chemistry of Electrolytes and Interfacial Nanosystems (PHENIX), Sorbonne Université, CNRS, 75005 Paris, France

[b] Laboratoire Léon-Brillouin (LLB), UMR12 CEA-CNRS, Université Paris-Saclay, F-91191 Gif-sur-Yvette CEDEX, France

[c] European Spallation Source (ESS) AB, Box 176, S-22100 Lund, Sweden

[d] Institut Laue Langevin (ILL), Grenoble F-38042, France

[e] Centre de Recherche Paul-Pascal, UMR Université de Bordeaux–CNRS 5031, 115 Avenue du Dr Albert Schweitzer, 33600 Pessac, France

E-mails: claire.hotton@universite-paris-saclay.fr (C. Hotton), yasine.sakhawoth@gmail.com (Y. Sakhawoth), anne-laure.rollet@sorbonne-universite.fr (A.-L. Rollet), juliette.sirieix-plenet@sorbonne-universite.fr (J. Sirieix-Plénet), lingsam.tea@sorbonne-universite.fr (L. Tea), sophie.combet@cea.fr (S. Combet), melissa.sharp@esss.se (M. Sharp), hoffmann@ill.fr (I. Hoffmann), frederic.nallet@u-bordeaux.fr (F. Nallet), natalie.malikova@sorbonne-universite.fr (N. Malikova)

Abstract. In this article, ion-specific effects in aqueous solutions of polyelectrolytes (PEs) are addressed. We focus on ionene cationic chains, featuring simple structure, absence of side groups, and very regular chain charge density. Ion-specific effects in ionene solutions are demonstrated using a series of monovalent (halide) counterions. The study combines both static and dynamic measurements by small angle neutron scattering, neutron spin echo (NSE), and pulsed field gradient NMR (PFG-NMR). Ion-specific effects are a phenomenon at high PE concentration, and the nature of the counterion is seen to influence drastically ionene chain–chain interactions via what we refer to as "ion-specific screening". The origin lies in the closer approach of large, highly polarisable counterions to the chain backbone, leading to more constricted counterion clouds. Equally affected is the local chain rigidity as well as collective and self-diffusion coefficients at larger scales. Small, nonpolarisable, strongly hydrating counterions, here F^- ions, lead to locally rigid chains. For such chains, the nm-scale collective dynamics as seen by NSE is the fastest while self-diffusion seen at the μm scale by PFG-NMR is the slowest. In other words, the loss of charge on the chain due to ion-specific screening has the opposite effect on collective diffusion and self-diffusion of the chains.

* Corresponding author

Keywords. Polyelectrolytes, Ion-specific effects, Inter-chain interactions, Self-diffusion, Collective diffusion, PFG-NMR, Neutron spin echo.

Manuscript received 31 January 2024, revised 7 June 2024, accepted 5 July 2024.

1. Introduction

Charged polymer chains, also referred to as "polyelectrolytes" (PEs), are omnipresent among natural compounds (e.g., polysaccharides, nucleic acids) or indeed among the many synthetic substances used in food, cosmetic, and packaging industries as well as in water treatment and several other fields [1]. The behaviour of charged polymer chains in solution, and we shall restrict ourselves here to aqueous solutions, is a world of its own. Two possible reference situations can be thought of as systems of departure: (a) electrolyte solutions and (b) uncharged polymer chains in solution. In the first case, we depart from a solution of atomic or molecular ions, both positive and negative, and we add connectivity between one type of these ions, be it positive or negative. This creates significant charge density inhomogeneities in the solution. This is by now not the common way of thinking about PE solutions, but it was indeed the point of departure for Fuoss et al. back in the 1950s [2]. In the second case, we start with the already quite complex case of a macromolecular chain in solution, where the quality of the solvent (good/theta/bad) decides the chain conformation. Next we add charge to a fraction of the monomers on the chain, but importantly we also need to introduce a population of counterions into the surrounding solution (and partially condensed onto the chain if the necessary conditions are met). Either way, the transition (adding connectivity or adding charge) is far from trivial and the consequences in terms of the structure, interactions, and dynamics of the (macro)molecular species present in such a solution are considerable. This is naturally reflected in the very different properties of the solution at the macroscopic scale, its phase diagram, rheology, osmotic pressure, heat of dilution, and other thermodynamic properties [1].

Historically, it is indeed through the macroscopic properties that the peculiar behaviour of PE solutions was uncovered, and the rheological studies of Fuoss et al. (the Fuoss law) remain a reference in this respect (e.g., [3]). Tremendous progress has been made since the Fuoss studies and we recommend two reviews, dating from very different times, to trace this progress [4,5]. A breakthrough came in the 1970s with the advent of scattering techniques, in particular neutron scattering, the seminal experimental work of Cotton, Jannink et al. [6,7], and the accompanying theoretical developments by de Gennes, Pfeuty et al. (the scaling approach) [8–10]. From this point onwards, the molecular level description of solutions of neutral polymers and later on PE solutions indeed began to emerge.

Overall, the scaling-based theory of PE solutions is highly successful, especially when dealing with the interpretation of neutron and X-ray scattering experiments. There, it predicts correctly the scaling of the PE chain correlations probed via the universally observed maximum in the scattering data, the so-called *polyelectrolyte peak*. This feature of the scattering curves is completely absent for solutions of neutral polymers and reflects the electrostatic repulsion present in solutions of *charged* chains. De Gennes' theory of PE solutions was later broadened by Dobrynin and Rubinstein to include all cases of solvent quality and concentration regimes [11]. The initial picture of dilute and semidilute concentration regimes was enriched (entangled and nonentangled subregimes) and extended, especially in the high concentration range [12–14]. The competition between electrostatic repulsion along a PE chain and hydrophobic (solvophobic) collapse has led to the prediction of the so-called *pearl-necklace* conformation of PE chains for the case of bad solvent [15], and these have indeed been observed experimentally [16,17] and by simulations [18].

Both the conformation of individual chains (the chain form factor) and the chain–chain correlations (the structure factor) modify the scattering curves. These individual contributions often cannot be distinguished easily, contrary to solutions of inorganic colloids for example. Polymer/PE solutions pose two problems in this respect: (1) the conformation of polymer/PE chains changes as a function of concentration and (2) polymer chains are penetrable objects, that is, individual chains can get entangled. A very elegant method, the zero average contrast (ZAC) method, is accessible in neutron scattering to highlight each of the contributions for cases where

deuteration of the polymer/PE chains is possible [19]. In the absence of such measurements, the main feature to account for in the scattering spectra remains the position of the scattering peak. As follows from the above paragraphs, the position of the PE peak depends not only on the PE concentration but also on the chain conformation (consequence of the effective chain charge and the solvent quality).

In our past publications, we have shown that for identical PE concentration, solvent quality, and valence of the counterions, the *chemical nature* of the counterion by itself can also influence the shape of the scattering curve, the position, and shape of PE peak and the extent to which this peak is indeed visible or not [20,21]. The origin lies in the hydration of the given ion, which leads to a different degree of screening of the chain charge. When the chemical nature of an ion is to "blame" for a given observation, biochemists and physical chemists refer to such instances as *ion-specific effects* [22–27]. It has been observed that ion-specific effects manifest themselves more strongly for anions than cations [28,29]. Solutions and gels based on cationic chains with compensating *anions*, such as ionenes [20,21,30], show indeed stronger ion-specific effects than anionic PEs, such as the widely studied polystyrene sulfonate [16, 31–34]. Needless to say, any purely electrostatic theory, such as the scaling approach of de Gennes, the Manning theory of counterion condensation [35,36], or indeed the Poisson–Boltzmann approach [32] cannot account for these effects as hydration properties of solvated ions do not come into consideration. Attempts in what seems the correct direction are theories accounting for local dielectric heterogeneities around the ions and the PE chains [37]. The orientation of the dipole moment of water molecules is indeed closely linked to the hydration and the polarisability of the hydrated species.

Ionenes, the focus of this study, are a group of water soluble cationic PEs with pH independent charge, based on quaternary ammonium charged centers linked by simple hydrocarbon chains. Ionenes have already several applications including ion exchange resins [38], water treatment in the oil industry [39], humidity sensors [40], organic templates in the synthesis of mesoporous silica [41], and anti-microbial agents [42]. Within the realm of PEs, ionenes present the advantage of a regular and tunable separation of charges on the backbone, as opposed to statistically distributed charges for other PEs. In our initial scattering studies on ionene aqueous solutions, we explored the transition from hydrophilic to hydrophobic polyelectolyte behaviour as the ionene charge density decreases. This transition was indeed found, however later than expected: the hydrophobicity of the hydrocarbon backbone of ionenes becomes "visible" when only 15% of the monomers are charged, not before [21]. Dramatic *ion-specific effects* in ionene aqueous solutions have been initially observed in thermodynamic properties [43–45] and later on the microscale by scattering for the particular case of two halide ions, F^- and Br^- [20,21]. In this contribution, we show how this generalises for an entire series of halide counterions and what consequences it has for the dynamics of the PE chains. We bring information on the chain dynamics at the microscopic (nm) scale, by the neutron spin echo (NSE) technique, and also on the mesoscopic (μm) scale, by pulsed field gradient NMR (PFG-NMR).

2. Experimental methods

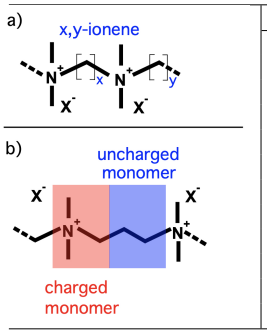
2.1. *Ionenes: synthesis and structural overview*

Ionenes and their precursors were synthesised using a procedure adapted from those described previously [20,21,45]. The details of the synthesis are provided in the SI file (part 1). The synthetic route leads invariably to ionenes with bromide counterions. The molecular weights of ionenes were determined by size exclusion chromatography (SEC) as described in [46]. The range of molecular weights is 20,000–60,000 g/mol, which corresponds to 100–300 nm in terms of chain length. SEC measurements on cationic PEs are very difficult [47] and, for us, were successful only for 6,9-ionenes. In the following, we consider that the above range of molecular weights applies also to ionenes of other charge densities, which were synthesised under identical conditions. We have indeed confirmation that the molecular weights of ionenes with different charge densities are of the same order of magnitude from the NMR signal of amine end groups, which allows estimation of the degree of ionene polymerisation [48].

Counterion exchange was performed by dialysis starting from Br-ionenes. Dialysis tubes (Sigma-Aldrich, MWCO = 12,000 $g{\cdot}mol^{-1}$) were filled with 0.02 M solutions of Br-ionenes and first dialysed

against 0.05 M solution of the desired NaX (3 weeks) to exchange anions and then dialysed against water (2 weeks) to remove sodium ions. All ionene solutions for neutron scattering and NMR measurements were prepared gravimetrically. Deuterated water (Eurisotop, 99.9%D) was used for neutron scattering samples as well as NMR samples. The pH of the solutions was close to neutral, and thus we estimate the effects of any dissolved carbonic acid as very small.

The general chemical formula of ionenes is $[-(CH_3)_2N^+-(CH_2)_x-(CH_3)_2N^+-(CH_2)_y-]_n$ for an x,y-ionene chain with Br^- or other counterions (Figure 1). Ionenes with X^- counterions are referred to as X-ionenes in the rest of the manuscript. Values x and y represent the number of $-CH_2-$ (methylene) units between adjacent charged centers (quaternary ammonium centers) and can be varied accurately by synthesis [49–51]. By increasing x and y, the ionene chain is less charged. In this manuscript, we discuss ionene chains for which x, y = 3,3 and 6,9 (referred to as 3,3-ionenes and 6,9-ionenes). The simple structure (absence of bulky side groups) and finely tunable and *regular* charge density are very interesting structural features of ionenes. Other PEs, including styrene, often present charge on (bulky) side groups and charged monomers are distributed statistically along the chains. In order to draw a parallel between ionenes on one side and polystyrene-based and other PEs on the other side, we may consider the structure of ionenes as a sequence of charged and uncharged "monomers" as depicted in Figure 1.



x,y-ionene	a (Å)	f_{chem}	ξ
3,3–	5.00	0.50	1.43
4,5–	6.88	0.36	1.04
6,6–	8.75	0.29	0.82
6,9–	10.63	0.24	0.67
6,12–	12.50	0.20	0.57
12,12–	16.25	0.15	0.43

Figure 1. *Left*: (a) Schematic view of an x,y-ionene chain. (b) Schematic view of a 3,3-ionene ($x = 3$, $y = 3$) chain, showing the definition of a charged and an uncharged monomer. *Right*: Ionene structural parameters: a is the charge separation on the chain, f_{chem} the fraction of charged monomers, and ξ the Manning charge density parameter, defined as $\xi = L_B/a$, where L_B is the Bjerrum length (7.14 Å in water at room temperature). While 4,5-ionenes are at the $\xi = 1$ limit (onset of Manning-type condensation), only 3,3-ionenes have sufficient charge density to induce significant condensation ($\xi > 1$) and decrease the chemical charge (f_{chem}) to an effective charge (f_{eff}).

2.2. *Neutron scattering*

Small angle neutron scattering (SANS) measurements were carried out on the PACE spectrometer at LLB-Orphée, Saclay, France. Using up to three different combinations of incident neutron wavelength (λ) and sample to detector distance, a wavevector (q) range of 0.01–0.45 $Å^{-1}$ was covered ($q = 4\pi\sin(\theta/2)/\lambda$). The detector efficiency was taken into account by normalisation of data with a flat (incoherent) signal from bulk light water. Ionene solutions (hydrogenated chains in D_2O solvent) were loaded into quartz cells with a path length of 1 or 2 mm. Due to the isotropic nature of our samples, data were grouped in concentric rings, each corresponding to a given q value. The measured scattered intensities were corrected for transmission, sample thickness, and incoherent and solvent background to yield the coherent scattered intensity I_{coh}. We checked the reproducibility of neutron scattering spectra by measuring samples from different synthesis batches.

Neutron Spin Echo (NSE) [52,53] experiments were carried out on the IN15 spectrometer in ILL, Grenoble, France. Samples (hydrogenated chains in D_2O solvent) were measured in 1 mm or 2 mm flat quartz cells. Using a combination of two neutron wavelengths, 6 Å and 9 Å, and detector angles of 3°, 5°, 7°, 10°, 13°, 19° at 6 Å and 3°, 6° at 9 Å, we achieved an accessible q range of 0.04 $Å^{-1}$ to 0.5 $Å^{-1}$ and time range of 0.07–11 ns at 6 Å and 0.25–36 ns at 9 Å. Both static (I_{coh}) and dynamic data ($I(q,t)$) in NSE were corrected for contributions from the quartz cell and the solvent (D_2O) background.

2.3. *Pulsed field gradient NMR*

The single pulse ^1H NMR spectra were recorded using a Bruker Avance III 300 MHz NB spectrometer

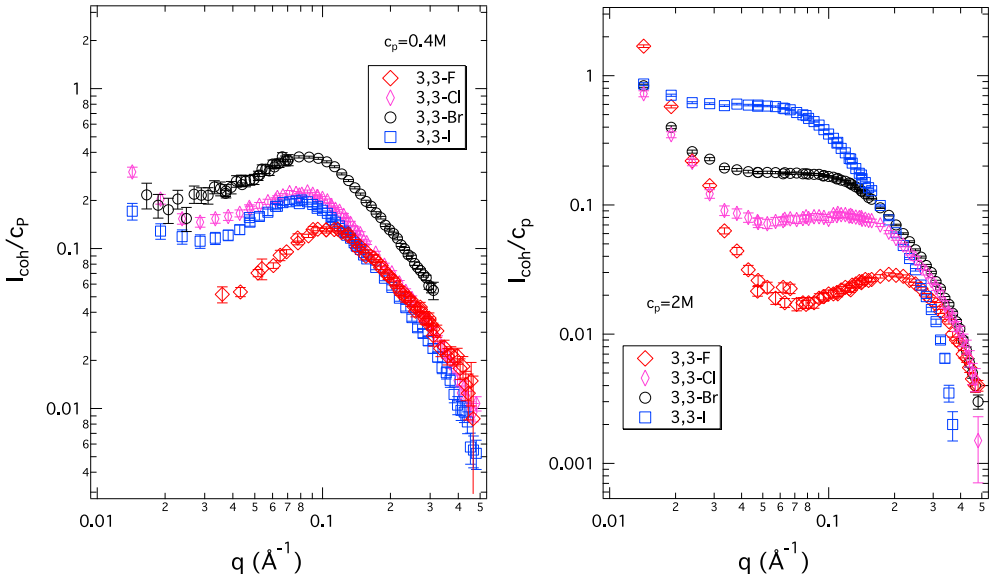


Figure 2. Coherent neutron scattering intensity normalised by ionene monomer concentration (I_{coh}/c_p) in arbitrary units versus scattering wavevector (q) for room temperature aqueous solutions (in D_2O) of 3,3 X-ionenes at two monomer concentrations, 0.4 M (left) and 2 M (right). These monomer concentrations correspond to volume fractions around 2% and 10%, respectively.

operating at 7.05 T. The lock was obtained with a sealed 2 mm capillary filled with D_2O inserted inside the NMR tube. The chemical shift was referenced to $CHCl_3$. The PFG_NMR experiments were performed using a BBFO probe equipped with a 55 G·cm^{-1} gradient coil. We used an NMR pulse sequence combining bipolar gradient pulses and stimulated echo. This sequence was repeated with 16 gradients of increasing strength from 2 to 50 G·cm^{-1} for a duration of 1.5 ms. The diffusion time was approximately 200 ms, which corresponds to a diffusion over a length scale of 1.5 µm. The self-diffusion coefficients are obtained by nonlinear least-square fitting of the echo attenuation, using the Bruker TopSpin software. All PFG-NMR data were measured at room temperature on ionene solutions in D_2O.

3. Results and discussion

SANS data of 3,3 X-ionenes with four different halide counterions ($X^- = F^-$, Cl^-, Br^-, I^-), at a moderate and a high monomer concentration, are shown in Figure 2. From our previous scattering studies on ionenes, we know that both of these concentrations are in the semidilute regime. The overlap concentration for ionenes synthesised using our protocols was estimated to be below $c_p = 0.07$ M [20]. In addition, viscosity data (see SI part 2) confirms that both F-ionenes and Br-ionenes are in the same concentration regime, that is, semidiluted. In the semidiluted regime, the position of the peak reflects the mesh size formed by the interpenetrating chains. At 0.4 M monomer concentration, the spectra of all systems show a well-defined PE peak (a clear maximum) at an almost identical position in the wavevector q (a slight shift towards higher q values is noticed for the F-ionene; see later). At high monomer concentration (2 M), the four systems feature very different scattering curves. Note that the increase in intensity in the small q region ($q < 0.03$ Å^{-1}) is due to large-scale heterogeneities in the system—a repeatedly observed feature for PE solutions/gels, which remains poorly understood. The changes that interest us most are thus confined to the q region roughly between 0.03 Å^{-1} and 0.4 Å^{-1}. In this central region, a clear PE peak remains visible only for the F-ionene. Its position is shifted to higher q values in comparison to data at 0.4 M, as expected, due to a denser mesh size at this higher concentration. In the sequence $F^- \rightarrow Cl^- \rightarrow Br^- \rightarrow I^-$, the central part of the spectrum gains in intensity and the intensity dip to the left of the PE peak seen for F-ionene gradually

disappears. As a consequence, the peak becomes highly asymmetric (in the case of Cl-ionene, still a slight maximum is observed). The peak disappears completely for the Br-ionenes and I-ionenes and instead a plateau is seen in the spectra, resembling a signal we would expect from a neutral polymer.

Figure 3 summarises the position of the PE peak for the four 3,3 X-ionenes seen in their SANS spectra. Up to 0.1 M concentration, the position of the peak is very close for all systems; for $c_p > 0.1$ M, the situation changes. Only the F-ionene follows the predicted $c_p^{1/2}$ law and the three other systems depart significantly from this description. Combining information from Figures 2 and 3, we observe that at 0.4 M, the PE peak position for 3,3 F-ionene is already somewhat higher than for all other systems; all curves present a well-defined PE peak. For higher concentrations, the *strong* departure from the $c_p^{1/2}$ law in Figure 3 for Cl-, Br-, and I-ionenes is mainly a consequence of the PE peak *disappearing* from the scattering signal. For a poorly defined asymmetric peak, the determination of its position is increasingly difficult (see the right side of Figure 2). Overall, clearly the ion-specific effect is a high concentration phenomenon and we can place the critical concentration at around 0.1 M. This corresponds to a charge concentration in the system of 0.05 M according to $c(X^-) = c(N^+) = c_p f_{\text{chem}}$.

In order to further investigate the ion-specific effect and its consequence on the scattering spectra, we carried out measurements for a series of ionenes with mixed counterion clouds at a constant monomer concentration. This was done for ionenes of different charge densities (3,3, 6,9, and 12,12). (Note that based on our previous scattering results, we know that only the 12,12-ionenes begin to show the signature of backbone hydrophobicity [21]. All ionene chains with higher charge densities behave as hydrophilic.) The most striking changes in the spectra are observed for the most highly charged chains (3,3-ionenes), and this system is shown in Figure 4. Additional data for 6,9-Br/F and 12,12-Br/F systems are included in the SI file (part 3). The two extreme systems ($xF = 1$ and $xF = 0$) are naturally identical to the F- and Br-ionene data in Figure 2. All intermediate systems place themselves logically between the two extremes, with gradual changes with increasing/decreasing xF. The complete disappearance of the PE peak (absence of a maximum) is only present for the pure Br system. As expected, the intermediate

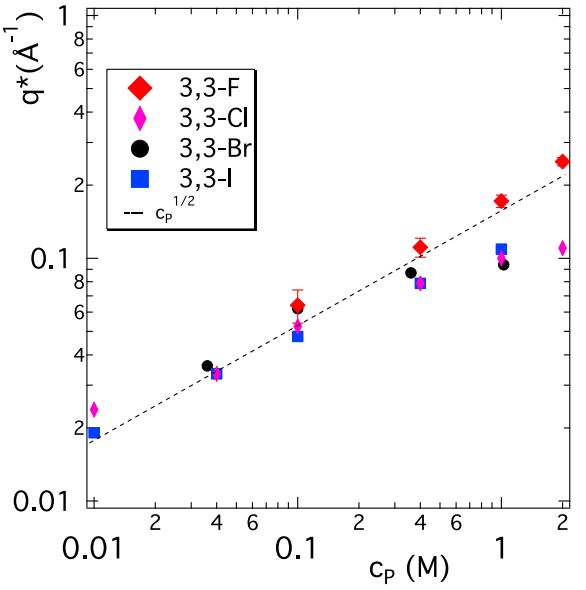


Figure 3. Position of the polyelectrolyte peak in SANS spectra (q^*) versus ionene monomer concentration (c_p) for 3,3-ionenes with different counterions. Typical error bars are represented on the 3,3-F data set. Dashed line is a guide to the eye representing a $c_p^{1/2}$ scaling law, expected in the semidilute concentration regime.

curves *cannot* be obtained by a linear combination of the curves corresponding to the $xF = 1$ and $xF = 0$ extremes [54]. For a given xF, each chain is surrounded by a mixed counterion cloud at the given ratio; there are no chains in a "pure F" or "pure Br" environment. Interestingly, as we decrease the charge density of the ionene chains (6,9- and 12,12-ionenes), the difference in scattered intensity between the pure F and pure Br extremes is diminished (see SI). This is probably due to a decreasing overall counterion concentration in the system as we move from 3,3- to 12,12-ionenes.

Having scattering data for ionene solutions across a whole series of halide counterions gives a very strong argument for the previously suggested origin in terms of a decreased effective charge of the ionene chains as we move towards larger, more polarisable ions with a lower hydration energy, that is, as we descend the halogen series in the periodic table. For completeness, radii of halide ions in solution, polarisability, and hydration energies are summarised for

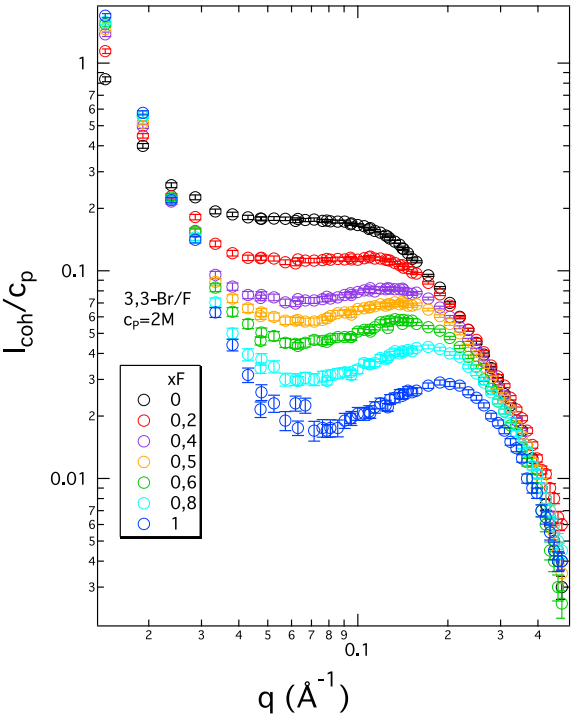


Figure 4. Coherent neutron scattering intensity normalised by ionene monomer concentration (I_{coh}/c_p) in arbitrary units versus scattering wavevector (q) for room temperature aqueous solutions (in D_2O) of 3,3-ionenes with mixed Br–F counterion clouds. The fraction of F^- counterions (xF) is shown in the legend. All systems are at 2 M monomer concentration.

the four halide ions used in Table 1. For the larger ions, the counterion atmosphere around the ionene backbone is more constricted as has also been clearly shown by previous molecular dynamics simulations on ionene solutions [55]. As soon as counterion clouds of adjacent chains do not overlap, the repulsion between the chains is no longer present and the PE peak in the scattering spectra disappears. This is a concentration-dependent phenomenon, accentuated at high PE (and thus counterion) concentration, which we can refer to as "ion-specific screening". It seems important to distinguish this from counterion condensation in the Manning sense of the word. As we have seen by osmotic pressure measurements in ionene solutions in the past [20] and as was equally observed for other systems [56], the counterions indeed still contribute to the osmotic pressure. This

Table 1. Ionic radii in solution (R_s), polarisabilities (α), and hydration free energies (ΔG_{hyd}) for halide ions

Ion	R_s [58] (Å)	α [59] (Å^3)	ΔG_{hyd} [59] (kcal/mol)	ΔG_{hyd} (k_BT/ion)
F^-	1.24	1.20	−112.1	−189.3
Cl^-	1.80	3.65	−82.4	−139.1
Br^-	1.98	4.96	−76.1	−128.5
I^-	2.25	7.30	−67.0	−113.1

is contrary to what has been seen for counterions in solutions of hydrophobic polystyrene-based PEs, where counterions are condensed as part of the "pearls" in the pearl-necklace conformation [57].

For completeness, we note that the interpretation of the SANS on ionene solutions, here and in our previous publications [20,21], is based on the assumption that the scattered signal is dominated by the ionene monomer–monomer correlations and that contributions of the counterions can be neglected. The underlying estimation of the relative intensities is provided in the SI (part 4). It shows that contributions of halide ions to the scattered intensity increase as we move from F^- to I^-. Importantly, for the purposes of the *qualitative* trends that we discuss here, and which are common to ionenes of all charge densities, the above assumption is indeed reasonable.

In the following, we are interested in exploring the rigidity of the ionene chains as a function of the above ion-specific screening. We have explored this for the case of F^- and Br^- counterions using the NSE technique [52,53]. Neutron spin echo gives access to the microscopic dynamics of the chain on the length scale of nm and a timescale of ps-ns. These are sufficiently short length scales to avoid the influence of large heterogeneities, which lead to the observation of a slow mode, a very common feature in dynamic light scattering studies on PE solutions [60,61]. The measured data in NSE is the intermediate scattering function $I(q,t)$, which is formally the spatial Fourier transform of the van Hove correlation function $g(r,t)$. We measure the dynamic data on the coherently scattered signal arising from the contrast between hydrogenated ionene chains in a deuterated solvent (D_2O), the same systems as those used previously for the small angle

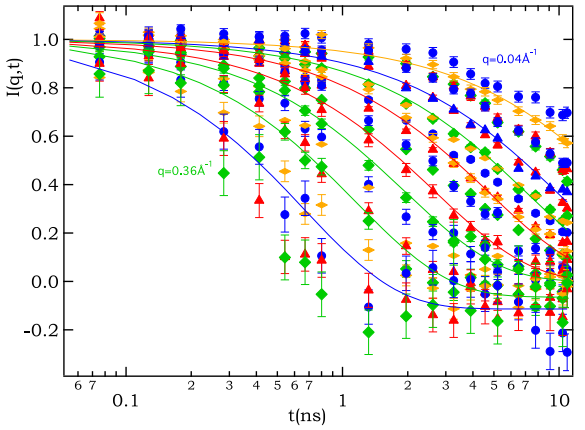


Figure 5. Intermediate scattering function $I(q,t)$ for 6,9-Br ionene at c_p = 0.4 M as measured by NSE. Each curve corresponds to a given q value in the range from 0.04 to 0.36 Å^{-1}. Lines are mono-exponential fits for a selection of the curves.

scattering experiments. The NSE data were collected for ionene chains with intermediate chain charge densities, 6,9-ionenes, at 0.4 M and 2 M monomer concentrations.

An example of NSE data is shown in Figure 5, where a series of $I(q,t)$ curves for 6,9-Br ionene at 0.4 M is presented, each curve corresponding to a given q value. Following a standard analysis, the data was modeled using mono-exponential decay to obtain a value of a characteristic relaxation time τ at a given q value. Under a simple diffusion model, this characteristic time is converted into an effective diffusion coefficient D_{eff} following the relation $1/\tau = D_{\text{eff}}q^2$. Note that for large q values, the background corrections (quartz cell and solvent) lead to an unphysical long-time asymptote of slightly less than 0. For these cases, the fitting parameters were relaxed to allow for a nonzero (slightly negative) constant background.

Figure 6 summarises the coherently scattered intensity (as measured by polarisation analysis on NSE) and the effective diffusion coefficients D_{eff} resulting from the mono-exponential fitting of the $I(q,t)$ curves. Data for all four systems studied by NSE are shown: 6,9-Br and 6,9-F ionenes, each at 0.4 M and 2 M monomer concentration. The position of the PE peak in the coherently scattered intensity observed by NSE (Figure 6, left) reflects what has been observed already by SANS (Figures 2 and 4). The data for Br- and F-ionenes at low monomer concentration show a peak at the same q position while at high concentration, the scattered intensity is radically different. For completeness, the relative contributions of coherent and incoherent scattering, as determined by polarisation analysis in NSE, are presented in SI (part 5).

In Figure 6 (left), we also indicate the high q intensity dependence. Assuming that the chain–chain correlations do not contribute in a significant way to the scattered intensity in the high q region (this is an approximation, as we are not working under the "zero average contrast" conditions here), the power law reflects the conformation of the individual chains. While a q^{-1} behaviour is characteristic of a rod-like conformation (and this is the case of the low concentration data for both Br- and F-ionenes), a q^{-2} behaviour corresponds to the signal of a Gaussian chain, in other words a neutral polymer in a Θ solvent ($q^{-1.7}$ indicates Gaussian chains with an excluded volume contribution). It is clear that compared to 0.4 M data, both 2 M data sets have a higher decay exponent ($q^{-1.42}$) at high q values, which indicates a less rod-like conformation. At the same time, we need to note that the high q dependence in the SANS spectra (Figures 2 and 4) does not agree with the NSE-determined decay exponents. The SANS data features much higher exponents (close to q^{-2} already for the 0.4 M data sets and even higher for the 2 M data sets). It is important to realise that the decay exponent at high q is very sensitive to the incoherent background subtraction. This is done in very different ways in SANS and in NSE. In SANS data reduction, the density of H atoms is first estimated from the concentrations of the hydrogenated chains in the deuterated solvent and compared to H atom density of pure water. The background constant to subtract is determined from the ratio of these two H atom densities. In NSE, the decomposition of the total scattered signal into coherent and incoherent contributions is measured *directly* using polarisation analysis, which relies on the spin flip of incoherently scattered intensity from H nuclei [62]. As a result, we consider the high q exponents determined from NSE as more reliable and we do not conclude on the chain conformation from the high q SANS signal.

There are several points to note regarding the effective diffusion coefficients in Figure 6 (right), all

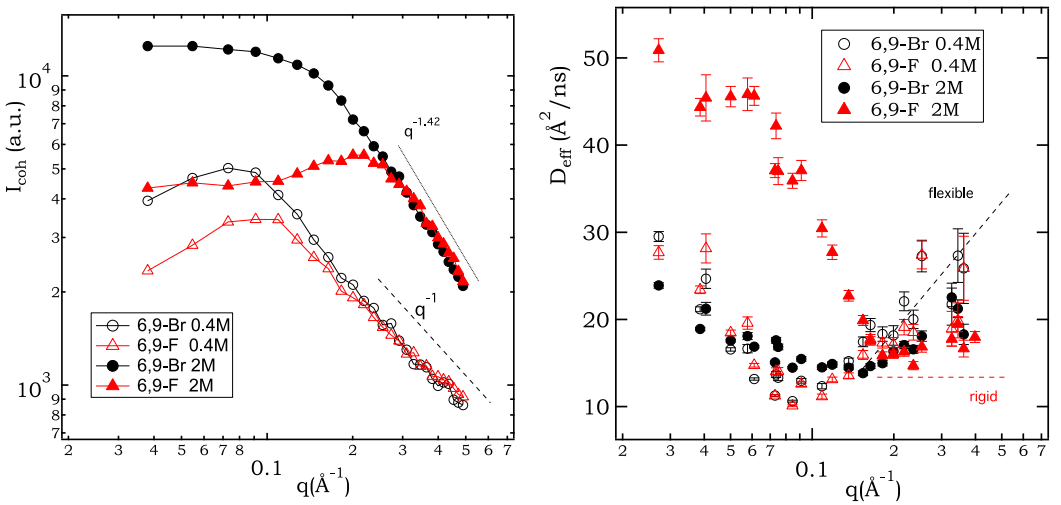


Figure 6. Coherently scattered intensity (polarisation analysis in NSE, left) and effective diffusion coefficient D_{eff} (right) as a function of the wavevector q for all systems studied by NSE. Note that 10 Å^2/ns = 10^{-10} m^2/s.

of the order of 10^{-10} m^2/s. We may consider the figure in two parts, below and above the q^* position or the high q end of the observed plateau, depending on the given system. Both theory [63] and previous experiments [64–66] on PE solutions show a rapid decrease in D_{eff} for $q < q^*$ and then a constant value for $q > q^*$. The constant value at high q reflects the rod-like conformation of a charged chain. On the contrary, the signal in a semidilute solution of a neutral polymer chain should show D_{eff} increasing linearly with q in the high q region [67]. The behaviour of D_{eff} in the case of charged rigid chains stems from the theoretical considerations of de Gennes and collaborators [63]. First, D_{eff} is expressed as $D_{\text{eff}} = k_{\text{B}}T\mu(q)/S(q)$, where $\mu(q)$ is the mobility and $S(q)$ is the static scattering signal [68]. This is indeed a mathematical expression of the so-called "de Gennes narrowing", which in a simplified way states that a slowing down of dynamics takes place in q regions where peaks in structural correlations are present. Clever modelling by de Gennes indicates that for locally rigid chains, $\mu = \mu(0)/(ql_{\text{p}})$, where l_{p} is the persistence length, that is, $\mu \propto q^{-1}$ for $ql_{\text{p}} \geq 1$. Taking into account that for locally rigid chains $S(q) \propto q^{-1}$ in that same high q range, the q dependence of D_{eff} cancels out and a constant D_{eff} in the high q region is recovered.

From the high q behaviour, the most rigid chain is indeed that of F-ionenes at 2 M concentration, as this data set features the most constant values beyond q^*. The remaining three systems (Br 2 M, Br 0.4 M, and F 0.4 M) all feature a more significant increase in D_{eff} beyond q^*. Although both 0.4 M data sets exhibited clear rod-like behaviour in the coherently scattered intensity (Figure 6, left), these chains do not show the highest rigidity as seen from D_{eff}. One difficulty in the interpretation is probably the intra- and inter-chain correlations mixing in both the static and dynamic signals. Measurements under ZAC would probably help disentangle the inter- and intra-chain correlations contributing to the scattered signal, at least for the static signal. NSE measurements under ZAC seem challenging.

Overall, we view the trends in Figure 6 (right) as consistent with the theoretical predictions outlined previously, which indicate above all a very different rigidity for the ionene chains with Br^- and F^- ions at 2 M concentration. The striking difference between Br and F is related to the very different scattered intensities: as q decreases below 0.2 Å^{-1} (a) the 2 M F system passes through a structural maximum and its dynamics increases very fast for all smaller q values; (b) the scattered intensity for the 2 M Br system continues to grow below 0.2 Å^{-1} to reach a plateau, and the dynamics in 2 M Br is significantly suppressed in comparison to 2 M F in the entire region below 0.2 Å^{-1}, but begins to rise slowly once the plateau is reached.

The mesoscopic dynamics of ionene chains in D_2O solutions was measured by PFG-NMR. Figure 7 summarises the D_{NMR} data for 3,3-ionenes with the four different halide counterions as a function of monomer concentration c_p. For comparison, this figure also features data from sodium polystyrene sulfonate (PSS), measured by PFG-NMR in H_2O, from reference [69]. The ionene and PSS data fall into the same range of D_{NMR} of the order of 10^{-11} m^2/s. At room temperature, the ratio of D_2O and H_2O viscosities is 1.25, and this conversion factor would have to be used for a detailed quantitative comparison of the two data sets. The ionene chains correspond to molecular weights of 20–60 kDa, which fall well within the range of molecular weights for the PSS data [69]. However, the size polydispersity (PID) of ionene chains (due to the poly-addition reaction as opposed to radical polymerisation for PSS) is significantly higher: PID(ionene) = 1.8–2.0 and PID(PSS) = 1.25–1.5. Given these differences, we do not dwell on a detailed quantitative PSS–ionene comparison. The general trend as a function of c_p is similar for the two types of PE chains, with a decrease in D_{NMR} for c_p above approximately 0.1–0.2 M.

Let us concentrate on the D_{NMR} data in Figure 7 for different counterions, especially in the higher c_p region (above 0.1 M). The ordering of D_{NMR} data for 3,3-ionenes does not follow in a simple way the halide anion series, contrary to what was seen in the SANS data. The order here is $D(\mathrm{F}) < D(\mathrm{Cl}) > D(\mathrm{Br}) \gg D(\mathrm{I})$. We see two phenomena behind this non-monotonous behaviour: (a) change in chain conformation (less rod-like as we move down the halide series towards larger anions) and (b) inter-chain aggregation (the system becomes less soluble as we move down the halide series). A change from rod-like to globular chain conformation leads to an *increase* in self-diffusion coefficient of the chain [70]; the inter-chain aggregation leads to its *decrease* (diffusion of larger objects). For the 3,3 I-ionenes, the inter-chain aggregation is indeed pronounced; 3,3 I-ionene shows a significantly lower D_{NMR} than all the other systems (roughly lower by a factor of 2) throughout the entire concentration range. This scenario is consistent with the evolution of viscosity (η) of aqueous solutions of a neighbouring cationic PE, poly(diallyldimethylammonium), along the halide series [71]. Indeed, η decreases between Cl^- and Br^- solutions of this PE (chains change from a rod-like

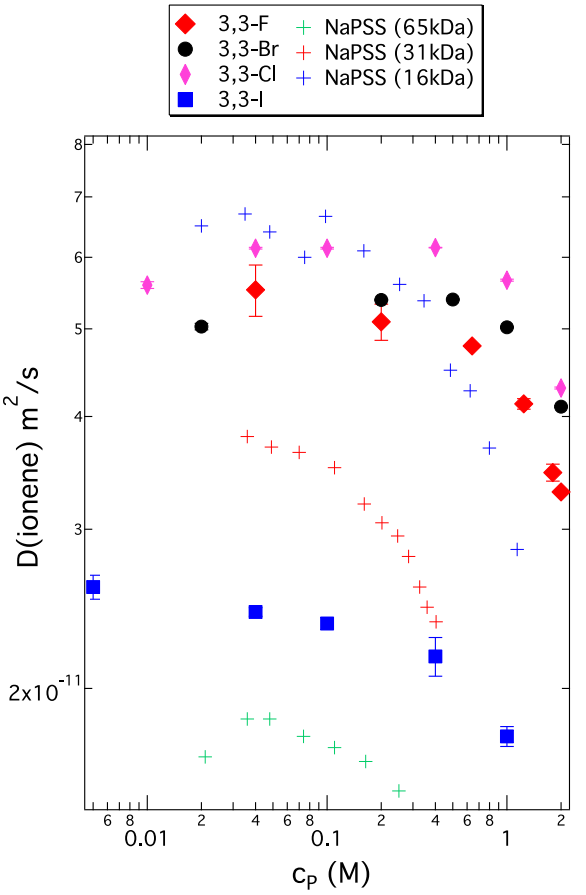


Figure 7. Self-diffusion coefficients of ionene chains for 3,3-ionenes with four different halide counterions as indicated versus monomer concentration c_p (solvent = D_2O) as measured by PFG-NMR. Crosses correspond to measurements for polystyrene sulfonate with Na^+ counterions (NaPSS) at three different molecular masses as indicated. NaPSS data from reference [69] (solvent = H_2O).

to a more globular conformation), and the system becomes insoluble with I^- (no F^- data are available in ref [71]). We attach importance to the fact that 3,3 F-ionenes show somewhat lower D_{NMR} than Br- and Cl-ionenes in the moderate to high c_p range. To ensure that this difference is real, we have explored changes in D_{NMR} for ionenes with mixed Br/F counterion clouds. The data is summarised in Figure 8 for two different ionene charge densities. Indeed, as we move from pure Br to pure F ionene chains, the chain dynamics clearly *decreases*.

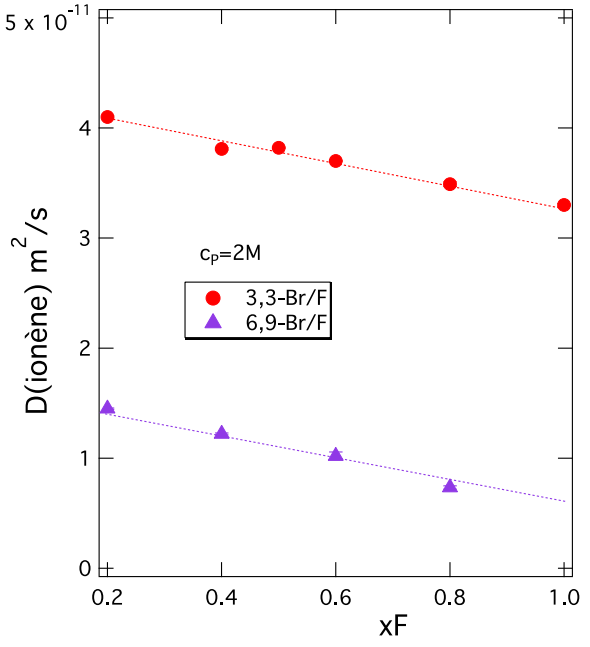


Figure 8. Self-diffusion coefficients of ionene chains for systems with mixed Br/F counterion atmospheres at a monomer concentration of 2 M as a function of F^- counterion fraction xF as measured by PFG-NMR. Two different ionene chain charge densities are presented.

4. Conclusion

The combination of SANS, NSE, and PFG-NMR provides us with several pieces of information on the state of the counterion atmosphere around positively charged PE chains in aqueous solution as a function of the counterion chemical nature, that is, *ion-specific effects*. A series of monovalent halide counterions is explored. A strong ion-specific effect is observed clearly at high PE monomer concentrations and is consistent with the picture of an increasingly more compact counterion atmosphere around the PE chain, a phenomenon we refer to as "ion-specific screening". A stronger counterion screening of the charge on the PE chains takes place as we move along the halide series towards larger, more polarisable and more weakly hydrated counterions. This can indeed be seen as another example of the Collins' concept of "matching water affinities" in which ions are characterised by their "softness" with consequences for favourable ion pairing between "soft" (weakly hydrated) anions and cations on one hand and "hard" (strongly hydrated) anions and cations on the other [26]. Within the Collins' classification, the quaternary ammonium groups on ionene PE chains are "soft" cations, and thus more favourable ion-paring is in place with "soft" halide ions, that is, larger and more polarisable anions. More effective screening of the PE chain charge is demonstrated by a reduced chain–chain repulsion in the system as shown by the disappearance of the PE structural peak in the scattering data. Importantly, the ion-specific effect has consequences for the chain rigidity and local and mesoscopic chain dynamics as shown further by NSE and PFG-NMR.

In principle, it is important to distinguish between the self-diffusion coefficient measured by PFG-NMR (due to the position encoding method using magnetic field gradients, PFG-NMR indeed measures self-diffusion) at the μm scale and the effective diffusion coefficient D_{eff} of the PE chains as measured by NSE (arising from the coherently scattered signal) at the nm scale. Depending on the length scale (q value), D_{eff} obtained by NSE indeed represents different quantities: (a) for length scales below the mesh size (q values above $q*$), it represents the local individual chain dynamics [63]; (b) for length scales above the mesh size (q values below $q*$), D_{eff} merges with $D_{\text{collective}}$ in the hydrodynamic limit $q \to 0$, (e.g., [72]). Let us now look at the different scales in turn, starting from the most local scale:

- Local scale below PE mesh size, probed by NSE: Here, the notion of *collective* motion seen via the coherent signal in NSE no longer applies; the scale probed is too small. NSE probes the dynamics of individual chains and informs us on their rigidity. We observe that PE chains retaining a strong chain–chain repulsion (i.e., with F^- counterions) show increased rigidity of the PE chains at high monomer concentration.
- nm scale larger than the PE mesh size, probed by NSE: Here, NSE dynamic data probe *collective* dynamics in the PE network, as the coherent signal is dominated by the *inter-chain* correlations at these length scales. For locally rigid chains (i.e., with F^- counterions), the collective dynamics at this scale is very fast. For more flexible chains

(i.e., with Br^- or Cl^- counterions), the dynamics here is suppressed. We see this as a case of "de Gennes narrowing", as the Br^- and Cl^- systems show an intense scattering signal at this spatial scale. The effective *collective* diffusion coefficients at the nm scale are all of the order of 10^{-10} m^2/s.

- μm scale, probed by PFG-NMR: On this largest scale, PFG-NMR probes the self-diffusion of the individual PE chains. Locally rigid charged chains, with fast collective dynamics seen at the nm scale (i.e., with F^- counterions), diffuse consistently slightly *slower* than locally more flexible chains with slow nm scale dynamics. The self-diffusion coefficients at the μm scale are all of the order of 10^{-11} m^2/s.

Dynamics in PE solutions has been probed extensively by DLS [60]. In the semidilute PE concentration regime, this techniques gives access to the *collective* dynamics (D_{coll}) at the mesoscopic scale; the closest length scale would be the PFG-NMR scale. Concentrating on the fast mode in the DLS signal, charged rigid PE chains show consistently *faster* collective dynamics compared to flexible neutral chains. This is again a demonstration of de Gennes narrowing, which leads to large values of D_{coll} for repulsive systems at length scales much larger than the PE mesh size (expressed in the scattering language, as $q < q*$, which is where DLS operates). For charged chains at high monomer concentration in the absence of salt, the order of magnitude for DLS-determined *collective* diffusion coefficients is 10^{-10} m^2/s [60]. This is the same order of magnitude as the *collective* diffusion coefficients measured here by NSE. Furthermore, the order of magnitude difference between self-diffusion and collective diffusion coefficients in PE solutions has been noted before ($D_{self} \ll D_{coll}$) [60,69]. Our new data sets (NSE, PFG-NMR) on ionene PE solutions are consistent with these observations. However, the effects of the counterion specificity on D_{self} measured here by PFG-NMR suggest that D_{self} of charged chains is *lower* than that of neutral chains. In other words, the loss of charge on the chain, due to counterion-specific screening of the chain charge, has the opposite effect on D_{self} and D_{coll}. This seems to be indeed in line with very recent DLS and PFG-NMR data on PSS [73]. Overall, the more strongly charged chains adopt a more extended conformation, resulting in a lower self-diffusion coefficient while the *collective* diffusion in these more repulsive systems is enhanced (de Gennes narrowing).

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organisation that could benefit from this article, and have declared no affiliations other than their research organisations.

Acknowledgements

The authors thank Sašo Čebašek at the University of Ljubljana for help with sample preparation, Matija Tomšič at the University of Ljubljana for help throughout NSE measurements at the ILL (Grenoble), and François Ribot at the College de France (Paris) for discussions and access to NMR spectrometers. The NSE data from ILL is available under doi:10.5291/ILL-DATA.9-11-1624.

Supplementary data

Supporting information for this article is available on the journal's website under https://doi.org/10.5802/crchim.326 or from the author.

References

[1] H. Dautzenberg, W. Jaeger, J. Kötz, B. Philipp, C. Seidel, D. Stscherbina, *Polyelectrolytes: Formation, Characterization and Application*, Hanser, Munich, 1994.
[2] D. Edelson, R. M. Fuoss, *J. Am. Chem. Soc.*, 1950, **72**, 306-310.
[3] R. M. Fuoss, *Discuss. Faraday Soc.*, 1951, **11**, 125-134.
[4] H. Eisenberg, *Biophys. Chem.*, 1977, **7**, 3-13.
[5] M. Muthukumar, *Macromolecules*, 2017, **50**, 9528-9560.
[6] J. P. Cotton, B. Farnoux, G. Jannink, *J. Chem. Phys.*, 1972, **57**, 290-294.
[7] M. Daoud, J. P. Cotton, B. Farnoux *et al.*, *Macromolecules*, 1975, **8**, 804-818.
[8] P. G. de Gennes, P. Pincus, R. M. Velasco, F. Brochard, *J. Phys.*, 1976, **37**, 1461-1473.
[9] P. Pfeuty, R. M. Velasco, P. G. de Gennes, *J. Phys. Lett.*, 1977, **38**, L5-L7.
[10] P. Pfeuty, *J. Phys. Colloq.*, 1978, **39**, C2-149-C2-160.
[11] A. V. Dobrynin, R. H. Colby, M. Rubinstein, *Macromolecules*, 1995, **28**, 1859-1871.
[12] M. Muthukumar, *J. Chem. Phys.*, 1996, **105**, 5183-5199.
[13] K. Nishida, K. Kaji, T. Kanaya, *J. Chem. Phys.*, 2001, **114**, 8671-8677.

[14] P. Lorchat, I. Konko, J. Combet, J. Jestin, A. Johner, A. Laschewski, S. Obukhov, M. Rawiso, *Europhys. Lett.*, 2014, **106**, article no. 28003.

[15] A. V. Dobrynin, M. Rubinstein, S. P. Obukhov, *Macromolecules*, 1996, **28**, 2974-2979.

[16] M. N. Spiteri, C. E. Williams, F. Boué, *Macromolecules*, 2007, **40**, 6679-6691.

[17] W. Essafi, M.-N. Spiteri, C. Williams, F. Boué, *Macromolecules*, 2009, **42**, 9568-9580.

[18] C. Holm, H. J. Limbach, K. Kremer, *J. Phys.: Condens. Matt.*, 2003, **15**, S205-S211.

[19] F. Boué, J. P. Cotton, A. Lapp, G. Jannink, *J. Chem. Phys.*, 1994, **101**, 2562-2568.

[20] N. Malikova, S. Cebasek, V. Glenisson, D. Bhowmik, G. Carrot, V. Vlachy, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12898-12904.

[21] N. Malikova, A.-L. Rollet, S. Čebašek, M. Tomšič, V. Vlachy, *Phys. Chem. Chem. Phys.*, 2015, **17**, 5650-5658.

[22] A. Salis, B. W. Ninham, *Chem. Soc. Rev.*, 2014, **43**, 7358-7377.

[23] W. Kunz, R. Neueder, *Specific Ion Effects*, World Scientific, New Jersey, 2009, 3-54 pages.

[24] Y. Marcus, *Chem. Rev.*, 2009, **109**, 1346-1370.

[25] P. Jungwirth, D. J. Tobias, *Chem. Rev.*, 2006, **106**, 1259-1281.

[26] K. D. Collins, *Biophys. J.*, 1997, **72**, 65-76.

[27] W. J. Xie, Y. Q. Gao, *J. Chem. Phys. Lett.*, 2013, **4**, 4247-4252.

[28] T. Morita, P. Westh, K. Nishikawa, Y. Koga, *J. Phys. Chem. B*, 2014, **118**, 8744-8749.

[29] N. Schwierz, D. Horinek, R. R. Netz, *Langmuir*, 2013, **29**, 2602-2614.

[30] C. Hotton, G. Ducouret, J. Sirieix-Plénet, T. Bizien, L. Porcar, N. Malikova, *Macromolecules*, 2023, **56**, 923-933.

[31] T. A. Waigh, R. Ober, C. E. Williams, J.-C. Galin, *Macromolecules*, 2001, **34**, 1973-1980.

[32] J. Combet, F. Isel, M. Rawiso, F. Boué, *Macromolecules*, 2005, **38**, 7456-7469.

[33] W. Essafi, N. Haboubi, C. Williams, F. Boué, *J. Phys. Chem. B*, 2011, **115**, 8951-8960.

[34] J. Combet, P. Lorchat, M. Rawiso, *Eur. Phys. J. - Spec. Top.*, 2012, **213**, 243-265.

[35] G. S. Manning, *J. Chem. Phys.*, 1969, **51**, 924-933.

[36] G. S. Manning, *J. Chem. Phys.*, 1969, **51**, 934-938.

[37] M. Muthukumar, *J. Chem. Phys.*, 2004, **120**, 9343-9350.

[38] M. P. Raskop, A. Grimm, A. Seubert, *Microchim. Acta*, 2007, **158**, 85-94.

[39] E. F. Lucas, C. R. E. Mansur, L. Spinelli, Y. G. C. Queiros, *Pure Appl. Chem.*, 2009, **81**, 473-494.

