Journal of Molecular Liquids 347 (2022) 117945

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

ComplexMixtures.jl: Investigating the structure of solutions of complex-shaped molecules from a solvent-shell perspective

Leandro Martínez

Institute of Chemistry, University of Campinas, Campinas, SP, Brazil
Center for Computational Engineering & Science, University of Campinas, Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 15 July 2021
Revised 10 October 2021
Accepted 22 October 2021
Available online 9 November 2021

ABSTRACT

Distribution functions are used to investigate the interactions between the components of condensed-phase systems, while allowing the computation of thermodynamic properties that can be probed experimentally. Radial distribution functions are the most fundamental and easily understood of these distributions, but fail to provide a molecular picture of the interactions when one or all species have complex shapes. On the other hand, regardless of the complexity of the molecular structures involved, minimum-distance distribution functions (MDDFs) can provide a molecular viewpoint on solute–solvent contacts. Here, we describe the ComplexMixtures.jl package, which provides a practical implementation of MDDFs and corresponding Kirkwood-Buff integrals to analyze Molecular Dynamics and Monte-Carlo simulations. Examples are provided for the study of macromolecules in solutions of multiple cosolvents, homogeneous systems, polymer solvation by organic solvents and lipid bilayer interactions with disruptive agents. The distribution functions can be examined using tools to assess the contributions of each atom, group of atoms, and amino acid residues, for example. ComplexMixtures.jl is free software and is compatible with the most common molecular simulation trajectory formats. The software is available as a Julia package with a comprehensive documentation at: http://m3g.iqm.unicamp.br/ComplexMixtures.

1. Introduction

Distribution functions are fundamental for the understanding of the structure, interactions and thermodynamics of solutions. They quantify the local fluctuations of some order parameter, usually the density of some reference atom or site in a solvent molecule relative to a solute, normalized by a reference state which can be computed from, for example, an ideal-gas distribution of the same system.

The radial distribution function [1–3] is the most commonly used measure of the solvent distribution surrounding a solvated species. Given a single reference site in the solute and solvent molecules (usually their centers of mass or one atom of interest), the probability density profile of finding a distance between these sites is computed for each distance and normalized by an equivalent distribution in an ideal state. This provides a quantitative measure of the solvent accumulation or depletion caused by the presence of the solute, directly connected to X-ray diffraction patterns [4]. Appropriate integrals of these distributions (the Kirkwood-Buff integrals) can be used to compute thermodynamic properties of the solutions to be compared with experiments [5–7]. Thus, simulations that correctly predict macroscopic properties of the solutions can be used to understand how observable properties emerge from the molecular distributions of their constituents [8,9].

If the molecules have complex structures, radial distribution functions are inconvenient for depicting solute–solvent interactions, and specialized representations of the density of the species in space are required. For example, distribution functions can be three-dimensional, representing the density of a solvent around a solute at each point in space [10,11]. However, if the solute is flexible these representations of the solvent density are not intuitive, because they are affected by the mobile molecular groups. A variety of different order parameters have been proposed to represent the molecular interactions of solutes and solvents in solutions [1,12–17]. Many-body distributions, which can provide insights into the correlation of multiple particles [18–20], angular distribution functions [21–23], and distribution functions of mixed geometrical properties [21,24,25] are examples. Additional order parameters can help to clarify the composition of, for example, the second solvation peak, or of the broadening of the first peak [24], but interpreting the

E-mail address: lmartine@unicamp.br

https://doi.org/10.1016/j.molliq.2021.117945
0167-7322/

distributions from a molecular standpoint can be challenging. Strategies for reducing the dimensionality of the representations of such distributions must be used [26–30].

Distribution functions that are computed from the shortest distance between a solute and a solvent reference site were proposed decades ago for calculating energetic and thermodynamic properties of solutions [1,12,13,31]. More recently, we investigated the use of generalized Minimum-Distance distribution functions (MDDFs) for the analysis of the interactions between complex solutes and solvent molecules, particularly biomolecules in solvent mixtures [32–35]. MDDFs are the distribution functions of the minimum distance between *any* solute and *any* solvent atom (being other distributions based on minimum-distances particular cases of MDDFs). The peaks and dips of the MDDFs are always associated with the closest interactions between solute and solvent atoms, and the interpretation of these distributions in terms of intermolecular interactions at nearest-neighbour solvent shells is very convenient. Normalization of the MDDFs, on the other hand, is difficult, requiring volume integration of the space associated with each solute atom as well as the probability of finding *any* solvent atom at each element volume [32]. However, with the right normalization strategy, the MDDFs allow the computation of KB integrals and thermodynamic properties, while providing a rich view of the molecular interactions regardless of the structural complexity of the species involved.

ComplexMixtures.jl is a practical implementation of MDDFs and associated analytical tools that enables a comprehensive understanding of solute–solvent interactions in solutions of complex-shaped molecules. We illustrate how the MDDFs can be used to understand the interactions in complex molecular systems by investigating the solvation of proteins in mixed solvents, homogeneous mixtures of small molecules, polymer solvation, and the interactions of a lipid bilayer with disruptive agents.

The package is distributed as free software under the MIT license at http://m3g.iqm.unicamp.br/ComplexMixtures.

2. Approach

The ComplexMixtures.jl package was implemented in the Julia language [36], allowing the development of a dynamical, flexible, and extendable library that is performant and parallelizable.

The efficient computation of minimum-distances between objects of arbitrary shape demands specialized methods. Here, the distribution function is computed up to a maximum distance defined by a cutoff. Distances greater than that of the cutoff are not considered, so cell lists [3] can be used to obtain the MDDFs in $O(n)$ time. However, because the minimum distance between any atoms of the molecules is desired, these short distances must be stored and sorted for each molecule, adding complexity and computational cost. In ComplexMixtures.jl, auxiliary arrays of distances and indices are dynamically allocated for that accumulation and sorting. This efficient cell list implementation is also available as a standalone library (https://github.com/m3g/CellListMap.jl) [37], allowing for the implementation of custom molecular simulation analysis routines. The Chemfiles library is used to obtain compatibility with the most popular simulation trajectory formats [38].

Normalization of the MDDFs must be accomplished through numerical integration. In ComplexMixtures.jl we generate many distributions of the solvent molecules with the proper bulk density, around a non-interacting solute [32]. When only single reference sites are considered in the solute and in the solvent, this reduces to the usual spherical-shell count of radial distribution functions. When considering the shortest distance between a single solvent atom and the atoms of the solute, this corresponds to the normalization of proximal, or solvent-shell, distribution functions [12,32,39].

The computation of the MDDF along a trajectory is almost embarrassingly parallel, because site counting can be performed independently for each frame. Coordinates for each frame are read sequentially, and asynchronous tasks are launched to compute the MDDF for each frame. The main thread remains in charge of reading the trajectory reading and launching the computations.

2.1. Setup of the calculation

Code 1 shows a minimal input file. The user must provide the system's structure as a PDB file, and select the atoms of the solute (the "protein") and of the solvent ("TMAO" molecules, in the example). Structure reading and atom selections are carried out here using the PDBTools.jl package (https://github.com/m3g/PDBTools.jl), which was also developed to provide convenient selection syntax for the current project. These selections are fed into an appropriate data structure (*Selection* type) that contains the indexes of the atoms and molecular identities. The *Trajectory* constructor is used to read the trajectory, and it is then passed to the *mddf* function for calculation. The distribution functions, KB integrals, and atomic contributions will be stored in the *results* variable. More than one molecule can combine to form both the solute set and the solvent set. If the solute and solvent comprise the same set of molecules, an auto-correlation function is computed.

