

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

A. Klamt^{*,1,2}, L. Koch¹, S. Terzi¹, U. Huniar¹, J. Schwöbel¹, and T. Gaudin^{1,3}

¹ COSMOlogic GmbH&CoKG, Imbacher Weg 46, D-51379 Leverkusen, Germany

² Institute of Physical and Theoretical Chemistry, Universität Regensburg, Germany

³ Université de Technologie de Compiègne

* e-mail: klamt@cosmologic.de

Abstract

During the past 20 years, the efficient combination of quantum chemical calculations with statistical thermodynamics COSMO-RS 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 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, as the prediction of micellar structure and critical micelle concentrations, finite loading effects in micelles and biomembranes, free energies and structure of liquid interfaces, micro-emulsions, and many more of similar problems, which often are of huge practical importance.

Introduction

The conductor-like screening model for realistic solvation¹⁻³ (COSMO-RS) is a combination of the dielectric continuum solvation model COSMO⁴ with an efficient statistical thermodynamic model of pairwise molecular surface interactions, known as quasi-chemical approximation⁵ or COSMOSPACE⁶. For the quantification of the surface interactions it uses the surface polarization charge densities σ 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 solution⁷⁻¹², and thus for all equilibrium distribution properties as partition coefficients, solubilities, vapor pressures, and related properties.

One of the limitations of COSMO-RS was its inability to handle inhomogeneous, structured liquid system. This was partially overcome by the introduction of the COSMOmic approach¹³, which uses external information about the structure of a micelle or biomembrane, usually taken from molecular dynamics (MD) simulations, in order to represent the micelle as a layered liquid of varying composition with respect to the COSMO polarization charge densities σ . 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 as achievable with MD simulations, at about 0.01 percent of the computational cost¹⁴⁻¹⁶. More recently, COSMOmic has been extended to ionic solutes^{17,18}, now allowing for the prediction of the biomembrane to water partition coefficients of neutral and ionic species with an accuracy of ~ 0.7 log.-units. Another extension of COSMO-RS towards inhomogeneous systems, i.e. to the prediction of interfacial tensions, has been developed by Andersson et al.¹⁹.

While the COSMOmic approach has been proven to be efficient and reliable with respect to the prediction of free energies of solutes in infinite dilution in micellar systems, its dependence on MD simulations for getting the structure of the micelle prevents its application to many questions of eminent practical importance, as the direct prediction of the critical micelle concentration (CMC) of new surfactants and surfactant mixtures²⁰, or the effects of finite concentration enrichment of solutes in micelles, as well as from the predictive simulation of liquid-liquid interfaces and micro-emulsions. 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. 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 call this new method COSMOplex, further on. In this paper, after a short introduction of the concepts of COSMO, COSMO-RS and COSMOmic, the theory of COSMOplex is presented, followed by a description of first application examples to biomembranes, micelles, micro-emulsions, and liquid-liquid interfaces.

Methods and definitions

COSMO

Quantum chemical calculations are performed with the dielectric continuum solvation model COSMO⁴ for each of the molecules, i.e. for the solute and solvent molecules under consideration. COSMO is used in the limit of the ideal conductor, i.e. with infinite dielectric constant. The result of such a COSMO calculation is a COSMO file, which 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 a listing of the COSMO surface segments with respect to positions, segment area and conductor screening charge on the segment.

For the purpose of COSMO-RS and especially of COSMOplex it is required that 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 COSMO calculation and is represented by a separate COSMO file, containing the surface polarization charge information and the energy $E_{COSMO}^{X_c}$ of conformation c of compound X . We will use the expression “molecule” here for each conformation, and the expression “compound” for the ensemble of all the considered conformations of a chemical entity. The way to determine the relative weights of the different conformations will be described in the COSMOmic paragraph. The COSMOconf software²¹ in combination with the quantum chemistry software package TURBOMOLE²² was used to generate the COSMO files of the

relevant conformations for the compounds considered in this paper. The BP86 density functional^{23,24} in combination with the TZVP basis set²⁵ were used for the COSMO calculations. COSMO*conf* was slightly modified to the needs of COSMOplex, in order to keep geometrically diverse conformations with very similar σ -profiles in the set, while these are filtered out with standard COSMO*conf*, because they are irrelevant in homogeneous liquids.

COSMO-RS

The conductor-like screening model for realistic solvation (COSMO-RS)¹⁻³ starts from COSMO files for all involved molecules, i.e. of solute and solvent molecules. COSMO-RS takes the COSMO surface segments and performs an efficient statistical thermodynamics algorithm, known as COSMOSPACE,⁶ which is equivalent to an exact solution of Guggenheim's quasi-chemical approximation.⁵ This results in a kind of thermodynamic continuum response function, the so-called σ -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.

The interaction energies of the surface segments, which are required for the calculation of the contact probabilities with COSMOSPACE, are based on local properties of the two surface segments involved in the contact. In the initial and most simple version of COSMO-RS, only the most important of these properties is considered, i.e. the conductor polarization charge density σ . 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-imbedded molecules, is expressed as a misfit energy

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

where $a_{contact}$ is the size of the surface contact and σ and σ' are the polarization charge densities of the two interacting segments, respectively. The next most important contribution to the interactions is hydrogen bonding, expressed as

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

where σ_{hb} is a kind of minimum polarity threshold for hydrogen bond formation. The temperature dependence of the hydrogen bond energy coefficient $c_{hb}(T)$ expresses the entropy loss going along 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 enthalpies²⁶ and with hydrogen bond enthalpies derived from FTIR-experiments²⁷. Some newer versions of COSMO-RS use a slightly more refined functional form for the hydrogen interactions, involving element specific parameters.

In the most simple “one-descriptor” version of COSMO-RS each molecule X can be described by its surface composition histogram with respect to σ , the so-called σ -profile $p^X(\sigma)$, and a pure or mixed liquid system is characterized by its solvent σ -profile

$$p_S(\sigma) = \sum_i \frac{x_i p^{X_i}(\sigma)}{x_i A^{X_i}} \quad (3)$$

where x_i is the mole fraction of molecule i and A_i is its COSMO surface area. Solving the COSMOSPACE equations, which in the one-descriptor version turn into

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

we yield the σ -potential $\mu_S(\sigma)$, which expresses, how much a solvent S likes surface segments of polarity σ . The σ -potential is a kind of thermodynamic continuum response function of the liquid with respect to the surface polarity of the molecules. Integrating the σ -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

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

where $\gamma_{S,combi}^X(A^X, V^X, S)$ can be any typically used expression for the combinatorial activity coefficient, depending on the composition and the surface areas and volumes of the solute and solvent molecules. The solvent dependence usually is 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. μ_s^x is the pseudo-chemical potential, expressing the chemical potential at a reference mole fraction of 1 mol/mol.