[40] T. Erdmenger, I. Perevyazko, J. Vitz, G. Pavlov, U. S. Schubert, *J. Mater. Chem.*, 2010, **20**, 3583-3585.

[41] S. Berezovska, V. Yanishpolski, V. A. Tertykh, M. V. Burmistr, K. M. Sukhyy, *J. Therm. Anal. Calorim.*, 2006, **86**, 93-96.

[42] E. Kiss, E. T. Heine, K. Hill *et al.*, *Macromol. Biosci.*, 2012, **12**, 1181-1189.

[43] M. Luksič, M. Bončina, V. Vlachy, M. Druchok, *Phys. Chem. Chem. Phys.*, 2012, **14**, 2024-2031.

[44] M. Seručnik, M. Bončina, M. Luksič, V. Vlachy, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6805-6811.

[45] S. Čebašek, M. Seručnik, V. Vlachy, *J. Phys. Chem. B*, 2013, **117**, 3682-3688.

[46] Y. Sakhawoth, L. J. Michot, P. Levitz, N. Malikova, *Chem. Phys. Chem.*, 2017, **18**, 2756-2765.

[47] J. M. Layman, E. M. Borgerding, S. R. Williams, W. H. Heath, T. E. Long, *Macromolecules*, 2008, **41**, 4635-4641.

[48] S. R. Williams, E. M. Borgerding, J. M. Layman, W. Wang, K. I. Winey, T. E. Long, *Macromolecules*, 2008, **41**, 5216-5222.

[49] A. Rembaum, H. Noguchi, *Macromolecules*, 1972, **5**, 261-269.

[50] A. Rembaum, H. Noguchi, *Macromolecules*, 1972, **5**, 253-260.

[51] S. R. Williams, T. E. Long, *Prog. Polym. Sci.*, 2009, **34**, 762-782.

[52] F. Mezei, *Neutron Spin Echo Spectroscopy*, Lecture Notes in Physics, vol. 128, Springer Verlag, 1980.

[53] C. Le Coeur, N. Jouault, N. Malikova, G. Sudre, *Actual. Chim.*, 2022, **477**, 21-31.

[54] Y. Sakhawoth, "Floculation - Formation et structure des agrégats entre les chaînes de polyélectrolytes et colloïdes argileux", Phd thesis, Université Pierre et Marie Curie - Paris VI, 2017, https://theses.hal.science/tel-01758928.

[55] M. Druchok, N. Malikova, A.-L. Rollet, V. Vlachy, *AIP Adv.*, 2016, **6**, article no. 065214.

[56] D. Qu, J. S. Pedersen, S. Garnier, A. Laschewsky, H. Möhwald, R. von Klitzing, *Macromolecules*, 2006, **39**, 7364-7371.

[57] W. Essafi, F. Lafuma, D. Baigl, C. E. Williams, *Europhys. Lett.*, 2005, **71**, 938-944.

[58] Y. Marcus, *Chem. Rev.*, 1988, **88**, 1475-1498.

[59] G. Lamoureux, B. Roux, *J. Phys. Chem. B*, 2006, **110**, 3308-3322.

[60] M. Sedlak, *Physical Chemistry of Polyelectrolytes: Structure and Dynamics of Polyelectrolyte Solutions by Light Scattering*, Surfactant Science Series, vol. 99, Marcel Dekker, 2001.

[61] K. L. Ngai, *Adv. Colloid. Interface Sci.*, 1996, **64**, 1-43.

[62] G. L. Squires, *Introduction to the Theory of Thermal Neutron Scattering*, Dover Publications, 1996.

[63] J. Hayter, G. Janninck, F. Brochard-Wyart, P. G. de Gennes, *J. Phys. Lett.*, 1980, **41**, L451-L454.

[64] F. Nallet, G. Janninck, J. Hayter, R. Oberthür, C. Picot, *J. Phys.*, 1983, **44**, 87-99.

[65] T. Kanaya, K. Kaji, R. Kitamaru, J. S. Higgins, B. Farago, *Macromolecules*, 1989, **33**, 1356-1359.

[66] S. Lee, V. R. Tirumala, M. Nagao, T. Tominaga, E. K. Lin, J. P. Gong, W. Wu, *Macromolecules*, 2009, **42**, 1293-1299.

[67] M. Adam, M. Delsanti, *Macromolecules*, 1977, **10**, 1229-1237.

[68] J. P. Hansen, I. R. McDonald, *Theory of Simple Liquids*, Academic Press, 1986.

[69] M. G. Oostwal, M. H. Blees, J. De Bleijser, J. C. Leyte, *Macromolecules*, 1993, **26**, 7300-7308.

[70] R. S. Koene, T. Nicolai, M. Mandel, *Macromolecules*, 1983, **16**, 227-231.

[71] A. Lezov, V. Rogozhin, A. Donets *et al.*, *J. Mol. Liq.*, 2022, **347**, article no. 118369.

[72] P. C. Martin, O. Parodi, P. S. Pershan, *Phys. Rev. A*, 1972, **8**, 2401-2420.

[73] E. Buvalaia, M. Kruteva, I. Hoffmann, A. Radulescu, S. Förster, R. Biehl, *ACS Macro Lett.*, 2023, **12**, 1218-1223.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 29-51
https://doi.org/10.5802/crchim.317

Research article

French Network on Solvation (GDR 2035 SolvATE)

Topological graphs: a review of some of our achievements and perspectives in physical chemistry and homogeneous catalysis

Sana Bougueroua [a,*], Ylène Aboulfath [b], Alvaro Cimas [a], Ali Hashemi [c], Evgeny A. Pidko [c], Dominique Barth [b] and Marie-Pierre Gaigeot [a,*,d]

[a] Université Paris-Saclay, Univ Evry, CY Cergy Paris Université, CNRS, LAMBE UMR8587, 91025 Evry-Courcouronnes, France

[b] Université Paris-Saclay, Univ Versailles Saint Quentin, DAVID, 78035 Versailles, France

[c] Inorganic Systems Engineering, Department of Chemical Engineering, Faculty of Applied Sciences, Delft University of Technology, 2629 HZ Delft, The Netherlands

[d] Institut Universitaire de France (IUF), 75005 Paris, France

E-mails: sana.bougueroua@univ-evry.fr (S. Bougueroua), mgaigeot@univ-evry.fr (M.-P. Gaigeot)

Abstract. This paper reviews some of our developments in algorithmic graph theory, with some applications in physical chemistry and catalysis. Two levels of granularity in the topological graphs have been developed: atomistic 2D-MolGraphs and coarse-grained polygraphs of H-bonded cycles. These graphs have been implemented with the key algorithms of isomorphism and polymorphism, in order to analyze molecular dynamics simulations of complex molecular systems. These topological graphs are transferable without modification from "simple" gas molecules, to liquids, to more complex inhomogeneous interfaces between solid and liquid for instance. We show hereby that the use of algorithmic graph theory provides a direct and fast approach to identify the actual conformations sampled over time in a trajectory. Graphs of transitions can also be extracted, showing at first glance all the interconversions over time between these conformations. H-bond networks in condensed matter molecular systems such as aqueous interfaces are shown to be easily captured through the topological graphs. We also show how the 2D-MolGraphs can easily be included in automated high-throughput in silico reactivity workflows, and how essential they are in some of the decisive steps to be taken in these workflows. The coarse-grained polygraphs of H-bonded cycles are shown to be essential topological graphs to analyze the dynamics of flexible molecules such as a hexapeptide in gas phase.

Keywords. Algorithmic graph theory, Conformational search, Molecular dynamics, Identification of conformers, Pathways, Reaction network.

Manuscript received 31 January 2024, revised 14 May 2024, accepted 15 May 2024.

* Corresponding authors

1. Introduction

The field of Theoretical and Computational Chemistry applies the laws of physics and chemistry coupled with computer programs to calculate the structures and chemical and physical properties of molecules in different states of matter, such as thermodynamic properties, spectroscopic signals, chemical reaction pathways, phase diagrams, etc. The arsenal of theoretical tools in computational chemistry has evolved in recent years, nowadays also including theoretical methods from Operation Research (OR), which uses algorithms to build solutions on well-formulated problems, or Artificial Intelligence (AI), which uses various Machine Learning methods, based in particular on neural networks, not only to predict new physical and chemical states, events and properties, but also to develop, for example, force fields or DFT functionals for simulations. This new era has triggered a revival in the field of theoretical and computational chemistry for research teams to develop new theoretical methods that include OR and/or AI to go beyond the simple use of "classical numerical methods".

Over the past decade, our group has approached this new era through the prism of algorithmic graph theory, in the context of OR and AI, based on the representation of matter in topological graphs [1–5].

A graph encodes topological properties of matter (in the same way for molecules, assemblies of molecules, liquids, solid materials, also interfaced with liquids) by means of vertices and edges that reflect the specific interactions (in pairs) between vertices. At the molecular level of representation, the vertices are usually associated to atoms while the edges report on interactions between atoms, e.g., chemical bonds and intermolecular interactions. Most graphs are defined in two dimensions (2D-graphs), any information related to, e.g., distances, angles, is usually not encoded into topological 2D-graphs unless vertices/nodes are specifically labeled with such information. Graphs can however be also three-dimensional, with an indication of coordinates in space that would thus encode intra- and inter-molecular interactions between atoms. Easier to obtain or to predict than 3D-graphs, 2D-graphs already carry information on the structure, the functional properties, and even the 3D shape of the materials they model. Examples include the classification of similar molecules according to their topology [6,7], the prediction of patterns in biological molecules [8], the prediction of the 3D structure of small molecules [9], etc. Molecular graphs are also commonly used in supervised machine learning algorithms, the framework of graph-based models for molecules is indeed naturally suited to carry out predictions in message-passing neural network schemes.

In bio-/chemo-/materials informatics, the challenge is to identify or design algorithms capable of obtaining molecular properties from input graphs and to follow these properties in time. We have developed a series of 2D-graphs, at various levels of granularity of representation, and associated algorithms in order to analyze physical and chemical structures and properties from atomistic molecular dynamics (MD) simulations (DFT-based MD and classical force field FF-MD). In these developments, our aim was to ensure that 2D-graphs and algorithms could be applied without any modification to "simple" isolated molecules, as well as to assemblies of molecules and to more complex liquid and solid states of matter, including inhomogeneous solid/liquid interfaces [1–5, 10,11].

This paper reviews some of our developments and achievements, which are also included in the GaTewAY software [2,3,12]. Other research groups, also experts in MD simulations, have worked on 2D-graphs. In the last decade, there have been developments of algorithmic graph theory devoted to the analysis of various types of molecular dynamics simulations, from, e.g., the conformational analysis of gas phase molecules and clusters, to their chemical reactivity, to the dynamics of H-bonds in liquids, to the dynamics of the solvation shells of ions in liquids, to the structural and dynamical analysis of complex aqueous interfaces in condensed matter [1,10,13–22].

Section 2 of this paper reviews the definitions used in the atomistic 2D-MolGraphs developed by our group [1,10], and the associated algorithms, including the key isomorphism algorithm ensuring that a whole trajectory can be analyzed in terms of the relationships between the conformations explored over time. The topological analyses hence provide a statistical view over the whole timescale of the trajectory and over the whole set of conformations explored. Statistics is crucial for the detailed knowledge of the dynamics of isolated molecules as well as for the

dynamics of molecules in the liquid state, as applications in Section 3 will show.

We also review our recent developments of 2D-graphs consisting of H-bonded cycles and the key polymorphism algorithm [4], which have been built to go beyond the atomistic representation in 2D-MolGraphs and to be able to represent (bio-)molecules whose 3D structures and dynamics are solely based on hydrogen bonds. The algorithms we have developed enable us to track the complex conformational dynamics of flexible H-bonded molecules in real time. Section 3 will show that, while the interpretation of the complex conformational dynamics of a highly flexible hexapeptide in the gas phase would remain elusive at the atomistic level of representation (2D-MolGraphs), it is well understood by means of coarse-grained graphs of H-bonded cycles (polygraphs) and by means of polymorphic "metastructures". Only the coarse-grained representation in the graphs allows such comprehension.

Section 3 will further show how atomistic 2D-MolGraphs can easily be included in automated high-throughput in silico reactivity workflows and how essential they are in some of the decisive steps to be taken in these workflows. We implemented the 2D-MolGraphs in the computational catalytic reaction space exploration method ReNeGate [5] and the high-throughput reactivity screening HiREX workflow [11], specifically designed to explore realistic catalytic systems and identify thermodynamically feasible chemical transformations, corresponding to secondary catalyst deactivation and inhibition paths.

Prospects and new developments in progress are discussed in Section 4, which, with this review of methods and applications, we hope will spark further requests for additional algorithms and new applications in our physical chemistry and chemistry communities.

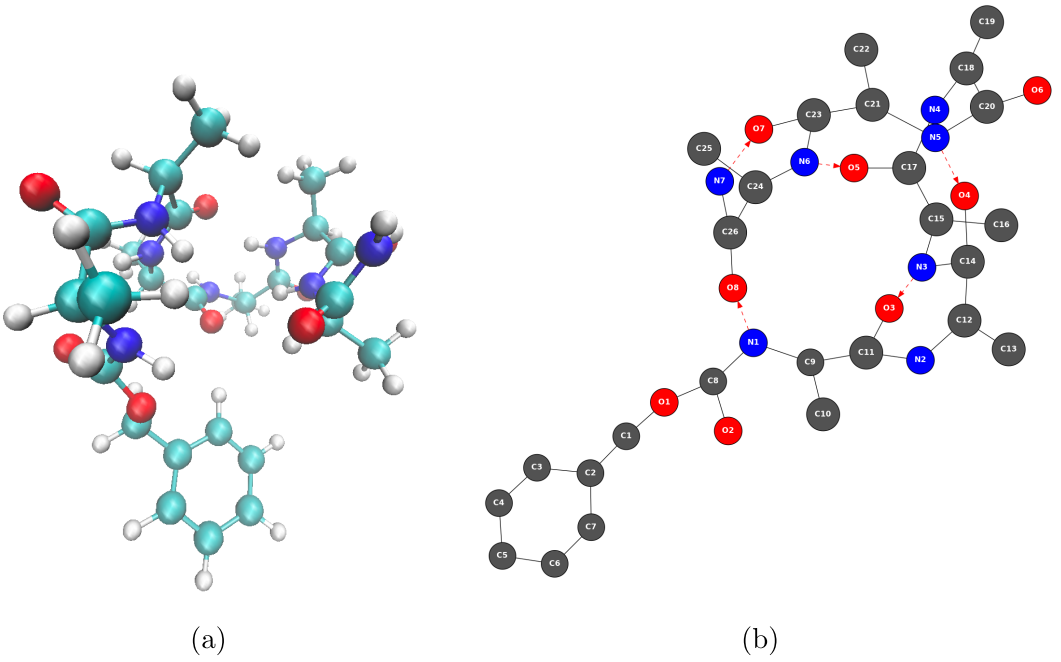
2. Methods

2.1. *2D-MolGraph for modeling a conformation*

Our developments have been aimed at defining topological 2D molecular graphs (labeled 2D-MolGraph) and associated algorithms in order to automatically analyze MD trajectories from the knowledge of the time evolution of the conformations and hence automatically detect conformational changes through topological changes. Our 2D-MolGraphs share a similar degree of granularity in representing the topology of molecular systems to the one used by previous implementations in the literature in chemistry [13,14, 17], i.e., a vertex in the 2D-MolGraph represents an atom or a molecule and an edge between two vertices represents the interactions/bonds (covalent bond, hydrogen bond...) between two atoms/molecules. Most of the literature on graphs in the chemical community does not consider the chemical nature of the atoms in the vertices, which is not efficient for recognizing identical structures where chemically identical atoms have been swapped. Moreover, they lack specific chemical information (e.g., covalent bonds, hydrogen bonds, exchange of atoms in homogeneous clusters, etc.) that might be relevant for a more detailed characterization of the structures.

One crucial step in our method has been to define a model that represents any molecular conformation with the right level of granularity and be transferable without any modification from gas phase molecules and clusters to the condensed phase (solids, liquids, interfaces between solids and liquids). To that end, we have chosen to define any molecular conformation by a colored mixed graph $G = (V, E_C, A_H, E_I, E_O)$, with both (directed) arcs and (undirected) edges. Such a graph is denoted 2D-MolGraph, in which each vertex represents an atom while the edges represent covalent bonds (E_C), hydrogen bonds (A_H), ionic (or electrostatic) interactions (E_I) typically between a cation/anion atom and other atoms, organometallic interactions (E_O) between metallic atoms and their surrounding. Only the hydrogen bonds are associated to directed edges (from the donor to the acceptor atom). The hydrogen atoms in a molecular system are not included in vertices of the 2D-MolGraph. Instead, their presence is solely known by directed edges. Hence, any hydrogen that is not involved in a hydrogen bond is not represented in the 2D-MolGraph.

The definition of bonds and interactions is mainly based on Euclidian distances. The Euclidian distance between a pair of atoms $[a, b]$ with respective Cartesian coordinates (x_a, y_a, z_a) and (x_b, y_b, z_b) is: $\sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2}$. There is a covalent bond or an interaction between two atoms if the Euclidean distance is less than a cutoff distance D_r. For covalent bonds, the algorithm defines the D_r distance by the sum of covalent radii of atoms a and b



(a) (b)

Figure 1. Illustration of the passage from a 3D conformation (a) to a 2D-MolGraph (b). (a) One snapshot in the 3D representation of the gas phase Z-Ala$_6$–COOH peptide ($C_{26}H_{39}N_7O_8$) extracted from a MD trajectory. Carbon atoms are colored turquoise, nitrogen atoms blue, oxygen atoms red, and hydrogen atoms light gray. (b) Associated topological 2D-MolGraph. Vertices are colored dark gray, blue, and red, corresponding to carbon, nitrogen and oxygen atoms, respectively. Hydrogen atoms are not included in vertices, their knowledge is included only through directed edges that represent hydrogen bonds. Edges are black lines for a covalent bond and red dashed lines (arcs) for hydrogen bonds, the latter directed from the donor to the acceptor of the H-bond.

with an additional 2% margin (that typically includes the effect on distances from vibrational motions). For hydrogen bonds, the algorithm sets up the default D_r value between the hydrogen atom (donor) and the acceptor atom (heavy atom) to 2.3 Å, which can be changed by the user. For the organometallic and ionic interactions, the user is free to set case-specific D_r distances. For example, the distance between manganese and oxygen atoms used in one of the applications in Section 3 was set to 2.44 Å. This choice was made because the developers assume that the covalent bonds are stronger than the other types of interactions between atoms. More details are found in [1,5,12].

One can easily define and implement new relevant interactions that are needed to describe a given molecular system, and hence augment the number of definitions for the edges in the 2D-MolGraphs.

In order to take into account the chemical type of the atoms in a 2D-MolGraph, we apply a special case of graph coloring, such that the vertices of a given 2D-MolGraph display the same color *if and only if* the corresponding atoms have the same chemical type (see Figure 1). Figure 2 illustrates an adjacency matrix built prior to the construction of a 2D-MolGraph. The matrix shown here is associated to a selected part of the Z-Ala$_6$–COOH peptide from Figure 1a (3D structure) and Figure 1b (2D-MolGraph). As the peptide contains 80 atoms, only a selected part of the peptide has been extracted here for this illustration. The figure shows that covalent bonds and hydrogen bonds of interest in the conformation of this peptide are encoded in the adjacency matrix (with crosses and circles, respectively) and that the matrix has the colored information of the actual chemical nature of the atoms. The graph on the right side of the figure is

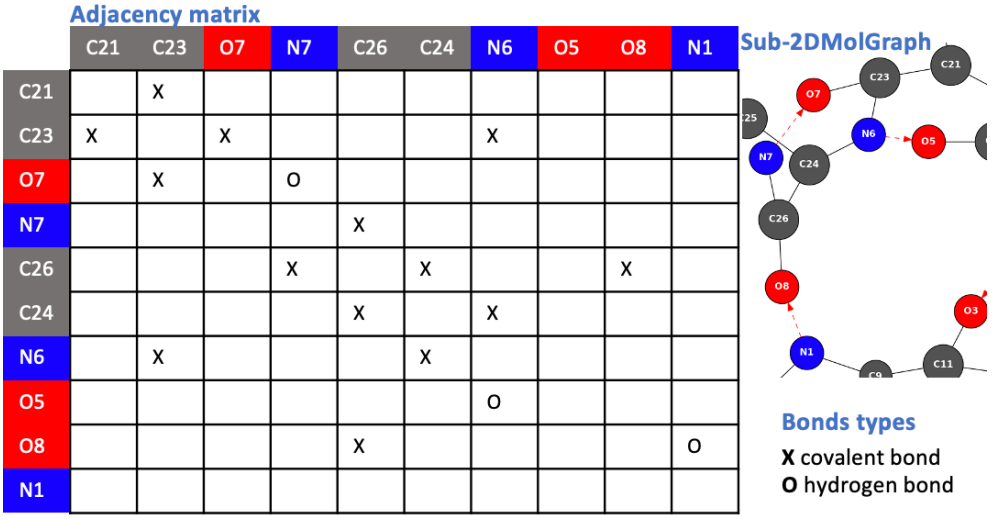


Figure 2. Illustration of an adjacency matrix for a selected portion of the Z-Ala$_6$–COOH peptide shown in Figure 1a (3D-structure) and Figure 1b (2D-MolGraph).

the subgraph of the 2D-MolGraph associated to the selected part of the molecule encoded in the matrix.

With 2D-MolGraphs in our hands, the exploration with time of molecular conformations along MD simulations can easily be seen as the exploration of graph topologies, that can be tracked using graph theory-based methods, such as isomorphism. Graph isomorphism as defined in [23] allows representation of each conformer with a fingerprint graph and comparisons between graphs. In our applications, isomorphism consists in comparing the distribution of edges between two 2D-MolGraphs: if the graphs compared have the same set of bonds/interactions connected to the same set of atoms (in terms of chemical types/colors for the graphs), these graphs are then isomorphic, i.e., the two graphs are identical. More formally, an isomorphism is a bijection between the vertex sets of the two graphs if and only if it induces a corresponding bijection between their edge sets (if such an isomorphism exists, the two graphs are said to be isomorphic). A (non-polynomial) algorithm to check if such an isomorphism exists is proposed in [23–26].

Isomorphism checks, together with keeping the chemical nature of the atoms, are the key components of the conformational search over MD trajectories. The changes in conformations are hence followed over time by scanning trajectories for changes in bonding patterns of choice (among hydrogen bonds, proton transfers, coordination numbers, covalent bonds, and organometallic interactions). Once the different conformations of the molecular system have been found by graph analysis, a graph of transitions can be generated, similar to the ones considered in different analysis or generation of temporal graph sequences [27–29]. This graph has its vertices composed of the conformations that have been identified and its edges are composed by the transitions found between conformations. Both vertices and edges in the graph of transitions contain information on the percentage of existence of a conformation over the duration of the trajectory (for the vertices) and the percentage of times a transition between two vertices/conformations has been seen (for the edges). In one glance, one can hence see the relationships between the conformations and associated statistics.

Figure 3 shows an example of a graph of transitions. This graph is composed of four vertices, each vertex represents one molecular conformation that has been identified by isomorphism along the trajectory. Each vertex has a number and a label that represents the conformation. For instance, one vertex in this picture is labeled with “1” and “N1-O1”, meaning this is the first conformation identified over the trajectory, and the associated conformation has one hydrogen bond between atoms “N1” and “O1”. The “68.97%” number in the vertex is the percentage of

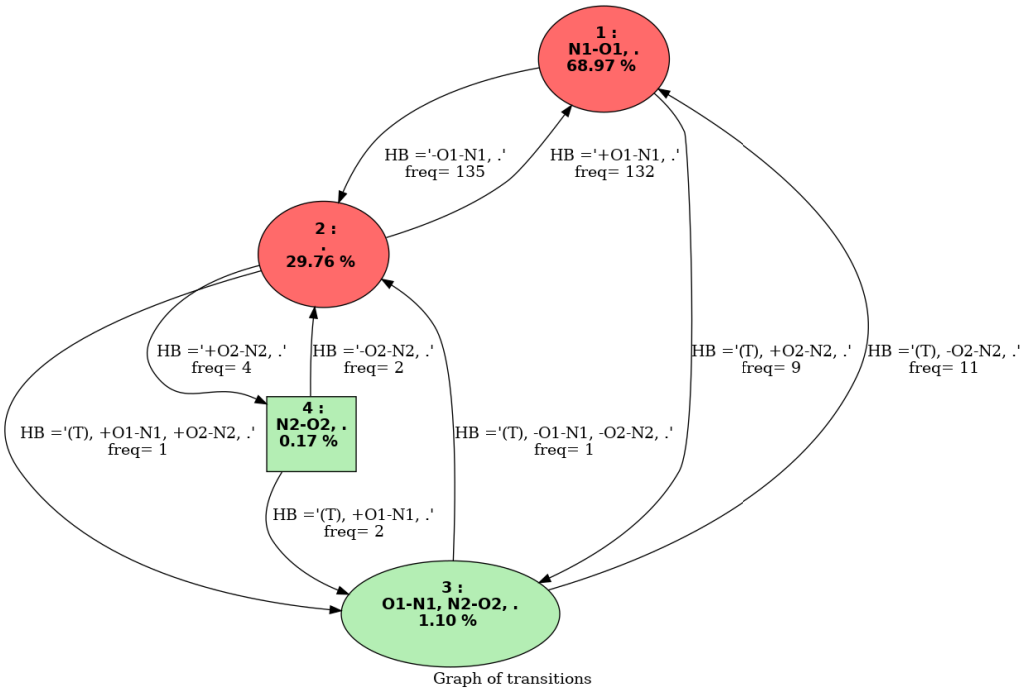


Figure 3. Graph of transitions. See text for nomenclature and colors.

time this conformation has been seen over the whole trajectory. Hence, in this graph of transition, conformation 1 is the most frequent conformation that has been observed over the trajectory, with a total percentage of appearance of ~69% over the trajectory. The edges between the vertices are labeled with two kinds of information: (1) the total frequency rate for going from one conformation to the other one, (2) the bond(s)/interactions(s) that have changed when going from one conformation to the other. For instance, one can observe a large conformational dynamics between conformations 1 and 2 in the graph of transitions in Figure 3; such an event occurs around 130 times, back and forth. Also of note, the hydrogen bond N1···O1 disappears when going from conformation 1 to conformation 2 (and appears on the reverse way). Conformation 2 in the graph of transitions is labeled with "2" and ".", the latter meaning that this conformation contains no hydrogen bond. Conformation 2 has been seen ~30% over the whole trajectory.

The colors of the vertices in the graph of transitions directly give the most relevant conformations in terms of appearance periods. We hence colored red the conformations that appear at least $P_{\min}$% (an input parameter that the user can change the default value of 4% has been used here) and the ones in green occur below this threshold. All the conformations explored along the MD simulations can be kept in the graph of transitions. Such information might indeed be useful for some analyses, typically when rare events (rare conformations) are investigated. The user can modify this at will.

2.2. *From a topological 2D-MolGraph to a coarse-grained graph of H-bonded cycles*

In [2,3] we have shown that the rationalization of the conformational dynamics becomes complex and almost impossible to achieve for flexible H-bonded molecules that isomerize frequently over time through the dynamics of breaking and forming of their H-bonds. As shown in these references, a simple short peptide such as Z-Ala$_6$–COOH (illustrated by one 2D-MolGraph in Figure 1) already shows a high flexibility of its network of N–H···O H-bonds at relatively low temperatures (gas phase MD trajectories). Numerous breaking/forming of

H-bonds were observed over the trajectories, which signaled the appearance and disappearance of several conformers of the gas phase peptide numerous times along the trajectory. This high dynamical flexibility of the H-bond network, however well represented by the 2D-MolGraphs and their graph of transitions [2,3], prevents a clear rationalization of the actual conformational dynamics of the peptide. In particular, some of the H-bonds in the network, though formed between different atoms, seem to play a similar role into the final 3D structure of the peptide molecule, i.e., into the final folded/semi-folded/unfolded skeleton. This is not captured by the 2D-MolGraphs at the atomic level of representation, but this similarity of H-bonded cycles can be captured by graphs defined at a higher/coarser granularity of representation. We therefore defined graphs in which the vertices are directly associated to the H-bonded cycles formed, whose polymorphism is furthermore taken into account. These concepts are now explained.

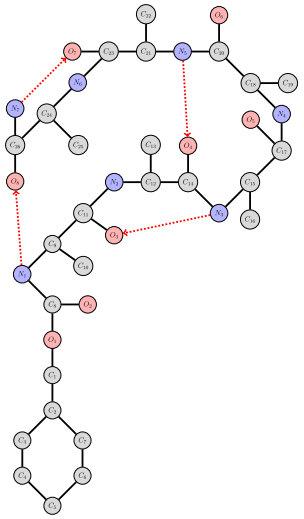
Cycles in molecular graphs have been shown to be good representations of the structure of molecular systems [30–32]. We hence proposed a representation of each conformation identified by a 2D-MolGraph over a trajectory based on a well chosen set of cycles. The cycles of interest to us are the ones formed by at least one H-bond. With these, we will be able to quantify and follow in time the changes in the H-bonded network of molecular structures. Our current developments have been done for gas phase molecules only [4]. Note that the number of cycles in a graph can be exponential with respect to the number of vertices. Therefore in our approach, we only consider a subset of cycles in the 2D-MolGraph, called "minimum cycle basis", restricted to the H-bonded cycles for each 2D-MolGraph of the trajectory. Given a graph, finding a minimum cycle basis, which is not necessarily unique, can be done in polynomial time with an evolution of the algorithm of Horton [33]. More details can be found in [4].

Given a 2D-MolGraph, we compute its minimum basis set of H-bonded cycles. The associated graph of cycles is defined as a graph in which the set of vertices is the cycle basis, and there is an edge between two vertices (cycles) if and only if these two cycles interact, i.e., they share *at least* one covalent bond or one hydrogen bond in the 2D-MolGraph. Hence each 3D-conformer in the trajectory of a molecule is represented by a graph of cycles. The evolution in time of the conformations in a MD trajectory can be represented by the sequence of their graphs of cycles.

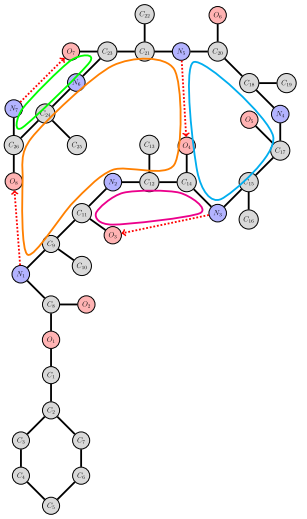
Figure 4 illustrates the transformation from a topological graph (Figure 4a) to a graph of cycles (Figure 4c). Considering the 2D-MolGraph in Figure 4a and the minimum cycle basis in Figure 4b, there are four H-bonded cycles, each of them of various size, composed of one or several hydrogen bonds. Hence, the pink vertex in Figure 4c is built on one H-bonded cycle composed of six vertices/atoms in the 2D-MolGraph (this is a six-membered H-bonded cycle), while the larger orange vertex/cycle is built upon two hydrogen bonds (see the two directed dashed red edges in Figure 4a). Figure 4c shows the graph of cycles obtained from the minimum cycle basis shown in Figure 4b. In this graph, the vertices are the H-bonded cycles, labeled by the heavy atoms involved in the hydrogen bond(s) producing them. Taking once again the examples of the pink and orange cycles/vertices depicted in Figures 4b–4c, the pink vertex is labeled N3O3 as it is built on the N3–H$\cdots$O3 hydrogen bond, while the orange vertex is labeled with its two constitutive H-bonds N1–H$\cdots$O8 and N5–H$\cdots$O4. The orange, pink and blue vertices/cycles interact with each other as seen through the edges connecting these three vertices (i.e., sharing at least one covalent bond or one hydrogen bond). On the other hand, the green vertex (related to the H-bonded cycle N7O7) interacts only with the orange one.

Given a MD trajectory, each 2D-MolGraph is now associated to a minimum cycle basis and to the corresponding graph of cycles. In other words, there is one graph of cycles per 2D-MolGraph (i.e., per identified molecular conformation).

The set of conformational isomers explored over the MD trajectory can furthermore be summarized by one single graph of cycles uniting those of all the identified conformers. This union takes into account all the possible H-bonded cycles as well as all the possible interactions between these cycles which were observed in all the identified conformers. Some of these H-bonded cycles can be identical in different conformers. Some of these H-bonded cycles can be similar to each others in the sense that they are built upon different donor/acceptor atoms but they are playing the same role in the final structure of the molecule. One therefore has to recognize the similarity between



(a) 2D-MolGraph of a given molecular conformation of the hexapeptide Z-Ala$_6$−COOH.



(b) Minimum H-bonded cycle basis computed for the 2D-MolGraph in Figure 4a.

(c) The graph of cycles for the basis set of H-bonded cycles found in Figure 4b.

Figure 4. Example for the coarse-grained representation of a 2D-MolGraph for one conformation of the gas phase peptide Z-Ala$_6$–COOH. See text for details.

the H-bonded cycles of the different miminum cycle bases (i.e., similarity between the H-bonded cycles in the different identified conformations).

To that end, the next stage was to develop an algorithm that can group together the cycles of different minimum cycle bases which are similar, i.e., the ones playing the same role in the structure of different conformers. The similarity constraint that we introduced is the basis to cluster the union of cycles in all the minimum cycle bases of all the conformers of the trajectory that we finally want to obtain. Each part in such a clustering is called a *polymorphic cycle*, in which all cycles are considered as different forms of a same cycle in the structures of the conformers it appears in. The hypothesis that we made is that a cycle can evolve over time, that is to say that its set of links can evolve while retaining its same role in the molecular structure. The trajectory can therefore be seen as the interaction of the H-bonded cycles evolving over time in their atomic structure (polymorphism) and still interacting in the same way, but with some of these cycles appearing or disappearing over time. The final graph thus obtained is called a polygraph, a contraction of “polymorphic cycle graph”; note that this definition is different from the one of a polygraph as a generalization of the directed graph [34], or that used in chemoinformatics related to polymers [35]. The whole algorithms have been detailed in [4].

We now show an application of this methodology for the gas phase chondroitin disulfate CS2S4S molecule, for which a 3D snapshot is reported in Figure 5. Figure 6 reports the polygraph obtained for a 600 K trajectory of this gas phase molecule. Each vertex of the polygraph is labeled by “**P**-XX” where **P** stands for polycycle and XX is the set of oxygen and/or nitrogen atoms (with their associated label number in the list of the molecule’s atoms) involved in the H-bond of the polycycle (i.e., one of the two partners which close the H-bonded cycle). Each polycycle has been assigned a given color. Note that a polycycle can be built over one or over several H-bonds (two to three for this molecule). In that case, the atoms involved in the series of H-bonds are written on a line with a “,” between them. See some examples in the vertex colored in violet for several instances of multiple H-bonds forming a polycycle.

Figure 6 shows that this polygraph is a complete graph made of five vertices. It is complete because there is an edge between each pair of vertices. As a reminder, there is an edge between two polycycles

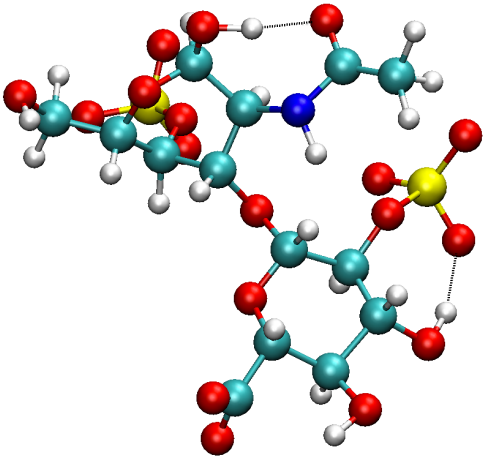


Figure 5. A 3D representation of chondroitin disulfate CS2S4S ($C_{14}H_{20}NO_{18}S_2$). Colors for the atoms: carbon in turquoise, nitrogen in blue, oxygen in red, hydrogen in light gray, and sulfur in yellow.

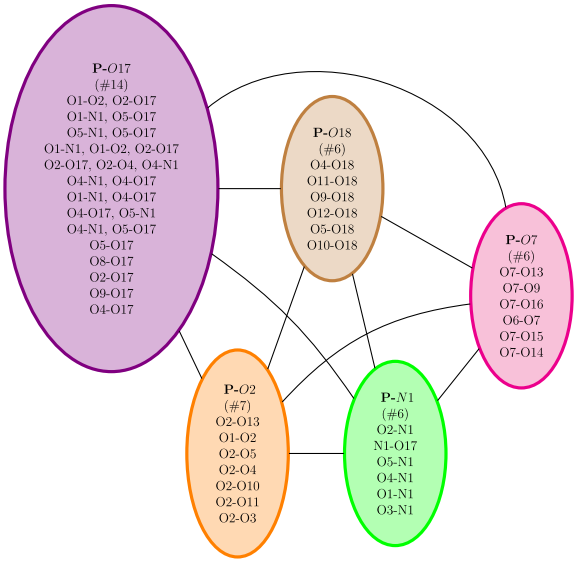


Figure 6. The polygraph obtained for a MD trajectory of the gas phase chondroitin disulfate CS2S4S molecule.

whenever these two cycles have atoms that share at least one covalent bond or one H-bond in the initial 2D-MolGraph (i.e., these polycycles are interacting with each other within the molecular structure). Because the polygraph is complete, the number of polycycles cannot be reduced. All vertices involve different polymorphic identities. Four of the five vertices have six or seven different actual identities for the H-bonded cycle that forms the polycycle (dark green, pink, green, and orange), while there is an even larger diversity of 14 different identities for polycycle **P**-O17 (purple). For instance, vertex **P**-O7 is associated to a polycycle that can have six different identities of the H-bond that closes the cycle/polygon with the O7 atom. As can be seen, this O7 can either H-bond to the H atom carried by the O13, or to O9, O16, O7, O15 and O14. In the vertex in the violet color, some of the lines report other identities than the **P**-O17 identity. For instance, one line reports O1-O2 and O2-O17, which means that the **P**-O17 polycycle is built over two simultaneous hydrogen bonds, i.e. the O2-O17 H-bond but also the O1-O2.

3. Review of selected applications

3.1. *H-bond dynamics of a flexible gas phase peptide and the relevance of the coarse-grained graphs of H-bonded cycles*

The conformational dynamics of the gas phase hexapeptide Z-Ala$_6$–COOH ($C_{26}H_{39}N_7O_8$, 80 atoms, 3D illustration in Figure 7) is analyzed hereby both in terms of the 2D-MolGraphs (atomistic level of topology representation) and in terms of the coarse-grained polymorphic cycles (coarse-grained level of topology representation) in order to present the strengths and limitations of each level. An ab initio MD (AIMD) trajectory of ~6.0 ps (12,294 snapshots, δt of 0.5 fs) at ~450 K serves as the basis for the conformational dynamics to analyze. This temperature has been chosen as a good example for the analysis of trajectories where several medium and large conformational changes are expected to occur, which are always rather hard to characterize without the help of topological graphs.

3.1.1. *A high conformational dynamics of Z-Ala$_6$–COOH provided by the 2D-MolGraphs*

The analysis of the trajectory in terms of topological molecular 2D-MolGraphs [1,3] indicates that 93 different conformations are sampled over ~6 ps (which hence shows a rather highly dynamical peptide), built over nine different H-bonds. The structure of each identified conformer is composed of one

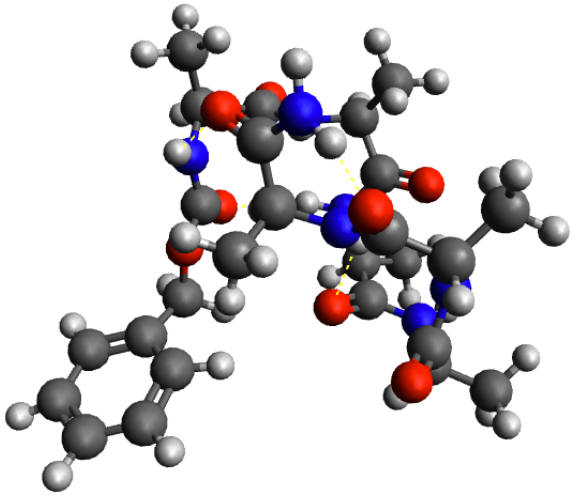


Figure 7. A 3D representation of the hexapeptide Z-Ala$_6$–COOH. Colors for the atoms: dark gray for carbon, dark blue for nitrogen, red for oxygen, white for hydrogen.

to six H-bonds that are formed simultaneously. The peptide is found in either opened structures where a low number of simultaneous H-bonds are present or in H-bonded folded structures.

As illustrations, Figure 8 presents two 2D-MolGraphs corresponding to two identified conformers of Z-Ala$_6$–COOH, respectively produced by five (Figure 8c) and three H-bonds (Figure 8d). Two of these H-bonds (N7$\cdots$O7 and N6$\cdots$O5) are present in both conformations.

Figure 9 is the graph of transitions that summarizes the whole 6 ps of conformational dynamics of Z-Ala$_6$–COOH at 450 K. As can be immediately seen, this graph of transitions is composed of an extremely high number of vertices and edges connecting these vertices, which is the signature of the high dynamicity and high flexibility of the peptide. There are 93 vertices in the graph for the 93 different conformers of Z-Ala$_6$–COOH found. In the graph of transitions reported in Figure 9, most of the identified conformers (in the green vertices) appear over very short periods of time (less than 0.24 ps) while two conformers (red vertices) appear over larger durations of time.

While the high number of vertices and edges in the graph of transitions in Figure 9 illustrates the high flexibility of the peptide at 450 K that is nicely captured by our 2D-MolGraph topological graphs, it also illustrates the limit of the atomic level of granularity used in the 2D-MolGraphs for highly flexible molecular systems where presumably too much information is provided in the graph of transitions and is hard to process. Furthermore, some of the periods of times that the 2D-MolGraphs analyze as breaking/forming covalent bonds and/or H-bonds correspond in practice to the dynamics of these bonds around the threshold values employed in the method for conformational recognition. Some of the transitions observed between conformers are thus in practice the actual signature of the existence of one single "meta-conformation" around which dynamicity and flexibility occur.

This information can however not be extracted from the atomistic topological 2D-MolGraphs. To get that information, one has to analyze the trajectory with a coarse-grained topology representation, whose results are described in the following section.

3.1.2. *Polygraphs are the good coarse-grained representation to analyze the conformational dynamics of Z-Ala$_6$–COOH*

To go beyond the limitations of the atomistic representation in the topology highlighted above in the case of the highly flexible Z-Ala$_6$–COOH, we now apply the coarse-grained topology representation described in Section 2.2 to analyze the conformational dynamics of this peptide: it now consists in representing a molecule through the ensemble of its H-bonded cycles and to apply an algorithm of polymorphism in order to recognize the cycles that are isomorphic to each others. Figures 10 and 11 illustrate the two immediate outputs of this analysis, respectively showing the global polygraph generated over the whole 450 K trajectory in Figure 10 and the chronogram (time evolution) of the nine identified polymorphic cycles in Figure 11.

There is a total of 23 H-bonded cycles over the 93 2D-MolGraphs extracted from the ~6 ps trajectory of Z-Ala$_6$–COOH at 450 K. Once polymorphism has been applied over these 23 H-bonded cycles, some of the cycles could be merged, thus resulting into nine polymorphic cycles. Each polycycle contains between one and nine possible identities. The obtained global polygraph that represents the whole 6 ps trajectory is shown in Figure 10, built over the nine vertices that represent the nine identified

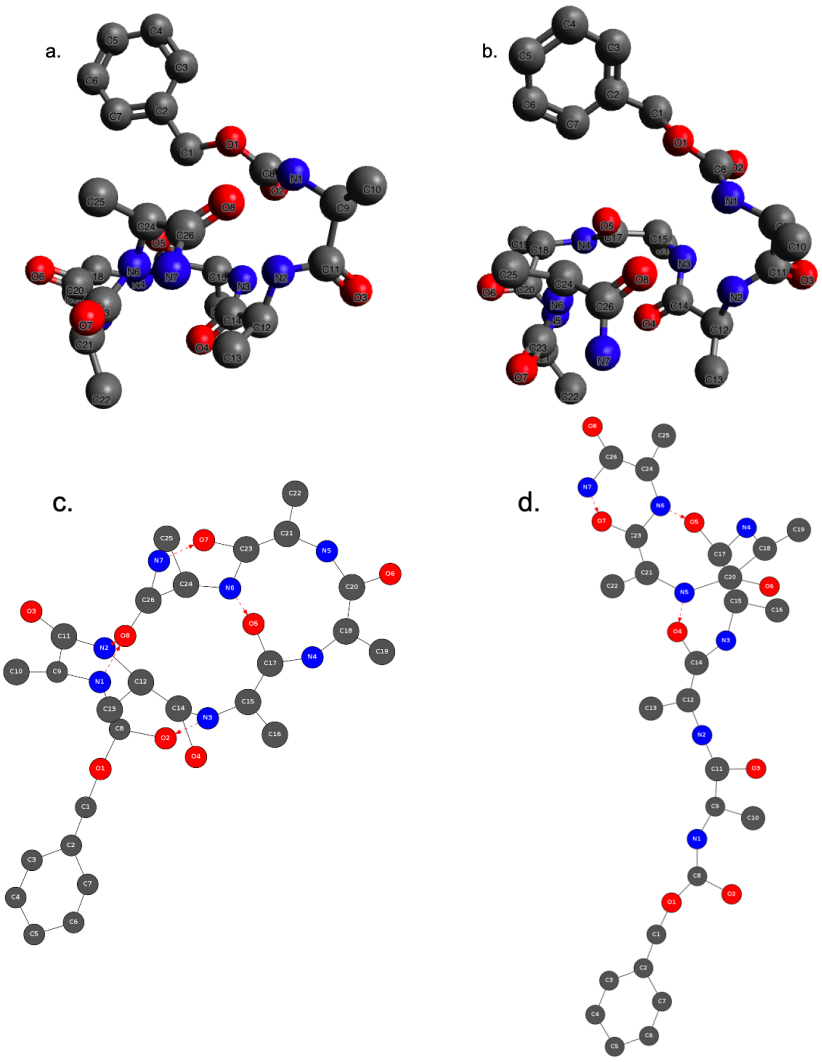


Figure 8. Two 2D-MolGraphs (c and d) of the gas phase peptide Z-Ala$_6$–COOH and their associated 3D structures (a and b) extracted from the MD trajectory analysis. Colors of the vertices: dark gray for C atoms, dark blue for N, red for O. No vertices for hydrogens, their knowledge is in the directed edges (arcs). Solid black edges in the graphs are for covalent bonds, the red arcs are the hydrogen bonds directed from the donor to the acceptor (heavy) atoms.

polymorphic H-bonded cycles. One can therefore see that five of the vertices/polycyles have only one possible identity in terms of the H-bonded cycle (note that the H-bonded cycle for the polycyle P-N1,O8,N2 is built upon two hydrogen bonds, i.e., N1$\cdots$O8 and N2$\cdots$O8), two other vertices have a limited number of isomorphic identities (two identities for the dark blue vertex P-O4,N5, three identities for the magenta vertex P-N3). Two vertices are much more polymorphic as they adopt four (cyan vertex P-O4,N5) and up to nine (orange vertex P-O8) different polymorphic identities. Furthermore, one can remark that several of these polycycles are built upon two or three H-bonds. For instance, in the orange P-O8 polycycle, the first identity "N1-O8,O4-N4" is built upon two H-bonds (N1$\cdots$O8 and O4$\cdots$N4), the first identity of the cyan polycycle P-O4,N5 is built upon three H-bonds (O2$\cdots$N3, N1$\cdots$O8, and O4$\cdots$N5).

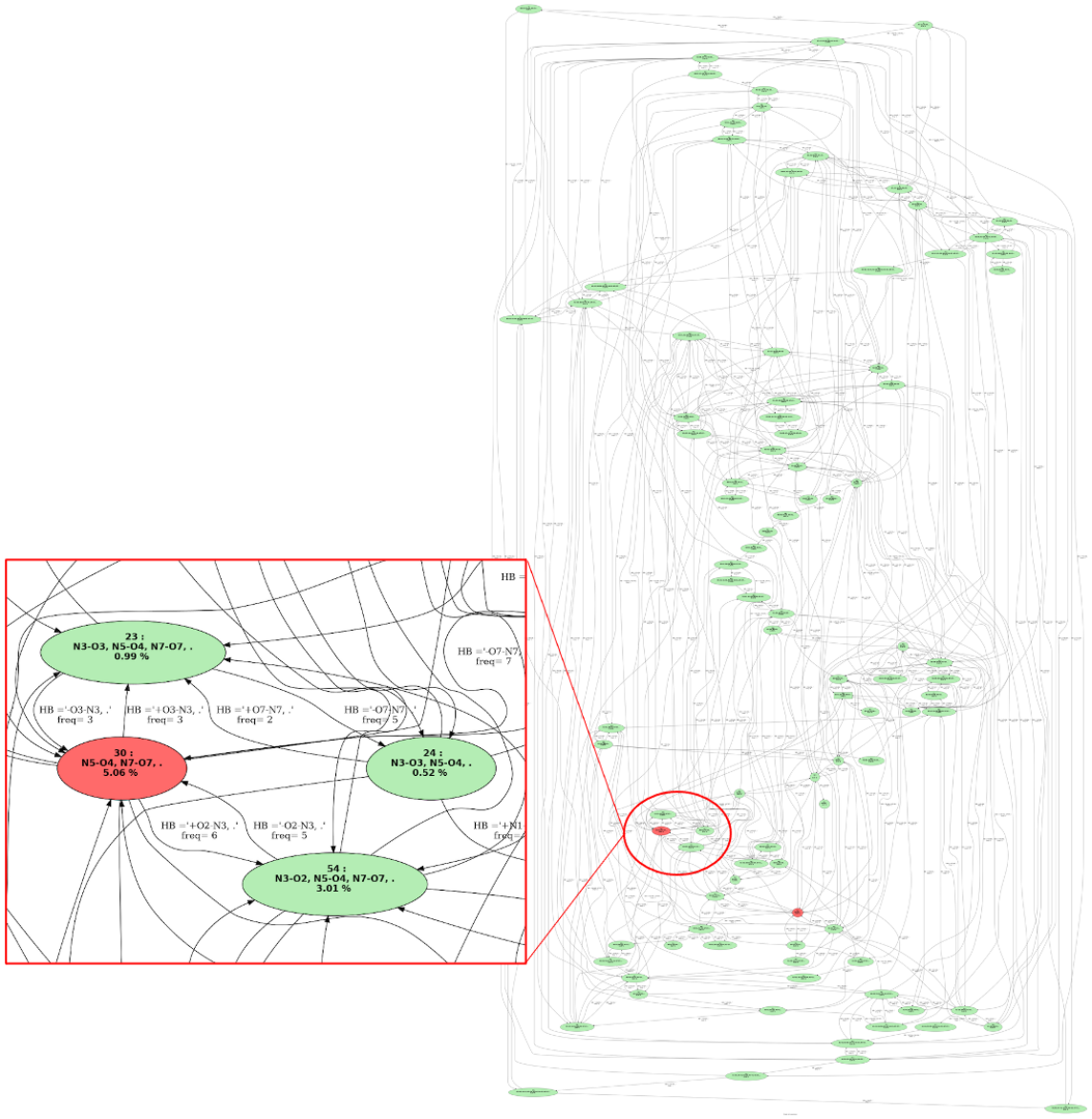


Figure 9. Graph of transitions for the 450 K dynamics of the gas phase Z-Ala$_6$–COOH peptide. The vertices indicate the explored conformers. Vertices in red are for conformers with a total percentage of appearance $P_{\min}$% greater than 4% of the dynamics time (parameter $P_{\min}$% can be modulated at will), vertices in green are for conformers with $P_{\min}$% < 4%. Directed edges between vertices indicate transitions between two conformers as observed over time. The labels on each edge provide the total percentage of occurrence of the transition and the associated chemical change(s) that occur. A zoom over a small portion of the graph of transitions is provided on the left-hand side.

The low/high number of identities within each polymorphic H-bonded cycle informs on the sections of the peptide with a low/high structure flexibility. The higher polymorphic nature of two of the polycycles in Z-Ala$_6$–COOH corresponds to high structure flexibility in these two zones, while

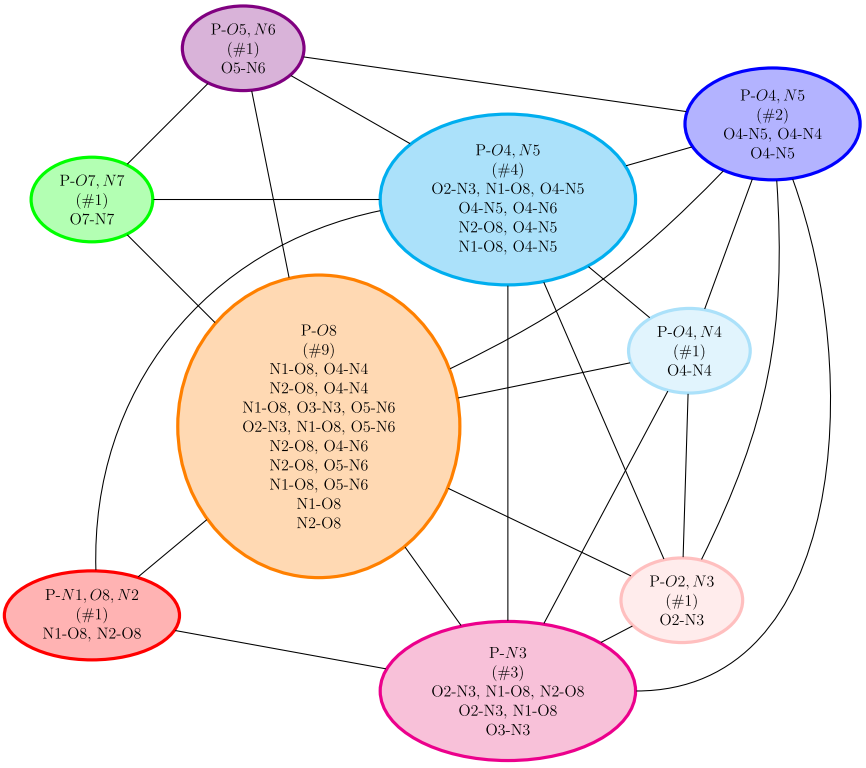


Figure 10. The global polygraph obtained over the 6 ps AIMD trajectory of the Z-Ala$_6$–COOH peptide at 450 K. The color coding of the vertices is identical to the color coding of the lines in the chronogram in Figure 11. Each vertex of the polygraph is labeled by "P-XX" where P stands for polycycle and XX is the set of the oxygen atom and/or the nitrogen atom (with their associated label number in the list of the atoms of the molecule) involved in the H-bond of the polycycle. Each vertex contains the list of the polymorphic cycles, each one being labeled by the atoms' names and labels in the H-bond(s) building the H-bonded cycle.

maintaining the same general role of the H-bonded cycle into the final 3D structure. This high flexibility was already observed in the previous section with the atomistic topological analysis of the 2D-MolGraphs. The analysis with the polymorphic H-bonded cycles immediately identifies the underlying "meta-structures" that are of interest for the comprehension of the conformational dynamics of the peptide.