Code 1. Input example for ComplexMixtures.jl, to investigate the protein-TMAO interactions in a solution. The trajectory is provided here in the DCD format.

```
using ComplexMixtures, PDBTools

# Load PDB file of the system
atoms = readPDB("./system.pdb")

# Select the protein and the TMAO molecules
protein = select(atoms,"protein")
tmao = select(atoms,"resname TMAO")

# Setup solute and solvent structures
solute = Selection(protein,nmols=1)
solvent = Selection(tmao,natomspermol=14)

# Setup the Trajectory structure
trajectory = Trajectory("./trajectory.dcd",solute,solvent)

# Run the calculation and get results
results = mddf(trajectory)

# Save results to json file
save(results,"results.json")
```

2.2. Output

The MDDFs, KB integrals, atomic contributions to the MDDF, atom counts, and other auxiliary variables are all stored in the results output variable in Code 1.

Code 2 shows the typical MDDF calculation output summary from ComplexMixtures.jl. The estimated molar volumes of the solute and solvent in the simulated system and in the bulk phase of the solution are given. Depending on its accumulation or depletion in the vicinity of the solute, the molar volume of the solvent in the bulk phase may be greater or smaller than the overall molar volume computed from its density in the simulation. The convergence analysis of the distribution functions allows for the consistency of the calculation to be checked. The RDF displayed is computed from a single reference atom, which can be specified by the user or is assumed to be the most central atom of the solvent molecule. The corresponding KB integral computed from this RDF is also computed and must converge to the same value as that computed from the MDDF over long distances.

The data can be analyzed directly within Julia scripts (comprehensive examples are provided in the user manual) or exported to standard ASCII files for use with other analysis software. The outcomes obtained from multiple trajectories can be concatenated.

Code 2. MDDF calculation output summary from ComplexMixtures.jl. A summary of solute and solvent properties, as well as the converged values of the distributions over long distances is provided.

```
------------------------------------------------------------

MDDF Overview:

Solvent properties:
-------------------

Simulation concentration: 48.746 mol L⁻¹
Molar volume: 20.514 cm³ mol⁻¹

Concentration in bulk: 51.196 mol L⁻¹
Molar volume in bulk: 19.532 cm³ mol⁻¹

Solute properties:
------------------

Simulation Concentration: 0.0022157 mol L⁻¹
Estimated solute partial molar volume: 21598.904 cm³ mol⁻¹

Using with dbulk = 10.0Å:
Molar volume of the solute domain: 79342.916 cm³ mol⁻¹

Auto-correlation: false

Trajectory files and weights:
  ./production1.dcd - w = 0.5
  ./production2.dcd - w = 0.5

Long range MDDF mean (expected 1.0): 1.004 +/- 0.009
Long range RDF mean (expected 1.0): 1.004 +/- 0.008

------------------------------------------------------------
```

3. Examples

In this section, we will look at some examples of how minimum-distance distribution functions can be used to better understand the solution structure and interactions of mixtures of complex molecules. We'll start with a common application: studying a biomolecule in solution. Then, a simple homogeneous binary mixture of water and glycerol is analyzed. The third example concerns the dimethylformamide solvation properties of a polyacrylamide model. Finally, we demonstrate the application of minimum-distance distribution functions to the investigation of the solvation of a POPC lipid bilayer by a mixture of water and ethanol, where ethanol is known to disrupt the membrane.

The entire set of examples, including input files and the analysis scripts can be found at http://github.com/m3g/ComplexMixturesExamples. Previous publications on the understanding of protein solvation by ionic liquids [34,40], and other denaturing or stabilizing osmolytes [32,33,35] provide additional examples.

3.1. Protein solvation by mixtures of cosolvents

The first example consists of a protein (Subtilisin Carlsberg [41]) dissolved in 50% by volume solution of water and glycerol. In Fig. 1A, we begin by demonstrating the most basic distribution function: that of water molecules relative to the protein. The total MDDF has two clearly discernible peaks, at ∼ 1.8 Å and at ∼ 2.6 Å. Specific (hydrogen-bonding) interactions are reflected by the peak at ∼ 1.8 Å. The MDDF can be decomposed into the contributions of the different types of water atoms. Both hydrogen and oxygen atoms contribute to hydrogen bonds at a rate of approximately

