



Cite this: *Phys. Chem. Chem. Phys.*,  
2019, 21, 9225

# COSMOplex: self-consistent simulation of self-organizing inhomogeneous systems based on COSMO-RS†

Andreas Klamt,<sup>a</sup> Johannes Schwöbel,<sup>a</sup> Uwe Huniar,<sup>a</sup> Larissa Koch,<sup>a</sup> Selman Terzi<sup>a</sup> and Théophile Gaudin<sup>a,c</sup>

During the past 20 years, the efficient combination of quantum chemical calculations with statistical thermodynamics by the COSMO-RS method has become an important alternative to force-field based simulations for the accurate prediction of free energies of molecules in liquid systems. While it was originally restricted to homogeneous liquids, it later has been extended to the prediction of the free energy of molecules in inhomogeneous systems such as micelles, biomembranes, or liquid interfaces, but these calculations were based on external input about the structure of the inhomogeneous system. Here we report the rigorous extension of COSMO-RS to a self-consistent prediction of the structure and the free energies of molecules in self-organizing inhomogeneous systems. This extends the application range to many new areas, such as the prediction of micellar structures and critical micelle concentrations, finite loading effects in micelles and biomembranes, the free energies and structure of liquid interfaces, microemulsions, and many more related topics, which often are of great practical importance.

Received 27th February 2019,  
Accepted 1st April 2019

DOI: 10.1039/c9cp01169b

rsc.li/pccp

## Introduction

The formation and properties of self-organizing systems are of tremendous importance in the chemical industry (e.g., enhanced oil recovery), the pharmaceutical industry and biochemistry (e.g., drug delivery formulations, biomembrane properties), and in daily life (e.g. components of washing powders, cosmetic ingredients). The interactions taking place are complex in nature and difficult to study experimentally and to describe mathematically due to the delicate balance of weak attractions in the lipophilic domain of surfactant molecules and much stronger interactions between the hydrophilic domain of surfactant molecules and the polar solvent and within the solvent environment.<sup>1,2</sup>

Several theoretical approaches are used to handle self-organizing systems. Due to the wide application area, the following selection of methods and references is by no means intended to be complete.

The simplest theoretical approaches are quantitative structure–property relationships (QSPR) between molecular structures and specific properties of interest,<sup>3</sup> e.g. for critical micelle concentrations (CMC).<sup>4,5</sup> Group contribution methods were used for the prediction

of micellar and microemulsion properties, for example applying the universal functional activity coefficient model (UNIFAC).<sup>6,7</sup> Another interesting thermodynamic approach is the integrated free energy model (IFEM) for microemulsions.<sup>1</sup> Here, the Gibbs free energy is split up into a balance of different lipophilic and hydrophilic energetic contributions. Thermodynamic activity coefficient and equation of state models were adapted to cover surfactant systems, such as the non-random two-liquid model (NRTL) or the statistical associating fluid theory (SAFT).<sup>8,9</sup> Dissipative particle dynamics (DPD) simulations provide mechanistic insight into micellar systems and their phase behaviour.<sup>10,11</sup> All these methods have in common that they are based on fitted interaction parameters or require experimental data.

Molecular dynamics (MD) simulations are widely used and considered as predictive to study micellar and biomembrane systems.<sup>12–15</sup> On the one hand, they operate mechanistically on an atomistic scale. On the other hand they are computationally demanding and, thus, they have not found extensive use in microemulsion modelling.<sup>1</sup>

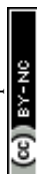
The conductor-like screening model for realistic solvation<sup>16–18</sup> (COSMO-RS) is a predictive, generally applicable and efficient computational model to handle fluid phase properties. COSMO-RS is a combination of the dielectric continuum solvation model COSMO<sup>19</sup> with an efficient statistical thermodynamic model of pairwise molecular surface interactions, known as the quasi-chemical approximation<sup>20</sup> or COSMOSPACE.<sup>21</sup> For the quantification of the surface interactions it uses the surface polarization

<sup>a</sup> COSMOlogic GmbH & Co KG, Imbacher Weg 46, D-51379 Leverkusen, Germany.  
E-mail: andreas.KLAMT@3ds.com

<sup>b</sup> Institute of Physical and Theoretical Chemistry, Universität Regensburg, Germany

<sup>c</sup> School of Chemistry and Chemical Engineering, Nanjing University, China

† Electronic supplementary information (ESI) available: A description of the AUTOBOX algorithm used for the preparation of COSMOplex input distributions and experimental reference values for different systems. See DOI: 10.1039/c9cp01169b



charge densities  $\sigma$  of each solute arising from quantum chemical COSMO calculations.

In several blind prediction challenges and benchmark studies COSMO-RS has been proven to be one of the most accurate tools for the prediction of the free energies of molecules in solution,<sup>22–27</sup> and thus for all equilibrium distribution properties, such as partition coefficients, solubilities, vapour pressures and related properties.

Several combinations of COSMO-RS and the methods mentioned above have been used already towards predictions of inhomogeneous systems: first, critical micelle concentrations or micelle: water partition coefficients in combination with MD and DPD simulations;<sup>13,28,41–43</sup> second, interfacial tensions, with COSMO-RS calculations fitted to experimental data or combined with DPD simulations;<sup>29,30</sup> and third, equivalent alkane carbon numbers (EACN) of microemulsion systems, with a QSPR model using COSMO-RS parameters for non-polar oil/surfactant/water systems.<sup>31,32</sup>

One of the limitations of COSMO-RS was its inability to handle inhomogeneous, structured liquid systems directly on a mechanistic basis. This was partially overcome by the introduction of the COSMOmic approach,<sup>33</sup> which uses external information about the structure of a micelle or biomembrane, usually taken from MD simulations, in order to represent the micelle as a layered liquid of varying composition with respect to the COSMO polarization charge densities  $\sigma$ . COSMOmic then calculates the free energies of solutes in such a layered liquid system by sampling over all relevant conformations, positions and orientations of the solute in this system, resulting in reliable predictions of micelle or membrane to water partition coefficients and free energy profiles of solutes in such systems.

While it is based on MD simulations as an initial step, COSMOmic has been demonstrated to yield at least comparably reliable results for the free energies of neutral solutes in micellar systems to those achievable with MD simulations, at about 0.01 percent of the computational cost.<sup>34–36</sup> More recently, COSMOmic has been extended to ionic solutes,<sup>37,38</sup> now allowing for the prediction of the biomembrane to water partition coefficients of neutral and ionic species with an accuracy of about  $0.7 \log_{10}$  units.

While the COSMOmic approach has been proven to be efficient and reliable with respect to the prediction of free energies of solutes at infinite dilution in micellar systems,<sup>41</sup> many applications of eminent practical importance could not be handled by the COSMO-RS model so far, such as the direct prediction of the critical micelle concentration (CMC) of new surfactants and surfactant mixtures,<sup>13</sup> or the effects of finite concentration enrichment of solutes in micelles, as well as the predictive simulation of liquid–liquid interfaces and microemulsions. As a COSMOmic calculation ends up with a detailed prediction for the probability distribution of each solute with respect to position, orientation and conformational population, it was tempting since the beginnings of COSMOmic to calculate the distribution of the constituents of a micellar system, *i.e.* usually surfactants and water, in the same way and thus to yield a self-consistent prediction of the structure of the inhomogeneous system. But the missing point was the inability of COSMOmic to

take into account the hard-core repulsion of molecules, *i.e.* COSMOmic would predict completely unrealistic, high concentrations of atoms in some layers of the simulation box, leaving other layers rather empty.

In this paper, we describe a fundamental way to overcome this problem by introducing a system pressure in each of the layers, similar to molecular dynamics or dissipative particle dynamics simulations. This system pressure is the response of the system to over-occupation, *i.e.* it is the result of the Pauli or Lennard-Jones repulsion of the molecules in the self-organizing system. We simultaneously converge a pressure function  $P(z)$  together with the COSMOmic distribution probabilities of the molecules in the system, leading to a self-consistent prediction of the structures and free energies of molecules in self-organizing systems. Since this method extends the COSMOmic approach beyond micellar systems to a much wider range of complex, self-organizing systems, we will label this new method COSMOplex, further on.

The theory as well as first application examples of COSMOplex for the prediction of properties of biomembranes, micelles, microemulsions, and liquid–liquid interfaces is described in this paper.

## Material and methods

In this section, after a short introduction of the concepts of COSMO, COSMO-RS and COSMOmic, the theory of COSMOplex is presented.

### COSMO

The conductor-like screening model (COSMO)<sup>19</sup> is a dielectric continuum solvation model using the limit of an ideal conductor, *i.e.* with an infinite dielectric constant. The result of such a quantum chemical COSMO calculation contains the self-consistently calculated energy of the molecule in the presence of a perfect conductor, embedding the molecule on the COSMO surface, together with individual COSMO surface segments and their positions, segment areas and conductor screening charge densities on the particular segments.

Each chemical compound is represented by an ensemble of the relevant conformations of the compound, *i.e.* of its relevant local geometry minima. Each of these conformations has to undergo a separate quantum chemical COSMO calculation, generating the surface polarization charge information and the energy  $E_{\text{COSMO}}^{\text{xc}}$  of conformation  $c$  of compound X. The expression “molecule” is used here for each distinct conformation, whereas the expression “compound” is used for all considered conformations of a chemical entity. The approach used to determine the relative weights of the different conformations is described in the COSMOmic subsection.

