BIOMEDICAL INFORMATICS

A statistical geometry analysis of simulated water-DMSO and water-MeCN binary mixtures for bio-molecular studies

Neetu Singh Yadav & Devapriya Choudhury*

[1]School of Biotechnology, Jawaharlal Nehru University, New Delhi-110067, India; Devapriya Choudhury – E-mail: devach@jnu.ac.in; Ph: +91 11 26704078; *Corresponding author

Abstract:
Water-Dimethylsulfoxide (DMSO) and water-Acetonitrile (MeCN) binary mixtures at various molar ratios ranging from 0 to 1 are studied using Molecular Dynamics (MD) simulations. Hydration properties of water in different regions of MeCN/DMSO are investigated by using the statistical geometry approach. The obtained results reveal that in water-DMSO simulations both water and solvent molecules prefer to be in mixed cluster forms, depending upon the concentration of DMSO. While in case of water-MeCN mixtures, self-association of water and acetonitrile molecules, take place, showing microheterogeneity associated with the water-MeCN binary mixtures. The results highlight the utility of statistical geometric analysis of MD simulation data of binary liquid mixtures for rapid screening of polar organic solvents in non-aqueous enzymology.

Abbreviations: POS - Polar Organic Solvent, χ_S - Mole fraction of Solvent.

Keywords: Statistical geometry, Microheterogeneity, Molecular Dynamics.

Background:
A key element of research in non-aqueous enzymology is therefore to screen for suitable organic solvents that have minimal effects on the structure of the enzyme of interest, or to screen for enzymes that are stable when high concentrations of polar organic solvents (POS) are added to the surrounding medium. While there have been a number of spectacular successes of the experimental screening approach [1], it is also important to develop *in silico* screening methods for POSs suitable for a particular protein, or even a rational approach to suggest mutations that can enable an enzyme protein to resist the harmful effects of added POSs. Molecular Dynamics (MD) simulations provide a promising approach for the study of enzymes in non-aqueous media. The force-field parameter set in use with much popular MD software has recently been enhanced with parameters for a large number of organic solvents [2, 3]. However, the very large number of particles that need to be taken into account in realistic MD simulations of proteins dissolved in partially for fully non-aqueous systems require immense

computational resources and often an inordinate amount of time, making a large scale MD based solvent screening program non-feasible.

A fairly large number of studies have accumulated that convincingly demonstrate that the deleterious effects of POSs on proteins stems from the effects that these solvents exert on the large-scale structure of water [4]. Hence one way to speed up the simulations would be to only consider mixtures of water and POSs and monitor the time-averaged large-scale structure of the mixtures and compare them with that of liquid water. Simply using mixtures of water and POSs would greatly reduce the particle numbers in the simulation and thus cut down the time required for analysis by a significant amount. We test this hypothesis by analyzing MD simulations of binary mixtures water and two organic solvents *viz.*, Acetonitrile (MeCN) and Di-methyl sulphoxide (DMSO), both of which are known to significantly affect the structure of proteins dissolved in them. We also tested the feasibility of using methods of statistical

BIOMEDICAL INFORMATICS

geometry to analyze the large-scale time-averaged structures of the binary mixtures as obtained from standard MD simulations.

Application of statistical geometry for the analysis of the large scale structure of molecular liquids were first developed by Bernal and others in the 1950s and was found to be a promising tool for the analysis and prediction of a number of experimentally measurable quantities [5]. Unfortunately, except for a few notable exceptions [6, 7], these methods have not been routinely used in bio-molecular simulations.

Systems and Methods:
Molecular Dynamics Simulations:
A series of MD simulations were carried out for mixtures of water and POS (MeCN and DMSO) where the concentration of the organic solvent was varied. Details of the MD simulation protocol and its validation are given in the **supplementary data**.