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

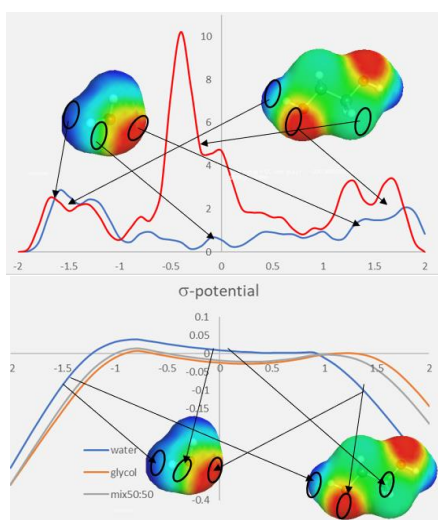


Figure 1: Schematic illustration of the COSMO-RS principles: First the COSMO surface polarization charge densities of the surface segments are converted into a histogram (σ -profile, upper diagram). Then the solvent σ -profile is converted into a σ -potential (lower diagram), and the chemical potential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

Meanwhile COSMO-RS has been refined by including the influence of other surface descriptors², 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 local surface properties of the surface segments, which can be derived from the COSMO calculation or the molecular structure and can be considered as known for each surface segment, and instead of using eq. 4 the statistical thermodynamics is done by the equivalent exact solution of the quasi-chemical approximation, which was published under the acronym COSMOSPACE⁶ for the generalized case. Nevertheless, for the sake of clarity in the following we will describe the concepts of COSMOmic and COSMOplex just using the terminology of polarization charge density interactions. The BP-TZVP-2018 parameterization of COSMOtherm²⁸ was used for the COSMO-RS calculations within COSMOplex in this article.

COSMOmic

While the original COSMO-RS was just applicable to homogeneous bulk liquids, it has been extended to inhomogeneous, micellar systems^{13,17} by considering such systems as layered liquids, each layer being considered as homogeneous and described by COSMO-RS. Nevertheless, 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 with the partial σ -profiles of the corresponding atoms in a COSMO file of the respective compound led to a specific σ -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 σ -sigma potentials, we can use the standard COSMO-RS algorithm to calculate a σ -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, center positions and orientations of the solute relative to the z -direction, is performed. In practice, the mid of each layer is tested as center position, and a grid of vectors on the unit sphere, typically $n_{\text{ori}} = 162$ directions, are used for the orientations of the micelle normal relative to the molecule. It should be noted that two different choices for the molecular center have been implemented, either the geometric center of the molecular surface area, or the center of polarity, defined as the squared polarization charge density σ . By default, we use the geometric surface center throughout this paper. For a given state, i.e. conformation, position and orientation of the compound, sometimes abbreviated as a *cpo* further on, each of the surface segments v has a coordinate z_v^{cpo} with respect to the z -direction of the micellar system. The chemical potential of this surface segment v is then evaluated by linear interpolation of the two adjacent layer- σ -potentials $\mu_M(\sigma_v, z)$. For negative values of z_v^{cpo} , which may appear in lamellar micelles geometries, a mirror boundary condition is applied, and for values of z_v^{cpo} larger than the outer boundary of the simulated box the σ -potential is assumed to be the same as in the pure solvent embedding the micelle, i.e. in most cases pure water. Then, in close analogy to eq. 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 $V(z)$ simply can be taken into account in the energy of each surface segment by the multiplication of the polarization charge density σ with the membrane potential and the segment area¹⁷.

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 DFT-COSMO calculation yields the total free energy of the compound for this special *cpo*.

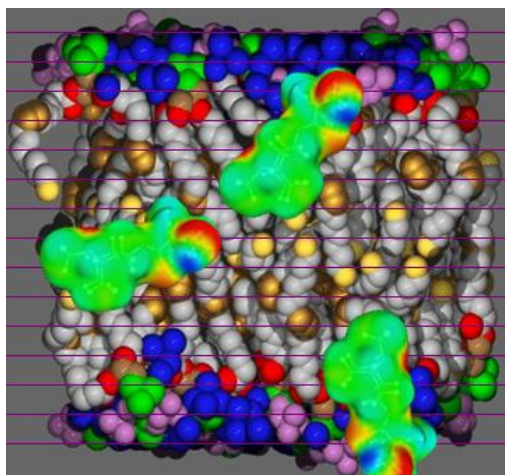


Figure 2: Schematic illustration of COSMOmic: A micelle is considered as layered liquid. The composition of each liquid is taken from MD simulations. After the calculation of σ -profiles and σ -potentials of the layers, all positions and orientations of a solute molecule are sampled, calculating the state specific free energy from the local σ -potentials.

Sampling of all positions, orientations, and conformations thus allows us to calculate the partition function for compound X in the micellar system, and from that we can derive all thermodynamic equilibrium properties, including 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^{13,16,17}, which are of big interest for biochemistry and pharmaceutical research. But also for other micelle-water partition systems COSMO-RS yields reliable partition coefficients^{15,24–26}. 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.

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

Based on the terminology introduced in the previous chapters we now can set the framework for COSMOplex. The workflow is outlined in figure 3, and we will refer to this figure and the numbering of steps indicated therein along the explanation of the method.

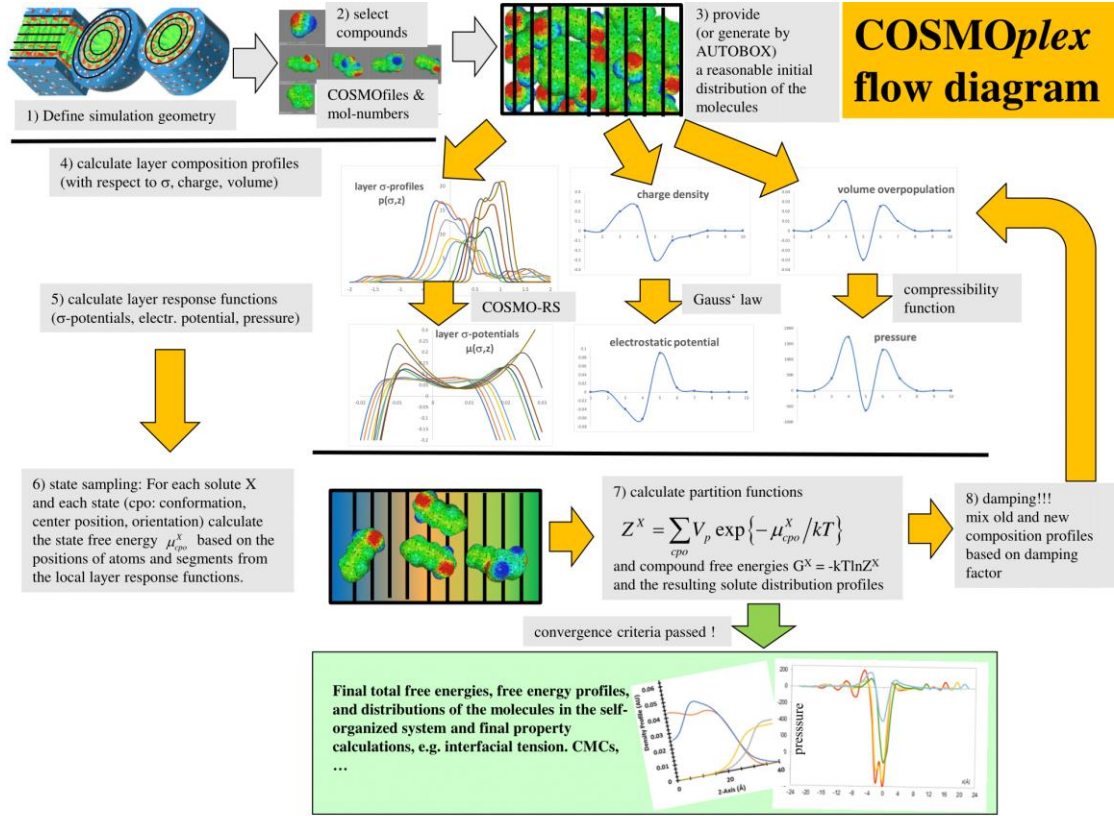


Figure 3: (will be printed full page) Workflow of the COSMOplex method

Step 1: Definition of geometry: 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_{geo}^{tot} virtually cut into L layers along the z -axis. For simplicity we assume that all layers have the same thickness δ , although varying thicknesses could be handled as well. By default, we shall assume a rectangular geometry, but cylindrical or spherical boxes are supported as well. In the latter cases the z -axis is the radial axis. Thus, the geometric volume of layer i is

