First-principles prediction of critical micellar concentrations for ionic and nonionic surfactants

Turchi, M.; Karcz, A. P.; Andersson, M.P.

Published in:
Journal of Colloid and Interface Science

Link to article, DOI:
10.1016/j.jcis.2021.08.044

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
First-principles prediction of critical micellar concentrations for ionic and nonionic surfactants

M. Turchi, A.P. Karcz, M.P. Andersson *

Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

GRAPHICAL ABSTRACT

Article history:
Received 9 April 2021
Revised 6 August 2021
Accepted 7 August 2021
Available online 10 August 2021

Keywords:
Critical micelle concentration
Interfacial tension
Nonionic
Ionic
Zwitterionic
COSMO-RS
Surfactant
Density functional theory
Phase separation

ABSTRACT

The concentration of surfactant in solution for which micelles start to form, also known as critical micelle concentration is a key property in formulation design. The critical micelle concentration can be determined experimentally with a tensiometer by measuring the surface tension of a concentration series. In analogy with experiments, in-silico predictions can be achieved through interfacial tension calculations.

We present a newly developed method, which employs first principles-based interfacial tension calculations rooted in COSMO-RS theory, for the prediction of the critical micelle concentration of a set of non-ionic, cationic, anionic, and zwitterionic surfactants in water. Our approach consists of a combination of two prediction strategies for modelling two different phenomena involving the removal of the surfactant hydrophobic tail from contact with water. The two strategies are based on regular micelle formation and thermodynamic phase separation of the surfactant from water and both are required to take into account a wide range of polarity in the hydrophilic headgroup.

Our method yields accurate predictions for the critical micellar concentration, within one log unit from experiments, for a wide range of surfactant types and introduces possibilities for first-principles based prediction of formulation properties for more complex compositions.

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Surfactants are commonplace in society and are the key components in formulated products, such as detergents, cosmetics, fibers, oilfield chemicals, pharmaceuticals and food [1–5]. One property specific to each surfactant is the critical micellar concentration (CMC), which is the concentration of surfactant in water, above which micelles spontaneously start to form. Any nominal increase in surfactant concentration beyond the CMC only leads to the formation of more micelles, while the real surfactant concentration in water remains at the CMC. Although the CMC itself is not causally
related to full formulation properties, any prediction of complex systems requires at least the CMC to be correctly predicted by the method. For modelers, besides the qualitative use for formulations, the CMC thus serves as a benchmark for predictions.

Attempts to predict the CMC have been published using different modelling techniques, such as quantitative structure–activity relationships (QSAR) models [6–10], group-contribution (GC) methods [11,12], dissipative particle dynamics (DPD) [13–16], and COSMO-RS-based models [17,18]. QSAR model parametrization relies substantially on experimental data regression of previously measured CMC. This feature makes the models suitable for predicting the CMC of surfactants, which have similar chemical compositions to the ones used in the regression. As an example, Huibers et al. needed to use different parameters and equations to develop distinct QSAR models that could reliably predict the CMC of nonionic surfactants belonging to the alkyl ethoxylates, CEₘ, family [7], and anionic surfactants of the sulfate class [6]. GC methods, similar to QSAR, highly rely on the parametrization of the interaction groups. While GC methods were employed to predict CMC of nonionic surfactants [12,19], to the best of our knowledge no GC was employed to predict CMC for ionic surfactants. This is probably due to the lack of an extensive collection of parametrized headgroups for describing the hydrophilic part of ionic surfactants. DPD is a coarse-grained molecular model that simulates the dynamics of atoms grouped into “beads.” As with any other molecular dynamic model, beads interact with each other by means of a force field. In this regard, an adequate parameterization of the head group is crucial for the prediction of the CMC. DPD was used for the prediction of the CMC of nonionic surfactants [16,20]. Anionic and cationic surfactants were also studied with DPD [13,14], although the parameterization of the bead describing the hydrophobic head group and the associated counterion prevents straightforward employment of this method to a wider range of surfactants. Usually, GC, QSAR and DPD models yield very accurate quantitative predictions, characterized by root-mean-square-error (RMSE) between 0.2 and 0.5 in logarithmic scale. On the other end, because of their highly empirical nature, they cannot be applied to different surfactant types (i.e., nonionic, anionic, cationic, and zwitterionic), using the same parametrization. In other words, the dataset used for the predictions needs to include molecules with similar chemical structures to the ones used for the dataset of the parametrization. Within COSMO-RS methods, COSMOplex [18] has recently emerged as a robust method for the prediction of CMC of nonionic surfactants. The method that combines molecular dynamics (MD) simulation with COSMOMic (MD/COSMOMic) [17] was also employed for the prediction of the CMC of nonionic surfactants belonging to the CEₘ series. So far, their usage is still limited to nonionic surfactants. COSMOplex and MD/COSMOMic yielded a prediction accuracy in predicting the CMC that is generally lower than GC, QSAR and DPD models (RMSE between 0.5 and 0.8 in the log scale), but in line with COSMO-RS models for the prediction of other thermodynamic properties, such as the thermodynamic partition coefficient [21–26]. Despite the lower prediction accuracy, as compared to GC, QSAR and DPD, models based on COSMO-RS have the key advantage of using a universal parametrization in every prediction, thus making them potentially applicable to any surfactant class and type.