Statistical Geometry:
Statistical Geometry analyses were initiated by first carrying partitioning the space contained in the simulation box by means of Delaunay Tessellation [5]. Delaunay tessellation in 3D partitions space defined by a cloud of points into a series of non-overlapping tetrahedral. The vertices of the tetrahedral are defined using the cloud of points given as input. Points that are not a vertex of a particular tetrahedron are not allowed to be anywhere within the volume in the tetrahedron. In our case, every non-hydrogen atom contained within the simulation box defined the clouds of points. Since the tetrahedral obtained by Delaunay tessellation do not contain any other point within them, points across any edge of a given tetrahedron are by definition nearest neighbors. We used this property to carry out nearest neighbor analysis of different atoms in our binary mixtures. For this we considered only two types of atoms, the first type denoted by W consisted of only the Oxygen atoms of water molecules and the second type, denoted by S could consist of any non-hydrogen atom from the POS, which in case of MeCN could either be carbon or nitrogen, or carbon, oxygen and sulphur in case of DMSO. With these designations, we can have five types of tetrahedral depending on the identity of the vertices, there are: W_4, $i.e.$, composed of four W atoms, W_3S, composed of three W and one S atom and similarly, W_2S_2, WS_3 and S_4 **(Figure 1)**. We then counted the numbers of each class of these tetrahedral from every snapshot of our MD simulations.

The numbers were further normalized in terms of a log-odds ratio (f) as follows:
$$f_i = P_{i,obs} / P_{i,exp}$$

$P_{i,obs}$ denotes the observed probability of finding a particular class of tetrahedron and is given by:
$$P_{i,obs} = n_i / \Sigma\, n_i$$

where, n_i denotes the number of tetrahedral belonging to the ith class.

$P_{i,exp}$ denotes the expected probability of finding a particular class of tetrahedron and is given by:
$$P_{i,exp} = 4_k p_w\, (1-p_w)^{(4-k)}$$

where, k ($0 \leq k \leq 4$) denotes the number of w atoms in the particular tetrahedron class, and p_w denotes the probability of finding a w atom and is given by:
$$P_w = n_w / n_T$$

where, n_w denotes the number of water oxygen atoms and n_T denotes the total number of non-hydrogen atoms in the simulation box.

The local structure of the mixture can be quantified by the distortion of the Delaunay tetrahedral, this has been defined in terms of a parameter called tetrahedrality, which has been defined as follows [8]:

$$T = \sum \left(\frac{(l_i - l_j)^2}{15\overline{l^2}} \right)$$

A perfectly regular tetrahedron has a tetrahedrality of 0, which increasing deviations from regularity causes a corresponding increase in the parameter.

Results and Discussion:
Figure 2A and B plots the log-odds ratio (f), for different tetrahedron types as function of added solvent concentration. As can be seen from **Figure 2A**, in case of MeCN, the log-odds ratio for W_4 and S_4 types of tetrahedral become increasingly dominant with increasing MeCN concentration. This indicates that water-water and MeCN-MeCN associations are more preferred over water-MeCN associations. In case of DMSO, the picture is quite different, firstly none of the tetrahedron classes dominate to the extent as W_4 and S_4 types dominate in case of MeCN. Furthermore, the W_4 type shows a significant decline. The preference of water-water self-association in water-MeCN binary mixtures had been previously suggested both from experimental [9–12] as well computational [13, 14] suggest that clusters of pure water and pure MeCN can co-exist in binary mixtures, creating a phenomenon termed "microheterogeneity". The dominance of W_4 and S_4 type tetrahedral over the W_3S, W_2S_2 and WS_3 types strongly supports these observations.