$$V_{geo}(z_i) = A\delta \quad (7)$$

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

$$V_{geo}(z_i) = \pi H \left(\left(z_i + \frac{\delta}{2} \right)^2 - \left(z_i - \frac{\delta}{2} \right)^2 \right) = 2\pi H z_i \delta \quad (7a)$$

for a cylindrical box of height H , and

$$V_{geo}(z_i) = \frac{4\pi}{3} \left(\left(z_i + \frac{\delta}{2} \right)^3 - \left(z_i - \frac{\delta}{2} \right)^3 \right) = 4\pi \left(z_i^2 \delta + \frac{\delta^3}{12} \right) \quad (7b)$$

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

Step 2: Selection of compounds: We consider the box to be filled with $N_{tot} = \sum_{k=1}^{n_c} N_k$ molecules of n_c different compounds. Each compound X_k may be represented by m_k conformations.

Each conformation X_k^j is represented by a 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 $V_{k,COSMO}^j$. The total volume is split into atom volumes V^α based on a relative closest distance criterium, so that we have atom volumes available for each atom α . More detail on atom volumes and atom centers is given in Appendix B.

Step 3: We generate a start distribution of the molecules in the box. This 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 (see SI for more information on AUTOBOX) is provided in the COSMOplex software to achieve a reasonable start distribution. The start distribution is represented as an atom probability distribution, as it was used as input for COSMOmic. More details of the AUTOBOX algorithm are given in the supporting information (SI). 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 may have 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 4: 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 σ -profiles of the atoms to the respective layers and yield layer σ -profiles $p_{sos}^0(\sigma, z_i)$, 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 $Q_{sos}^0(z_i)$ of layer charge for all layers by summing up the product of polarization charge density times segment area for all segments in a layer.

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

$$\tilde{\rho}^0(z_i) \equiv \frac{V_{occ}^0(z_i)}{V_{geo}(z_i)} \quad (8)$$

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 1 indicate an overpopulation of the respective layer, space filling ration lower than 1 indicate underpopulation.

The average molecular surface area and the average molecular volume is also calculated in each layer (see Appendix C).

Step 5: Next we calculate the three continuum response functions corresponding to the three composition functions calculated in step 4. The response functions for the layer σ -profiles are the corresponding σ -potentials $\mu_{sos}^0(\sigma, z_i)$. The response function for the charge distribution is the electrostatic potential, which can be calculated from charge distribution according to Gauss law. For details see Appendix D. The response function the space filling is pressure. It can be calculated from the compressibility. For details see Appendix E.

The combinatorial contribution, a kind of size correction, may be considered as an additional response function. Since its contribution is small we refer to Appendix F for details.

Step 6: As a next step we scan all possible states of each compound X in the simulation box. The states are enumerated by the conformation index (c), the center 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. For the latter a homogeneous grid of vectors on the unit sphere as introduced in the original COSMO paper⁴ is used. By default, 162 orientations are considered.

For each state cpo the free energy $\mu_{sos}^{X_{cpo}}$ is initialized with the conformational energy of the compound. For the COSMO-RS part of the free energy we determine the position of each surface segment and of each atom of the compound X. Using a linear interpolation between the two neighboring layers of each segment, we determine the value of the σ -potential of the segment at its position and the corresponding electrostatic potential and the respective contributions of each segment to the total free energy $\mu_{sos}^{X_{cpo}}$ of the state. In addition, we determine the position of each atom. The atom volume is multiplied by the interpolated pressure of the two nearest layers and this pressure penalty is added to $\mu_{sos}^{X_{cpo}}$. Finally, the combinatorial free energy is added (see Appendix F). Hence, the free energy of the state cpo is

$$\mu_{SOS}^{X_{cpo}} = E_{COSMO}^{X_c} + \mu_{SOS,res}^{X_{cpo}} + E_{SOS,estat}^{X_{cpo}} + E_{SOS,press}^{X_{cpo}} + \mu_{S,combi}^{X_{cpo}} \quad (9)$$

Step 7) Knowing the individual free energies of all states cpo, we can build the compound partition functions as sum of the Boltzmann weights:

$$Z^X = \sum_{cpo} N_p^{mol} \exp\left\{\frac{-1}{kT} \mu_{SOS}^{X_{cpo}}\right\} \quad (10)$$

The multiplication by the number of molecules N_p^{mol} in layer p is required, since the combinatorial contribution of COSMO-RS is gauged to a mol/mol concentration scale. The total free energy of each compound X is calculated as

$$G^X = -kT \ln\{Z^X\} \quad (11)$$

And the total free energy of the system as

$$G_{tot} = \sum_{k=1}^{n_c} N_k \mu_{SOS}^X = -RT \sum_{k=1}^{n_c} N_k \ln\{Z_k\} - E_{Coulomb}^{corr} - E_{press}^{corr} \quad (12)$$

where the corrections for the electrostatic and pressure energy are required in order to avoid double counting. For details see Appendix G.

The individual probability to find compound X in state cpo is

$$w_{cpo}^X = N_p^{mol} \exp\left\{\frac{-1}{kT} \mu_{SOS}^{X_{cpo}}\right\} / Z^X \quad (13)$$

By knowing the individual state probabilities, we can construct new layer composition functions, i.e. new layer σ -profiles, charge distribution, and layer pressure profile.

Step 8) We mix the new and the previous composition function by using a damping factor λ as weight for the old composition and $1 - \lambda$ as weight for a new composition, i.e. we know all the three composition functions as they were achieved in step 4. Thus we can restart the COSMOplex cycle at step 5, i.e. with the calculation of new response functions. By that 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 probably are subject to future improvement. Since the final results do not depend on the convergence algorithm, we skip a description of the damping algorithms in this article.

With step 8 we have closed the iterative cycle for the recursive calculation of the distribution and free energies of molecules in the self-organizing system. Finally, after several hundreds or thousands of iterations the system converges to a minimum of the total free energy. If the convergence criteria, which may depend on the application, are met, the iterative cycle is stopped and final property evaluations, as the calculation of the interfacial tension as integral of the pressure profile, can be easily performed.