Taking advantage of the high flexibility of COSMO-RS based models, in this paper we present a prediction approach that employs interfacial tension calculations based on quantum chemical density functional theory (DFT) and the COSMO-RS [18] implicit solvent model. This approach involves a combination of two prediction strategies for modelling two different phenomena involving the removal of the surfactant hydrophobic tail from contact with water. The first strategy models the formation of ordered micelles via preferential hydrophobic tail solvation for strongly polar hydrophilic head groups, such as ionic surfactants. The second strategy models the critical micellar concentration as the surfactant concentration in a two-phase surfactant/water system in thermodynamic equilibrium, which proves to work well for non-ionic surfactants without particularly polar headgroups. Examples in this study include the well-known alkyl ethoxylates, CEₘ, a few other nonionic surfactants, and several ionic surfactants that are representative of anionic, cationic, and zwitterionic classes (see Table 1 for surfactants used in this study). Key features of our predictive approach are its robustness, computational speed, and general applicability to all the aforementioned surfactant classes. This flexibility, which to the best of our knowledge is not offered by any other CMC prediction model, combined with its computational efficiency, makes the model applicable to a wide range of surfactant systems. The overall predicted CMC is the lowest concentration of the two individual predictions of the computed CMC values from Strategy 1 and Strategy 2.

2. Methods
2.1. Quantum chemical calculations

We performed the quantum chemical DFT calculations with geometry optimization in order to compute the molecular screening charge surface, also known as σ-surface or COSMO surface, within COSMO-RS [27] theory for all the surfactants in the predictions. Water and n-decane were taken from the COSMOTHERM database. The COSMOconf package [28] was used to generate stable conformers for each surfactant. The whole set of generated conformers was used in the COSMO-RS calculations, and for alkyl ethoxylates, the conformers included those with and without internal hydrogen bonding of the terminal OH group to the ether groups.

All the DFT calculations were performed with the Turbomole software, version 7.4.0 [29]. Initially, DFT calculations with geometry optimization were carried out using a Becke-Purdue [30,31] (BP) functional, the TZVP basis set [32], resolution of identity (RI) approximation together with the COSMO [33] implicit solvent model, and an infinite dielectric constant (needed for subsequent COSMO-RS calculations [27]). Additional single point energy calculations were performed on the optimized geometries with the TZVPD basis set, in order to be able to use the “BP_TZVPD_FINE_19” parameterization in the COSMOTHERM software.

2.2. COSMO-RS calculation within COSMOTHERM software

In order to compute the interfacial tension (IFT), we used our developed COSMO-RS-based method [34], which builds upon the flatsurf module of COSMOTHERM [35]. In brief, flatsurf calculates the free energy of transferring a molecule from a bulk phase to a featureless interface. The liquid–liquid IFT method adds a third (virtual) surface phase, which is placed between the two bulk phases. The liquid–liquid interface thus has two interfaces, both of which contribute to the total IFT and therefore allows for the partial implicit solvation in two adjacent phases. In particular, the addition of the surface phase, in the spirit of Guggenheim, allows for treating surface enrichment of surface-active species, like surfactants. All the details about the liquid–liquid IFT can be found in the original paper [34]. Of special interest is that the method requires only neutral species to be present, which means that all ionic surfactants in the study include the counterion specifically bonded to the ionic surfactant [36]. We employed the script by L.V. Nikolajsen for the calculations [37].

Using the liquid–liquid IFT method, two different strategies were utilized for the prediction of the CMC of anionic, cationic,
zwitterionic and non-ionic surfactants at $T = 298$ K, relevant to the majority of experimental data collected. Fig. 1 presents a schematic description of the two strategies.