In case of DMSO-water binary mixtures, some previous studies suggested the presence of $(H_2O)_2$:DMSO type of complexes at low concentrations of DMSO (χ_{DMSO} ~0.33) and $(DMSO)_2$:H_2O clusters at high concentrations (χ_{DMSO} ~0.6) [15], while others suggested the presence of clusters with a complex composition of water and DMSO [16]. Thus it can be concluded that significant amount of water-DMSO interaction occurs leading to water-DMSO complexes of various stoichiometries. Our statistical geometry analysis supports this view by showing that none of the five types of tetrahedral dominate to the extent observed in water-MeCN mixtures. The small but significant decline of W_4 tetrahedral points towards the possibility that pure water clusters are disfavored in water-DMSO binary mixtures.
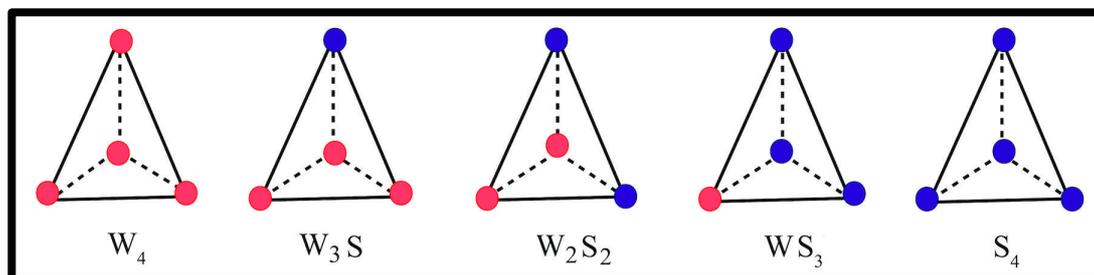
Figure 1: Five types of water-co-solvent association patterns expressed as vertices of Delaunay tetrahedral. Red and Blue colors denote water (W) and POS (S) atoms respectively.
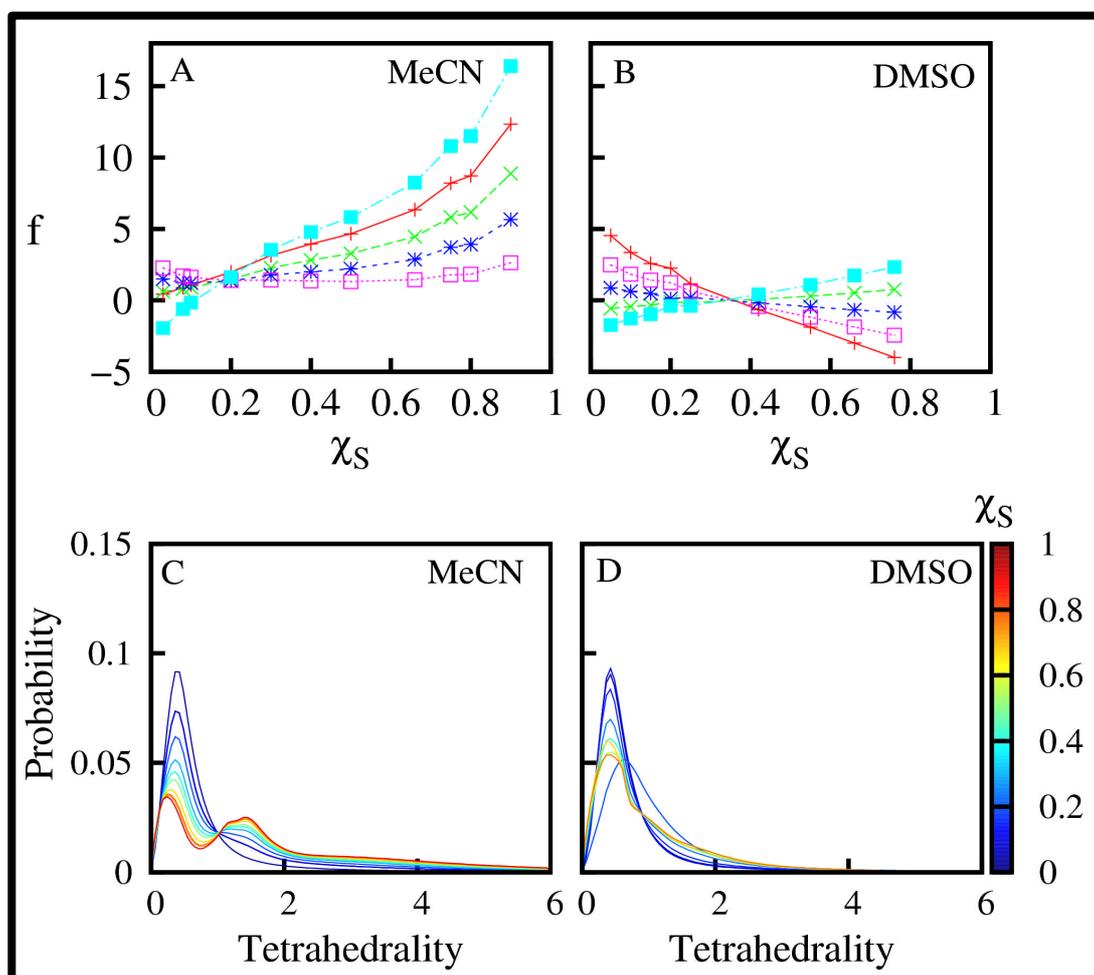