### COSMO-RS

The conductor-like screening model for realistic solvation (COSMO-RS)<sup>16–18</sup> starts from quantum chemical COSMO calculations for all involved molecules in the system. COSMO-RS



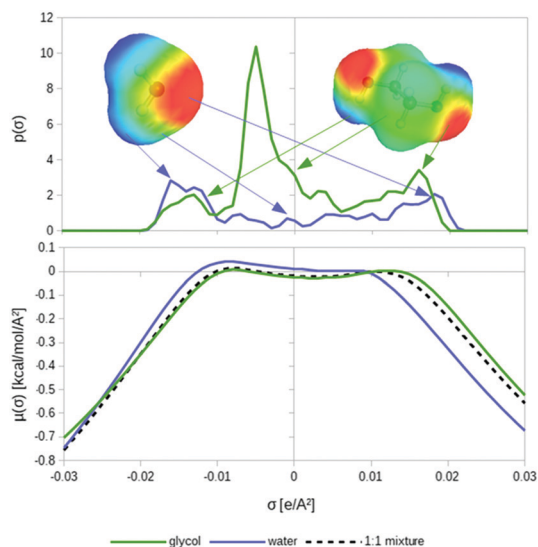
takes the COSMO surface segments and utilizes an efficient statistical thermodynamics algorithm called COSMOSPACE,<sup>21</sup> which is equivalent to an exact solution of Guggenheim's quasi-chemical approximation.<sup>20</sup> This results in a kind of thermodynamic continuum response function, the so-called  $\sigma$ -potential, comprising the free energies of the different types of surface segments in the ensemble of pairwise interacting surface segments. From the sum of the segment free energies, COSMOSPACE directly yields free energies or chemical potentials of the compounds in a pure or mixed liquid. A schematic illustration of the COSMO-RS method is shown in Fig. 1.

The interaction energies of the surface segments are based on the local properties of the two surface segments involved in the contact. In the initial one-descriptor version of COSMO-RS, only the most important of these properties is considered, *i.e.* the conductor polarization charge density  $\sigma$ . This is an excellent descriptor for the local polarity of the molecular surface at the position of the segment. The short range electrostatic interactions, or more precisely the local deviation of the electrostatic interaction of the two segments compared with the reference state of the conductor-embedded molecules, is

$$\Delta E_{\text{misfit}} \cong a_{\text{contact}} c_{\text{misfit}} (\sigma + \sigma')^2 \quad (1)$$

Here,  $a_{\text{contact}}$  is the size of the surface contact, and  $c_{\text{misfit}}$  is the misfit energy coefficient, taking into account the average electronic polarizability of the neighbouring molecules.  $\sigma$  and  $\sigma'$  are the polarization charge densities of the two interacting segments, respectively. The next most important contribution to the interactions is hydrogen bonding, expressed as

$$\Delta E_{\text{hb}} \cong a_{\text{contact}} c_{\text{hb}}(T) \min(0, \sigma\sigma' - \sigma_{\text{hb}}^2) \quad (2)$$



**Fig. 1** Schematic illustration of the COSMO-RS principles: first the COSMO surface polarization charge densities of the surface segments are converted into a histogram ( $\sigma$ -profile, upper diagram). Then the solvent  $\sigma$ -profile is converted into a  $\sigma$ -potential (lower diagram), and the chemical potential of each molecule in this solvent is calculated by applying the solvent  $\sigma$ -potential to the surface segments of the solute.

where  $\sigma_{\text{hb}}$  is a kind of minimum polarity threshold for hydrogen bond formation. The temperature dependence of the hydrogen bond energy coefficient  $c_{\text{hb}}(T)$  expresses the entropy loss associated with hydrogen bonding. While this hydrogen bond expression initially was a heuristic assumption, its functional form meanwhile has been excellently confirmed by comparisons with quantum chemical calculations of hydrogen bond enthalpies<sup>39</sup> and with hydrogen bond enthalpies derived from FTIR-experiments.<sup>40</sup> Some newer versions of COSMO-RS use a slightly more refined functional form for the hydrogen interactions, involving element specific parameters.

Each molecule X is described by its surface composition histogram with respect to  $\sigma$ , the so-called  $\sigma$ -profile  $p^X(\sigma)$ , and a pure or mixed liquid system is characterized by its solvent  $\sigma$ -profile

$$p(\sigma) = \sum_k \frac{x_k p^{X_k}(\sigma)}{x_k A^{X_k}} \quad (3)$$

where  $x_k$  is the mole fraction of the  $k$ th molecule  $X_k$  and  $A_k$  is its COSMO surface area. Solving the COSMOSPACE equations, turning into

$$\mu_S(\sigma) = -\frac{kT}{a_{\text{eff}}} \ln \left[ \int d\sigma' p_S(\sigma') \exp \left\{ \frac{-a_{\text{eff}}(e_{\text{int}}(\sigma, \sigma') + \mu_S(\sigma'))}{kT} \right\} \right] \quad (4)$$

we obtain the  $\sigma$ -potential  $\mu_S(\sigma)$ . This expresses the attraction of a solvent S to surface segments of polarity  $\sigma$ . The  $\sigma$ -potential is a kind of thermodynamic continuum response function of the liquid with respect to the surface polarity of the molecules. Integrating the  $\sigma$ -potential over the surface of a solute X leads to the residual free energy of the solute X in the solvent, which has to be supplemented by a combinatorial contribution

$$\mu_S^X = \mu_{S,\text{res}}^X + \mu_{S,\text{combi}}^X = \int d\sigma p_S(\sigma) \mu_S(\sigma) + kT \ln \gamma_{S,\text{combi}}^X(A^X, V^X, S) \quad (5)$$

where  $\gamma_{S,\text{combi}}^X(A^X, V^X, S)$  can be any typically used expression for the combinatorial activity coefficient. These expressions depend on the composition of the liquid, and the surface areas and volumes of the solute and solvent molecules. The solvent dependence is usually described by the average molecular volume  $\bar{v}_s$  and average molecular surface area  $\bar{q}_s$ . Volumes and surface areas are taken from the COSMO cavities of the molecules.  $\mu_S^X$  is the pseudo-chemical potential, expressing the chemical potential (at a reference mole fraction of 1 mol mol<sup>-1</sup>).

It should be noted that in general multiple conformations of a molecule are considered. Knowing the internal energy of each conformation from the initial quantum chemical COSMO calculation and the individual pseudo-chemical potentials from eqn (5), all thermodynamic equilibrium properties of the conformational ensemble are calculated from the conformational partition function.

Meanwhile COSMO-RS has been refined by more advanced variants of eqn (1)–(4), including the influence of other surface descriptors,<sup>17</sup> *e.g.*, local polarizability and element specific hydrogen bonding and dispersion parameters, in order to improve



the accuracy of the surface interaction expressions. But the basic scheme of COSMO-RS is conserved, *i.e.* the interactions are calculated based on the local surface properties of the surface segments, which can be derived from quantum chemical COSMO calculations.

For the sake of clarity, we describe the concepts of COSMOmic and COSMOplex solely in terms of polarization charge density interactions; the results presented in this paper are obtained by inclusion of the more advanced surface properties.<sup>17,18</sup>

### COSMOmic

While the original COSMO-RS was applicable only to homogeneous bulk liquids, it has been extended to inhomogeneous, micellar systems<sup>33,37</sup> by considering such systems as layered liquids, each layer being considered as homogeneous and described by COSMO-RS. An illustration of the COSMOmic method is shown in Fig. 2.

In COSMOmic, information about the individual composition of each layer has to be derived from external information, usually from snapshots of molecular dynamics simulations of a micellar system *M*. The distribution of atoms over the layers of a micelle was converted into an atom probability distribution. Multiplication of the individual atom probabilities by the partial  $\sigma$ -profiles of the corresponding atoms of a COSMO calculation leads to a specific  $\sigma$ -profile for each layer, *i.e.* to a surface polarity distribution  $p_M(\sigma, z)$  depending on the coordinate  $z$  directed along the outward normal vector of the micelle layers. In the case of spherical or cylindrical systems,  $z$  corresponds to the radial direction. Given these layer  $\sigma$ -profiles, we can use the standard COSMO-RS algorithm to calculate the  $\sigma$ -potential  $\mu_M(\sigma, z)$  of each layer.

For the calculation of the chemical potential of a compound *X* at virtual infinite dilution in such a layered liquid system, a systematic sampling of all possible states, *i.e.* of conformations, centre positions and orientations of the solute relative to the  $z$ -direction, is performed. In practice, the middle of each layer is tested as the centre position, and a grid of vectors on the unit

sphere is used for the orientations of the micelle normal relative to the molecule (herein, this grid is defined by a number of  $n_{\text{ori}} = 162$  orientation vectors). Different choices for the exact definition of the molecular centre have been implemented into COSMOmic. We use the geometric surface centre throughout this paper. For a given state, *i.e.* conformation, position and orientation of the compound, abbreviated as a cpo state further on, each of the surface segments  $\nu$  has a coordinate  $z_\nu^{\text{cpo}}$  with respect to the  $z$ -direction of the micellar system. The chemical potential of this surface segment  $\nu$  is then evaluated by linear interpolation of the two adjacent layer- $\sigma$ -potentials  $\mu_M(\sigma_\nu, z)$ . Additional boundary conditions are explained further in Appendix A. Then, in close analogy to eqn (5), all segment chemical potentials of each conformation  $X^c$  of a compound *X* are summed to yield the residual chemical potential of  $X^c$ . It should be noted that an electrostatic membrane potential  $\phi(z)$  simply can be taken into account in the energy of each surface segment by the multiplication of the polarization charge density  $\sigma$  by the membrane potential and the segment area.<sup>37</sup>

The combinatorial contribution is estimated based on appropriate averages of the surface areas and volumes of the molecules contributing to each layer and from the volume contribution of the solute in each layer. Combining the residual and the combinatorial chemical potentials with the inner energy of the conformation resulting from the quantum chemical COSMO calculation yields the total free energy of the compound for a particular cpo state.

Sampling of all positions, orientations, and conformations thus allows us to calculate the partition function of compound *X* in the micellar system *M*, and from that we can derive all thermodynamic equilibrium properties, including the partition coefficients of the solute between the solvent, usually water, and the micellar system. Because most experimental data is available for these systems, COSMOmic has been validated to a large degree based on biomembrane–water partition coefficients,<sup>33,36,37</sup> which are of interest for biochemistry and pharmaceutical research. But also for other micelle–water partition systems COSMO-RS yields reliable partition coefficients.<sup>15,24–26</sup> No special parameterization of COSMOmic was required beyond the underlying COSMO-RS parameterization, except for the inclusion of ions, for which the membrane potential needed to be derived from fitting of an appropriate functional form with three adjustable parameters.<sup>37</sup>

### COSMOplex: simulation of complex, self-organizing systems based on COSMO-RS

The terminology introduced in the previous chapters sets the framework for COSMOplex. The general workflow is outlined in Fig. 3, and we refer to this figure and the steps indicated therein throughout the explanation of the method.