**Strategy 1:** The first strategy considers ternary systems composed of the surfactant, $n$-decane (i.e., the oil component), and water separated in two phases. Phase 1 consists of a mixture of water and surfactant, while Phase 2 consists of pure $n$-decane. The strategy is intended for surfactants with strongly polar hydrophilic headgroups, where regular micelles are expected to form, in which surfactant tails are preferentially solvated by other surfactant hydrophobic tails with little interaction between tail and headgroup (Fig. 1). It should be noted that the oil chosen in the calculations should also match the surfactant tail chemistry. All of the surfactants with a strongly polar headgroup in our study are dominated by alkane chains, so any alkane with a low solubility in water may be used. Strategy 1 employs an iterative approach involving a series of IFT calculations, whereby the molar fraction of the surfactant in the aqueous phase is effectively increased until the value of the IFT reaches 0 mN/m. At this point (red square in Fig. 2), the molar fraction of the surfactant in the water-rich phase is assumed to yield the CMC. Fig. 2 describes how the iterative approach works, and the surfactant SDS is taken as a reference, while profiles for other surfactants are given in Supplementary Figures S1-S5. From Fig. 2, we see how, at low molar fractions, the IFT profile is characterized by positive values around 50 mN/m, which is consistent with a water/alkane interface without surfactant. In the positive range of the IFT profile, where the interfacial tension between the oil and water-rich phase is higher than 0, the micellization process is energetically unfavorable; therefore, micelle formation does not occur. At the CMC point, IFT becomes zero, and the creation of a new surface area between the phases can be done without any penalty to free energy. Thus, micellization may begin. In the calculations beyond the CMC point (red square), the IFT reaches negative values, which seems unphysical. However, this is based on a situation where computational equilibrium without dynamics is taken into account. This should be interpreted as follows: For a situation with negative IFT, the system would gain free energy by creating more surface area between the water and the oil.

### Table 1

<table>
<thead>
<tr>
<th>Surfactant Molecular formula</th>
<th>Surfactant type</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sodium dodecylsulfate</td>
<td>CH$_3$(CH$<em>2$)$</em>{11}$SO$_4$Na</td>
<td>anionic SDS</td>
</tr>
<tr>
<td>sodium 2-dodecylsulfate (one ethoxyl unit)</td>
<td>CH$_3$(CH$<em>2$)$</em>{10}$(CH$_2$OCH$_2$CH$<em>3$)$</em>{1}$SO$_4$Na</td>
<td>anionic SLES$_{EO1}$</td>
</tr>
<tr>
<td>sodium 2-[2-(dodecylsulfate)ethoxy]ethoxy]ethylsulfate (three ethoxyl units)</td>
<td>CH$_3$(CH$<em>2$)$</em>{10}$(CH$_2$OCH$_2$CH$<em>3$)$</em>{3}$SO$_4$Na</td>
<td>anionic SLES$_{EO3}$</td>
</tr>
<tr>
<td>potassium octa decanoate</td>
<td>C$<em>{18}$H$</em>{35}$O$_2$K</td>
<td>anionic PS</td>
</tr>
<tr>
<td>sodium octyl 1-sulfonate</td>
<td>C$<em>{18}$H$</em>{35}$NaO$_3$S</td>
<td>anionic SDS</td>
</tr>
<tr>
<td>hexadecyltrimethylammonium bromide</td>
<td>C$<em>{16}$H$</em>{33}$BrN</td>
<td>cationic CTAB</td>
</tr>
<tr>
<td>hexadecyltrimethylammonium chloride</td>
<td>C$<em>{16}$H$</em>{33}$ClN</td>
<td>cationic CTAC</td>
</tr>
<tr>
<td>1-hexadecylpyridin-1-ium chloride</td>
<td>C$<em>{16}$H$</em>{33}$NCl</td>
<td>cationic CPC</td>
</tr>
<tr>
<td>2-[3-[(dodecanoylamino)propyl-dimethylazaniumyl]acetate (no counterions)]</td>
<td>C$<em>{18}$H$</em>{36}$NaO$_3$</td>
<td>zwitterionic CAPB (no counterions)</td>
</tr>
<tr>
<td>2-[3-[(dodecanoylamino)propyl-dimethylazaniumyl]acetate (with counterions)]</td>
<td>C$<em>{18}$H$</em>{36}$NaCl</td>
<td>zwitterionic CAPB (with counterions)</td>
</tr>
<tr>
<td>octadecyl benzyl dimethylammonium chloride</td>
<td>C$<em>{18}$H$</em>{36}$NCl</td>
<td>cationic ODBMC</td>
</tr>
<tr>
<td>2-[2-(hexoxyethoxy)ethoxy]ethanol</td>
<td>C$<em>{12}$H$</em>{26}$O$_4$</td>
<td>nonionic C$_6$E$_3$</td>
</tr>
<tr>
<td>2-octoxylethanol</td>
<td>C$<em>{10}$H$</em>{20}$O$_4$</td>
<td>nonionic C$_8$E$_1$</td>
</tr>
<tr>
<td>2-[2-(2-octoxyethoxy)ethoxy]ethanol</td>
<td>C$<em>{14}$H$</em>{28}$O$_6$</td>
<td>nonionic C$_{8}E_2$</td>
</tr>
<tr>
<td>2-[2-[2-(2-descxyethoxy)ethoxy]ethoxy]ethoxy]ethoxy]ethoxy]ethanol</td>
<td>C$<em>{14}$H$</em>{28}$O$_6$</td>
<td>nonionic C$_{12}E_2$</td>
</tr>
<tr>
<td>(2R,3S,5S,6R)-2-[2-(hydroxymethyl)-6-octoxyxane-3,4,5-triol</td>
<td>C$<em>{14}$H$</em>{28}$O$_6$</td>
<td>nonionic C$_{10}E_2$</td>
</tr>
<tr>
<td>(2R,3S,5S,6R)-2-dodecylethanol</td>
<td>C$<em>{14}$H$</em>{28}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>2-oxoxygen-propane-1,3-diol</td>
<td>C$<em>{11}$H$</em>{22}$O$_4$</td>
<td>nonionic C$_{8}E_1$</td>
</tr>
<tr>
<td>N,N-dimethylpropylamine-1-amine oxide</td>
<td>C$<em>{12}$H$</em>{24}$N$_2$O</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>buty 2-ethyl hexa ethylene glycol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>undecyl N,N-diethanolamine</td>
<td>C$<em>{11}$H$</em>{23}$C(O)N(C$_2$H$_4$OH)$_2$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>2-[2-[2-[2-(hexadoxyethoxy)ethoxy]ethoxy]ethoxy]ethoxy]ethoxy]ethanol</td>
<td>C$<em>{30}$H$</em>{62}$O$_8$</td>
<td>nonionic C$_{16}E_2$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
<tr>
<td>(2R,3R,5R,6R)-2-dodecylethanol</td>
<td>C$<em>{12}$H$</em>{24}$O$_6$</td>
<td>nonionic C$_{12}E_1$</td>
</tr>
</tbody>
</table>
oil (surfactant tails). This can only occur through the creation of more micelles. In the process of creating micelles, and in order to take advantage of the negative IFT, the mass balance in the system must be upheld, and any micelles created will deplete surfactants from the aqueous phase. This process will thus continue to generate micelles, reducing the aqueous surfactant concentration until it reaches the CMC, whereby no new surface area is created or destroyed.