Figure 2: Over (under) abundance of different water-solvent association patterns in water-MeCN (A) and water-DMSO (B) mixtures as a function of MeCN and DMSO concentration (χ_S). A positive Log-Odds value (f) indicates a greater than expected occurrence of a particular association pattern and *vice-versa* for negative f values. Red, Green, Blue, Pink and Cyan color lines denotes W_4, W_3S, W_2S_2, WS_3 and S_4 types of tetrahedron respectively. Panel C, D plots the distribution of Tetrahedrality values for all water-solvent tetrahedral with increasing concentration (χ_S) of MeCN (Panel C) and DMSO (Panel C). For this, all types of tetrahedral were clubbed together and their tetrahedrality values were calculated.

The local spatial arrangement of the binary mixtures can be gauged from the tetrahedrality distributions shown in **Figure 2C and 2D**. In pure aqueous medium or in the presence of low concentrations of organic solvents, the peak of the tetrahedrality distributions are close to 0, indicating the dominance of nearly regular tetrahedral. This is expected behavior, since water molecule tends to make hydrogen bonded clusters with four other water molecules, and these hydrogen bonds are directed

towards the vertices of a regular tetrahedron. As the organic solvent content increases, the water-water hydrogen bond network begins to get disrupted which leads to increasing distortion in shape of the tetrahedral. What is interesting however is that in case of water-DMSO binary mixtures, increase in organic solvent content merely increases the spread of the tetrahedrality values. However, the dominating component is made up of nearly regular tetrahedral. In case of water-MeCN mixtures, the distributions turn bi-modal with a second peak with a tetrahedrality value near 1.75 becomes co-dominant with the first peak with tetrahedrality value near 0. An explanation of this difference becomes apparent when the tetrahedrality distribution for each class of tetrahedral is calculated separately **(Figure S2 – see supplementary data).** The figure shows that in water-DMSO binary mixtures, the tetrahedrality distribution of solvent free W_4 type of tetrahedral gets perturbed with increasing concentrations of DMSO. This behavior is not seen in case of water-MeCN mixtures. The perturbation in the tetrahedrality distribution of W_4 tetrahedral by DMSO indicates that the solvent can interfere with water-water association network. It should be noted that the water molecules that are participating in a W_4 tetrahedron could also be part of other mixed water-DMSO tetrahedral simultaneously. The perturbation of W_4 tetrahedral may therefore be due to the fact that molecules involved the formation of pure-water cluster, simultaneously be part of neighboring water-DMSO clusters, causing a spatial perturbation in the water-water interaction geometry. The geometry of pure S_4 tetrahedral also shows a dramatic concentration dependent change. At high water contents, there is a very sharp peak in the S_4 tetrahedrality distribution indicating a very specific DMSO-DMSO interaction geometry. With increase in DMSO content various other forms of interactions develop with the consequent increase in spread of the tetrahedrality values. Among the mixed type of tetrahedral, *i.e.*, W_3S, W_2S_2 and WS_3, in water-DMSO mixtures, one finds increase in DMSO content progressively increase the spread of tetrahedrality values. Taken together, one observes a diversity of water-DMSO association patterns in binary mixtures, which result in the observed spread of tetrahedrality values.