**Step 1: physical conditions.** First we have to select the overall geometry and the structure of our simulation box. For the self-organizing system (SOS) we consider a box of total volume  $V_{\text{geo}}^{\text{tot}}$  virtually cut into  $L$  layers along the  $z$ -axis. For simplicity we assume that all layers have the same thickness  $\delta$ , although varying thicknesses could be handled as well. By default, we shall assume a rectangular geometry, but cylindrical or spherical boxes

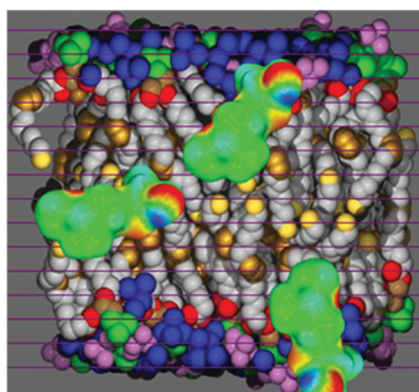


Fig. 2 Schematic illustration of COSMOmic: a micelle is considered as a layered liquid. The composition of each liquid is taken from MD simulations. After the calculation of the  $\sigma$ -profiles and  $\sigma$ -potentials of the layers, all positions and orientations of a solute molecule are sampled, calculating the state specific free energy from the local  $\sigma$ -potentials. (Biomembrane snapshot taken from ref. 59).





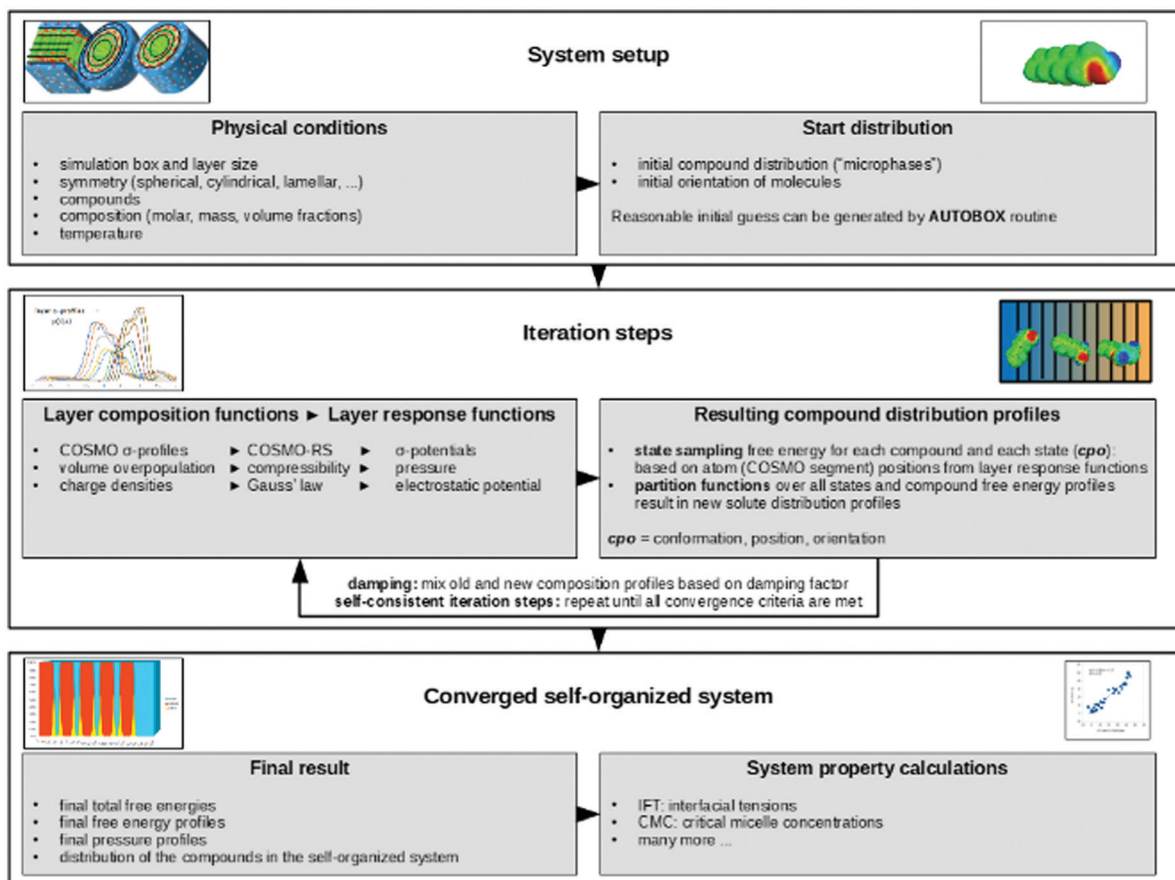


Fig. 3 Workflow of the COSMOplex method.

are supported as well. In the latter cases the  $z$ -axis is the radial axis. Thus, the geometric volume of layer  $i$  is

$$V_{\text{geo}}(z_i) = A\delta \quad (6a)$$

for the case of a rectangular box with cross-section area  $A$ ,

$$V_{\text{geo}}(z_i) = \pi H \left[ \left( z_i + \frac{\delta}{2} \right)^2 - \left( z_i - \frac{\delta}{2} \right)^2 \right] = 2\pi H z_i \delta \quad (6b)$$

for a cylindrical box of height  $H$ , and

$$V_{\text{geo}}(z_i) = \frac{4\pi}{3} \left[ \left( z_i + \frac{\delta}{2} \right)^3 - \left( z_i - \frac{\delta}{2} \right)^3 \right] = 4\pi \left( z_i^2 \delta + \frac{\delta^3}{12} \right) \quad (6c)$$

for a spherical system. For the boundary condition of the boxes see Appendix A.

At a particular temperature  $T$ , we consider the box to be filled with  $N_{\text{tot}} = \sum N_k$  molecules of  $n_c$  different compounds. Each compound  $X_k$  may be represented by  $m_k$  conformations.

In our workflow, each conformation  $X_k^J$  (with a conformer index  $J$ ) is represented by a particular COSMO file, which provides all information about its internal energy  $E_k^J$  in the COSMO reference state, the elements and coordinates of the atoms, and the coordinates, sizes, and screening charges of all segments on the COSMO cavity. The sum of all segment areas is considered as the COSMO surface area  $A_k^J$ . The enclosed volume

of the COSMO surface is  $V_{k,\text{COSMO}}^J$ . The total volume is split into atom volumes  $V^\alpha$  based on a relative closest distance criterion, so that we have atom volumes available for each atom  $\alpha$ . More detail on atom volumes and atom centres is given in Appendix B.

**Step 2: start distribution.** The molecules selected above need to be initially distributed and oriented in the virtual simulation box. This initial guess is basically a free choice of the user. It is advisable to precondition the system towards the expected self-organization, for example to place the more polar molecules with higher concentration on one side and the less polar molecules more on the other side. Furthermore, it is advisable to aim for a reasonably homogeneous space filling in all layers. An empirical tool named AUTOBOX is provided in the COSMOplex software to achieve a reasonable start distribution, box and layer sizes. The start distribution is represented as an atom probability distribution, as it was used as an input for COSMOmic. More details of the AUTOBOX algorithm are given in the ESI.† Nevertheless, it should be emphasized that, as in almost all non-linear optimization methods, and as in MD simulations, the details of the start distribution have an influence on the required number of steps and on the achieved simulation minimum to which COSMOplex converges, but otherwise the properties of a converged COSMOplex simulation should be independent of the start distribution.



**Step 3: layer composition functions.** Given a start distribution of the molecules and their atoms throughout the layered system, we can calculate three layer composition functions.

First, we assign the partial  $\sigma$ -profiles of the atoms to the respective layers and generate layer  $\sigma$ -profiles  $p_{\text{sos}}^0(\sigma, z_i)$ , in the same way as it was done in COSMOmic. The index zero indicates that this is just the initial guess.

As a second composition function we calculate the expectation values  $Q_{\text{sos}}^0(z_i)$  of the layer charge for all layers by summing the product of the polarization charge density and the segment area for all segments in a layer.

As a new ingredient of COSMOplex vs. COSMOmic, the space filling ratio

$$\tilde{\rho}^0(z_i) \equiv \frac{V_{\text{occ}}^0(z_i)}{V_{\text{geo}}(z_i)} \quad (7)$$

*i.e.* from the ratio of the occupied volume and the available geometric volume, is considered as a third composition function. Space filling ratios larger than one indicate an overpopulation of the respective layer; space filling ratios lower than one indicate underpopulation.

The average molecular surface area and the average molecular volume are also calculated in each layer, see Appendix C.

**Step 4: layer response functions.** Next we calculate the three continuum response functions corresponding to the three composition functions calculated in step 3. The response functions for the layer  $\sigma$ -profiles are the corresponding  $\sigma$ -potentials  $\mu_{\text{sos}}^0(\sigma, z_i)$ . The response function for the charge distribution is the electrostatic potential  $\varphi(z_i)$ , which can be calculated from the charge distribution according to Gauss's law. For details see Appendix D. The response function for the space filling ratio (eqn (7)) is the layer pressure  $P(z_i)$ . It can be calculated from the compressibility. For details see Appendix E.

The combinatorial contribution  $\mu_{\text{combi}}$ , a kind of size correction, may be considered as an additional response function. Since its contribution is small we refer the reader to Appendix F for details.

**Step 5: resulting compound distribution profiles.** As a next step we scan all possible states of each compound X in the simulation box. As explained in the COSMOmic section above, the states are enumerated by the conformation index (c), the centre position (p) and by the orientation (o) of the z-axis of the system relative to the coordinate system of the molecule in the COSMO file; abbreviated as a cpo state in the following sections. For the latter a homogeneous grid of vectors on the unit sphere as introduced in the original COSMO paper<sup>19</sup> is used.