Strategy 2: The second strategy considers binary systems composed of two pure phases, Phase 1 constituted of pure surfactant and Phase 2 constituted of pure water. In this approach, the thermodynamic two-phase equilibrium is calculated using COSMO-RS, and the IFT between the two phases is also calculated as in Fig. 1. The molar fraction of the surfactant in the water-rich phase is assumed to yield the CMC. As a consequence of this approach, the value of the IFT may not always be zero, and the implications discussed later.

While it will be made clear later in the results and discussion, we recommend that in order to predict the CMC for a general surfactant, both strategies should be employed, and the lowest predicted value of the two should be chosen as the predicted CMC.

3. Results and discussion

The two prediction strategies were tested for a set of anionic, cationic, zwitterionic, and nonionic surfactants. The predicted values are reported together with experimental data in Table 2.

From Table 2 it is evident that Strategy 1 performed best in predicting the CMC for ionic surfactants, whereas Strategy 2 generally yielded more accurate predictions for nonionic surfactants. We ascribe this marked difference in the prediction accuracy of the two strategies depending on the type of surfactant to the polarity of the head groups. In particular, the higher polarity of the head groups of ionic surfactants makes the distinction between the hydrophobic carbon chain and the hydrophilic head more pronounced, and it disfavors a “normal” phase separation into two bulk phases because of the very unfavorable interactions between the hydrophilic and hydrophobic parts of the surfactant. Consequently, the n-decane oil phase represents better the hydrophobic environment of the ionic surfactant micelle chains, and the phenomenon that best describes ordered micelle formation by ionic surfactants is their preferential tail interaction via the creation of the virtual oil phase in Strategy 1. In the case of nonionic surfactants, the lower polarity of the head groups leads to more balanced interactions between the hydrophilic and hydrophobic parts of the surfactants, and, consequently, the micellization process is better described by a phase separation in the simple binary surfactant/water system.

As the polarity of the hydrophilic headgroup of the surfactant in principle dictates which strategy should be utilized to predict the CMC, we investigated the polarity of different surfactants by looking at the screening charge density (σ) profiles from COSMO-RS calculations. Fig. 3 shows the σ-profiles of some representative surfactants selected as references.

As shown in Fig. 3a, all the ionic surfactants exhibit a primary peak around zero, which is from the nonpolar carbon tail. A secondary peak at σ value slightly lower than 0.02 e/Å² describes the negatively charged Cl⁻ counterion of CTAC, whereas a weakly pronounced peak at ~0.01 e/Å² is due to the positively charged headgroup. For SDS, a secondary peak relative to the negatively charged head group is located around 0.015 e/Å², and another secondary peak due the sodium counterion is found around ~0.025 e/Å². Comparing the CAPB profiles, we note that the CAPB without explicit counterions shows the σ-profile extending beyond the value of 0.02 e/Å² (area marked with a red arrow), whereas the σ-profile for CAPB with counterions shows a peak at around 0.02 e/Å² followed by a sudden drop beyond. CAPB with counterions, similar to SDS, exhibits another secondary peak in the vicinity of ~0.025 e/Å², which comes from the added sodium counterion.