In the polygraph, there is an edge between two vertices whenever two polycycles interact together, i.e., the conformations belonging to each vertex share at least one covalent bond or one H-bond. The polygraph in Figure 10 is not complete, i.e., not all the polycycles are directly connected in pairs by an edge. This incompleteness can be explained by the following two reasons that are related to the rules applied for building a polygraph: (i) either there is a conformation for which the identities of the two polycycles appear simultaneously, without interaction (the criterion for interaction/edge is the sharing of at least one covalent bond or one H-bond); or (ii) there is no conformation for which the two identities appear simultaneously, however their merging does not respect the polycycle rules. For example, we found that the blue P-O4,N5 polycycle/vertex and the orange P-O8 polycycle/vertex do not interact because there is at least one identity from P-O4,N5 and one identity from P-O8 that appear simultaneously in one conformation of Z-Ala$_6$–COOH. As these two polycycles do not share one covalent bond/H-bond, there is thus no interaction/edge in the polygraph.

The chronogram of the Z-Ala$_6$–COOH polycycles, presented in Figure 11, shows the evolution in time of the nine polycycles at the temperature of 450 K.

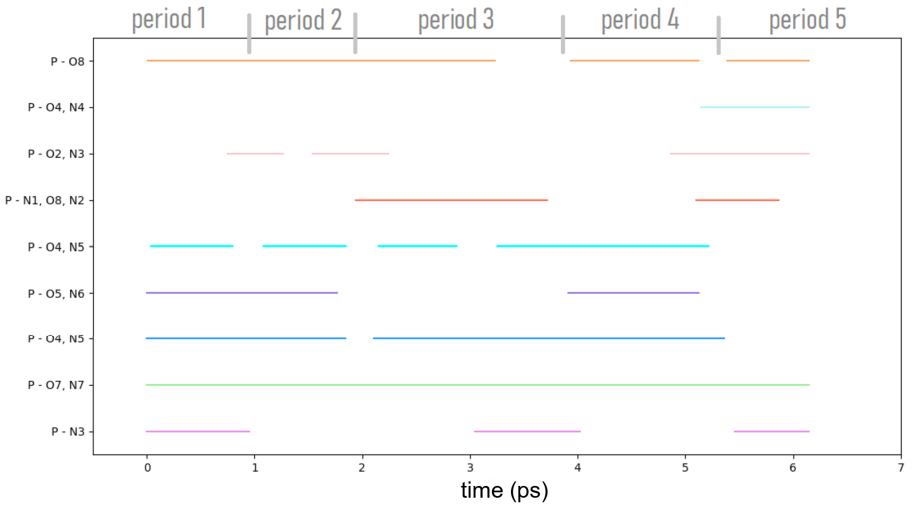


Figure 11. Chronogram obtained over the 450 K AIMD trajectory of the Z-Ala$_6$–COOH peptide. Five time periods have been identified (see text for details). The color coding of the lines is identical to the color coding of the vertices in the polygraph in Figure 10.

The reader has to keep in mind that each polycycle can have several identities in terms of the atoms that form the H-bonded cycle (see discussion above). Five distinct periods of time can be extracted, labeled as period 1 to period 5 in the figure. Each time period is associated to a different structural organisation of the nine polycycles in Z-Ala$_6$–COOH. During period 1, six polycyles coexist. Period 2 starts as the polycycle P-N3 disappears and polycycle P-O2,N3 appears. There is then the simultaneous disappearance of this polycycle at time ~2.2 ps and appearance of polycycle P-N1,O8,N2 at the slightly earlier time ~1.9 ps, which marks a clear change in the 3D structure of the peptide, and thus the end of period 2 and start of period 3. Interestingly, polycycles P-O7,N7 (green line), P-O4,N5 (dark blue) and P-O4,N5 (cyan) are formed over the whole 6 ps trajectory (or almost always, with smallish periods of time of interruption). These H-bonded polycycles thus form strong pillars for the global 3D structure. See more details in [4]. One has to remark that one of these three polycycles (P-O4,N5, darker blue) can adopt up to four different identities. This is an important pillar of the 3D structure with high flexibility.

Beyond the time evolution, one further crucial information extracted from the chronogram is which polycycles can be formed simultaneously and which ones cannot be present simultaneously in the 3D conformation of the peptide. For example, it is possible to form the polycycle P-O4,N4 (light blue) at the start of period 5 only if none of the P-O4,N5 (cyan) and P-O5,N6 (purple) polycycles are present. P-O4,N4 seems to coexist only with P-O2,N3 (light pink) and P-N1,O8,N2 (dark orange). Polycycles P-O4,N4, P-O2,N3, P-N1,O8,N2, P-O5,N6, and P-N3 need hence certain structural conditions for them to be exist. This can be explained by the location of H-bonded polycyle P-O4,N4 in regard to polycycles P-O4,N5 (both of them) and P-O5,N6 within the 3D structure of Z-Ala$_6$–COOH.

The polygraph made of nine vertices and their pairwise connected edges in Figure 10 gives the global/statistical view of the whole 6 ps trajectory in terms of a “global metastructure” of the peptide. However, the actual details of the dynamics are lacking in this global polygraph. This could be inferred from the chronogram in Figure 11 where we already saw that not all the polycycles/vertices of the global polygraph could coexist simultaneously over time.

Figure 12 now presents one illustration of the evolution with time of the sub-polygraphs from the global polygraph over a short time period, i.e., between times 2.19 ps and 2.51 ps (period 2 and period 3 in Figure 11). The sequence of sub-polygraphs is reported with the following conventions of colors: the polycycles that are present among the nine

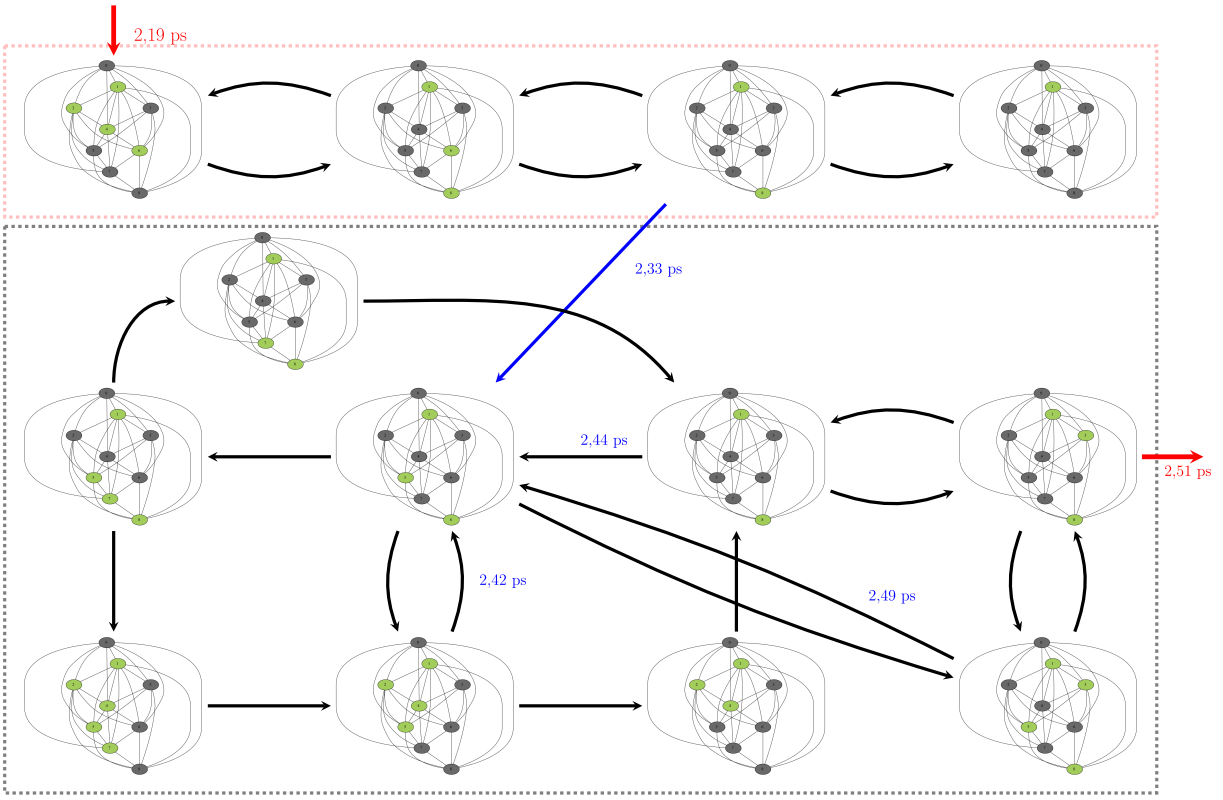


Figure 12. Sequence in time of the sub-polygraphs of the H-bonded polycycles of Z-Ala$_6$–COOH. See text for all details and comments of this chart.

possible polycycles are colored in green while the ones that are absent are colored in dark gray. The arrows between the sub-polygraphs can be colored in black, in blue, or in red. The two arrows in red mark the sequence of starting time of the period under scrutiny (2.19 ps) and ending time (2.51 ps), respectively. The arrow in black marks "simple transitions" between sub-polygraphs/3D-structures/conformations of Z-Ala$_6$–COOH. These arrows can be found forward (→) and backward (←), as there are multiple forward/backward isomerisations over time. Marking each one of these transitions over the "real time" would be too messy and would make the plot too cumbersome. The blue arrow marks a distinct transition in the sequence of sub-polygraphs, i.e., in the time sequence of the 3D conformations adopted by the peptide over this period of time. There is indeed no transition observed back to structures adopted previously (in time). Therefore, we marked the whole chart by two distinct rectangles, one outlined with red dots (top of the figure) and one with gray dots (bottom). In the gray rectangle, one can now see the sequence between the sub-polygraphs by following the dark arrows, some of them going in one direction only, others with back/forth directions. Some of the key moments in time are indicated in blue over certain dark arrows.

The conformation dynamics seen in this figure over a very short period of time nicely illustrates the complex dynamics in the H-bonds of Z-Ala$_6$–COOH at 450 K, and the high flexibility of the peptide that we have already discussed several times above. Figure 12 also nicely shows the fast exchange between polycycles.

3.2. *Topological 2D-MolGraphs easily capture complex condensed phase H-bonded networks*

Condensed phase, where liquid water is present, is another area where the 2D-topological graphs

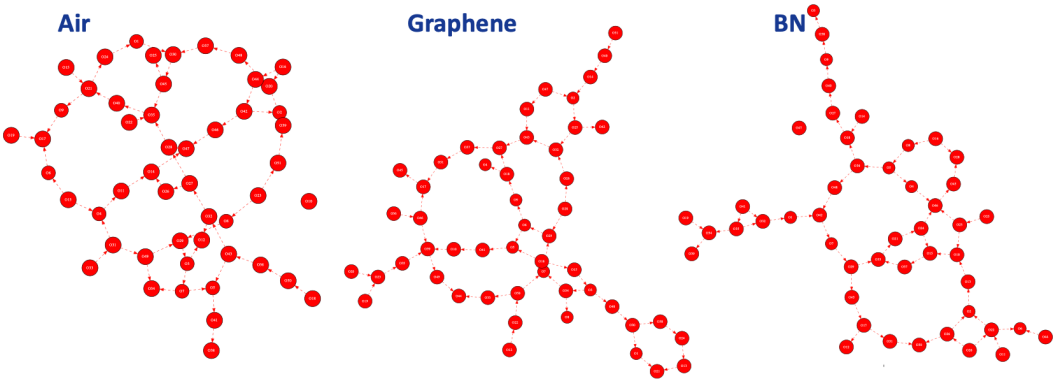


Figure 13. Illustration of one 2D-MolGraph per hydrophobic aqueous interface: air/liquid water (left), graphene/liquid water (middle), BN/liquid water (right). Only the water molecules located in the BIL [36,37] are taken into account for the 2D-MolGraph analysis. Vertices of the graph represent the oxygen atoms of the water molecules (red); the dashed red arcs represent the H-bonds between two water molecules oriented from donor to acceptor.

can be of great help in understanding H-bond networks. Here we present the use of the atomistic 2D-MolGraphs in order to unravel the structure of water in inhomogeneous molecular systems made of an interface between liquid water and another medium. We specifically focus on three hydrophobic aqueous interfaces, i.e., air/liquid water, graphene/liquid water, and boron nitride BN/liquid water, for which we want to characterize the organization of liquid water at the interface with air or the solid. To that end, our methodology of atomistic topological 2D-MolGraphs is applied on AIMD trajectories of these three aqueous interfaces (~50 ps time-length MD). These three interfaces have been shown to be hydrophobic by independent molecular analyses in [37] based on a molecular descriptor of hydrophobicity developed in this latter paper and in the follow-up paper [38].

We have previously shown in [10,37,39] that liquid water in contact with hydrophobic surfaces forms a two-dimensional (2D) highly collective H-bonded network made by the water molecules in the layer in direct contact with the hydrophobic surface (i.e., water located in the BIL–Binding Interfacial Layer– as defined in [36,37]), in which the water–water H-bonds are formed parallel to the surface. This water-collective 2D-Hbonded-Network is the molecular signature of surface hydrophobicity [37,38].

Here, we illustrate the recognition of this 2D-HBonded-Network using topological 2D-MolGraphs, and how these graphs can provide details on the organization of water molecules in this collective H-bond network. Three DFT-MD trajectories have been analyzed using our graph theory algorithm: air/liquid water as the prototype of hydrophobic surfaces, graphene/liquid water, and BN/liquid water. For each trajectory, 400 snapshots were extracted and analyzed (from a total of 50 ps trajectory per system). This corresponds to roughly one snapshot every 0.1 ps of dynamics, which represents a good statistical sampling regarding the dynamics of H-bonds. The graph analyses are carried out on the BIL-interfacial region only, in which there is roughly an average of 48 water molecules (all simulation boxes are roughly equivalent in sizes).

The 2D-MolGraphs in Figure 13 show the very specific structural property of the water molecules in the BIL of the hydrophobic interfaces: a collective arrangement of the water molecules in terms of H-bonded polygons (or rings) adjacent to each others. This collective arrangement is called the 2D-HBonded-Network [10,37,39].

To obtain the statistical view on the number of water molecules that are interconnected within the 2D-HBonded-Network, the (identified non-isomorphic) 2D-MolGraphs can be analyzed in terms of the size of the connected components, i.e., the set of subgraphs in which all vertices are connected to each others without interruption. Figure 14 illustrates the

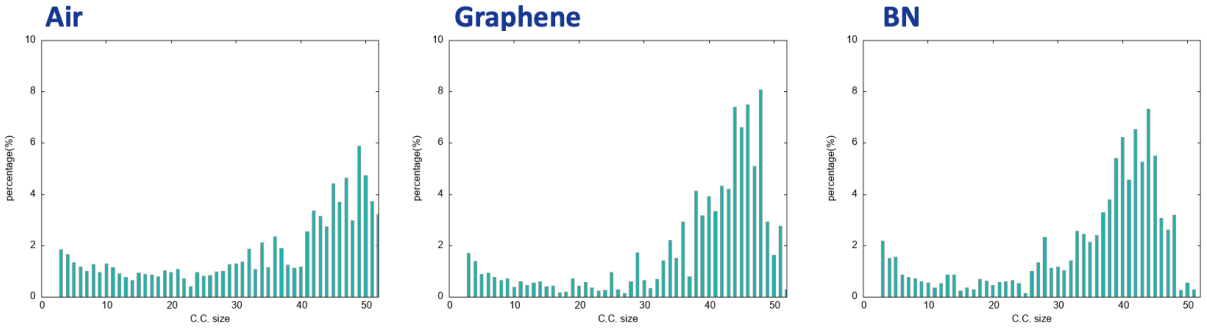


Figure 14. Distribution of the connected components of the 2D-MolGraphs (see text for details) for the air/liquid water interface (left), the graphene/liquid water (middle), and the BN/liquid water (right), obtained over ~50 ps AIMD trajectories.

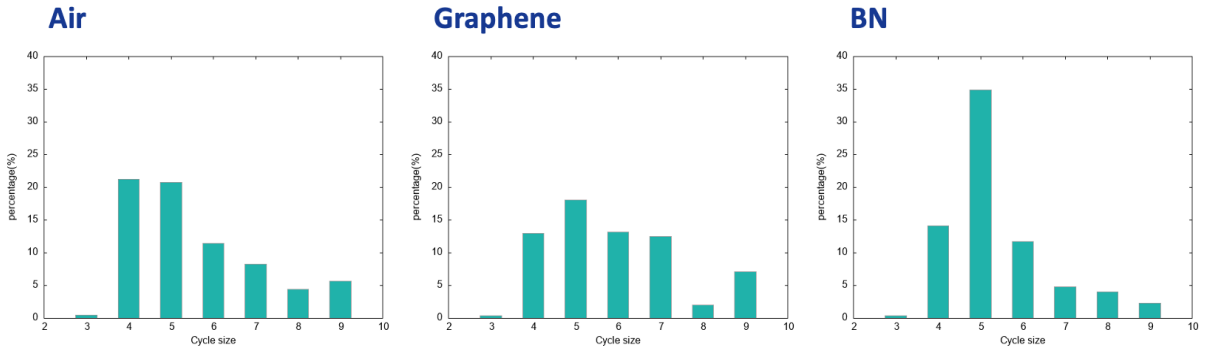


Figure 15. Distribution of the size of H-bonded rings/cycles formed by the water molecules in the 2D-MolGraphs. (left): air/water interface; (middle): graphene/water interface; (right): BN/water interface.

distribution of the connected components for the 400 2D-MolGraphs extracted for each trajectory of the three interfaces.

One can immediately see that the 2D-HBonded-Network is extended over ~90–95% of the water molecules located in the BIL for the three interfaces, hence with a high degree of interconnectivity (very collective network between the water molecules). One can thus conclude that the water molecules in the BIL are statistically organized with the same collective HB-Network in all these hydrophobic interfaces.

The 2D-MolGraphs provide further details on the organization of water molecules in this collective H-bonded network. One can see from the 2D-MolGraphs shown in Figure 13 that water molecules are organized in polygons/rings formed by H-bonds. Using the Horton algorithm [40], we analyzed the 2D-MolGraphs in terms of the size of H-bonded polygons/rings formed by the water molecules in the BIL for the three hydrophobic interfaces. The results are presented in Figure 15.

Very interestingly, though the water molecules in the three investigated BILs are assembled with the same collective 2D-Hbonded-Network, the distribution of sizes of the H-bonded polygons that build these networks are non-identical between the three hydrophobic interfaces. On the one hand, the sizes of the H-bonded polygons are centered on four to six for the air/water and BN/water interfaces. There is a clear dominant component related to H-bonded pentagons made by the water molecules at the interface with the BN surface while the formation of H-bonded tetragons and pentagons is found equivalent at the interface with air. On the other hand, the 2D-Hbonded-Network made by the water molecules at the surface of graphene is more homogeneous in terms of sizes of the polygons, where tetragons,

hexagons, and heptagons have roughly the same probability of appearance, and H-bonded pentagons dominate slightly more. We hence see that water molecules predominantly form five-membered H-bonded rings/polygons at the interface with the BN surface, which can be associated to the hexagonal-templated structure of BN. The length of the C–C covalent bonds in BN is shorter than the O···H hydrogen bonds: the best arrangement for water molecules is thus into five-membered H-bonded rings rather than six-membered rings.

Moreover, we find that the percentage of water molecules giving rise to the polygons within the 2D-HBonded-Network is around 30–40% for the three interfaces, with the following interesting ranking: one finds a larger percentage of water in the 2D polygons for the air/water interface (~43%) than for the graphene/water interface (~36%) interface, and the BN/water interface (~34%). Such percentages might explain the strength of the 2D-HBonded-Network found at each interface. The work in [37] indeed showed that the strength of the 2D-HBonded-Network can be ranked as air > graphene > BN. In other words, the more water molecules forming rings within the 2D-Hbonded-Network (i.e., the more rings being formed), the stronger the 2D-HBonded-Network, and the more hydrophobic the interface.

3.3. *Integration of 2D-MolGraphs in workflows in high-throughput in silico chemical reactivity*

We further explored the utility of our 2D-MolGraph approach as the core for automated reaction network analysis workflows suitable for implementation in high-throughput in silico reactivity screening of complex multicomponent homogeneous catalytic systems [41]. We implemented the 2D-MolGraph approach in the computational catalytic reaction space exploration method ReNeGate [5] (see the workflow in Figure 16) and in the high-throughput reactivity screening HiREX workflow [11] specifically designed to explore realistic catalytic systems and identify thermodynamically feasible chemical transformations, corresponding to secondary catalyst deactivation and inhibition paths.

The method was validated by case studies on representative multicomponent (de)hydrogenation catalytic systems based on Mn(I) coordination complexes with a special focus on probing unconventional and less expected reaction channels, which could be responsible for loss of catalytically potent species during the initial pre-catalyst activation. An illustrative example is our computational analysis of the activation of manganese pentacarbonyl bromide ($Mn(CO)_5Br$) with inorganic alkoxide base KOiPr that is a common protocol for the experimental screening and in situ generation of Mn-based homogeneous catalysts [42]. The reactivity exploration was carried out with parallel metadynamics simulations on a minimal model constituted by the two main reagents only (Figure 17a) using the CREST functionality (metaD/CREST) at the GFN2-xTB level [43]. The reactive trajectories were populated using the root-mean-square-deviation (RMSD) in Cartesian space as a metric for the collective variables [43], while the pushing and pulling strengths (k and α) were systematically varied over the parallel simulations.

The analysis of the reactive trajectories with the 2D-MolGraphs yielded a reaction network containing 12 conformers, which after trimming the edges exceeding an arbitrary threshold of 25 kcal/mol, produced the reaction network shown in Figure 17b. State (1) corresponds to the starting configuration with unreacted $Mn(CO)_5Br$ and KOiPr, which can transform to one of the new species identified by the 2D-MolGraphs from the reactive trajectories (Figure 17c).

Our automated procedure revealed that all reaction paths involve the reaction of the alkoxide nucleophile with the Mn(I)-bound carbonyl ligand to form a Mn–acyl complex (2). Due to its approximate nature, the GFN2-xTB method incorrectly predicts further migratory insertion of CO with the –C(O)OiPr species yielding species (4) and (8) to be also thermodynamically favorable. Subsequent energy refinement at the DFT level restores the agreement with the experimental observations. Similar reaction paths were identified for more complex catalytic model containing the molecularly defined Mn(I)-catalyst stabilized by a bidentate diamino ligand, the alkoxide base, two isopropanol solvent molecules and acetophenone as a model substrate. Simulations suggested that the nucleophilic attack of the alkoxide anion by the Mn-bound carbonyl ligand may initiate reaction paths resulting in a (partial) decoordination of the organic ligand, which can be considered the onset of catalyst deactivation [5].

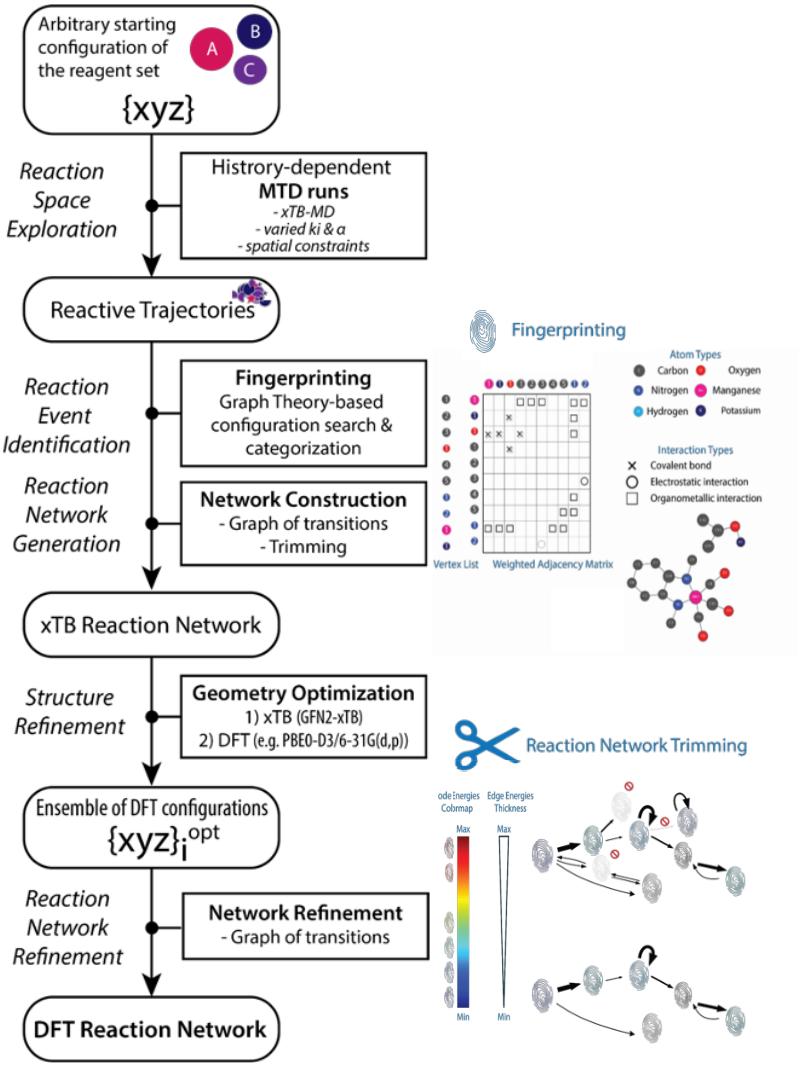


Figure 16. Schematic representation of the ReNeGate workflow from [5] involving the sequential reactive space exploration, the structure analysis using the 2D-MolGraphs in the fingerprinting and network construction parts, with the reaction network generation and refinement steps (trimming).

These computational insights have inspired the experimental finding on the stabilizing role of Lewis acid additives mediating the nucleophilicity of the alkoxide base allowing to considerably extend the lifetime of the homogeneous Mn(I) catalysts [44] and the discovery of new Mn-mediated C–C coupling chemistry [45].

The application of our reaction network analysis approach was further extended towards high-throughput computational reactivity exploration and automation identification of the classes of reactivity patterns within specific catalyst groups (Figure 18a) [11]. We have applied this workflow to a virtual library containing 576 Mn-pincer complexes corresponding to four distinct pincer families with varied functionalization of the ligand backbone (R1 and R2 functionalities) and Mn coordination (X) (Figure 18b). The 2D-MolGraphs were used to analyze the reactive trajectories, as well as to featurize and label the discovered new configurations and intermediates following the changes in the interaction patterns observed during the transformations.

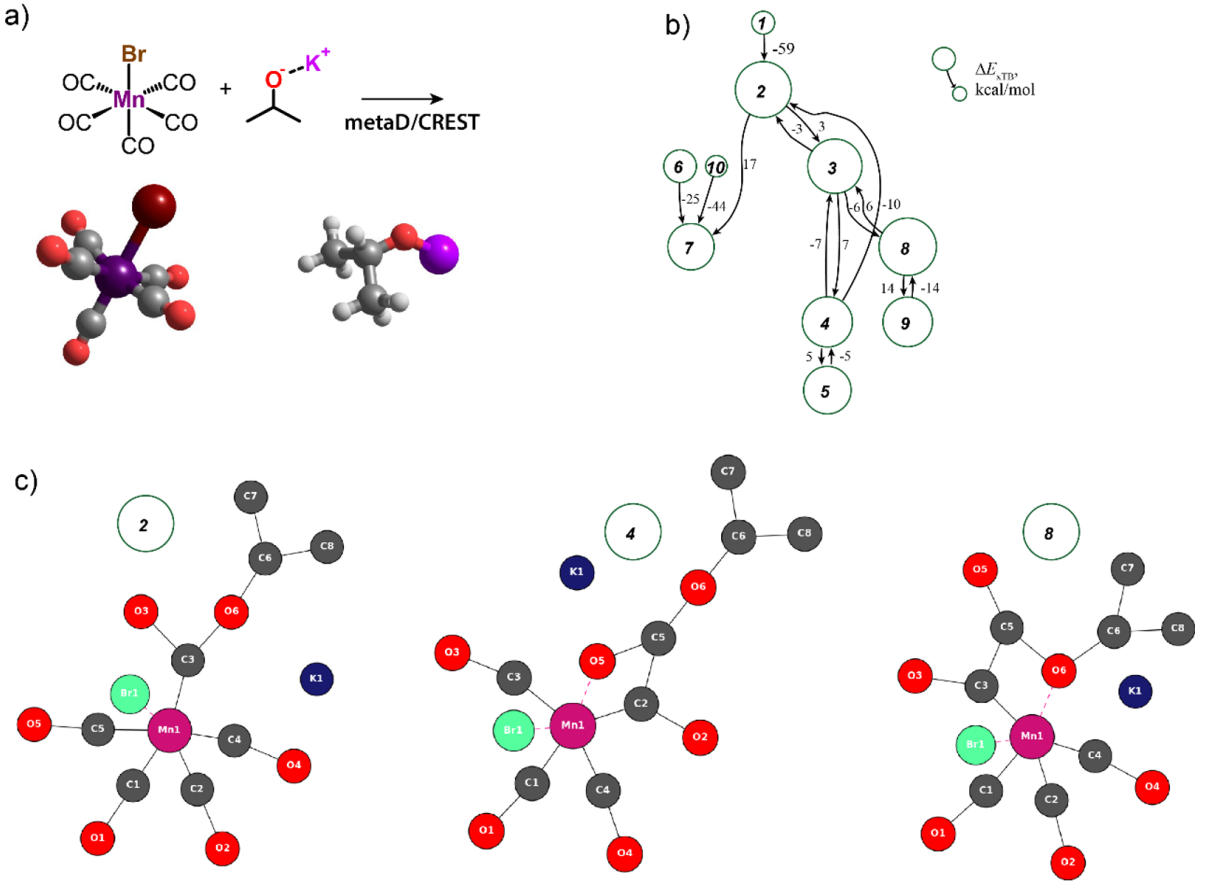


Figure 17. (a) Metadynamics simulations of the reaction between Mn(CO)5Br and KOiPr yield reactive trajectories, from which (b) a reaction network has been constructed with the respective three representative 2D-MolGraphs presented in (c). Colors for the 2D-MolGraph vertices: dark gray for carbons, red for oxygens, dark blue for postassium, pink for the manganese, and green for the bromide. Solid black edges in the graphs represent covalent bonds, the red dashed edges represent the organometallic interactions.

K-mode clustering analysis on the resulting labeled database (Figure 18b) has provided new insights into the reactivity of Mn(I) pincers and how it is affected by the structural modification of the ligand backbone.

Specifically, the calculations indeed revealed multiple paths involving the nucleophilic attack on the carbonyl ligand and decoordination of the pincer ligand. Depending on the ligand's nature and structure, the thermodynamic favorability of such secondary conversion paths varies greatly providing thus an opportunity to tune the stability and reactivity of the transition metal catalyst, and guide the exploration of new chemical conversion paths.

4. Prospects and new developments in progress

With this review, we believe that the topological 2D-MolGraphs and associated graph algorithms have been shown to be powerful tools for analyzing atomistic molecular dynamics trajectories and extracting the actual conformations sampled over time. Demonstrations were carried out on gas phase flexible molecules and inhomogeneous aqueous interfaces in the condensed phase. We have also shown the relevance of different levels of granularity to be used in the topological graphs. In particular, the coarse-grained approach of graphs of polymorphic H-bonded cycles was shown decisive in order to

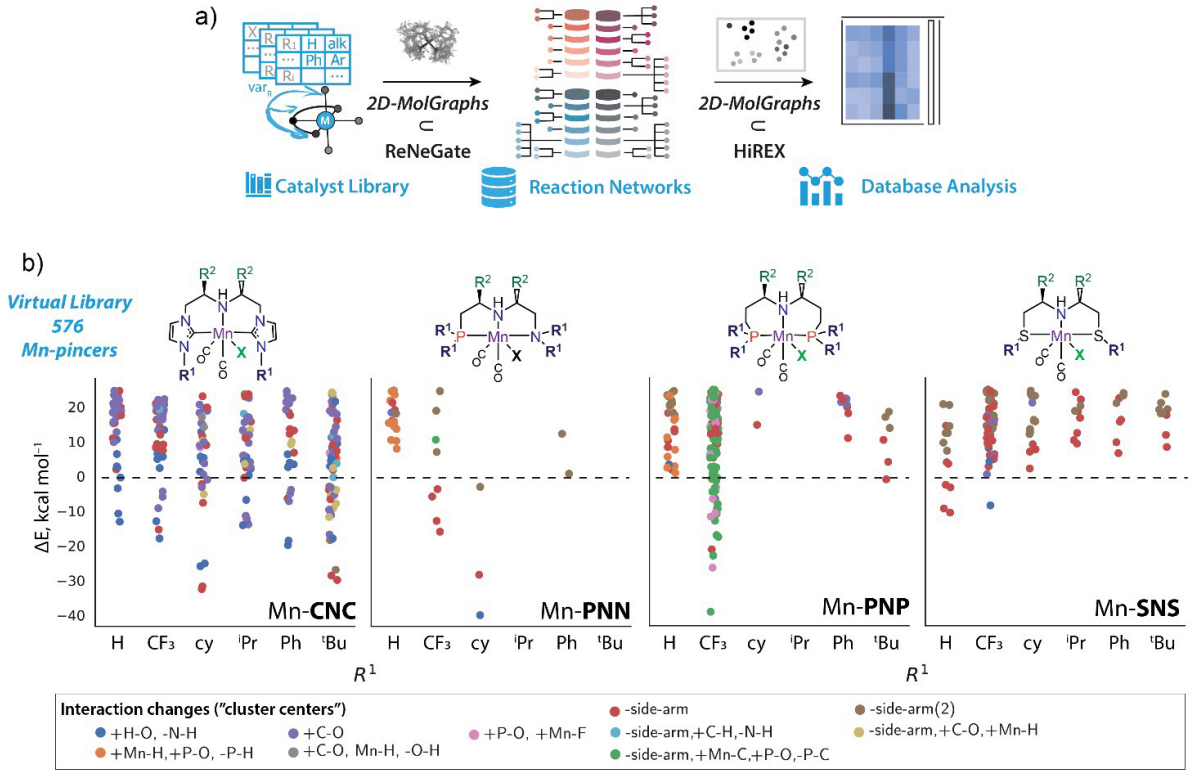


Figure 18. (a) A perspective workflow for the chemical space exploration and automated reactivity analysis of virtual catalyst libraries involving 2D-MolGraphs as the core component of the dynamic trajectory analysis and structure featurization. The method was used for (b) the reactivity screening of a virtual library of 576 Mn(I) pincer complexes followed by K-mode clustering analysis of the resulting data. The figure presents top 10 clusters showing the most frequent types of interaction changes as a function of different functionalization at R1 site for each pincer class. The color coding is given at the bottom of the figure.

rationalize the dynamics of a flexible hexapeptide. The "metastructures" over which the dynamics of this peptide is built could be found only at this level of topology representation.

We have also shown that the 2D-MolGraph approach can be easily coupled to global workflows that include several theoretical methods to sample conformational and reactive chemical spaces. The topological graphs were inserted in high-throughput in silico chemical reactive workflows in homogeneous catalysis. The conformational fingerprinting provided by the 2D-MolGraphs was decisive in several steps of these workflows. The outcome of these reactive workflows would not have been as easy and successful without the topological graphs. These works continue.

Ongoing works and developments also include for instance the use of the 2D-MolGraphs to extract the knowledge of structural motifs at aqueous interfaces. For instance, it is of utmost importance to reveal the motifs formed between the sites at the surface of a solid (for instance silica oxide as in [46,47]) and the water molecules located at the interface with the solid in the Binding Interfacial Layer (BIL). These motifs are not only responsible for the spectroscopic signatures recorded at aqueous interfaces, typically by SFG (Sum Frequency Generation) spectroscopies, they are also involved in the chemical reactivity of these aqueous interfaces. Motifs and their vibrational fingerprints have already been discussed in [47] for aqueous silica interfaces. We are developing algorithms that

can automatically recognize and classify these motifs from the 2D-MolGraphs, for instance in terms of the sizes of the H-bonded cycles formed between surface sites and water molecules and in terms of their statistical distribution in space within the BIL. The same algorithms will also be applied to the water molecules in the BIL of biomolecules. This is also ongoing work, using the atomistic 2D-MolGraphs and the coarse-grained polygraphs of H-bonded cycles.

Naturally, databases of 2D-MolGraphs and coarse-grained polygraphs can be built up and connected to AI (Artifical Intelligence) and machine learning techniques. This is where our next step will take us.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

[1] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, M.-P. Gaigeot, *J. Chem. Phys.*, 2018, **149**, article no. 184102.

[2] S. Bougueroua, M. Bricage, Y. Aboulfath, D. Barth, M.-P. Gaigeot, *Molecules*, 2023, **28**, article no. 2892.

[3] S. Bougueroua, Y. Aboulfath, D. Barth, M.-P. Gaigeot, *Mol. Phys.*, 2023, **121**, article no. e2162456.

[4] Y. Aboulfath, S. Bougueroua, A. Cimas, D. Barth, M.-P. Gaigeot, *J. Chem. Theor. Comput.*, 2024, **20**, 1019-1035.

[5] A. Hashemi, S. Bougueroua, M.-P. Gaigeot, E. A. Pidko, *J. Chem. Theor. Comput.*, 2022, **18**, 7470-7482.

[6] S. N. Ilemo, D. Barth, O. David, F. Quessette, M.-A. Weisser, D. Watel, *PLoS One*, 2019, **14**, article no. e0226680.

[7] C. Gianfrotta, V. Reinharz, D. Barth, A. Denise, *Proceedings of the 19th International Symposium on Experimental Algorithms (SEA 2021), June 2021, Nice (virtual), France*, 2021.

[8] D. Barth, O. David, F. Quessette, V. Reinhard, Y. Strozecki, S. Vial, in *Experimental Algorithms. SEA 2015* (E. Bampis, ed.), Lecture Notes in Computer Science, vol. 9125, Springer, Cham, 2015.

[9] A. Lamiable, F. Quessette, S. Vial, D. Barth, A. Denise, *IEEE ACM Trans. Comput. Biol. Bioinform.*, 2013, **10**, 193-199.

[10] A. Serva, S. Pezzotti, S. Bougueroua, D. R. Galimberti, M.-P. Gaigeot, *J. Mol. Struct.*, 2018, **1165**, 71-78.

[11] A. Hashemi, S. Bougueroua, M.-P. Gaigeot, E. A. Pidko, *J. Chem. Inf. Model.*, 2023, **63**, 6081-6094.

[12] S. Bougueroua, F. Quessette, D. Barth, M.-P. Gaigeot, "GaTewAY : Graph theory-based software for automatic analysis of molecular conformers generated over time", 2022, ChemRxiv, https://doi.org/10.26434/chemrxiv-2022-1d5x8.

[13] B. L. Mooney, L. R. Corrales, A. E. Clark, *J. Comput. Chem.*, 2012, **33**, 853-860.

[14] A. Ozkanlar, A. E. Clark, *J. Comput. Chem.*, 2014, **35**, 495-505.

[15] K. Han, R. M. Venable, A.-M. Bryant *et al.*, *J. Phys. Chem. B*, 2018, **122**, 1484-1494.

[16] C. M. Tenney, R. T. Cygan, *J. Phys. Chem. C.*, 2013, **117**, 24673-24684.

[17] F. Pietrucci, W. Andreoni, *Phys. Rev. Lett.*, 2011, **107**, article no. 085504.

[18] M. Hudelson, B. L. Mooney, A. E. Clark, *J. Math. Chem.*, 2012, **50**, 2342-2350.

[19] F. Pietrucci, W. Andreoni, *J. Chem. Theory. Comput.*, 2014, **10**, 913-917.

[20] E. Martínez-Núñez, *Phys. Chem. Chem. Phys.*, 2015, **17**, 14912-14921.

[21] E. Martínez-Núñez, *J. Comput. Chem.*, 2015, **36**, 222-234.

[22] A. Jindal, V. Arunachalam, S. Vasudevan, *J. Phys. Chem. B.*, 2021, **125**, 5909-5919.

[23] B. D. McKay, *Congressus Numerantium*, vol. 30, Department of Computer Science, Vanderbilt University Tennessee, US, 1981, Chapter 2, 47-87 pages.

[24] E. M. Luks, *J. Comput. Syst. Sci.*, 1982, **25**, 42-65.

[25] B. D. McKay, A. Piperno, *J. Symb. Comput.*, 2014, **60**, 94-112.

[26] S. G. Hartke, A. Radcliffe, *Communicating Mathematics*, vol. 479, Minnesota, American Mathematical Society, 2009, Chapter 8, 99-111 pages.

[27] A. Casteigts, K. Meeks, G. B. Mertzios, R. Niedermeier, *Temporal Graphs: Structure, Algorithms, Applications. Report from Dagstuhl Seminar, April 25–30*, 2021, http://www.dagstuhl.de/21171.

[28] P. Liu, A. E. Sariyüce, *KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, 1501-1511 pages.

[29] P. Plamper, O. J. Lechtenfeld, P. Herzsprung, *Environ. Sci. Technol.*, 2023, **57**, 18116-18126.

[30] F. Berger, P. Gritzmann, S. de Vries, *Networks*, 2017, **70**, 116-131.

[31] P. M. Gleiss, P. F. Stadler, A. Wagner, D. A. Fell, *Adv. Complex Syst.*, 2001, **04**, 207-226.

[32] S. N. Ilemo, D. Barth, O. David, F. Quessette, M.-A. Weisser, D. Watel, *PLoS One*, 2019, **14**, article no. e0226680.

[33] J. D. Horton, *SIAM J. Comput.*, 1987, **16**, 358-366.

[34] D. Babi, A. Graovac, B. Mohar, T. Pisanski, *Discrete Appl. Math.*, 1986, **15**, 11-24.

[35] M. Juvana, B. Moha, *Discrete Appl. Math.*, 1997, **80**, 57-71.

[36] S. Pezzotti, D. Galimberti, Y. Shen, M.-P. Gaigeot, *J. Phys. Chem. Chem. Phys.*, 2018, **20**, 5190-5199.

[37] S. Pezzotti, A. Serva, F. Sebastiani *et al.*, *J. Phys. Chem. Lett.*, 2021, **12**, 3827-3836.

[38] W. Chen, S. E. Sanders, B. Ozdamar, D. Louaas, F. S. Brigiano, S. Pezzotti, P. B. Petersen, M.-P. Gaigeot, *J. Phys. Chem. Lett.*, 2023, **14**, 1301-1309.

[39] S. Pezzotti, D. R. Galimberti, M.-P. Gaigeot, *J. Phys. Chem. Lett.*, 2017, **8**, 3133-3141.

[40] J. D. Horton, *SIAM J. Comput.*, 1987, **16**, 358-366.

[41] A. V. Kalikadien, A. Mirza, A. N. Hossaini, A. Sreenithya, E. A. Pidko, *ChemPlusChem*, 2024, **89**, article no. e202300702.

[42] W. Yang, G. A. Filonenko, E. A. Pidko, *Chem. Commun.*, 2023, **59**, 1757-1768.

[43] S. Grimme, *J. Chem. Theor. Comput.*, 2019, **15**, 2847-2862.
[44] R. van Putten, "Catalysis, chemistry, and automation: Addressing complexity to explore practical limits of homogeneous Mn catalysis", Dissertation, TU Delft, The Netherlands, 2021.
[45] R. Van Putten, G. A. Filonenko, A. M. Krieger, M. Lutz, E. A. Pidko, *Organometallics*, 2021, **40**, 674-681.
[46] M. Sulpizi, M.-P. Gaigeot, M. Sprik, *J. Chem. Theor. Comput.*, 2012, **8**, 1037-1047.
[47] S. Pezzotti, D. R. Galimberti, M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2019, **21**, 22188-22202.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 53-65
https://doi.org/10.5802/crchim.314

Research article

French Network on Solvation (GDR 2035 SolvATE)

Hydrogen bond interactions of coumarin-153 in molecular solvents: molecular dynamics and principal component analysis

Kateryna Goloviznina [a], Dmytro Dudariev [b,c], François-Alexandre Miannay [b], Oleg Kalugin [c], Volodymyr Koverga [d,e], Toshiyuki Takamuku [f], Raffaele Vitale [a] and Abdenacer Idrissi [*,b]

[a] Sorbonne Université, CNRS, Physicochimie des Électrolytes et Nanosystemes Interfaciaux, PHENIX, F-75005 Paris, France

[b] Univ. Lille, CNRS, UMR 8516 - LASIRe - Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, F-59000 Lille, France

[c] Department of Inorganic Chemistry, V.N. Karazin Kharkiv National University, Svobody sq. 4, Kharkiv, 61022, Ukarine

[d] Department of Chemical Engineering, University of Illinois Chicago, Chicago, IL, 60608, USA

[e] Materials Science Division, Argonne National Laboratory, Lemont, IL, 60439, USA

[f] Department of Chemistry and Applied Chemistry, Faculty of Science and Engineering, Saga University, Honjo-machi, Saga 840-8502, Japan

E-mail: nacer.idrissi@univ-lille.fr (A. Idrissi)

Abstract. Hydrogen bond interactions significantly affect the coumarin-153's (C153) photophysics, including its ability to act as a donor of weak hydrogen bonds via its 14 C–H bonds and as an acceptor via its O atoms in the ester and the carbonyl groups, as well as via its F atom in the trifluoromethyl group. The distances between the donor atoms and their closest electronegative neighbor atom served as descriptors of the hydrogen bond interactions. These descriptors were calculated using the nearest neighbor radial distribution approach. Principal component analysis (PCA) was then performed on these distances to compare the unique structures surrounding donor bond atoms and identify patterns in the interactions between C153 and various solvent, such as acetonitrile, butyrolactone, propylene carbonate, methanol, ethanol, propanol, and butanol.

Our findings demonstrate that, when C153 acts as a hydrogen bond donor, the interaction behavior of the H atoms that are close to the N atom and that of the H atom close to the trifluoromethyl F atom of C153 is substantially different. More specifically, the former H atoms interact preferentially with the hydroxyl oxygen atom of the solvent while the H_{10} atom interacts preferentially with the ester oxygen atoms of propylene carbonate.

Moreover, when C153 behaves as a hydrogen bond acceptor, PCA shows that the carbonyl O atom of C153 interacts preferentially with the hydroxyl H atom of the alcohols, while the F atoms mostly interact with the other ethyl and methyl H atoms of the solvent.

* Corresponding author

Keywords. Coumarin-153, Solvation, Molecular dynamics, Photophysics, Hydrogen bond interactions, Nearest neighbor radial distribution, Principal component analysis (PCA).

Funding. French National Agency for Research (ANR-19-CE05-0009-01).

Manuscript received 25 December 2023, revised 7 March 2024, accepted 25 April 2024.

1. Introduction

Solvent environment and, particularly, the presence of hydrogen bonds-donating or -accepting interactions as well as the occurrence of stacking and dipole–dipole interactions are expected to have a significant impact on the photophysics of fluorophores like coumarin 153 (C153) [1–4]. For this reason, the effect of the solvent on the photophysics of C153 has been explored in a range of solvents, including alcohols [5–7], acetonitrile [2,8–10], propylene carbonate [6], dimethyl sulfoxide, formamide, nitromethane, acetone [2], methanol, ethanol and fluorinated ethanol solvents [11], in mixtures such as dioxane–water [12], acetonitrile–benzene [13], hexane–propionitrile [12], tert-butyl alcohol–water and trimethylamine N-oxide–water [14], acetonitrile–propylene carbonate [15] dimethyl sulfoxide–glycerol [16,17], toluene–acetonitrile [18], and cyclohexane–phenol solvent [19]. These studies were extended to aqueous and non-aqueous reverse micelles [20], Triton X-100-cyclohexane microemulsions [21], and also to various ionic liquids, such as 1-butyl-3-methylimidazolium hexafluorophosphate [22], 1-hexyl-3-methylimidazolium hexafluorophosphate [8], 1-butyl-3-methylimidazolium tetrafluoroborate [10], hydroxylfunctionalized ionic liquids [23], ionic liquids containing fluoroalkylphosphate [24] and tetraalkylammonium bromide [25], 1-dodecyl-3-methylimidazolium bis(trifluoromethylsulfonyl)amide with benzene, chloroform, propylene carbonate solvents [26], and deep eutectic solvents such as choline chloride [27] and acetamide–urea [28]. One should notice that the properties of the mixtures are modulated through the change of the mixture composition. The aforementioned works showed that the strength and nature of the hydrogen bonds can influence the electronic transitions of C153, resulting in either fluorescence quenching or enhancement. Indeed, the efficiency of fluorescence, expressed in terms of quantum yield, can be affected by hydrogen bond interactions as they may promote non-radiative decay processes and reduce the quantum yield. It is therefore essential to take into account the specific molecular structure of C153 (namely, the donor or acceptor nature of its atoms) and the characteristics of the solvent in experimental studies to understand how hydrogen bond interactions influence C153 photophysics in a given environment. This can be illustrated through the following examples. First, Maronceli et al. showed that the presence of hydrogen bonds between C153 and alcohols results (beyond the effects of solvent polarity) in a small additional shift in both the absorption and emission spectra of C153. However, when examining the solvation dynamics of C153 in N-methylpropionamide, a solvent with hydrogen bond donor ability, distinct differences in dynamics were observed compared to the case when alcohol solvents were considered [6]. On the other hand, in propylene carbonate, a non-associated solvent lacking hydrogen bond donor capability, the behavior of the solvation function closely resembles that observed in alcohols. Other studies, instead, showed that coumarin 102 and C153 have different photophysical behaviors in the same solvent and that these different behaviors are due to their methyl CH_3 and trifluoryl CF_3 substituents, respectively [29,30]. Computer molecular modeling could corroborate these experimental results, since simulation approaches can provide deep atomistic insights into the local structure of the solvent around the donor and acceptor regions of C153, insights that cannot be obtained by any kind of experimental strategy. As a matter of fact, the microscopic environment surrounding the C153 molecule was also investigated through quantum-chemical calculations and molecular dynamics (MD) simulations [11,31], which permitted to analyze interatomic and intermolecular radial distribution functions (RDF)—considering either the center-of-masses or specific atoms like O, F, N, and C—without explicitly accounting for the hydrogen atoms of C153 and the solvent molecules. Notably, these investigations were conducted for both the ground state (GS) and excited state (ES) of C153. Based on molecular dynamics simulations of C153 in 1,4-dioxane, Cinacchi et al. [32] suggested that changes in the solvation shell around the GS and ES of C153 are primarily due to alterations in the orientation of the solvent molecules. In other MD studies,

the local structure around the hydrogen atoms of C153 has been explored [26,31]. It has been observed that C153 is solvated by alcohols or water through hydrogen bonds, specifically between the hydroxyl OH group of the solvents and the carbonyl oxygen atom of C153. Furthermore, MD simulations of C153 in dimethyl sulfoxide (DMSO)–water mixtures have shown that the hydration structures in GS and ES differ from those in pure water due to the significant influence of DMSO molecules [27]. This was attributed to the formation of a hydrogen bond network between DMSO and water molecules upon mixing. In the case of imidazolium ionic liquids (ILs), MD studies of C153 in 1-ethyl-3-methylimidazolium tetrafluoroborate ($EmimBF_4$) [22] demonstrated a preferential solvation of C153 by the $Emim^+$ cation. In a study by Barman [33], MD simulations, quantum-chemical calculations and infrared (IR) spectroscopy were combined to investigate the formation of hydrogen bonds between aniline and C153. It was concluded that the carbonyl group of C153 serves as a primary hydrogen bond-accepting site and shows a higher bond strength in the ES. Furthermore, the formation of a C=O···H–N hydrogen bond was confirmed by the observation of the IR absorption band at 1736 cm^{-1} in the IR spectrum of C153 in the aniline–cyclohexane mixture, shifted compared to the stretching band of the C=O group (1748 cm^{-1}) which is commonly measured in pure cyclohexane. Similar experiments were conducted for the coumarin C102–aniline system [34], and the formation of the stronger hydrogen bond for C153 in the ES was also observed in methanol solutions [35]. Furthermore, the solvation of C153 in ionic 1-butyl-3-methylimidazolium tetrafluoroborate–propylene carbonate mixtures was investigated by coupling MD simulations, time-correlated photon counting and fluorescence up-conversion techniques. It was established that the solvation of C153 is determined by its interaction with the ions at high IL content, and with the solvent molecules at a IL molar fraction lower than 0.2 [36]. Xu et al. [37] conducted B3LYP/TZVP calculations to thoroughly explore the formation of hydrogen bonds between C153 and ethanol in both GS and ES. They observed that the hydrogen bond C=O···H–O is strengthened in the electronic ES, i.e., the bond length decreases from 1.867 Å in GS to 1.813 Å in ES, as also shown by Han and coworkers [38], This observation suggests that the hydrogen bond O···H–O in the C153–EtOH complex in the ES must be strengthened. This suggestion was also substantiated by IR experiments conducted in this study [38].

