Artificial Water Channels Form Precursors to Sponge-Like Aggregates in Water–Ethanol Mixtures

Arthur Hardiagon, Marc Baaden, Fabio Sterpone

Artificial Water Channels form Precursors to Sponge-Like Aggregates in Water-Ethanol Mixtures

Arthur Hardiagon, Marc Baaden, and Fabio Sterpone*

Université Paris Cité, CNRS, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, F-75005, Paris, France.

E-mail: sterpone@ibpc.fr

Abstract

Self-assembled Artificial Water Channels (AWCs) are reshaping current water desalination technologies. Recently, the improvements achieved by incorporating hydrophilic compounds into polyamide membranes (PA) at the interface have been experimentally confirmed. However, the determination of the nanoscale structure of AWCs remains unclear. An important step in the preparation of PA membranes is solubilization by colloidal suspension of the solid phase in a water-ethanol mixture. We perform molecular dynamics simulations to study the nanoscale structure of AWC aggregates. We characterize the size and shape of the aggregates at several key locations in the ternary phase diagram. The role of ethanol in forming the interface between the solvent and the solute phase is highlighted. We found that the structure of the aggregates formed in the ternary solution resembles the disordered sponge-like structures observed when AWCs were inserted into lipid membranes. Such permeable sponge architectures allow the passage of water despite their non-crystalline organization and have been previously shown to be consistent with AWC permeation measurements in membrane environments.

Introduction

Designing optimal desalination technologies is a challenge for the coming years. Currently available technologies are still too energy consuming for widespread and practical application. Artificial Water Channels (AWC), based on the introduction of self-assembling molecules into membranes, represent a promising solution to increase the reverse osmosis capability of the filtering membranes. Recent advances have been made to improve performance in terms of ion blockage and water permeability.[1–3]

The transport properties of synthetic self-assembled channels have been characterized in liposome vesicles and in polyamide membranes (PA).[3] The preparation of PA membranes involves two steps. The first is to dissolve the self-assembling molecules in the form of powder in an ethanolic solution. The second step is the polymerization of the components of PA into a phase containing the AWCs for highly selective transport of water. The best results for AWC formation and incorporation into the PA phase were obtained by maximally solubilizing large droplets of the elemental AWC-forming molecules in the ethanolic solution.[2] Dilution of the binary mixture with water leads to the formation of a stable colloidal solution. Using dynamic light scattering (DLS), the experimenters found that the size of the droplets increases with the volume ratio of water to ethanol until precipitation occurs.[3]

At this ns, we lack a microscopic view of the aggregation process that leads to the formation of AWCs in PA membranes. In our previous work, we used molecular dynamics simulations (MD) to study the behaviour of AWCs in lipid membranes by assuming their crystallographic X-ray structures[4–6] as the initial state for the AWCs. Investigation of the aggregation process can provide additional insight into the structures of functional AWCs in membranes without relying on information from the separate X-ray structure. To this end, in the present work we model the ternary mixtures consisting of the AWCs forming molecular aggregates with water and ethanol.

The phenomenon of spontaneous emulsification induced by rapid dilution of an ethanolic solution is known as the "ouzo effect".[7] It applies to many ternary mixtures containing an

insoluble species in water and a mixture of water and a so-called "hydrotropic" solvent such as ethanol, i.e., a species that mixes well with both water and the solute. The formation of a colloidal solution with droplets of size $100 - 500nm$ and its long-term metastability are beyond the scope of this study; the characteristic size and time scale of the process are beyond the current capabilities of all-atom molecular simulations. Nevertheless, the study of the homogeneous liquid phase is possible with MD simulations. In particular, one can characterize the phase diagram of the ternary systems and predict the aggregation patterns, which could be experimentally validated.[8,9]

In this work we have performed MD simulations to study the aggregation of small amphiphilic molecules forming AWCs in a ternary mixture of water, ethanol and solute. We address the following questions: (i) the effect of concentration and water-ethanol ratio on the stability and aggregation mechanism, (ii) the interplay between the solute and solvents in the aggregates, (iii) the similarity between the aggregate structures in the ternary solution and the relaxed structures of AWCs in lipid membranes.

Simulation Methods

Chemical compounds

We have studied the behaviour of four different compounds sharing a hydrophobic tail with eight carbons. As shown in Figure 1, we consider one molecule from the hydroxy channels and two molecules from the imidazole family. We have previously reported the filtering properties of AWCs formed by these compounds in bilayer lipid membranes. For comparison, we have also studied the behaviour of the 1-octanol compound known to form *Ouzo* emulsions and *pre-Ouzo* structures in the single-phase system.[10]

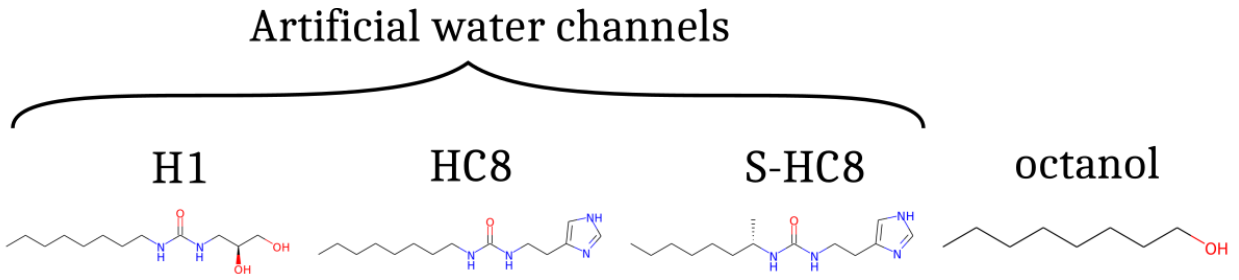


Figure 1: Compounds used in this study. H1 is part of the hydroxychannel family, HC8 and S-HC8 are imidazole-based compounds and octanol, a long-chain alcohol, is used here as a control compound. These compounds have previously been studied by MD simulations in refs.[4,5,11]

Molecular Dynamics Simulation Setup

All simulations were performed using GROMACS 2020.3 software.[12] For each system and phase ns in the phase diagram, we fixed the initial number of solute molecules (from 100 to 463 molecules) and the size of the cubic simulation box ($6nm$ and $10nm$). The solute concentration is controlled by these two parameters. After the solute molecule was inserted at a random position using Packmol software,[13] the system was solvated in a pre-equilibrated solvent mixture with a given fraction of water in ethanol.