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. Nevertheless, if starting from a reasonable initial distribution of the molecules, we found that in many cases the system directly converges to the global minimum.

DIRPLEX: Directional segment extension

Although the COSMO*plex* 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 COSMO*plex* 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 system two segments need to have opposite orientations in order to form an interacting segment pair. But in COSMO*plex* we consider an anisotropic liquid, with a specially preferred direction, the z -direction of the SOS. Thus, it may be that some segment types preferably point to the positive z -directions, and others preferably point in lateral or opposite direction. As an example, we may consider the interface of water to nonpolar solvent, e.g. cyclohexane, as illustrated in figure 4. If the alkane phase is in the region of low z -values and water at high z -value, then the layer accommodating the relatively sharp interface of the liquids will mainly have nonpolar surface segments with 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 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 introduced into COSMO-RS by considering the segment normal direction as an additional segment property. For the purpose of COSMO*plex* it is sufficient to split each segment v into two subsegments in the following way: If the scalar product of the segment normal vector with the z -direction ($n_{v,z}$) is positive, the portion $n_{v,z}$ is added to a z^+ -sub-ensemble of surface segments, and the portion $(1 - n_{v,z})$ is added to the z^0 -sub-ensemble. If $n_{v,z}$

is negative, the portion $|n_{v,z}|$ is added to a z^- -sub-ensemble of surface segments, and the portion $(1 + n_{v,z})$ 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 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 neighboring layers which is necessary to yield identical amounts of z^+ - and z^- -areas in each layer.

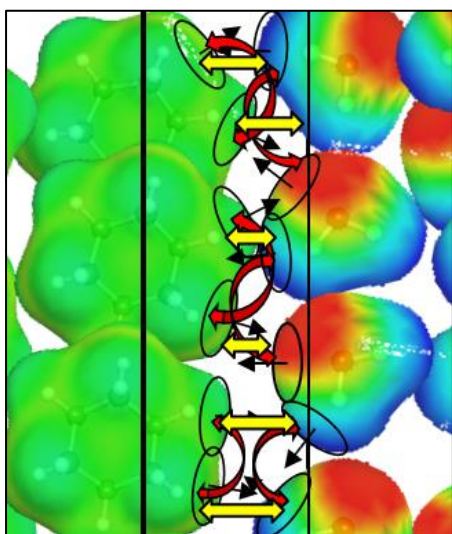


Figure 4: Schematic illustration of the DIRPLEX extension: Without the DIRPLEX constraint, the polar water segments in the middle layer would preferentially interact with each other, and the green alkane segments would form nonpolar pairs (red arrows). The directionality constraint enforces the formation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

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 examples section we compare COSMOplex with and without the DIRPLEX extension for the DMPC biomembrane and for the cyclohexane-water interface. Larger differences are only found in the latter case. Nevertheless, for the sake of being on the save side, DIRPLEX should be used by default, if the computational resources allow for the increase of the computation time by roughly a factor of 3 caused by DIRPLEX.

Application examples

In this section, we describe a range of different applications. All applications are considered as proofs of principle, not expecting that the optimal performance of COSMOplex has yet been achieved.

Self-organization of a DMPC bilayer and DMPC-water partition coefficients

As a first example, we apply COSMOplex to the self-organization of a DMPC biomembrane, which was the most important system studied by COSMOmic^{13,17}. A first COSMOplex simulation was started from a AUTOBOX generated input. Three more runs were started, adding the DIRPLEX option, adding the self-consistent electrostatic potential calculation, and adding both options, respectively.

Figure 5 shows the volume distribution profiles for DMPC and water, respectively, resulting from the four COSMOplex simulations. The distribution resulting from the MD simulations³², 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 snap shot deviates a bit from the idealized flat and reflection symmetric 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 negligible effect. Figure 5 also shows the self-consistent electrostatic potentials as arising from the two last options, in comparison to the empirically fitted potential fitted to experimental DMPC-water partition coefficients of ions¹⁷. Given the fact that most reported DMPC MD simulations yield values between 1.0 and 1.3 V in the membrane center^{17,33}, 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 raise 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.

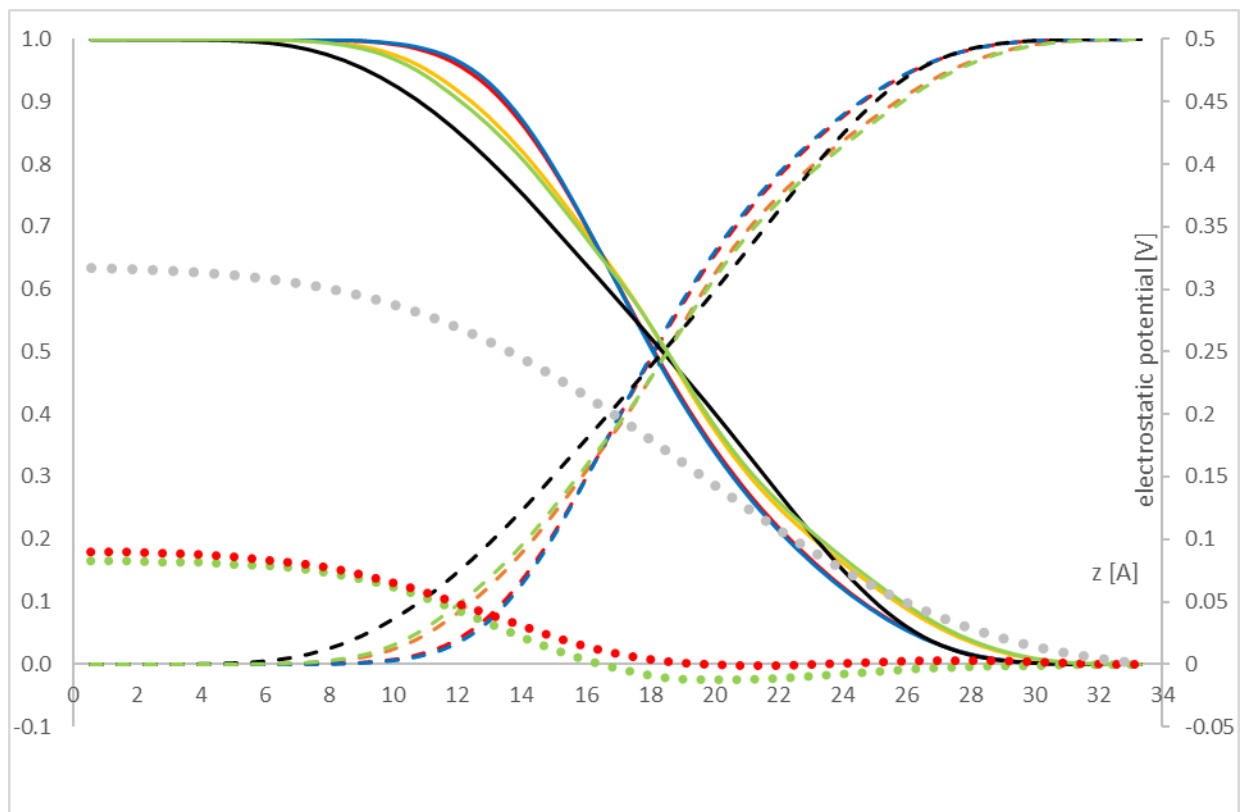


Figure 5: Volume fractions of DMPC (full lines) and water (dashed lines) in a DMPC membrane. The center of the bilayer is at the left end of the diagram. The MD composition³² is marked in black, bare COSMOplex in blue, COSMOplex with self-consistent electrostatics in red, COSMOplex with DIRPLEX in yellow, and COSMOplex with DIRPLEX self-consistent electrostatics in green. The dotted lines show the two self-consistent electrostatic potentials, with the fitted, pseudo-experimental potential¹⁷ is given for reference in grey.