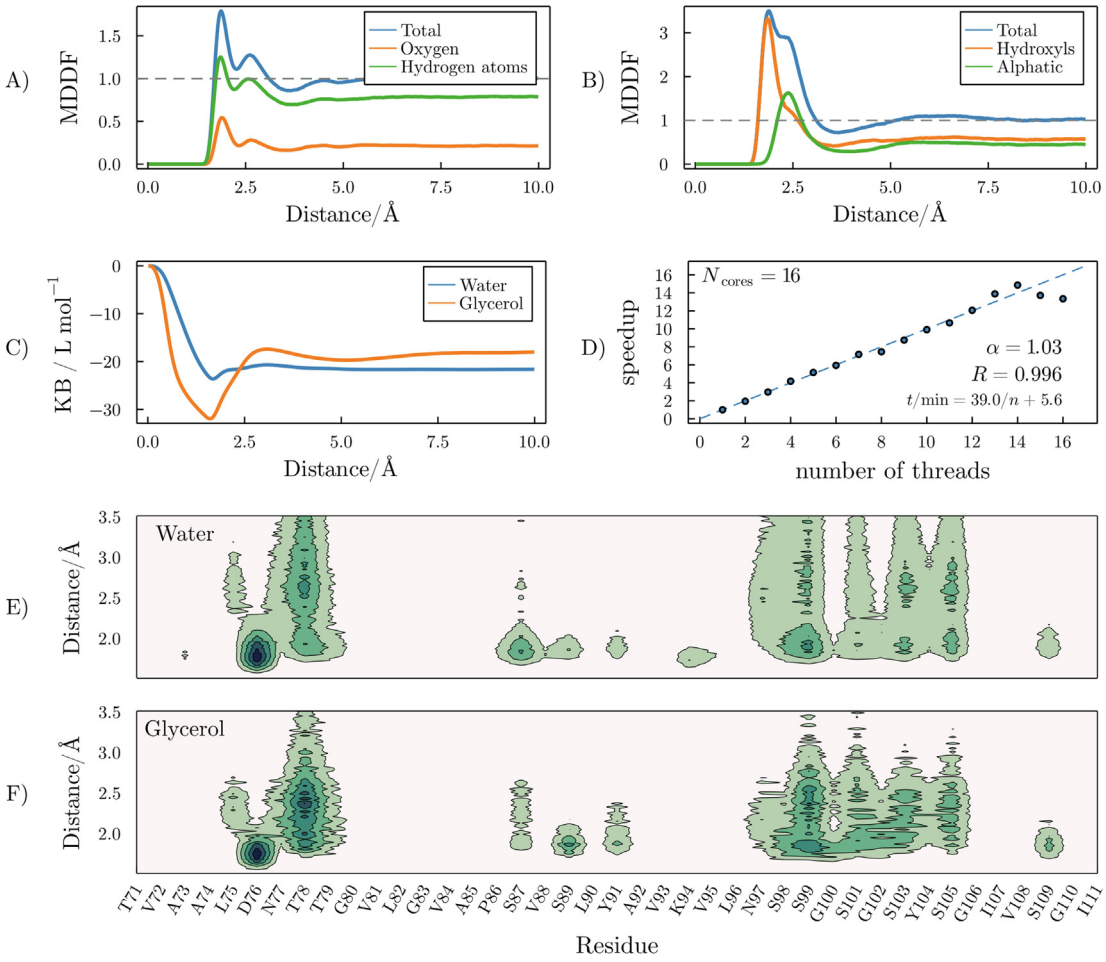

Fig. 1. Minimum-distance distribution functions of A) water and B) glycerol relative to a protein and their decomposition based on the solvents' atom types. C) Kirkwood-Buff integrals associated with the MDDFs shown in A and B. D) Calculation time and scaling for a 60-thousand-atom system for a 2-thousand-frame trajectory. In today's personal computers, typical analysis time is in the order of a few minutes and scales linearly with the number of available processors. E) and F) Solvation of each amino acid residue by water and glycerol, respectively.

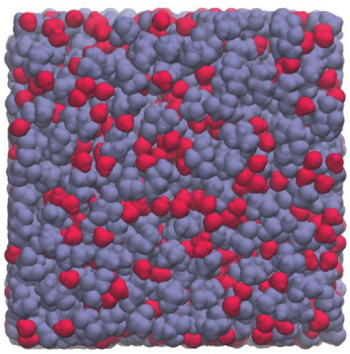

Fig. 2. A homogeneous mixture of 1000 water molecules (red) and 1000 glycerol molecules (purple). The glycerol molecular structure is shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2:1. As a result of the geometry of the water molecule, the contributions of hydrogen and oxygen atoms converge to ∼ 0.79 and ∼ 0.21 at long distances, respectively: the contributions of each atom for the MDDF at long distance depend on the fraction of the molecule's surface area that is associated with each atom (see the Appendix A of [32]).

Fig. 1B depicts the glycerol distribution relative to the protein, which is also decomposed into group contributions. We spit the glycerol MDDF into the hydroxyl and aliphatic contributions. As expected, a hydrogen-bonding peak is observed, which is completely determined by the interactions of the hydroxyl groups. Interestingly, both the hydroxyl and aliphatic groups of glycerol contribute to the second peak, which is associated with non-specific interactions.

Fig. 1C shows the Kirkwood-Buff integrals of water and glycerol calculated from the MDDFs of Fig. 1A and 1B. The distance dependence of the KB integrals computed from the MDDFs is interesting because it reflects the excluded volume associated with the solute and solvent molecules at short distances, and the possible compensation of this excluded volume by favorable solvation interactions.

The local density augmentation of the solutes, associated with the MDDF peaks at short distances, compensate (here partially) for the excluded volumes. However, in this case, both KB integrals are negative. Since the KB integral for glycerol is greater (less negative) than that of water, the protein is preferentially solvated by glycerol. This is a known problem with the CHARMM36 forcefield for carbohydrates [42], as glycerol is preferentially excluded and protects the protein from unfolding in experiments. Other protective osmolytes, such as TMAO, can exhibit density augmentation at the protein surface even when the protein is preferentially hydrated, this being consistent with experimental findings [32].

Fig. 1E and 1F illustrate the fact that the MDDFs can be decomposed into the contributions of the solute atoms or groups of solute atoms. In this case, the MDDFs of glycerol and water are decomposed into the contributions of the atoms of each protein amino acid residue. The densities of solvent-protein minimum distances are plotted as a function of the distance to the solute group in the form of contour plots. Residues that do not contribute to the MDDF are shielded from the solvent. Some residues in this example stand out in their ability to interact with water but not with glycerol (for example, K94 and V95). Three-dimensional representations of the density can also be obtained, as described in the user manual.

Finally, Fig. 1D highlights the performance and parallel scalability of the package. With the exception of the trajectory-reading step, the problem is embarrassingly parallel, as the minimum-distance count can be performed independently for each frame of the trajectory. Scaling is linear with the number of computer cores available while not being constrained by other running processes or trajectory reading from the disk. On a personal computer, a typical trajectory analysis of 2-thousand frames of a system with 60 thousand atoms will currently take a few minutes. If necessary, for very long trajectories, this can be sped up by lowering the precision of the numerical integration of the volumes on each frame, because averaging over many frames improves the sampling by itself.

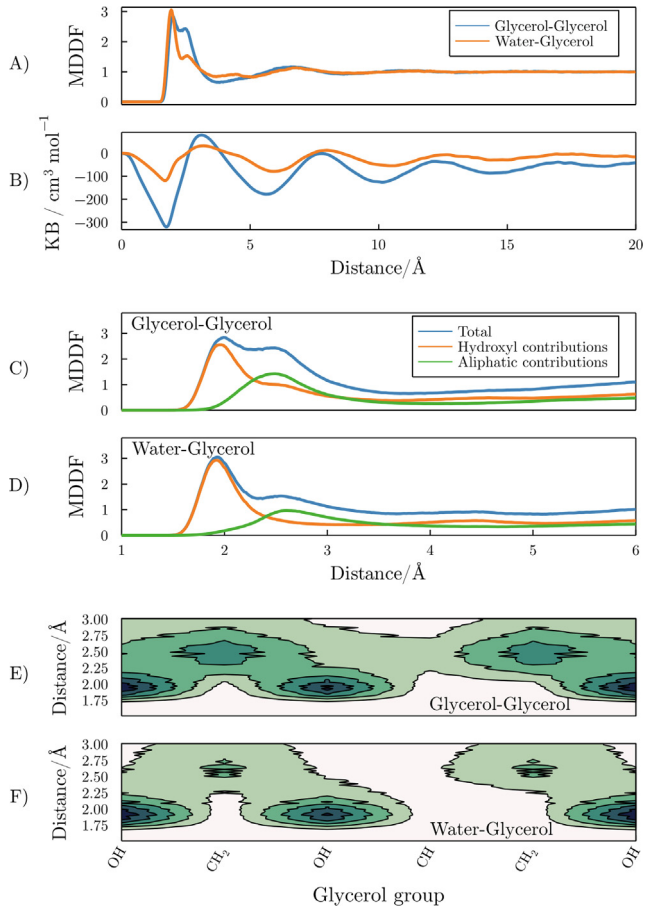

Fig. 3. Structure of a 1:1 (mol/mol) solution of water and glycerol. A) Glycerol and water minimum-distance distribution functions relative to glycerol. Hydrogen-bonding is observed in both distributions, but non-specific interactions are more prominent in the glycerol auto-correlation function. B) Kirkwood-Buff integrals of: Glycerol molecules are possibly slightly preferentially hydrated, as indicated by the greater KB integral for water at large distances. C) Group contributions to the glycerol auto-correlation function. The hydrogen-bonding peak is associated to hydroxyl groups, and the aliphatic groups contribute significantly only to the second solvation peak. D) Glycerol group contributions, as a solute, to the water-glycerol distribution function. The hydroxyls are responsible for the specific interactions with water at hydrogen-bonding distances. E) and F) Glycerol group contributions to the glycerol auto-correlation and glycerol-water correlation functions. It becomes apparent that the second peak of the distributions, associated with non-specific interactions, corresponds to the solvation of the CH_2 groups. The CH group is largely protected from the solvents.