We used the force field parameters already tested in previous studies of AWCs in lipid bilayers.[4,6] We used the CHARMM general force field (CgenFF)[14,15] automatic procedure to assign the non-bonded parameters and partial charges of the ethanol molecule. In coherence with the force field for solutes and ethanol, we used the TIP3P model for water.[16] Note that refinement of the interactions between water and ethanol, as proposed for the study of phase transitions,[17] is beyond the scope of this work.

All molecular simulations were performed under periodic boundary conditions with constant pressure and an integration time step of 2 fs and using the Leap-Frog integrator. Bonds involving hydrogen atoms were constrained using the LINear Constraint Solver algorithm.[18] The Particle Mesh Ewald method was used to treat Coulomb interactions with a real space cutoff of 12 Å. For the LJ interactions, a force switching function was applied over the 10-12 Å range and a long range correction was used to account for the LJ interactions beyond the

cutoff distance. The Nosé-Hoover thermostat and Parinello-Rahman barostat were used to maintain temperature and pressure at $T = 310$K and $P = 1$bar, respectively. All systems were minimized for 400000 steps using position restraints of 1000 $kJ.mol^{-1}$ for the water molecules. Production runs were extended by 500 ns without any restraints.

Simulation and Data Analysis

The detection of the aggregates is based on a clustering algorithm, DBSCAN (Density-Based Spatial Clustering of Applications with Noise).[19] We chose this method because it is easy to implement via the library scikit-learn. It detects clusters based on the maximum distance between two neighbors in the same cluster, taking into account the *minimum-image* convention. The accuracy of the clustering method depends mainly on the cutoff parameter, which we set to 3 Å. A more detailed description of the clustering method can be found in the supporting information. We compute the number of clusters, the size of the maximum cluster, and the size distribution of the clusters. Note that the time evolution of the number of clusters and the size of the maximal cluster provides a good indication of the convergence of the state of the mixtures. We verified that the number of clusters remained constant after a few nanoseconds of simulation without any constraints, see Figure S1.

To investigate the structure of the micelles, we calculated the eccentricity e of the aggregate as described in Ref.[20] When e tends to zero, the aggregate has a spherical shape. When e tends to one, the aggregate has a dominant dimension like a 1D object. This analysis is performed using the MDAnalysis,[21] Numpy, and Numba libraries. For more details, see the supporting information.

We can characterize each water molecule as a function of its immediate environment: We calculate for each water molecule the distribution $\phi(n_s, n_w, n_e)$ of the number of nearest neighbors in a coordination shell, either solute (n_s), water (n_w), or ethanol (n_e) molecules. In this way, one can detect the water molecule in the bulk ($n_s = 0$), the water molecule at the interface (medium n_s values), and the water molecules trapped in the aggregates (large

n_s values).

In addition, we have studied the penetration of the solvent into the aggregate phase. For this purpose, we use statistics on a few hundred recorded configurations for a selected aggregate of a given size. We calculated the number of molecules of each component of the ternary mixture (solute, water, and ethanol) in concentric spheres around the center of mass of the aggregate.

We also investigate the structural features of the aggregates by measuring the radial distribution functions (RDF) between different chemical groups common to all compounds. For this purpose, we use customized scripts adapted from the MDAnalysis library.

All molecular visualisations shown in figures were produced using VMD. [22]

Results and Discussion

Phase Separation in Ternary AWC-ethanol-water Mixtures

Aggregation of hydroxy channel molecules is studied for different water/ethanol molar ratios at fixed solute concentration, $C = 769\,mmol/L$, see pathway I in Figure 2a. The initial size of the box is fixed at $6\,nm$ and the total number of solute molecules in the box is $N = 100$.

When the water fraction is small ($f < 0.83$), the system consists of a single liquid phase, and the solute is dissolved by the solvent mixture, see Figure 2d. When the water fraction increases at a fixed solute concentration, a phase transition occurs in the system. At larger water fraction ($f > 0.83$), the system separates into a solute-rich phase and a solute-poor phase. The size distribution of the aggregates and their eccentricity are shown in Figure 2b and Figure 2c. When the size distribution peaks at $N = 100$ molecules, the system has reached the percolation limit, with all molecules belonging to the same cluster. It is worth noting that the single-phase system has a tendency to form large clusters. Nevertheless, the eccentricity of these aggregates is zero, which is due to a uniform distribution of the solute in the box rather than a spherical shape. This is mainly due to the high solute concentration