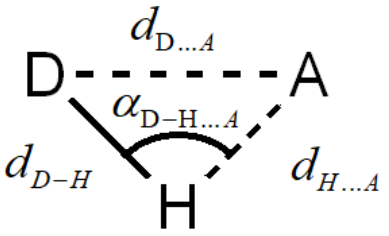


Figure 1. Definition of the distances describing the hydrogen bond interactions between a generic donor D–H and a generic electronegative acceptor A. The distances $d_{H...A}$ and $d_{D...A}$ are calculated using molecular dynamic simulations relying on the nearest neighbor approach.

In order to retrieve essential information about hydrogen bond interactions, geometric data [39,40], spectroscopic data [41–47], physical chemical data [48], and MD simulation data [49–53] have been usually analyzed by means of principal components analysis (PCA). PCA can provide a comprehensive overview of hydrogen bond interactions offering insights into dominant spectral features, their variations, global trends, and correlations, and aiding in the identification and understanding of hydrogen bond-related patterns in the collected data.

This article provides a thorough study of the hydrogen bond interactions between C153 and various solvents, including methanol, ethanol, 1-propanol, 1-butanol, acetonitrile, γ-butyrolactone, and propylene carbonate. These solvents exhibit distinct properties such as dipole moment, viscosity, dielectric constant, density, and the capacity to form hydrogen bonds, as well as stacking interactions and dipole–dipole interactions. In this study, we combine MD simulations with PCA. Atomistic simulations provide the coordinates of atoms that are involved in the hydrogen bond interactions. The generic hydrogen bond between a donor D–H and acceptor atom A is described by the configuration given in Figure 1.

The D–H and A moieties may belong to any of the mixture's constituents. The intramolecular distance $d_{D–H}$ is assumed to remain constant in our simulations, while $d_{H...A}$ and $d_{D...A}$ are determined as the average distances of the nearest neighbor radial distributions of the first neighbor electronegative atom A

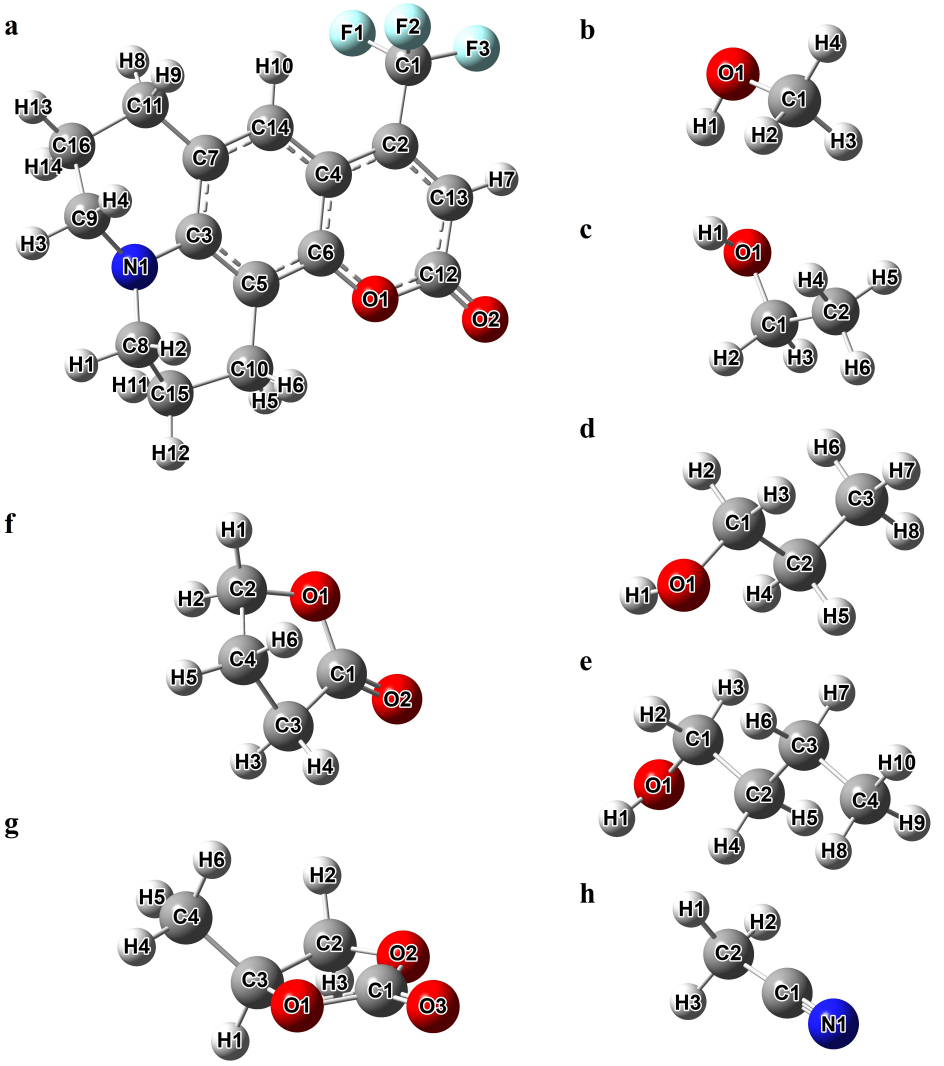


Figure 2. Atomic labels for Coumarin-153 (a) and the investigated solvents ((b) methanol, (c) ethanol, (d) 1-propanol, (e) 1-butanol, (f) γ-butyrolactone, (g) propylene carbonate, (h) acetonitrile).

with respect to the H and D, respectively. It is worth noting that once $d_{\mathrm{D–H}}$, $d_{\mathrm{H...A}}$ and $d_{\mathrm{D...A}}$ are determined, the angle α can be calculated using the law of cosines (see Figure 1).

The average values of distances $d_{\mathrm{H...A}}$ and $d_{\mathrm{D...A}}$ were calculated for both GS and ES of C153 and serve as indicators of the strength of the hydrogen bond interactions between C153 and the solvent molecules. Importantly, the GS and ES of C153 were simulated by considering their corresponding charge distributions, which were determined through quantum calculations (see Figure 3). These distances were calculated for cases where C153 acted as either a hydrogen bond donor or acceptor. In the former scenario, the 14 C–H bonds of C153 were considered, and these distances were calculated from electronegative atoms such as the N_1 atom of acetonitrile, the O_1 atoms of the alcohol solvents, and the O_1, O_2, and O_3 of the other solvents. In the latter case, these distances were calculated between each of the 39 C–H bonds of the solvent molecules and the 6 electronegative atoms of C153, i.e., the N_1, $O_{1–2}$ and $F_{1–3}$ atoms (see Figure 2 for more details on the numbering of the atoms).

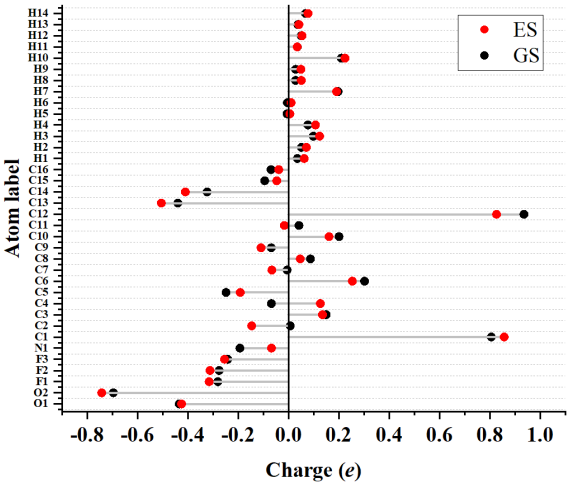


Figure 3. Charge values associated to the atoms of C153 in the ground state and excited state as determined from quantum-chemical calculations performed with B3LYP/6-311+G(d).

PCA was then applied to analyze the two distance matrices obtained in the two aforementioned cases. The first matrix contained 10 rows corresponding to the 10 electronegative atoms of the different solvents under study and 28 columns corresponding to the 14 C–H bonds of C153 in both GS and ES. The second matrix contained 43 rows corresponding to the 39 C–H and the 4 O–H bonds of the different solvents under study and 12 columns corresponding to the 6 electronegative atoms of C153 in both GS and ES. PCA helps in understanding the structural patterns encoded by the investigated data and extracting essential information regarding the nature of the hydrogen bond interactions as captured by the calculated distance values. The questions we would like to answer in this paper are: When C153 is acting as a hydrogen bond donor, are all its H atoms equivalent or do some of them exhibit a particular behavior? Is there any difference among the acceptor atoms of the solvents? What happens to the hydrogen atoms that are close to the electronegative atoms of C153? Is there a difference in the behavior of the H atoms of the solvents depending on their respective location (e.g., close to or far from the solvent electronegative atoms)?

The paper is organized as follows: the second section details the MD simulation procedure and the calculation of the distance matrices which were used to investigate the hydrogen bond interactions of C153 in the selected solvents. In the third section, the results of this study are discussed. The fourth section provides conclusions and future perspectives.

2. Methodology

A single C153 molecule was immersed into pure solvents-acetonitrile (AN), γ-butyrolactone (GBL), propylene carbonate (PC), methanol (MeOH), ethanol (EtOH), 1-propanol (PrOH), and 1-butanol (BuOH). Both solvent and solute molecules were modeled as non-polarizable, rigid bodies containing a collection of interaction sites. All-atom models were used for the solvents and the solute. Initial configuration and force field files were prepared using PACKMOL [86] and DL_FIELD (version 3.3), respectively [54], and the simulations were carried out in DL_POLY (version 4.07) [55]. These simulations were performed at constant values of the number of molecules, N, the pressure P, and the temperature T on systems containing 1 molecule of solute and 863 molecules of solvents, placed in a cubic box with periodic boundary conditions at an average temperature of 298 K and pressure 1 atm. The NPT ensemble was maintained using a Berendsen thermobarostat, with a relaxation time of 0.1 and 0.2 ps for the thermostat and the barostat, respectively. The Lennard-Jones forces were cut off at 15 Å and long-range Coulomb interactions were treated using the Ewald sum method. The equations of motion were solved using combined SHAKE and velocity algorithms. The equilibration of the systems as well as the production of the HISTORY file for further analysis were performed with a time step of 0.0005 ps and a total number of steps equal to 1,000,000. The configurations of the system were recorded in the HISTORY file every 10 steps, which produced 100,000 configurations that were treated by TRAVIS (version 17/08/12) [56]. This permitted better statistics and a noise level decrease for the radial distribution functions calculated between a single C153 molecule and the solvents.

All the calculations were performed in both GS and ES of C153. The charges of the equilibrated system in GS were switched to that of the ES (without changes in the C153 geometry) and the system was

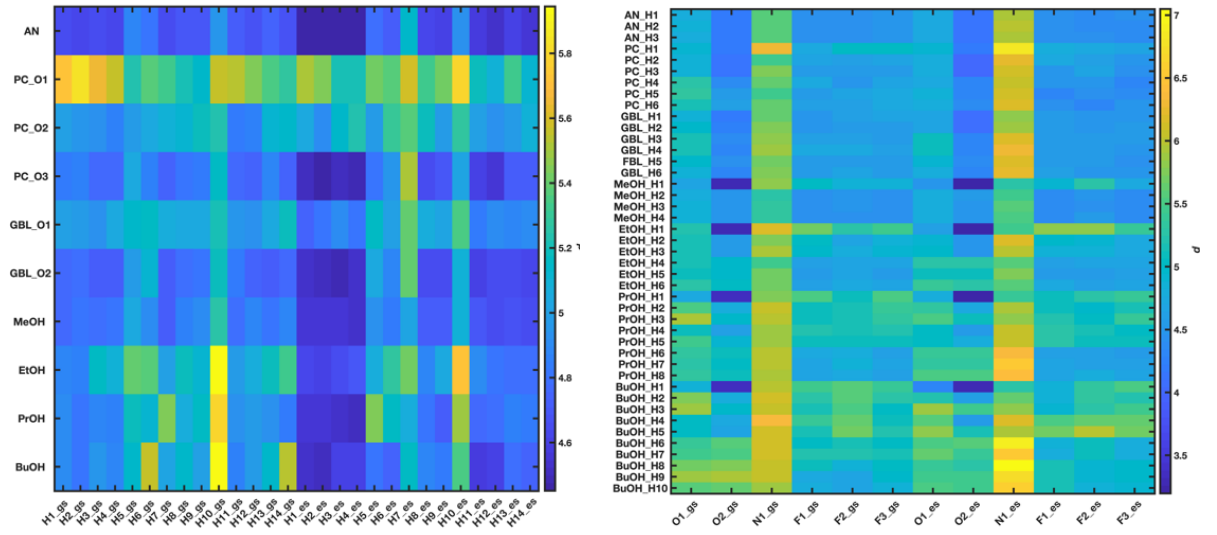


Figure 4. Raw d values obtained through molecular dynamics simulations when considering C153 as a hydrogen bond donor (left) or acceptor (right). The numbering of atoms is given in Figure 2.

left to evolve under the new solute-solvent interactions for 1,000,000 steps with a time step of 0.0005 ps. The geometry of the molecule did not undergo any change.

We used the force field developed by Cinacchi for C153. The atoms' numbering is given in Figure 2 [32].

Quantum-chemical calculations were performed with the B3LYP/6-311+G(d) level of theory using Gaussian09W [91] to obtain the atom positions as well as the Mulliken atomic charges for GS and ES of C153 [41]. The values of these charges are displayed in Figure 1. The OPLS2005 force field was used for alcohols [57], while the force field proposed by Koverga was used for propylene carbonate (PC), γ-butyrolactone (GBL), and acetonitrile (AN) [58].

PCA [96] was then used to get insights into the solvation structure of C153 in GS and ES. Indeed, for each possible hydrogen bond D–H$\cdots$A, the nearest neighbor radial distribution of the electronegative atom of a given compound around either the H or D atoms of the donor compound was calculated. Based on this distribution, the average interatomic distances $d_{\mathrm{H...A}}$ and $d_{\mathrm{D...A}}$ were computed and merged for the sake of simplicity in a unique distance metric d as follows:

$$d_{\mathrm{D...A}} = \sqrt{d^2_{\mathrm{D-A}} + d^2_{\mathrm{H...A}}}$$

The resulting distance values were finally gathered into two different data matrices, one containing the d values estimated when C153 was assumed to act as a donor of hydrogen bonds and the other when C153 was assumed to act as a hydrogen bond acceptor. The first matrix contains 10 rows corresponding to the 10 electronegative atoms of the different solvents under study and 28 columns corresponding to the 14 C–H bonds of C153 in both GS and ES. Each one of its elements represents the distance between a given electronegative atom and one of the 28 C–H bonds of C153. The second matrix contained 43 rows corresponding to the 39 C–H and the 4 O–H bonds of the different solvents under study and 12 columns corresponding to the 6 electronegative atoms of C153 in both GS and ES. Each one of its elements represents the distance between one of the 43 solvent bonds and a given electronegative atom of C153. Both distance matrices were double-centered and finally subjected to PCA by means of in-house-developed Matlab routines. Raw data are displayed in Figure 4.

3. Results

3.1. *Coumarin-153 as hydrogen bond donor*

Figure 5 displays the first/second and the third/fourth principal component biplots resulting from the PCA decomposition of the distance matrix obtained in the case C153 acts as hydrogen bond

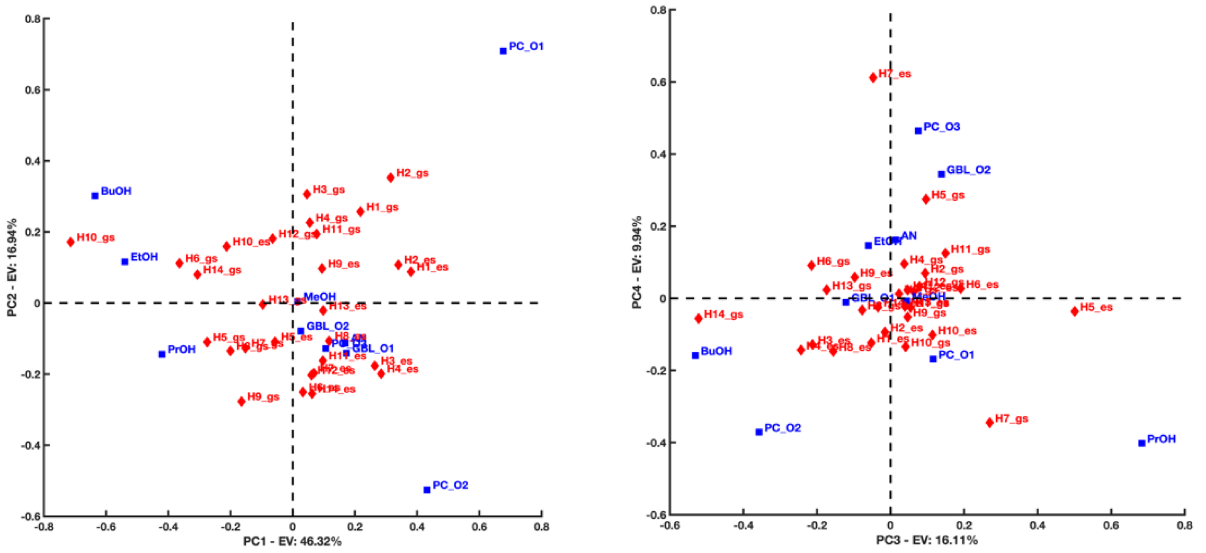


Figure 5. First/second (left) and third/fourth (right) principal component biplots resulting from the PCA decomposition of the distance matrix obtained when considering C153 as a hydrogen bond donor in both its ground and excited states. Blue squares refer to the solvent's electronegative atoms (acceptors), while red diamonds refer to C153's hydrogen atoms (donors).

donor[1]. PC1 highlights a difference in the interacting trend of the H_1 and H_2 atoms located near the nitrogen atom of the quinolizidine heterobicyclic group of C153 with respect to the H_{10} atom that is close to the trifluoromethyl group of C153. The former seems to interact preferentially with the hydroxyl O_1 atom of BuOH, EtOH, and PrOH and be located relatively far from the O atoms (O_1 and O_2) of the PC. The latter exhibits the opposite behavior. This is corroborated by the nearest neighbor distributions represented in Figure 6 and calculated using the following configurations: C_{14}–$H_{10}\cdots$A and C_8–$H_2\cdots$A (A = O_1 or O_2 of PC or O_1 of BuOH), respectively.

Interestingly, PC2 suggests that the distances between the H_3 and H_4 atoms of C153 to the O_1 atom of PC are shortened when C153 goes from GS to ES. The opposite occurs when considering the distances of these H atoms to the O_2 atom of PC. This is in good agreement with the changes observed for the nearest neighbor radial distributions concerning these H atoms and the O_1 and O_2 of PC (see Figure 7). Figure 7 left highlights, in fact, the emergence of a short-distance contribution that reduces the average distance between H_3 and H_4 to the O_1 of PC. Conversely, from Figure 7 right, it is evident that a long-distance contribution makes the average distance between H_3 and H_4 to the O_2 of PC increase. It is interesting to notice that the same conclusion cannot be drawn for the H_1 and H_2 atoms of C153 (their projection coordinates remain positive along PC2 when C153 goes from GS to ES). This suggests that the distance $H_1\cdots O_1$, $H_1\cdots O_2$, $H_2\cdots O_1$ and $H_2\cdots O_2$ do not vary as much as for H_3 and H_4 when C153 goes from GS to ES (see also Figure 8).

Finally, PC3 highlights a difference in the interacting behavior of the H_5 atom of C153 in ES and the H_{14} atom of C153 in GS. It seems that H_5 interacts preferentially with the O atom of BuOH and the O_2 atom of PC and is located relatively far from the O atom of PrOH and the O_1 atom of PC. An opposite interaction trend is instead observed for H_{14}.

It is also interesting to notice that PC4 points out a difference in the interacting behavior of the H_7 atom when C153 goes from GS to ES. When C153 is in GS, H_7 seems to be located quite far from the O of PrOH. In contrast, this distance seems to significantly decrease when C153 is in ES. Additionally, the PC3–PC4 biplot suggests that the interacting behavior of

[1] In this case, the first four principal components account for approximately 90% of the total data variation.

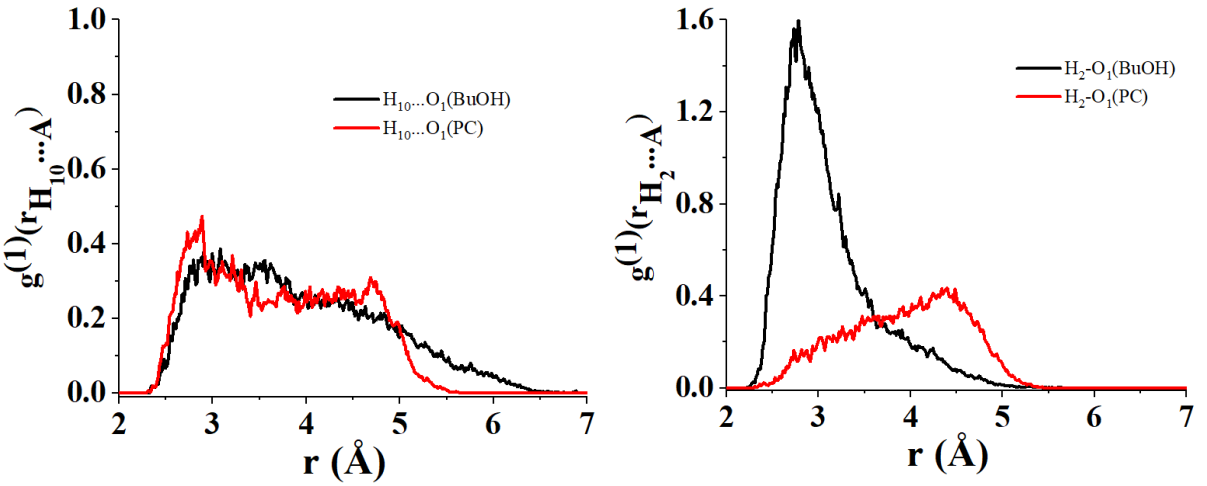


Figure 6. Radial distribution of the nearest neighbor for the atomic couples (i) O_1 of BuOH/H_{10} of C153 (left black), (ii) O_1 of PC/H_{10} of C153 (left red), (iii) O_1 of BuOH/H_2 of C153 (right black) and (iv) O_1 of PC/H_2 of C153.

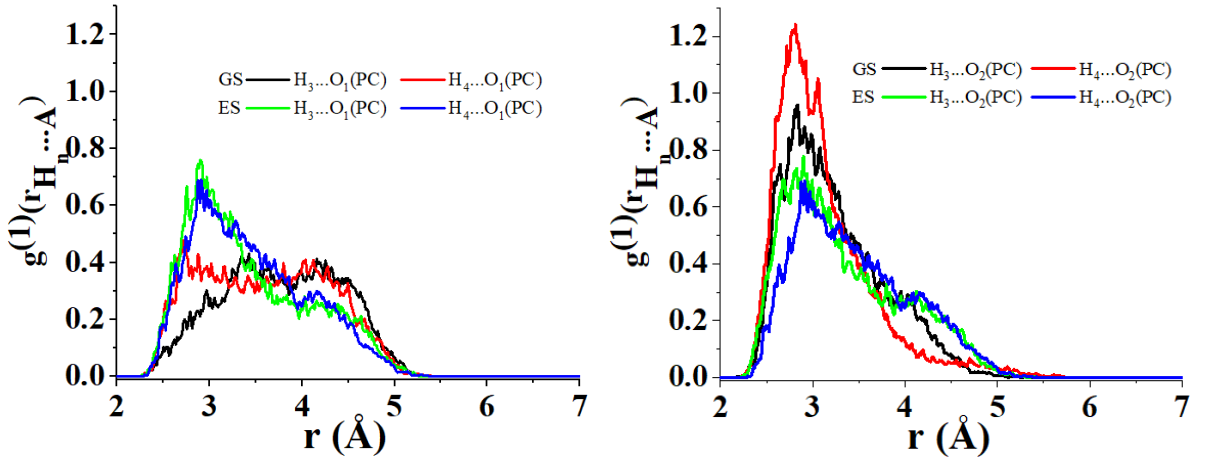


Figure 7. Radial distribution of the nearest neighbor for the atomic couples (i) O_1 of PC/H_3 of C153 in its ground state (GS) (left black), (ii) O_1 of PC/H_4 of C153 in GS (left red), (iii) O_1 of PC/H_3 of C153 in its excited state (ES) (left green), (iv) O_1 of PC/H_4 of C153 in ES (left blue), (v) O_2 of PC/H_3 of C153 in GS (right black), (vi) O_2 of PC/H_4 of C153 in GS (right red), (vii) O_2 of PC/H_3 of C153 in ES (right green) and (viii) O_2 of PC/H_4 of C153 in ES (right blue).

the atoms O_3 in PC and O_2 in GBL might exhibit similarities.

3.2. *Coumarin-153 as hydrogen bond acceptor*

Figure 9 displays the first/second principal component biplot resulting from the PCA decomposition of the distance matrix obtained in the case C153 acts as a hydrogen bond acceptor[2]. PC1 highlights a difference in the interacting trend of the carboxyl O_2 atom of C153 with respect to the F atoms of the trifluoromethyl group of C153 in both GS and ES. The O_2

[2]In this case, the first two principal components account for approximately 86% of the total data variation.

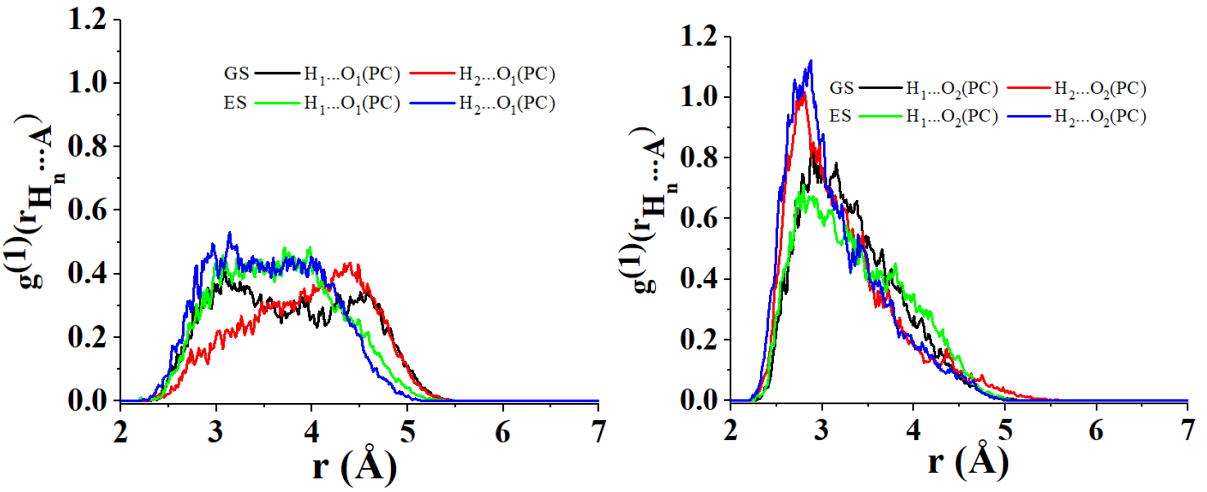


Figure 8. Radial distribution of the nearest neighbor for the atomic couples (i) O_1 of PC/H_1 of C153 in its ground state (GS) (left black), (ii) O_1 of PC/H_2 of C153 in GS (left red), (iii) O_1 of PC/H_1 of C153 in its excited state (ES) (left green), (iv) O_1 of PC/H_2 of C153 in ES (left blue), (v) O_2 of PC/H_1 of C153 in GS (right black), (vi) O_2 of PC/H_2 of C153 in GS (right red), (vii) O_2 of PC/H_1 of C153 in ES (right green) and (viii) O_2 of PC/H_2 of C153 in ES (right blue).

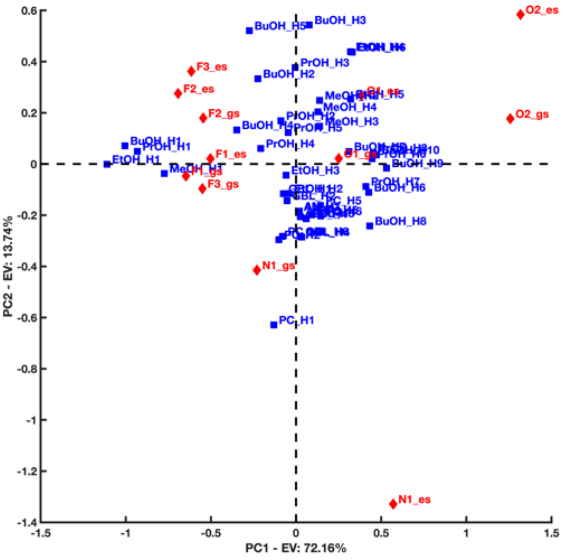


Figure 9. First/second component biplot resulting from the PCA decomposition of the distance matrix obtained when considering C153 as a hydrogen bond acceptor in both its ground and excited states. Blue squares refer to the solvent's hydrogen atoms (donors), while red diamonds refer to C153's electronegative atoms (acceptors).

atom, in fact, seems to interact preferentially with the H_1 atoms of the hydroxyl groups of the alcohol solvents from which the F_{1-3} atoms appear relatively far. The latter, on the other hand, were found to generate preferential interactions with the terminal methyl or ethyl groups of some of the investigated solvents. Furthermore, PC1 also suggests that, overall, the distance between the N_1 atom of C153 and the terminal H atoms of some of the investigated solvents (e.g., PrOH and BuOH) increases when C153 goes from GS to ES. This is clearly corroborated by the nearest neighbor distributions displayed in Figure 10.

The following paragraphs will focus on the hydrogen-bonding interactions involving O_2 of C153 and H_1 of the hydroxyl group of the alcohol solvent on the one hand and H_1 of the PC and GBL solvents on the other hand. We then calculated the values of the distances $d_{H...A}$ and $d_{D...A}$ describing their interactions. More precisely, we calculated these distances for the fifth neighbors. These distances were compared to those representative of two extreme configurations of D–H corresponding to linear and bent geometries. In the former, angle α is equal to 180°, which implies $d_{D...A} = d_{D-H} + d_{H...A}$. This corresponds to a strong interaction. In the latter, α is equal to 90° implying $d_{D...A} = \sqrt{d^2_{D-A} + d^2_{H...A}}$. This

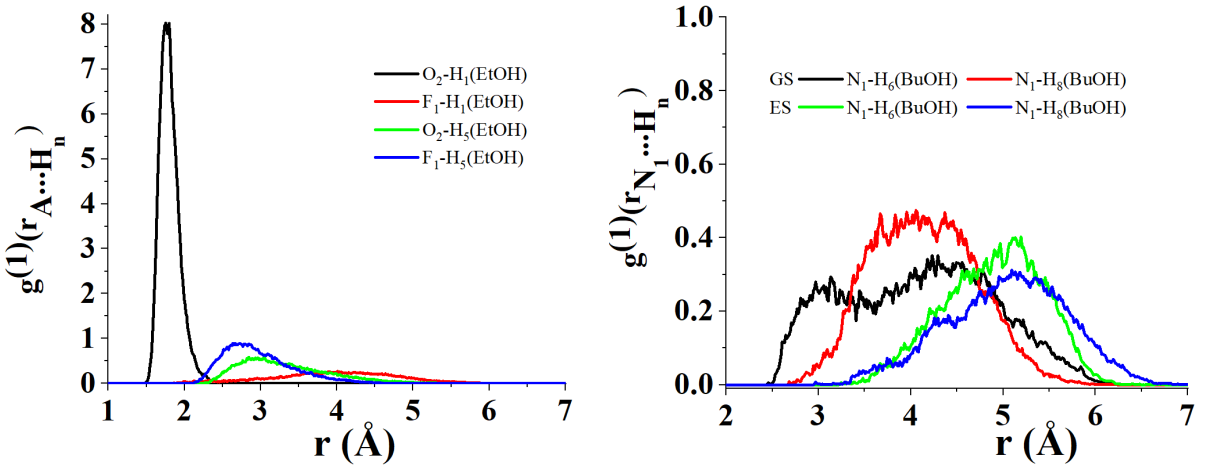


Figure 10. Radial distribution of the nearest neighbor for the atomic couples (i) O_2 of C153 in its ground state (GS) and excited state (ES)/H_1 of EtOH (left black), (ii) F_1 of C153 in GS/ES/H_1 of EtOH (left red), (iii) O_2 of C153 in GS/ES/H_5 of EtOH (left green), (iv) F_1 of C153 in GS/ES/H_5 of EtOH (left blue), (v) N_1 of C153 in GS/H_6 of BuOH (right black), (vi) N_1 of C153 in GS/H_8 of BuOH (right red), (vii) N_1 of C153 in ES/H_6 of BuOH (right green) and (viii) N_1 of C153 in ES/H_8 of BuOH (right blue).

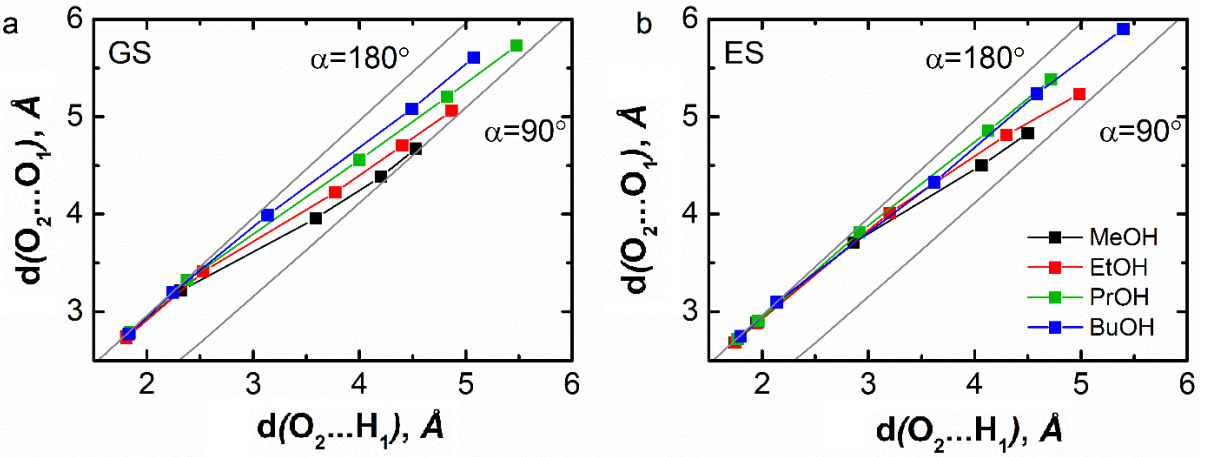


Figure 11. Graphical representation of the two distance values characteristics of the hydrogen-bonding interaction between the carbonyl O_2 atom of C153 and the hydroxyl H_1 atom of alcohols for the five nearest neighbors in the ground (a) and excited state (b) of C153. The line corresponding to $\alpha = 180°$ reflects a linear O_1–$H_1\cdots O_2$ configuration (i.e., $d_{O_2\ldots O_1} = d_{O_2\ldots H_1} + d_{O_1}$–$H_1$), while the line corresponding to $\alpha = 90°$ reflects a bent O_1–$H_1\cdots O_2$ configuration (i.e., $d_{O_2\ldots O_1} = \sqrt{d^2_{O_2\ldots H_1} + d^2_{O_1-H_1}}$).

corresponds to a weak interaction. Figure 11 shows that for all the alcohol solvents, the values of the first two neighbor distances are similar to those typical of a linear geometry of the O_1–$H_1\cdots O_2$. In the ES, this similarity is even more pronounced, which indicates a general reinforcement of the interactions between the solvents and C153. This is in accordance with the findings of Cerezo et al. [33].

Conversely, regarding AN, GBL, and PC (see Figure 12), the calculated distances are very close to

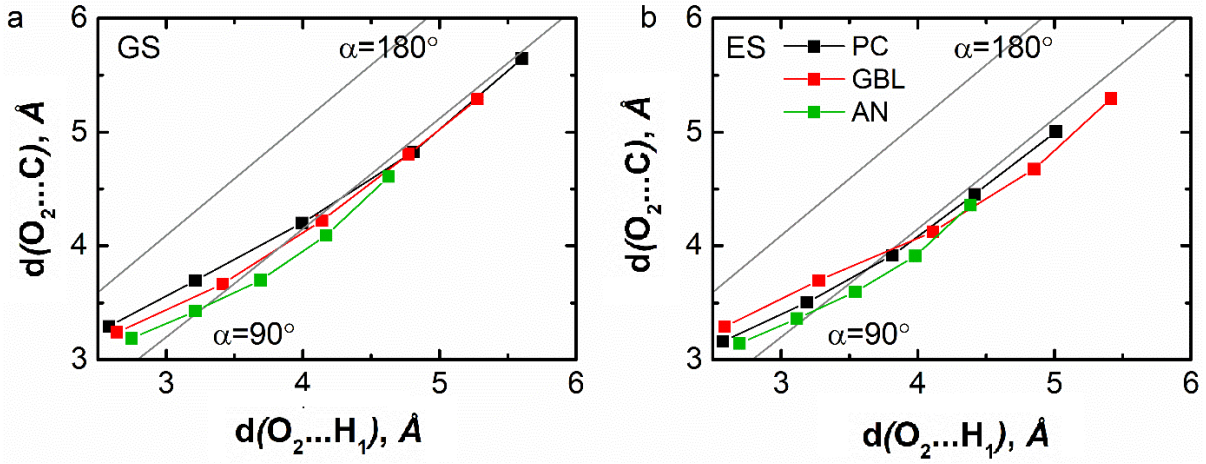


Figure 12. Graphical representation of the two distance values characteristics of the hydrogen-bonding interaction between the carbonyl O_2 atom of C153 and the H_1 atom of propylene carbonate, γ-butyrolactone, and acetonitrile for the five nearest neighbors in the ground (a) and excited state (b) of C153. The line corresponding to $\alpha = 180°$ reflects a linear C–$H_1 \cdots O_2$ configuration (i.e., $d_{O_2 \ldots O_1} = d_{O_2 \ldots H_1} + d_{O_1 - H_1}$), while the line corresponding to $\alpha = 90°$ reflects a bent C–$H_1 \cdots O_2$ configuration (i.e., $d_{O_2 \ldots O_1} = \sqrt{d^2_{O_2 \ldots H_1} + d^2_{O_1 - H_1}}$).

those characteristic of a bent geometry which emphasizes the occurrence of weak interactions between C153 and these solvents. Notice that when C153 is in ES, these distances do not vary and remain similar to those typical of a bent geometry.

4. Conclusions

The aim of this work was to study the hydrogen bond interactions of C153 in various solvents, including alcohols, acetonitrile, γ-butyrolactone, and propylene carbonate. To do this and highlight the existence of similarity patterns in the interacting behavior of donor and acceptor moieties, we combined MD simulations and PCA. Hydrogen bonds were here described by the distances between donor and acceptor atoms. These distance values were calculated using the nearest neighbor radial distribution approach in both GS and ES of C153. These two states were modeled with different charge distributions but retained the geometrical structure of C153. PCA highlighted that, when C153 acts as a hydrogen bond donor:

- the H_1 and H_2 atoms of C153 interact preferentially with the hydroxyl O_1 atom of BuOH, EtOH, and PrOH;
- the H_{10} atom of C153 is instead located at a short distance to the O_1 and O_2 of the PC;
- the distances between the H_3 and H_4 atoms of C153 to the O_1 atom of PC are shortened when C153 goes from GS to ES. The opposite occurs when considering the distances of these H atoms to the O_2 atom of PC;
- the H_5 atom of C153 interacts preferentially with the O_1 atom of BuOH and the O_2 atom of PC and is located relatively far from the O_1 atom of PrOH and the O_1 atom of PC. An opposite interaction trend is instead observed for H_{14};
- when C153 is in GS, its H_7 atom seems to be located quite far from the O_1 of PrOH. In contrast, this distance seems to significantly decrease when C153 is in ES.

In addition, it was observed that when C153 acts as a hydrogen bond acceptor:

- the carboxyl O_2 atom of C153 interacts preferentially with the H_1 atoms of the hydroxyl groups of the alcohol solvents;
- the F_{1-3} atoms of C153 were found to generate preferential interactions with the termi-

nal methyl or ethyl groups of some of the investigated solvents;
- the distance between the N_1 atom of C153 and the terminal H atoms of several of the solvents under study (e.g., PrOH and BuOH) increases when C153 goes from GS to ES.

The hydrogen bond interactions of C153 were also assessed via the behavior of the fifth first neighbors. Our results show that the distance values involving the hydroxyl H atoms of the alcohols are close to those associated with strong hydrogen bond interactions (linear configuration of $O_2\cdots H_1–O_2$). These distances are even shortened in the ES of C153 which indicates a further reinforcement of the hydrogen bonds. On the other hand, the distance values involving the H atoms of PC and GBL are closer to those associated with weak hydrogen bond interactions (bent configuration of $O_2\cdots H_1–C$). Moreover, they do not vary when C153 goes from GS to ES.

Declaration of interests

The authors do not work for advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

The authors acknowledge financial support from the French National Agency for Research (ANR-19-CE05-0009-01 Ultrafast photoinduced processes of organic dyes in ionic liquid/molecular solvent mixtures designed for dye solar cells).

Acknowledgments

This work was performed using computational facilities of the Centre de Resources Informatiques (CRI) de l'Université de Lille and Centre Régional Informatique et d'Applications Numériques de Normandie (CRIANN) which are thankfully acknowledged for the CPU time allocation.

References

[1] R. S. Moog, D. D. Kim, J. J. Oberle, S. G. Ostrowski, *J. Phys. Chem. A*, 2004, **108**, 9294-9301.
[2] M. L. Horng, J. A. Gardecki, A. Papazyan, M. Maroncelli, *J. Phys. Chem.*, 1995, **99**, 17311-17337.
[3] K. Das, B. Jain, H. S. Patel, *J. Phys. Chem. A*, 2006, **110**, 1698-1704.
[4] M. Glasbeek, H. Zhang, *Chem. Rev.*, 2004, **104**, 1929-1954.
[5] H. Shirota, E. W. Castner, *J. Chem. Phys.*, 2000, **112**, 2367-2376.
[6] M. Maroncelli, G. R. Fleming, *J. Chem. Phys.*, 1987, **86**, 6221-6239.
[7] C. F. Chapman, R. S. Fee, M. Maroncelli, *J. Phys. Chem.*, 1995, **99**, 4811-4819.
[8] D. Chakrabarty, A. Chakraborty, D. Seth, N. Sarkar, *J. Phys. Chem. A*, 2005, **109**, 1764-1769.
[9] Z. Guang-Jiu, H. Ke-Li, *ChemPhysChem*, 2008, **9**, 1842-1846.
[10] Y. Kimura, A. Kobayashi, M. Demizu, M. Terazima, *Chem. Phys. Lett.*, 2011, **513**, 53-58.
[11] S. Mondal, R. Halder, B. Biswas, B. Jana, P. C. Singh, *J. Chem. Phys.*, 2016, **144**, article no. 184504.
[12] T. Molotsky, D. Huppert, *J. Phys. Chem. A*, 2002, **106**, 8525-8530.
[13] B. M. Luther, J. R. Kimmel, N. E. Levinger, *J. Chem. Phys.*, 2002, **116**, 3370-3377.
[14] D. Banik, S. Bhattacharya, P. K. Datta, N. Sarkar, *ACS Omega*, 2018, **3**, 383-392.
[15] J. A. Gardecki, M. Maroncelli, *Chem. Phys. Lett.*, 1999, **301**, 571-578.
[16] H. Kaur, S. Koley, S. Ghosh, *J. Phys. Chem. B*, 2014, **118**, 7577-7585.
[17] G. Angulo, M. Brucka, M. Gerecke, G. Grampp, D. Jeannerat, J. Milkiewicz, Y. Mitrev, C. Radzewicz, A. Rosspeintner, E. Vauthey, P. Wnuk, *Phys. Chem. Chem. Phys.*, 2016, **18**, 18460-18469.
[18] R. Królicki, W. Jarzęba, M. Mostafavi, I. Lampre, *J. Phys. Chem. A*, 2002, **106**, 1708-1713.
[19] D. Liu, A. Chen, L. Sui, S. Li, D. Ke, Q. Li, Y. Jiang, M. Jin, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.*, 2019, **219**, 68-73.
[20] P. Hazra, D. Chakrabarty, N. Sarkar, *Chem. Phys. Lett.*, 2003, **371**, 553-562.
[21] D. Chakrabarty, D. Seth, A. Chakraborty, N. Sarkar, *J. Phys. Chem. B*, 2005, **109**, 5753-5758.
[22] N. Ito, S. Arzhantsev, M. Maroncelli, *Chem. Phys. Lett.*, 2004, **396**, 83-91.
[23] S. Li, A. Yu, R. Lu, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.*, 2016, **165**, 161-166.
[24] S. K. Das, P. K. Sahu, M. Sarkar, *J. Phys. Chem. B*, 2013, **117**, 636-647.
[25] S. S. Hossain, S. Paul, A. Samanta, *J. Phys. Chem. B*, 2020, **124**, 2473-2481.
[26] Z. L. Terranova, S. A. Corcelli, *J. Phys. Chem. B*, 2013, **117**, 15659-15666.
[27] A. H. Turner, D. Kim, *J. Chem. Phys.*, 2018, **149**, article no. 174503.
[28] A. Das, S. Das, R. Biswas, *J. Chem. Phys.*, 2015, **142**, article no. 034505.
[29] M. Sayed, D. K. Maity, H. Pal, *J. Photochem. Photobiol. A: Chem.*, 2023, **434**, article no. 114265.
[30] B. Biswas, P. C. Singh, *J. Fluorine Chem.*, 2020, **235**, article no. 109414.
[31] F. Cichos, R. Brown, P. A. Bopp, *J. Chem. Phys.*, 2001, **114**, 6834-6842.

[32] G. Cinacchi, F. Ingrosso, A. Tani, *J. Phys. Chem. B*, 2006, **110**, 13633-13641.
[33] J. Cerezo, S. Gao, N. Armaroli, F. Ingrosso, G. Prampolini, F. Santoro, B. Ventura, M. Pastore, *Molecules*, 2023, **28**, article no. 3910.
[34] Y. Liu, J. Ding, D. Shi, J. Sun, *J. Phys. Chem. A*, 2008, **112**, 6244-6248.
[35] X. Jinmei, C. Junsheng, D. Shunle *et al.*, *J. Phys. Org. Chem.*, 2016, **29**, 305-311.
[36] Y. Smortsova, F.-A. Miannay, T. Gustavsson, F. Sauvage, F. Ingrosso, O. Kalugin, A. Idrissi, *J. Mol. Liquids*, 2021, **340**, article no. 117163.
[37] J. Xu, J. Chen, S. Dong, A. Fu, H. Li, T. Chu, *J. Phys. Org. Chem.*, 2016, **29**, 305-311.
[38] G. Zhao, K. Han, *ChemPhysChem*, 2008, **9**, 1842-1846.
[39] C. Laurence, M. Berthelot, M. Helbert, K. Sraidi, *J. Phys. Chem.*, 1989, **93**, 3799-3802.
[40] T. M. Krygowski, H. Szatyłowicz, J. E. Zachara, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 2077-2082.
[41] R. Kersaudy, E. Gagnière, N. Caillol, D. Colson, J. P. Valour, D. Mangin, *J. Crystal Growth*, 2023, **618**, article no. 127329.
[42] H. Maeda, Y. Ozaki, M. Tanaka, N. Hayashi, T. Kojima, *J. Near Infrared Spectrosc.*, 1995, **3**, 191-201.
[43] H. Koch, K. Noack, S. Will, *Phys. Chem. Chem. Phys.*, 2016, **18**, 28370-28375.
[44] S.-Y. Yang, C.-D. Eom, Y. Han, Y. Chang, Y. Park, J.-J. Lee, J.-W. Choi, H. Yeo, *Wood Fiber Sci.*, 2014, **46**, 138-147.
[45] M. Hu, M. Tang, H. Wang, M. Zhang, S. Zhu, Z. Yang, S. Zhou, H. Zhang, J. Hu, Y. Guo, X. Wei, Y. Liao Terahertz, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.*, 2021, **254**, article no. 119611.
[46] Z. Zhou, J. Li, Y. Wang, Z. Wang, Y. Yu, *J. Mol. Struct.*, 2024, **1295**, article no. 136677.
[47] R. Watanabe, A. Oishi, S. Nakamura, H. Hagihara, H. Shinzawa, *Polymer*, 2023, **283**, article no. 126243.
[48] P. Ouiyangkul, V. Tantishaiyakul, N. Hirun, *Int. J. Pharmaceut.*, 2020, **587**, article no. 119630.
[49] B. L. Groot, X. Daura, A. E. Mark, H. Grubmüller, *J. Mol. Biol.*, 2001, **309**, 299-313.
[50] M. Yuguang, P. H. Nguyen, S. Gerhard, *Prot.: Struct., Funct. Bioinform.*, 2004, **58**, 45-52.
[51] G. G. Maisuradze, A. Liwo, H. A. Scheraga, *J. Mol. Biol.*, 2009, **385**, 312-329.
[52] S. A. M. Stein, A. E. Loccisano, S. M. Firestine, J. D. Evanseck, *Annual Reports in Computational Chemistry. 2*, Elsevier, 2006, Chapter 13, 233-261 pages.
[53] I. Daidone, A. Amadei, *WIREs Comput. Mol. Sci.*, 2012, **2**, 762-770.
[54] C. W. Yong, *J. Chem. Inform. Model.*, 2016, **56**, 1405-1409.
[55] W. Smith, I. T. Todorov, *Mol. Simul.*, 2006, **32**, 935-943.
[56] M. Brehm, B. Kirchner, *J. Chem. Inform. Model.*, 2011, **51**, 2007-2023.
[57] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225-11236.
[58] V. Koverga, "Local structure organization in ionic liquids and molecular solvents mixtures: A molecular dynamics simulation", Doctoral thesis, University of Lille, Lille, 2017.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 67-76
https://doi.org/10.5802/crchim.324

Research article

French Network on Solvation (GDR 2035 SolvATE)

On the phase behavior of sorbitol/water/H_2/CO_2 mixtures at high pressures and temperatures by in situ infrared spectroscopy

Isaline Bonnin [a, b], Raphaël Mereau [a], Karine De Oliveira Vigier [b] and Thierry Tassaing [*, a]

[a] Institut des Sciences Moléculaires (ISM), Univ. Bordeaux, CNRS, Bordeaux INP, ISM, UMR 5255, F-33400 Talence, France

[b] Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Univ. de Poitiers, UMR CNRS 7285, 1 rue Marcel Doré, 86073 Poitiers Cedex 9, France

E-mail: thierry.tassaing@u-bordeaux.fr (T. Tassaing)

Abstract. The thermodynamic behavior of multi-component system, which includes sorbitol dissolved in water under CO_2 and H_2 pressure, is investigated by in situ infrared spectroscopy. Our study focuses on the gas phase of binary, ternary, and quaternary systems at temperatures ranging from 40 °C to 220 °C and at CO_2 pressures between 30 and 120 bar, while maintaining a constant pressure of H_2 at 30 or 60 bar. The influence of addition of sorbitol in the liquid aqueous phase and of H_2 in the CO_2-rich gas phase on the mutual solubility of CO_2 and water is evaluated. In our experimental conditions (T< 220 °C and P< 120 bar), it is found that in such multi-component system, a water-rich liquid phase always coexists with a CO_2-rich gas phase. Presence of sorbitol in the water-rich phase and H_2 in the CO_2-rich phase has a limited impact on the mutual solubility of water and CO_2. Furthermore, sorbitol is not soluble in the CO_2-rich phase and instead remains in the water-rich liquid phase over the thermodynamic range investigated.

Keywords. Sorbitol, CO_2, Hydrogen, Water, High pressures, Infrared spectroscopy.

Funding. Region Nouvelle Aquitaine, University of Poitiers, MITI of CNRS, INCREASE Federation, GDR 2035 SolvATE.

Manuscript received 19 February 2024, revised 15 May 2024 and 17 June 2024, accepted 19 June 2024.

1. Introduction

In the context of biomass valorization towards substitution of petroleum-based materials, sorbitans represent bio-sourced alternatives to a large number of petroleum-based molecules; and can be used in a wide range of applications such as additives, pharmaceuticals and monomers for polymer industries [1–4]. In particular, the hydrogenation of glucose into sorbitol and the dehydration of sorbitol with various catalytic systems have been studied in depth [5,6]. However, the synthesis of these diols remains, to this day, a technological challenge due to the cost of energy-consuming processes, use of expensive and toxic solvents as well as the intensive purification steps.

* Corresponding author

To this regard, it is worthwhile to develop "one-pot" protocols for the obtention of sorbitans starting from glucose combining heterogeneous catalytic hydrogenation and an acid-catalyzed dehydration reaction [5,7]. Conventionally, strong mineral acids are used and they limit the selective obtention of sorbitans to favor its dehydration to isosorbide (produced by double dehydration of sorbitol). Being non-toxic and non-corrosive, gaseous CO_2 has attracted attention in replacing environmentally unsafe mineral acid catalysts thanks to the reversible control of reaction mixture's pH.

In order to study this one-pot reaction (hydrogenation and dehydration), we combined in situ spectroscopic analysis with molecular modeling to better understand the mechanisms and the kinetics in the formation of different reactions' intermediates [8–10]. However, the thermodynamics of these multi-component mixtures are not clearly known. We recall that the aqueous dehydration of sorbitol was performed under H_2 and CO_2 pressure (between 30 and 120 bar); and at high temperature (up to 220 °C). The phase behavior and the mutual solubility of all the components of this multi-component mixture depend strongly upon the pressure/temperature conditions which in turn can significantly influence the selectivity and yield of the sorbitans synthesis during hydrogenation and dehydration occurring in the liquid phase.