For each cpo state, the total free energy of the state ( $\mu_{\text{sos}}^{\text{Xcpo}}$ ) is initialized with the conformational energy of the compound X ( $E_{\text{COSMO}}^{\text{Xc}}$ ). For the COSMO-RS part of the free energy ( $\mu_{\text{sos, res}}^{\text{Xcpo}}$ ) we determine the position of each surface segment and of each atom of a particular compound. Using a linear interpolation between the two neighbouring layers of each segment, we determine the value of the  $\sigma$ -potential of the segment at its position and the corresponding electrostatic potential  $\varphi$  and the respective contributions of each segment ( $E_{\text{sos, \varphi}}^{\text{Xcpo}}$ ) to the

total free energy of the state. In addition, we determine the position of each atom. The atom volume is multiplied by the interpolated pressure  $P$  of the two nearest layers and this pressure penalty ( $E_{\text{sos, P}}^{\text{Xcpo}}$ ) is added to the total free energy of the state. Finally, the combinatorial free energy ( $\mu_{\text{sos, combi}}^{\text{Xcpo}}$ ) is added, see Appendix F. Hence, the free energy of a particular cpo state is

$$\mu_{\text{sos}}^{\text{Xcpo}} = E_{\text{COSMO}}^{\text{Xc}} + \mu_{\text{sos, res}}^{\text{Xcpo}} + E_{\text{sos, \varphi}}^{\text{Xcpo}} + E_{\text{sos, P}}^{\text{Xcpo}} + \mu_{\text{sos, combi}}^{\text{Xcpo}} \quad (8)$$

Based on these individual free energies of all cpo states, we can build the compound partition functions as a sum of the Boltzmann weights:

$$Z^{\text{X}} = \sum_{\text{cpo}} N_i^{\text{mol}} \exp\left(-\frac{\mu_{\text{sos}}^{\text{Xcpo}}}{kT}\right) \quad (9)$$

A multiplication by the number of molecules  $N_i^{\text{mol}}$  in layer  $i$  is performed (since the combinatorial contribution of COSMO-RS is gauged to a mol mol<sup>-1</sup> reference state). The total free energy of each compound X is calculated as

$$G^{\text{X}} = -kT \ln(Z_k) \quad (10)$$

and the total free energy of the system as

$$G_{\text{tot}} = \sum_{k=1}^{n_c} N_k \mu_{\text{sos}}^{\text{X}} = -RT \sum_{k=1}^{n_c} N_k \ln(Z_k) - E_{\varphi}^{\text{corr}} - E_P^{\text{corr}} \quad (11)$$

where the corrections for the electrostatic energy ( $E_{\varphi}^{\text{corr}}$ ) and pressure energy ( $E_P^{\text{corr}}$ ) are required in order to avoid double counting. For details see Appendix G.

The individual probability to find compound X in state cpo is

$$w_{\text{cpo}}^{\text{X}} = N_i^{\text{mol}} \exp\left(-\frac{\mu_{\text{sos}}^{\text{Xcpo}}}{kT}\right) / Z^{\text{X}} \quad (12)$$

Based on these individual state probabilities, we can construct new layer composition functions, *i.e.* new layer  $\sigma$ -profiles, charge distribution, and layer pressure profile.

**Step 6: damping and self-consistent iteration steps.** We mix the new and the previous composition function by using a damping factor  $\lambda$  as a weight for the old composition and  $1 - \lambda$  as a weight for a new composition. Now we have arrived at three new composition functions as calculated in step 3 before. Thus we can restart the COSMOplex cycle at step 4, *i.e.* with the calculation of new response functions. In that way an iterative, recursive cycle is defined, which has to be repeated until the composition of the system and the chemical free energies of the compounds are sufficiently converged. Damping is required in order to avoid oscillations which tend to occur due to the extreme pressure response on overpopulation of layers. Different damping schemes, which vary the damping factor based on different measures of the degree of oscillation of the past cycles, have been implemented, but these schemes are probably subject to future improvements. Since the final results do not significantly depend on the convergence algorithm, we skip a description of the damping algorithms in this article.



**Final step: converged self-organized system.** With step 6 we have closed the iterative cycle for the recursive calculation of the distribution and free energies of molecules in the self-organizing system. After several hundred or thousand iterations the system converges to a minimum of the total free energy. If all convergence criteria are met, which may depend on the application of interest, the iterative cycle is stopped and final, system dependent property calculations can be performed. For example, the interfacial tension is calculated by using the integral of the converged pressure profile.

As in every non-linear optimization procedure, the achieved minimum can either be the global minimum, or just a local minimum. Starting from different start conditions, or by using different damping schemes, it should be checked whether the global optimum has been found. According to our findings, the systems converge directly to the global minimum in the majority of cases, if they are set up from a reasonable initial distribution of the molecules, which is facilitated by the AUTOBOX routine.

### Directional segment extension (DIRPLEX)

Although the COSMOplex algorithm as described above already yields qualitatively and in many cases also quantitatively plausible results, it is necessary to introduce a further extension of the COSMOplex algorithm which concerns the directionality of the surface segments. Since in a bulk, homogeneous and isotropic solvent no direction is preferred, the orientation of surface segments does not require special attention, although even in such systems two segments need to have opposite orientations in order to form an interacting segment pair. But in COSMOplex we consider an anisotropic liquid, with a specially preferred direction, defined as the  $z$ -direction of the SOS further on. Thus, it may be that some segment types preferably point in the positive  $z$ -direction, and others preferably point in lateral or opposite directions. As an example, we may consider the interface of water with a non-polar solvent, *e.g.* cyclohexane, as illustrated in Fig. 4. If the alkane phase is in the region of low  $z$ -values and water at high  $z$ -values, then the layer accommodating the relatively sharp interface of the liquids will mainly have nonpolar surface segments with a positive  $z$ -normal direction, while the polar water segments will have an excess of negative segment normal directions. Standard COSMO-RS thermodynamics for the segments in such a layer would allow for large numbers of contacts between the polar water surface segments and between the nonpolar alkane segments, but geometry prohibits such contacts. This constraint can be implemented into COSMO-RS by considering the segment normal direction as an additional segment property. For the purpose of COSMOplex it is sufficient to split each segment  $\nu$  into two subsegments in the following way: if the scalar product of the segment normal vector with the  $z$ -direction ( $n_{\nu,z}$ ) is positive, the portion  $n_{\nu,z}$  is added to a  $z^+$ -sub-ensemble of surface segments, and the portion  $(1 - n_{\nu,z})$  is added to the  $z^0$ -sub-ensemble. If  $n_{\nu,z}$  is negative, the portion  $|n_{\nu,z}|$  is added to a  $z^-$ -sub-ensemble of surface segments, and the portion  $(1 + n_{\nu,z})$  is added to the  $z^0$ -sub-ensemble. Then we solve the COSMOSPACE equations with the additional boundary condition that surface segments from the  $z^+$ -sub-ensemble may only interact with surface segments of the  $z^-$ -sub-ensemble and *vice versa*, while segments from the

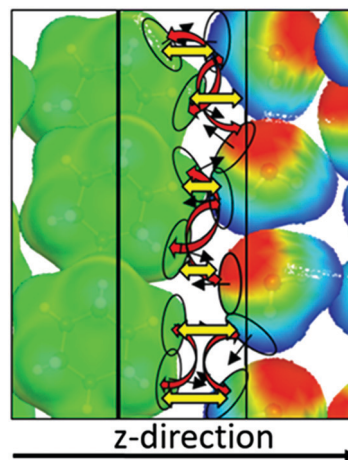


Fig. 4 Schematic illustration of the DIRPLEX extension: without the DIRPLEX constraint, the polar water segments in the middle layer would preferentially interact with each other, and the green alkane segments would form nonpolar pairs (red arrows). The directionality constraint enforces the formation of polar–non-polar contacts (yellow arrows). The black arrows indicate the segment normal vectors.

orthogonal  $z^0$ -sub-ensemble can still freely interact with each other. It should be noted that the COSMOSPACE equations for the coupled  $z^+$ - and  $z^-$ -sub-ensembles do only converge if the amount of segment surface is exactly identical in both sub-ensembles. In order to achieve this in every layer, we solve a system of linear equations yielding the minimum transfer of  $z^+$ - and  $z^-$ -segments from the neighbouring layers which is necessary to yield identical amounts of  $z^+$ - and  $z^-$ -areas in each layer. The acronym DIRPLEX will be used in the following sections for this directional segment extension.

Basically, segment directionality should already have been taken into account in the COSMOmic method, but apparently COSMOmic worked well without this extension. The reason is that in COSMOmic we mostly studied soft, surfactant-rich interfaces, at which apparently the segment pairing is mainly controlled by the segment polarity, and less by directionality. In the results section we compare COSMOplex with and without the DIRPLEX extension for a phospholipid biomembrane and for the cyclohexane–water interface. Larger differences are only found in the latter case. In general, DIRPLEX causes an increase of the computation time by roughly a factor of three.

### Software

Quantum chemical density functional theory (DFT) calculations are performed for each of the molecules, *i.e.* for the solute and solvent molecules under consideration. The COSMOconf software<sup>44</sup> in combination with the quantum chemistry software package TURBOMOLE<sup>45</sup> was used to generate the COSMO files of the relevant conformations for the compounds considered in this paper. The BP86 density functional<sup>46,47</sup> in combination with the TZVP basis set<sup>48</sup> was used for the COSMO calculations. COSMOconf was slightly modified to the needs of COSMOplex, in order to keep geometrically diverse conformations with very similar  $\sigma$ -profiles in the set, while these are filtered out with





standard COSMOconf settings, because they are irrelevant in homogeneous liquids. The BP\_TZVP\_2019 parameterization of COSMOtherm<sup>49</sup> was used for the COSMO-RS calculations within COSMOplex in this article.

## Results and discussion

In this section, we describe a range of different application examples. All applications are considered as proof of principle, and therefore it should not be expected that the optimal performance of COSMOplex has yet been achieved.