3.2. Homogeneous mixtures

This example illustrates how to use ComplexMixtures.jl to investigate the solution structure of a crowded (1:1 M) solution of glycerol and water, at room temperature and pressure. The system simulated is illustrated in Fig. 2. It consists of 1000 water molecules and 1000 glycerol molecules. We compute the distribu-

tion functions associated to water-glycerol interactions, and also the glycerol-glycerol autocorrelation functions. The purpose of this example is to demonstrate how to obtain a detailed molecular picture of the solvation structures in a homogeneous solution.

As shown in Fig. 3A, both water and glycerol form hydrogen bonds with glycerol molecules, as shown by the peaks at ∼ 1.8 Å. The auto-correlation function of glycerol shows a distinct second peak corresponding to non-specific interactions, which are most likely associated with its aliphatic groups. The KB integrals associated to these distributions are shown In Fig. 2B, with the KB integral for water being slightly greater. This means that glycerol molecules are preferentially hydrated from a macroscopic standpoint, though the difference is small and could be due to model limitations.

The distribution functions of Fig. 3A can be decomposed into group contributions of the solute or solvent molecules. Here, we partition the distributions into the contributions of the chemical groups of the solute (glycerol in both cases). The contributions of hydroxyls and aliphatic groups are depicted in Fig. 2C and 2D, respectively. The interactions with the glycerol hydroxyl groups are responsible for the first peaks of both distributions. The second peaks contain significant contributions from all groups, particularly the aliphatic groups of glycerol. These non-specific interactions are stronger for the glycerol auto-correlation, resulting in the larger second peak of the MDDF.

The above distributions can be decomposed further into the contributions of each glycerol chemical group, as shown in Fig. 2E and 2F. Two additional insights into the solution structures can be obtained from these figures: first, the second solvation peaks, at ∼ 2.5 Å, are centered on the CH_2 groups of glycerol. According to the maps, the solvent molecules at these positions may be interacting with the molecules at the first solvation shell, where water or glycerol form hydrogen bonds with the solute (an additional order parameter would have to be computed to confirm this hypothesis, as suggested in [24]). The second distinguishing feature of the maps is that the CH group of glycerol is particularly protected from both solvents.

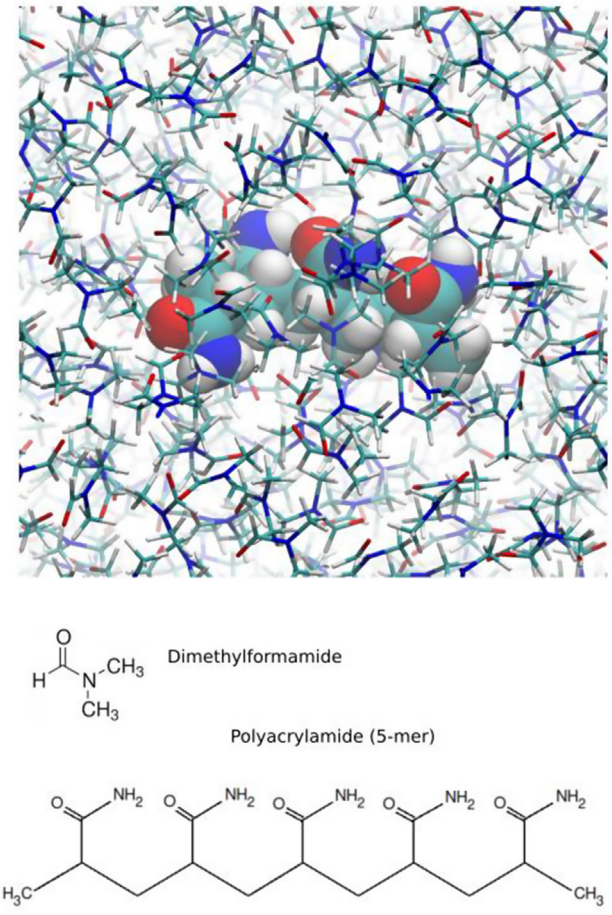

Fig. 4. A 5-mer of polyacrylamide capped by methyl groups (spheres), in a solution of dimethylformamide (sticks). The system is composed of 1 PolyACR chain and 5000 DMF molecules. The molecular structures of DMF and of the polyacrylamide model are depicted.

3.3. Solvation of a polymer by an organic solvent

In this example we illustrate how the solvation structure of a polymer can be studied with ComplexMixtures.jl. The system is a 5-mer segment of polyacrylamide capped with methyl groups (PolyACR), solvated by 2000 dimethylformamide (DMF) molecules. The system is interesting because of the various functional groups and polarities involved in the DMF-PolyACR interactions, as well as the structural complexity of this solute. A snapshot of the system is shown in Fig. 4.

Fig. 5A shows the minimum-distance distribution function of DMF molecules relative to polyacrylamide. A peak at ∼ 2.3 Å indicates that the solvent and the polymer have favorable non-specific interactions. The peak is followed by a distinct dip, but there is no

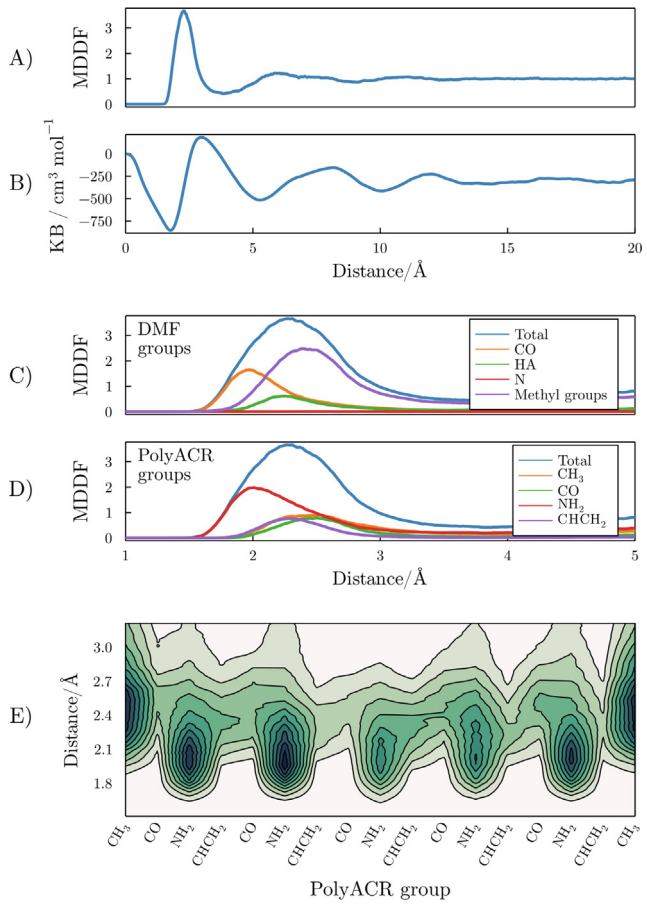

Fig. 5. Polyacrylamide solvation by dimethylformamide. A) Minimum-distance distribution function and B) Kirkwood-Buff integral of DMF relative to PolyACR. C) Contributions of the DMF groups to the DMF-PolyACR distribution function, showing that the broad peak of the distribution is a combination of hydrogen-bonding interactions of PolyaACR with the carbonyl and of non-specific interactions with the other chemical groups of DMF. D) Decomposition of the MDDF into the chemical groups of PolyACR, demonstrating that hydrogen bonds occur through the polymer's amine groups. E) PolyACR group contributions to the MDDF: the terminal methyl groups and the amines contribute the most. The central mer is shielded from the solvent by partial polymer folding.