The aim of this paper is to determine the thermodynamic behavior of binary, ternary, and quaternary systems composed of sorbitol/water/H_2/CO_2 at temperatures ranging from 40 °C to 220 °C; at CO_2 pressure between 30 bar to 120 bar; H_2 pressure at 30 or 60 bar. Additionally, we have focused our investigation on the gas phase using infrared spectroscopy which performs well in situ and in operando investigations. Of particular note, the evolution of the intensity of selected vibrational modes of CO_2 and H_2O have also been analyzed as a function of temperature and pressure in order to determine the evolution of H_2O and CO_2 concentrations in binary, ternary, and quaternary mixtures. From the results obtained, it was possible to show that in such multi-component system, a water-rich liquid phase coexists with a CO_2-rich gas phase under our experimental conditions ($T < 220$ °C and $P < 120$ bar). In addition, it was found that sorbitol remains in the water-rich liquid phase over the thermodynamic range investigated.

2. Experimental setup

2.1. *High pressure in situ infrared absorption setup*

The in situ analysis of sorbitol–water–CO_2–H_2 mixture was done with IR spectroscopy in one high-pressure cell (Figure 1). Due to strong absorption of water, the liquid-phase analysis is best performed using ATR-FTIR equipment which can withstand high temperature and pressure. Due to the lack of such equipment, we have focused our study on the gas phase using a home made HP cell which can withstand high temperature of about 250 °C and pressure up to 20 MPa. The home-made stainless-steel cell is composed of three cylindrical windows (one sapphire window for visual observation and two silicon windows for IR absorption with a pathlength of 26 mm). Windows were positioned on the flat surface of an inconel plug with a Kapton foil placed between the window and the plug to compensate for imperfections at the two surfaces (unsupported area principle). Flat Kapton rings were used to ensure sealing between the plug and the cell body. Heating was achieved using four cartridge heaters inside the body of the cell and a thermocouple located close to one cartridge was used to regulate the temperature with an accuracy of $\Delta T = \pm 0.5$ °C. The cell was connected via a stainless capillary to a pressurizing system which allows the regulation of pressure with an accuracy of $\Delta P = \pm 0.1$ bar.

For IR absorption measurements, a Nicolet 6700 FTIR spectrometer equipped with a Globar source, a KBr/Ge beamsplitter and a DLaTGS/KBr thermal detector was used to determine the spectral range, from 400 to 6500 cm^{-1}. Single beam spectra with 4 cm^{-1} resolution were obtained by Fourier transformation of 100 accumulated interferograms in order to improve the signal to noise ratio.

2.2. *Experimental procedure*

For the quaternary mixture, the lower part of the cell was filled with 1 ml of an aqueous solution of 30 wt% of sorbitol (purchased from Sigma Aldrich (>98% purity)). The experiments were performed by initially adding a predetermined pressure of H_2 (purchased from Air Liquid (99.9999% purity)); CO_2 (purchased from Air Liquid (99.95% purity)) was then added to

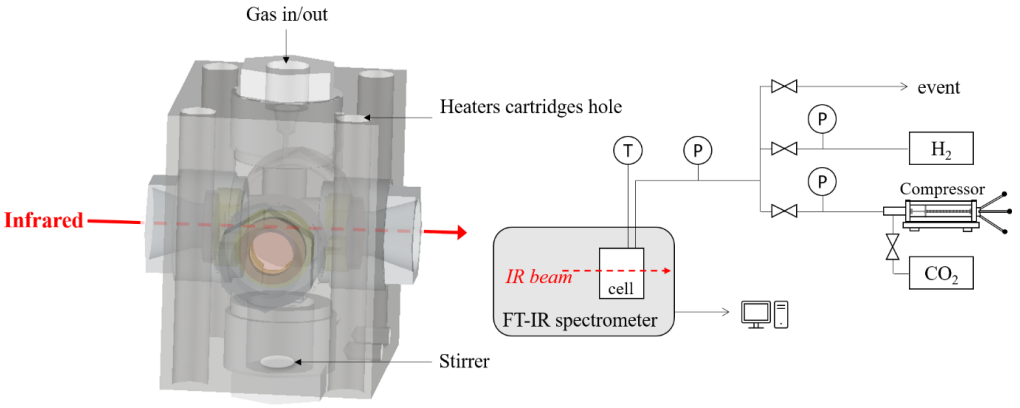


Figure 1. Design of the high-pressure experimental device with a high-pressure cell coupled with IR spectrometer for the in situ measurements of the CO_2-rich phase of the sorbitol/water/H_2/CO_2 mixtures.

this high-pressure cell using a manual pump (TOP industrie) up to the desired pressure. We mention that the notation 30/30 bar CO_2/H_2 means that 30 bar of H_2 is introduced and then the cell was filled with CO_2 until the total pressure reached 60 bar. Making the hypothesis that both CO_2 and H_2 behave as ideal gases and that their pressures can be added, we assumed that 30 bar of CO_2 was added to the cell. The cell volume (8 ml) is fixed and only the pressure can be adjusted.

2.3. *Infrared absorption spectra and calibration*

Figure 2 illustrates (for the case of a binary H_2O/CO_2 mixture) the spectral changes of the infrared absorption spectra in the CO_2-rich phase that occur with an increase of the temperature from 40 to 220 °C at a constant pressure of 60 bar.

In that region, fundamental and combination bands of CO_2 and H_2O were observed. However, only the spectral range between 2500 and 3200 cm^{-1}, and between 4500 and 7500 cm^{-1} are of interest for our purpose under our experimental conditions. Indeed, the strong absorption of CO_2 and H_2O precludes any quantitative analysis outside this spectral range.

For CO_2, the region from 4750 to 5200 cm^{-1} with overlapping peaks of CO_2 at 4800 cm^{-1}, 4950 cm^{-1} and 5100 cm^{-1} are of interest and correspond to the combination modes $4\nu_2+\nu_3$, $\nu_1+2\nu_2+\nu_3$ and $2\nu_1+\nu_3$, respectively. ν_1, ν_2 and ν_3 are the fundamental vibrational modes of CO_2 with symmetric stretch, bending mode and antisymmetric stretch, respectively [11–14]. The weak peak observed at 6950 cm^{-1} is assigned to the $3\nu_3$ overtone of CO_2.

For H_2O, the band at 1600 cm^{-1} which is saturated at temperatures above 120 °C is related to the bending mode ν_2 of water [15]. The profile at 5300 cm^{-1} having an increasing intensity with the temperature, is assigned to the $\nu_2+\nu_3$ combination mode of water [16,17]. In 7000–7500 cm^{-1} spectral range, the profile with a doublet structure observed at 7200 cm^{-1} with an enhanced intensity with temperature is assigned to the $2\nu_3$ overtone of water.

In order to quantify the concentration of CO_2, we selected the vibrational mode $2\nu_1+\nu_3$ of CO_2 and determined its epsilon value from its integrated area (from 4900 to 4740 cm^{-1}). This was selected as it negligibly overlaps with any contribution of water. We, then measured the infrared absorption spectra of neat gaseous CO_2 at constant pressures of 30, 60 and 120 bar at 40, 80, 120, 160, 200 and 220 °C. Knowing the concentration of CO_2 (in mol·L^{-1}) from the NIST database (NIST Chemistry Webbook [18]), the pathlength of the cell (l) and measuring the integrated area (A) in 4900 to 4740 cm^{-1} range; the molar extinction coefficient (epsilon = ε) values are then calculated by applying the Beer–Lambert law: $A=\varepsilon\cdot l\cdot c$ (see Table 1). For all temperatures and pressures, it can be seen that there is a significant variation of the epsilon values with increase in pressure. Hence, these values are used as reference for the same given temperature and pressure to determine CO_2 concentration in all analyzed mixtures.

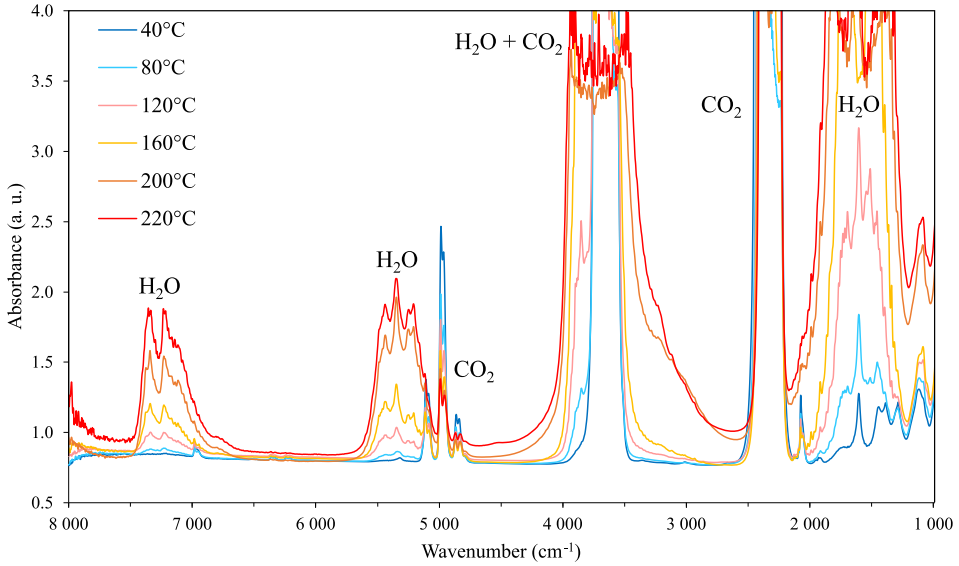


Figure 2. Infrared spectra of the CO_2-rich phase of the H_2O–CO_2 binary system at a constant pressure of CO_2 of 60 bar; and temperature from 40 °C to 220 °C.

Table 1. Molar extinction coefficients (in L·mol^{-1}·cm^{-2}) of the $2\nu_1 + \nu_3$ mode of CO_2 integrated in 4900 to 4740 cm^{-1} range

	40 °C	80 °C	120 °C	160 °C	200 °C	220 °C
30 bar CO_2	2.010	2.101	2.114	2.014	1.931	1.806
60 bar CO_2	1.916	2.013	1.970	1.903	1.791	1.675
120 bar CO_2	1.341	1.645	1.688	1.646	1.532	Saturate

By the same token, to quantify the concentration of water, selecting a vibrational mode of water and determining its epsilon value is required. In water–CO_2 mixtures, the molar fraction of CO_2 can be calculated from the concentration of CO_2 and water. We selected the $\nu_2 + \nu_3$ combination mode of water and used its integrated area from 5800 to 5347 cm^{-1} to calculate the concentration of water. We emphasize that only the high frequency wing of the mode of water has been considered for integration as the low frequency range is superimposed with a contribution of CO_2. Then, the epsilon of this band associated with water is fitted to obtain an experimental molar fraction of water calculated from the experimental water and CO_2 concentration, that is consistent with the molar fraction in the literature [19,20]. As shown in the Figure 3, a good fit is obtained in the whole temperature and pressure range specific to this study when a unique epsilon value of water is used and found to be equal to 140 L·mol^{-1}·cm^{-2}. Thus, we have evaluated a relative uncertainty of ±5% on the concentration measurements using our setup and data processing methodology.

Finally, it is worth noting that if sorbitol is soluble in the CO_2-rich phase, one should observe a peak in the alkane region (in the range 2800–3000 cm^{-1}) related to C–H stretching vibrations with a detection limit of about 10^{-4}–10^{-5} $g_{\text{sorbitol}}/g_{CO_2}$. We emphasize that the initial total amount of sorbitol is excessive in comparison to the minimum amount of sorbitol that could be detected.

In a few experiments, the collision-induced IR spectral band of H_2 is also observed at about 4150 cm^{-1} (see Figures 5 and 12 in the results and discussion section). This band has been reported previously for H_2 diluted in monoatomic gases under high pressure conditions [21,22]. Indeed, as the H_2 stretching vibration is inactive in IR spectroscopy,

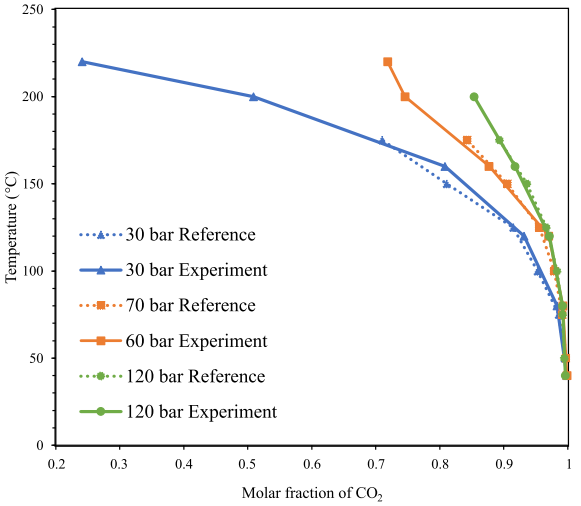


Figure 3. Comparison of the molar fraction of CO_2 in the CO_2-rich phase of H_2O–CO_2 mixtures calculated from our infrared spectra with literature data.

the observed H_2 spectrum in the CO_2-rich phase is due to an induced dipole moment of the H_2 molecule that results from its interaction with the surrounding CO_2 molecules. Therefore, the induced molar extinction coefficient of the H_2 peak strongly depends on the temperature and pressure and precludes the use of this peak for quantitative analyses [23].

3. Results and discussion

3.1. *H_2O–CO_2 binary system*

Preliminary studies are done in H_2O–CO_2 binary system to analyze the influence of water on the CO_2 concentration in the CO_2-rich phase based on the pressure of CO_2 and temperature. Infrared spectra of the gas phase of the binary system are illustrated in Figure 2 and has been described in earlier sections. Beer–Lambert law is applied to determine the evolution of the concentration of CO_2 in the CO_2-rich phase of the H_2O–CO_2 binary mixture and compare it with the concentration of neat CO_2 in the same thermodynamic range from 40 to 220 °C at constant CO_2 pressures (30, 60 and 120 bar as shown in Figure 4).

Looking initially at CO_2 alone (Figure 4, dotted lines), it can be inferred that the concentration of CO_2 gas decreases when the temperature rises.

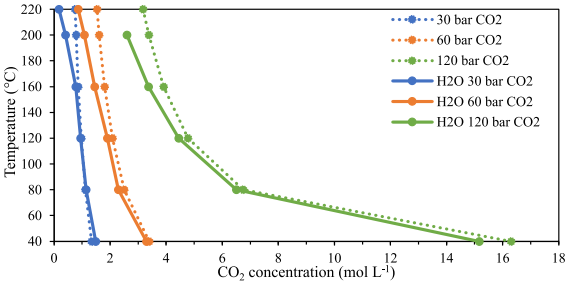


Figure 4. Temperature versus CO_2 concentration at constant pressure in neat CO_2 and in the CO_2-rich phase of the binary system H_2O–CO_2.

This effect is very pronounced at high pressure of 120 bar and less marked at 60 and 30 bar. In the presence of water (Figure 4, solid line), the same effect is observed but between 160 and 220 °C, the concentration of CO_2 decreases further. This effect can be explained by two phenomena: as temperature increases, CO_2 dissolves in water and consequently, its concentration in the gas phase decreases and concomitantly, water is "evaporated" in the gas phase leading to a decrease in CO_2 concentration. Both phenomena governing the mutual solubility of water and CO_2 are consistent with data reported in the literature [19,20].

3.2. *CO_2–H_2 binary system*

The evolution of the infrared spectra of the binary mixture CO_2/H_2 (30 bar/60 bar) with temperature is elucidated in Figure 5. We emphasize that above the critical point of CO_2, the binary mixture H_2/CO_2 displays only one gas phase as previously reported in thermodynamic studies on H_2/CO_2 mixtures [24–26]. Interestingly, a barely detectable band at 4100 cm^{-1} in neat H_2 is clearly observed after adding CO_2 in the cell at 40 °C. Then, its intensity decreases continuously as the temperature rises. This band is assigned to the vibrational stretching mode (vibron) of H_2 interacting with surrounding CO_2 molecules and the effect of CO_2 on this band has been discussed by our group in a recent article [23]. As the intensity of this band is related to "induced" effects, it is not possible to accurately calculate the concentration of H_2 in the mixture.

However, in order to determine the effect of H_2 on the CO_2 concentration, we determined the evolution

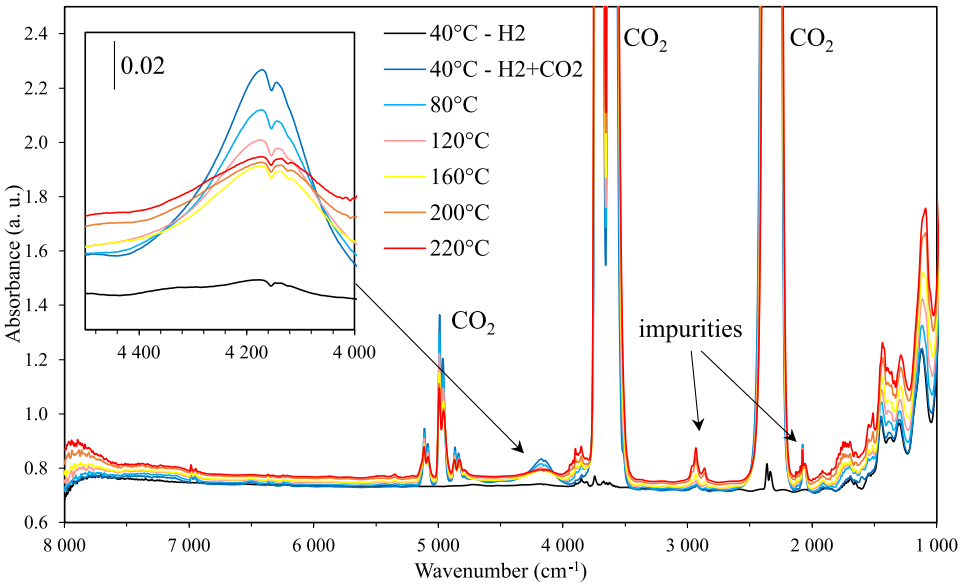


Figure 5. Infrared spectra of the binary system CO_2–H_2 at 90 bar from 40 °C to 220 °C.

of the concentration of CO_2 in the H_2–CO_2 binary mixture and compared it with the concentration of neat CO_2 in the same thermodynamic range (from 40 to 220 °C) at constant pressures of CO_2 (30 and 60 bar) as shown in Figure 6. The concentration of CO_2 appears to decrease at 60 bar of H_2 (yellow line in Figure 6), while at 30 bar (of H_2) negligible effect is observed (grey line in Figure 6). When the pressure of CO_2 is doubled from 30 to 60 bar while keeping the pressure of H_2 at 30 bar results in a weak decrease of the concentration of CO_2 in comparison with neat CO_2 at low temperature (below 120 °C). This effect could be attributed to the intermolecular interactions between H_2 and CO_2 as evidenced by "induced" effects on the vibrational stretching mode of H_2.

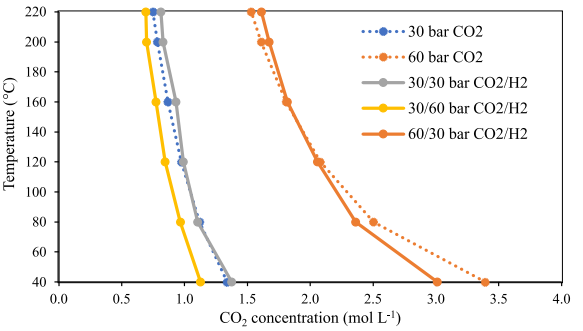


Figure 6. Temperature versus CO_2 concentration at constant pressure in neat CO_2 and in the CO_2-rich phase of the binary system CO_2–H_2.

3.3. *H_2O–CO_2–H_2 ternary system*

The binary system comprising CO_2 and H_2 is compared to the ternary system where water is added to the high-pressure cell at a given pressure of H_2 and CO_2 (Figure 7). At temperatures below 150 °C, concentration of CO_2 in the H_2/CO_2 binary mixture is similar to that of the ternary mixture with water. However, at temperatures above 150 °C, a significant decrease in CO_2 concentration is observed in the CO_2-rich phase of the ternary mixture when compared to CO_2-rich phase of the binary system CO_2–H_2.

To determine if the presence of H_2 has a significant effect on the solvation of water in the gas phase, for a given ratio H_2/CO_2 (see Figure 8), the evolution of water concentration according to the temperature for the binary (H_2O–CO_2) system is compared with the ternary system (H_2O–CO_2–H_2). First, the increase of the CO_2 pressure from 30 bar to 60 bar slightly increases the concentration of water (dotted lines). For example, at a constant temperature of

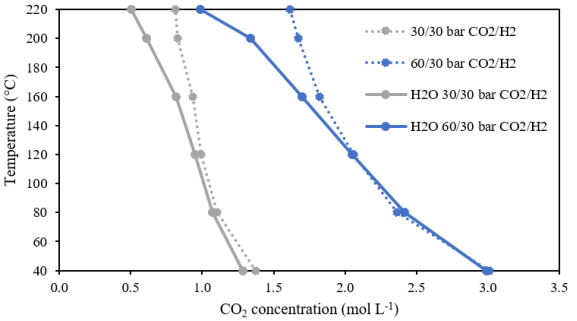


Figure 7. Temperature versus CO_2 concentration at constant pressure in the CO_2-rich phase of the binary (CO_2–H_2) and ternary (H_2O–CO_2–H_2) systems.

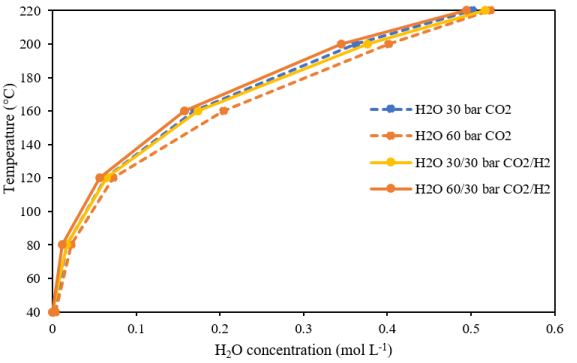


Figure 8. Temperature versus H_2O concentration at constant pressure in the CO_2-rich phase of the binary (H_2O–CO_2) and ternary (H_2O–CO_2–H_2) systems.

200 °C, the water concentration in the vapor phase is 0.36 mol·L^{-1} and 0.41 mol·L^{-1} for 30 bar and 60 bar of CO_2, respectively. When H_2 is added to the system (full lines), negligible effect is seen when 30 bar of CO_2 mixes with 30 bar of H_2 (yellow line). However, in the system comprising 60 bar of CO_2, the addition of 30 bar of H_2 leads to a slight decrease of the solubility water in the CO_2 phase (orange lines) and is similar to that measured with 30 bar of CO_2. Thus, the lower solubility of water in the CO_2 phase in the presence of H_2 could be due to the lower CO_2 concentration in the CO_2 phase as shown for the binary system CO_2–H_2 (cf. Figure 7) as well as to the fact that H_2 is a highly hydrophobic molecule that is expected to limit the solubility of water in the CO_2 phase.

3.4. *H_2O–sorbitol–CO_2 ternary system*

Figure 9 displays the infrared spectra of the gas phase of a ternary system, an aqueous solution of 30 wt% of sorbitol under CO_2 pressure. It is worth noting that the evolution of the profiles is similar to that observed when the temperature is increased in H_2O–CO_2 binary system (cf. Figure 2). This suggests that sorbitol is not detected in the CO_2-rich phase (under our detection limit of 10^{-4}–10^{-5} $g_{sorbitol}/g_{CO_2}$) as we have not observed any new peaks related to sorbitol between 2900 cm^{-1} and 1100 cm^{-1}.

The CO_2 concentration in the CO_2-rich phase of the H_2O–sorbitol–CO_2 ternary mixture is compared to that of the H_2O–CO_2 binary system (Figure 10). No significant difference in the CO_2 concentration was seen at 30 or 60 bar of CO_2. However, at a higher CO_2 pressure of 120 bar, the concentration of CO_2 slightly increases (green, full line) when sorbitol is added to water. It can be inferred that the presence of sorbitol in water has no effect on the CO_2 concentration in the gas phase at low CO_2 pressures and higher pressure of 120 bar, sorbitol modifies the water–CO_2 equilibrium.

In order to assess the modification of the water–CO_2 equilibrium due to the presence of sorbitol in the aqueous phase, the evolution of the concentration of water in the CO_2-rich phase for the H_2O–sorbitol–CO_2 ternary mixture with temperature is observed (see Figure 11) and compared with the results obtained for the H_2O–CO_2 binary system. As shown in Figure 10, no effect is seen at CO_2 pressure < 60 bar, as the concentration of water in the CO_2-rich phase is similar for the binary and ternary systems. However, at CO_2 pressure of 120 bar, the water concentration decreases from 0.3 mol·L^{-1} to 0.24 mol·L^{-1} at 160 °C. Therefore, it can be inferred that sorbitol prevents the solvation of water in the CO_2-rich phase by "keeping" water in the liquid phase and leads to a higher concentration of CO_2 in the CO_2-rich phase.

3.5. *H_2O–sorbitol–CO_2–H_2 quaternary system*

Figure 12 shows the evolution of the infrared spectra in the CO_2-rich phase as a function of temperature of the quaternary system comprising an aqueous solution of 30 wt% of sorbitol and the mixture of CO_2 and H_2 gases. The "induced" peak of H_2 at 4150 cm^{-1} is only observed at low temperatures. With increase in

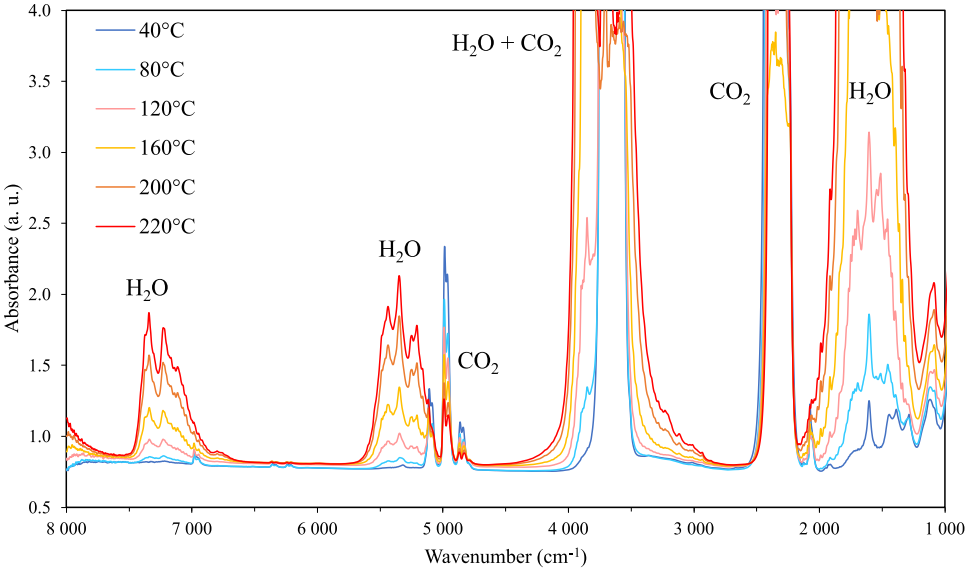


Figure 9. Infrared spectra of the CO_2-rich phase of the ternary system H_2O–sorbitol–CO_2 at 60 bar from 40 °C to 220 °C.

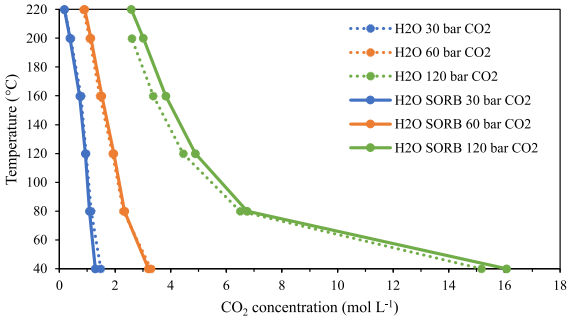


Figure 10. Temperature versus CO_2 concentration at constant pressure in the CO_2-rich phase of the binary (H_2O–CO_2) and ternary (sorbitol–H_2O–CO_2) systems.

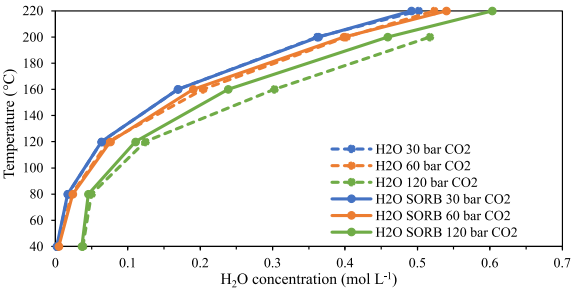


Figure 11. Temperature versus H_2O concentration at constant pressure in the CO_2-rich phase of the binary (H_2O–CO_2) and the ternary (sorbitol–H_2O–CO_2) systems.

temperature, the band of water increases and overlaps H_2 while the bands of CO_2 decrease. As seen in the ternary system (Figure 9), sorbitol is not detected in the CO_2-rich phase.

In Figure 13, the concentration of CO_2 in the CO_2-rich phase for the quaternary mixture is compared with the ternary system H_2O–CO_2–H_2. It appears that at the pressure range investigated (corresponds to conditions used for the catalytic reactions reported previously [8,10]), the addition of sorbitol does not have a significant effect on the CO_2 concentration in the presence of H_2.

By the same token, the quaternary system is compared to the ternary system H_2O–sorbitol–CO_2 to study the effect of H_2 on water concentration in the CO_2-rich phase (cf. Figure 14). Almost no effect is observed, which is consistent with the results reported in Figure 13.

Therefore, at the pressure and temperature range investigated where two distinct water-rich and CO_2-rich phase are observed, the presence of sorbitol in the water-rich phase and H_2 in the CO_2-rich phase have a limited impact on the mutual solubility of water and CO_2.

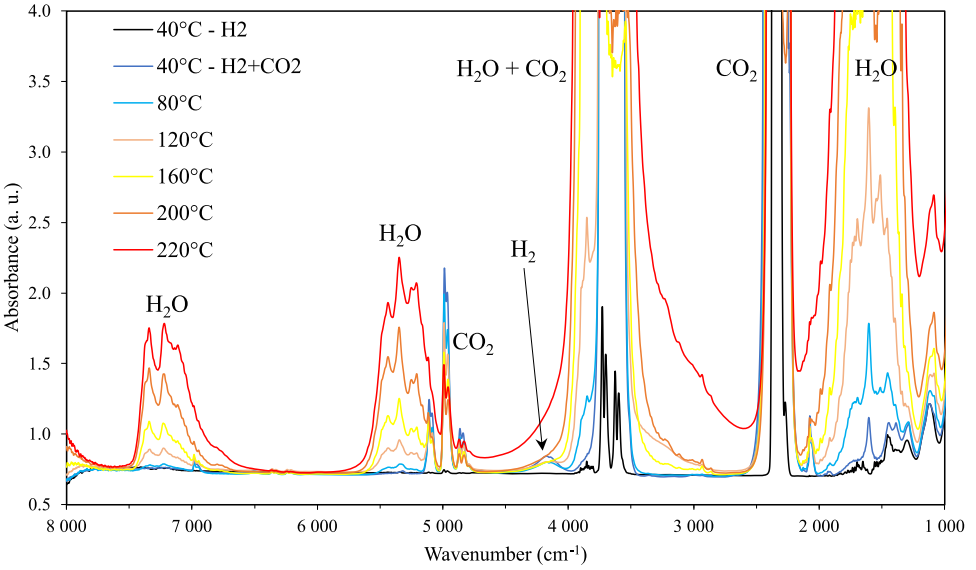


Figure 12. Infrared spectra of the CO_2-rich phase of the quaternary system H_2O–sorbitol–CO_2–H_2 at 90 bar at temperatures from 40 °C to 220 °C.

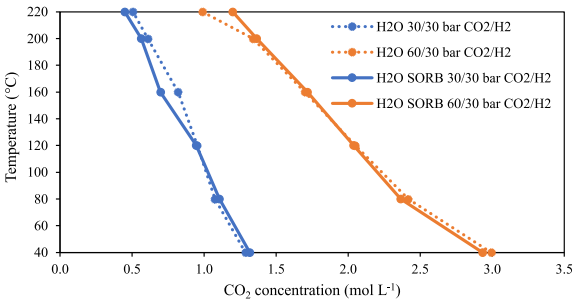


Figure 13. Temperature versus CO_2 concentration at constant pressure in the CO_2-rich phase of the ternary (H_2O–CO_2–H_2) and quaternary (H_2O–sorbitol–CO_2–H_2) systems.

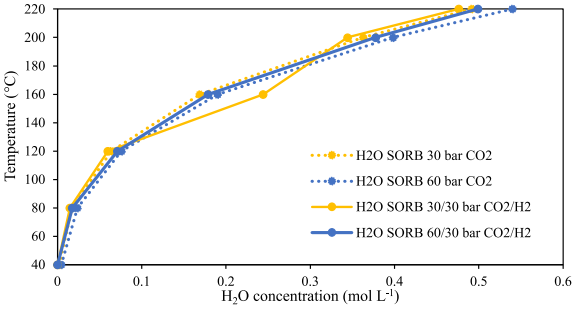


Figure 14. Temperature versus H_2O concentration at constant pressure and increasing temperature in the CO_2-rich phase of the ternary (H_2O–sorbitol–CO_2) and quaternary (H_2O–sorbitol–CO_2–H_2) systems.

4. Conclusion

The aim of this paper was to investigate, using in situ infrared spectroscopy, the thermodynamic behavior of the H_2O–sorbitol–CO_2–H_2 system which is involved in the one-pot catalytic reaction for hydrogenation of glucose to sorbitol and the acid-catalyzed dehydration of sorbitol by CO_2 [8,10]. Specifically, our study focused on the gas phase of binary, ternary and quaternary mixtures and selected vibrational modes of CO_2 and H_2O that have been analyzed in order to determine the evolution of the H_2O and CO_2 concentrations as a function of the temperature and pressure.

At a given pressure of CO_2, with increasing temperature, the concentration of water in the gas phase increases with a concomitant decrease of the CO_2 concentration when compared with neat CO_2. Addition of H_2 at a pressure of 30 bar to H_2O/CO_2 binary system lowered the concentration of both CO_2 and water in the gas phase. On the other hand, the addition of sorbitol to water appears to prevent the mutual solubility of CO_2 and water, an effect that is

significant only at high CO_2 pressure (120 bar) investigated in this study.

Finally, the purpose of these thermodynamic measurements was to determine whether sorbitol, present in the liquid water phase at high temperature, could dissolve in the gas phase and reduce the sugar concentration in water, thus affecting the catalytic reactions mentioned earlier. However, an inverse relationship was observed, where both sorbitol and H_2 limited the water solubility in the gas phase to some extent. Therefore, we confirm that under our experimental conditions (T < 220 °C and P < 120 bar), the ternary and quaternary mixtures display two distinct phases, sorbitol/water-rich liquid and CO_2/H_2-rich gas phase, both reactions of hydrogenation and dehydration taking place in the liquid phase.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

The authors acknowledge the Region Nouvelle Aquitaine and the University of Poitiers for their financial support (funding of the PhD of IB). The authors are also grateful to the program “Instrumentation in situ en conditions extrêmes” of the MITI of CNRS for financial support towards infrared and Raman equipment, the INCREASE Federation and the GDR 2035 SolvATE.

References

[1] H. Kobayashi, A. Fukuoka, *Green Chem.*, 2013, **15**, 1740-1763.
[2] Y. Zhu, M. Durand, V. Molinier, J.-M. Aubry, *Green Chem.*, 2008, **10**, 532-540.
[3] P. Tundo, F. Aricò, G. Gauthier, L. Rossi, A. E. Rosamilia, H. S. Bevinakatti, R. L. Sievert, C. P. Newman, *ChemSuschem*, 2010, **3**, 566-570.
[4] G. Flèche, M. Huchette, *Starch - Stärke*, 1986, **38**, 26-30.
[5] I. Bonnin, R. Mereau, T. Tassaing, K. De Oliveira Vigier, *Beilstein J. Org. Chem.*, 2020, **16**, 1713-1721.
[6] C. Dussenne, T. Delaunay, V. Wiatz, H. Wyart, I. Suisse, M. Sauthier, *Green Chem.*, 2017, **19**, 5332-5344.
[7] P. Barbaro, F. Liguori, C. Moreno-Marrodan, *Green Chem.*, 2016, **18**, 2935-2940.
[8] I. Bonnin, “Conversion catalytique de sucres en sorbitans : utilisation du CO_2”, Phd thesis, University of Poitiers, 2022.
[9] I. Bonnin, R. Mereau, K. De Oliveira Vigier, T. Tassaing, *Mol. Catal.*, 2023, **547**, article no. 113279.
[10] I. Bonnin, R. Méreau, T. Tassaing, K. De Oliveira Vigier, *ACS Sustain. Chem. Eng.*, 2024, **12**, 7276-7288.
[11] G. Herzberg, *Molecular Spectra and Molecular Structure: Infrared and Raman Spectra of Polyatomic Molecules*, D. Van Nostrand Company, Inc., Princeton, NJ, 1956, 273-277 pages.
[12] R. Oparin, T. Tassaing, Y. Danten, M. Besnard, *J. Chem. Phys.*, 2005, **122**, 94505-94512.
[13] R. Oparin, T. Tassaing, Y. Danten, M. Besnard, *J. Chem. Phys.*, 2004, **120**, 10691-10698.
[14] M. Buback, J. Schweer, H. Yups, *Z. Naturforsch.*, 1986, **41a**, 505-511.
[15] T. Tassaing, R. Oparin, Y. Danten, M. Besnard, *J. Supercrit. Fluids*, 2005, **33**, 85-92.
[16] H. G. Kjaergaard, B. R. Henry, H. Wei, S. Lefebvre, T. Carrington, O. S. Mortensen, M. L. Sage, *J. Chem. Phys.*, 1994, article no. 6228.
[17] R. Oparin, T. Tassaing, Y. Danten, M. Besnard, *J. Chem. Phys.*, 2005, **123**, article no. 224501.
[18] NIST, http://webbook.nist.gov/chemistry/ (accessed 09/2021).
[19] S.-X. Hou, G. C. Maitland, J. P. M. Trusler, *J. Supercrit. Fluids*, 2013, **73**, 87-96.
[20] M. A. Pigaleva, I. V. Elmanovich, Y. N. Kononevich, M. O. Gallyamov, A. M. Muzafarov, *RSC Adv.*, 2015, **5**, 103573-103608.
[21] R. Coulon, L. Galatry, J. Robin, B. Vodar, *J. Phys. Radium*, 1955, **16**, 728-729.
[22] H. L. Welsh, *J. Atmos. Sci.*, 1969, **26**, 835-840.
[23] I. Bonnin, R. Mereau, K. De Oliveira Vigier, T. Tassaing, *J. Mol. Liq.*, 2022, **360**, article no. 119455.
[24] C. Y. Tsang, W. B. Street, *Chem. Eng. Sci.*, 1981, **36**, 993-1000.
[25] J. O. Spano, C. K. Heck, P. L. Barrick, *J. Chem. Eng. Data*, 1968, **13**, 168-171.
[26] K. Bezanehtak, G. B. Combes, F. Dehghani, N. R. Foster, D. L. Tomasko, *J. Chem. Eng. Data*, 2002, **47**, 161-168.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 77-93
https://doi.org/10.5802/crchim.315

Review article

French Network on Solvation (GDR 2035 SolvATE)

Neural network potentials for exploring condensed phase chemical reactivity

Axel Gomez[#,a], **Miguel de la Puente**[#,a], **Rolf David**[a] **and Damien Laage**[*,a]

[a] PASTEUR, Department of Chemistry, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France
URL: https://www.chimie.ens.fr/recherche/laboratoire-pasteur/chimie-theorique/
E-mails: axel.gomez@ens.psl.eu (A. Gomez), miguel.delapuente@ens.psl.eu (M. de la Puente), rolf.david@ens.psl.eu (R. David), damien.laage@ens.psl.eu (D. Laage)

Abstract. Recent advances in machine learning offer powerful tools for exploring complex reaction mechanisms in condensed phases via reactive simulations. In this tutorial review, we describe the key challenges associated with simulating reactions in condensed phases, we introduce neural network potentials and detail how they can be trained. We emphasize the importance of active learning to construct the training set, and show how these reactive force fields can be integrated with enhanced sampling techniques, including transition path sampling. We illustrate the capabilities of these new methods with a selection of applications to chemical reaction mechanisms in solution and at interfaces.

Keywords. Chemical reactivity, Machine learning, Molecular simulations.

Funding. PSL OCAV (Idex ANR-10-IDEX-0001-02PSL), GENCI-IDRIS (Grants 2022-A0110707156, 2023-A0130707156 and 2024-A0150707156).

Manuscript received 31 January 2024, revised 17 April 2024, accepted 14 May 2024.

1. Introduction

The elucidation of reaction mechanisms is at the heart of chemistry. A mechanism identifies a molecular pathway involving a series of successive structures explaining the chemical reorganization caused by a reaction. Determining the mechanism, the order in which the elementary changes occur, and which of them causes the rate-limiting energy barrier is essential to guide the work of chemists to control reactivity and act on it, e.g., by catalyzing a reaction step. Organic physical chemists have achieved remarkable successes and determined the mechanisms of fundamental reactions such as nucleophilic substitutions and Diels–Alder reactions [1–4]. This was obtained via a combination of approaches, including for example, linear free energy relationships, substituent effects and isotopic effects on the rate constant [3, 4]. However, these experimental approaches provide only an incomplete and often ambiguous description of the mechanism. In addition, it is well appreciated that reaction mechanisms as drawn in chemistry textbooks are largely idealized and that reactions actually proceed along pathways which can significantly deviate from these simplified representations. An in-depth determination of reaction mechanisms, taking into account path variability, would therefore be desirable, but remains almost inaccessible experimentally.

[#] Contributed equally

[*] Corresponding author

Numerical simulations appear as a perfectly suited technique for the determination of reaction mechanisms, since they naturally provide the desired atomic resolution. However, simulations face multiple challenges when it comes to modelling chemical reactivity in condensed phases such as liquids, solid/liquid and vapor/liquid interfaces, and biochemical environments. We now describe four of these challenges.

The first computational difficulty is that chemical reactions typically involve the breaking and making of covalent bonds, which imply electronic structure rearrangements. The mechanical description of molecular energies provided by classical force fields with fixed bond patterns and fixed partial atomic charges is therefore not adequate, and the electronic structure should be determined along the reaction pathway by solving (approximately) the electronic Schrödinger equation at every step. Density functional theory (DFT) is nowadays a method of choice for chemists, since it offers an attractive balance between cost and accuracy [5].

However, a second difficulty arises from the large size of the systems of interest. Studying chemical reactivity typically requires very large molecular systems, since one should not only describe the reaction partners, but also the surrounding solvent molecules which impact the properties of the reagents. This implies that systems of at least hundreds or thousands of atoms should be included in the electronic structure calculation. While very good DFT functionals can be used for small molecular systems (with a few tens of atoms), larger systems induce very high computational costs.

A third challenge is that reaction equilibrium constants and reaction rate constants are governed by differences in free energies, respectively between reactant and product and between reactant and transition state, and not by potential energies. At ambient temperature, the entropic contribution to these free energies can thus be important, and determining these free energies requires an extensive sampling of the configurational space. This therefore imposes long (more than nanosecond-long) molecular dynamics simulations propagated with forces obtained from DFT calculations. This implies that the costly electronic structure calculations must be repeated millions of times. Pioneering studies using *ab initio* (i.e., DFT-based) molecular dynamics [6] have already provided valuable insight in a broad range of chemical reaction mechanisms, showing the potential of this approach (see, e.g., Refs [7–12]). However, they often face limitations in the system size and in the trajectory length, and therefore in the precision on the computed free energies. Systems of a few hundreds of atoms simulated for a few hundreds of picoseconds typically require millions of CPU hours on a single computer (i.e., more than a hundred years) and remain little practical despite the availability of parallel computing architectures.

The accessible simulation times thus remain much shorter than typical reaction times. It is therefore very unlikely that a trajectory initiated in the reactant state will spontaneously cross the reaction free energy barrier(s) to form the product during the available simulation time. Elucidating the reaction mechanism thus imposes to force the reactants to move along the reaction coordinate. However, this requires identifying the reaction coordinate and this is a fourth challenge. While chemical intuition can be used for simple reactions like acid/base proton transfers, for complex mechanisms determining the reaction coordinate is precisely what chemists try to do and for which the assistance of simulations is requested. Enhanced sampling methods have been proposed to identify pathways on complex energy surfaces [13–17], but they require extensive simulations, and have so far been mostly used with classical force field descriptions of the energies and forces (e.g., for protein folding) and very little for chemical reactions because of their prohibitive computational cost.

Recent developments in machine learning now provide a solution to all these challenges [18–20]. Among the many machine learning approaches that have been applied to interatomic potentials, deep neural networks have been particularly successful [18–20]. In this paper, we describe how these recent developments now allow training neural networks which provide the energies and forces traditionally obtained from an expensive resolution of the Schrödinger equation, and how they can be combined with enhanced sampling methods to determine the mechanisms of complex reactions in condensed phases. We introduce the structure of neural network potentials in Section 2, then we describe how they are trained in Section 3, and present applications to chemical reactions in Section 4. We finally

offer some concluding remarks in Section 5. We emphasize that our goal is to provide a tutorial review and we will often direct the reader to excellent, more specialized reviews on the specific aspects that will be discussed.

2. Neural network potentials

As described in the introduction, the first challenge in performing molecular simulations of a reactive system is to obtain the energies and forces. This can be done either "on-the-fly" or *a priori*.

In Born–Oppenheimer *ab initio* molecular dynamics [9,21], the Schrödinger equation is solved "on-the-fly" for the valence electrons at a given electronic structure level (typically DFT) to obtain the forces on the ions (formed by the nuclei and core electrons) in a given configuration and to propagate the trajectory to the next step. One advantage is that forces are only determined for configurations which are visited by the trajectory. However, a limitation is that forces must be evaluated at every step of the simulation, and the resulting computational cost has so far limited the size of the systems and the length of the trajectories. Reactive extensions of classical force fields [22,23] have also been used: because they are parametrized to describe some pre-defined chemical reaction, they do not require expensive electronic structure calculations. While these approaches have been successful for some combustion reactions and enzymatic reactions, they remain less accurate than *ab initio* approaches and can exhibit parameterization biases.

Another approach for the calculation of forces and energies relies on the *a priori* calculation and fitting of the potential energy surface (PES) on which the system is evolving. While this approach has been successfully used for gas phase reactions and spectroscopy [24], it was for some time limited to small systems, due to the complexity of fitting a highly multidimensional PES. However, the idea of *fitting* the PES of a molecular system around a reduced number of configurations of known energy (computed at the given reference level of theory) to gain access to entire regions of the PES with a reduced computational cost but comparable accuracy has known a growing interest for several decades [18,25–28]. The mapping of atomic configurations to the PES was initially performed through analytical functions [24,25], but this approach was limited to low dimensionality problems. Modern machine learning (ML) techniques are ideal tools to perform these mappings in more complex situations, and have been used since the early 2000s in the context of computational chemistry [18,27]. Two very successful ML tools to fit high-dimensional PES of molecular systems in the condensed phase are kernel-based regression methods (including, e.g., (symmetric) Gradient Domain ML [29,30] and Gaussian Approximation Potentials [31]) and neural network (NN) based methods [18,28,32–35]. The advantages and limitations of each family of methods have been discussed in great detail in excellent reviews [19,36]. Very briefly, the choice of a given method typically depends on the systems and objectives of the study. Kernel-based regression methods present a series of advantages: they tend to outperform NN-based methods in the low-data regime and the lower architecture complexity makes them more easily interpretable (noticeably, Gaussian Approximation Potentials [31] are able to provide analytical uncertainties on predicted quantities). However, NN-based methods are generally more efficient in the high-data regime and tend to present higher prediction efficiencies and scaling with, e.g., system size and number of chemical elements.

Here we focus on the applications of NN-based techniques to chemical reactivity in the condensed phase. Several approaches have been proposed, including, e.g., the SchNet architecture [34], ANI [37], and the DeepPot-SE scheme [20,32,33]. Most of them rely on the High-Dimensional NN (HDNN) concept introduced by Behler and Parrinello [18]. HDNN provides an elegant solution to one of the main difficulties in mapping the nuclear configurations of atoms to the PES with NN. In earlier implementations, because many of the PES symmetries (e.g., invariance upon translation, rotation, permutation of identical atomic species) were not directly encoded in the usual representations of configurations via Cartesian or internal coordinates, these invariances had to be learned directly by the NN [26] which limited their accuracy and scalability with system size. In HDNN, the total energy of the system is decomposed into atomic contributions $E = \sum_{i \in \text{atoms}} E_i$, which are predicted by a single NN per atomic species. This decomposition makes the energy invariant upon permutation of the coordinates

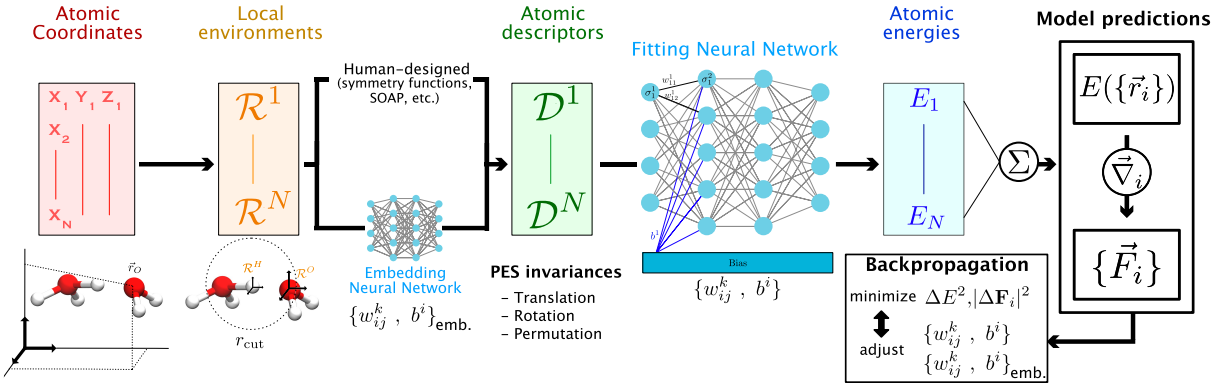


Figure 1. Schematic representation of a NNP workflow to map the atomic configurations of a system to the PES. In a given configuration, a local environment matrix is defined for every atom $\mathscr{R}^i$: the latter contains the chemical element and the positions in the atomic frame of reference of every neighboring atoms within a given cutoff distance. These local environments are transformed (either manually or through the use of embedding NN) into descriptor vectors $\mathscr{D}^i$, which preserve the invariances of the PES. These descriptors are then transformed into atomic energies E_i by fitting NN (one per type of atomic species), which yield the configuration energy $E(\{\vec{r}_i\})$ after summation. This energy can be analytically differentiated to obtain the atomic forces $\vec{F}_i$ needed to propagate MD. The prediction error on both of these quantities is used to adjust the parameters of the fitting (and, if needed, those of the embedding) network during backpropagation.

of identical atomic species and provides a NN potential (NNP) model (the set of NN for each individual species) that can in principle be used to study systems with an arbitrary number of atoms. The other invariances of the PES can then be either learned directly from the nuclear positions (generally in the local frames of reference of every atom) in what are sometimes called end-to-end NNPs [33], or encoded in human-designed *descriptors* of the atomic configurations including, e.g., symmetry functions [38]. This total energy decomposition assumes that the atomic energy depends only the atomic environment within a finite range, but long-range effects important for, e.g., electrostatics can be considered separately and we will return to this point at the end of this section. The general workflow of a NNP for the prediction of molecular energies and forces is presented in Figure 1.

In the following, we focus on the implementation of end-to-end NNPs provided in the Deep Potential - Smooth Edition (DeepPot-SE) model [33], within the DeePMD-kit package [20,32]. This choice is motivated by the successful applications of this approach to chemical reactions recently obtained in our group [39–41] and because it presents a good trade-off between prediction accuracy and computational efficiency [42]. However our discussion remains general since our conclusions could have been obtained with other families of end-to-end NNPs. This model has been presented in detail elsewhere [20,33], and here we only summarize its most important features, some of which are shared with other NNP families. The philosophy of the DeepPot-SE scheme is to limit human intervention as much as possible in the construction of the PES model. The complex relation between the local atomic environment and the energy is therefore learned through intermediary *embedding* NNs that map the coordinates of the system onto non-linear descriptors, produced as the output layer of the embedding NN. These descriptors are then provided as inputs to the *fitting* NNs that predict the atomic energies (see Figure 1). In order to minimize the amount of information that needs to be learned by the NNPs, the descriptors of every atom are usually built by the embedding NNs from the positions of the other atoms within its local frame of reference (which is already a translationally invariant representation). Furthermore, because the energy of a given atom is mostly determined by its local environment, only the

coordinates of atoms within a given cutoff distance (typically 6 Å) are used as input of the embedding NN. While this is an approximation to which we will return, this has the key advantage of limiting the computational cost and ensuring that the model can be used for systems of arbitrarily large sizes.

The parameters of both NNs (called *weights* and *biases*) are simultaneously optimized during the *training* process: this consists in the minimization of the following loss function L defined as the sum of mean square errors (MSE) of the NNP predictions on several target quantities:

$$L(p_E, p_f) = \frac{p_E}{N}\Delta E^2 + \frac{p_f}{3N}\sum_i |\Delta \mathbf{F}_i|^2 \qquad (1)$$

N is the number of atoms in the system and ΔE^2 and $|\Delta \mathbf{F}_i|^2$ are the MSE on the energy and on the atomic forces. An additional term involving the system virial can be introduced in the loss function, and is typically employed for solid state systems. Since in practice it has little impact for reactive models in the liquid phase we remove it here for clarity reasons. The prefactors p_E and p_f determine the relative importance of each target quantity; they are chosen by the user and typically evolve during the minimization procedure following the adaptive *learning rate* [43]. During this minimization, the MSE are computed from the differences between the target quantity predicted by the NNP and the reference value for a random subsample of configurations (often called a batch) and the *weights* and *biases* of both NNs are modified by *backpropagation*. The atomic forces are computed by differentiation of the predicted total energy with respect to the atomic coordinates, which imposes that this quantity vary smoothly enough to always be differentiable. This is enforced in the Smooth-Edition scheme by applying weighting functions (e.g., cosine [32] or polynomial [20] functions) to the radial components of the local atomic coordinates that are used to build the descriptors in the embedding step.