Similarly, for the ionic surfactants in Fig. 3b, because of the hydrophobic carbon tail, all the nonionic surfactants exhibit a primary peak at around 0.00 e/Å². For the C8E5 surfactant, we observe a secondary peak due to the polarity of the hydrophilic head group. C8E5 is taken to be representative for all the other nonpolar surfactants, which indeed show similar σ-profiles, except C8C2NO. With regard to C8C2NO, we observe that, like CAPB without counterions, the σ-profiles extends beyond the value of 0.02 e/Å² (area marked with a red arrow).

As we will discuss in the next sections, we found the optimal range of σ-profile for reliable CMC prediction accuracy to be between the most negative value of the σ-profile of the sodium cation (red line) and the most positive value of the σ-profile of the chloride anion (green line). Consequently, for surfactants characterized by σ-profiles, which extend beyond this range (i.e., CAPB without counterions and C8C2NO), the prediction accuracy can be expected to be lower.

3.1. Ionic surfactants

For ionic surfactants, we comment exclusively on predicted values of Strategy 1, as the predictions from Strategy 2 were consistently unreliable indicators of the CMC.

Anionic surfactants: With regard to anionic surfactants (i.e., SDS, SLES-EO1, SLES-EO3, PS and SOS), the CMC predicted for SDS agrees with experimentation. Concerning SLES-EO1 and SLES-EO3 surfactants, Aoudia et al. [39] reported values for log(CMC) very close to each other, ~3.10 and ~3.09. Contrarily, Vleugels et al. [46] observed a marked decrease in CMC with an increase in the number of EO groups for SLES surfactants having the same carbon chain length. In particular, log(CMC) values for SLES-EO2 and SLES-EO4 were ~3.00 and ~3.33 respectively. Our predictions yielded values of ~2.40 for SLES-EO1 and ~3.42 for SLES-EO3, which suggest a decrease in the CMC value for an increase of the EO units and, therefore, are in agreement with the trend observed by Vleugels et al. [46]. Furthermore, in the case of SLES-EO3, Strategy 2 yields a predicted value of ~2.75, thus achieving a prediction accuracy comparable to Strategy 1. For SLES-EO3, the predicted value of
Strategy 1 is 0.33 lower than experimentation, and Strategy 2 is 0.34 higher than the experimental value. This could indicate that CMC for SLES surfactants with even larger numbers of EO units could potentially be better predicted with Strategy 2. With regard to PS surfactant, the predicted value of $/C0^3.35$ is in excellent agreement with the experimental value of $/C0^3.30$. The hydrophobic tail of PS is constituted of 18 carbons (C18) instead of the C12 structures of SDS and SLES. This demonstrates that our method is able to capture the hydrophobic effect of different carbon chain lengths. Notably, the experimental measurement of the CMC for PS was conducted at 60 °C and consequently we performed the IFT calculations at the same temperature. The excellent prediction accuracy for PS at 60 °C shows the robustness of the model in performing accurate predictions at temperatures that are considerably higher than ambient conditions. Finally, the SOS surfactant was tested due to its high CMC experimental value (-0.81 in the log scale). Again, Strategy 1 confirmed to be the most accurate, yielding a predicted value of 0.12 which is 0.94 higher than the experimental value. The difference with the experimental value is higher than the rest of anionic surfactants, but still within 1 unit in the log scale.