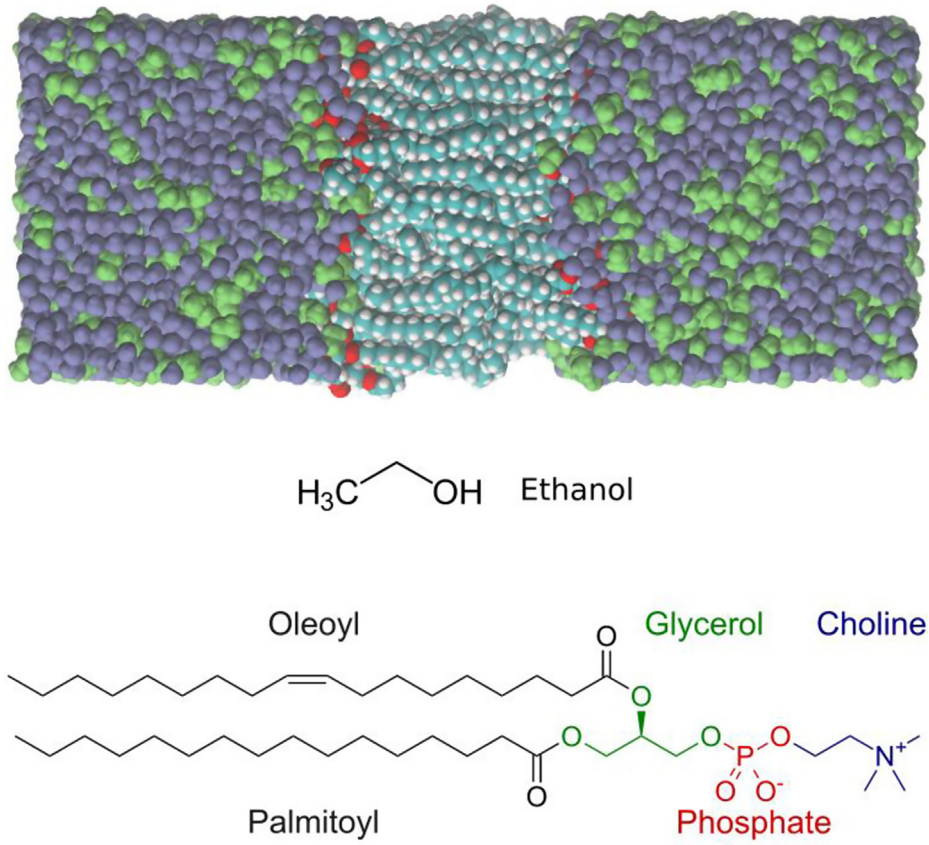

Fig. 6. POPC lipid bilayer solvated by 5000 water molecules (purple) and 1000 ethanol molecules (green). The structures of ethanol and of the POPC lipid are illustrated. The POPC functional groups are: the charged Choline and Phosphate groups, the polar Glycerol group, and the lipid radicals Oleoyl (which contains the insaturation) and Palmitoyl. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

discernible structure at greater distances. Thus, the DMF molecules are structured around the polymer, but essentially only in its first solvation shell.

From the KB integral it is possible to obtain the apparent molar volume of the solute. The DMF-PolyACR KB integral is negative in this case, as shown in Fig. 5B, indicating that the accumulation of DMF in the polymer's first solvation shell is insufficient to compensate for the excluded volume of the solute. In this solution, the PolyACR model has an apparent molar volume of about 300 $cm^3\ mol^{-1}$.

The MDDF can be subdivided into the contributions of the DMF chemical groups, and of the PolyACR chemical groups. The contributions of DMF to the distribution function are shown in Fig. 5C. Specific interactions at distances less than 2 Å indicate that hydrogen-bonds form between the carbonyl group of DMF and PolyACR. Less specific interactions, which peak at $\sim$ 2.3 Å are dominated by the methyl groups of DMF, with minor contributions from all other chemical groups of the solvent molecules.

The decomposition of the MDDF in the contributions of the polymer's chemical groups, shown in Fig. 5D, enables the establishment of some intriguing correlations with the contributions of the DMF groups. The polymer amine groups are associated with the specific hydrogen-bonding interactions. These, too, contribute to non-specific interactions at greater distances, but only partially as part of a pool of contributions from all polymer groups, polar or aliphatic.

We can decompose the MDDF into the contributions of each polymer chain segment. Fig. 5E's contour plot depicts the contributions of each chemical group of the polymer, but now divided into each mer of the polymer chain. The terminal methyl groups have a strong interaction with DMF. Significant local density augmentations, corresponding to hydrogen-bonding interactions, are visible, particularly on top of the amine groups. Surprisingly, the DMF molecules are excluded from the polymer's aliphatic and carbonyl groups.

Finally, the central acrylamide mer is more weakly solvated by DMF than the mers closer to the polymer chain's extremes. This is most likely due to the polymer's partial folding. Thus, the MDDFs provide insights not only into the interactions between solute and solvent molecules related to the chemical nature of the groups involved, but also into the system's configurational equilibrium, as illustrated here by solvent exposure of the polymer chain conformations.

3.4. Lipid bilayer interactions with disruptive agents

The solutes under investigation can be a collection of molecules in a heterogeneous mixture. Here, ComplexMixtures.jl is used to investigate the interactions of a POPC bilayer with a mixture of 20% (mol/mol) of ethanol in water. Ethanol destabilizes the membrane at this concentration [43,44]. The molecular system simulated is illustrated in Fig. 6. The lipid molecules are regarded as a single solute in this context. Importantly, the computation of the MDDF does not need to be adapted to account for anisotropy of the system: The normalization of the distribution is obtained by generating random configurations of the solvent molecules in the simulation box, with bulk density, and by computing the equivalent distance histogram that is computed from the actual simulation.