In practice, despite the end-to-end philosophy of the DeepPot-SE scheme, many hyperparameters of the model need to be chosen by the user. These include, e.g., the number of layers and neurons of both NN, the prefactors of the target quantities in the loss function, the batch size, and the total number of training iterations. Systematic grid search procedures can be used to determine these hyperparameters. However, this can require large amounts of validation data. Fortunately, the NNPs performances have been found to remain robust over a reasonable range of these hyperparameters. This implies that it is often possible to use similar sets of hyperparameters for training NNPs on different chemical systems.

We quickly present typical sets of hyperparameters that we have successfully used in our group to train NNPs for reactive systems in the liquid phase. For the NN architectures, three hidden layers of some tens of neurons for the embedding network and four hidden layers of some hundreds of neurons for the fitting network are enough to yield NNPs of the same quality as that obtained with larger architectures which induce larger computational costs. For the prefactors of the loss function components, we have found that using a much larger prefactor for the atomic forces than for the energies at the beginning of the training procedure typically yields NNPs that are more stable during the MD simulations. We typically employ p_f and p_E prefactors that range from approximately 1000 and 0.01 respectively at the beginning of the training procedure to 1 and 0.1–1 at the end. The total number of training iterations can be set between some hundreds of thousands to some millions depending on the training set size (the number of reference configurations used for the training) and batch sizes below 5 all give comparable NNPs. Finally, the learning rate used for the parameter optimization is adapted during the training procedure and we have found that limiting values of 1×10^{-3} and 1×10^{-8} at the beginning and at the end of the training procedure give satisfactory results.

As described above, the DeepPot-SE scheme corresponds to what are sometimes called "second generation" NNPs [36], which currently are the most commonly used NNPs to study reactivity in the liquid phase. The main limitation of this NNP generation is the absence of explicit long-range effects in the construction of the descriptors, which can lead to unphysical behaviors in situations where these effects can significantly affect the energy of a configuration (most noticeably at interfaces in the presence of charges). However, we stress that long-range interactions are present in the reference energies and forces used for the training, and are therefore described in a mean-field manner. Adding configurations with long-range interactions to the training set can be enough to correctly reproduce the physics of

such systems, as these long-range effects can be indirectly learned by the NNs [44]. Long-range effects can also be explicitly included in the NNP description of the system in what are called "third generation" NNPs [36] (which have been implemented with the DeepPot-SE model in the DeePMD-kit package [45]). However, these have so far had limited applications to realistic reactive systems and we will therefore focus on examples of applications of the "second generation" DeepPot-SE models.

3. Training set construction

Neural network potentials exhibit high accuracy in predicting energies and forces for configurations closely resembling those in their training set. However, deviations from these trained configurations lead to a rapid deterioration in the prediction accuracy of the neural network model—a challenge known as generalization performance [43]. The careful selection of configurations to be included in the training set is therefore critical in determining the quality of the NNP. This section outlines systematic approaches aimed at enhancing the accuracy of NNP for structures encountered in chemical reactions, with a specific focus on active learning procedures that iteratively refine NNP accuracy.

Identifying and generating relevant chemical structures for the training set is a formidable task, due to the overwhelming dimensionality of the chemical space. Several approaches can be followed. The first one is that adopted by several on-going efforts to generate general-purpose NNPs aimed at providing a universal force-field [37]. The training sets of these NNPs are typically parts of existing very large structure datasets of realistic synthesizable molecules. One such example is the publicly available GDB-11 dataset [46], which contains all molecules made of up to 11 atoms of carbon, nitrogen, oxygen or fluorine and includes 26.4 million distinct structures. However, this approach is not practical for reactivity in condensed phase for two main reasons. The first one is that solvent arrangements need to be included in the configurations, which considerably increases the possibilities to be considered. Different strategies have been proposed to overcome this limitation and generalize NNPs trained on small systems to be transferable to larger molecular systems, including the condensed phase [47–49]. For example, in the same fashion as hybrid QM/MM techniques, hybrid NNP and molecular mechanics (NNP/MM) simulations are being developed to treat a solute by machine learning models while the solvent is described with a classical force-field [50,51]. However, these setups cannot easily describe reactions that involve solvent molecules. The second very important difficulty arises because the study of chemical reactivity requires to include high-energy transition-state geometries, which are not present in databases of stable structures.

Simulating chemical reactions in condensed phases therefore requires the specific training of the NNP on configurations that are representative of the region of phase space explored during the reaction. The training set should contain all kinds of structures that will be encountered during the reactive simulations, including typical configurations of the reactants, transition states and products, but also intermediate structures visited along the reaction coordinate.

The *a priori* identification of all relevant configurations to be included in the training set, e.g., from an *ab initio* MD simulation, is usually not possible. One therefore uses an iterative approach to progressively enrich the training set and systematically improve its prediction accuracy. Such an approach was for example implemented in the deep potential generator (DP-GEN) [52].

We now describe this active learning strategy combined with a query by committee [53] estimation of the NNP quality. As described in Figure 2A, one starts from a small training set including configurations that are representative of limited regions of the phase space, typically stable reactant configurations easily sampled in short DFT-based MD simulations. A few hundreds of configurations with energies and forces at a selected reference level of theory (DFT with a hybrid functional, for example) typically constitutes a valid initial training set. Here it is important to select uncorrelated structures in order to maximize the amount of information present in the dataset. Based on this initial set of configurations, a first approximate NNP can be trained.

The next step is to discriminate between configurations whose forces and energies are properly predicted by the NNP and configurations which are unknown and poorly described. Contrasting the NNP prediction with reference DFT calculations would be

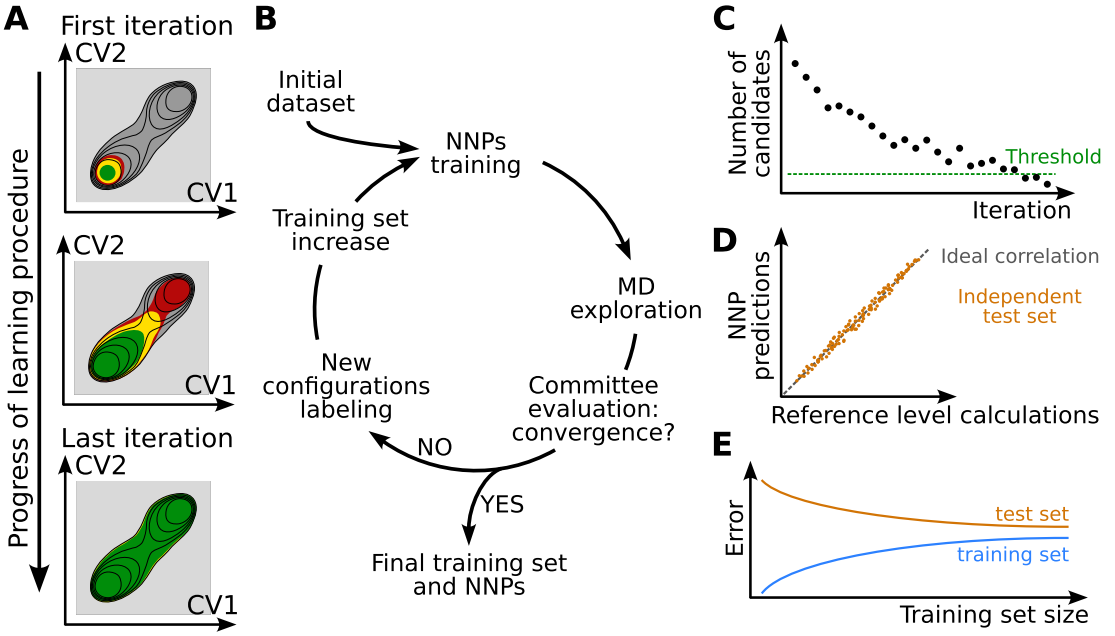


Figure 2. Concurrent learning scheme. (A) Iterative exploration of the phase space until convergence. The relevant region of phase space for the chemical reaction of interest is in dark gray. The black lines represent schematically a free energy surface. The part of configurational space known to the NNPs at each iteration is in green, the intermediate region that provides new structures to be added to the training set is in yellow. The red area corresponds to regions where the NNPs extrapolate significantly but that may be accessible during the exploration phase. (B) Training set construction algorithm. (C) NNPs committee evaluation test. The number of candidates, i.e. configurations selected to be labeled and added to the training set, decreases over iterations. A threshold value can be decided as an interruption criterion. (D) Correlation between the NNP predictions and the reference calculations, for example for energies or forces. (E) Schematic representation of a learning curve where the error on the training set and the test set are represented as a function of the number of data used for training.

extremely expensive. One therefore uses the query by committee approach [53] to estimate the quality of the prediction provided by the NNP. Several distinct NNPs are trained on the same training set but initialized with different random parameters. After training, their predictions all converge to a similar answer in regions where they are properly optimized, but significantly differ for configurations far from these in the training set. This deviation between NN predictions is therefore exploited to identify which configurations are already known by the NNP model and which ones are unknown and should be added to the training set. In our studies, three or more NNP models are typically trained on the same training set at each step of the iterative training procedure.

An additional advantage of this concurrent learning approach is that it greatly mitigates the risk of instabilities recently found [42] for models trained with different approaches and for which the average values of different observables (including, e.g., radial distribution functions) fluctuate unphysically during the molecular dynamics trajectory. The iterative nature of the training set construction, guided by collective variables based on chemical knowledge, ensures that the chemical space accessible to the system is sampled systematically, which greatly reduces the risks of encountering configurations leading to instabilities during production. Furthermore, monitoring the deviation between the committee of NNPs during the molecular dynamics trajectory allows an "on-the-fly" detection of these instabilities, which

might not be directly detectable from the model performance over a possibly incomplete test set.

At each iteration of the active learning, the ensemble of NNPs is then used to propagate short MD trajectories. The goal is to explore the phase space around the region of the training set, in order to identify configurations where the NNP prediction is poor and which should thus be added to the training set. The maximal deviation on the atomic forces computed by all models serves as a proxy to measure whether the configurations are already known or should be added to the training set to improve it. MD configurations that are very similar to a subset of the training set will exhibit a low maximal deviation on forces, typically of the order (or below) of 0.1 eV·Å^{-1}. Frames containing new structures will be associated to a large maximal deviation on forces. New configurations to be added to the training set are usually selected at the border between the region already included in the training set (Figure 2A, green region) and the unknown (see Figure 2A, yellow region). This corresponds to intermediate values of maximal deviation on forces, typically between 0.1 and 0.8 eV·Å^{-1}. In this region, the NNP model is not accurate but it does not completely fail or break. In the red area of Figure 2A, the committee of NNPs gives a very large maximal deviation. This corresponds to configurations that, although accessed during the exploration phase, might be highly non-physical and where the NNPs cannot be trusted at all. The new structures with intermediate deviations, called *candidates*, are usually sub-sampled to ensure a proper decorrelation, before being labeled at the same reference level as the training set and added to it. The improved dataset is used for training new NNP models for the next iteration of the procedure (Figure 2B). As illustrated schematically in Figure 2A, each iteration pushes further the limits of the region of configurational space that is properly described by the NNP.

A key feature of the training set is that it should cover all the structures to be visited during the subsequent simulations. For chemical reactions, this implies that unbiased MD simulations during the active learning iterations are not adequate. Typical barriers between the reactant and product basins are significantly larger than the thermal energy $k_B T$, where k_B is the Boltzmann constant and T the temperature, and the transition state region will not be explored spontaneously. The active learning explorations are therefore performed with enhanced sampling techniques, including, e.g., metadynamics [13], on-the-fly probability enhanced sampling (OPES) [17] or umbrella sampling [54], or at elevated temperatures to enhance the sampling. The choice of the enhanced sampling technique—and of the relevant collective variables when applicable—is crucial for an efficient exploration step for cases where the free energy landscape features larger than thermal energy barriers. In addition, if the NNP is developed for path integral MD (PIMD) [55,56] simulations for an explicit description of nuclear quantum effects, the training set should include typical configurations of the path integral beads, since such simulations evaluate the forces on the beads. Alternatively, nuclear quantum effects can also be learned directly in the framework of machine-learned centroid molecular dynamics schemes [57].

The iterative enrichment of the training set is stopped when the quality of the NNP is considered to have converged. The number of configurations that are retained to be labeled and added to the training set after an exploration phase is a good indicator of the convergence. If less than a user-specified number of structures are pointed out by the committee of neural networks, then it can be assumed to be converged (Figure 2C). A converged concurrent learning process can generate thousands or tens of thousands configurations for the training set, depending on the chemical complexity of the system.

One should then validate the quality of the NNP. There is no absolute condition to assess the reliability of the optimized NNP but a collection of metrics that confirm the quality of the model is usually deemed satisfactory [58]. The NNP validation is assessed on an independent test set of configurations. The latter includes configurations sampled within the region of configurational space of interest and whose energies and forces are labeled at the reference level of theory. This validation set is a separate dataset that is not included in the training set. The NNPs are used to predict energy and forces of the configurations in the test set and the predictions are compared with the reference values (Figure 2D). The correlation is considered to be acceptable when the average prediction errors on forces are typically below 0.1 eV·Å^{-1}. This is complemented by the analysis of the learning curve, which represents a measure of the relative error, e.g., the root-mean squared NNP error on the energy or

on the forces divided by the standard deviation of the reference data, both on the training set and on the independent test set as a function of the size of the training set used (Figure 2E). For small training sets, prediction errors are typically much lower on known configurations (included in the training set) than on unknown ones (such as those in test sets), which is known as *overfitting* [36]. This difference of performance usually results from an imbalance between the model complexity (featuring too many optimizable parameters) and the amount of information present in the training set. It can thus be monitored while the training set size increases by comparing the prediction errors over the training set and over an independent test set. As the amount and diversity of data included in the training set increase, the gap between the prediction errors on training and test sets should decrease (see Figure 2E). This is typically referred to as a decrease in *overfitting* or, equivalently, an improvement of the *generalization performance* of the model.

Practically, the concurrent learning procedure depicted in Figure 2B is performed via a combination of software: one to train the NNP, for example DeePMD-kit [32], a MD engine interfaced with the NNP software to propagate MD simulations, which can be LAMMPS [59] for classical nuclei MD simulations or i-PI [60] for PIMD, optionally coupled with PLUMED [61] for enhanced sampling, and an electronic structure code for energy and forces calculation at the chosen reference level of theory, e.g., CP2K [62]. A series of scripts are needed to prepare the files and extract the configurations to be added to the training set. For some simple applications, this procedure has been implemented in the DP-GEN software [63]. For more complex reactions which require, e.g., several enhanced sampling techniques or the explicit consideration of nuclear quantum effects, we have developed our own suite of scripts (https://github.com/arcann-chem/).

4. Applications of NNP for reactivity

4.1. *Brief overview of recent applications*

The application of NNPs to chemical reactivity in solution and in materials is expanding extremely rapidly. For example, they have been successfully used to study proton transfer reactions, including water self-dissociation in the bulk [64] and at the air–water interface [40] and proton transport in porous materials [65]. NNPs have also been applied to heterogeneous catalysis reactions, including water dissociation on TiO_2 surfaces [66–69] and ammonia decomposition [70]. They have been used to investigate the mechanisms of chemical reactions, such as Diels–Alder cycloadditions [71], prebiotic amino acid synthesis [72], urea decomposition [73], and different reactions relevant to atmospheric chemistry [74,75]. They have even been extended to the very complex reaction schemes involved in combustion [76,77]. While we cannot provide a comprehensive review of these applications, in the following we illustrate the capabilities and limitations of this method for a number of reactions that have been studied in our group.

4.2. *Application to a simple reaction*

We first describe how NNPs have been applied to study the dissociation of formic acid in water and at the air–water interface [39]. Extensive experimental and simulation studies had concluded that simple acids including formic acid experience a change in their acidity between the bulk solution and the air–water interface. This question attracted an important interest because of the major implications for the description of chemical reactions at the surface of atmospheric aerosols. However, both experiments and simulations led to contrasted results, reporting both acidity enhancement and reduction for the same acid, depending on the technique being used.

For simulation studies, the challenge is twofold. First, the simulations have to be able to describe the dissociation reaction, so that DFT-based molecular dynamics simulations would be typically used. Second, in order for the comparison between the acidity in the bulk and at the surface to be conclusive, the uncertainties on the acid dissociation free energies should be much smaller than the difference between the dissociation free energies in the bulk and at the water-interface. This requires the careful convergence of these calculations, and therefore long simulations. However, the computational cost of DFT-based simulations typically imposes strong limitations on the trajectory length, and thus on the precision of the calculated free energy difference.

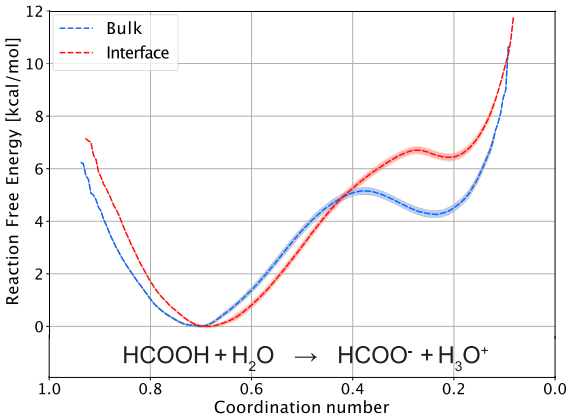


Figure 3. Formic acid dissociation free energy profile in the bulk (blue) and at the air–water interface (red) along the hydrogen coordination number around the acidic oxygen atom of formic acid [39].

We therefore trained a NNP to describe the dissociation of formic acid both in the bulk and at the air–water interface, at the BLYP-D3 level of theory. The training set was constructed iteratively as described above and includes both bulk and interface environments, and configurations at different stages of the dissociation reaction. Here the reaction mechanism is known and can be described with a simple reaction coordinate, that we chose to be the hydrogen coordination number around the acidic oxygen atom of formic acid, which decreases from 1 in the acidic form to 0 in the conjugate base formate ion. Biased sampling along this coordinate was used to generate training set configurations. We note that while the NNP does not include long-range interactions because of its short-range cutoff, long-range interactions are accounted for in the reference energies and forces calculated for the training set configurations, so that long-range interactions between charged species for example are described by the NNP in a mean-field manner.

Typical dissociation reaction free energy profiles are shown in Figure 3 in the bulk and at the air–water interface. The thickness of each curve reports the 95% confidence interval on the free energy, which shows the excellent convergence provided by our long simulations. These calculations could unambiguously reveal that formic acid dissociation is less favorable at the air–water interface than in the bulk, in agreement with vibrational sum frequency generation spectroscopy and with X-ray photoemission spectroscopy [39]. We could then use these dissociation free energy profiles as the starting point of a model to predict the dissociation free energy based on the respective solvation free energies of the reaction partners (formic acid and water reactants, formate ion and hydronium products). This approach was also extended to the dissociation of the strong inorganic nitric acid in the bulk and at the interface [39]. In a subsequent study, we followed a similar approach to study the self-dissociation of water which governs the pH of neat water, and we showed how it is affected at the air–water interface, and how this determines the pH of small aqueous aerosols [40].

4.3. *Application to a more complex reaction*

While the aforementioned examples highlight the potential of NNPs to study chemical reactivity in the bulk and at interfaces, the cases described so far involve simple single-step reactions, with a well-defined reaction coordinate. However, many reactions of interest are far more complex. They can involve multiple intermediates and transition states, a series of concerted or sequential steps, and several competing pathways leading to the same or to different products. Well-known examples of such important but complex reactions are the Wittig reaction, the aldol condensation, and the Michael addition, for which changing the conditions can lead to different products, and the S_NAr reactions and amine protection–deprotection reactions which involve multiple steps [78].

The importance of the proper identification of the reaction coordinate is illustrated by the model free energy surface in Figure 4A. For the reaction transforming the reactant **A** into the product **B**, two transition states $\mathbf{TS}_1$ and $\mathbf{TS}_2$ are revealed when the free energy surface is represented along the two collective variables (CV1 and CV2). However, if the CV2 coordinate is ignored, these two pathways cannot be distinguished and the two transition states will appear as a single saddle point in the one-dimensional profile along CV1 (Figure 4B). If the two transition state structures have different entropic properties, they will be affected differently by temperature, changing

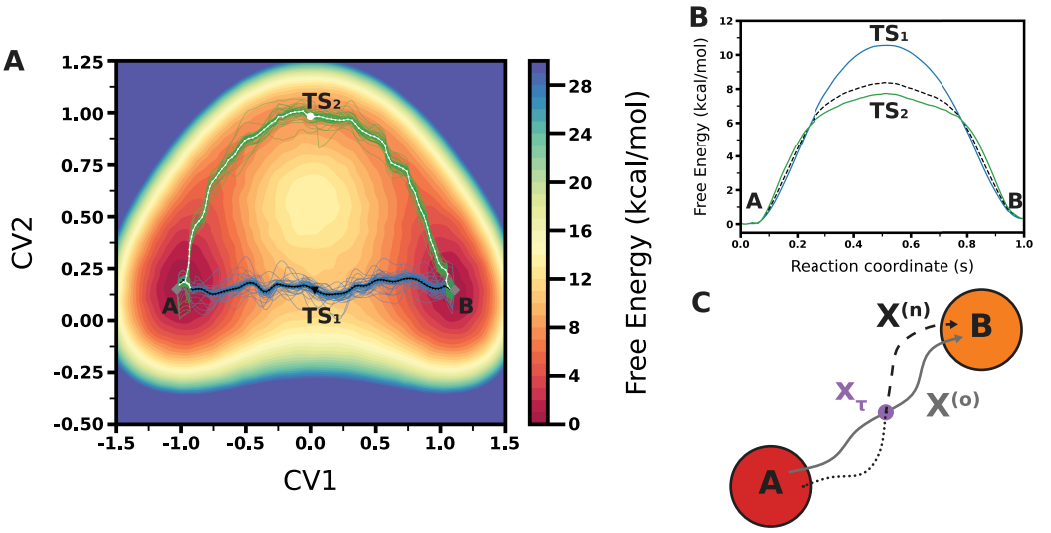


Figure 4. (A) Model free energy surface along CV1 and CV2. The gray square represents the two minima (A and B). The two minimum free energy profiles are represented with black and white solid lines. The two transition states are represented by a black triangle and a white square. The path ensembles are represented by the blue and green lines. (B) Free energy profiles of the same model free energy surface represented along the *s* reaction coordinate for both the blue and green paths and along CV1 when CV2 is not considered explicitly (dashes). (C) Scheme of a shooting move in Transition Path Sampling, with the initial path (solid gray line), the shooting point (purple circle), the new path (dashes for the forward time propagation and dots for the backward time propagation).

the relative importance of the two pathways, and this could not be understood by considering the profile along CV1 only. In addition, a proper sampling of the two transition state structures is likely difficult if the CV2 coordinate is not considered explicitly.

A number of methods have been proposed to identify reaction pathways on complex energy surfaces. When a very limited number of coordinates is to be considered, traditional multidimensional free energy surface calculations using umbrella sampling or metadynamics can be used, and the pathway can be determined *a posteriori*. When a larger number of coordinates is involved, these methods become impractical because of the necessary computational effort, and other methods which do not require computing the full dimensional free energy hypersurface have been proposed. An attractive approach is provided by the adaptive string method (ASM) [79] which focuses on the minimal free energy path (MFP) and determines the free energy surface only in the vicinity of the reaction path. Based on the string method [80], this method considers a string that spans from A to B. This string is then optimized iteratively, moving on the free energy surface towards the MFP. When convergence is reached, the string is then used to identify the path collective variables (parallel *s* and transverse *z*) [81] corresponding to the MFP, to determine the reaction pathway, and to calculate the free energy profile with the reaction transition states and intermediates. While this method can consider a large number of collective variables (typically from 2 to 10), it remains critical for the quality of the final result to include all relevant coordinates. Considering the model surface in Figure 4A, both CV1 and CV2 should be considered to identify the two pathways, and a string defined with CV1 but not CV2 would probably only identify the blue reaction path.

It is therefore interesting to consider other methods which do not require any *a priori* selection of coordinates in the generation of reaction pathways. One such very popular method is transition path sampling (TPS) [14,15,82–84]. TPS only requires a set of conditions that define the starting state A and

the ending state B, together with one pathway connecting A and B. The latter does not need to be the minimum free energy path, and it is typically generated by high temperature or biased simulations. Using a Monte Carlo approach, TPS then iteratively generates the statistical ensemble of pathways connecting A and B. TPS is described in detail in excellent reviews [14,15,84,85] and we only summarize its simplest form. The initial path ($X^{(o)}$) is discretized in L snapshots, x_i, with $0 \leq i < L$. One snapshot $x_\tau^{(o)}$ is selected, and a shooting move is performed by propagating a new trajectory forward and backward in time from this snapshot. One considers if this new trajectory connects A and B, and it is accepted according to a Metropolis criterion. One then performs another iteration, selecting a snapshot either from the new path ($X^{(n)}$) if it has been accepted or from the old path ($X^{(o)}$) if it has been rejected, until an adequate number of decorrelated paths are obtained. TPS has been further extended to more complex situations [85–89]. TPS therefore provides the ensemble of reaction pathways, with the important advantage that they are generated from unbiased trajectories and thus with the true dynamics. The transition path ensemble can then be *a posteriori* projected on a selection of collective variables (CVs) to identify the most relevant ones and define a simplified reaction coordinate (but this choice does not affect the transition path generation). It can further be used for a committor analysis, to identify the transition state location, and the ensemble of pathways can be used to determine the free energy profile [90]. We note that other related techniques including transition interface sampling [91,92] and forward flux sampling [93,94] are also available to explore the free energy landscape (see, e.g., the review in [95]).

TPS has been extensively used for biophysical phenomena occurring on rough free energy surfaces, including protein folding, but it remains a computationally demanding technique due to the very large number of trajectories required. While TPS has been successfully applied to some chemical reactions [96–98], the cost of propagating these many trajectories with a description that accounts for chemical reactions, i.e., typically DFT-based MD, has so far hindered its broad application to reactivity.

Combining TPS and NNP therefore provides a promising approach to explore chemical reactivity in complex systems. In the following, we describe a study performed in the group [41] where we used the TPS implementation of OpenPathSampling [99,100] together with the OpenMM [51] MD engine interfaced with DeePMD-kit [20,32] and the biased sampling tools provided by PLUMED [61].

We present the application of this method to study the formation of a peptide bond. The condensation of an amine and a carboxylic acid to yield an amide is a key step in the formation of polypeptides, since the latter are polymers of amino acids linked together by peptide bonds. In addition, amide formation is one of the most common reactions in organic synthesis. In bulk aqueous solution, this reaction is extremely unfavorable. In living organisms, it is catalyzed by enzymes and in organic synthesis it is typically performed after activating the reactants, for example by replacing the carboxylic acid with an acyl chloride or an acyl anhydride. Understanding the mechanism of this reaction is therefore of great interest for organic synthesis, but also in the context of prebiotic chemistry.

The reaction involves several rearrangements: the formation of the C–N bond, the breaking of the C–O bond, and at least two proton transfers (see Figure 5). The sequence of these steps is not clear, and it has been a subject of debate since a mechanism (see Figure 5A) was proposed by Jencks in the late sixties [101,102]. The mechanism is considered to start with the formation of the C–N bond, yielding a tetrahedral species ($T^{\pm}$) which has been proposed as either a long-lived intermediate or a short-lived transient structure [101–104], followed by the cleavage of the C–O bond. Numerous theoretical studies have been published on this reaction, using a variety of methods ranging from gas-phase DFT calculations to QM/MM simulations, in order to investigate the mechanism and the role of the solvent [103–108].

We first describe the training of the NNP. Following the active learning approach described above, we have constructed our training set iteratively, with reference energy and force calculations performed at the DFT-GGA level of theory (BLYP-D3). The final training set contains an ensemble of approximately 75,000 diverse configurations encompassing reactants, products, intermediates, and transition states, all solvated in a cubic box with 300 water molecules [41]. These reactive structures were sampled via enhanced sampling using multiple collective variables to describe the reaction, including

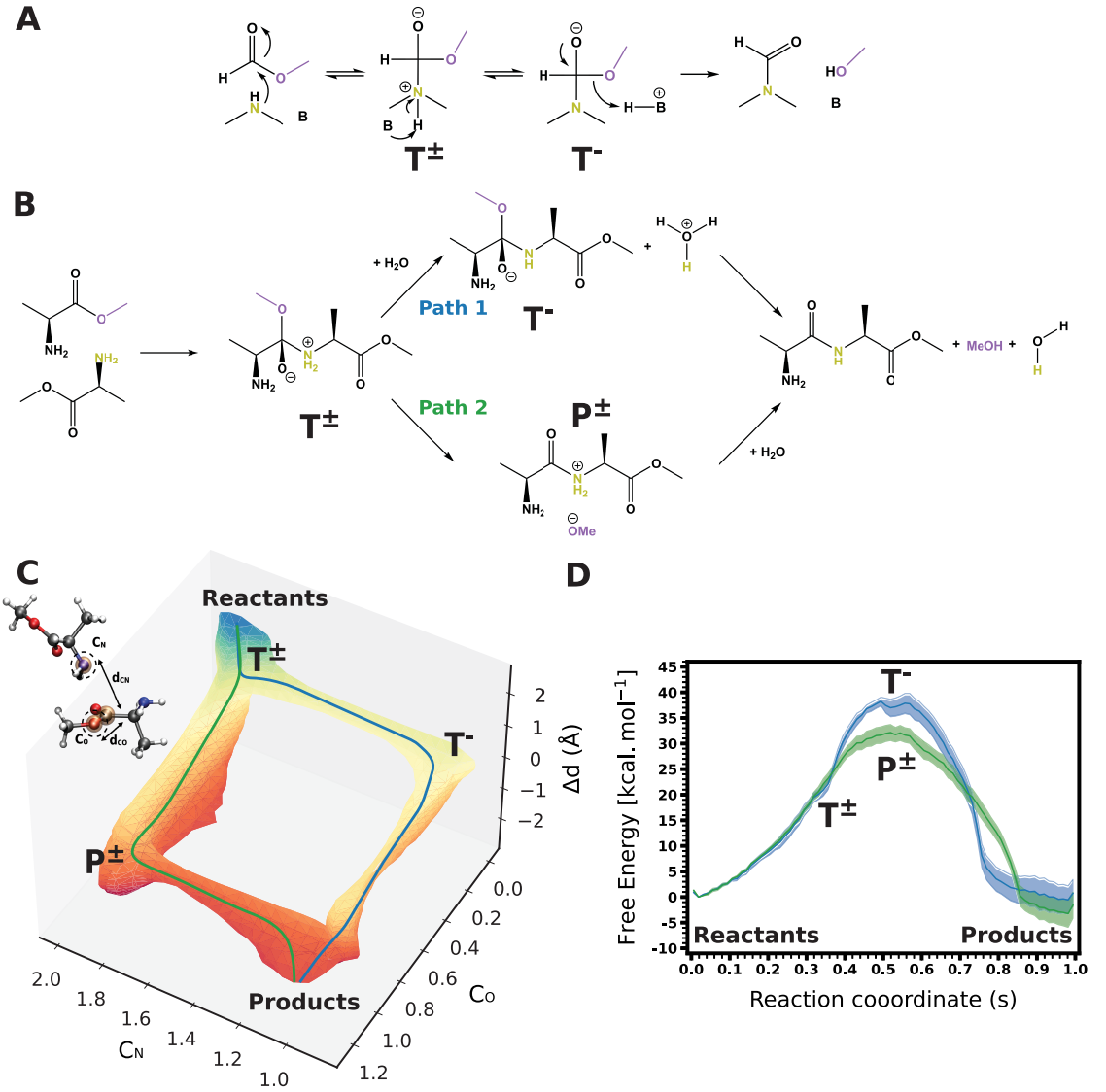


Figure 5. (A) The amide formation mechanism proposed by Jencks for the aminolysis of acetate ester (with the nucleophilic amine in olive and the leaving group in violet). (B) The two pathways for peptide bond formation identified in our calculations: the general-base catalysis mechanism similar to Jencks' proposal (path 1) and mechanism with no acid-base catalysis which is the most favorable one in pH-neutral conditions (path 2). The same color code is used for the reacting groups as in (A). (C) Transition path ensemble probability obtained from our TPS results for the peptide bond formation as a function of three CVs, respectively the $\Delta d = d_{CO} - d_{CN}$ difference of bond lengths, and the $C_{O,N}$ hydrogen coordination numbers around O and N where the path 1 is represented in blue and path 2 in green. (D) Free energy profiles for the two pathways obtained by umbrella sampling along the path collective variable of each mechanism, represented with the same color code as in (C).

$\Delta d = d_{CO} - d_{CN}$, the difference in the lengths of the forming C–N bond (d_{CN}) and of the breaking C–O bond (d_{CO}), and $\Delta C = C_N - C_O$, the difference in coordination numbers between the nucleophilic amine nitrogen and the leaving group oxygen with any hydrogen atom, which are "chemically intuitive" variables. Several additional collective variables, such as the protonation states of the solvent, were also used to ensure an even training of the NNP on all mechanisms proposed in the literature.

After the NNP was trained, TPS was used to generate the transition path ensemble and to determine the reaction pathway, without any variable-induced bias. The ensemble of reactive trajectories consists of more than 50,000 trajectories, for a total of ≈0.5 μs. These trajectories were then projected along several CVs to visualize the most probable "reaction tubes". The results in Figure 5C revealed two very distinct pathways [41]. The relevant CVs were the difference between the formed and cleaved bond lengths Δd, the C_N labile hydrogen coordination of the nucleophilic nitrogen (of the attacking amine) and the C_O labile hydrogen coordination of the oxygen of the ester. We subsequently constructed a path-collective variable based on these four coordinates (d_{CN}, d_{CO}, C_N and C_O) for each of the two paths, and we determined the free energy profile along each pathway using umbrella sampling (Figure 5D).

The first mechanism (Path 1, Figure 5B) starts with the nucleophilic attack of the amine, forming a transient tetrahedral species ($T^{\pm}$), followed by the transfer of a proton from the amine to the solvent, thereby forming the transition state (T^{-}) and H_3O^+. The succession of these two steps is the rate-limiting step of this mechanism. The reaction then evolves barrierlessly with the departure of the leaving group and its subsequent protonation by H_3O^+, leading to the formation of the products. This general base-catalyzed mechanism resembles the mechanism previously proposed by Jencks [101,102], with the notable difference that both tetrahedral species $T^{\pm}$ and T^{-} are not long-lived intermediates. However, our calculations revealed that it is not the most favorable under neutral pH conditions.

The second mechanism (Path 2, Figure 5B), previously unreported, was found to be the most favorable at neutral pH, as seen in the free energy profile along the reaction coordinate (Figure 5D). It does not involve any acid or base catalysis. The rate-limiting step only involves the nucleophilic attack of the amine, followed by the departure of the leaving group forming the transition state $P^{\pm}$, with a strong ionic pair character. This is followed by proton rearrangements, which occur without any barrier. The overall free energy barrier is consistent with experimental rate constants for similar uncatalyzed amino acid condensations [109].

The presence of these two competing mechanisms provides an explanation to the experimental Kinetic Isotope Effects (KIE) measurements [110] conducted in different pH conditions. Previous proposals had suggested that the mechanism does not change with pH but that the two successive barriers (before and after the proposed long-lived tetrahedral intermediate) are affected differently by pH. In contrast, our calculations show that two separate single-step mechanisms co-exist and that pH affects their barriers differently: the mechanism without acid-base catalysis via $P^{\pm}$ is favored at moderately basic pH while the base-catalyzed mechanism via T^{-} is expected to be increasingly favored as pH is increased [41]. These results therefore show that combining NNP and TPS can provide a powerful method to explore reaction mechanisms.

5. Concluding remarks

In this tutorial review, we have presented a brief overview of the significant potential offered by neural-network potentials in simulating chemical reactions in the condensed phase. We have described their training process and their combination with enhanced sampling techniques, such as transition path sampling, to investigate complex reaction mechanisms. This field is growing extremely rapidly. On-going developments include the advancement of next-generation NNPs including long-range interactions for example for an improved description of electrostatics [36], the implementation of Δ-learning to facilitate the training and increase the range of accessible accuracies [111], embedding schemes combining ML potentials with classical molecular mechanics (MM) force fields for mixed ML/MM methods [112] equivalent to hybrid quantum mechanics/molecular mechanics (QM/MM) schemes, the combination of NNPs with neural networks to learn observables relevant for molecular spectroscopy [113], and the application of

deep learning to identify relevant reaction coordinates [114–116].

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Dedication

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was supported by PSL OCAV (Idex ANR-10-IDEX-0001-02PSL) and HPC allocations from GENCI-IDRIS (Grants 2022-A0110707156, 2023-A0130707156 and 2024-A0150707156).

References

[1] E. D. Hughes, C. K. Ingold, *J. Chem. Soc.*, 1935, 244-255.
[2] R. B. Woodward, *J. Am. Chem. Soc.*, 1942, **64**, 3058-3059.
[3] T. H. Lowry, K. S. Richardson, *Mechanism and Theory in Organic Chemistry*, Benjamin-Cummings Publishing Company, New York, NY, 1987.
[4] E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, Sausalito, CA, 2006.
[5] N. Mardirossian, M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315-2372.
[6] R. Car, M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471-2474.
[7] P. Carloni, U. Rothlisberger, M. Parrinello, *Acc. Chem. Res.*, 2002, **35**, 455-464.
[8] P. Minary, M. E. Tuckerman, *J. Am. Chem. Soc.*, 2004, **126**, 13920-13921.
[9] R. Iftimie, P. Minary, M. E. Tuckerman, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 6654-6659.
[10] D. Marx, *ChemPhysChem*, 2007, **8**, 209-210.
[11] O. Isayev, L. Gorb, M. Qasim, J. Leszczynski, *J. Phys. Chem. B*, 2008, **112**, 11005-11013.
[12] L. Pierce, P. Markwick, J. McCammon, N. Doltsinis, *J. Chem. Phys.*, 2011, **134**, article no. 174107.
[13] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12562-12566.
[14] C. Dellago, P. G. Bolhuis, P. L. Geissler, in *Advances in Chemical Physics* (I. Prigogine, S. A. Rice, eds.), vol. 123, John Wiley & Sons, Ltd, New York, 2002, 1-78.
[15] P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291-318.
[16] A. Gil-Ley, G. Bussi, *J. Chem. Theory Comput.*, 2015, **11**, 1077-1085.
[17] M. Invernizzi, M. Parrinello, *J. Phys. Chem. Lett.*, 2020, **11**, 2731-2736.
[18] J. Behler, M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, article no. 146401.
[19] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, K.-R. Müller, *Chem. Rev.*, 2021, **121**, 10142-10186.
[20] J. Zeng, D. Zhang, D. Lu, P. Mo, Z. Li, Y. Chen, M. Rynik, L. Huang, Z. Li, S. Shi, Y. Wang, H. Ye, P. Tuo, J. Yang, Y. Ding, Y. Li, D. Tisi, Q. Zeng, H. Bao, Y. Xia, J. Huang, K. Muraoka, Y. Wang, J. Chang, F. Yuan, S. L. Bore, C. Cai, Y. Lin, B. Wang, J. Xu, J.-X. Zhu, C. Luo, Y. Zhang, R. E. A. Goodall, W. Liang, A. K. Singh, S. Yao, J. Zhang, R. Wentzcovitch, J. Han, J. Liu, W. Jia, D. M. York, E. Weinan, R. Car, L. Zhang, H. Wang, *J. Chem. Phys.*, 2023, **159**, article no. 054801.
[21] D. Marx, J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, Cambridge, 2009.
[22] A. Warshel, R. M. Weiss, *J. Am. Chem. Soc.*, 1980, **102**, 6218-6226.
[23] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama, A. C. T. van Duin, *NPJ Comput. Mater.*, 2016, **2**, article no. 15011.
[24] C. Qu, Q. Yu, J. M. Bowman, *Annu. Rev. Phys. Chem.*, 2018, **69**, 151-175.
[25] J. H. McCreery, G. Wolken Jr., *J. Chem. Phys.*, 1975, **63**, 2340-2349.
[26] T. B. Blank, S. D. Brown, A. W. Calhoun, D. J. Doren, *J. Chem. Phys.*, 1995, **103**, 4129-4137.
[27] S. Lorenz, A. Groß, M. Scheffler, *Chem. Phys. Lett.*, 2004, **395**, 210-215.
[28] S. Käser, L. Vazquez-Salazar, M. Meuwly, K. Töpfer, *Digit. Discov.*, 2023, **2**, 28-58.
[29] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller, *Sci. Adv.*, 2017, **3**, article no. e1603015.
[30] S. Chmiela, H. E. Sauceda, A. Tkatchenko, *Nat. Commun.*, 2018, **9**, article no. 3887.
[31] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, article no. 136403.
[32] H. Wang, L. Zhang, J. Han, W. E, *Comput. Phys. Commun.*, 2018, **228**, 178-184.
[33] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, W. E, in *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett, eds.), vol. 31, Curran Associates, Inc., Montréal, 2018.
[34] K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, eds.), vol. 30, Curran Associates, Inc., Longbeach, CA, 2017.
[35] E. Kocer, T. W. Ko, J. Behler, *Annu. Rev. Phys. Chem.*, 2022, **73**, 163-186.
[36] J. Behler, *Chem. Rev.*, 2021, **121**, 10037-10072.
[37] J. S. Smith, O. Isayev, A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192-3203.
[38] J. Behler, *J. Chem. Phys.*, 2011, **134**, article no. 074106.

[39] M. de la Puente, R. David, A. Gomez, D. Laage, *J. Am. Chem. Soc.*, 2022, **144**, 10524-10529.
[40] M. de la Puente, D. Laage, *J. Am. Chem. Soc.*, 2023, **145**, 25186-25194.
[41] R. David, I. Tuñón, D. Laage, *J. Am. Chem. Soc.*, 2024, **146**, 14213-14224.
[42] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, T. S. Jaakkola, *Transact. Mach. Learn. Res.*, 2023, https://openreview.net/pdf?id=A8pqQipwkt.
[43] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, 2016.
[44] S. Yue, M. C. Muniz, M. F. Calegari Andrade, L. Zhang, R. Car, A. Z. Panagiotopoulos, *J. Chem. Phys.*, 2021, **154**, article no. 034111.
[45] L. Zhang, H. Wang, M. C. Muniz, A. Z. Panagiotopoulos, R. Car, W. E, *J. Chem. Phys.*, 2022, **156**, article no. 124107.
[46] T. Fink, J.-L. Reymond, *J. Chem. Inf. Model.*, 2007, **47**, 342-353.
[47] J. Daru, H. Forbert, J. Behler, D. Marx, *Phys. Rev. Lett.*, 2022, **129**, article no. 226001.
[48] T. Plé, L. Lagardère, J.-P. Piquemal, *Chem. Sci.*, 2023, **14**, 12554-12569.
[49] S. Zhang *et al.*, *Nat. Chem.*, 2024, **16**, 727-734.
[50] R. Galvelis, A. Varela-Rial, S. Doerr, R. Fino, P. Eastman, T. E. Markland, J. D. Chodera, G. De Fabritiis, *J. Chem. Inf. Model.*, 2023, **63**, 5701-5708.
[51] P. Eastman, R. Galvelis, R. P. Peláez, C. R. A. Abreu, S. E. Farr, E. Gallicchio, A. Gorenko, M. M. Henry, F. Hu, J. Huang, A. Krämer, J. Michel, J. A. Mitchell, V. S. Pande, J. P. Rodrigues, J. Rodriguez-Guerra, A. C. Simmonett, S. Singh, J. Swails, P. Turner, Y. Wang, I. Zhang, J. D. Chodera, G. De Fabritiis, T. E. Markland, *J. Phys. Chem. B*, 2024, **128**, 109-116.
[52] L. Zhang, D. Y. Lin, H. Wang, R. Car, E. Weinan, *Phys. Rev. Materials*, 2019, **3**, 1-9.
[53] A. Krogh, J. Vedelsby, in *Advances in Neural Information Processing Systems* (G. Tesauro, D. Touretzky, T. Leen, eds.), vol. 7, MIT Press, Denver, CO, 1994.
[54] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187-199.
[55] R. P. Feynman, A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, 1965.
[56] T. E. Markland, M. Ceriotti, *Nat. Rev. Chem.*, 2018, **2**, article no. 0109.
[57] T. D. Loose, P. G. Sahrmann, G. A. Voth, *J. Chem. Theory Comput.*, 2022, **18**, 5856-5863.
[58] A. M. Tokita, J. Behler, *J. Chem. Phys.*, 2023, **159**, article no. 121501.
[59] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, S. J. Plimpton, *Comput. Phys. Commun.*, 2022, **271**, article no. 108171.
[60] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. Van Speybroeck, M. Ceriotti, *Comput. Phys. Commun.*, 2019, **236**, 214-223.
[61] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604-613.
[62] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, J. Hutter, *J. Chem. Phys.*, 2020, **152**, article no. 194103.
[63] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, W. E, *Comput. Phys. Commun.*, 2020, **253**, article no. 107206.
[64] M. Calegari Andrade, R. Car, A. Selloni, *Proc. Natl. Acad. Sci. USA*, 2023, **120**, article no. e2302468120.
[65] M. Bocus, R. Goeminne, A. Lamaire, M. Cools-Ceuppens, T. Verstraelen, V. Van Speybroeck, *Nat. Commun.*, 2023, **14**, article no. 1008.
[66] B. Wen, M. F. Calegari Andrade, L.-M. Liu, A. Selloni, *Proc. Natl. Acad. Sci. USA*, 2023, **120**, article no. e2212250120.
[67] M. F. Calegari Andrade, H.-Y. Ko, L. Zhang, R. Car, A. Selloni, *Chem. Sci.*, 2020, **11**, 2335-2341.
[68] Y.-B. Zhuang, R.-H. Bi, J. Cheng, *J. Chem. Phys.*, 2022, **157**, article no. 164701.
[69] Z. Zeng, F. Wodaczek, K. Liu, F. Stein, J. Hutter, J. Chen, B. Cheng, *Nat. Commun.*, 2023, **14**, article no. 6131.
[70] M. Yang, U. Raucci, M. Parrinello, *Nat. Catal.*, 2023, **6**, 829-836.
[71] T. A. Young, T. Johnston-Wood, H. Zhang, F. Duarte, *Phys. Chem. Chem. Phys.*, 2022, **24**, 20820-20827.
[72] T. Devergne, T. Magrino, F. Pietrucci, A. M. Saitta, *J. Chem. Theory Comput.*, 2022, **18**, 5410-5421.
[73] M. Yang, L. Bonati, D. Polino, M. Parrinello, *Catal. Today*, 2022, **387**, 143-149.
[74] M. Galib, D. T. Limmer, *Science*, 2021, **371**, 921-925.
[75] D. Xia, J. Chen, H.-B. Xie, J. Zhong, J. S. Francisco, *J. Am. Chem. Soc.*, 2023, **145**, 4791-4799.
[76] J. Zeng, L. Cao, M. Xu, T. Zhu, J. Z. H. Zhang, *Nat. Commun.*, 2020, **11**, article no. 5713.
[77] J. Zeng, L. Zhang, H. Wang, T. Zhu, *Energy Fuels*, 2021, **35**, 762-769.
[78] J. Boström, D. G. Brown, R. J. Young, G. M. Keserü, *Nat. Rev. Drug Discov.*, 2018, **17**, 709-727.
[79] K. Zinovjev, I. Tuñón, *J. Phys. Chem. A*, 2017, **121**, 9764-9772.
[80] E. Weinan, W. Ren, E. Vanden-Eijnden, *J. Phys. Chem. B*, 2005, **109**, 6688-6693.
[81] D. Branduardi, F. L. Gervasio, M. Parrinello, *J. Chem. Phys.*, 2007, **126**, article no. 054103.
[82] P. G. Bolhuis, C. Dellago, D. Chandler, *Faraday Discuss.*, 1998, **110**, 421-436.
[83] C. Dellago, P. G. Bolhuis, F. S. Csajka, D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964-1977.
[84] C. Dellago, P. G. Bolhuis, in *Advanced Computer Simulation Approaches for Soft Matter Sciences III* (C. Holm, K. Kremer, eds.), Springer, Berlin, Heidelberg, 2009, 167-233.
[85] P. G. Bolhuis, D. W. H. Swenson, *Adv. Theory Simul.*, 2021, **4**, article no. 2000237.
[86] P. G. Bolhuis, *J. Phys.: Condens. Matter*, 2003, **15**, S113-S120.

[87] J. Rogal, P. G. Bolhuis, *J. Chem. Phys.*, 2008, **129**, article no. 224107.
[88] Z. F. Brotzakis, P. G. Bolhuis, *J. Chem. Phys.*, 2016, **145**, article no. 164112.
[89] P. G. Bolhuis, G. Csányi, *Phys. Rev. Lett.*, 2018, **120**, article no. 250601.
[90] Z. F. Brotzakis, P. G. Bolhuis, *J. Chem. Phys.*, 2019, **151**, article no. 174111.
[91] T. S. van Erp, D. Moroni, P. G. Bolhuis, *J. Chem. Phys.*, 2003, **118**, 7762-7774.
[92] T. S. van Erp, P. G. Bolhuis, *J. Comput. Phys.*, 2005, **205**, 157-181.
[93] R. J. Allen, P. B. Warren, P. R. Ten Wolde, *Phys. Rev. Lett.*, 2005, **94**, article no. 018104.
[94] R. J. Allen, D. Frenkel, P. R. Ten Wolde, *J. Chem. Phys.*, 2006, **124**, article no. 024102.
[95] S. W. Hall, G. Díaz Leines, S. Sarupria, J. Rogal, *J. Chem. Phys.*, 2022, **156**, article no. 200901.
[96] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, M. Parrinello, *Science*, 2001, **291**, 2121-2124.
[97] D. Zahn, *J. Chem. Theory Comput.*, 2006, **2**, 107-114.
[98] A. Tiwari, B. Ensing, *Faraday Discuss.*, 2016, **195**, 291-310.
[99] D. W. H. Swenson, J.-H. Prinz, F. Noe, J. D. Chodera, P. G. Bolhuis, *J. Chem. Theory Comput.*, 2019, **15**, 837-856.
[100] D. W. H. Swenson, J.-H. Prinz, F. Noe, J. D. Chodera, P. G. Bolhuis, *J. Chem. Theory Comput.*, 2019, **15**, 813-836.
[101] G. M. Blackburn, W. P. Jencks, *J. Am. Chem. Soc.*, 1968, **90**, 2638-2645.
[102] A. C. Satterthwait, W. P. Jencks, *J. Am. Chem. Soc.*, 1974, **96**, 7018-7031.
[103] S. Chalmet, W. Harb, M. F. Ruiz-López, *J. Phys. Chem. A*, 2001, **105**, 11574-11581.
[104] K. Świderek, I. Tuñón, S. Martí, V. Moliner, J. Bertrán, *Chem. Commun.*, 2012, **48**, article no. 11253.
[105] J. H. Jensen, K. K. Baldridge, M. S. Gordon, *J. Phys. Chem.*, 1992, **96**, 8340-8351.
[106] S. Ilieva, B. Galabov, D. G. Musaev, K. Morokuma, H. F. Schaefer, *J. Org. Chem.*, 2003, **68**, 1496-1502.
[107] P. K. Sharma, Y. Xiang, M. Kato, A. Warshel, *Biochemistry*, 2005, **44**, 11307-11314.
[108] K. Świderek, I. Tuñón, S. Martí, V. Moliner, J. Bertrán, *J. Am. Chem. Soc.*, 2013, **135**, 8708-8719.
[109] K. Sakata, N. Kitadai, T. Yokoyama, *Geochim. Cosmochim. Acta*, 2010, **74**, 6841-6851.
[110] J. F. Marlier, B. A. Haptonstall, A. J. Johnson, K. A. Sacksteder, *J. Am. Chem. Soc.*, 1997, **119**, 8838-8842.
[111] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 2087-2096.
[112] L. Böselt, M. Thürlemann, S. Riniker, *J. Chem. Theory Comput.*, 2021, **17**, 2641-2658.
[113] M. de la Puente, A. Gomez, D. Laage, *J. Phys. Chem. Lett.*, 2024, **15**, 3096-3102.
[114] L. Bonati, E. Trizio, A. Rizzi, M. Parrinello, *J. Chem. Phys.*, 2023, **159**, article no. 014801.
[115] H. Chen, B. Roux, C. Chipot, *J. Chem. Theory Comput.*, 2023, **19**, 4414-4426.
[116] H. Jung, R. Covino, A. Arjun, C. Leitold, C. Dellago, P. G. Bolhuis, G. Hummer, *Nat. Comput. Sci.*, 2023, **3**, 334-345.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 95-106
https://doi.org/10.5802/crchim.283

Research article

French Network on Solvation (GDR 2035 SolvATE)

On the validity of some equilibrium models for thermodiffusion

Mario Araujo-Rocha[a], **Alejandro Diaz-Marquez**[a] **and Guillaume Stirnemann**[*,a,b]

[a] CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, Université Paris Cité, 13 rue Pierre et Marie Curie, 75005, Paris, France

[b] PASTEUR, Département de chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France

E-mails: araujorocha@ibpc.fr (M. Araujo-Rocha), diazmarquez@ibpc.fr (A. Diaz-Marquez), guillaume.stirnemann@ens.psl.eu (G. Stirnemann)

Abstract. When applied to binary solutions, thermal gradients lead to the generation of concentration gradients and thus to inhomogeneous systems. While being known for more than 150 years, the molecular origins for this phenomenon are still debated, and there is no consensus on the underlying physical models or theories that could explain the amplitude of the concentration gradient in response to a given temperature gradients. Notably, there have been some attempts to relate this non-equilibrium, steady-state manifestation, to equilibrium properties of these solutions, for example, to the temperature dependence of the self-diffusion coefficient or to the solvation free energies of each of their components. Here, we use molecular dynamics simulations on dilute solutions containing molecular-size solutes, both in a thermophoretic setting as well as under equilibrium conditions, to test the validity of such models. We show that these approaches are inadequate and lead to completely uncorrelated estimates as compared to those based on the out-of-equilibrium measurements. Crucially, they fail to explain the strong mass dependence (to which thermodynamics or single-particle diffusion are insensitive) observed in the simulations and measured in the experiments. However, our results suggest an interesting correlation between the amplitude of the short-time molecular motion and that of the concentration gradient that would deserve future investigations.