<table>
<thead>
<tr>
<th>Surfactant</th>
<th>Exp.</th>
<th>Predicted (Strategy 1)</th>
<th>Predicted (Strategy 2)</th>
<th>IFT (Strategy 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS</td>
<td>-2.07 (a)</td>
<td>-2.07</td>
<td>-1.15</td>
<td>-</td>
</tr>
<tr>
<td>SLES_EO1</td>
<td>-3.10 (b)</td>
<td>-2.4</td>
<td>-1.58</td>
<td>-</td>
</tr>
<tr>
<td>SLES_EO3</td>
<td>-3.09 (b)</td>
<td>-3.42</td>
<td>-2.75</td>
<td>-</td>
</tr>
<tr>
<td>PS</td>
<td>-3.30 (c)</td>
<td>-3.35</td>
<td>-1.46</td>
<td>-</td>
</tr>
<tr>
<td>SOS</td>
<td>-0.81 (a)</td>
<td>0.12</td>
<td>0.67</td>
<td>-</td>
</tr>
<tr>
<td>CTAB</td>
<td>-3.04 (c)</td>
<td>-3.40</td>
<td>-1.38</td>
<td>-</td>
</tr>
<tr>
<td>CTAC</td>
<td>-2.96 (d)</td>
<td>-3.38</td>
<td>-0.94</td>
<td>-</td>
</tr>
<tr>
<td>CPC</td>
<td>-3.00 (e)</td>
<td>-3.45</td>
<td>-1.02</td>
<td>-</td>
</tr>
<tr>
<td>CAPB (no counterions)</td>
<td>-3.55 (e)</td>
<td>-0.88</td>
<td>0.46</td>
<td>-</td>
</tr>
<tr>
<td>CAPB (with counterions)</td>
<td>-2.15 (f)</td>
<td>-1.7</td>
<td>-1.82</td>
<td>-3.38</td>
</tr>
<tr>
<td>OBDMC</td>
<td>-5.15 (e)</td>
<td>-5.22</td>
<td>-5.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>C4E5</td>
<td>-1 (a)</td>
<td>0.61</td>
<td>-1.9</td>
<td>8.08</td>
</tr>
<tr>
<td>C4E1</td>
<td>-2.31 (a)</td>
<td>0.7</td>
<td>-2.85</td>
<td>10.65</td>
</tr>
<tr>
<td>C4E3</td>
<td>-2.12 (a)</td>
<td>0.56</td>
<td>-2.32</td>
<td>-0.07</td>
</tr>
<tr>
<td>C10E2</td>
<td>-3.22 (a)</td>
<td>0.51</td>
<td>-3.33</td>
<td>1.29</td>
</tr>
<tr>
<td>C12E5</td>
<td>-4.06 (a)</td>
<td>0.37</td>
<td>-4.05</td>
<td>0.37</td>
</tr>
<tr>
<td>C6BG1</td>
<td>-1.6 (a)</td>
<td>0.61</td>
<td>-1.1</td>
<td>-1.73</td>
</tr>
<tr>
<td>C10BG1</td>
<td>-3.72 (a)</td>
<td>0.46</td>
<td>-2.81</td>
<td>-3.76</td>
</tr>
<tr>
<td>C6C6H12E2</td>
<td>-3.88 (a)</td>
<td>0.53</td>
<td>-3.66</td>
<td>9.49</td>
</tr>
<tr>
<td>C6BGLYE</td>
<td>-2.24 (a)</td>
<td>0.69</td>
<td>-2.11</td>
<td>3.97</td>
</tr>
<tr>
<td>C6C5NO</td>
<td>-1.27 (a)</td>
<td>-0.30</td>
<td>0.68</td>
<td>-</td>
</tr>
<tr>
<td>IC6E6</td>
<td>-1.00 (e)</td>
<td>0.22</td>
<td>-1.99</td>
<td>5.43</td>
</tr>
<tr>
<td>C11CONEO</td>
<td>-3.58 (b)</td>
<td>0.55</td>
<td>-2.25</td>
<td>1.64</td>
</tr>
<tr>
<td>C6C7E7</td>
<td>-5.76 (a)</td>
<td>0.45</td>
<td>-6.43</td>
<td>3.21</td>
</tr>
</tbody>
</table>

Table 2
Comparison between the predicted values for the log(CMC) and IFT for the two approaches and the experimental value of log(CMC). The colors are used to guide the reader. Green implies good prediction accuracy (<1 log unit deviation), orange indicated inaccurate predictions (>1 log unit deviation), and blue refers to IFT values for Strategy 2 that are larger than 3 mN/m. The bold values of the CMC predictions in the table are the best outcomes.

(a) Ref. [38], (b) Ref. [39], (c) Ref. [40], (d) Ref. [41], (e) Ref. [42], (f) Ref. [43], (g) Ref. [44], (h) Ref. [45].
Cationic surfactants: In the case of cationic surfactants, the influence of the counterion on the CMC value was investigated considering CTAC, CTAB, CPC and OBDMC surfactants. Our predictions yielded values of $-3.4$ and $-3.38$ for CTAB and CTAC, respectively, which are in good agreement with experimental data of Wei et al. [40] for CTAB ($-3.04$) and Li et al. [41] for CTAC ($-2.96$). Although the difference between the log(CMC) of CTAB and CTAC are negligible for both experimental and predicted values, the trend in predictions is consistent with experiments, yielding lower values of log(CMC) for CTAB as compared to CTAC. Concerning CPC, the predicted value of $-3.45$ is in good agreement with the experimental value of $-3.00$. Notably, the CPC hydrophilic head is constituted by the heterocyclic pyridine acid group, which is substantially different from the trimethylammonium group of CTAB and CTAC. This further validates the consistency of the method in predicting the CMC of cationic surfactants with different hydrophilic headgroups. Finally, the OBDMC surfactant was tested due to its low experimental CMC value ($-5.15$ in the log scale). The predicted CMC of $-5.22$ agreed very well with experiments, which further confirmed the reliability of the model in predicting also quite low values of CMC.