Fig. 7A displays the minimum-distance distribution functions of water and ethanol relative to the lipid bilayer. Both solvents

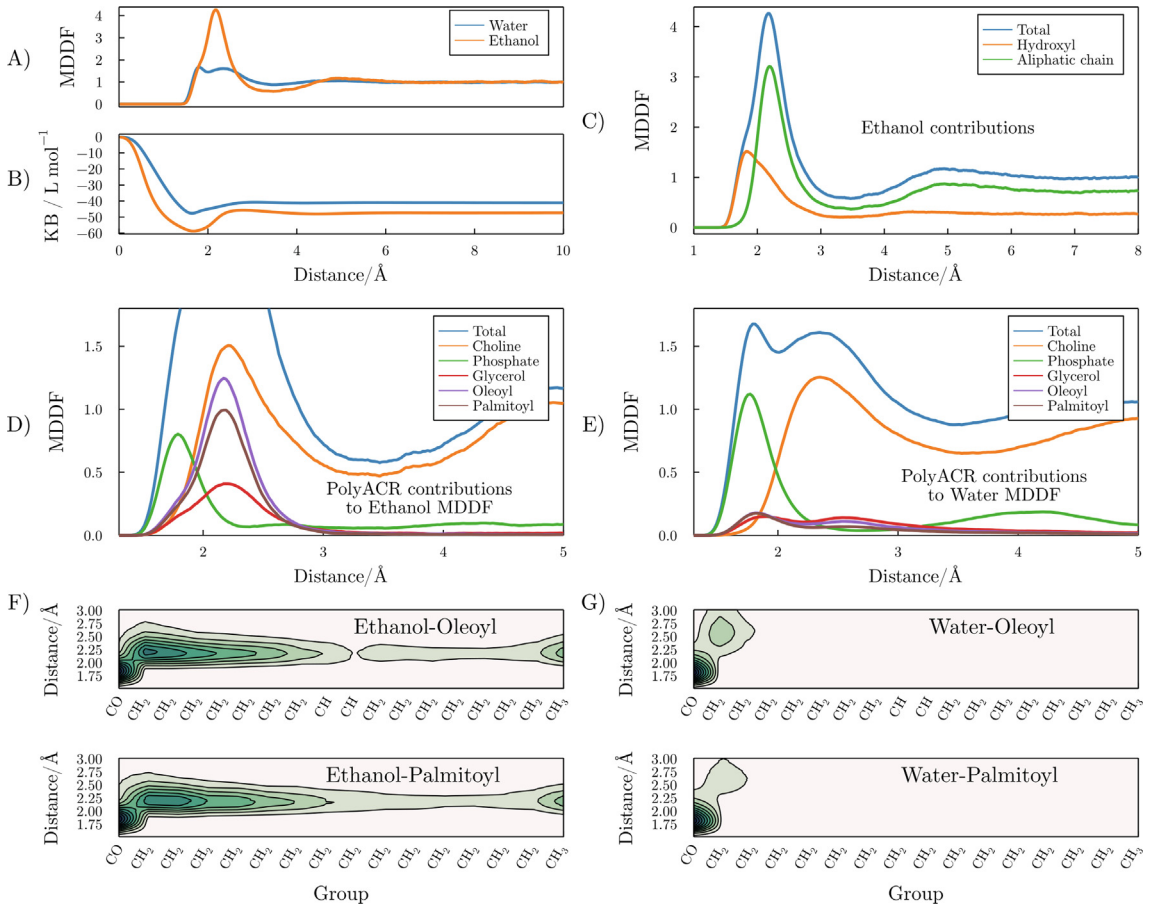

Fig. 7. POPC lipid bilayer solvated by a mixture of water and ethanol. A) Minimum-distance distribution functions of water and ethanol relative to the POPC lipids. B) Kirkwood-Buff integrals, suggesting that the bilayer is preferentially hydrated, despite the ethanol local density augmentation. C) Decomposition of the MDDF into the contributions of the chemical groups of ethanol, revealing that the peak associated with non-specific interactions is mostly determined by the aliphatic chain of the solvent. D) Decomposition of the ethanol-membrane MDDF into the chemical groups of the POPC lipids: ethanol penetrates into the membrane and interacts with the Glycerol, Oleoyl and Palymitoyl groups. E) POPC contributions to the MDDF of water relative to the membrane reveal important contributions only of the charged groups of the lipids. F) Interactions of ethanol with the lipid tails of POPC, showing that the insaturation leads to a discontinuity in the solvation structure, which is not present in the palmitoyl group. G) Water does not penetrate into the membrane, avoiding interactions with the lipids' tails.

exhibit a local density augmentation in the range of ~ 1.7 Å to ~ 3 Å. The water MDDF shows a clear peak at hydrogen-bonding distances, whereas the profile of ethanol distribution is dominated by non-specific interactions. At the first solvation shell, ethanol has a greater local density augmentation than that of water. However, the KB integrals associated with these distribution functions, shown in Fig. 7B, indicate that the membrane is preferentially hydrated (the KB integral of water is greater than that of ethanol). In comparison with the MDDFs, the distance dependence of the KB integrals is interesting to interpret: the excluded volume associated with the membrane-ethanol interaction is greater than that associated with membrane-water interactions (because of the larger ethanol volume). This leads to the smaller KB integral for ethanol despite its greater local density augmentation at the first membrane solvation shell.

Fig. 7C shows the ethanol group contributions to the MDDF. The hydroxyl group of ethanol, as expected, forms specific interactions with the membrane. The aliphatic ethanol groups dominate the peak associated with non-specific interactions. Fig. 7D and E show the POPC chemical groups to the distribution functions of ethanol (Fig. 7D) and water (Fig. 7E). Clearly, ethanol molecules can interact with the glycerol and lipid groups of POPC molecules, while water molecules cannot.

The contribution of each chemical group of the lipid tails to the MDDFs can be computed to further investigate the penetration of the solvents into the membrane. These are shown as contour plots in Fig. 7F and 7G. Ethanol and water interact with the carbonyl group at the head of the lipid groups, but ethanol is found all the way down the lipid tails. Water, on the contrary, is excluded from the lipid tails. Interestingly, ethanol does not solvate Oleoyl and Palmitoyl equally: Oleoyl's unsaturation is less affine to ethanol than the other groups. Another intriguing finding is that the terminal methyl groups of the lipids have significant interactions with ethanol. These interactions occur primarily as a result of the bending of the lipid tails towards the membrane polar regions, reaching portions of the membrane that are rich in ethanol molecules (not shown). As a result, the MDDFs provide a detailed chemical picture of membrane solvation, as well as insights into the mechanisms of membrane disruption by cosolvents.

4. Conclusion

A package was created to study the solvation of molecules with complex shapes via molecular dynamics or Monte-Carlo simulations. It is primarily targeted, but not exclusively, at understanding the solvation structures of biomolecules and other macromolecules. Minimum-distance distribution functions provide an intuitive representation of molecular solute–solvent interactions. The Kirkwood-Buff solution theory can be used to compute thermodynamic properties from the MDDFs, after the definition of a

proper reference state. Tools are provided to investigate the contributions of chemical groups, atoms, or any other subsets of the solute and solvent molecules to the overall solvation structures, providing a detailed picture of their interactions.

Funding

The author acknowledges the financial support of Fapesp (2010/16947–9, 2018/24293–0, 2013/08293–7, 2018/14274–9) and CNPq (302332/2016–2).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A:. Methods

MDDF distributions are more expensive to compute than radial distribution functions because the normalization depends on the generation of random configurations of the solvents. Cell lists can be used to obtain a practical implementation of this computation [4]. For this package, a fast implementation of cell lists was implemented in Julia, using modern algorithms [45]. This implementation was later split into an independent package, CellListMap.jl [37], which can be used to implement diverse analyses and simulations based on cutoff-restricted particle interactions. On shared memory architectures, the performance CellListMap.jl is comparable to that of state-of-the-art simulations packages such as NAMD [46] (see https://github.com/m3g/2021_FortranCon/tree/main/cellistmap_vs_namd). All graphics were prepared with the Plots.jl package. The noise of the histograms was smoothed by computing a moving average of 10 points, using EasyFit.jl. Molecular images were produced with VMD [47].