Keywords. Thermophoresis, Solvation free energy, Translational diffusion.

Funding. European Research Council under the European Union's Eighth Framework Program (H2020/2014-2020)/ERC Grant Agreement No. 757111, "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01).

Manuscript received 13 October 2023, revised 20 December 2023, accepted 10 January 2024.

1. Introduction

When subjected to an inhomogeneous spatial temperature distribution, a liquid mixture's local composition usually becomes temperature- (and thus space-) dependent. This phenomenon is known as thermodiffusion, thermophoresis (especially when considering large particles in a solvent), or Soret effect [1–3]. In order to explain such experimental observations, that can be quantified with a wide range of techniques [4], a typical phenomenological approach consists of considering a particle current that is proportional to the concentration and to the temperature gradient (through a factor known as "thermal diffusion coefficient") and which adds up with the regular Fickian diffusion current, proportional

* Corresponding author

(through the regular diffusion coefficient) to the concentration gradient [2]. At steady state, the absence of net fluxes results in an exponentially distributed concentration of particles as a function of temperature, which is consistent with the experimental measurements. The temperature dependence of the concentration is called the Soret coefficient, and it corresponds to the ratio of thermal and regular diffusion coefficients. This phenomenon is observed for many different systems, across many lengthscales, and the present contribution will exclusively discuss and focus on molecular-size diffusing objects.

These considerations remain purely phenomenological. A vast body of experimental and computational work has attempted to go beyond this picture and characterise the molecular determinants of thermodiffusion (see, e.g., [2]). While being an "old" problem, this remains a partly open question and a timely topic, with very recent work testing phenomenological thermodynamic models using molecular dynamics simulations of Lennard-Jones particles [5–7].

It is now widely accepted that the Soret coefficient can be decomposed as a sum of a purely chemical component (which depends on the interatomic interactions between the components of the mixture), a mass-dependent component, sensitive to the difference in masses between the mixture particles, and a component sensitive to the difference in the moments of inertia [2,8,9]. This decomposition was extensively studied through molecular simulations of hard-spheres and Lennard-Jones binary fluids [7]. Very recent work completed this picture by highlighting the effect of the mass dipole [10] (i.e., the first moment of the mass distribution in the molecules). If we take the example of aqueous mixtures, the so-called chemical component has been related to the hydrophobic/hydrophilic nature of the solutes, in particular, to the strength of their hydrogen-bond interactions with water [3,11,12].

Although interesting, making such connections does not really provide an underlying molecular picture of the phenomenon. At the turn of the 20th century, Einstein's work paved the way for a molecular understanding of single particle diffusion [13]. The diffusion coefficient was thus connected to the mean square displacement of molecules, as well as to the atomistic details of the diffusing particle (e.g., its radius) and of the diffusing medium (e.g., its viscosity). As opposed to regular diffusion, a major challenge is that thermophoresis is, in essence, a phenomenon out of thermodynamic equilibrium, which considerably limits the theoretical framework and tools that can be used. While many attempts have been made for thermodiffusion [2,3,14–19] there is currently no "universal", widely-accepted theoretical model that is able to quantitatively connect the Soret coefficient or the thermal diffusion coefficient to the structural and thermodynamical characteristics of the mixture, especially for molecular-size solutes. We note however that models that contain at least some empirical or phenomenological ingredients were shown to correlate well with the thermophoretic behavior of hard-spheres and Lennard-Jones particles [7], although some limitations were also pointed out [5].

Despite these practical limitations, several appealing models have been proposed that rely on equilibrium considerations to explain the steady-state of a liquid system under a thermal gradient. For example, Eastman derived the following equation (written here with modern notations) for a particle immersed in a fluid [14]:

$$S_T = \frac{1}{k_B T} \frac{\mathrm{d}G}{\mathrm{d}T} \tag{1}$$

where G is the free energy of the particle in the solution. This idea was later reinforced by a number of experimental results from the Braun group on colloidal suspensions [16,20,21] that were shown to agree with such models, which therefore assume local equilibria along the temperature gradient. In other words, the particles in a solution would, therefore, tend to accumulate in the regions where the system can minimize its free energy; that is, it follows the gradient of solvation free energy. The validity of such a picture has often been questioned, and it was shown that its applicability could depend on the lengthscale of the system considered [19].

Another interesting theory relates the Soret coefficient of particle 2 in a solution with 1 to the differences in the activation energy for single-particle diffusion E_a of the two particle types 1 and 2 in the mixture [17]:

$$S_T = \frac{E_a^2 - E_a^1}{RT^2} \tag{2}$$

In other words, the molecules with the highest activation energies for diffusion would tend to accumulate in the cold regions (positive Soret coefficients).

This equation echoes a phenomenological expression where the diffusion activation energy is replaced by the heat of transport of the species in the mixture. It is important to note that there seems to be a sign typo in the original publication [17], the final equation (7.9 in the original article) not being consistent with the discussion in the text.

Measuring the single-particle diffusion of molecular systems or their solvation free energy is not straightforward in the experiments; the validity of these models has often been discussed based on some measured correlations, for example, with the solute surface area [16,20], or based on theoretical arguments only [14]. For example, one obvious limitation of these models is that their ingredients are mass-independent, i.e. they cannot explain, without further corrections, the mass dependence of the Soret coefficient. In order to test these models more thoroughly, we used molecular dynamics in explicit solvent, replicating experimental thermophoretic settings and equilibrium conditions. We used two types of solutions, starting from binary mixtures of Lennard-Jones (LJ) particles as well as aqueous solutions. We focused on the dilute regime (from the perspective of the solute).

We show that these equilibrium models are not adequate. There is no obvious correlation between the Soret coefficient and the solvation free energy of the dilute particle or the difference between its activation energy for diffusion and that of the solvent. These models fail to reproduce qualitatively the measured trend, and their predictions are sometimes off by one order of magnitude. We finish by discussing some interesting correlations between the Soret coefficient in these mixtures, which suggest that, for very different systems ranging from the binary LJ solutions to several aqueous mixtures and for a wide range of molecular masses that can be artificially tuned in the simulation, the Soret coefficient is directly correlated to the relative difference between the amplitude of the solute motion at short timescales and that of the solvent.

2. Methods

2.1. *Systems*

The simulations of the LJ binary mixtures considered Argon-like particles [22], with the solvent being described with ϵ = 0.2381 kcal/mol, σ = 3.405 Å and m = 39.94 u. Solute particles were identical and only differed in their ϵ values and sometimes their mass. The reference temperature was thus $T = \epsilon/k_B$ = 119.8 K, with a corresponding T^* = 1 in reduced units. As detailed below, simulations were typically performed at reduced temperatures ranging from T^* = 0.9 to 1.1 (107.82 to 131.78 K) and reduced pressure P^* = 0.8 (33.52 MPa). We have thoroughly checked that the system remains in a liquid phase in these conditions.

The all-atom molecular simulations of the binary mixtures were performed at temperatures ranging from T = 300 K to 360 K and a pressure of 1 bar. The force field parameter models for trimethylamine N-oxide (TMAO) [23,24], urea [25], methanol [26], glucose [27] and water [28] were extracted directly from the available literature which fulfill the working temperature and pressure conditions in aqueous solutions.

Solvent particles were randomly integrated into the simulation box through the utilization of the Packmol package [29]. Simultaneously, the solute particles were uniformly inserted along the z-direction of the thermal gradient, leading to an initial configuration characterized by a flat concentration profile. The Coulombic cut-off was 8.5 Å and the LJ cut-off was 9 Å, with particle–particle particle–mesh (PPPM) [30] solver for the long-range electrostatic forces, in the case of dilute aqueous solutions. The particles were randomly generated using LAMMPS software version 07-Aug-19's [31] internal random atom generator for the LJ systems. The cut-off for non-bonded interactions was set to 2.5σ (9.534 Å) for the LJ systems. Timesteps were 1 fs in all cases.

2.2. *Thermophoretic simulations*

All simulations under a thermal gradient were performed following a protocol described before [32], with 12 steps from preparation to production, using the LAMMPS software version 07-Aug-19 [31] with the eHex algorithm to generate a temperature gradient [33]. These include some energy minimizations, equilibration steps in successive thermodynamic ensembles, activation of a heat exchange algorithm [33] and a subsequent equilibration of the temperature and concentration gradients, and finally a production stage.

For the LJ mixtures, we started with boxes of $7.65\sigma \times 7.65\sigma \times 50\sigma$, containing 2048 atoms of the solvent and 80 atoms of the solute. The same aforementioned protocol [32] was performed on these systems, adapting the lengths of each step to their reduced time equivalents. The thermal gradient was attained by applying the eHex algorithm of heat exchange [33] in two 4σ-long slabs over the z-axis; kinetic energy was removed from the atoms present in the region between $z = 10.5\sigma$ and 14.5σ and applied to the atoms present in the region between $z = 35.5\sigma$ and 39.5σ. A scheme of the system and of the temperature gradient is shown in Figure 1.

The thermophoretic simulations of the molecular binary mixtures started with boxes with 25 Å length along the x and y axes and 50 Å along z. They contain 1024 molecules of water and 40 molecules of the solute. The resulting systems are identical, in terms of construction, to the LJ mixtures shown in Figure 1. The thermostatted regions are positioned along the z-axis at 12.5 Å and 27.5 Å with a thickness of 4 Å, creating two symmetrically located regions of 25 Å.

The exchanged heat flux in the eHex algorithm was set such that the generated temperature differences between the cold and hot slab were 60 K for the aqueous solutions ($Q = 0.0375$ kcal·mol^{-1}·fs^{-1}) and 0.2 reduced temperature-unit (23.96 K) for the LJ systems ($Q = 5\epsilon/\tau$, or 0.00054 kcal·mol^{-1}·fs^{-1}).

All presented error bars for the Soret coefficient are calculated from the standard deviation of multiple simulations with different starting configurations.

2.3. *Free energy perturbation*

We first detail the hydration free energy calculations for the solute molecules in dilute aqueous solutions. The simulation process involved employing cubic simulation boxes of 33.15 Å-side, filled with solvent molecules inserted randomly without spatial constraints. The system was minimized, the velocities were set to correspond to the target median temperature followed by an isobaric-isothermal NPT equilibration of 200 ps at 100 kPa. The decoupling process involved progressively reducing Coulombic and LJ interactions over 40 steps using soft-core potentials by introducing a tunable parameter (λ). These simulations were performed without an induced temperature gradient.

Simulations were executed using both the LAMMPS software version 07-Aug-19 [31], with its FEP package, and the GROMACS package version 2019.4 [34] to validate the results. LAMMPS simulations were performed at 8.5 Å Coulombic cutoff, 9 Å LJ cutoff and using the particle–particle particle–mesh (PPPM) solver for long-range electrostatic forces. The damping parameters for the Nose–Hoover [35,36] barostat and thermostat are $P_{\text{damp}} = 1$ ps and $T_{\text{damp}} = 0.1$ ps respectively, with 3 Nose–Hoover chains and the velocity Verlet algorithm. Then, Coulombic interactions were switched off (λ_{Coul}) over 20 steps ($\delta\lambda_{\text{Coul}} = -0.05$). Once the Coulombic term was completely decoupled, the LJ potential was decoupled in the same way over the remaining windows. During this process, intermolecular non-bonded interactions were also affected, so another decoupling simulation was performed in the gas phase, and the desolvation free-energy was obtained as a difference between these two processes.

GROMACS simulations were performed at 9 Å Coulombic cutoff, 9 Å LJ cutoff and the particle–mesh Ewald (PME) solver [37]. The damping parameters for the Parrinello-Rahman [38] barostat and the Nose–Hoover thermostat at 2 ps, and the barostat compressibility were set at 4.46×10^{-5} bar^{-1}, using the leap-frog integrator and 1 Nose–Hoover chain. The Coulombic and LJ interactions were decoupled over 14 steps. First, the Coulombic interactions were switched off (λ_{Coul}) over 4 steps ($\delta\lambda_{\text{Coul}} = -0.25$). Once the Coulombic term was completely decoupled, the LJ potential was decoupled, where (λ_{LJ}) decreased over 10 steps ($\delta\lambda_{\text{LJ}} = -0.1$).

For GROMACS simulations, the Alchemical Analysis tool was employed [39], capable of handling various free-energy methods, including BAR (Bennet acceptance ratio, [40]), MBAR (Multistate Bennett acceptance ratio, [41]) and TI (thermodynamic integration [42]), while LAMMPS simulations utilized a separate tool within the FEP package [43], with our own adaptations made to facilitate compatibility with the Alchemical Analysis tool.

For the LJ binary mixtures, the analyses of free energy perturbation method were done using the FEP-package on the LAMMPS software version 07-Aug-19 [31]. As in the molecular mixtures, the system was initially minimized, then the velocities were then set to the target temperatures. This was followed by a NPT equilibration step of 1000 reduced

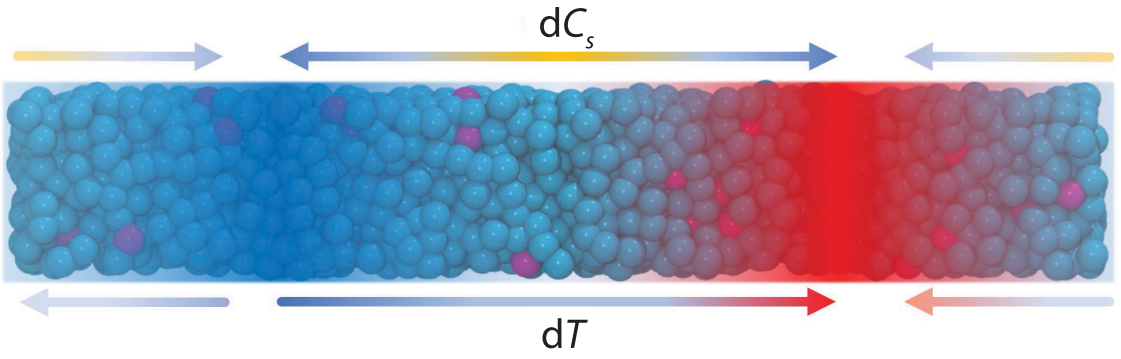


Figure 1. Schematic representation of a thermophoretic simulation (here, for LJ binary mixture). Cold and hot regions are shown in blue and red, respectively. They are positioned at one-fourth and three-fourths of the simulation box along its long z axis so as to generate the temperature gradient "dT" twice, thanks to periodic boundary conditions. This in turn generates a concentration gradient "dC_s" of solute particles (magenta) in the solvent (blue particles).

time units (2.180 ns) with $T^* = 1.0$ and $P^* = 0.8$. The damping parameters for the Nose–Hoover barostat and thermostat were $P^*_{\text{damp}} = T^*_{\text{damp}} = 1$ reduced unit time. The LJ potential in this method is modified by a soft core to avoid singularities while annihilating atoms [44]. The simulations were performed with a LJ cutoff (r_c) of 2.5σ (9.534 Å). Contrary to the method used for the molecular mixtures, there was no need of two decoupling steps, as there were no Coulombic interactions. The LJ potential was decoupled through the factor λ_{LJ} in 20 steps ($\delta\lambda_{\text{LJ}} = -0.05$). As in the aqueous solution systems, these simulations were performed without an induced temperature gradient. In both the aqueous solution systems and the LJ systems, the presented error bars for the desolvation energies are calculated from the standard deviation of multiple simulations with different starting configurations.

2.4. *Mean square displacement calculations*

A cubic box (33.15 Å-side) was considered for the molecular binary mixtures. The equilibration part is replicated from thermophoretic simulations [32], which are steps 1 to 8, in order to obtain a system in the microcanonical NVE ensemble with a volume and total energy corresponding to the averages in the NPT ensemble, such that the average pressure and temperature matched the targets of the chosen thermodynamic conditions. The diffusion coefficients were determined by measuring the mean squared displacements (MSD) of the centers of mass for the solute and the solvent molecules/particles at the median temperature of 330 K. An equivalent protocol was followed for the LJ binary mixtures, with a starting simulation cubic box of size 25σ, at their median temperature of $T^* = 1.0$. As it is well established [45], the diffusion coefficient is expected to be slightly box size-dependent because of periodic boundary effects. However, assuming that the diffusion coefficient of these molecular species can be described by the Debye–Stokes–Einstein model, the size correction is temperature-independent [45], and therefore, the activation energy for diffusion is expected to be size-independent. For the determination of error bars, we used 5 replicas of 5 ns each, analyzed by blocks of 1 ns.

3. Results and discussions

3.1. *Determination of Soret coefficients*

We first study the thermophoretic behaviour of dilute aqueous solutions and of dilute LJ binary mixtures. In both cases, we systematically vary the solute nature and/or interaction parameters in order to simulate a range of different chemical properties for the solutes while keeping the solvent nature intact. Because these two types of solutions obviously have very different phase diagrams, we adapted the reference temperature, density and temperature gradients to remain in the liquid phase.

In silico thermophoretic settings were modelled as detailed in the Methods section. In short, we used a heat exchange algorithm [33] in the microcanonical ensemble, which pumps kinetic energy in one portion of the simulation box (which becomes colder than the average) and injects it in another region, which becomes the hot section of the system (Figure 1). In between, a temperature gradient is generated. The affected regions are chosen such that the gradient is unidirectional along one axis of the system (typically chosen as the z axis). In general, the temperature-gradient is established after a short equilibration timescale on the order of a few tens of picoseconds for the investigated systems [32]. A subsequent and much longer equilibration time [32] is then required to establish the concentration gradient, and only then is data acquired to determine the Soret coefficient. As shown previously, these timescales correspond to those of the heat and mass transport dynamics, respectively [32]. Simulations were replicated several times (typically 10 or more) in order to determine converged values and statistical uncertainties.

The Soret coefficient is determined after fitting the concentration profile with the following expression:

$$\frac{\mathrm{d}\ln C_s}{\mathrm{d}T} = -S_T \tag{3}$$

As shown before, under the conditions employed here, the log-scale plot of the solute concentration (defined here as its molality) is typically linear outside the thermostatted regions [32].

In Table 1, we show the values of the simulated Soret coefficients for several dilute aqueous solutions and binary LJ mixtures. The values are typically on the order of 10^{-3}–10^{-2} K^{-1}. Although Soret coefficients are most often measured for polymer or hydrocarbon mixtures [8,46–48] or large colloidal particles [16,20,21,49], there are a few experimental measurements for aqueous solutions, including urea [11, 12] and ethanol [50–53] (which is expected to be very close in behavior to methanol) that suggest that the Soret coefficient of the solute in these mixtures is indeed close to the values found here. Similar observations can be made for LJ mixtures [15,54]. This comparison thus validates our approach and we thereby consider the simulated values as the reference that will now use to test several theoretical models.

Table 1. Molecular masses and Soret coefficients of molecular and LJ solutes

Solute	M ($\mathrm{g{\cdot}mol^{-1}}$)	S_T (10^{-3} K^{-1})
Methanol	32.0	1.5 ± 0.7
Urea	60.1	4.4 ± 0.7
TMAO	75.1	5.0 ± 0.9
Glucose	180.2	3.1 ± 1.1
LJ Spheres ($\epsilon = 0.5$)	39.9	-7.6 ± 0.7
LJ Spheres ($\epsilon = 1.5$)	39.9	7.5 ± 2.9

Data for molecular solutes is reported at the median simulation temperature of 330 K in water, while that for LJ solutes is in a binary mixture of LJ particles with solvent particles of the same size and with $\epsilon = 1$ at $T^* = 1.0$.

3.2. *Solvation free energy*

In order to determine the solvation free energy at several different temperatures, we used standard approaches based on the progressive decoupling of the interactions between the target molecule and its environment. Simulations were performed by different simulation codes. For the aqueous solution simulations, as a natural choice, we first used the same software that was employed for thermophoretic simulations, but we also complemented our study with simulations performed with Gromacs [34]. For each approach, we compared three different algorithms in order to estimate the free energy cost associated with the process: TI, BAR and MBAR. Further details on these can be found in the Methods section.

We first take the example of a dilute aqueous solution of TMAO to compare the results from these different approaches (Figure 2A). Even when using the same raw data from the simulations, we notice that different algorithms give slightly different values, but the employed simulation codes give consistently similar answers. All these variations are typically within error bars, and lies within 0.4 kcal/mol from each other at 300 K and less than 1 kcal/mol at 330 and 360 K. Noticeably, they all quantitatively agree for the observed trend upon increasing temperature, with a ≈2 kcal/mol increase in the solvation free energy.

A second important step is to verify that our simulations agree with previous simulations and experimental data when available. Because we were

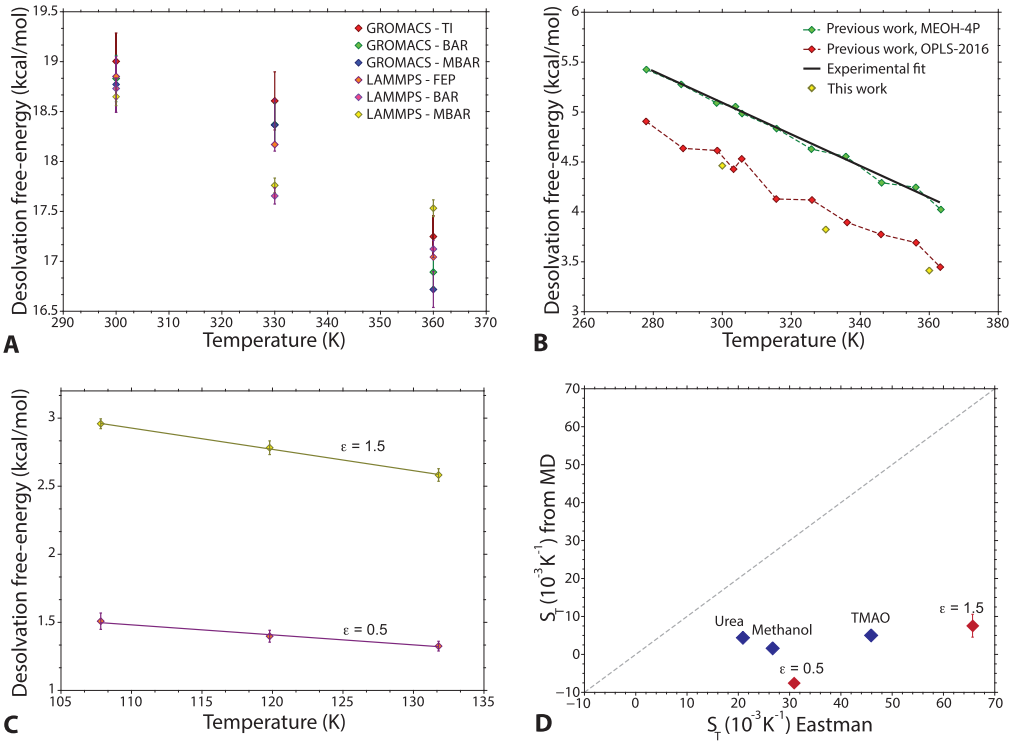


Figure 2. Solvation free energy approach. (A) Desolvation free energy of TMAO as a function of temperature in the dilute aqueous solution. Two different programs, as well as different algorithms, are compared. (B) Desolvation free energy of methanol in the dilute aqueous solution as a function of temperature, comparing results from this work (yellow) to those of previous ones (green and red, using two different forcefields) as well as the experimental estimate (black line). (C) Desolvation free energy for the solute particles as a function of temperature for the binary LJ mixtures. (D) Comparison between the direct estimate of the Soret coefficient in the investigated dilute solutions (vertical axis) and the predictions from the free energy model (horizontal axis).

lacking such data for TMAO, we do this comparison for methanol, using the LAAMPS code as the reference method (Figure 2B). Our simulations are in excellent agreement with previous simulation results using the OPLS force field for methanol [55], and are smaller in absolute value than the experimental values [56] (by about 0.5 kcal/mol). However, a very good agreement with experimental data was reported using a specifically reparametrized force field [55]. Despite these small differences, a critical aspect is that the experimental temperature dependence is very well reproduced by the previous simulations and the current ones. This provides a further validation of our approach.

We now turn to the determination of the temperature dependence of the solvation free energy for a variety of solutes. Figure 2C shows the temperature dependence for LJ solutes as a function of temperature. In addition to TMAO and methanol, we also performed simulations on aqueous solutions of urea. Overall, we notice that the solvation free energy becomes favorable as temperature increases, i.e., the solvation entropy is negative. In aqueous solutions, this is typically what is observed in the experiments [56] and in previous simulation work [57–59], as there is an entropic cost to create a cavity accommodating the solute, and also a reduced entropy due to sometimes strong interactions between the water molecules and some of the solute chemical groups (for example, H-bond acceptors).

The positive slope of the free energy as temperature increases is thus compatible with the

thermophobic nature of these solutes; however, testing Equation (1) reveals that the amplitude of these variations with temperature (Figure 2D) would lead to Soret coefficients at least one order of magnitude higher to what is measured in the thermophoretic simulations. We do not observe strong variations of the results by changing the algorithm used to estimate the free energies, and similar results are obtained for a range of solutes. In addition, we checked that performing the simulations with two different programs gave the same qualitative answer. Therefore, such a model does not seem appropriate to explain thermodiffusion and the value of the Soret coefficient.

3.3. *Temperature dependence of single-particle diffusion*

We then examine, for the same solutions, the temperature dependence of the translational single-particle diffusion coefficient of both the solute and the solvent particles or molecules. As done for the solvation free energy calculations, we simulated the different systems under equilibrium conditions at several temperatures (see Methods). Diffusion coefficients were obtained from mean square displacement calculations. Their temperature dependence was very Arrhenius-like, enabling the determination of the corresponding activation energies to test the Equation (2) model.

Starting with the aqueous solutions, we find that the activation energy for the solute diffusion is generally higher than that of water. As typical solutes display positive Soret coefficients, i.e., they accumulate in the cold regions: this is consistent with the original idea of Prigogine's work that a higher activation energy for diffusion would imply that molecules would be more easily trapped in the cold regions [17]. For the investigated solutes, the differences in the diffusion activation energy with the solvent are on the order of RT, which gives rise to the right orders of magnitude for S_T. However, as evidenced in Figure 3, we fail to find any correlation between the magnitude of the difference between the activation energies for the diffusion of the solute and of the solvent and the Soret coefficient, and Equation (2) therefore does not appear to be valid for these solutions. Using the same methodology, we reach similar conclusions for the investigated LJ dilute mixtures (Figure 3), albeit in obviously very different temperature conditions.

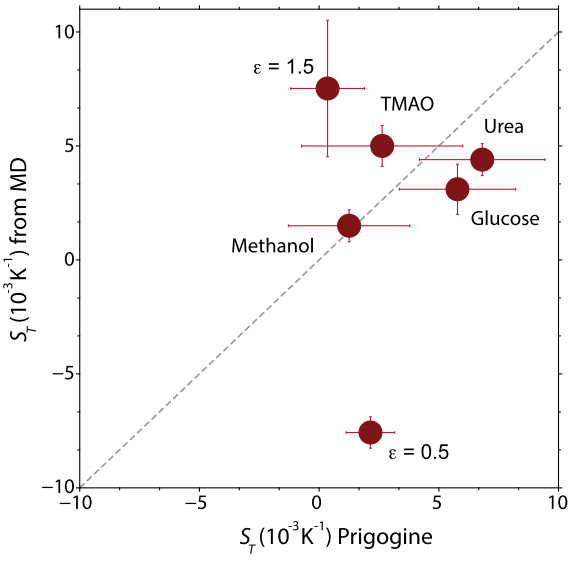


Figure 3. Activation energy for diffusion approach from Equation (2). Comparison between the direct estimate of the Soret coefficient in the investigated dilute solutions at T = 330 K (molecular solutes) and T^* = 1 (LJ solutions) (vertical axis) and the predictions from the activation energy for diffusion model (horizontal axis).

As suggested before, Prigogine's model is not able to explain the effects of mass on thermophoresis [18]. In particular, diffusion coefficients of a dilute particle in a given solvent are in principle mass-independent (note, however, that changing the solvent's mass could alter the solvent's viscosity and thus the diffusion coefficients of both the solute and the solvent). However, observations made from experimental and simulation measurements suggest that the mass-dependencies of thermophoresis follow phenomenological relationships in which one component of the Soret coefficient is proportional to the relative mass difference between the components of the binary mixtures [2]. The asymmetry between a change of mass in the solvent and a change of mass in the solute from the perspective of the diffusion coefficient (the first one having a small effect, the later one no effect) is a first hint that the Prigogine model would not be able to

explain the phenomenological mass-dependencies. Moreover, the variations of the diffusion coefficient temperature-dependence are too small to match the large effects of mass found before.

In order to account for the mass dependencies, a correction to the original Progogine model has been suggested and tested on binary LJ mixtures [18]. The Soret coefficient of particle 2 in a mixture with 1 was written as [18]:

$$S_T = \frac{E_a^2 - E_a^1}{RT^2} + \frac{M^2 - M^1}{M^1 + M^2}\frac{E_a^2 + E_a^1}{RT^2} \quad (4)$$

(we note that Prigogine's original expression seems to be written with the wrong sign, but was taken as such when the derivation of this expression was made, we also corrected their expression when writing it here). This model was again tested for our solutions and largely overestimated the measured Soret coefficients. Indeed, for molecular solutes that are typically several times the mass of a water molecule, the mass correction simply becomes the sum of the activation energies divided by RT^2. However, since the activation energies for diffusion are often several times the thermal energy, the correction typically amounts to several times $1/T$, whereas S_T is on the order of $1/T$ for the dilute molecular solutes studied here.

3.4. *Correlation with short-term molecular motion*

As we failed to quantitatively explain the Soret coefficients in dilute LJ mixtures and aqueous solutions with some of the previous theoretical equilibrium models, we tried to investigate possible correlations between S_T and other equilibrium observables at a given temperature. Keeping in mind Prigogine's conceptual ideas, i.e., that an imbalance in the temperature dependence of the molecular motion of molecules/particles would explain the onset of a concentration gradient, but seeking for quantities that would depend on mass, we focused on the initial regime of the molecular motion. For example, we looked at the average root mean squared displacement at a time interval corresponding to the onset of the diffusive regime after the initial ballistic region. In order to generate more data points, we (artificially) varied the solute masses and repeated the simulations for several solutes. As seen in Figure 4A, S_T is inversely correlated to the amplitude of this motion, with molecules faster molecules associated with a lower S_T value.

By taking a different perspective, we can also look at the average time it takes for a molecule to travel a certain distance. Because this picture probably makes more sense for particles of similar sizes, we performed these measurements on simulations of binary LJ mixtures for which we systematically varied the masses and the interaction energies of the solute particles. As shown in Figure 4B, we again find a clear correlation between S_T and the timescale of a short-range initial motion covering 1σ. Interestingly, the crossover between positive and negative S_T values quantitatively agrees with the timescale of this initial motion being slower or faster as compared to that of the solvent particles. While a correlation does not imply that two quantities are directly connected to each other, we found that these observations, made here for very different types of dilute solutions, are encouraging and would deserve future investigations.

4. Conclusions

In this work, we investigate the validity of some of the models that have been proposed to explain thermophoresis based on equilibrium considerations, working here exclusively on dilute binary mixtures (where the dilute particle is considered as a solute and the concentrated one as the solvent). The first of these models is based on an idea first put forward by Eastman in the 1940s [14], and which has gained considerable attention in the last 15 years thanks to important results from the Braun group [16,20,21]. Based on experimental measurements on small colloidal particles, it has been suggested that the Soret coefficient could be explained in terms of the temperature dependence of the solute's solvation free energy. The second investigated model connects the Soret coefficient to the temperature dependence of the diffusion coefficients of the solute and of the solvent molecules and was originally suggested by Prigogine [17], with some later refinements [18].

We use molecular dynamics to compute both non-equilibrium properties (such as thermodiffusion) and equilibrium ones (such as translational dynamics and solvation free energies) and to compare them. As already shown in our own work on dilute aqueous solutions [32], as well as in previous contributions focusing on LJ particles mixtures [15,54],

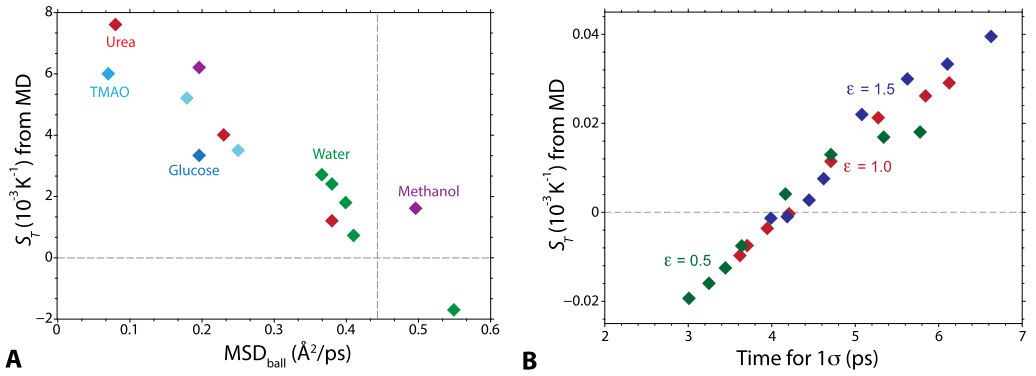


Figure 4. Correlation between the measured Soret coefficients for the investigated systems (varying both the chemical nature and the relative mass differences of the solute particles/molecules) and the short-term motion, seen from two different perspectives. (A) Correlation between S_T at a median temperature of 330 K with the mean square displacement of the solute at 330 K after 170 fs (corresponding to a change in sign in the second derivative of the MSD, indicative of a change of regime, that was observed to be almost system-independent for the dilute aqueous solutions shown here). Data for "water" solutes in water, with masses corresponding to 0.5, 1.5, 2.0, 4.0 and 8.0 times that of a regular water mass; for other solutes (except glucose, where the only data point corresponds to its natural mass): mass of a water molecule, regular mass, or 4 times the regular mass. (B) Soret coefficient at a median temperature of $T^* = 1.0$ vs the typical time to move by 1σ at the same temperature. These values were obtained through the MSD measurement of simulations as described in Section 2.4, tested over three values of ϵ for the solute (0.5, 1.0 and 1.5) and eight values of solute mass (0.33, 0.5, 0.75, 1.00, 2.00, 4.00, 6.00 and 8.00).

we first demonstrate how out-of-equilibrium simulations with an active heat exchange scheme can generate temperature gradients in molecular dynamics systems, which, in turn, leads to the establishment of concentration-gradients whose amplitude is compatible with experimental measurements.

We then run and analyse molecular dynamics trajectories on the same systems but in equilibrium conditions. By using different software and algorithms, we show that we can obtain reliable and converged estimates of the solute solvation free energies, and we examine their temperature dependence. Both for the aqueous solutions as well as for the LJ mixtures, we find that these cannot explain the Soret coefficients and that the Eastman model is not valid for the investigated molecular systems and predicts values that are one order of magnitude different from the steady-state measurements in non-equilibrium simulations. We cannot rule out that it would be correct for different systems, but we unambiguously show that this is not the case here. The same conclusions apply to Progogine's model and its variant to include mass dependencies. We find that the Soret coefficient does not correlate with the imbalance in the activation energy for diffusion between solute and solvent molecules.

These conclusions could have perhaps been expected. Experiments as well as simulations have evidenced the critical mass effects on the amplitude of the concentration gradient [2]. However, free energy considerations, as well as single-particle diffusion properties, are in principle mass-independent. A decomposition of the Soret coefficient into different terms corresponding to the intrinsic so-called chemical contribution, mass contribution, moment of inertia component, or, as recently discussed, a mass dipole contribution, is an interesting phenomenological perspective but lacks theoretical grounds; hence, a model encompassing all these effects at once is still lacking. We finish our discussion with some interesting correlations we found between the asymmetry of the short-term motion of solvent and solute molecules and the amplitude of the Soret coefficient in these solutions, which we believe would deserve further investigations as they cover a wide range of chemistry and masses in very different

solutions. In particular, the effect of different molecular chemistries on water dynamical properties and H-bond exchange kinetics [60–63] could maybe be related to the thermal transport properties of the molecular solutes in aqueous solutions.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

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 (GS). This work was also supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01 to GS).

Acknowledgments

The simulations presented here benefited from a local computing platform administered by G. Letessier, and benefited from the HPC resources of TGCC under the allocation A0070811005 made by GENCI (Grand Equipement National de Calcul Intensif).

References

[1] M. A. Rahman, M. Z. Saghir, *Int. J. Heat Mass Transf.*, 2014, **73**, 693-705.
[2] W. Köhler, K. I. Morozov, *J. Non-Equil. Thermodyn.*, 2016, **41**, 151-197.
[3] D. Niether, S. Wiegand, *J. Phys. Condens. Matter*, 2019, **31**, article no. 503003.
[4] W. Köhler, A. Mialdun, M. M. Bou-Ali, V. Shevtsova, *Int. J. Thermophys.*, 2023, **44**, article no. 140.
[5] N. E. Zimmermann, G. Guevara-Carrion, J. Vrabec, N. Hansen, *Adv. Theory Simul.*, 2022, **5**, article no. 2200311.
[6] O. R. Gittus, F. Bresme, *Phys. Chem. Chem. Phys.*, 2023, **25**, 1606-1611.
[7] H. Hoang, G. Galliero, *Eur. Phys. J. E*, 2022, **45**, article no. 2200311.
[8] W. M. Rutherford, *J. Chem. Phys.*, 1987, **86**, 5217-5218.
[9] W. M. Rutherford, *J. Chem. Phys.*, 1989, **90**, 602-603.
[10] O. R. Gittus, F. Bresme, *J. Chem. Phys.*, 2023, **159**, article no. 114503.
[11] D. Niether, S. Di Lecce, F. Bresme, S. Wiegand, *Phys. Chem. Chem. Phys.*, 2018, **20**, 1012-1020.
[12] D. Niether, H. Kriegs, J. K. Dhont, S. Wiegand, *J. Chem. Phys.*, 2018, **149**, article no. 044506.
[13] A. Einstein, *Ann. Phys.*, 1905, **322**, 549-560.
[14] E. D. Eastman, *J. Am. Chem. Soc.*, 1926, **48**, 1482-1493.
[15] P.-A. Artola, B. Rousseau, *Mol. Phys.*, 2013, **111**, 3394-3403.
[16] S. Duhr, D. Braun, *Phys. Rev. Lett.*, 2006, **96**, article no. 168301.
[17] I. Prigogine, L. de Brouckere, R. Buess, *Physica*, 1952, **18**, 915-920.
[18] P. A. Artola, B. Rousseau, G. Galliéro, *J. Am. Chem. Soc.*, 2008, **130**, 10963-10969.
[19] A. Würger, *C. R. Mec.*, 2013, **341**, 438-448.
[20] S. Duhr, D. Braun, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 19678-19682.
[21] M. Reichl, M. Herzog, A. Götz, D. Braun, *Phys. Rev. Lett.*, 2014, **112**, article no. 198101.
[22] L. Rowley, D. Nicholson, N. Parsonage, *J. Comput. Phys.*, 1975, **17**, 401-414.
[23] K. M. Kast, J. Brickmann, S. M. Kast, R. S. Berry, *J. Phys. Chem. A*, 2003, **107**, 5342-5351.
[24] C. Hölzl, P. Kibies, S. Imoto, R. Frach, S. Suladze, R. Winter, D. Marx, D. Horinek, S. M. Kast, *J. Chem. Phys.*, 2016, **144**.
[25] S. Weerasinghe, P. E. Smith, *J. Phys. Chem. B*, 2003, **107**, 3891-3898.
[26] D. Gonzalez-Salgado, C. Vega, *J. Chem. Phys.*, 2016, **145**, article no. 034508.
[27] S. H. Jamali, T. van Westen, O. A. Moultos, T. J. H. Vlugt, *J. Chem. Theory Comput.*, 2018, **14**, 6690-6700.
[28] J. L. Abascal, C. Vega, *J. Chem. Phys.*, 2005, **123**, article no. 234505.
[29] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157-2164.
[30] E. Polak, G. Ribiere, *Esaim Math. Model. Numer. Anal.*, 1969, **3**, 35-43, http://www.numdam.org/item/M2AN_1969__3_1_35_0/.
[31] S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1-19.
[32] A. Diaz-Marquez, G. Stirnemann, *J. Chem. Phys.*, 2021, **155**, article no. 174503.
[33] P. Wirnsberger, D. Frenkel, C. Dellago, *J. Chem. Phys.*, 2015, **143**, article no. 124104.
[34] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.
[35] S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511-519.
[36] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695-1697.
[37] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
[38] M. Parrinello, A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182-7190.
[39] P. V. Klimovich, M. R. Shirts, D. L. Mobley, *J. Comput. Aided Mol.*, 2015, **29**, 397-411.
[40] C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245-268.
[41] M. R. Shirts, J. D. Chodera, *J. Chem. Phys.*, 2008, **129**, article no. 124105.
[42] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 1st ed., Academic Press, Inc., USA, 1996.
[43] R. W. Zwanzig, *J. Chem. Phys.*, 2004, **22**, 1420-1426.

[44] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, *Chem. Phys. Lett.*, 1994, **222**, 529-539.
[45] I.-C. Yeh, G. Hummer, *J. Phys. Chem. B*, 2004, **108**, 15873-15879.
[46] D. Alonso De Mezquia, M. Mounir Bou-Ali, J. A. Madariaga, C. Santamaría, *J. Chem. Phys.*, 2014, **140**, article no. 084503.
[47] G. Wittko, W. Köhler, *J. Chem. Phys.*, 2005, **123**, article no. 014506.
[48] G. Wittko, W. Köhler, *Eur. Phys. J. E*, 2006, **21**, 283-291.
[49] J. K. Dhont, S. Wiegand, S. Duhr, D. Braun, *Langmuir*, 2007, **23**, 1674-1683.
[50] S. Wiegand, H. Ning, H. Kriegs, *J. Phys. Chem. B*, 2007, **111**, 14169-14174.
[51] A. Königer, B. Meier, W. Köhler, *Phil. Mag.*, 2009, **89**, 907-923.
[52] P. Kolodner, H. Williams, C. Moe, *J. Chem. Phys.*, 1988, **88**, 6512-6524.
[53] K. J. Zhang, M. E. Briggs, R. W. Gammon, J. V. Sengers, *J. Chem. Phys.*, 1996, **104**, 6881-6892.
[54] D. Reith, F. Müller-Plathe, *J. Chem. Phys.*, 2000, **112**, 2436-2443.
[55] M. Martínez-Jiménez, H. Saint-Martin, *J. Chem. Theory Comput.*, 2018, **14**, 2526-2537.
[56] J. Staudinger, P. V. Roberts, *Chemosphere*, 2001, **44**, 561-576.
[57] D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, K. A. Dill, *J. Chem. Theory Comput.*, 2009, **5**, 350-358.
[58] M. M. Kubo, E. Gallicchio, R. M. Levy, *J. Phys. Chem. B*, 1997, **101**, 10527-10534.
[59] E. Gallicchio, M. M. Kubo, R. M. Levy, *J. Phys. Chem. B*, 2000, **104**, 6271-6285.
[60] D. Laage, G. Stirnemann, J. T. Hynes, *J. Phys. Chem. B*, 2009, **113**, 2428-2435.
[61] F. Sterpone, G. Stirnemann, J. T. Hynes, D. Laage, *J. Phys. Chem. B*, 2010, **114**, 2083-2089.
[62] G. Stirnemann, J. T. Hynes, D. Laage, *J. Phys. Chem. B*, 2010, **114**, 3052-3059.
[63] G. Stirnemann, F. Sterpone, D. Laage, *J. Phys. Chem. B*, 2011, **115**, 3254-3262.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 107-115
https://doi.org/10.5802/crchim.304

Research article

French Network on Solvation (GDR 2035 SolvATE)

Thermodynamic investigations on host/guest complexation in deep eutectic solvent/water mixtures

Lamia Nakhle [a,b], Miriana Kfoury [a,*], Sophie Fourmentin [a], Hélène Greige-Gerges [b] and David Landy [a]

[a] Univ. Littoral Côte d'Opale, UR 4492, UCEIV, Unité de Chimie Environnementale et Interactions sur le Vivant, 59140 Dunkerque, France

[b] Bioactive Molecules Research Laboratory, Faculty of Sciences, Lebanese University, 90656 Jdeidet el-Metn, Lebanon

E-mail: miriana.kfoury@univ-littoral.fr (M. Kfoury)

Abstract. The development of formulations based on cyclodextrins (CDs) and deep eutectic solvents (DESs), is an interesting approach to expand the applications of host–guest chemistry. However, there is a lack of information in the literature on the complexation mechanism in these non-conventional media. The aim of this study is to provide, for the first time, fundamental information on the influence of the type and the wt% (weight percent) content of DES in the mixture on the thermodynamic properties (binding strength and complexation mechanism) of a model inclusion complex, β-CD/adamantanol by means of isothermal titration calorimetry (ITC).

Keywords. Adamantanol, Cyclodextrins, Deep eutectic solvents, Inclusion thermodynamics, Isothermal titration calorimetry.

Funding. National Council for Scientific Research of Lebanon (CNRS-L) and Université du Littoral Côte d'Opale (ULCO), CPER (Contrat de Plan Etat-Région) research project IRenE (Innovation et Recherche en Environnement), French Ministère de l'Enseignement Supérieur, region Hauts-de-France and European Regional Development Fund.

Manuscript received 27 September 2023, revised 3 January 2024 and 26 February 2024, accepted 5 March 2024.

1. Introduction

Cyclodextrins (CDs) are cyclic oligosaccharides obtained from the enzymatic degradation of starch [1]. Their three-dimensional structure forms a truncated-cone-shape due to the chair conformation of glucose residues. The external surface of the torus is hydrophilic and the central cavity is hydrophobic. This allows them to encapsulate guests and consequently form stable inclusion complexes via noncovalent interactions. The three native CDs, α-CD, β-CD and γ-CD have six, seven and eight D-glucose units, respectively. They consequently possess different cavity sizes allowing the encapsulation of a wide range of guest molecules of various structures. The attractive features of CDs, low price, high availability, ease of production, large-scale synthesis and safe nature, make them the most investigated

* Corresponding author

macrocycles [2,3]. They have been widely used in medicine, food, environment, cosmetics, chemical analysis, catalysis, and other fields. For more than 130 years, the mechanism of formation of an inclusion complex between CD and a guest has been explored and explained mainly in aqueous solution [4] and to a less extent in some organic solvents [5–7]. Lately, a new generation of green solvents, deep eutectic solvents (DESs), has gained interest as an alternative to water for applications that require drier conditions and for organic solvents that are being phased out in favor of more environmentally friendly alternatives. They are gaining interest in different fields such as gas absorption, solvation, extraction, formulation, synthesis, electrochemistry and so on [8–10]. A DES is a mixture of pure compounds for which the eutectic point temperature is below that of an ideal liquid mixture [11]. DESs are classified into five different types (I–V) based on the nature of compounds used for their preparation [12]. The nature of the constituents and the addition of a third component (i.e. water) can modulate the functionalities and physicochemical properties of the solvent [13,14]. Recently, CDs were used to provide supramolecular properties to these sustainable solvents [15–17]. This could be achieved either by dissolving CDs in previously prepared DES (ternary systems) or using CDs as a partner of the DES constitution (binary systems) [17]. In both cases, the development of these new supramolecular mixtures requires the study and understanding of the CD/guest inclusion mechanism in these non-conventional media. However, to the best of our knowledge, no study has been conducted to investigate the thermodynamics of the inclusion mechanism of CD in DES/water mixtures. The present paper investigates, for the first time, the thermodynamic parameters of the CD inclusion complex in different DES/water mixtures. β-CD/adamantanol was chosen as a model inclusion complex due to the high recognition ability of β-CD towards this guest (Figure 1). Eight choline chloride (ChCl)-based DES described in the literature were prepared [18] (Figure 1). Then, the obtained liquid solvents were mixed with different amounts of water to form various DES/water mixtures. The competitive and solvation contributions of these mixtures were then investigated. The results were compared with those obtained in water, individual components of DESs and ethanol/water mixtures.

2. Experimental section

2.1. *Materials*

Choline chloride (ChCl, 98%) was purchased from Sigma-Aldrich, China. Urea (99%) was purchased from Sigma-Aldrich, USA. Propane-1,3-diol, glycerol, D-(−)-sorbitol, D-(−)-fructose and D-(+)-glucose anhydrous were provided by VWR Chemicals, Belgium. Xylitol and (±)-butane-1,3-diol were provided by Alfa Aesar, Germany. β-CD was provided by Wacker-Chemie, Lyon, France. 1-Adamantanol was purchased from Acros, Belgium. ChCl was dried at 60 °C for at least 2 weeks before use. All other compounds were used as received. Distilled water was used throughout this work.

2.2. *Preparation of deep eutectic solvents*

DESs were prepared using the heating method by mixing ChCl as the HBA and different HBDs (urea, glycerol, xylitol, sorbitol, butane-1,3-diol, propane-1,3-diol, glucose and fructose) (Figure 1) under stirring until the formation of a clear and homogenous liquid (Table 1). HBA: HBD molar ratios and temperature used for DES preparation are presented in Table 1. The water content of the prepared DESs were determined using the Karl Fischer titration method (Mettler Toledo DL31) and listed in Table 1.

2.3. *Isothermal titration calorimetry (ITC)*

ITC experiments were conducted using a VP-ITC, MicroCal Inc., USA. Titration experiments were carried at three different temperatures (15, 25 and 35 °C) in water, 25 wt% (weight percent) of each studied DES component alone in water, DES/water mixtures (25, 50 and 70 wt% DES) and ethanol/water mixtures (25, 50 and 70 wt% ethanol). Solutions of β-CD (5 mM) and adamantanol (0.5 mM) were individually prepared in each studied mixture and loaded respectively in the syringe and the cell of the calorimeter. When the obtained K_f value was less than 3000 M^{-1}, release experiments were also performed in order to increase the accuracy of the determined thermodynamic parameters. For release experiments, the solution of inclusion complex (5 mM of β-CD with 5 mM of adamantanol) was loaded in the syringe and the corresponding solution (the studied mixture) in the cell.

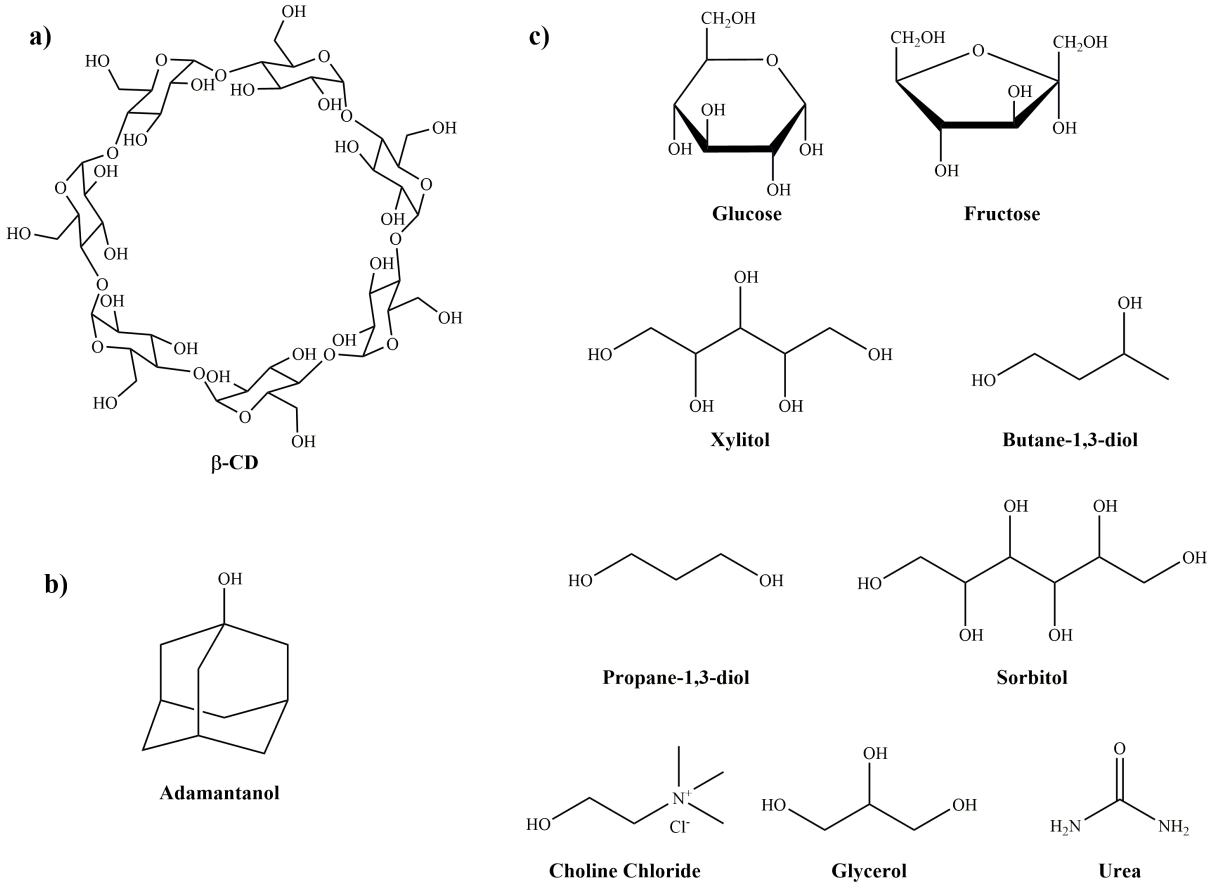


Figure 1. Chemical structure of (a) β-CD, (b) adamantanol and the (c) DESs components.