In case of MeCN the situation is quite different, there is almost no change in the shape of the tetrahedrality distribution of W_4 and W_3S types of tetrahedral. While W_2S_2 and WS_3 types show a clear bi-modal distribution with increasing organic solvent concentration. In case of the S_4 tetrahedral, type distribution is quite spread out, but one the modes observed in W_2S_2 and WS_3 distribution is clearly visible. The observed pattern indicates that there is a specific pattern of MeCN-MeCN association which is getting more pronounced with increasing MeCN concentration.

X-ray diffraction studies [12] indicate the formation of specific zig-zag clusters of MeCN formed bi dipolar interactions both in the pure liquid as well as in binary mixtures with high solvent content. The role of tetrahedral with tetrahedrality of 1.75 in the formation of MeCN clusters is for further investigation.

Conclusion:
Statistical geometric analysis of MD simulation data of binary mixtures was able to quickly detect significant differences in the structure of water-DMSO and water-MeCN binary mixtures. One can, in principle, apply the method of MD simulations to large number of binary mixtures with different organic solvent components. Statistical geometric analysis on the simulation data would provide a common set of parameters to classify such binary mixtures into different categories, and one or a few examples from each category can be subjected to further intense study regarding their effects on protein structures. The combinatorial problem of a large number of solvents and a large number of concentration regimes for each solvent is mitigated. Thus, it will be a reasonably rapid tool for screening of organic solvents in non-aqueous enzymology.

References:
[1] Matthijs N *et al*. J. Chromatogr. A 2006, **111**:48. [PMID: 16483582].
[2] Luzar A *et al*. J. Chem. Phys. 1993, **98**:8160.
[3] Grabuleda X *et al*. J. Comput. Chem. 2000, **21**:901.
[4] Arakawa T *et al*. Biophys. Chem. 2007, **131**: 62. [PMID: 17904724]
[5] Bernal JD. Nature. 1959, **183**:141.
[6] Vaisman II *et al*. J. Am. Chem. Soc. 1992, **114**:7889.
[7] Singh RK *et al*. J. Comput. Biol. 1996, **3**:213. [PMID: 8811483]
[8] Medvedev N *et al*. J. Non-Cryst. Solids 1987, **94**:402.
[9] Moreau C *et al*. J. Chem. Thermodyn. 1976, **8**:403.
[10] Jamroz D *et al*. J. Am. Chem. Soc. 1993, **115**:6165.
[11] Bertie JE *et al*. J. Phys. Chem. B 1997, **101**:4111.
[12] Takamuku T *et al*. J. Phys. Chem. B 1998, **102**:8880.
[13] Kovacs H *et al*. J. Am. Chem. Soc. 1991, **113**:5596.
[14] Mountain RD. J. Phys. Chem.A, 1999, **103**:10744.
[15] Luzar A *et al*. J. Chem. Phys. 1993, **98**:8160.
[16] Kirchner B *et al*. Chem. Phys. Lett. 2002, **364**:497.

Supplementary Data:

Methodology:
Two sets of MD simulations were performed. In first set, twelve MeCN-water mixtures were simulated. The concentration of MeCN was varied from mole-fraction of 0.03 to 1. In case of DMSO, eleven simulations were carried out where the concentration of DMSO was varied between molefractions 0.05 to 1. For comparison, a simulation of pure water was also carried out.

The number of water and POS required for a particular concentration were calculated by the following formula:
$C_{vol} = (M_{os}\rho_{os})n_{os} / (M_{os}\rho_{os})n_{os} + (M_w\rho_w)n_w$
Where, M, ρ, and n represent the molecular weight, density and number of molecules respectively.