In all cases we also calculated the DMPC-water partition coefficients for the 207 neutral solutes considered by Endo et al.³⁴. 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 input.

Model	mean dev	RMSD*	slope	shift	r ²	RMSD
MD_based	-0.72	0.86	1.24	-1.56	0.78	0.81
COSMO $plex$	-0.15	0.90	1.15	-0.61	0.75	0.88
&Epot	-0.21	0.88	1.16	-0.70	0.76	0.86
&DIRPLEX	-1.23	0.94	1.05	-1.45	0.71	0.94
&DIRPLEX&Epot	-1.24	0.89	1.11	-1.68	0.75	0.88

Table 1: Statistical performance of the 5 different DMPC models for the prediction of 207 logarithmic DMPC-water partition coefficients³⁴. RMSD* is the root mean squared deviation after correction for the mean deviation, the last four columns give the coefficients and performance from linear regression.

Prediction of critical micelle concentrations

The prediction of critical micelle concentrations (CMCs) is of special interest in many areas of formulation science. Here we consider the prediction of the CMCs of seven diverse neutral surfactants. For each surfactant a spherical aqueous micelle model was built with AUTOBOX based on the surfactant geometry, and a COSMO $plex$ calculation with self-consistent electrostatic field was started. As shown in table 2 and figure 6, the trends in CMC are reasonable well predicted by COSMO $plex$. It needs to be emphasised, that further studies, investigating the influence of system size, layer width, more conformations and other options of COSMO $plex$, are required in order to ultimately determine the accuracy of the COSMO $plex$ for CMC prediction. Initial simulations for ionic surfactants also converged nicely and showed reasonable results. These systems will be studied in in more detail in a forthcoming paper.

Surfactant	COSMO $plex$	exp.
nonylphenyl pentaethylene glycol	-5.01	-4.26 ³⁵
dodecyl- β -D-glucoside	-3.42	-3.77 ³⁶
dodecyl- β -D-maltoside	-4.01	-3.70 ³⁷
octyl glyceryl ether	-2.48	-2.24 ³⁷
tetraethylene glycol monoethyl ether	-1.99	-2.15 ³⁵
octyl- β -D-glucoside	-2.86	-1.80 ³⁸
octanoyl-N-methylglucamine	-2.68	-1.16 ³⁹

Table 2: COSMO $plex$ results for CMCs ($\log \text{CMC}/[\text{mol/l}]$) of seven diverse surfactants.

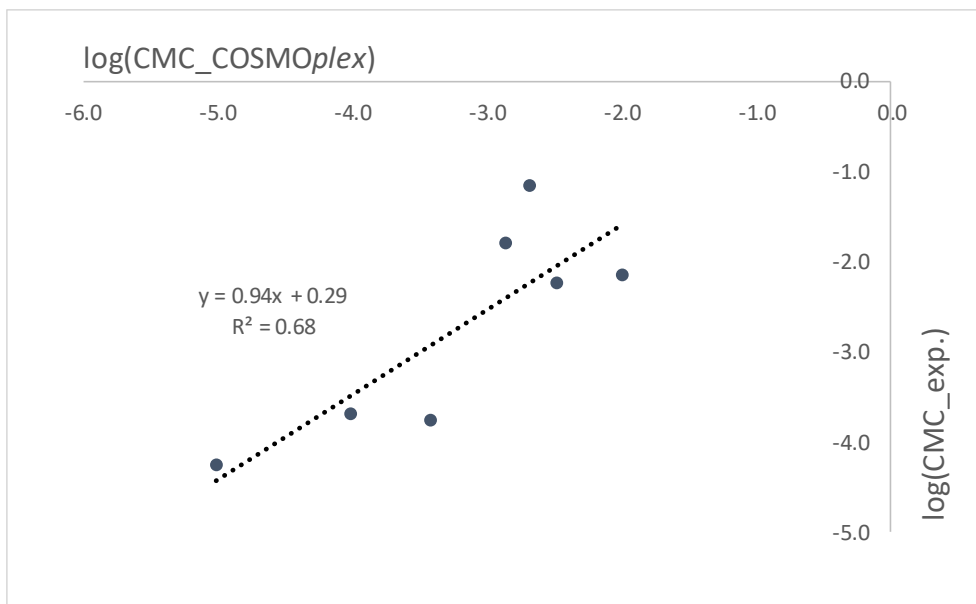


Figure 6: COSMOplex results for CMCs (log CMC/[mol/l]) of seven diverse surfactants.

Microemulsion systems

Microemulsions are of special interest for many areas of formulation science and for the important field of enhanced oil recovery. If such system, composed of a less polar compound, usually called oil, a surfactant, and water, is mixed in a test tube by shaking, three phases form, i.e. an oil-rich phase at the top, an aqueous phase at the bottom, and a micellar phase in the middle. As a proof of principle, we applied COSMOplex to a typical microemulsion system, i.e. octane, C8E3, and water at 288 K. We initialized the system as 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 at these conditions⁴⁰. And indeed, after heavy initial fluctuations it really formed three phases with surfactant saturated oil and water phases, and a middle region in which oil, surfactant and water self-organize to multiple micellar layers (see figure 7). This means that by 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. 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, because such investigations are practically impossible with other simulation techniques.