Zwitterionic surfactants: The surfactant CAPB was chosen as a reference example for zwitterionic surfactants. Staszak et al. [42] measured the CMC for CAPB at different NaCl concentrations, reporting values for log(CMC) ranging from $-2.15$ (2 M NaCl water solution) to $-3.55$ (pure water). We performed two different predictions: one considering the pure CAPB molecule and a second one adding two counterions in proximity of the charged groups (i.e., $\text{Cl}^-$ anion next to the amine cation and $\text{Na}^+$ cation next to the carboxyl acid anion). The addition of the counterions enabled us to simulate the presence of NaCl in solution as well as to be consistent with predictions of the other ionic surfactants (including counterions bound to charged sites at the surfactant). Predictions of CAPB with counterions yielded a value of log(CMC) equal to $-1.7$, which is in good agreement with the experimental value of $-2.15$ (CAPB in 2 M water solution). However, this is considerably far from the experimental value of $-3.55$ in pure water. Predictions performed with pure CAPB molecules yielded a value of log(CMC) of $-0.88$, which is drastically higher than the experimental values. Notably, for the zwitterionic surfactant with counterions, the prediction of Strategy 2 was slightly more accurate than Strategy 1, yielding a value of the log(CMC) equal to $-1.82$. Overall, for zwitterionic surfactants we found that the addition of counterions was crucial to perform accurate predictions and that the performance of Strategy 1 and Strategy 2 were comparable. In particular, the addition of the sodium counterion entails the decrease of the
screening charge density (σ) of the negatively charged carboxylate to a value lower than 0.02 e/Å², which we anticipate will impact prediction accuracy.

Finally, the effect of the parametrization was briefly investigated by comparing the predictions of log(CMC) for CTAB, SDS, and all the nonionic surfactant using the BP_TZVP_19 parametrization to our chosen TZVPD_FINE_19. The TZVP parametrization yielded lower predictions accuracy than TZVPD-FINE, yet was still within 1.5 log units. Predicted values for the TZVP parametrization are reported in Table S1 of the Supplementary material.

3.2. Nonionic surfactants

Linear and branched polyoxyethylenated (POE) surfactants: C6E3, C6E1, C6E2, C10E3, C12E6, C16E7 and IC6E6 (the branched POE) are polyoxyethylated surfactants with different methylene/ethylene oxide (CH2/EO) ratios and, therefore, different hydrophobic/hydrophilic balances. For all the POE surfactants, Strategy 2 predicts CMC with good accuracy, always within one unit in the logarithmic scale. Notably the values of the log(CMC) span between −1 of C6E3 and −5.76 of C6E1. Therefore our model was tested in the whole spectrum of CMC of most common surfactants. Concerning the IFT predictions, all the values except for C6E3, IC6E6 and C6E1 are fairly close to zero. For C6E3 (10.65 mN/m), IC6E6 (5.43 mN/m) and for C6E1 (8.08 mN/m), the IFT is significantly above zero. Rosen et al. [45] observed that when the hydrophobic tail of the surfactant is too short there is little tendency for micellization to occur because the hydrophobic group does not manage to efficiently distort the solvent structure. This is indeed the case for C6E3 and IC6E6 surfactants, which have the same CMC experimental value and for which analogous considerations can be made. In particular, Telgmann et al. [47] observed that for C6E1, oligomeric structures rather than proper micelles occur in correspondence of the CMC. The presence of unstructured oligomers instead of micelles can explain why our predicted IFT was significantly higher than zero. In principle, our method predicts that no C6E3 micelles would form at the “CMC,” which is consistent with experiments. For C6E1, the single EO group of the hydrophilic group might be too small to interact efficiently with the water phase, thus, causing phase separation (instead of a micellization phenomenon) between water and the C6E1 molecules.

Glucoside hydrophilic groups: For C8B1 and C12B1, predicted values of the CMC are in good agreement with experimental values, again within one log unit. The values of the IFT are negative for both surfactants. This substantial reduction of the IFT can be explained with water molecules also hydrating part of the methylene units. This was indeed reported for C8BG1 micelles by Paula et al. [48], who observed that water hydrates between two and three methylene groups of the surfactant hydrophilic tail.

C9C2NO: The predicted CMC value of this surfactant is the least accurate of the dataset and the only one differing by more than 1 log unit from the experimental value. Our hypothesis, considering that COSMO-RS theory is known to reliably describe the amide groups, is that the model tends to overestimate the contribution due to the double bond (C=C) in the hydrophobic carbon chain. Rosen et al. [45] reported that when a carbon–carbon double bond is present in the hydrophobic chain, the CMC is generally higher than the one of the corresponding saturated compound. Our model predicts a CMC that is 1.2 log units higher than the experimental value. Extended studies on more similar surfactants would be necessary to confirm or deny our hypothesis, but it is outside the scope of this paper.

The focus of this paper is the prediction of CMC, which is substantiated by experimental studies. The qualitative interpretation of the calculations for CMC predictions with Strategy 2, where the IFT is significantly greater than zero, would be that the overall surface area between the surfactant and water phases needs to be smaller, so as not to pay too large of an energy penalty for forming the contact area. This would mean that only for surfactant/water systems where IFT between the two phases is below or near zero, small micelles would form. For systems where IFT is significantly larger than zero, much larger micelles would be formed or other phenomena would be observed, such as oligomerization [47] or nanoparticle formation consisting of multiple micelles to reduce the surface area [36]. We are not aware of experimental measurements of drop or micelle sizes for all these systems, but present the implications of our predictions here nonetheless.