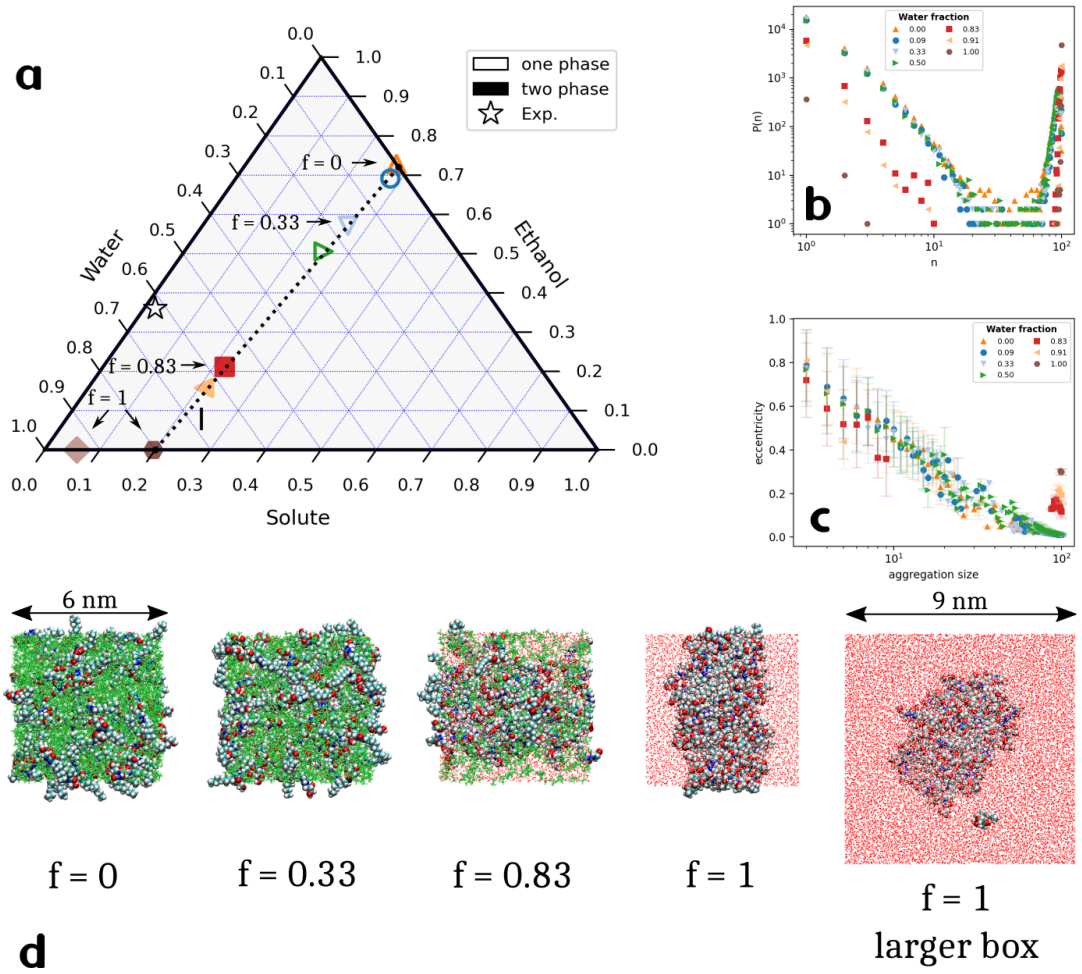


Figure 2: (a) Phase diagram of a ternary solution. The system along path I was studied in molecular dynamics simulations. (b) Distribution of aggregate sizes in log-log scale and (c) eccentricity of aggregates over the time interval 50-500 ns for a composition range along path I. The aggregate size, n, is the number of molecules in the same cluster. (d) The final snapshots for simulations at different water-ethanol molar ratios f. The solute molecules are represented by Van der Waals spheres with a colour associated with the atom type, and the solvent molecules are represented by red (green) sticks for water (ethanol). The rightmost panel corresponds to a lower solute concentration (brown diamond in the phase diagram) due to a larger solvation box.

in the box ($C = 769\ mmol/L$) and the cutoff used in the clustering algorithm. Therefore, the broad peak at $n \approx N$ is understood to be a homogeneous and dense liquid phase that approximates a bicontinuous phase (see also the log-linear plot in Figure S2).

As shown in Figure 2c and Figure 2d, the large aggregates have higher eccentricity when the water content is high. However, the formation of a cylindrical structure connecting the opposite sides of the simulation box in the extreme case $f = 1$ is an artefact caused by the periodic boundary conditions. We performed an additional simulation with a lower solute concentration, using the cylindrical structure as the initial solute configuration and a larger solvation box. The final configuration we obtained after 500 ns is shown in Figure 2d. The largest aggregate relaxes to an ellipsoidal shape with a lower eccentricity, see Figure S3. Note that this system corresponds to a different region of the phase diagram with a lower solute concentration. Interestingly, the largest aggregate is stable and includes the total number of solute molecules. Moreover, the shape of the formed micelle is flat and one can distinguish the hydrophobic core, two opposite sides enriched with hydrophilic regions, and a lateral less hydrophobic region exposed to the solvent. For more microscopic views, see Figure S3.

We extended the analysis of pathway I to the other compounds mentioned in Fig.1, i.e., HC8, S-HC8, and 1-octanol. The detailed results are shown in Figure S4. We show that all the compounds forming AWCs have similar phase diagrams. In contrast, the 1-octanol molecule shows a different behaviour. We ns out a difference in the phase state at the lower molar water-to-ethanol fraction: The octanol is completely solvated by the solvent mixture and the cluster analysis does not indicate large coherent clusters. This is somehow related to the low hydrophilicity of the octanol compound, which has only one hydroxyl group, insufficient for the formation of a large hydrogen-bonded network.

The structure of the observed large aggregates could be related to the inner cores of the droplets of the experimentally observed emulsion, although further characterization of the local density of the individual components of the emulsion is lacking at this stage.

Mesoscale Aggregate Formation and Ouzo Effects

Colloidal emulsions usually consist of two phases, a high-concentration solute and a low-concentration solute. The latter could also be related to the early stage of AWC-ethanolic solutions before dilution. Moreover, the experimental observations of the colloidal phase[3] were made in a region of the phase diagram corresponding to a lower concentration of AWC solute, see Figure 2a. In order to study the low concentration solute phase, we considered an additional path in the diagram. This choice allows to study aggregates of medium size ($< 10nm$) and to characterize their structure.

In Figure 3 we show the results along the path II, for a range of solute concentrations at a fixed molar water-ethanol ratio $f = 0.83$. We examine the size distribution and shape of the hydroxychannel solute. For this set of simulations, the box length is fixed at $10nm$ to provide a better sampling of aggregate sizes. In Figure 3a, we distinguish three regimes. The first regime is molecular solubilization at small concentrations ($C = 42, 83mmol/L$). It is characterized by a monotonic decay of the distribution of size aggregates.

In contrast, the regime of phase separation is achieved at the highest concentration, $C = 769mmol/L$. Namely, in this regime a large cluster is formed, which permeates all molecules of the system (463 molecules).

An intermediate state is observed at the other concentrations. For example, at the concentrations $C = 166mmol/L$ and $C = 382mmol/L$ we observe a deviation from the monotonic decay of the aggregate size distribution, indicating a high probability of formation of aggregates with large size, i.e., $\sim 20 - 80$ molecules for $C = 166mmol/L$. At $C = 382mmol/L$, a positive deviation of the decay at $n > 10$ is observed, as well as an additional peak around $n = 200$ molecules, which is close to the total number of molecules for this simulation. This latter peak is of the same type as those observed in the pathway I simulations at a small water-to-ethanol ratio, and is related to the percolation of solute molecules induced by the finite size. Over the size range of $20 < n < 80$, the eccentricity of the clusters is constant (about 0.4). This is consistent with visual inspection of the mesoscale aggregates showing

nonspherical objects with a very irregular surface exposed to the solvent mixture.