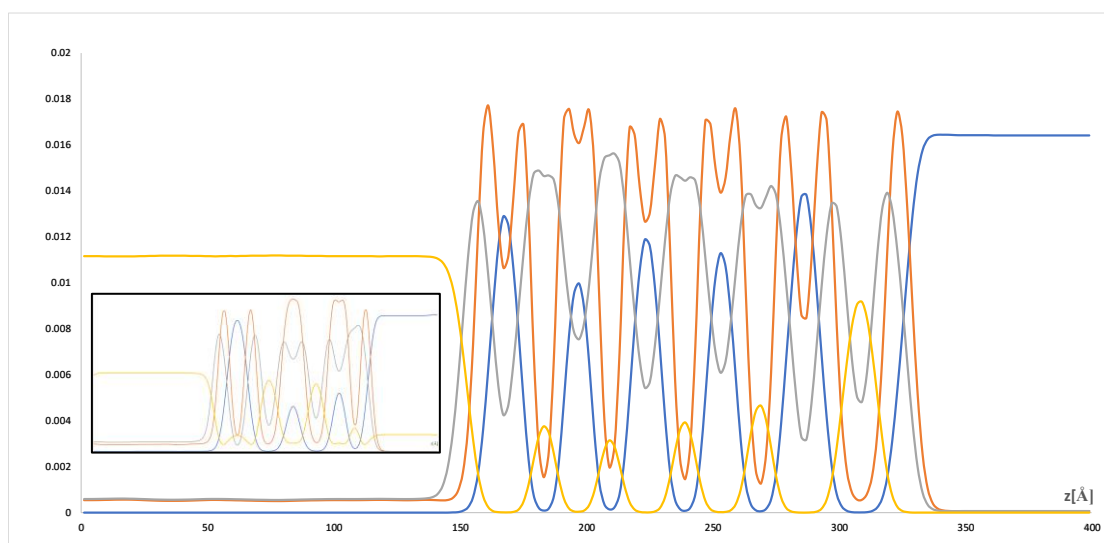


Figure 7: Spontaneous formation of a microemulsion phase in an oil (octane, 35 %-vol), surfactant (C8E3, 30 %-vol) and water (35 %-vol) system: The lines give the relative atom probability for octane atoms (yellow), water oxygen (blue), surfactant oxygens (red) and surfactant carbon atoms (grey). Five double layers built up for COSMOplex. If we use the DIRPLEX extension less (two or three) and broader double layers are formed (see inset).

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 simulations⁴¹, the IFT can be simply calculated as the z-integral of the lateral pressure.

The initialization of the rectangular simulation box with mirror boundary conditions is relatively simple. For details see the AUTOBOX section in the SI. In figure 8 three typical interfacial pressure profiles are shown. The profiles have been base-line corrected. In figure 9 we show the results for the IFT prediction of 40 aqueous systems taken from the paper of Andersson et al.¹⁹ 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 figure 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 are much worse and strongly underestimate the experimental interfacial tensions by a factor of almost 2. The use of the self-consistent electrostatic potential has almost no influence on the interfacial tension prediction.

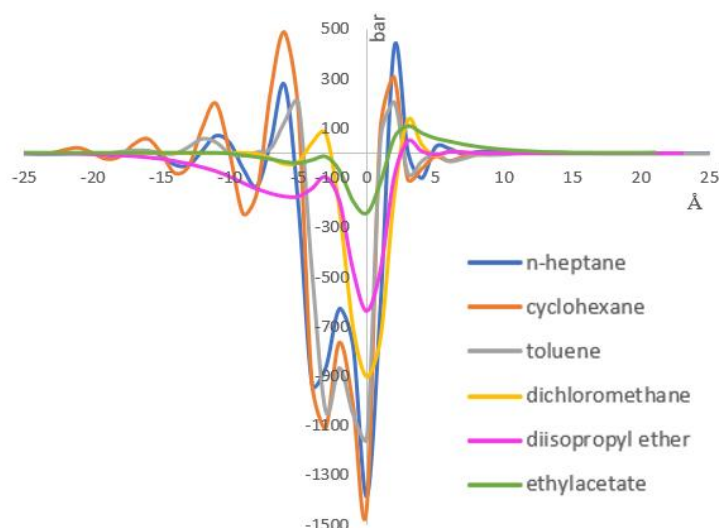


Figure 8: COSMOplex pressure profiles for six interfacial tension calculations: The pressure drops in the interfacial region. For cyclohexane and toluene pressure fluctuations arising from the first, second and third molecular layer can be observed. The pressure profiles are base-line corrected.

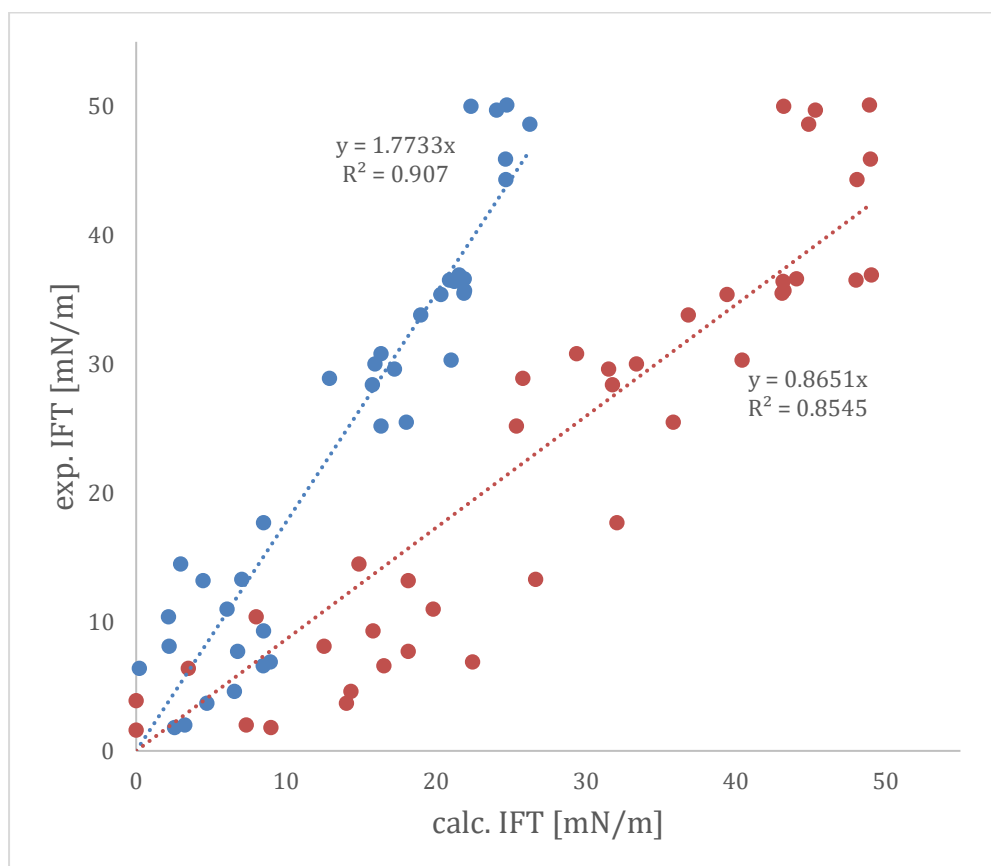


Figure 9: Experimental¹⁹ and predicted interfacial tensions for 40 aqueous systems: COSMOplex predictions without DIRPLEX extensions are in blue, with DIRPLEX in orange. (For compound list and data see Table SI1 in the supporting information)

Summary and outlook

The COSMO-RS method was originally developed for homogenous, bulk liquids. By the COSMOmic method it was to the simulation of solutes in micellar systems but being limited to infinite dilution and requiring computationally expensive molecular dynamics input, describing 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 self-organizing molecular systems. First applications to micelle formation, micro-emulsion 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 10 hours on a single laptop CPU. Thus, the simulations are 3 to 4 orders of magnitude less CPU demanding than comparable molecular dynamics simulations. Furthermore, without any doubt, COSMOplex still has considerable potential for tuning its performance. By that COSMOplex enables many simulation applications of self-organizing which so far seem to be completely unfeasible.

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

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 nonionic 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, as well as 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 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.

Supporting information: A description of the AUTOBOX algorithm used for the preparation of COSMOplex input distributions is provided as supporting information.