### Self-organization of a phospholipid bilayer and phospholipid–water partition coefficients

As a first example, we apply COSMOplex to the self-organization of a biomembrane. The phospholipid 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) was the most important system studied by COSMOmic.<sup>33,37</sup> Four COSMOplex simulations were started from an AUTOBOX generated input: one calculation with the extended DIRPLEX routine, one with an optional self-consistent electrostatic potential switched on, one with both options, and one standard calculation with none of these options.

Fig. 5 shows the volume distribution profiles for DMPC and water, respectively, resulting from the four COSMOplex simulations. The distribution resulting from the MD simulations,<sup>50</sup> as used in COSMOmic, is shown for comparison. As can be seen, all distributions are in reasonable agreement with each other, but the COSMOplex distributions are generally a bit sharper than the distribution derived from the MD simulations. Most likely this is a result of the geometry scatter in MD, *i.e.* the fact that the geometry and symmetry of the DMPC double layer in each snapshot deviate a bit from the idealized flat and reflection symmetry arrangement. This already caused some uncertainty in the position of the reference plane, and some stretching or compression. Within the COSMOplex generated models, the use of the DIRPLEX extension causes some broadening of the profiles, while the use of the self-consistent electrostatic potential has

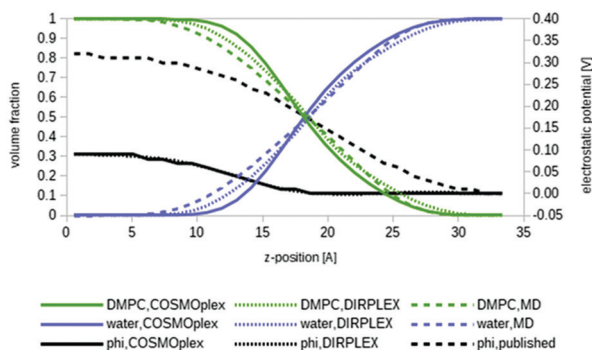


Fig. 5 Volume fractions of the phospholipid DMPC (green lines) and water (blue lines) in a simulated biomembrane system. The centre of the bilayer is at the left end of the diagram. Bare COSMOplex is marked by continuous lines, COSMOplex with DIRPLEX by dotted lines, and the MD composition<sup>50</sup> by dashed lines. The continuous and dotted black lines show the two particular self-consistent electrostatic potentials, the fitted, pseudo-experimental potential<sup>37</sup> is given for reference (dashed black line).

Table 1 Statistical performance of the 5 different DMPC biomembrane models for the prediction of 207 logarithmic DMPC–water partition coefficients.<sup>52</sup> RMSD\* is the root mean squared deviation from the experiment after correction for the mean deviation, the last four columns give the coefficients and performance from linear regression. The column  $\phi$  indicates the used electrostatic potential

Model	$\phi$	Mean dev.	RMSD*	Slope	Shift	$r^2$	RMSD
MD based	Fitted	−0.72	0.86	1.24	−1.56	0.78	0.81
COSMOplex	Off	−0.15	0.90	1.15	−0.61	0.75	0.88
COSMOplex	Self-consistent	−0.21	0.88	1.16	−0.70	0.76	0.86
DIRPLEX	Off	−1.23	0.94	1.05	−1.45	0.71	0.94
DIRPLEX	Self-consistent	−1.24	0.89	1.11	−1.68	0.75	0.88

negligible effect. Fig. 5 also shows the self-consistent electrostatic potentials arising from the two last options, in comparison to the empirically fitted potential, more in detail, fitted to experimental DMPC–water partition coefficients of ions.<sup>37</sup> Given the fact that most reported DMPC MD simulations yield values between 1.0 and 1.3 V in the membrane centre,<sup>37,51</sup> the difference between the COSMOplex result of 0.1 V and the “true” experimental value of 0.3 V must be considered as an encouraging result. Comparing the COSMOplex potentials with the membrane potential generated by the fit to ion partition coefficients, it is apparent that the latter has its strongest increase in the region of the zwitterionic groups ( $z \sim 17$  Å), while the COSMOplex potentials rise at  $z \sim 12$  Å and thus seem to arise from the ester groups, which are located in that region of the membrane. More detailed analysis will be presented in forthcoming papers.

In all cases we also calculated the DMPC–water partition coefficients for the 207 neutral solutes considered by Endo *et al.*<sup>52</sup> The statistics for the different models is given in Table 1. The MD-based COSMOmic model is slightly better than the COSMOplex models. Within the COSMOplex models the DIRPLEX model without electrostatics seems to be slightly worse than the other three models. Nevertheless, all differences in the performance for the partition coefficient models are very small, and we may conclude that COSMOplex-based models are essentially as good as the COSMOmic models, which required CPU intensive MD simulations as an input.

### Prediction of critical micelle concentrations

The prediction of critical micelle concentrations (CMCs) is of special interest in many areas of formulation science.<sup>13</sup> Here we consider the prediction of the CMCs of ten diverse neutral surfactants. For each surfactant a spherical aqueous micelle model was built with AUTOBOX based on the surfactant geometry, and a COSMOplex calculation with a self-consistent electrostatic field was started. As shown in Table 2 and Fig. 6, the trends in the CMC are reasonably well predicted by COSMOplex. Initial simulations for ionic surfactants also converged nicely and showed reasonable results (details not shown). These systems will be studied in more detail in a forthcoming paper.

### Microemulsion systems

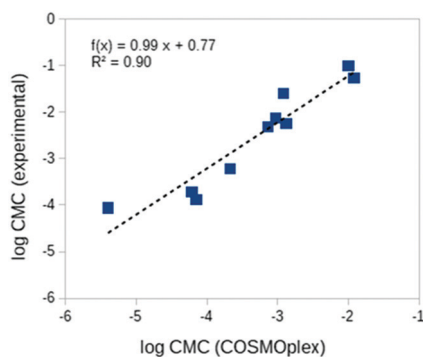
Microemulsions are of special interest for many areas of formulation science and for the important field of enhanced oil





**Table 2** COSMOplex results for CMCs ( $\log_{10}$  CMC in  $[\text{mol l}^{-1}]$  units) of ten diverse non-ionic surfactants

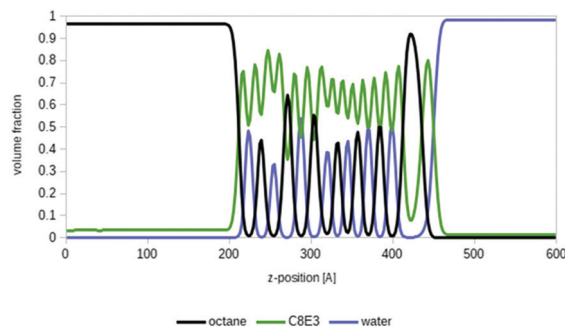
Surfactant	Abbreviation	$\log_{10}$ CMC (COSMOplex)	$\log_{10}$ CMC (experimental) <sup>58</sup>
Triethylene glycol monoheptyl ether	C <sub>6</sub> E <sub>3</sub>	-2.00	-1.00
Dimethyl nonylamine oxide	C <sub>9</sub> C <sub>2</sub> NO	-1.92	-1.27
Octyl $\beta$ -D-glucoside	C <sub>8</sub> BG <sub>1</sub>	-2.92	-1.60
Triethylene glycol monoheptyl ether	C <sub>8</sub> E <sub>3</sub>	-3.03	-2.12
Octyl $\alpha$ -glyceryl ether	C <sub>8</sub> BGLYE	-2.88	-2.24
Octyl glycol ether	C <sub>8</sub> E <sub>1</sub>	-3.14	-2.31
Triethylene glycol monodecyl ether	C <sub>10</sub> E <sub>3</sub>	-3.67	-3.22
Dodecyl $\beta$ -D-glucoside	C <sub>12</sub> BG <sub>1</sub>	-4.21	-3.72
Octoxynol-2	C <sub>8</sub> (C <sub>6</sub> H <sub>4</sub> )E <sub>2</sub>	-4.15	-3.88
Hexaethylene glycol monododecyl ether	C <sub>12</sub> E <sub>6</sub>	-5.40	-4.06



**Fig. 6** COSMOplex results for CMCs ( $\log_{10}$  CMC in  $[\text{mol l}^{-1}]$  units) of ten diverse surfactants.

recovery.<sup>7</sup> Deeper insights into microemulsion systems and the physical-chemical properties involved are provided elsewhere, for example on the definition of Winsor types, fish tail points and fish diagrams, or equivalent alkane carbon numbers.<sup>2,7,53</sup> In short, a typical microemulsion system is composed of a less polar compound (usually called oil), a surfactant and water. After shaking this mixture in a test tube, three phases show up: an oil-rich phase at the top, an aqueous phase at the bottom, and a micellar, surfactant-rich phase in the middle.

As a proof of principle, we applied COSMOplex to a typical microemulsion system, *i.e.* *n*-octane, a surfactant C<sub>8</sub>E<sub>3</sub>, and water at 288 K. We initialized the system as a three-zone system of oil (35 vol%), surfactant (30 vol%) and water (35 vol%). Experimentally, this system is expected to form an oil-rich phase, a microemulsion phase and a water rich phase under these conditions.<sup>54</sup> And indeed, after initial fluctuations it really formed three phases with surfactant saturated oil and water phases, and a middle region in which the oil, surfactant and water self-organize into multiple micellar layers (see Fig. 7). This means that with COSMOplex – without any special pre-conditioning – we were able to investigate the self-organization of such complicated systems *in silico* with simulation times in the range of a few hours on a single CPU.



**Fig. 7** Spontaneous formation of a microemulsion phase in an oil (octane, 35 weight%, black line), surfactant (C<sub>8</sub>E<sub>3</sub>, 30 weight%, green line) and water (35%-weight, blue line) system: the lines show the position dependent particular volume fractions in a virtual test tube, simulated by COSMOplex.

Even though the quantitative agreement with experimental microemulsion phase diagrams is not yet perfect, COSMOplex may become an important tool for the systematic investigation of microemulsion formation. In forthcoming publications, the prediction of more sophisticated fish diagrams<sup>53</sup> by COSMOplex will be described in detail. These are calculated by temperature and concentration grids at the areas of phase transitions between a single phase, two phase systems and three phase systems.