We report here the occurrence of mesoscale aggregates in MD simulations over a broad size spectrum rather than specific stabilized sizes over the time scale of the simulation. Local peaks in the size distribution have been referred to as "pre-ouzo" regimes in other surfactant-free microemulsions (SFME) such as ethanol-water-1-octanol systems.[11] It should be noted that a broader size distribution for aggregation has been reported for the same system in the presence of ions.[23]

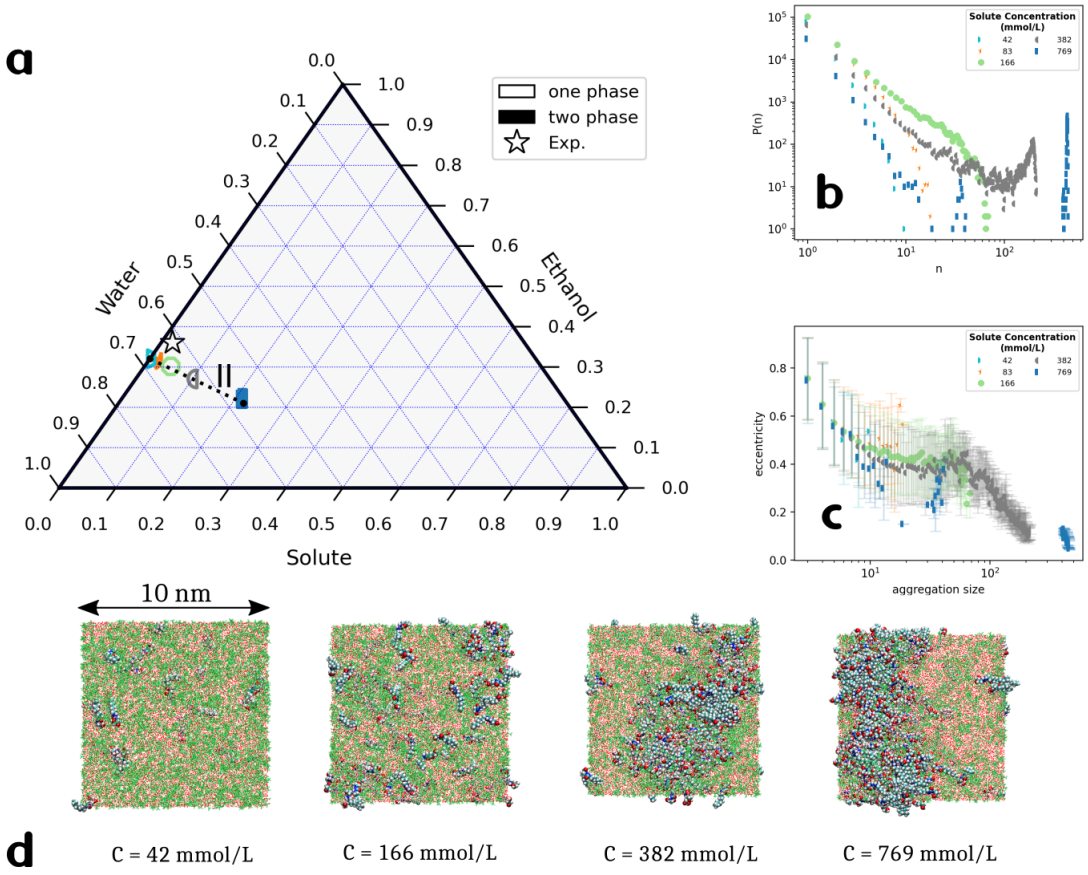


Figure 3: (a) Phase diagram and path II explored in molecular dynamics simulations with coordinates related to the weight% of the three components. (b) Distribution of aggregate sizes in log-log and (c) eccentricity of aggregates over the time interval 50-500 ns for a composition range following the path II. (d) The final snapshots of a selection of simulations with increasing solute concentration C. The molecular representations are depicted in the same way as in Figure 2d).

At this stage, we show that aggregation at the mesoscale is possible near the miscibility

gap. The micelle-like mixture is highly polydisperse near the transition.

Molecular Insights and Determinants of Aggregate Formation

To understand the relative stability of intermediate and large aggregates, we now focus on the structure of the solvent around the hydroxy channel aggregates as well as the local ordering of H1 aggregates within the large aggregates.

The Role of Ethanol

We characterize the mutual presence of water and ethanol at the interface of an aggregate: we have considered all water molecules in the system and characterized their local environment by counting the number of surrounding neighbors.

In Figure 4 we show the 3D distribution of the number of neighbors around all water molecules. To perform this analysis, we consider only the heavy atoms (CNO atoms). A whole molecule is considered a neighbor if at least one of its heavy atoms has a distance less than $d = 6.2$Å from the oxygen atom of the water molecule. This distance corresponds to the first minimum observed in the RDFs of heavy atom pairs between solute and water molecules, see Figure S6.

Figure 4 illustrates the maximum area of the distribution of the number of neighbors of water molecules for different water fractions. In the presence of a single homogeneous phase (for a water fraction $f \leq 0.50$), the number of ethanol neighbors (y-axis) decreases as the number of solute neighbors increases. On the contrary, for $f \geq 0.83$ we observe that the number of ethanol neighbors is slightly higher for an average number of solute neighbors ($n_{H1} = 4$, mean contour) than in the case of bulk ($n_{H1} = 0$) or an isolated water in the core of the aggregate ($n_{H1} = 8$). The intermediate values of n_{h01} correspond to the interface between the aggregate and the bulk. Thus, when a large dominant cluster forms along pathway I, hydration of the interface regions appears to be accompanied by ethanol enrichment.