Table 1. Molar ratio of preparation, heating temperature and water content (%) of the studied DESs

DES	Molar ratio	Heating temperature (°C)	Water content (%)
ChCl:urea	1:2	60	0.05
ChCl:glycerol	1:2	Room temperature	0.20
ChCl:xylitol	1:2	80	0.07
ChCl:sorbitol	1:1	80	0.36
ChCl:butane-1,3-diol	1:2	40	0.08
ChCl:propane-1,3-diol	1:1	80	0.21
ChCl:glucose	2:1	80	0.26
ChCl:fructose	2:1	80	2.15

All experiments consisted of an initial injection of 25 μL followed by 10 injections of 25 μL (delivery time of each injection: 50 s; interval between two consecutive injections: 60 s; stirring speed: 1000 rpm). Blank experiments were conducted and the heats of dilution were subtracted from the measured heat of the corresponding titration or release experiment. Formation constants (K_f) and thermodynamic parameters (ΔH, ΔS and ΔG) were determined by global analysis of all isotherms obtained for

Table 2. Formation constant (K_f) values of β-CD/adamantanol inclusion complex in the absence and the presence of 25 wt% of each DES component at 25 °C

Medium		K_f (M^{-1})
Water	-	29,700 ± 400
25 wt% DES component	ChCl	23,000 ± 2000
	Urea	29,000 ± 1000
	Glycerol	16,800 ± 900
	Xylitol	24,000 ± 1000
	Sorbitol	34,000 ± 2000
	Butane-1,3-diol	2000 ± 300
	Propane-1,3-diol	3800 ± 400
	Glucose	44,000 ± 2000
	Fructose	27,000 ± 2000

a given solvent, using an algorithmic treatment as described in [19].

3. Results and discussion

3.1. *Influence of DES components on adamantanol complexation in cyclodextrin*

First, the effect of the presence of each DES component on the stability of the β-CD/adamantanol inclusion complex was evaluated. ITC titration experiments were carried out in water containing 25 wt% of each DES component at different temperatures (15, 25 and 35 °C). The K_f values determined at 25 °C are listed in Table 2 in comparison to those obtained in water. All thermodynamic parameters (K_f, ΔH, ΔS and ΔG) obtained at 15, 25 and 35 °C are given in Table S1.

The obtained K_f value of β-CD/adamantanol at 25 °C in water was equal to 29,700 M^{-1}. This high K_f value is due to the strong binding of the adamantyl moiety to the cavity of β-CD [20]. A decrease in the K_f value indicates that the DES component has a competitive effect for the β-CD cavity while an increase in the magnitude of this thermodynamic parameter reveals a stabilizing effect of the studied compound.

As can be seen in Table 2, ChCl, the HBA used for the preparation of the eight DESs in this study, has very low competitive effect on the β-CD/adamantanol inclusion complex, as K_f value was slightly affected by its presence in solution. The results collected for the HBD showed diverse effects for the eight compounds. Some presented a strong (butane-1,3-diol, propane-1,3-diol) or weak (glycerol) competitive effect, while others showed either no effect (urea, xylitol, fructose) or a weak stabilizing effect (sorbitol, glucose). Additionally, results showed that the presence of each of the studied DES components at 25 wt% maintained the enthalpy driven mechanism of the inclusion complexation of adamantanol in β-CD (Table S1).

3.2. *Influence of DES on adamantanol complexation in cyclodextrin*

Titration and release experiments were performed in different DES/water and ethanol/water mixtures (25, 50 and 70 wt% solvent) at different temperatures (15, 25 and 35 °C). It should be emphasized that ITC studies were not carried out above 70 wt% due to the excessive viscosity of the mixtures, which prevented reliable recording of heat release. An example of the results of titration and release experiments for β-CD/adamantanol inclusion complex in ChCl:butane-1,3-diol (70 wt%) at 25 °C is represented in Figure 2. The obtained results were treated simultaneously using a global analysis and compared with those found in water. The K_f values of β-CD/adamantanol inclusion complex determined at 25 °C in the different mixtures are listed in Table 3. All thermodynamic parameters obtained at 15, 25 and 35 °C are represented in Tables S2, S3 and S4, respectively.

In the presence of 25 wt% DES, the results (Table 3) indicated that only ChCl:butane-1,3-diol and ChCl:propane-1,3-diol significantly reduced the stability of the inclusion complex. This is coherent with the findings obtained in the presence of 25 wt% of individual DESs components as the lowest K_f values were observed in the presence of butane-1,3-diol and propane-1,3-diol (Table 2). This observation led us to suggest that one can infer the competitive behavior of a specific DES/water mixture (at least for weak DES content) by testing the competitive behavior of the aqueous solutions of each component. As for butane-1,3-diol and propane-1,3-diol, ethanol showed also a strong competitive effect on the complexation of adamantanol in β-CD.

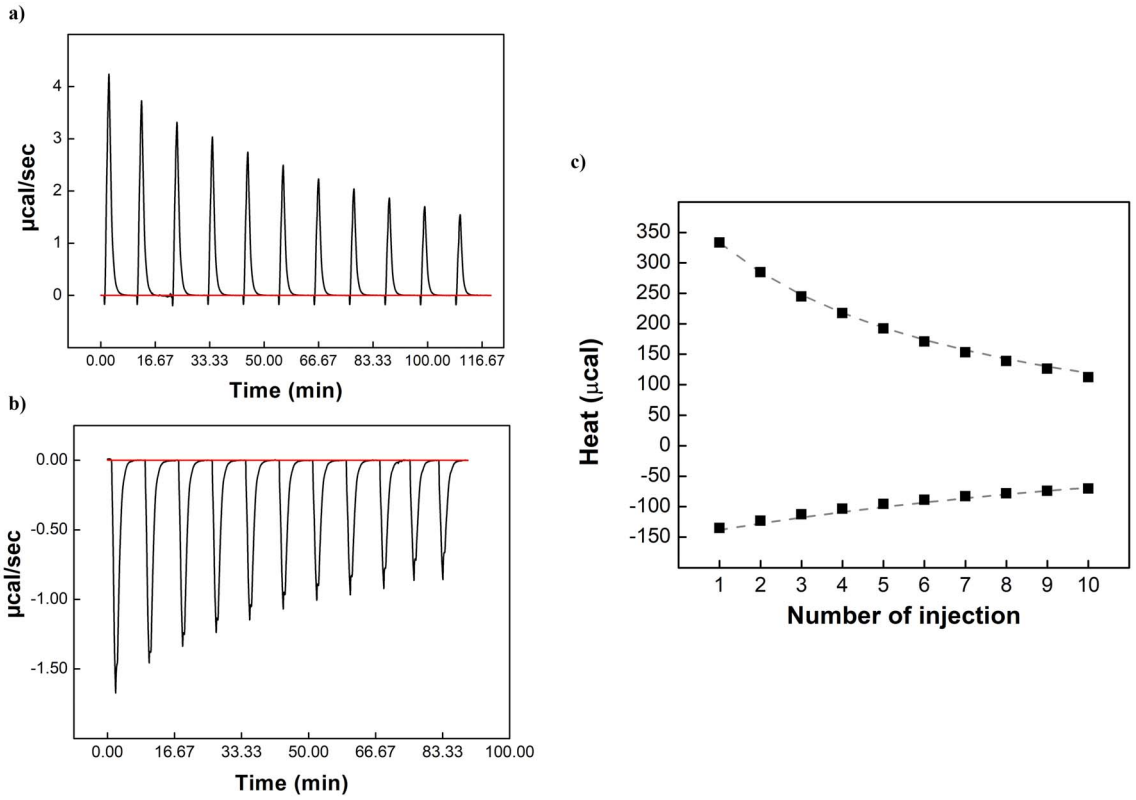


Figure 2. Typical isothermal titration calorimetry data obtained for the binding interaction of β-CD and adamantanol in ChCl:butane-1,3-diol DES/water 70/30 wt% at 25 °C. (a) Release thermogram, (b) titration thermogram (c) corresponding isotherms presenting the integrated heat data; the dashed curves correspond to the theoretical curves.

Table 3. K_f values for β-CD/adamantanol inclusion complex determined in the different DES/water and ethanol/water mixtures at 25 °C in comparison to water

	K_f (M^{-1})		
Water	29,700 ± 400		
Solvent wt%	25 wt%	50 wt%	70 wt%
ChCl:urea	30,000 ± 2000	21,000 ± 1000	12,100 ± 700
ChCl:glycerol	21,000 ± 1000	10,500 ± 600	5600 ± 300
ChCl:xylitol	28,000 ± 2000	17,000 ± 1000	ND
ChCl:sorbitol	31,000 ± 2000	19,000 ± 1000	ND
ChCl:butane-1,3-diol	4300 ± 500	1500 ± 300	500 ± 20
ChCl:propane-1,3-diol	10,100 ± 800	4300 ± 400	2600 ± 400
ChCl:glucose	31,000 ± 2000	15,000 ± 1000	ND
ChCl:fructose	26,000 ± 2000	12,500 ± 100	ND
Ethanol	1700 ± 100	1800 ± 400	100 ± 10

ND: Not determined (ITC experiments were not carried out because of the high viscosity of these media).

This observation is in good agreement with previous studies reporting a decrease in K_f in the presence of ethanol [7,21,22].

Concerning the effect of the wt% of DES, results showed that as the wt% of the solvent increased, the stability of the β-CD/adamantanol inclusion complex decreased progressively. This was observed for the eight DESs and not only for those that showed a competitive effect. This indicated that at higher DES content (>25 wt%), the destabilization of the inclusion complex was not solely induced by the competitive effect of DES components.

Moreover, a negative heat capacity change (ΔC_p) was observed in water for β-CD/adamantanol inclusion complex (-95 J·K^{-1}). This is typical of macromolecular associations in water and evidenced the significant involvement of the hydrophobic interactions in the binding process [23,24]. In DES based media (70 wt% DES content), ΔC_p values ranged between -51 and -16 J·K^{-1}. This increase in ΔC_p values confirmed the disruption of the solvatophobic forces in the presence of DES leading to the decrease in the stability of the inclusion complex.

Altogether, these results proved that at high DES wt% content, an additional phenomenon related to the solvation of CD and adamantanol affected the stability of the inclusion complex. When thoroughly adding DES to water, β-CD and adamantanol experiment more favorable interactions with the added surrounding solvent molecules, which in turn disfavor the complex formation. This induces a decrease in the solvatophobic forces, the governing forces for the inclusion complex formation. Among all DESs and for all DES wt% content, the lowest K_f values were observed for ChCl:butane-1,3-diol, consecutively to the highest solvation potency and competitive behavior of butane-1,3-diol. The destabilizing effect of ethanol was even more pronounced, for all solvent contents, in accordance with literature. This is due to the fact that ethanol is able to effectively solvate guest molecules and consequently subtract them from the complexation in the CD cavity [7,21,22].

3.3. *Influence of DES on the complexation mechanism of adamantanol in cyclodextrin*

The thermodynamic parameters of β-CD/adamantanol inclusion complex (ΔH, −TΔS and ΔG) were determined in water, in the different DES/water and ethanol/water mixtures at different temperatures. Results obtained at 15, 25 and 35 °C are presented in Table S2, S3 and S4, respectively. Figure 3 illustrates the results determined at 25 °C in the presence of 25 wt% DES in comparison to water and 25 wt% ethanol.

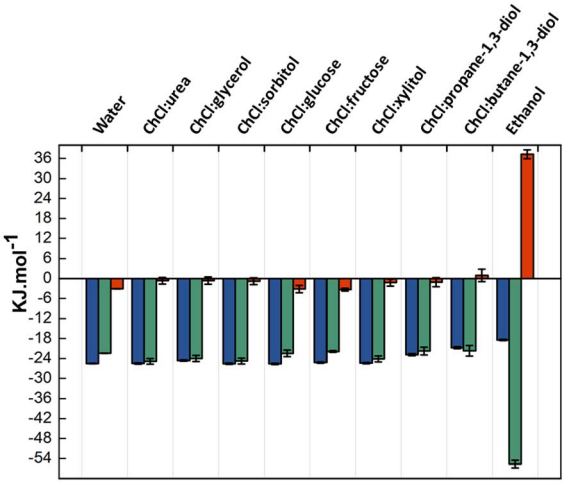


Figure 3. Evolution of the ΔG, ΔH and −TΔS in the different DESs (25 wt%) in comparison to water and ethanol (25 wt%) at 25 °C.

The exothermic character (ΔH < 0) of the binding process between β-CD and adamantanol explained the observed decrease in K_f values when increasing temperatures in water and all studied mixtures [25].

In water and for all DES/water mixtures, the entropy effect (−TΔS) of the β-CD/adamantanol complex was fairly close to zero over the studied temperature range. Accordingly, the entropy variation makes a minor contribution to the negative values of standard Gibbs energy (ΔG), to the contrary of the enthalpic term. The observed entropy variation may be due to the compensation between the release of water molecules from the cavity, which has positive contribution to entropy [25,26] and the β-CD/adamantanol complex formation, which has negative contribution to entropy. Indeed, once adamantanol is encapsulated in the cavity of the CD, it leads to a significant loss in the degree of reagent freedom, which consequently, causes the loss of entropy (ΔS < 0, −TΔS > 0) [27]. Altogether, these observations indicated that the binding event between β-CD and adamantanol in water and all

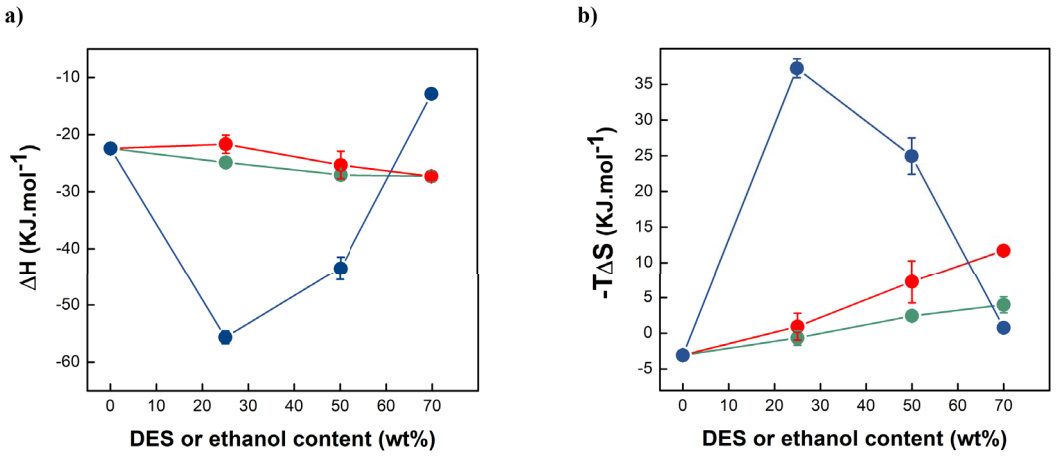


Figure 4. Evolution of (a) ΔH and (b) −TΔS with increasing the wt% of ChCl:urea, ChCl:butane-1,3-diol and ethanol at 25 °C.

DES/water mixtures was similar and was enthalpy driven (i.e., exothermic binding events with $|\Delta H| > |-T\Delta S|$) with low entropic assistance [28].

In ethanol/water mixtures, a different thermodynamic signature was detected. An entropy–enthalpy compensation of the binding process appeared at all temperatures and for all ethanol wt%. This could be explained by the fact that when moving from water to a hydroalcoholic medium, the thermodynamic profile of the inclusion complex formation can be influenced by the variation of the solvation of the guest (adamantanol). Ethanol solubilizes adamantanol by displacing water molecules in its hydration shell, affecting consequently both the enthalpy and entropy of the complexation. Enthalpy–entropy compensation was reported in literature when water was progressively replaced by an organic solvent [29,30]. This can be confirmed by the fact that increasing ethanol wt% prevents the formation of inclusion complex (lowest K_f values) by effectively solvating guest molecules and thus by subtracting them from the complexation. This observed difference in the enthalpic and entropic contributions to Gibbs energy of complexation between DES/water and ethanol/water mixtures underlined the strong difference in solvation mechanisms between the DES and ethanol solutions.

The evolution of the thermodynamic parameters (ΔH and −TΔS) as function of solvent content wt% (Figure 4) additionally emphasized on the pronounced competition and solvation effects of ethanol, which led respectively to a strong increase then decrease of enthalpic (ΔH up to -55.7 kJ·mol^{-1} at 25 wt% and down to -12.9 kJ·mol^{-1} at 70 wt%) and entropic ($-T\Delta S$ up to 37.2 kJ·mol^{-1} at 25 wt% and down to 0.8 kj·mol^{-1} at 70 wt%) contributions. On one hand, since the inclusion of ethanol in CDs is generally favored by entropy (i.e., $\Delta H = -2.4$ kJ·mol^{-1}; $-T\Delta S = -3.3$ kJ·mol^{-1} for α-CD/ethanol [31]) and the inclusion of adamantanol by enthalpy (Figure 3), replacement of ethanol molecules by an adamantanol molecule is characterized by strong enthalpy and entropy variations, thus explaining the thermodynamic signature of competitive phenomena (25 wt%). On the other hand, the fact that both enthalpy and entropy of inclusion were strongly reduced at high ethanol content (70 wt%) tends to demonstrate that adamantanol inclusion in CD cavity is somewhat analogous to its solvation by ethanol.

The evolution of both ΔH and −TΔS was far less pronounced for both non-competitive (ChCl:urea) and weakly competitive (ChCl:butane-1,3-diol) DESs, with a moderate increase of enthalpy contribution overbalanced by a negative contribution of entropy, with increasing DES content. Such results showed that adamantanol interacted more favorably with β-CD than with the DES components, but that solvation by DES induced more disorder than inclusion in β-CD.

4. Conclusion

In this study, the thermodynamic parameters of the β-CD/adamantanol inclusion complex in non-conventional green media were investigated for the first time. The studied media included water, DES/water mixtures, aqueous solutions of DES components and ethanol/water mixtures. Eight different DESs were investigated at different wt% content (25, 50 and 75 wt%). According to ITC results, the type and wt% of DES affected differently the inclusion complex stability. A destabilizing effect was observed for all DESs in media with a high concentration of co-solvent (>50 wt%). However, DESs were much less destabilizing than ethanol. In addition, DESs did not disturb the complexation mechanism as observed in the presence of ethanol. Also, the thermodynamic signature of the complexation in DES/water mixtures was similar to that in water and was enthalpy driven with low entropic assistance, in contrast to ethanol/water mixtures. Taken together, the results proved that replacing organic solvents with DESs could be advantageous in maintaining the complexation and solvation capacity of CDs in a drier environment. Nevertheless, further investigations and refinements with other guest species, e.g. with lower binding affinity to CDs, should be explored to confirm whether they could lead to the same thermodynamic behavior of the complexation in the presence of DESs. In addition, other DES systems, belonging to the five DES classes, should be investigated in a broader comparative study.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

The authors would like to acknowledge the National Council for Scientific Research of Lebanon (CNRS-L) and Université du Littoral Côte d'Opale (ULCO) for granting a doctoral fellowship to LN. This work is a contribution to the CPER (Contrat de Plan Etat-Région) research project IRenE (Innovation et Recherche en Environnement) and is supported by the French Ministère de l'Enseignement Supérieur, the region Hauts-de-France and the European Regional Development Fund.

Supplementary data

Supporting information for this article is available on the journal's website under https://doi.org/10.5802/crchim.304 or from the author.

References

[1] G. Crini, *Chem. Rev.*, 2014, **114**, 10940-10975.
[2] G. Crini, *Progr. Polym. Sci.*, 2005, **30**, 38-70.
[3] T. Loftsson, H. H. Sigurdsson, P. Jansook, *Materials*, 2023, **16**, article no. 2223.
[4] N. Morin-Crini, S. Fourmentin, É. Fenyvesi, E. Lichtfouse, G. Torri, M. Fourmentin, G. Crini, *Environ. Chem. Lett.*, 2021, **19**, 2581-2617.
[5] Y. Zheng, L.-N. Dong, M. Liu, A. Chen, S. Feng, B. Wang, D. Sun, *J. Agric. Food Chem.*, 2014, **62**, 244-250.
[6] A. Illapakurthy, C. Wyandt, S. Stodghill, *Eur. J. Pharm. Biopharm.*, 2005, **59**, 325-332.
[7] T. Usacheva, D. Kabirov, D. Beregova, G. Gamov, V. Sharnin, M. Biondi, L. Mayol, F. D'Aria, C. Giancola, *J. Therm. Anal. Calorim.*, 2019, **138**, 417-424.
[8] A. P. Abbott, *Curr. Opin. Green Sustain. Chem.*, 2022, **36**, article no. 100649.
[9] A. R. Jesus, A. Paiva, A. R. C. Duarte, *Curr. Opin. Green Sustain. Chem.*, 2023, **39**, article no. 100731.
[10] L. Nakhle, M. Kfoury, I. Mallard, D. Landy, H. Greige-Gerges, *Environ. Chem. Lett.*, 2021, **19**, 3747-3759.
[11] M. A. R. Martins, S. P. Pinho, J. A. P. Coutinho, *J. Solution Chem.*, 2019, **48**, 962-982.
[12] D. O. Abranches, J. A. P. Coutinho, *Curr. Opin. Green Sustain. Chem.*, 2022, **35**, article no. 100612.
[13] L. Nakhle, M. Kfoury, I. Mallard, H. Greige-Gerges, D. Landy, *J. Mol. Liquids*, 2022, **359**, article no. 119371.
[14] T. El Achkar, H. Greige-Gerges, S. Fourmentin, *Environ. Chem. Lett.*, 2021, **19**, 3397-3408.
[15] T. El Achkar, T. Moufawad, S. Ruellan, D. Landy, H. Greige-Gerges, S. Fourmentin, *Chem. Commun.*, 2020, **56**, 3385-3388.
[16] P. Janicka, M. Kaykhaii, J. Płotka-Wasylka, J. Gębicki, *Green Chem.*, 2022, **24**, 5035-5045.
[17] M. Kfoury, D. Landy, S. Fourmentin, *Curr. Opin. Green Sustain. Chem.*, 2022, **36**, article no. 100630.
[18] Y. Dai, J. van Spronsen, G.-J. Witkamp, R. Verpoorte, Y. H. Choi, *Anal. Chim. Acta*, 2013, **766**, 61-68.
[19] E. Bertaut, D. Landy, *Beilstein J. Org. Chem.*, 2014, **10**, 2630-2641.
[20] K. Leko, M. Hanževački, Z. Brkljača, K. Pičuljan, R. Ribić, J. Požar, *Chem. Eur. J.*, 2020, **26**, 5208-5219.
[21] L. Nakhle, M. Kfoury, H. Greige-Gerges, S. Fourmentin, *J. Mol. Liquids*, 2020, **310**, article no. 113156.
[22] M. Kfoury, C. Geagea, S. Ruellan, H. Greige-Gerges, S. Fourmentin, *Food Chem.*, 2019, **278**, 163-169.

[23] M. M. Lopez, G. I. Makhatadze, *Biophys. Chem.*, 1998, **74**, 117-125.
[24] T. K. Dam, S. Oscarson, C. F. Brewer, *J. Biol. Chem.*, 1998, **273**, 32812-32817.
[25] D.-Z. Sun, L. Li, X.-M. Qiu, F. Liu, B.-L. Yin, *Int. J. Pharm.*, 2006, **316**, 7-13.
[26] M. Wszelaka-Rylik, P. Gierycz, *J. Therm. Anal. Calorim.*, 2013, **111**, 2029-2035.
[27] A. Cooper, *Curr. Opin. Chem. Biol.*, 1999, **3**, 557-563.
[28] A. Cooper, C. M. Johnson, J. H. Lakey, M. Nöllmann, *Biophys. Chem.*, 2001, **93**, 215-230.
[29] A. F. Danil de Namor, R. Traboulssi, D. F. V. Lewis, *J. Am. Chem. Soc.*, 1990, **112**, 8442-8447.
[30] P. L. Irwin, G. King, K. B. Hicks, *Carbohydr. Res.*, 1996, **282**, 65-79.
[31] M. V. Rekharsky, Y. Inoue, *Chem. Rev.*, 1998, **98**, 1875-1918.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Chimie
2024, Vol. 27, Special issue S5, p. 117-123
https://doi.org/10.5802/crchim.300

Research article

French Network on Solvation (GDR 2035 SolvATE)

How NaCl addition destabilizes ionic liquid micellar suspension until phase separation

Jean-François Dufrêche [a], Marie Plazanet [*, b], Gautier Meyer [b] and Isabelle Billard [c]

[a] ICSM, Univ Montpellier, CEA, CNRS, ENSCM, Bagnols-sur-Cèze, France

[b] Laboratoire Interdisciplinaire de Physique, LIPhy, CNRS & Univ. Grenoble-Alpes, Grenoble, France

[c] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, Grenoble INP, LEPMI, 38000 Grenoble, France

E-mails: jean-francois.dufreche@icsm.fr (J.-F. Dufrêche), marie.plazanet@univ-grenoble-alpes.fr (M. Plazanet), gautier.meyer@univ-grenoble-alpes.fr (G. Meyer), isabelle.billard@grenoble-inp.fr (I. Billard)

Abstract. The ionic liquid tributyltetradecylphosphonium chloride ($[P_{4,4,4,14}]Cl$) forms micelles in water, with a very low CMC, below 1 wt%. The solution is macroscopically homogeneous, even large amounts of $[P_{4,4,4,14}]Cl$ in water do not induce any phase separation. The ternary system $[P_{4,4,4,14}]Cl/NaCl/H_2O$ instead displays a LCST (Lower Critical Separation Temperature) behavior, being monophasic at low T and experiencing phase separation when T is increased. This phenomenon has been ascribed to the T-increased adsorption onto the micellar surface of these additional chloride ions. The lowering of the repulsive interactions between micelles finally allows coalescence and thus phase separation. In this work, we explore the impact of NaCl addition onto the phase separation, at fixed T. Specific chloride electrode allows the determination of chloride counterion adsorption for different samples in the phase diagram, all of them being single-phase. A simple theory based on the Poisson–Boltzmann equation and with charge regulation is proposed. The only fitted parameter is the chloride adsorption constant. It enables to model the different populations of ions in the solution and at the micelle surface in different conditions. Considering the effective charge of the micelles with respect to the ionic strength of the solution, it moreover provides a key element in the prediction of phase separation.

Keywords. Aqueous biphasic solution, Extraction, Ionic liquid, Regulation charge theory.

Funding. French National Agency for Research (Grant No. ANR-ITALLIX-22-CE29-0023-01).

Manuscript received 1 November 2023, revised 8 December 2023, accepted 9 February 2024.

1. Introduction

Aqueous biphasic systems (ABS) recently deserve a lot of attention thanks to their environmental friendly composition, with a large water content. They have been investigated for multiple extraction purposes in order to substitute the use of traditional highly polluting organic solvents [1,2]. Among ABS, the family comprising the ionic liquid (IL) tributyltetradecylphosphonium chloride ($[P_{4,4,4,14}]Cl$) mixed with water and a strong acid has been proposed for metallic ion extraction [3–5]. Metallic ions have preferential phases, as for example Co(II) prefers the

* Corresponding author

upper, ionic liquid-rich phase while Ni(II) prefers the lower highly acid solution. Beside the interest in metal extraction, these ternary systems are surprising with a very rich phase diagram, either in presence of a strong acid or a salt (for example, HCl or NaCl). The solutions are thermomorphic, i.e. reversibly change from a monophasic to a biphasic liquid state with temperature, and with the particularity to separate upon an increase of temperature, meaning that the biphasic region of the phase diagram increases with temperature, therefore also having a *Lower Solution Critical Temperature* (LCST). The mono- or biphasic states are related to the composition and, at any fixed temperature, the increase of both IL or salt/acid content triggers the phase separation. The metallic ions extracted will therefore play themselves a determinant role in their own extraction [6–8]. Understanding the mechanisms of phase separation in such family of systems is therefore a challenging task that we partially address in this paper. Our previous investigation [9] was focused on the structural organization of the acidic (HCl, H_2SO_4) solutions with concentration and temperature. We highlighted that the IL $[P_{4,4,4,14}]Cl$, which could be assimilated to a cationic surfactant, forms spherical micelles in solution. Upon temperature rise, the micelles aggregate until eventually causing the phase separation. The aggregation is due to an adsorption of chloride ions at the micelles surface with temperature, as proved by the titration of free chloride ions in solution: the adsorption of ions causes a variation in the Electric Double Layer (EDL) composition around the micelles and a screening of electrostatic repulsion bewteen micelles. Our previous structural investigation also shows an aggregation of the micelles when the acid content increases. Here, we question the similarity of the phase separation with temperature and salt concentration in the system. Using the former chloride titration technique at fixed temperature but variable salt content, we aim at rationalizing the electrostatic interactions responsible for the state of the solution. After describing the experimental methods, we present the experimental results: the investigated points are presented in the Figure 1 on the phase diagram of the particular solution investigated here: $[P_{4,4,4,14}]Cl$, NaCl and water. We then detail the theoretical approach based on a classical charge regulation theory to model the results. We eventually discuss the validity and limits of the model.

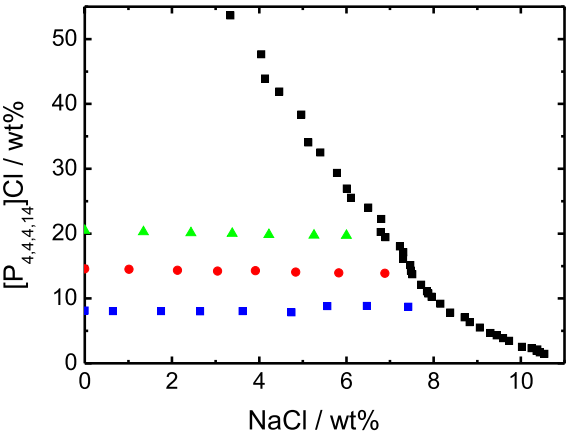


Figure 1. Binodal curve of the system $[P_{4,4,4,14}]Cl$, NaCl and water, and points investigated by chloride titration. Binodal data (black squares) from [10]. The three series of points correspond to the following IL wt%: 8.12 (blue squares), 14.55 (red circles) and 20.48% (green triangles).

2. Materials and methods

2.1. *Chemicals*

Tributyltetradecylphosphonium chloride ($[P_{4,4,4,14}]Cl$) has been provided by Interchim, while NaCl was purchased from Honeywell. All chemicals were used as received. All samples were made using ultra-pure water (Millipore system, 18 MΩ).

2.2. *Samples*

Desired amounts of NaCl and of a concentrated aqueous solution of $[P_{4,4,4,14}]Cl$ were weighted in a 10 mL gauged flask and then ultra-pure water was added to the line. The composition of all the samples can be found in the ESI.

2.3. *Apparatuses and methods*

A balance (Fisherbrand, Analytical Series, precision 0.0001 g) was used to prepare the samples. Free chloride ions concentration (M) was determined by use of a chloride specific electrode (ThermoScientific, chloride half cell Orion 9417SC and reference cell), which was calibrated by seven standard aqueous solutions of NaCl with concentrations ranging

from 6×10^{-3} M to 1.1 mol·L^{-1}. The calibration was found linear with very good regression parameter and under the experimental procedure followed, reproducibility is within ±2%. For both calibration and measurement of the samples, temperature was controlled at $T = (25 \pm 0.5)$ °C by a thermostated bath. Densities at $T = 25$ °C were measured by use of a density-meter (Anton-Paar, DMA 4001, precision 10^{-4} g·cm^{-3}). Concentrations in moles per liter (M) were calculated from the weighted masses and flask volume while NaCl and $[P_{4,4,4,14}]Cl$ wt% were calculated from added masses, density and flask volume. Samples are sorted within three series, for which a fixed mass of IL is mixed with increasing amounts of NaCl. To label the series, we use the IL wt% in the absence of NaCl, i.e. 8.12/14.55/20.48 wt% of IL. This corresponds respectively to 0.185, 0.329 and 0.462 mol·L^{-1}.

The binodal data obtained by Schaeffer et al. [10] have been obtained with Iolitech as the IL provider, while our samples are prepared with an Interchim batch of IL. Nevertheless, both data sets agree very well in terms of monophasic domain (see Figure 1). In addition, we prepared one sample supposedly being biphasic and very close to the binodal, according to the data by Schaeffer et al. and it actually appears to be biphasic (see ESI).

3. Experimental results

In the Figure 2 are plotted the experimental measurements of the chloride concentrations in the solutions. The first point of each experimental series corresponds to the case without any addition of NaCl and the horizontal lines are the experimental averages of each series. The line $x = y$ would be the behavior without any adsorption at the micelle surface (see caption). The chloride specific electrode is sensitive to the chemical potential of the ion, which is converted directly into the equivalent concentration of chloride ion in an aqueous electrolyte solution in equilibrium with the system. By neglecting the activity coefficient corrections, which are relatively small in this chloride concentration range [11], one can consider that this quantity corresponds to the bulk free chloride ion concentration $[Cl^-]_{free}$, i.e. the Cl^- concentration far away from the micelles. The quantity of chlorides adsorbed at the micelle surface is therefore equal to $[Cl^-]_{ads} = [Cl^-]_{tot} - [Cl^-]_{free}$. The

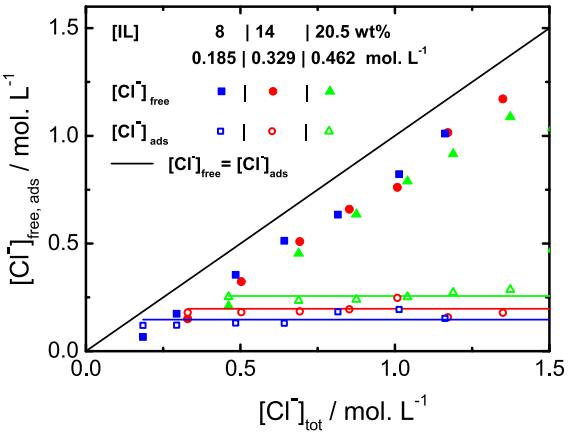


Figure 2. Chloride titration for three different ionic liquid contents, as a function of total amount (in mol·L^{-1}) of chloride. The $[Cl^-]_{free}$ are directly measured and the $[Cl^-]_{ads}$ are deduced from the measurement by the relation $[Cl^-]_{tot} = [Cl^-]_{ads} + [Cl^-]_{free}$. The line $x = y$ would be the behavior without any adsorption at the micelle surface.

data are presented up to the limit of linearity of the electrode, that does not allow a reliable measurement of the region too close to the binodal line. For this reason, more points are plotted in the Figure 1 than in the Figure 2.

As observed in the Figure 2, without any addition of NaCl, about 40% of the Cl^- are free in the solution, whatever the mass fraction of ionic liquid. More precisely, the fraction of adsorbed chloride at the surface follows the linear behaviour given by the equation: $[Cl^-]_{ads} = 0.39 \cdot [LI] + 0.071$, both concentration in mol·L^{-1}. The micelles therefore carry an effective charge for any fraction of IL. This result is consistent with the structural characterization of the binary mixture (IL + water) [9], where a strong correlation peak between micelles is observed, arising from the electrostatic repulsion between the objects in solution in absence of additional charges. Upon addition of NaCl, the adsorbed quantity of chloride is constant, all the ions are dispersed into the solution, in opposition to what is observed with temperature.

4. Theory

As previously expressed, we consider that the quantity measured by the electrode corresponds to the

bulk free chloride ion concentration far away from the micelles, where the aqueous solution plays the role of a reservoir. The measurement can therefore be used to establish a link between the total concentration of introduced chlorides $[Cl^-]_{tot}$ and $[Cl^-]_{free}$.

This result can then be compared with those obtained using an Electrical Double Layer (EDL) model with charge regulation. The total chloride concentration reads

$$[Cl^-]_{tot} = [Cl^-]_\chi + [Cl^-]_{EDL} + [Cl^-]_\infty \tag{1}$$

where $[Cl^-]_\chi$ represents the concentration of the ions that are specifically bound to the surface, on particular adsorption sites. $[Cl^-]_{EDL}$ is the excess concentration of chlorides in the electrical double layer. This corresponds to the ions electrostatically bound to the surface, beyond the Stern layer represented by $[Cl^-]_\chi$. $[Cl^-]_\infty$ is the global concentration of the free ions far away from the micelle. It is not exactly equal to the reservoir concentration $[Cl^-]_{free}$ because the free ions cannot penetrate the micelles.

The mass action law for the equilibrium between free ions and ions bound specifically to the surface reads as follows:

$$\rho_\chi^S = K° C_0 e^{\frac{e\psi_S}{k_B T}} \tag{2}$$

ρ_χ^S is the area number density (i.e. the number of bound ions per unit area). e is the elementary charge, $k_B T$ is the thermal energy and ψ_S is the electrostatic potential at the micelle surface. $K°$ is the mass action law constant of the adsorption reaction. $C_0 = [Cl^-]_{free}$ is the chloride concentration in the reservoir where the electrostatic potential $\psi_S = 0$. We consider the regime [12] of strong electrostatic screening for which $\kappa R > 1$ where R is the micelle radius and κ the Debye parameter. If $R \approx 2$ nm, this corresponds to the case where the salt concentration is greater than 2×10^{-2} mol·L^{-1}. In this regime the interface is almost flat compared to the Debye distance. Thus the Gouy–Chapman solution of the Poisson–Boltzmann equation can be used.

The SI practical unit of the mass action law constant $K°$ is meter if the reservoir salt concentration $C_0 = [Cl^-]_{free}$ is the volume number density (i.e. the number of electrolyte per unit volume). The effective charge of the micelles is then:

$$\sigma^{eff} = \sigma - eK° C_0 e^{\frac{e\psi_S}{k_B T}} \tag{3}$$

with σ the bare charge. The diffusive part of the EDL is modelled by the Poisson–Boltzmann equation. Thus the effective charge can also be calculated thanks to the Grahame equation:

$$\sigma^{eff} = \sqrt{8\epsilon_0 \epsilon_r k_B T C_0} \sinh\left(\frac{e\psi_S}{2k_B T}\right) \tag{4}$$

$\epsilon_0 \epsilon_r = \epsilon$ is the permittivity of water. The two equations (3) and (4) have to be solved numerically to obtain σ^{eff} and ψ_S. We finally obtain from this self-consistent calculation:

$$\rho_\chi^S = \frac{\sigma - \sigma^{eff}}{e} \tag{5}$$

Then we calculate the chloride excess in the EDL. Considering the Gouy–Chapmann equation [13] yielding the anion concentration $C_-(x)$ as a function of the position x with respect to the surface, we obtain the excess chloride area number density

$$\rho_{EDL}^S = \int_0^{+\infty} (C_-(x) - C_0)\, dx = \frac{4AC_0}{(1-A)\kappa} \tag{6}$$

with $A = \tanh(e\psi_S/4k_B T)$ and the Debye parameter $\kappa = (2e^2 C_0/\epsilon_0 \epsilon_r k_B T)^{1/2}$.

These area number densities ρ_χ^S and ρ_{EDL}^S must then be transformed into volume number densities to obtain the total experimental concentration $[Cl^-]_{tot}$. So multiplying by the specific surface area of the micelles S_V^{mic}, we finally obtain from (1):

$$[Cl^-]_{tot} = S_V^{mic}\left(\rho_\chi^S + \rho_{EDL}^S\right) + (1-\eta)C_0 \tag{7}$$

where η is the volume fraction of the micelles supposed to be spherical. Here are the model parameters we used. Scattering experiments [9] allow us to specify the radius of the micelles $R = 18$ Å, the aggregation number $N_{agg} = 30$. The resulting volume fractions η for the three series are 9.1, 16.1 and 22.7% and the resulting surface charge density is $\sigma = eN_{agg}/4\pi R^2 = 0.737$ e·nm^{-2}. The only unknown parameter in the model is therefore the adsorption constant $K°$. This value has been fitted from the experimental curve at the lower IL concentration (8.16 wt%). We obtained $K° = 1$ nm. The other curves are therefore true predictions, since they were calculated without any adjustable parameters.

In the Figure 3, we inverted the axes in order to stick to the experimental observables. Despite the simplicity of the model we have a quantitative agreement with the data for the range of chloride and ionic liquid concentrations investigated, although some deviations are observed and discussed below.

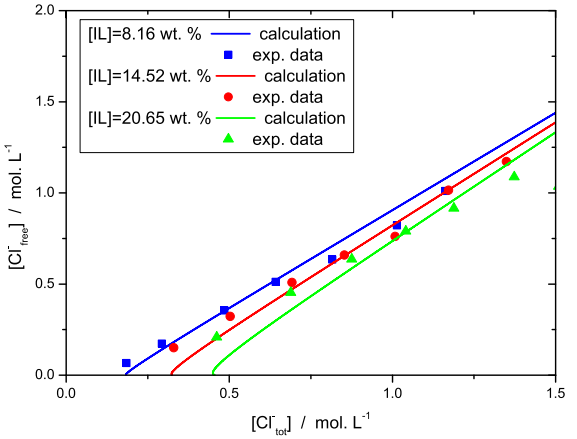


Figure 3. Free chloride in the solution (i.e. chloride concentration in the reservoir) as a function of total amount of chloride in the systems, as measured by titration and computed, for the three sets of samples (different colors/symbols).

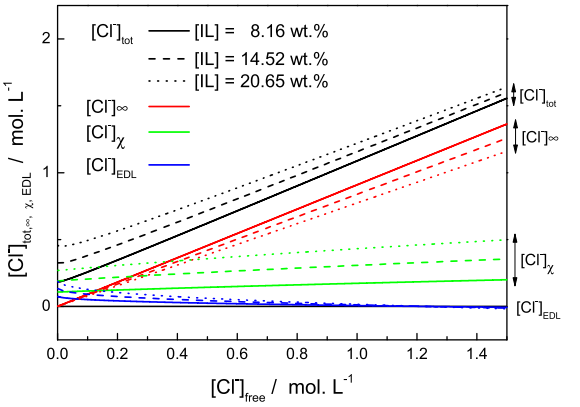


Figure 4. Different chloride populations (colors/groups) for three solutions of various IL mass fraction (full, dashed and dotted lines) as function of the free chloride in the solution (chloride concentration in the reservoir).

5. Discussion

The Figure 4 enables to analyse the different chloride populations. When no salt (NaCl) is added to the solution, the reservoir salt concentration $C_0 = [Cl^-]_{free}$ is 0. In this limit the model is not rigorously valid for two reasons. First, as long as the salt concentration is less than 2×10^{-2} mol·L^{-1}, the strong electrostatic screening regime is not reached. Second we neglected the critical micellar concentration (cmc) and the concentration of free IL in solution. The cmc of $[P_{4,4,4,14}]Cl$ was evaluated from the variation of the surface tension as a function of IL wt% in aqueous mixture, measured by pending drop method (see SI). It was found to be equal to 0.05 ± 0.02 wt%, and drops consequently (<0.02 wt%) when acid or salt is added to the solution. There is therefore a proportion of free chloride not taken into account by the model for these two effects for low salt concentration $C_0 = [Cl^-]_{free}$, as shown in Figure 3.

Adding more chloride into the solution, therefore having more chloride into the reservoir, leads to a slight increase of the bounded ions while the concentration into the EDL decreases. For a reservoir concentration of 1.25 mol·L^{-1}, the EDL concentration $[Cl^-]_{EDL}$ even becomes negative. $[Cl^-]_{EDL}$ is actually an excess term with respect to bulk concentration. This means that above this limit, adsorption onto the surface is strong enough so that it reverses the surface charge, which becomes negative. We note that in absence of well-defined adsorption sites, there is no chemical saturation of the surface. Adsorption is therefore not limited and inversion is possible. In practice, this means that the electrostatic repulsion between micelles breaks down. The system can then become destabilised because the attractions due to Van der Waals forces are no longer counterbalanced by electrostatic forces. Phase separation is then possible. This is precisely what happens experimentally. Despite its simplicity, the model allows us to represent the destabilisation mechanism of the solutions and the appearance of this phase separation. At some chloride addition, the effective charge of the micelles becomes very small, leading to a complete screening of the electrostatic repulsion between the objects and therefore a flocculation of the micelles followed by the phase separation.

This effect is due to the increase in $[Cl^-]_\chi$ population. Although the increase in $[Cl^-]_\chi$ population in the Figure 4 is fairly small, it is enough to neutralise the charge on the micelles. Because of the high charge of the micelles due to their high aggregation number, adsorption is already high in the absence of salt, even if the $K°$ constant is low. The addition of chloride further increases this phenomenon and the charge is eventually reversed.

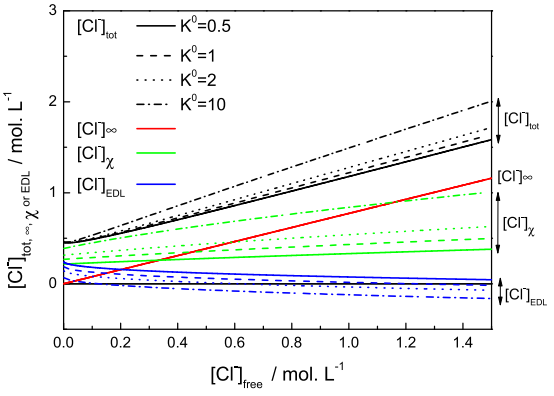


Figure 5. Different chloride populations (colors/groups) for the the 20.65 wt% IL solution as functions of the free chloride in the solution (chloride concentration in the reservoir). The values are given for different binding constants $K°$ (full, dashed, dotted or dash-dotted lines).

We eventually examined the influence of the bounding constant $K°$. The Figure 5 shows the effect of varying the binding constant between 0.5 and 10 on the solution at 8.6 wt% of IL. The figure is understandable if we remember that $[Cl^-]_{free} = C_0$ is the concentration of chloride in an electrolyte solution in equilibrium with the system. Thus the free ions far from the micelles $[Cl^-]_\infty$ are directly proportional to it and do not depend on the adsorption constant which only controls the adsorbed Cl^- and the ions in the Debye EDL. Overall, an increase in $K°$ logically increases chloride ions adsorption. This then reduces the counter-ions concentration and the co-ion depletion in the EDL: the effective charge being lower, fewer counter-ions are needed to screen the micelles and the repulsive force on cations is weaker.

We first remark that increasing the value of $K°$ indeed increases the slope of $[Cl^-]_\chi$ with the chloride reservoir concentration, as previously mentioned. But the main effect is that increasing the binding constant will increase $[Cl^-]_\chi$, reducing the effective potential at the micelle surface, therefore shifting to lower concentration the EDL and the term $[Cl^-]_{EDL}$. This leads to the effective charge of the micelle to be screened at lower NaCl addition, then the phase separation to be induced at lower NaCl or acid content. This behaviour is exactly what is observed upon increase of temperature, where the free chloride concentration decreases with temperature until phase separation. The higher is the temperature, the lower the salt (or acid) addition needed to observe the phase separation. The increase of temperature can therefore be modelled by an increase of $K°$. All this means that the adsorption reaction is favoured at high temperature and is therefore endothermic. It is this phenomenon that allows us to understand why this system shows a *Lower Solution Critical Temperature* (LCST). Eventually, we could assume that if $K°$ increases enough, i.e. at high enough temperature, the charge inversion of the micelles could lead to repulsion again and the system would turn homogeneous again upon NaCl addition. This was however never tested experimentally, since the solubility of NaCl limits the phase diagram.

6. Conclusion

The aqueous biphasic systems formed by the mixture of the ionic liquid $[P_{4,4,4,14}]Cl$ with salt or acid present a strong interest from a fundamental as well as potential application point of view. The phase diagram is complex, with self aggregation and organisation varying with the addition of charges and temperature in different ways. Following a detailed structural investigation of acidic solutions with temperature and an insight into the phase transition mechanisms upon heating, we complete here our rationalisation of electrostatic effects. We propose a simple still relevant charge regulation model to describe the electrical double layer and screening of electrostatic interactions in the solution with addition of ions. The only fitted parameter is the binding constant taking into account the chemically or, in this case, physically bound ions to the micelle surface. The model is not rigorously valid at low salt concentration. Nevertheless, it provides a qualitative but strong insight into the mechanisms at play between the micelles upon addition of charges or increase of temperature, leading to the phase separation.

This approach could therefore be the first step in describing the phase diagram and thermodynamics of these complex systems, for which electrolyte theory can explain their behaviour and in particular their phase diagram. It would thus be possible

to take into account more complex geometries of aggregates (e.g., beyond a spherical organisation) or to look at the limit of weaker electrostatic screening and the influence of the non-micellised part of the ionic liquid. A model combining the DLVO approach and Hamaker's constant could thus enable the full prediction of the phase diagram versus ions concentration and temperature. Investigating the effect of salt versus acid, H^+ versus Na^+ or other ions, will eventually lead to another world of physical chemistry.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Dedication

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This research was funded by the French National Agency for Research (Grant No. ANR-ITALLIX-22-CE29-0023-01). A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

Supplementary data

Supporting information for this article is available on the journal's website under https://doi.org/10.5802/crchim.300 or from the author.

References

[1] J. González-Valdez, K. Mayolo-Deloisa, M. Rito-Palomares, *J. Chem. Technol. Biotechnol.*, 2018, **93**, 1836-1844.

[2] R. Karmakar, K. Sen, *J. Mol. Liquids*, 2019, **273**, 231-247.

[3] M. Gras, N. Papaiconomou, N. Schaeffer, E. Chaînet, F. Tedjar, J. A. Coutinho, I. Billard, *Angew. Chem. Int. Ed.*, 2018, **57**, 1563-1566, HAL Id: hal-01898075.

[4] N. Schaeffer, M. Gras, H. Passos, V. Mogilireddy, C. M. N. Mendonça, E. Pereira, E. Chainet, I. Billard, J. A. Coutinho, N. Papaiconomou, *ACS Sustain. Chem. Eng.*, 2019, **7**, 1769-1777.

[5] A. R. Carreira, H. Passos, N. Schaeffer, L. Svecova, N. Papaiconomou, I. Billard, J. A. Coutinho, *Sep. Purif. Technol.*, 2022, **299**, article no. 121720.

[6] D. Dupont, D. Depuydt, K. Binnemans, *J. Phys. Chem. B*, 2015, **119**, 6747-6757, PMID: 25978001.

[7] E. Sinoimeri, A.-C. Pescheux, I. Guillotte, J. Cognard, L. Svecova, I. Billard, *Sep. Purif. Technol.*, 2023, **308**, article no. 122854.

[8] E. Sinoimeri, V. Maia Fernandes, J. Cognard, J. F. B. Pereira, L. Svecova, I. Guillotte, I. Billard, *Phys. Chem. Chem. Phys.*, 2020, **22**, 23226-23236.

[9] G. Meyer, R. Schweins, T. Youngs, J.-F. Dufrêche, I. Billard, M. Plazanet, *J. Phys. Chem. Lett.*, 2022, **13**, 2731-2736, PMID: 35312328.

[10] N. Schaeffer, H. Passos, M. Gras, V. Mogilireddy, J. P. Leal, G. Pérez-Sánchez, J. R. B. Gomes, I. Billard, N. Papaiconomou, J. A. P. Coutinho, *Phys. Chem. Chem. Phys.*, 2018, **20**, 9838-9846.

[11] V. M. M. Lobo, *Electrolyte Solutions, Data on Thermodynamic and Transport Properties*, vol. I-II, Coimbra Editora, Lisbon, 1984.

[12] M. Jardat, J.-F. Dufrêche, V. Marry, B. Rotenberg, P. Turq, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2023-2033.

[13] J. Lyklema, *Fundamentals of Interface and Colloid Science*, Academic Press, 1995.

Comptes Rendus

Chimie

Objectif de la revue

Les *Comptes Rendus Chimie* sont une revue électronique évaluée par les pairs de niveau international, qui couvre l'ensemble des domaines des sciences chimiques.

Ils publient des numéros thématiques, des articles originaux de recherche, des articles de synthèse, des mises au point, des mises en perspective historiques, des textes à visée pédagogique, ou encore des actes de colloque, en anglais ou en français, sans limite de longueur et dans un format aussi souple que possible (figures, données associées, etc.).

Depuis 2020, les *Comptes Rendus Chimie* sont publiés avec le centre Mersenne pour l'édition scientifique ouverte, selon une politique vertueuse de libre accès diamant, gratuit pour les auteurs (pas de frais de publication) comme pour les lecteurs (accès libre, immédiat et pérenne).

À propos de la revue

Les *Comptes Rendus Chimie* sont exclusivement publiés au format électronique.

Toutes les informations sur la revue, ainsi que le texte intégral de l'ensemble des articles, sont disponibles sur son site internet, à l'adresse https://comptes-rendus.academie-sciences.fr/chimie/.

Informations pour les auteurs

Pour toute question relative à la soumission d'un manuscrit, merci de consulter le site internet de la revue : https://comptes-rendus.academie-sciences.fr/chimie/.

Contact

Académie des sciences
23 quai de Conti
75006 Paris (France)
cr-chimie@academie-sciences.fr

Comptes Rendus
de l'Académie des sciences

Chimie

Volume 27, nº S5, 2024

Special Issue / *Numéro spécial*

French Network on Solvation (GDR 2035 SolvATE) / *Réseau thématique sur la solvatation (GDR 2035 SolvATE)*

Guest Editor / *Rédactrice en chef invitée*

Francesca Ingrosso (Université de Lorraine–CNRS, LPCT UMR 7019, Nancy, France)

Cover illustration / *Illustration de couverture*

Cover illustration: graphical abstract from included article by Axel Gomez et al.

Contents / *Sommaire*

1878-1543 (electronic)

Académie des sciences — Paris

COMPTES RENDUS
DE L'ACADÉMIE DES SCIENCES
Chimie
Volume 27, nº S5, 2024