It should be noted that curvatures of micellar systems are not directly implemented in COSMOplex, beyond the three base geometries (sphere, cylinder and disc). System properties of a curvature could be possibly derived by derivatives of the micelle radius and accounting for the varying micelle:water interface areas and corresponding free energies. As an alternative, a linear combination of properties derived from different geometries could be used, where the pre-factors of this linear combination could be related to a comparison of the particular geometry dependent free energies. Also, the critical packing parameter could be a useful guide in the initial choice of the micellar shape.<sup>55</sup> This is, however, a matter for future investigations.

### Interfacial tensions

Another promising application area of COSMOplex is the prediction of the self-organization of molecules at a liquid-liquid interface and of the resulting interfacial tension (IFT). As introduced in the context of MD simulations,<sup>56</sup> the IFT can be simply calculated as the *z*-integral of the lateral pressure.

The initialization of a rectangular simulation box with mirror boundary conditions is performed by the AUTOBOX routine. In Fig. 8 three typical interfacial pressure profiles are shown. The profiles have been base-line corrected. In Fig. 9 we show the results for the IFT prediction of 40 aqueous systems taken from the paper of Andersson *et al.*<sup>29</sup> Since negative predictions must result from numerical noise, we replaced them by zero. The results, which take into account the segment directionality (DIRPLEX), are in good agreement with experimental interfacial tensions (see Fig. 9). It should be noted that no parameters have been adjusted to fit the COSMOplex results to the experimental data. As explained in the DIRPLEX section, results without the DIRPLEX extension underestimate



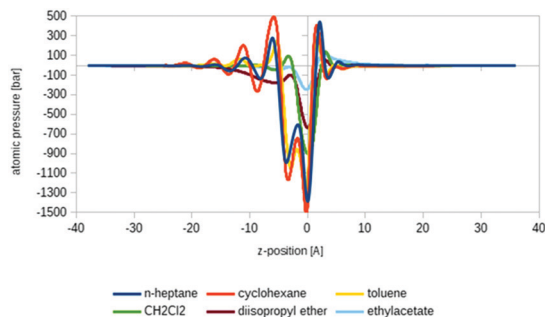


Fig. 8 COSMOplex atomic pressure profiles for six interfacial tension calculations. The pressure drops in the interfacial region. For cyclohexane and toluene pressure fluctuations arising from the first, second and third molecular layer can be observed. The pressure profiles are base-line corrected.

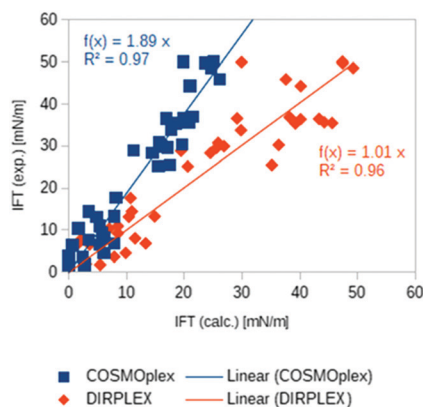


Fig. 9 Experimental<sup>29</sup> and predicted interfacial tensions for 40 aqueous systems: COSMOplex predictions without DIRPLEX extensions are in blue, with DIRPLEX in red. For the compound list and data see Table S1 in the ESI.†

the experimental interfacial tensions by a factor of almost two. The use of the self-consistent electrostatic potential has almost no influence on the interfacial tension prediction.

## Conclusions and outlook

The COSMO-RS method was originally developed for homogenous, bulk liquids. The COSMOmic method extended the application of COSMO-RS to the simulation of solutes in micellar systems. However, it was limited to infinite dilution and required computationally expensive molecular dynamics input to describe the structure of the system. By the introduction of a layer specific pressure, the applicability of the method – now named COSMOplex – has been extended to the self-consistent simulation of a wide range of self-organizing molecular systems.

First applications to micelle formation, microemulsion systems, biomembrane–water partitioning, and liquid–liquid interfacial tension predictions yield promising results in good agreement with experimental data. All simulations only took between 10 minutes and 24 hours on a single laptop CPU. Thus, the simulations are three to four orders of magnitude less CPU demanding than comparable molecular dynamics simulations.

No adjustment of empirical parameters was required. All interactions are directly derived from the surface segment interactions by the COSMO-RS method, supplemented by parameter-free expressions for the pressure and the long-range electrostatic interactions.

Overall, COSMOplex enables many simulation applications of self-organizing systems at significantly lower computational costs and – at the same time – potentially achieving better agreement with experiments due to its sound quantum chemical and statistical thermodynamic fundamentals. These potential applications include – to mention just a few examples – the direct physical prediction of complex so-called fish diagrams for microemulsion systems, microemulsion systems formed by hydrotropes,<sup>60</sup> the temperature or salt concentration dependence of surfactant properties, or even the fast self-consistent generation of bio-membrane systems, for example to investigate the effect of penetration enhancers in human skin.<sup>61</sup>

Furthermore, COSMOplex has still considerable potential for tuning its performance. Even though the method is already capable of treating the electrostatic potential of self-organizing systems self-consistently, the COSMOplex application examples studied in this article are restricted to non-ionic surfactants and neutral solutes. Ionic surfactants and ionic solutes will be considered in a forthcoming paper. Extensions to gas–liquid interfaces and solid–liquid interfaces, confined liquids, and liquid crystals should be possible with minor effort and will be considered soon. By the implementation of logarithmically scaling layer widths, COSMOplex should be extendable to the simulation of macroscopic distances, e.g. for the complete simulation of the volume between the anode and cathode of an electrochemical cell. As a distant future goal, the simulation of 3D-inhomogeneous systems seems to be possible, optimistically enabling a completely new way of simulating water and drug molecules in enzyme pockets.

## Conflicts of interest

The authors declare the following competing financial interests: Andreas Klamt is chief executive officer and all other authors are active or former employees of COSMOlogic GmbH&CoKG. COSMOlogic commercially distributes the software used in this paper.

## Appendix

### Appendix A: boundary conditions in COSMOplex

Currently the following boundary conditions are supported by COSMOplex:

(1) Mirror boundary conditions for rectangular boxes: if a coordinate is outside the box, it is transformed into the box by reflection. This assumes mirror symmetry perpetuation of the system outside the box. By reflection boundary conditions, most systems, which usually are simulated with periodic boundary conditions in MD simulations, can be reduced to half of the simulation volume, saving simulation time and memory.

(2) Pseudo-mirror boundary conditions for cylindrical and spherical simulation boxes: in spherical and cylindrical systems



coordinates cannot escape at the inner boundary. But at the outer boundary a molecule, while centered in one of the outer layers, may have segments and atoms sticking out of the simulated system. In this case we transform them back into the system in the same way as we do with mirror boundaries for rectangular boxes, *i.e.* a coordinate  $z > z_{\max}$  is transformed to  $2 \cdot z_{\max} - z$ . In order to be physically correct this boundary condition requires that the system does not show major changes any more at the outer boundary.

(3) Periodic boundary conditions for rectangular boxes: if the  $z$ -coordinate of an atom or segment exceeds the outer box limit  $z_{\max}$ , it is translated back into the simulation volume by subtraction of one box length, *i.e.* by  $z_{\max} - z_{\min}$ . Analogously, coordinates escaping the box on the other side are translated by addition of a box length.

If not stated otherwise, all COSMOplex calculations in this paper are performed using the first boundary condition for rectangular boxes and the second one for spherical and cylindrical systems. In classic COSMOmic calculations, for negative values of  $z$ , which may appear in lamellar micelle geometries, a mirror boundary condition is applied, and for values of  $z$  larger than the outer boundary of the simulated box the  $\sigma$ -potential is assumed to be the same as in the pure solvent containing the micelle, *i.e.* in most cases pure water.

#### Appendix B: details of atomic volume assignment

The geometric centre position of the associated volume of an atom  $\alpha$  is introduced as  $z_{\text{vol}}^{\alpha}$ . This slightly deviates from the atom position  $z^{\alpha}$  which usually is defined by the position of the nucleus of the atom.

In a next step, we define rescaled volumes for the molecules and atoms. The rescaled molecular volume  $V_k$  of a compound is defined as the liquid volume at room temperature, *i.e.* molecular weight divided by liquid density. The liquid density can either be given as experimental input data, or it can be estimated within the COSMOtherm software.<sup>49</sup> All atom volumes are rescaled by the ratio of  $V_k$  and  $V_{k,\text{COSMO}}$ , where the latter is the thermodynamic average of the COSMO volumes of the conformations according to their Boltzmann population in the pure liquid  $k$ .

Next, we adjust either the size of the simulation box or the number of molecules in the box in order to achieve equality of the box volume and the sum of the molecular volumes of all molecules in the system. It should be noted that fractional numbers for the counts of molecules are allowed in COSMOplex.

#### Appendix C: calculation of the number of molecules per layer and of the average molecular surface area and volume in each layer

The average molecular size parameters, *i.e.* the average molecular volume  $\bar{v}_{\text{sol}}(z_i)$  and average molecular surface area  $\bar{q}_{\text{sol}}(z_i)$ , are calculated as well. These are required for the calculation of the combinatorial free energy term in eqn (5), *i.e.* for the combinatorial response function of the layered liquid continuum. For this purpose, each molecule in each state is assigned to a layer  $i$  according to the amount of surface area residing in layer  $i$  divided by the total surface area of the molecule. Summation over all molecules and their state

probabilities yields the number of molecules per layer  $N_i^{\text{mol}}$ . Multiplying the fraction of surface in layer  $i$  by the molecular volume yields the average volume  $\bar{v}_{\text{sol}}(z_i)$ , and multiplying the fraction by the molecular area yields the average surface area  $\bar{q}_{\text{sol}}(z_i)$ .