We performed a standard analysis by isolating an aggregate and considering the environ-

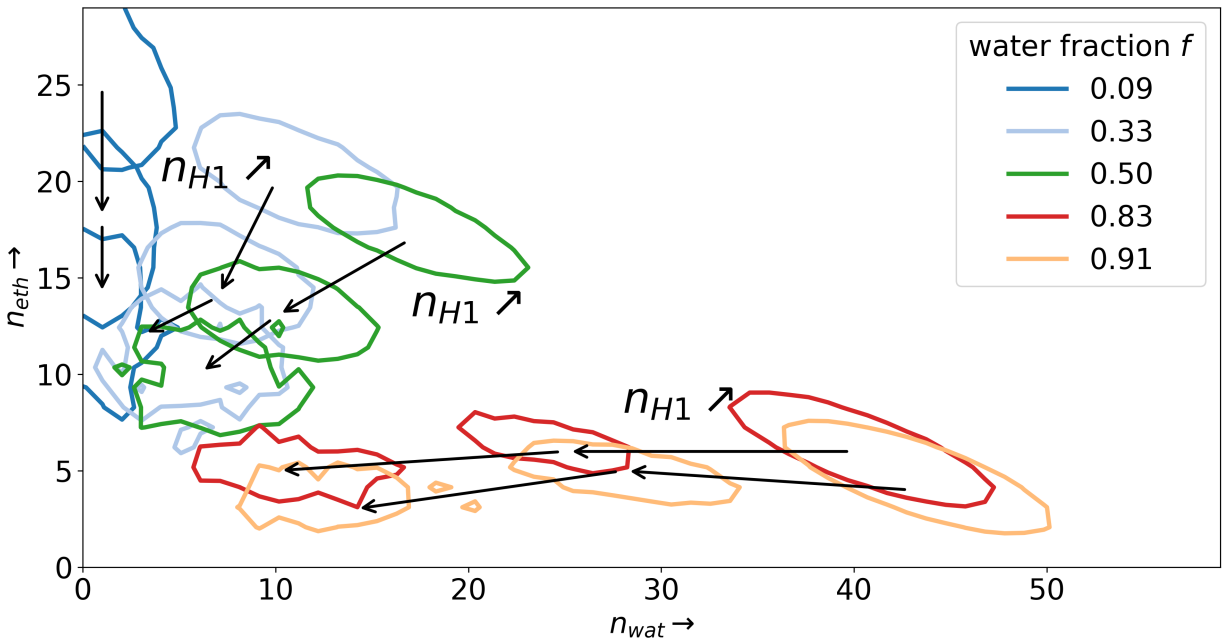


Figure 4: Distributions of the number of neighbors of water molecules $\phi(n_{H1}, n_{wat}, n_{eth})$ in a coordination shell, for the simulations of path I. Shown are only the contours delimiting the region around the maximum of the distribution. The arrows connect the maxima of the distributions for increasing numbers of solute neighbors n_{H1}. Three values for n_{H01} are shown, namely 0, 4, and 8. See also Figure S7 for a complete plot of the distributions of the number of neighbors. Note that in this representation, the center of the contour do not necessarily match the maximum of the distribution.

ment as a function of distance from the center of mass of the aggregate[11,20]. For this purpose, we focused on aggregates much smaller than the simulation box to avoid artifacts of finite size. We isolated an aggregate from a simulation of path II ($C = 166 mmol/L$, $f = 0.83$) that has a relevant population of intermediate aggregate sizes (see Figure 3). In Figure 5 we give the radial density of all three components around an aggregate of 20 molecules. It shows that water and ethanol are present near the center of the aggregate. Also, we observed that the ethanol distribution has a maximum around 15 Å. We made sure that other clusters of different intermediate sizes behave similarly (Figure S8).

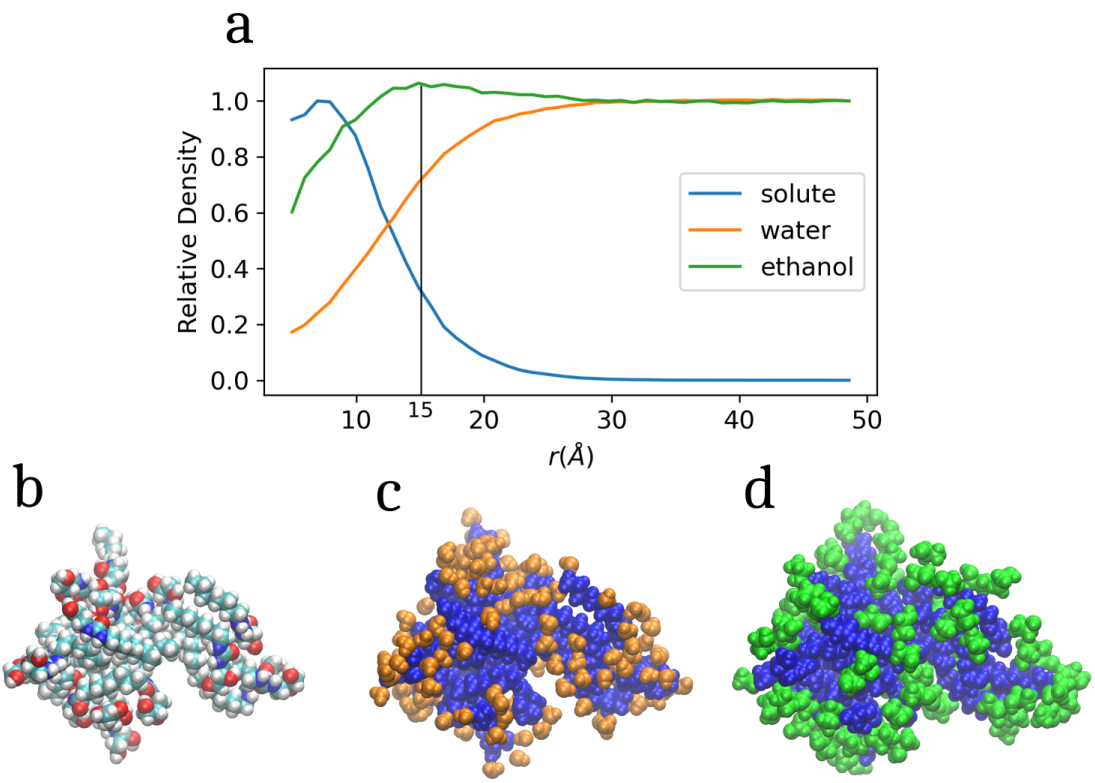


Figure 5: (a) Radial density around the center of mass of aggregates with 20 hydroxyl channel molecules in the absence of large scale phase separation ($C = 166 mmol/L; f = 0.83$). Snapshots of an aggregate with (b) solute molecules, (c) solute(blue)+water(orange) and (d) solute(blue) + ethanol(green)

Eccentricity measurements and visual inspection of the isolated clusters, as shown in Figure 5, indicate that the structure of the mesoscale aggregates is disaggregated, with disordered interfaces and solvent molecules penetrating the inner core of the aggregate.