All of the simulations in the examples were run with NAMD [46], with CHARMM36 parameters [48–50] for proteins, lipids, and carbohydrates, and the TIP3P model for water [51]. The molecular structures of DMF and PolyACR were built with the JSME tool [52], the POPC lipid bilayer was generated with the VMD membrane plugin, and the entire systems were finally constructed with Packmol [53,54]. The densities of aqueous solutions of glycerol were obtained from [55], the density of DMF was obtained from ref. [56]. The protein simulated in the first example is Subtilisin Carlsberg (pdb id. 1SBC) [41]. All systems were designed to be large enough such that solute minimum images were at least 30 Å apart. In all examples we used the parameter *dbulk* = 20 Å, implying that the bulk densities of the solvents were estimated from the number of solvent molecules farther than that distance from the solute.

Details of the simulations of the first example are described elsewhere [57,58]. All other simulations were performed specifically for the present work with the following protocol: The initial systems were minimized by 5000 steps of conjugate-gradient minimization, followed by 10 ns of simulation with temperature and pressure controls at 298.15 K and 1 Bar for equilibration. Next, 200 ns of simulation were performed also at constant temperature and pressure, and the analyses were performed on these last trajectories. The temperature was controlled by a Langevin thermostat with a coupling constant of 10/ps, and the pressure with a Langevin barostat with a piston period of 200 fs and a damping scale of 100 fs. Short-ranged interactions were cutoff at 12 Å, and long-range electrostatics was computed using the Particle-Mesh Ewald Sum method [59].

References

[1] P.K. Mehrotra, D.L. Beveridge, Structural analysis of molecular solutions based on quasi-component distribution functions. Application to [H2CO]aq at 25. degree.C, J. Am. Chem. Soc. 102 (13) (1980) 4287–4294.
[2] B.M. Ladanyi, M.S. Skaf, Computer Simulation of Hydrogen-Bonding Liquids, Annu. Rev. Phys. Chem. 44 (1) (1993) 335–368.
[3] D. Frenkel, B. Smit, Understanding Molecular Dynamics Simulations, Academic Press, 2002.
[4] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Oxford University Press, 2017.
[5] J.G. Kirkwood, Theory of Solutions, CRC Press, 2001.
[6] A. Ben-Naim, Molecular Theory of Solutions, Oxford University Press on Demand, 2006.
[7] S. Shimizu, C.L. Boon, The Kirkwood-Buff theory and the effect of cosolvents on biochemical reactions, J. Chem. Phys. 121 (18) (2004) 9147–9155.
[8] E. Ruckenstein, I.L. Shulgin, Thermodynamics of Solutions: From Gases to Pharmaceutics to Proteins, Springer Science & Business Media, 2009.
[9] J.J. Booth, M. Omar, S. Abbott, S. Shimizu, Hydrotrope accumulation around the drug: the driving force for solubilization and minimum hydrotrope concentration for nicotinamide and urea, Phys. Chem. Chem. Phys. 17 (2015) 8028–8037.
[10] T. Yoshidome, M. Ikeguchi, M. Ohta, Comprehensive 3D-RISM analysis of the hydration of small molecule binding sites in ligand-free protein structures, J. Comput. Chem. 41 (28) (2020) 2406–2419.
[11] D. Roy, A. Kovalenko, Biomolecular Simulations with the Three-Dimensional Reference Interaction Site Model with the Kovalenko-Hirata Closure Molecular Solvation Theory, Int. J. Mol. Sci. 22 (10) (2021) 5061, https://doi.org/10.3390/ijms22105061.
[12] S.-C. Ou, B.M. Pettitt, Solute-Solvent Energetics Based on Proximal Distribution Functions, J. Phys. Chem. B 120 (33) (2016) 8230–8237.
[13] B. Lin, B. Montgomery Pettitt, Note: On the universality of proximal radial distribution functions of proteins, J. Chem. Phys. 134 (2011) 106101.
[14] B.M. Baynes, B.L. Trout, Proteins in Mixed Solvents: A Molecular-Level Perspective, J. Phys. Chem. B 107 (2003) 14058–14067.
[15] V. Zeindlhofer, D. Khlan, K. Bica, C. Schröder, Computational analysis of the solvation of coffee ingredients in aqueous ionic liquid mixtures, RSC Adv. 7 (6) (2017) 3495–3504.
[16] V. Zeindlhofer, M. Berger, O. Steinhauser, C. Schröder, A shell-resolved analysis of preferential solvation of coffee ingredients in aqueous mixtures of the ionic liquid 1-ethyl-3-methylimidazolium acetate, J. Chem. Phys. 148 (2018) 193819.
[17] A.C. Furlan, F.W. FÁvero, J. Rodriguez, D. Laria, M.S. Skaf, Solvation in Supercritical Fluids, Challenges and Advances in Computational Chemistry and Physics. (2008) 433–453
[18] D. Ben-Amotz, A. Stamatopoulou, B.J. Yoon, Three-body distribution functions in hard sphere fluids. Comparison of excluded-volume-anisotropy model predictions and Monte Carlo simulation, The Journal of Chemical Physics. 107 (1997) 6831–6838.
[19] B.E. Clements, C.E. Campbell, P.J. Samsel, F.J. Pinski, Molecular-dynamics simulation of the static pair-pair correlation function for classical fluids, Phys. Rev. A. 44 (2) (1991) 1139–1147.
[20] Anirban Bhattacharjee, Thomas S. Hofer, Bernd M. Rode, Local density corrected three-body distribution functions for probing local structure reorganization in liquids, Phys. Chem. Chem. Phys. 10 (44) (2008) 6653, https://doi.org/10.1039/b807305h.
[21] A.K.H. Weiss, T.S. Hofer, Exploiting the capabilities of quantum chemical simulations to characterise the hydration of molecular compounds, RSC Adv. 3 (2013) 1606–1635.
[22] Paul J. Steinhardt, David R. Nelson, Marco Ronchetti, Bond-orientational order in liquids and glasses, Physical Review B. 28 (2) (1983) 784–805.
[23] A.D. Santis, A. De Santis, D. Rocca, Angular distribution functions and specific local structures in liquid water, J. Chem. Phys. 107 (1997) 10096–10101.
[24] S.V. Sukhomlinov, M.H. Müser, A mixed radial, angular, three-body distribution function as a tool for local structure characterization: Application to single-component structures, J. Chem. Phys. 152 (2020) 194502.
[25] Clovis Quintale, Léo Degrève, Angular—radial distribution functions in liquid water, J. Mol. Struct. (Thoechem) 282 (1-2) (1993) 159–165.
[26] A.D.J. Haymet, Stuart A. Rice, William G. Madden, An accurate integral equation for the pair and triplet distribution functions of a simple liquid, J. Chem. Phys. 74 (5) (1981) 3033–3041.
[27] R.Z. Troitzsch, G.J. Martyna, S.E. McLain, A.K. Soper, J. Crain, Structure of Aqueous Proline via Parallel Tempering Molecular Dynamics and Neutron Diffraction, J. Phys. Chem. B 111 (28) (2007) 8210–8222.
[28] O.M.D. Lutz, T.S. Hofer, B.R. Randolf, B.M. Rode, Hydration of trivalent lanthanum revisited - An QMCF-MD approach, Chem. Phys. Lett. 536 (2012) 50–54.
[29] J.A. Krumhansl, S. Wang, Triplet Correlation in Liquid Argon by Monte Carlo Method: Low Densities, The Journal of Chemical Physics. 56 (1972) 2034–2041.
[30] W.J. McNeil, W.G. Madden, A.D.J. Haymet, S.A. Rice, Triplet correlation functions in the Lennard-Jones fluid: Tests against molecular dynamics simulations, J. Chem. Phys. 78 (1983) 388–398.
[31] K. Coutinho, R. Rivelino, H.C. Georg, S. Canuto, The Sequential qm/mm Method and its Applications to Solvent Effects in Electronic and Structural Properties of