#### Appendix D: calculation of the electrostatic potential from layer charges

The electrostatic potential  $\varphi_{\text{sol}}(z_i)$  arising from the net charges in the layers is calculated, taking into account the geometry of the simulation box and the electronic polarizability, *i.e.* the infinite frequency dielectric constant  $\epsilon_{\infty} = n^2$ , of the layers. The latter is calculated from pure compound refractive indices and averaged by using particular atom volume probabilities in the layers. The pure compound refractive indices can either be given as an experimental input, or they are calculated from an increment scheme.

The reduction of the electric field resulting from the re-orientational polarizability should already be taken into account by the self-consistent organization of the molecules, since they feel the electrostatic potential.

In the planar geometry (rectangular box), the change of the electric field  $\Delta E(z_i)$  within a layer  $i$  is given by the volume charge density, *i.e.* the layer charge  $Q_{\text{sol}}(z_i)$  divided by the layer volume, multiplied by the layer width  $\delta$ . Hence, we have for the electric field at the interface between layer  $i$  and  $i - 1$ , *i.e.* at  $z_i - \delta/2$ ,

$$E_i = E\left(z_i - \frac{\delta}{2}\right) = \sum_{j=i}^L \frac{Q_j}{A_j} \quad (\text{D1})$$

The summation is from the outer end downwards, because we assume that the electric field and the potential at the right end of the box ( $i = L$ ) are zero, *e.g.*, in the water phase of a biomembrane bilayer. The electrostatic potential is the integral of the potential in the middle of layer  $i$  (with  $\epsilon_{\infty} = n^2$ ):

$$\varphi(z_i) = \sum_{k=i+1}^L \frac{E_k}{n^2(z_i)} \delta \quad (\text{D2})$$

In the spherical geometry the electric field at a radius  $z$  is the same as the electric field of a point charge at the centre which has the value of all included charge  $Q^{\text{inc}}(z_i)$ . Outer charge does not contribute. Hence, we can easily get the electrostatic potential by summing up the small changes of the electric field  $\Delta E(z_i)$  reduced by  $n^2(z_i)$ , starting at 0 and ending up at the outer shell, at which the field should have dropped to zero, because the total charge in the simulated sphere must be zero.

$$\varphi(z_i, z_{i+1}) = \frac{E_k\left(z_i + \frac{\delta}{2}\right)}{n^2\left(z_i + \frac{\delta}{2}\right)} \delta = \frac{Q^{\text{inc}}(z_i)}{\left(z_i + \frac{\delta}{2}\right)^2 n^2\left(z_i + \frac{\delta}{2}\right)} \delta \quad (\text{D3})$$

$Q^{\text{inc}}(z_i)$  is the sum of all layer charges up to  $z$ . To calculate the potential  $\varphi(z_i)$ , we just have to start at the centre and sum up all changes of the electric field.

In the cylindrical geometry the electric field at a radius  $z$  is the same as the electric field  $E(z)$  of a line charge at the centre





which has the value of all included charge. In close analogy to the spherical case we get (with  $H$  being the height of the simulated cylinder)

$$\varphi(z_i, z_{i+1}) = \frac{E_k \left( z_i + \frac{\delta}{2} \right)}{n^2 \left( z_i + \frac{\delta}{2} \right)} \delta = \frac{Q^{\text{inc}}(z_i)}{2H \left( z_i + \frac{\delta}{2} \right) n^2 \left( z_i + \frac{\delta}{2} \right)} \delta \quad (\text{D4})$$

To calculate the potential  $\varphi(z_i)$  we just have to start at the centre and sum up all changes of the electric field.

### Appendix E: pressure as a response to volume occupation

The new continuum response function considered in COSMOplex is the pressure function  $P(z_i)$ . The pressure response in a layer  $z_i$  is calculated *via* the space filling ratio

$$\tilde{\rho}^1(z_i) \equiv \frac{V_{\text{occ}}^1(z_i)}{V_{\text{geo}}(z_i)} \quad (\text{E1})$$

*i.e.* from the ratio of the occupied volume and the available geometric volume, by the compressibility function. The space filling can be very inhomogeneous, since too many molecules may prefer certain regions of space, which thus get overcrowded, while other regions may be underpopulated. For the calculation of the layer pressures from the space filling ratio we fitted a Lennard-Jones-type compressibility relation to typical pressure-dependent compressibility data of organic liquids<sup>57</sup>

$$P(\tilde{\rho}) = \frac{d}{d\tilde{\rho}} \left( \frac{\tilde{\rho}^9 - 2\tilde{\rho}^{4.5} + 1}{9 \times 4.5 c_{\text{comp}}} \right) = \frac{\tilde{\rho}^8 - \tilde{\rho}^{3.5}}{4.5 c_{\text{comp}}} \quad (\text{E2})$$

in which  $c_{\text{comp}}$  is the compressibility, which we set to  $10^{-6} \text{ kPa}^{-1}$  throughout this article for all compounds except for water, for which we use  $5 \times 10^{-7} \text{ kPa}^{-1}$ . As long as the compressibility function is steep enough, *i.e.* the compressibility is low, this function is like a steep wall and the results do not strongly depend on the details of the compressibility function and on the exact value of the compressibility. Nevertheless, a reasonable and smooth compressibility function is required for the convergence of the COSMOplex algorithm. The local, *i.e.*  $z$ -dependent, compressibility is calculated as a volume based average of the pure liquid compressibilities of the molecules in each layer. Details regarding derivation of the pressure function are provided in the ESI.†

### Appendix F: combinatorial contribution

The combinatorial contribution to the free energy of a molecule  $X$  in layer  $i$  is calculated utilizing the surface area and volume of molecule  $X$  and the average molecular volume and surface area in layer  $i$ , as described in Appendix C. For each state (cpo) of molecules  $X$  the combinatorial contribution is assigned according to the fraction of surface residing in layer  $i$ .

### Appendix G: total free energy corrections for the Coulomb energy and the compression energy

In variational, self-consistent field quantum chemical algorithms, the total energy of the system usually is calculated as a sum of the expectation values of the energies of the electrons. But this has to be corrected by adding half of the Coulomb energy, because

otherwise the electrostatic interaction of each electron is double counted. In other words, the electrostatic energy included in the expectation values has to be replaced by the correct value of the Coulomb energy. The same correction is required in COSMOplex for the Coulomb energy, *i.e.* the total Coulomb energy has to be reduced by a factor 1/2 ( $E_{\varphi}^{\text{corr}}$ ). The correction is also required for the compression energy, because the compression energy included in the partition functions of the molecules reflects the interaction of each atom volume with the final, self-consistent pressure, while the compression related energy ( $E_P^{\text{corr}}$ ) has a different value given by

$$E_P^{\text{corr}} = \sum_{i=1}^L \left[ \left( \frac{\tilde{\rho}_i^9 - 2\tilde{\rho}_i^{4.5} + 1}{9 \times 4.5 \times c_{\text{comp}}} \right) - P(\tilde{\rho}_i) \right] \langle V_i \rangle \quad (\text{G1})$$

where  $\langle V_i \rangle$  is the expectation volume of the molecular (or equivalently the atomic) volume in a layer  $i$ , and the expression in parentheses is the Lennard-Jones like expression of the compression energy, which was used in eqn (8) for the derivation of the pressure formula.

## Acknowledgements

We thank Steven Abbott, Jean-Marie Aubry and Maximilian Hahn for valuable discussions and feedback during the development of COSMOplex.

## References

- 1 A. Boza Troncoso and E. Acosta, Integrated free energy model (IFEM) for microemulsions, *J. Colloid Interface Sci.*, 2016, **466**, 400–412.
- 2 R. Nagarajan and E. Ruckenstein, Molecular theory of microemulsions, *Langmuir*, 2000, **16**, 6400–6415.
- 3 J. Hu, X. Zhang and Z. Wang, A review on progress in QSPR studies for surfactants, *Int. J. Mol. Sci.*, 2010, **11**, 1020–1047.
- 4 P. D. Huibers, V. S. Lobanov, A. R. Katritzky, D. O. Shah and M. Karelson, Prediction of critical micelle concentration using a quantitative structure–property relationship approach. 1. Nonionic surfactants, *Langmuir*, 1996, **12**, 1462–1470.
- 5 P. D. T. Huibers, V. S. Lobanov, A. R. Katritzky, D. O. Shah and M. Karelson, Prediction of critical micelle concentration using a quantitative structure–property relationship approach, *J. Colloid Interface Sci.*, 1997, **187**, 113–120.
- 6 H. Cheng, G. M. Kontogeorgis and E. H. Stenby, Prediction of micelle formation for aqueous polyoxyethylene alcohol solutions with the UNIFAC model, *Ind. Eng. Chem. Res.*, 2002, **41**, 892–898.
- 7 A. Boza Troncoso and E. Acosta, The UNIFAC model and the partition of alkyl and alkylphenol ethoxylate surfactants in the excess phases of middle phase microemulsions, *Fluid Phase Equilib.*, 2015, **397**, 117–125.
- 8 C.-C. Chen, Molecular thermodynamic model for Gibbs energy of mixing of nonionic surfactant solutions, *AIChE J.*, 1996, **42**, 3231–3240.