Conflicts of Interest: The authors declare the following competing financial interests: Andreas Klamt is chief executive officer, and all other authors are or were employees of COSMOlogic. COSMOlogic commercially distributes the software used in this paper.

Appendix A: Boundary conditions in COSMOplex

Currently in COSMOplex the following boundary conditions are supported:

- 1) 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.
- 2) Mirror boundary conditions for rectangular boxes. If a coordinate is outside the box, it is transformed into the box by reflection. This assumes mirror symmetric 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.
- 3) 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 it 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.

Appendix B: Details of atomic volume assignment

The geometric center position of the associated volume of an atom α is introduced as $\underline{x}_{vol}^{\alpha}$. This slightly deviates from the atom position \underline{x}^{α} 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 software⁴². All atom volumes are rescaled

by the ratio of V_k and $\langle V_{k,COSMO} \rangle_{pure}$, 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 $\bar{r}_{sos}(z_i)$ and average molecular surface area $\bar{q}_{sos}(z_i)$, are calculated as well. These are required for the calculation of the combinatorial free energy term in eq. 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^{mol} . Multiplying the fraction of surface in layer i by the molecular volume yields the average volume $\bar{r}_{sos}(z_i)$, and multiplying the fraction by the molecular area yields the average surface area $\bar{q}_{sos}(z_i)$.

Appendix D: Calculation of the electrostatic potential $V_{sos}(z_i)$ from layer charges $Q_{sos}(z_i)$

The electrostatic potential $V_{sos}(z_i)$ 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 the pure compound refraction indices and averaged based on the atom volume probabilities in the layers. The pure compound refraction indices can either be given as experimental input, or they are calculated from an increment scheme. The values used in the examples are:

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, if they feel the electrostatic potential.

In **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 $Q_{sos}(z_i)$ divided by layer volume, multiplied by the layer width δ . Hence, we have for the electric field at the interface between layer i and $i-1$, i.e. at $z_i - \delta/2$

$$E_i = E(z - \delta/2) = \sum_{j=i}^L \frac{q_j}{A} \quad (D1)$$

In **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 $Q_{sos}(z_i)$ divided by layer volume, multiplied by the layer width δ . Hence, we have for the electric field at the interface between layer i and $i-1$, i.e. at $z_i - \delta/2$

$$E_i = E(z - \delta/2) = \sum_{j=i}^L \frac{q_j}{A} \quad (D2)$$

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 ($i = L$) is zero (e.g. in the water phase of a DMPC bilayer). The electrostatic potential is the integral of the Hence the potential in the middle of layer i is

$$V(z_i) = \sum_{k=i+1}^L D_k \delta = \sum_{k=i+1}^L \frac{E_k}{n^2(z_k)} \delta \quad (D3)$$

In **spherical geometry** the electric field at a radius z is the same as the electric field of a point charge at the center which has the value of all included charge $Q^{inc}(z_i)$. 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.

$$\Delta V(z_i, z_{i+1}) = E(z_i + \frac{\delta}{2}) \frac{\delta}{n^2(z_i + \frac{\delta}{2})} = \frac{Q^{inc}(z_i)}{(z_i + \frac{\delta}{2})^2} \frac{\delta}{n^2(z_i + \frac{\delta}{2})} \quad (D4)$$

$Q^{inc}(z_i)$ is the sum of all layer charges up to z . To calculate the potential $V(z_i)$, we just have to start at the center and sum up all potential changes.

In **cylindrical geometry** the electric field as a radius z is the same as the electric field $E(z)$ of a line charge at the center which has the value of all included charge. In close analogy to the spherical case we get (with R_2 being the length of the simulated cylinder)

$$\Delta V(z_i, z_{i+1}) = E(z_i + \frac{\delta}{2}) \frac{\delta}{n^2(z_i + \frac{\delta}{2})} = \frac{Q^{inc}(z_i)}{2R_2(z_i + \frac{\delta}{2})} \frac{\delta}{n^2(z_i + \frac{\delta}{2})} \quad (D5)$$

To calculate the potential $V(z_i)$ we just have to start at the center and sum up all potential changes.

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

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

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 liquids⁴³

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

in which c_{comp} is the compressibility, which we set to 10^{-4} throughout this article for all compounds except for water, for which we use $0.5 * 10^{-4}$. 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 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 SI.

Appendix F: Combinatorial contribution

The combinatorial contribution to the free energy of a molecule X in layer i is calculated based on 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 from 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. It is also required for the compression energy, because the compression energy included in the partition functions of the molecules reflect the interaction of each atom volume with the final, self-consistent pressure, while the compression energy corresponding has a different value given by

$$E_{comp}^{corr} = \sum_{l=1}^{\#layer} \left[\left(\frac{\tilde{\rho}_l^9 - 2\tilde{\rho}_l^{4.5} + 1}{9 * 4.5 c_{comp}} \right) - P(\tilde{\rho}_l) \right] \langle V_l \rangle \quad (G5)$$

where $\langle V_l \rangle$ is the expectation volume of the molecular (or equivalently the atomic) volume in a layer l, and the expression in parentheses is the Lennard-Jones like expression of compression energy, which was used in eq. 9 for the derivation of the pressure formula.

References

- (1) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, 99 (7), 2224–2235.
- (2) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, 102 (26), 5074–5085.
- (3) Klamt, A. *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, The Netherlands; Boston, MA, USA, 2005.
- (4) Klamt, A.; Schüürmann, G. 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**, 1993 (5), 799–805.
- (5) Guggenheim, E. A. *Mixtures*; Clarendon Press: Oxford, 1952.