Inter-molecular Solute Contacts Drive Aggregation

In this section, we focus on the influence of water to induce intermolecular contacts between solute molecules.

In Figure 6 we show the evolution of RDF functions and coordination number between oxygen atoms of the solute along path I. The first peak of $g(r)$ at $r = 4.7$Å is associated with hydrogen bonding between a donor and an acceptor or the oxygen atom and the hydrogen atoms of the urea-containing part of the molecule.

We used the oxygen atoms to compare with the well-defined X-ray structure. As in the crystalline arrangement, in the case of urea-urea association, a second peak should occur at twice the distance of the first peak if the aggregates have long-range order.

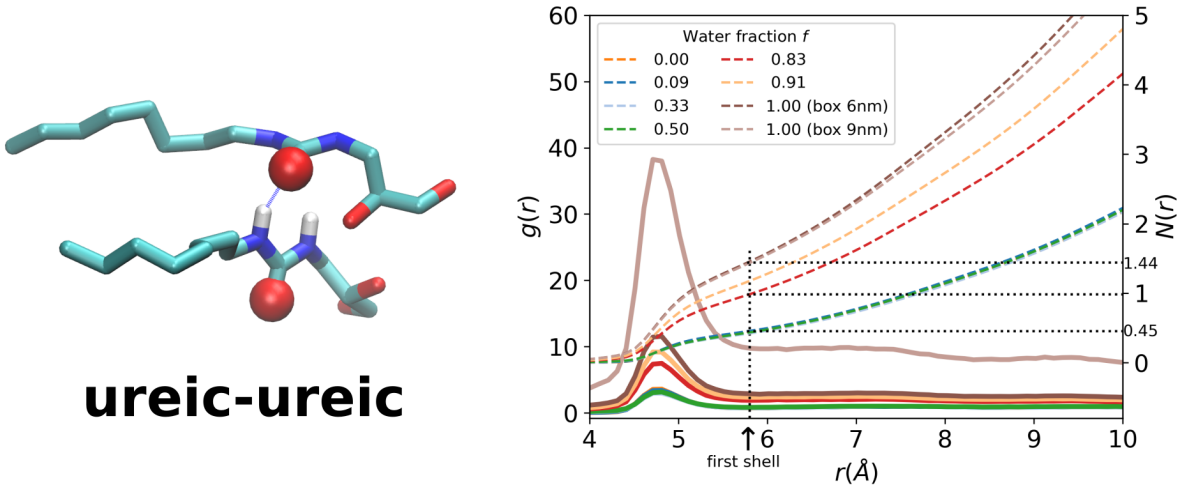


Figure 6: Radial distribution function $g(r)$ (plain line) and coordination number $N(r)$ (dashed line), calculated for all pairs of oxygen atoms in the urea-containing part of hydroxy channel molecules at different water fraction (path I).

The RDF curves of the homogeneous single-phase systems ($f \geq 0.50$) were superimposed. We observe a significant increase in the amplitude of the first peak of the ureic-ureic RDF function when the phase transition occurs ($f \geq 0.83$), as well as an increase in the coordination number, i.e., the average number of ureic neigbhors in the first shell. The second peak of the pair-pair correlation function $g(r)$ around 7Å is much broader. The absence of a clear

peak around 9Å seems to indicate that the inter-ureic interactions are not favored, even at a high water fraction. Other pairs of atoms in the ureic part and in the hydrophilic head were considered, see Figure S9: the first peak in the RDFs between the ureic part and the other alcoholic functions does not vary much. This observation indicates that the intermolecular contacts between the alcohols and ureic parts are independent of the degree of hydration of the aggregate interface.

We also verified that the ureic-ureic contacts are maintained when the cylindrical shape is relaxed to the flat micelle in the binary mixture water solution ($f = 1.0$; box 9nm in Fig.6). Both the cylindrical ($f = 1.0$; box 6nm) and flat micelles ($f = 1.0$; box 9nm) are characterised by the same ureic coordination number in the first shell, about 1.44, which is larger than in the other systems with the presence of ethanol. This suggests that hydration of the aggregate interface favors inter-ureic contacts.

At this stage, we have not discussed the positions of the RDF maxima. These peaks are related to the degree of disorder of the solute phase. In the next section, we compare the structure of the aggregates obtained in this study with those observed in previous simulations in a lipid membrane environment, i.e., whether the aggregates in water-ethanol mixtures can be precursors for functional water-permeable aggregates in a lipid bilayer.

Comparison with Previous Studies in Lipid Environments

In Figure 7, we compare the RDF functions of four different structures, the crystallographic one, the crystal structure relaxed in a lipid membrane obtained by MD simulation, the sponge-like structure embedded in the membrane structure also observed in our previous simulation, and the flat micelle structure obtained in the ternary mixture in the current study.[5] The second peak of the crystal-like bilayer structure in the membrane is not present in the bulk aggregate (see dashed and red lines in Figure 7). This peak is related to the lateral spacing between ureic cores when the molecules are aligned along their hydrophobic tails. See Figure S10 for more details on the nature of the peaks in the RDFs of the crystal-like

structure. The third peak of the crystal structure in the bilayer membrane is less pronounced but still present in the aggregate structure. This is related to the formation of a hydrogen-bonded chain from three adjacent ureic patterns.