- 9 X.-S. Li, J.-F. Lu and Y.-G. Li, Study on ionic surfactant solutions by SAFT equation incorporated with MSA, *Fluid Phase Equilib.*, 2000, **168**, 107–123.
- 10 E. Ryjkina, H. Kuhn, H. Rehage, F. Müller and J. Peggau, Molecular dynamic computer simulations of phase behavior of non-ionic surfactants, *Angew. Chem., Int. Ed.*, 2002, **41**, 983–986.
- 11 A. Vishnyakov, M.-T. Lee and A. V. Neimark, Prediction of the critical micelle concentration of nonionic surfactants by dissipative particle dynamics simulations, *J. Phys. Chem. Lett.*, 2013, **4**, 797–802.
- 12 S. A. Sanders and A. Z. Panagiotopoulos, Micellization behavior of coarse grained surfactant models, *J. Chem. Phys.*, 2010, **132**, 114902.
- 13 S. Jakobtorweihen, D. Yordanova and I. Smirnova, Predicting critical micelle concentrations with molecular dynamics simulations and COSMOmic, *Chem. Ing. Tech.*, 2017, **89**, 1288–1296.
- 14 D. Lopes, S. Jakobtorweihen, C. Nunes, B. Sarmiento and S. Reis, Shedding light on the puzzle of drug-membrane interactions: experimental techniques and molecular dynamics simulations, *Prog. Lipid Res.*, 2017, **65**, 24–44.
- 15 M. Lundborg, A. Narangifard, C. L. Wennberg, E. Lindahl, B. Daneholt and L. Norlén, Human skin barrier structure and function analyzed by cryo-EM and molecular dynamics simulation, *J. Struct. Biol.*, 2018, **203**, 149–161.
- 16 A. Klamt, Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
- 17 A. Klamt, V. Jonas, T. Bürger and J. C. Lohrenz, Refinement and parametrization of COSMO-RS, *J. Phys. Chem. A*, 1998, **102**, 5074–5085.
- 18 A. Klamt, *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier, Amsterdam, The Netherlands, Boston, MA, USA, 2005.
- 19 A. Klamt and G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
- 20 E. A. Guggenheim, *Mixtures*, Clarendon Press, Oxford, 1952.
- 21 A. Klamt, G. J. P. Krooshof and R. Taylor, COSMOSPACE: alternative to conventional activity-coefficient models, *AIChE J.*, 2002, **48**, 2332–2349.
- 22 C. C. Bannan, K. H. Burley, M. Chiu, M. R. Shirts, M. K. Gilson and D. L. Mobley, Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge, *J. Comput.-Aided Mol. Des.*, 2016, **30**, 927–944.
- 23 A. Klamt, F. Eckert, J. Reinisch and K. Wichmann, Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set, *J. Comput.-Aided Mol. Des.*, 2016, **30**, 959–967.
- 24 J. Reinisch, A. Klamt and M. Diedenhofen, Prediction of free energies of hydration with COSMO-RS on the SAMPL3 data set, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 669–673.
- 25 J. Zhang, B. Tuguldur and D. van der Spoel, Force field benchmark of organic liquids. 2. Gibbs energy of solvation, *J. Chem. Inf. Model.*, 2015, **55**, 1192–1201.
- 26 J. Zhang, B. Tuguldur and D. van der Spoel, Correction to force field benchmark of organic liquids. 2. Gibbs energy of solvation, *J. Chem. Inf. Model.*, 2016, **56**, 819–820.
- 27 A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco and F. J. Luque, On the performance of continuum solvation methods. A comment on “Universal approaches to solvation modeling”, *Acc. Chem. Res.*, 2009, **42**, 489–492.
- 28 R. Oviedo-Roa, J. M. Martínez-Magadán, A. Muñoz-Colunga, R. Gómez-Balderas, M. Pons-Jiménez and L. S. Zamudio-Rivera, Critical micelle concentration of an ammonium salt through DPD simulations using COSMO-RS-based interaction parameters, *AIChE J.*, 2013, **59**, 4413–4423.
- 29 M. P. Andersson, M. V. Bennetzen, A. Klamt and S. L. S. Stipp, First-principles prediction of liquid/liquid interfacial tension, *J. Chem. Theory Comput.*, 2014, **10**, 3401–3408.
- 30 H. Alasiri and W. G. Chapman, Dissipative particle dynamics (DPD) study of the interfacial tension for alkane/water systems by using COSMO-RS to calculate interaction parameters, *J. Mol. Liq.*, 2017, **246**, 131–139.
- 31 T. Lukowicz, A. Benazzouz, V. Nardello-Rataj and J.-M. Aubry, Rationalization and prediction of the equivalent alkane carbon number (EACN) of polar hydrocarbon oils with COSMO-RS  $\sigma$ -moments, *Langmuir*, 2015, **31**, 11220–11226.
- 32 T. Lukowicz, E. Illous, V. Nardello-Rataj and J.-M. Aubry, Prediction of the equivalent alkane carbon number (EACN) of aprotic polar oils with COSMO-RS sigma-moments, *Colloids Surf., A*, 2018, **536**, 53–59.
- 33 A. Klamt, U. Huniar, S. Spycher and J. Keldenich, COSMOmic: a mechanistic approach to the calculation of membrane-water partition coefficients and internal distributions within membranes and micelles, *J. Phys. Chem. B*, 2008, **112**, 12148–12157.
- 34 M. Palonciová, R. DeVane, B. Murch, K. Berka and M. Otyepka, Amphiphilic drug-like molecules accumulate in a membrane below the head group region, *J. Phys. Chem. B*, 2014, **118**, 1030–1039.
- 35 E. Ritter, D. Yordanova, T. Gerlach, I. Smirnova and S. Jakobtorweihen, Molecular dynamics simulations of various micelles to predict micelle water partition equilibria with COSMOmic: influence of micelle size and structure, *Fluid Phase Equilib.*, 2016, **422**, 43–55.
- 36 S. Jakobtorweihen, A. C. Zuniga, T. Ingram, T. Gerlach, F. J. Keil and I. Smirnova, Predicting solute partitioning in lipid bilayers: free energies and partition coefficients from molecular dynamics simulations and COSMOmic, *J. Chem. Phys.*, 2014, **141**, 045102.
- 37 K. Bittermann, S. Spycher, S. Endo, L. Pohler, U. Huniar, K.-U. Goss and A. Klamt, Prediction of phospholipid-water partition coefficients of ionic organic chemicals using the mechanistic model COSMOmic, *J. Phys. Chem. B*, 2014, **118**, 14833–14842.
- 38 A. Klamt, COSMO-RS for aqueous solvation and interfaces, *Fluid Phase Equilib.*, 2016, **407**, 152–158.
- 39 A. Klamt, J. Reinisch, F. Eckert, A. Hellweg and M. Diedenhofen, Polarization charge densities provide a predictive quantification of hydrogen bond energies, *Phys. Chem. Chem. Phys.*, 2012, **14**, 955–963.



- 40 A. Klamt, J. Reinisch, F. Eckert, J. Graton and J.-Y. Le Questel, Interpretation of experimental hydrogen-bond enthalpies and entropies from COSMO polarisation charge densities, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7147–7154.
- 41 T. Ingram, S. Storm, L. Kloss, T. Mehling, S. Jakobtorweihen and I. Smirnova, Prediction of micelle/water and liposome/water partition coefficients based on molecular dynamics simulations, COSMO-RS, and COSMOmic, *Langmuir*, 2013, **29**, 3527–3537.
- 42 L. Mokrushina, M. Buggert, I. Smirnova, W. Arlt and R. Schomäcker, COSMO-RS and UNIFAC in prediction of micelle/water partition coefficients, *Ind. Eng. Chem. Res.*, 2007, **46**, 6501–6509.
- 43 M. Buggert, C. Cadena, L. Mokrushina, I. Smirnova, E. J. Maginn and W. Arlt, COSMO-RS calculations of partition coefficients: different tools for conformation search, *Chem. Eng. Technol.*, 2009, **32**, 977–986.
- 44 COSMOconf 4.3, COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2018, <http://www.cosmologic.de>.
- 45 TURBOMOLE V7. 3, University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007, Karlsruhe, Germany, 2018; available from <http://www.turbomole.com>.
- 46 A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
- 47 J. P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822–8824.
- 48 A. Schäfer, C. Huber and R. Ahlrichs, Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr, *J. Chem. Phys.*, 1994, **100**, 5829.
- 49 COSMOtherm, Release 19, COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2019, <http://www.cosmologic.de>.
- 50 S. Jakobtorweihen, T. Ingram and I. Smirnova, Combination of COSMOmic and molecular dynamics simulations for the calculation of membrane-water partition coefficients, *J. Comput. Chem.*, 2013, **34**, 1332–1340.
- 51 L. Wang, Measurements and implications of the membrane dipole potential, *Annu. Rev. Biochem.*, 2012, **81**, 615–635.
- 52 S. Endo, B. I. Escher and K.-U. Goss, Capacities of membrane lipids to accumulate neutral organic chemicals, *Environ. Sci. Technol.*, 2011, **45**, 5912–5921.
- 53 S. Queste, J. L. Salager, R. Strey and J. M. Aubry, The EACN scale for oil classification revisited thanks to fish diagrams, *J. Colloid Interface Sci.*, 2007, **312**, 98–107.
- 54 A. Benazzouz, L. Moity, C. Pierlot, V. Molinier and J.-M. Aubry, Hansen approach versus COSMO-RS for predicting the solubility of an organic UV filter in cosmetic solvents, *Colloids Surf., A*, 2014, **458**, 101–109.
- 55 M. J. Rosen and J. T. Kunjappu, *Surfactants and Interfacial Phenomena*, Wiley, Hoboken, NJ, USA, 4th edn, 2012.
- 56 Y. Zhang, S. E. Feller, B. R. Brooks and R. W. Pastor, Computer simulation of liquid/liquid interfaces. I. Theory and application to octane/water, *J. Chem. Phys.*, 1995, **103**, 10252–10266.
- 57 *CRC Handbook of Tables for Applied Engineering Science*, ed. R. E. Bolz and G. L. Tuve, CRC Press, Boca Raton, Florida, 2nd edn, 2000.
- 58 P. Mukerjee and J. Mysels, Critical micelle concentrations of aqueous surfactant systems. Prepared under contract for the Office of Standard Reference Data, National Bureau of Standards of NSRDS-NBS 36, Washington, DC, *J. Pharm. Sci.*, 1972, **61**, 319.
- 59 D. P. Tieleman, private communication.
- 60 M. Hahn, S. Krickl, T. Buchecker, G. Jost, D. Touraud, P. Bauduin, A. Pfitzner, A. Klamt and W. Kunz, Ab-initio prediction of structuring/mesoscale inhomogeneities in surfactant-free microemulsions and hydrogen-bonding-free microemulsions, *Phys. Chem. Chem. Phys.*, DOI: 10.1039/C8CP07544A.
- 61 J. A. H. Schwöbel and A. Klamt, Mechanistic skin penetration model by the COSMOperm method: Routes of permeation, vehicle effects and skin variations in the healthy and compromised skin, *Comput. Toxicol.*, 2019, **11**, 50–64.