- (6) Klamt, A.; Krooshof, G. J. P.; Taylor, R. COSMOSPACE: Alternative to Conventional Activity-Coefficient Models. *AIChE J.* **2002**, *48* (10), 2332–2349.
- (7) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. Blind Prediction of Cyclohexane-Water Distribution Coefficients from the SAMPL5 Challenge. *J. Comput. Aided Mol. Des.* **2016**, doi:10.1007/s10822-016-9954-8.
- (8) Klamt, A.; Eckert, F.; Reinisch, J.; Wichmann, K. Prediction of Cyclohexane-Water Distribution Coefficients with COSMO-RS on the SAMPL5 Data Set. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 959–967.
- (9) Reinisch, J.; Klamt, A.; Diederhufen, M. Prediction of Free Energies of Hydration with COSMO-RS on the SAMPL3 Data Set. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 669–673.
- (10) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation. *J. Chem. Inf. Model.* **2015**, *55* (6), 1192–1201.
- (11) Zhang, J.; Tuguldur, B.; van der Spoel, D. Correction to Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation. *J. Chem. Inf. Model.* **2016**, *56* (4), 819–820.
- (12) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. On the Performance of Continuum Solvation Methods. A Comment on “Universal Approaches to Solvation Modeling.” *Acc. Chem. Res.* **2009**, *42* (4), 489–492.
- (13) Klamt, A.; Huniar, U.; Spycher, S.; Keldenich, J. 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* (38), 12148–12157.
- (14) Paloncýová, M.; DeVane, R.; Murch, B.; Berka, K.; Otyepka, M. Amphiphilic Drug-like Molecules Accumulate in a Membrane below the Head Group Region. *J. Phys. Chem. B* **2014**, *118* (4), 1030–1039.
- (15) Ritter, E.; Yordanova, D.; Gerlach, T.; Smirnova, I.; Jakobtorweihen, S. Molecular Dynamics Simulations of Various Micelles to Predict Micelle Water Partition Equilibria with COSMOmic: Influence of Micelle Size and Structure. *Fluid Phase Equilibria* **2016**, *422*, 43–55.
- (16) Jakobtorweihen, S.; Zuniga, A. C.; Ingram, T.; Gerlach, T.; Keil, F. J.; Smirnova, I. Predicting Solute Partitioning in Lipid Bilayers: Free Energies and Partition Coefficients from Molecular Dynamics Simulations and COSMOmic. *J. Chem. Phys.* **2014**, *141* (4), 045102.
- (17) Bittermann, K.; Spycher, S.; Endo, S.; Pohler, L.; Huniar, U.; Goss, K.-U.; Klamt, A. Prediction of Phospholipid-Water Partition Coefficients of Ionic Organic Chemicals Using the Mechanistic Model COSMOmic. *J. Phys. Chem. B* **2014**, *118* (51), 14833–14842.
- (18) Klamt, A. COSMO-RS for Aqueous Solvation and Interfaces. *Fluid Phase Equilibria* **2016**, *407*, 152–158.
- (19) Andersson, M. P.; Bennetzen, M. V.; Klamt, A.; Stipp, S. L. S. First-Principles Prediction of Liquid/Liquid Interfacial Tension. *J. Chem. Theory Comput.* **2014**, *10* (8), 3401–3408.
- (20) Jakobtorweihen, S.; Yordanova, D.; Smirnova, I. Predicting Critical Micelle Concentrations with Molecular Dynamics Simulations and COSMOmic. *Chem. Ing. Tech.* **2017**, *89* (10), 1288–1296.
- (21) *COSMOconf 4.0*; COSMOlogic GmbH & Co. KG; <http://www.cosmologic.de>: Leverkusen, Germany, 2015.
- (22) *TURBOMOLE V7.2*; University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>: Karlsruhe, Germany, 2017.
- (23) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.
- (24) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33* (12), 8822–8824.

- (25) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100* (8), 5829.
- (26) Klamt, A.; Reinisch, J.; Eckert, F.; Hellweg, A.; Diedenhofen, M. Polarization Charge Densities Provide a Predictive Quantification of Hydrogen Bond Energies. *Phys. Chem. Chem. Phys.* **2012**, *14* (2), 955–963.
- (27) Klamt, A.; Reinisch, J.; Eckert, F.; Graton, J.; Le Questel, J.-Y. Interpretation of Experimental Hydrogen-Bond Enthalpies and Entropies from COSMO Polarisation Charge Densities. *Phys. Chem. Chem. Phys.* **2013**, *15* (19), 7147–7154.
- (28) *COSMOtherm, Release 18.01*; COSMOlogic GmbH & Co. KG; <http://www.cosmologic.de>; Leverkusen, Germany, 2018.
- (29) Ingram, T.; Storm, S.; Kloss, L.; Mehling, T.; Jakobtorweihen, S.; Smirnova, I. Prediction of Micelle/Water and Liposome/Water Partition Coefficients Based on Molecular Dynamics Simulations, COSMO-RS, and COSMOmic. *Langmuir* **2013**, *29* (11), 3527–3537.
- (30) Mokrushina, L.; Buggert, M.; Smirnova, I.; Arlt, W.; Schomäcker, R. COSMO-RS and UNIFAC in Prediction of Micelle/Water Partition Coefficients. *Ind. Eng. Chem. Res.* **2007**, *46* (20), 6501–6509.
- (31) Buggert, M.; Cadena, C.; Mokrushina, L.; Smirnova, I.; Maginn, E. J.; Arlt, W. COSMO-RS Calculations of Partition Coefficients: Different Tools for Conformation Search. *Chem. Eng. Technol.* **2009**, *32* (6), 977–986.
- (32) Jakobtorweihen, S.; Ingram, T.; Smirnova, I. Combination of COSMOmic and Molecular Dynamics Simulations for the Calculation of Membrane-Water Partition Coefficients. *J. Comput. Chem.* **2013**, *34* (15), 1332–1340.
- (33) Wang, L. Measurements and Implications of the Membrane Dipole Potential. *Annu. Rev. Biochem.* **2012**, *81* (1), 615–635.
- (34) Endo, S.; Escher, B. I.; Goss, K.-U. Capacities of Membrane Lipids to Accumulate Neutral Organic Chemicals. *Environ. Sci. Technol.* **2011**, *45* (14), 5912–5921.
- (35) Sarmoria, C.; Puvvada, S.; Blankschtein, D. Prediction of Critical Micelle Concentrations of Nonideal Binary Surfactant Mixtures. *Langmuir* **1992**, *8* (11), 2690–2697.
- (36) Persson, C. M.; Kjellin, U. R. M.; Eriksson, J. C. Surface Pressure Effect of Poly(Ethylene Oxide) and Sugar Headgroups in Liquid-Expanded Monolayers. *Langmuir* **2003**, *19* (20), 8152–8160.
- (37) Puvvada, S.; Blankschtein, D. Molecular-Thermodynamic Approach to Predict Micellization, Phase Behavior and Phase Separation of Micellar Solutions. I. Application to Nonionic Surfactants. *J. Chem. Phys.* **1990**, *92* (6), 3710–3724.
- (38) Matsumura, S.; Imai, K.; Yoshikawa, S.; Kawada, K.; Uchibor, T. Surface Activities, Biodegradability and Antimicrobial Properties of n-Alkyl Glucosides, Mannosides and Galactosides. *J. Am. Oil Chem. Soc.* **1990**, *67* (12), 996–1001.
- (39) Okawauchi, M.; Hagio, M.; Ikawa, Y.; Sugihara, G.; Murata, Y.; Tanaka, M. A Light-Scattering Study of Temperature Effect on Micelle Formation of N-Alkanoyl-N-Methylglucamines in Aqueous Solution. *Bull. Chem. Soc. Jpn.* **1987**, *60* (8), 2719–2725.
- (40) Benazzouz, A.; Moity, L.; Pierlot, C.; Molinier, V.; Aubry, J.-M. Hansen Approach versus COSMO-RS for Predicting the Solubility of an Organic UV Filter in Cosmetic Solvents. *Colloids Surf. Physicochem. Eng. Asp.* **2014**, *458*, 101–109.
- (41) Zhang, Y.; Feller, S. E.; Brooks, B. R.; Pastor, R. W. Computer Simulation of Liquid/Liquid Interfaces. I. Theory and Application to Octane/Water. *J. Chem. Phys.* **1995**, *103* (23), 10252–10266.
- (42) *COSMOtherm C3.0, Release 17.04*; COSMOlogic GmbH & Co. KG; <http://www.cosmologic.de>; Leverkusen, Germany, 2017.
- (43) *CRC Handbook of Tables for Applied Engineering Science*, 2nd ed., 14. pr.; Bolz, R. E., Tuve, G. L., Eds.; CRC Press: Boca Raton, Florida, 2000.