Subscripts os and w denotes organic solvent and water respectively. In case of water-DMSO mixtures, the density of the system, below 50 % v/v DMSO is ~1 gm/ml which increases to 1.1 gm/ml at concentrations >50 % v/v DMSO. However, in water-MeCN, the density of the system decreases from 1 gm/ml with increasing MeCN concentration. Adequate numbers of water, DMSO and MeCN, at indicated concentrations were placed randomly, in a cubic box by using PACKMOL [1]. The number of Water, DMSO and MeCN molecules, along with the edge length of boxes are given in Tables S1, S2. The TIP4P model [2] of water and for the organic solvents, GAFF force-field parameters [3] was used. All atom parameters for DMSO, were downloaded from the website (www.virtualchemistry.org) [4], while for MeCN the parameters were taken from (http://research.bmh.manchester.ac.uk/bryce/amber/) site [5].

The simulation protocol consisted of three major steps: (a) In the first step, the box was energy minimized for 10,000 steps using steepest Descent method. (b) In the second step, the equilibration ran for 15 ns in isothermal-isobaric (NPT ensemble). During equilibration, the temperature and pressure were kept constant by using Berendsen's thermostat (τt) and barostat (τp) with time constants for temperature and pressure held at 0.1ps and 1.0ps respectively. (c) The production simulation was carried out for 15 ns under constant volume and temperature. Structures were saved at every ps. Equations of motion were integrated using the leapfrog algorithm, with time-step of 1 fs. Short-range electrostatics cut-off was set as 15Å. Particle-Mesh Ewald algorithm was used to account for long-range electrostatics [6]. van der Waal's energy cut-off was set at 15Å with dispersion corrections on energy and pressure. Neighbour lists were updated at every 10th step.

Table S1: Number of DMSO and Water molecules in the cubic simulation box.

Mole-Fraction (DMSO)	Water Numbers	DMSO Numbers	Box Vectors (nm)
0.0	1728	-	3.72, 3.72, 3.72
0.05	1298	86	3.65, 3.65, 3.65
0.1	1071	130	3.60, 3.60, 3.60
0.15	924	163	3.58, 3.58, 3.58
0.2	784	198	3.63, 3.63, 3.63
0.25	670	224	3.72, 3.72, 3.72
0.3	594	244	3.55, 3.55, 3.55
0.42	918	660	4.66, 4.66, 4.66
0.55	440	550	4.22, 4.22, 4.22
0.66	320	660	4.39, 4.39, 4.39
0.76	230	727	4.46, 4.46, 4.46
1.0	-	1000	4.82, 4.82, 4.82

Table S2: Number of MeCN and Water molecules in the simulation box.

Mole-Fraction (MeCN)	Water Numbers	MeCN Numbers	Box-Vector (nm)
0.0	1728	-	3.72, 3.72, 3.72
0.03	1000	40	3.50, 3.50, 3.50
0.08	900	80	4.00, 4.00, 4.00
0.1	900	100	4.00, 4.00, 4.00
0.2	800	200	4.00, 4.00, 4.00
0.3	680	340	4.00, 4.00, 4.0
0.4	570	420	4.00, 4.00, 4.00
0.5	430	430	3.50, 3.50, 3.50
0.66	280	550	4.00, 4.00, 4.00
0.75	150	550	3.82, 3.82, 3.82
0.8	120	520	4.00, 4.00, 4.00
0.9	50	550	4.00, 4.00, 4.00
1.0	-	500	3.62, 3.62, 3.62

Validation of MD simulations:

MD simulations were validated by calculating various concentration dependent physico-chemical parameters like density, viscosity, self-diffusion coefficient and compared with experimental values (**Figure S1**). In case of water-DMSO binary mixtures as the number of DMSO increases the density of system decreases (**Figure S1A**). The calculated density values are continuously over predicted from experimentally observed values with the error rate of 2.1-4.8%. While in the case of Water-MeCN binary mixtures, density of system decreases linearly with the rise in concentration of MeCN (**Figure S1B**), here the calculated density values are under-predicted as compared to experimental values but trend wise, they are similar to the experimentally observed values [7]. The self-diffusion coefficient is a dynamic property of the system. In case of water-DMSO binary mixtures, diffusion curve of water and DMSO molecules shows a progressive decrease up-to χDMSO ~0.33, but further increment of DMSO (χDMSO ≥ 0.3) concentration seems to have no effect on the diffusion of both molecules (**Figure S1C**). In the water-MeCN binary mixture, the diffusion of water decreases from up-to χMeCN ~0.4. While at high MeCN concentration, the diffusion remains unaffected by MeCN molecules (**Figure S1D**).

Unlike, water the diffusion of MeCN molecules, in the water-MeCN binary mixture, shows no such drastic reduction. Here, the diffusion values remain constant between 2.92 to 3.98, overall the concentration range.

The viscosity variation of water-DMSO and water-MeCN binary mixture (**Figure S1 E, F** respectively), apart from showing a linear decrease, as in experimental values [8, 9], is showing a zig-zag pattern with long error bars. This high error may be due to either the deficiencies of the parameters, or the improper equilibration of the system. The other possible reason for this absurd behavior of the viscosity variation may be the slow convergence of the formula, because of the fluctuation of the pressure and volume in the simulation box. Though the pressure of the system was kept constant in the NPT ensemble, still remarkable fluctuations occur throughout the simulation, which leads to the asymptotic behaviour of the pressure autocorrelation [10].

The calculated values of various physico-chemical parameters of the binary mixtures except perhaps for viscosity are in reasonable agreement with the experimental results, hence simulations were considered accurate enough for further analysis **(Figure S1).**
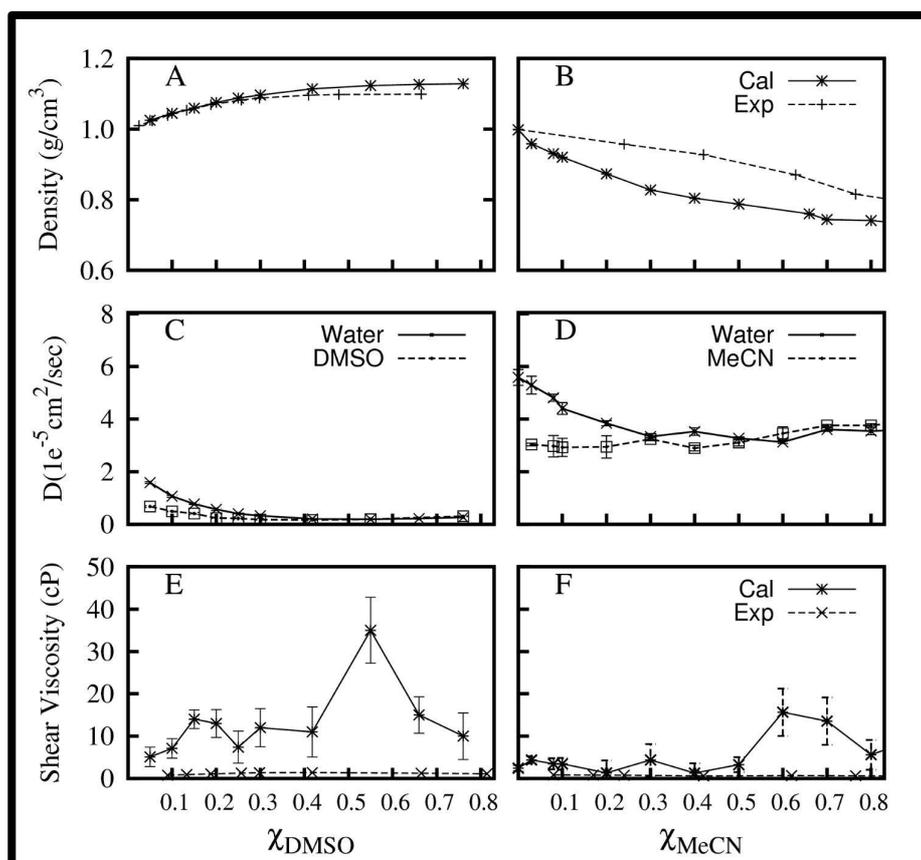