3.3. Overall prediction accuracy

The best predictions for all surfactants (ionic and nonionic) are compared to experimental data in Fig. 4.

All CMC predictions are rather accurate, and the errors are less than one logarithmic unit away from the experimental value, with the exception of the C11CONEO surfactant. The overall root-mean-square-error (RMSE) is 0.62, which is in the prediction range of alternate prediction methods based on COSMO-RS theory (which only included nonionic surfactants) [49]. Looking at Table 2, we can infer how Strategy 1 is generally better for ionic surfactants and Strategy 2 is generally better for nonionic surfactants. This is in line with the underlying physical phenomena that are the foundations of the two strategies, and they consider the influence of the polarity of the hydrophilic headgroup. Because Strategy 2 is very fast (typically, a few seconds), and significantly faster than Strategy 1 (in the order of few minutes), we suggest that it be always calculated, regardless of surfactant. Furthermore, the most accurate value for CMC prediction, when considering both strategies, is always the lowest value of the two. This could be interpreted as a competition between mechanisms for micellization, such that whichever one occurs for the lowest surfactant concentration first will determine the CMC. If Strategy 1 provides the lowest CMC, ordered micelles are predicted to form, while if Strategy 2 provides the lowest CMC, less ordered micelles are predicted to form. }


matically, it would therefore be advantageous to consistently use both strategies and choose whichever provides the lowest CMC concentration as the best prediction. This can certainly be relevant for borderline cases (i.e., CAPB and C₆C₆NO), where it is not obvious beforehand which strategy may yield the most accurate prediction. We would like to point out that unlike the more heavily parameterized computational methods, the underlying COSMO-RS based method we employ in this paper does not rely on any experimental data on interfacial tension or CMC – only on other thermodynamic properties.

We showed how our method is capable of predicting CMC also for other temperatures. This was expected based on the fact that the underlying COSMO-RS method [27] and IFT prediction method [34] gave good results also for temperatures that deviate substantially from 298 K.

4. Conclusion

We have presented a robust and fast method for predicting the critical micellar concentration (CMC) of surfactants in water based on density functional theory and the COSMO-RS implicit solvent model. CMC for ionic, zwitterionic, and nonionic surfactants are all predicted accurately, within one log unit in concentration of experimental observations, with the only exception of C₁₁CONEO, for which the predicted value of the CMC is 1.2 log units higher than the experimental value.

For nonionic surfactants, the prediction accuracy of our method, based on RMSE values, is slightly lower although comparable to QSAR [7], GC [12,19], and DPD models [16,20], whereas it is in line with the performance of COSMO-RS based models [17,18]. The highest deviation from the experimental value occurred for the C₁₁CONEO surfactant. For ionic surfactants, our method yields predicted values which are in line with previous QSAR [6] and DPD studies for SDS [13,14] and CTAB [14].

Apart from the largest deviation of the study, C₁₁CONEO, our model yielded less accurate predictions for the C₆E₃ and IC₆E₆ surfactants, which are characterized by a high experimental value of the CMC (−1.00 in the logarithmic scale). Nevertheless, for these two surfactants (characterized by a rather short hydrophobic tail), also, the accuracy of the experimental data could be questioned. This is suggested by the low tendency for micellization to occur when the short hydrophobic group does not manage to efficiently distort the solvent structure, as reported by Rosen et al. [45] and by the observation of oligomeric structures, rather than proper micelles corresponding with the CMC for C₆E₃ reported by Telgmann et al. [47].

The major drawback of previous prediction methods, except for COSMOplex, is their lack of immediate transferability to surfactant systems for which they have not been specifically parametrized. In particular, QSAR models need distinct regression to experimental values for surfactants with dissimilar chemical composition. GC models require parametrization of specific groups for treating diverse surfactants and are still limited to nonionic surfactants. DPD needs a non-trivial parametrization of the force field for bead interactions, whenever a surfactant with a new head group is considered. The MD/COSMOmic approach, similar to DPD, relies on

![Fig. 4. Optimal set of predicted values of the log(CMC) (Strategy1 for ionic surfactants and Strategy2 for nonionic surfactants) are compared to experimental values of log (CMC).](image)
force field parametrization at the MD level. COSMOplex, although in principle having the same flexibility of our method, is significantly more computationally demanding and still restricted to nonionic surfactants.

The generality of our method, combined with the applicability of COSMO-RS to multi-component systems, opens possibilities to predict formulation properties beyond CMC for more complex systems, including mixed surfactant systems as well as water–oil–surfactant systems.

CRediT authorship contribution statement

M. Turchi: Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing - original draft, Writing - review & editing. A.P. Karcz: Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

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

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcis.2021.08.044.

References