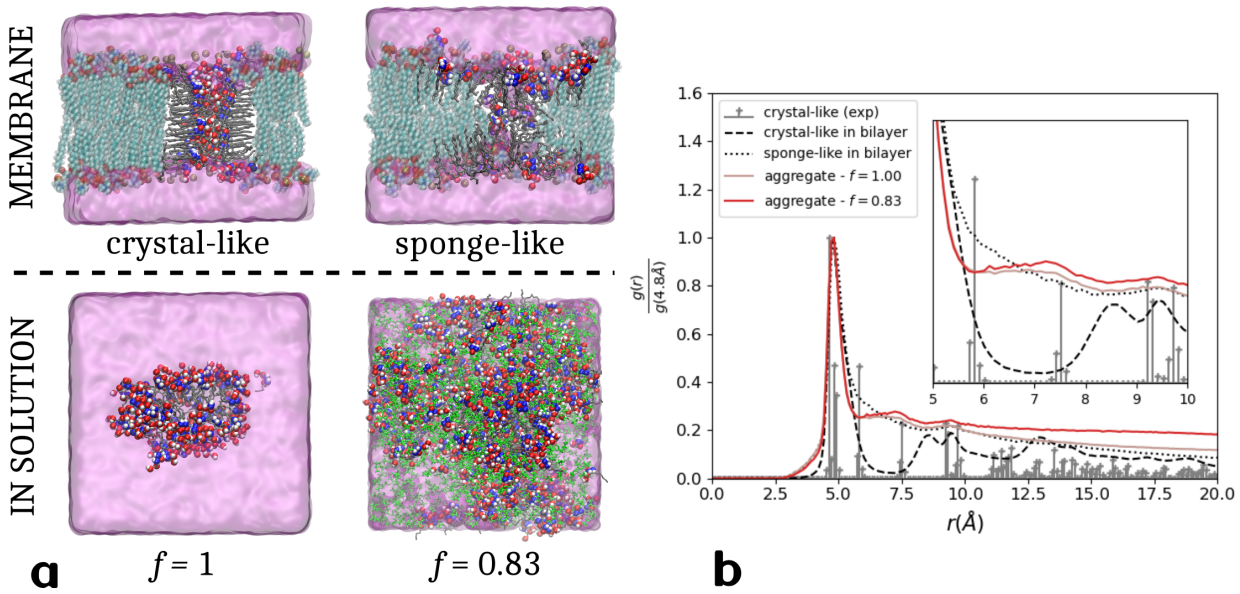


Figure 7: (a) Aggregation in different environments studied with molecular dynamics. The snapshots of the crystal-like and sponge-like systems are taken from our previous work.[5] The aggregation in solution refers to this study, and the snapshots were extracted from the simulation in a pure water solvent ($f = 1$) and in a solvent mixture with a water-to-ethanol ratio of $f = 0.83$. (b) Comparison of the RDF between the urea-containing parts in the AWCs structures shown in (a). The inset shows a zoom of the 5-10Å region. RDF functions were scaled by the height of the first peak for clarity.

We observe a great similarity between the RDF curve of the aggregate structure in pure water as solvent and that of the sponge-like structure. This result is not surprising when comparing the orientation of the solute molecules in an aggregate. Indeed, the hydrophobic tails tend to align with the membrane in the sponge-like aggregate and the latter geometry is then close to the flat micelle geometry with alcoholic functions aligned outward to the interface in contact with water. Therefore, interureic contacts are more likely to occur in this geometry than in a spherical geometry. As can be seen in the inset of Figure 7, the aggregates and the sponge-like structure differ mainly in the region between 6-8 Å. A peak at $r = 7.2$Å is present only for the flat micelle structure. We verified that this peak corresponds

to a stable arrangement of two water molecules bridging two adjacent solute molecules via hydrogen bonds with their urea-containing cores (see Figure S10c).

It should be noted that, using the same force field parameters as in this study, the sponge-like structure is difficult to sample due to the unsealing of the patch in the membrane after 1µs of simulation.[5] As long as the water manages to flow through the system, the sponge-like patch in the membrane remains dense and non-negligible inter-ureic interactions occur along the plane of the membrane.

Overall, the AWC sample preparation in ethanol-water mixtures may indeed favor the formation of precursors of sponge-like aggregates, which subsequently mature to fully membrane permeating sponge networks when inserted in a bilayer environment.

Conclusion and Perspectives

AWCs can be dispersed in ethanolic solutions to facilitate their incorporation into PA membranes. In this study, we used molecular dynamics to investigate the size, shape, and local structure of AWC aggregates in an ethanol-water mixture. As reported for other surfactant-free microemulsions,[10,11] the formation of stable aggregates is related to specific solvation effects. The solvation layers at the interfaces and the penetration of the solvent components into the aggregate phase are the key features of these new materials. If ethanol plays the role of a surfactant preventing phase separation and favouring solubilization of the mesoscale aggregates, water favours the alignment of the small solute molecules along their ureic cores. Moreover, the comparison between the aggregates in the ethanol-water mixture and the previously studied aggregates in bilayer membranes[4,5,24] shows that the sponge-like structure is formed at the preliminary stage of bulk aggregation. This suggests a pathway for the formation of loose aggregates of small hydrophilic molecules that exhibit enhanced water transport once inserted into a membrane matrix.[1,3] Our numerical approach makes several quantitative predictions about the structural states of self-assembling AWCs in solution and

these hypotheses could be tested experimentally in the future.[9]

Acknowledgement

We gratefully acknowledge support from ANR grant ANR-18-CE06-0004 (WaterChannels). and the "Initiative d'Excellence" program of the French government (Grant "DYNAMO", ANR-11-LABX-0011-01). This work was carried out with the HPC resources of GENCI and LBT. We thank Geoffrey Letessier for technical support at LBT. We are grateful to Mihail Barboiu and Libo Huang for stimulating discussions.

Supporting Information Available

Supplementary Text : Analysis Code ; Clustering Algorithm ; Chemical details

Table S1: List of compounds used in this study and related chemical properties.

Figure S1: Biggest cluster and number of clusters timeseries.

Figure S2: Cluster size and cluster eccentricities in linear.

representation.

Figure S3: Finite-size effects.

Figure S4: Cluster sizes for different compounds.

Figure S5: Effect of the distance cutoff.