Figure S1: Polar organic solvent concentration dependent variation of densities (A: DMSO, B: MeCN), Self Diffusion Coefficient (C: DMSO, D: MeCN) and Viscosity (E: DMSO, F: MeCN). In case of water-DMSO binary mixtures the experimental density values, for density and viscosity were taken from ref **[11],** while for water-MeCN experimental values were obtained from ref **[7].**
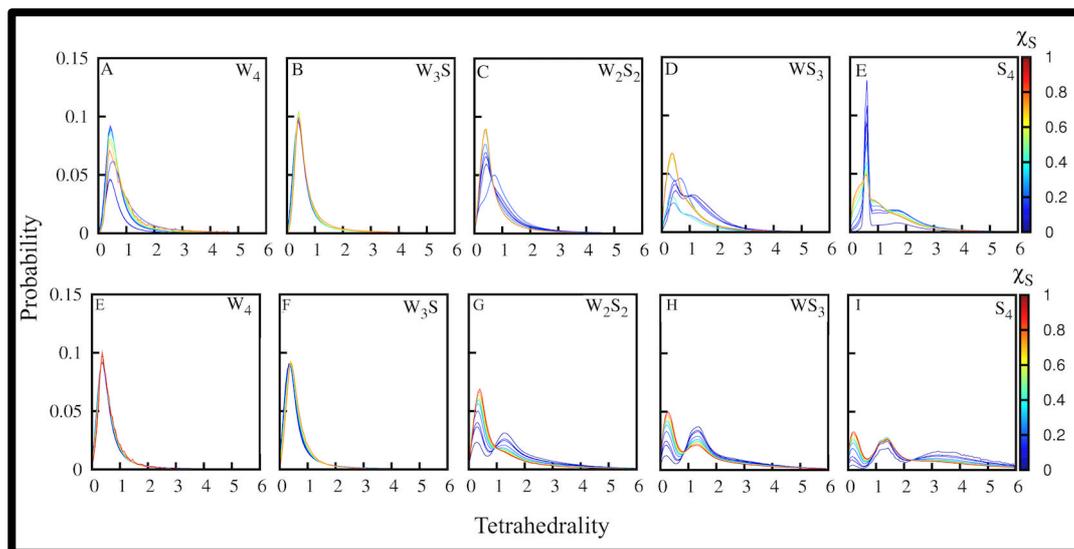
BIOMEDICAL INFORMATICS

©2018

Figure S2: Tetrahedrality variation of all types of tetrahedron in the increasing concentration of water-DMSO (Panels A to E) and water-MeCN (Panels E to I). The color bar indicates the concentration (χS).

References:
[1] Martínez L *et al.* J Comput Chem. 2009, **30:**2157. [PMID: 19229944]
[2] Jorgensen WL *et al.* J Chem Phys. 1983, **79:**926.
[3] Wang J *et al.* Journal of Computational Chemistry. 2004, **25:**1157. [PMID: 15116359]
[4] Luzar A & Chandler D. J Chem Phys. 1993, **98:**8160.
[5] Grabuleda X *et al.* J Comput Chem. 2000, **21:**901.
[6] Darden T *et al.* J Chem Phys. 1993, **98:**10089.
[7] Cunningham GP *et al.* J Chem Eng Data, 1967, **12:**336.
[8] Gabrielian L & Markarian S. J Mol Liq. 2004, **112:**137.
[9] Tsierkezos NG *et al.* J Chem Eng Data. 2000, **45:**395.
[10] Hess B. J Chem Phys. 2002, **116:**209.
[11] Cowie J & Toporowski P. Can J Chem. 1961, **39:**2240.

Journal

Bioinformation 14(7): 350-356 (2018)