Solutes, Challenges and Advances in Computational Chemistry and Physics. (2008) 159–189.
[32] L. Martínez, S. Shimizu, Molecular Interpretation of Preferential Interactions in Protein Solvation: A Solvent-Shell Perspective by Means of Minimum-Distance Distribution Functions, J. Chem. Theory Comput. 13 (2017) 6358–6372.
[33] I.P. de Oliveira, I.P. de Oliveira, L. Martínez, The shift in urea orientation at protein surfaces at low pH is compatible with a direct mechanism of protein denaturation, Phys. Chem. Chem. Phys. 22 (2020) 354–367.
[34] V. Piccoli, L. Martínez, Correlated counterion effects on the solvation of proteins by ionic liquids, J. Mol. Liq. 320 (2020) 114347.
[35] Ivan P. Oliveira, Leandro Martínez, Molecular basis for competitive solvation of the Burkholderia cepacia lipase by sorbitol and urea, Phys. Chem. Chem. Phys. 18 (31) (2016) 21797–21808.
[36] J. Bezanson, A. Edelman, S. Karpinski, V.B. Shah, Julia: A Fresh Approach to Numerical Computing, SIAM Rev. 59 (2017) 65–98.
[37] L. Martínez, CellListMap.jl: Flexible implementation of cell lists to map the calculations of short-ranged particle-pair dependent functions, such as forces, energies, neighbor lists, etc, Zenodo, 2021. https://doi.org/10.5281/ZENODO.4927541.
[38] G. Fraux, J. Fine, ezavod, G.P. Barletta, M. Dimura, F.X. Coudert, pelsa, M. Levesque, S. Guionniere, jmintser, chemfiles/chemfiles: Version 0.10.0, Zenodo, 2021. https://doi.org/10.5281/ZENODO.4540492.
[39] W. Song, R. Biswas, M. Maroncelli, Intermolecular Interactions and Local Density Augmentation in Supercritical Solvation: A Survey of Simulation and Experimental Results, The Journal of Physical Chemistry A. 104 (2000) 6924–6939.
[40] T.A. Shmool, L.K. Martin, L. Bui-Le, I. Moya-Ramirez, P. Kotidis, R.P. Matthews, G.A. Venter, C. Kontoravdi, K.M. Polizzi, J.P. Hallett, An experimental approach probing the conformational transitions and energy landscape of antibodies: a glimmer of hope for reviving lost therapeutic candidates using ionic liquid, Chem. Sci. 12 (2021) 9528–9545.
[41] David J. Neidhart, Gregory A. Petsko, The refined crystal structure of subtilisin Carlsberg at 2.5 Å resolution, "Protein Engineering, Design and Selection". 2 (4) (1988) 271–276.
[42] Theresa Cloutier, Chaitanya Sudrik, Hasige A. Sathish, Bernhardt L. Trout, Kirkwood-Buff-Derived Alcohol Parameters for Aqueous Carbohydrates and Their Application to Preferential Interaction Coefficient Calculations of Proteins, J. Phys. Chem. B. 122 (40) (2018) 9350–9360.
[43] M. Patra, E. Salonen, E. Terama, I. Vattulainen, R. Faller, B.W. Lee, J. Holopainen, M. Karttunen, Under the influence of alcohol: the effect of ethanol and methanol on lipid bilayers, Biophys. J. 90 (2006) 1121–1135.
[44] M. Ghorbani, E. Wang, A. Krämer, J.B. Klauda, Molecular dynamics simulations of ethanol permeation through single and double-lipid bilayers, J. Chem. Phys. 153 (2020) 125101.
[45] James S. Willis, Matthieu Schaller, Pedro Gonnet, Richard G. Bower, Peter W. Draper, An Efficient SIMD Implementation of Pseudo-Verlet Lists for Neighbour Interactions in Particle-Based Codes, in: S. Bassini, M. Danelutto, P. Dazzi (Eds.), Parallel Computing Is Everywhere, IOS Press, 2018, pp. 507–516.
[46] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, NAMD2: Greater Scalability for Parallel Molecular Dynamics, J. Comput. Phys. 151 (1999) 283–312.
[47] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[48] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E.M. Lopes, Jeetain Mittal, Michael Feig, Alexander D. MacKerell, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles, J. Chem. Theory Comput. 8 (9) (2012) 3257–3273.
[49] Alexander D. MacKerell, Michael Feig, Charles L. Brooks, Improved treatment of the protein backbone in empirical force fields, J. Am. Chem. Soc. 126 (3) (2004) 698–699.
[50] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell Jr, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, J. Phys. Chem. B. 114 (2010) 7830–7843.
[51] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, Michael L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (2) (1983) 926–935.
[52] B. Bienfait, P. Ertl, JSME: a free molecule editor in JavaScript, J. Cheminform. 5 (2013) 24.
[53] L. Martínez, R. Andrade, E.G. Birgin, J.M. Martínez, PACKMOL: a package for building initial configurations for molecular dynamics simulations, J. Comput. Chem. 30 (2009) 2157–2164.
[54] José Mario Martínez, Leandro Martínez, Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking, J. Comput. Chem. 24 (7) (2003) 819–825.
[55] Koichi Takamura, Herbert Fischer, Norman R. Morrow, Physical properties of aqueous glycerol solutions, J. Petrol. Sci. Eng. 98-99 (2012) 50–60.
[56] J. Manuel Bernal-García, Adriana Guzmán-López, Alberto Cabrales-Torres, Alejandro Estrada-Baltazar, Gustavo A. Iglesias-Silva, Densities and viscosities of (N, N-dimethylformamide + water) at atmospheric pressure from (283.15 to 353.15) K, J. Chem. Eng. Data. 53 (4) (2008) 1024–1027.
[57] Alvaro J. Lopez, Emília P. Barros, Leandro Martínez, On the Interpretation of subtilisin Carlsberg Time-Resolved Fluorescence Anisotropy Decays: Modeling with Classical Simulations, J. Chem. Inf. Model. 60 (2) (2020) 747–755.
[58] E.P. de Barros, Estudo computacional da solvatação e anisotropia de fluorescência resolvida no tempo de subtilisina Carlsberg em misturas de água e glicerol, University of Campinas, Master, 2015, https://doi.org/10.47749/t/unicamp.2015.955269.
[59] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.