Figure S6: RDF plots and solvation shell of the solute.

Figure S7: Distribution maps of water neighbors.

Figure S8: Radial density around cluster COM at several cluster size.

Figure S9: RDF plots between polar groups of the solute.

Figure S10: Structural features in ordered and disordered aggregates.

References

1. Di Vincenzo, M.; Tiraferri, A.; Musteata, V.-E.; Chisca, S.; Sougrat, R.; Huang, L.-B.; Nunes, S. P.; Barboiu, M. Biomimetic Artificial Water Channel Membranes for Enhanced Desalination. *Nat. Nanotechnol.* **2021**, *16*, 190–196.

2. Di Vincenzo, M.; Tiraferri, A.; Musteata, V.-E.; Chisca, S.; Deleanu, M.; Ricceri, F.; Cot, D.; Nunes, S. P.; Barboiu, M. Tunable Membranes Incorporating Artificial Water Channels for High-Performance Brackish/Low-Salinity Water Reverse Osmosis Desalination. *PNAS* **2021**, *118*, e2022200118.

3. Huang, L.-B.; Di Vincenzo, M.; Ahunbay, M. G.; van der Lee, A.; Cot, D.; Cerneaux, S.; Maurin, G.; Barboiu, M. Bilayer Versus Polymeric Artificial Water Channel Membranes: Structural Determinants for Enhanced Filtration Performances. *J. Am. Chem. Soc.* **2021**, *143*, 14386–14393.

4. Licsandru, E.; Kocsis, I.; Shen, Y.-x.; Murail, S.; Legrand, Y.-M.; van der Lee, A.; Tsai, D.; Baaden, M.; Kumar, M.; Barboiu, M. Salt-Excluding Artificial Water Channels Exhibiting Enhanced Dipolar Water and Proton Translocation. *J. Am. Chem. Soc.* **2016**, *138*, 5403–5409.

5. Huang, L.-B.; Hardiagon, A.; Kocsis, I.; Jegu, C.-A.; Deleanu, M.; Gilles, A.; van der Lee, A.; Sterpone, F.; Baaden, M.; Barboiu, M. Hydroxy Channels–Adaptive Pathways for Selective Water Cluster Permeation. *J. Am. Chem. Soc.* **2021**, *143*, 4224–4233.

6. Hardiagon, A.; Murail, S.; Huang, L.-B.; van der Lee, A.; Sterpone, F.; Barboiu, M.; Baaden, M. Molecular Dynamics Simulations Reveal Statistics and Microscopic Mechanisms of Water Permeation in Membrane-Embedded Artificial Water Channel Nanoconstructs. *J. Chem. Phys.* **2021**, *154*, 184102.

7. Vitale, S. A.; Katz, J. L. Liquid Droplet Dispersions Formed by Homogeneous Liquid-Liquid Nucleation: "The Ouzo Effect". *Langmuir* **2003**, *19*, 4105–4110.

8. Schöttl, S.; Lopian, T.; Prévost, S.; Touraud, D.; Grillo, I.; Diat, O.; Zemb, T.; Horinek, D. Combined Molecular Dynamics (MD) and Small Angle Scattering (SAS) Analysis of Organization on a Nanometer-Scale in Ternary Solvent Solutions Containing a Hydrotrope. *J. Colloid Interface Sci.* **2019**, *540*, 623–633.

9. Prévost, S.; Krickl, S.; Marčelja, S.; Kunz, W.; Zemb, T.; Grillo, I. Spontaneous Ouzo Emulsions Coexist with Pre-Ouzo Ultraflexible Microemulsions. *Langmuir* **2021**, *37*, 3817–3827.

10. Zemb, T. N.; Klossek, M.; Lopian, T.; Marcus, J.; Schöettl, S.; Horinek, D.; Prevost, S. F.; Touraud, D.; Diat, O.; Marčelja, S. et al. How to Explain Microemulsions Formed by Solvent Mixtures without Conventional Surfactants. *PNAS* **2016**, *113*, 4260–4265.

11. Schöttl, S.; Marcus, J.; Diat, O.; Touraud, D.; Kunz, W.; Zemb, T.; Horinek, D. Emergence of Surfactant-Free Micelles from Ternary Solutions. *Chem. Sci.* **2014**, *5*, 2949–2954.

12. Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*, 134110.

13. Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

14. Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

15. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM

General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

16. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

17. Chen, B.; Potoff, J. J.; Siepmann, J. I. Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *J. Phys. Chem. B* **2001**, *105*, 3093–3104.

18. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

19. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Portland, Oregon, 1996; pp 226–231.

20. Luft, C. M.; Munusamy, E.; Pemberton, J. E.; Schwartz, S. D. A Classical Molecular Dynamics Simulation Study of Interfacial and Bulk Solution Aggregation Properties of Dirhamnolipids. *J. Phys. Chem. B* **2020**, *124*, 814–827.

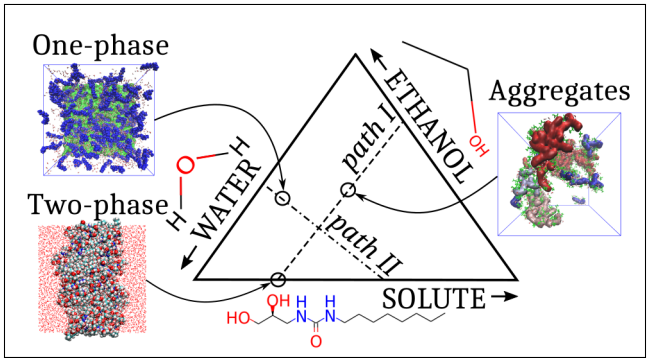
21. Beckstein, O.; Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S. et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. pp 98 – 105.

22. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

23. Schöttl, S.; Horinek, D. Salt Effects in Surfactant-Free Microemulsions. *J. Chem. Phys.* **2018**, *148*, 222818.

24. Murail, S.; Vasiliu, T.; Neamtu, A.; Barboiu, M.; Sterpone, F.; Baaden, M. Water Permeation Across Artificial I-Quartet Membrane Channels: from Structure to Disorder. *Faraday Discuss.* **2018**, *209*, 125–148.

TOC Graphic



For Table of Contents Only